High-frequency trading and networked markets

Federico Musciotto[a], Jyrki Piilo[b], and Rosario N. Mantegna[a,c,d,1]

[a]Dipartimento di Fisica e Chimica, Università degli Studi di Palermo, I-90128 Palermo, Italy; [b]Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014 Turun Yliopisto, Finland; [c]Complexity Science Hub Vienna, A1080 Vienna, Austria; and [d]Department of Computer Science, University College London, London WC1E 6BT, United Kingdom

Edited by Simon Asher Levin, Princeton University, Princeton, NJ, and approved March 5, 2021 (received for review August 14, 2020)

Financial markets have undergone a deep reorganization during the last 20 y. A mixture of technological innovation and regulatory constraints has promoted the diffusion of market fragmentation and high-frequency trading. The new stock market has changed the traditional ecology of market participants and market professionals, and financial markets have evolved into complex sociotechnical institutions characterized by a great heterogeneity in the time scales of market members' interactions that cover more than eight orders of magnitude. We analyze three different datasets for two highly studied market venues recorded in 2004 to 2006, 2010 to 2011, and 2018. Using methods of complex network theory, we show that transactions between specific couples of market members are systematically and persistently overexpressed or underexpressed. Contemporary stock markets are therefore networked markets where liquidity provision of market members has statistically detectable preferences or avoidances with respect to some market members over time with a degree of persistence that can cover several months. We show a sizable increase in both the number and persistence of networked relationships between market members in most recent years and how technological and regulatory innovations affect the networked nature of the markets. Our study also shows that the portfolio of strategic trading decisions of high-frequency traders has evolved over the years, adding to the liquidity provision other market activities that consume market liquidity.

complex networks | financial markets | high-frequency trading | statistically validated networks

The last 20 y have seen deep changes in the way financial markets operate (1). The adoption of the regulation about the national market system ("NMS") for equity securities (2) from the Securities and Exchange Commission of the United States and a similar adoption subsequently taken by the European Securities and Market Authority have affected the structure and practice of trading of equity securities in US, European, and several other markets worldwide. The most evident change has been the proliferation of market venues in a given country or in a group of countries, usually addressed as fragmentation of markets. A related change has been the specialization of a number of market participants in high-frequency traders (HFTs). HFTs are professional traders able to use high speed in the generation, routing, and execution of orders (3). The amount of transactions performed by HFTs is today estimated to be around 50% in most markets (4). Typical response time of these traders to a market state or information can be as fast as a few microseconds. This exceptional time performance is often achieved by colocating technical infrastructures of HFTs near the computer infrastructure of large market venues. The influence of regulatory changes and technological innovations have changed financial markets into complex sociotechnical institutions (5–9).

There is no shared view about how changes occurring in markets have modified the basis of financial asset trading. One view is that the presence of HFTs makes markets more efficient by decreasing the transaction cost per unit of transaction and by facilitating price discovery (10–13). Another view is that HFTs provide liquidity only under normal market conditions whereas their trading is not guaranteed in exceptional market states, making the markets more fragile and prone to flash or microflash crashes (14). There is also empirical evidence that HFTs competition is deteriorating liquidity provision (15) and that the interaction of HFTs with orders of large institutions is performed in a strategic way (16).

The technical and regulatory changes observed in markets in the last 20 y are not minor but rather they deeply affect the strategic behavior of market participants (5). An example of this huge impact concerns market making and liquidity provision. Market making is the trading activity providing liquidity to the market (i.e., the trading activity allowing one to find quickly a counterpart for a transaction in a market). In the past market making was done by institutionalized figures (called specialists at the New York Stock Exchange and jobbers at the London Stock Exchange [LSE]) that paid fees to the market and had privileges and obligations for their role. Today, in most settings, market making is not institutionalized and it is freely strategically performed by specialized market participants. Technical innovations (e.g., the use of computer algorithms and the fast access and process of market quotes) and regulatory changes (e.g., market fragmentation and changes in the information production and dissemination) provide a changing environment deeply affecting the profile and ecology of different classes of investors and the way they perform their strategic choices. We believe this extraordinary transition into a changing financial market setting is a clear case study of evolution and adaptation of the ecology of market participants to new states of the financial world (17–19).

Significance

During the last two decades, technological innovation and regulatory requirements have deeply changed the way financial markets work. Today, financial markets are characterized by the presence of high-frequency traders (able to perform financial transactions at a submillisecond time scale) and market fragmentation. Using methods of complex networks, we show that some market participants (specifically so-called market members) preferentially interact with or avoid other market members persistently over a time scale extending up to several months. By investigating two financial venues at three different periods of time from two different decades, we show that the persistent networked nature of today's markets is most pronounced since the diffusion of high-frequency trading and market fragmentation.

Author contributions: F.M., J.P., and R.N.M. designed research; F.M. analyzed data; F.M. performed numerical simulations; F.M., J.P., and R.N.M. analyzed results; and F.M. and R.N.M. wrote the paper with contributions from J.P.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: rosario.mantegna@unipa.it.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2015573118/-/DCSupplemental.

Published June 25, 2021.

Recently the empirical characterization of the trading decisions of investors (20) has made it possible to detect an ecology of investors (21, 22). In the present study, we analyze two different venues of stock markets in three different periods of time at the level of market members, i.e., at the level of those companies that are acting for their proprietary trading or are acting as brokers or dealers for the trading of customers. Our aim is to investigate whether the liquidity provision of the most active market members (notably the HFTs) presents a networked structure (i.e., it is provided under a framework of strategic decisions). With the term "networked structure" we mean that market members trading in a venue of a fragmented market in the presence of HFTs might establish statistically validated preferential or avoided trading relationships with specific market members. By using tools of complex networks (23), we detect a networked structure among market members and we document that this networked structure has evolved over the past 20 y from a poor and dynamically changing to a richer and dynamically more persistent network structure highlighting an ever-increasing strategic provision of liquidity.

Financial transactions occurring in financial markets can be described in terms of trading networks (23). Examples are trading networks occurring in the Interbank market (24) and in equity markets (25). Specifically, we investigate the degree and persistence of pairwise trading relations between market members of two European stock market venues in 2004 to 2006, 2010 to 2011, and 2018. The first one is the electronic order book of the LSE during the time period 2004 to 2006 and the second one is the Stockholm venue of the Nasdaq OMX market during the years 2010 to 2011 and 2018. Nasdaq OMX is a subsidiary of Nasdaq, Inc. operating in European Nordic countries. The first set of data refers to a period when the high-frequency trading practice and its infrastructures were still developing and HFTs were expanding their share of trading. During those years, market fragmentation was still pretty limited due to the fact that the regulation about the competition between market venues was not yet issued by the European Securities and Market Authority. Market fragmentation and the diffusion of high-frequency trading expands in Europe starting from 2009 (15, 26) and the presence of market members performing HFTs became widespread in European financial markets at the same time. Since 2010, the Nasdaq OMX has provided to its market members the same technology used in the main US Nasdaq venue, including INET platform and colocation services, ensuring HFTs access to the order book within microseconds or less. Therefore, our three sets of data cover one period when the practice, regulatory framework, and technology of HFTs were still developing and two periods characterized by key regulatory and technical changes.

Our analysis and statistical characterization of trading networks show that financial markets have continuously evolved into ever more complex sociotechnical systems (5–9) with a persistent networked structure which is present in the interaction of the different types of market members, making the liquidity provision a sophisticated strategic activity. In other words, since the first decade of this century markets have changed deeply and the ecological profile of market members has experienced a profound mutation. With our results we show that network-based studies are able to characterize the ecological fingerprint of some market members acting in this highly competitive sociotechnical system.

Data and Methodology

We investigate the LSE and the Nasdaq OMX Stockholm venues. The LSE dataset we investigate is a special release of the Rebuild Order Book. Specifically, we investigate 20 highly liquid stocks, traded during the 2004 to 2006 y. The Rebuild Order Book contains information on price, volume, submission, transaction, or cancelation time for each order. In addition to its standard information, our special release of the database contains the anonymized identity of the market member who placed the order. Here we investigate all orders submitted through the electronic order book. With this unique dataset we track the trading activity of anonymized market members of the LSE in the investigated years. For the Nasdaq OMX Stockholm venue, it was possible to obtain historical information about the identity of market members performing each transaction. Having this type of information is a rather unique characteristic of this venue and for this reason the venue has been investigated in several papers dealing with HFTs (15, 16, 27–31). The dataset for the Stockholm venue of the Nasdaq OMX is the ITCH INET data stream maintained and provided by the Nasdaq OMX (32). For this venue, we investigate the 20 most liquid stocks, traded during the February 2010 to December 2011 and the January 2018 to December 2018 time periods. The ITCH data stream captures the complete dynamics occurring on the electronic order book through a set of messages specifically formatted for each kind of action.

For both periods, we are able to reconstruct the full dynamics of the book in terms of submission (except for a small set of transactions of the order of less than 5% of all transactions), modification, and deletion of limit orders, together with the occurrence of transactions as a result of market orders. Matching the data with ticker data, we are then able to assign the identity of market members involved in each transaction and uncover their role both as aggressor or counterpart and as buyer or seller. A market participant is acting as aggressor when the participant submits a market order, i.e., an order of buy or sell to be instantly executed at the best offer or ask. The other market member participating in the transaction is called the counterpart. Counterparties submit quotes, called limit orders, to the order book signaling the willingness to buy or sell a given number of shares of an asset at a given price. Traditionally, market members submitting limit orders were considered as market participants providing liquidity to the market whereas market members submitting market orders were considered as participants taking liquidity from the market. The market evolution of the last 20 y has made this distinction less clear and today it is rather complex to discriminate between market members making or taking liquidity (5, 33, 34). While quotes, i.e., limit orders, of ITCH are anonymous, for the time period February 2010 to December 2011 the ITCH carried the identity of both market members involved in a transaction. This policy changed in March 2014 when it introduced the possibility for market members to be anonymous when trading large cap and OMXS30 shares.

In the period 2004 to 2006, LSE data have a temporal resolution of 1 s for the order book dynamics. The time resolution improves to 1 ms for the Nasdaq data of the Stockholm venue in 2010 to 2011 and further reduces to 1 μs for the same venue in 2018. These temporal resolutions are indicative of the fastest time scales present at the venues at the time of data recording although orders faster than the highest resolution might have been executed in all cases.

For each venue, each time period, and each stock we build two directed trading networks. One network is a buyer–seller (BS) directed network where we put an arc between two market members when they perform a mutual transaction within a given period. The period we choose in this study is a calendar month. For each calendar month, we determine the buyer–seller trading network of market members trading a specific stock. The weight of each arc is given by the number of transactions observed between the two market members when the first one is acting as a buyer and the second one is acting as a seller. In addition to the buyer–seller trading network, we also consider the aggressor–counterpart (AC) directed network to investigate the liquidity provision relationships between each pair of market members.

The first test we perform is designed to characterize patterns of transactions occurring between market members with respect to their categorization in terms of their ability to act as HFTs. Categorizing market members as HFTs directly from data by defining typical intervals for indicators such that the time to fill of orders or the short-term inventory balance is today poorly effective due to continuous innovation and specialization of HFTs. The time to fill is the time elapsed between the submission of the limit order (sent from the counterpart) and the submission of the matching market order (sent from the aggressor). In *SI Appendix*, section 1, we show deciles of the time to fill of active market members trading three representative stocks during a specific month. The deciles of time to fill are in the range from 1 s or less to thousands of seconds for December 2006, in the range from 1 ms or less to thousands of seconds for December 2011 and in the range from 1 μs or less to thousands of seconds for December 2018. The presence of very low values for the deciles of time to fill does not directly imply that a market member is a HFT. In fact, HFTs can take the role of both aggressor and counterpart in a transaction and therefore observing a low value of deciles provides only evidence that at least one of the two market members entering in the transaction is a HFT.

For this reason, as in other studies (16, 31), we use public information present in news, web pages, and technical literature to identify market members that are acting as HFTs at the Stockholm Stock Exchange. The list of market members we categorize as primarily HFTs is given in *SI Appendix*, section 2. We do not perform the present test on the LSE venue because we know only the anonymized identity of market members for this venue.

For the Nasdaq OMX Stockholm venue, each month, and each stock, we compute the number of transactions occurring between the two categories of market members. They perform four types of transactions in the AC (BS) network: 1) type 1, both aggressor (buyer) and counterpart (seller) are HFTs; 2) type 2, the aggressor (buyer) is a HFT and the counterpart (seller) is not; 3) type 3, the aggressor (buyer) is not a HFT and the counterpart (seller) is; 4) type 4, neither the aggressor (buyer) nor the counterpart (seller) is a HFT.

In our test, for each type of transaction we compare the empirically observed number with those computed in random simulations of the so-called configuration model of the same empirical network (35). A random realization of the configuration model is obtained starting from the empirical network of individual market members by randomizing the links of the network while preserving the number of transactions performed by each market member both as aggressor (buyer) and as counterpart (seller). Thus, random realizations of our null model preserve the number of transactions at the individual market member level while destroying preferential or avoided interactions between pairs of market members. Multiple links and self-loops are allowed in random realizations as market members perform several transactions with the same counterpart and with themselves. We iterate the randomizing procedure at least 100 times for each network and we compute for each of the four types of transactions a z score and a P value. In *SI Appendix*, section 3 we show a numerical example of the probability density function of the number of transactions observed in random realizations together with a Gaussian fitting. Due to the fact that we are performing a multiple-hypothesis test comparison, we evaluate the null hypothesis of random pairing of the four types of transactions by adopting the control for the false discovery rate (36). Our test tells us that transactions of types 2 and 3, i.e., transaction between a HFT and a non-HFT, are more common than expected in a random pairing whereas transactions of types 1 and 4, i.e., transaction between two HFTs or between two non-HFTs, are less common than expected.

By performing our test for each month and each stock, we verify that the rejection rate of the null hypothesis of the configuration model is higher than or equal to 65% for the AC network and higher than or equal to 75% for the BS network in 2010. The rejection rate becomes 100% for all months and all stocks in 2011 and 2018.

Having detected statistical evidence of the networked structure of transactions occurring between pairs of market members acting at the Nasdaq OMX we now investigate the nature and persistence of these overexpressed and underexpressed interactions in a more systematic way. Specifically, to capture the networked structure arising from the specialized interaction of market members, we use the methodology of statistically validated networks (37, 38). Statistically validated networks are those filtered networks where pairs of nodes of the network are selected when a given null hypothesis is rejected for them (23). Examples are statistically validated networks in bipartite systems (37, 38) and networks obtained by filtering the so-called backbone of a network (39–41). In our study, we use statistically validated networks as defined in ref. 38 because we are able to detect both overexpressed and underexpressed pair trading relationships with this method. This property of the methodology allows us to highlight both preferred and avoided trading relationships.

Specifically, for each pair of market members we compare the number of transactions occurring between them with the expected number they would get if each market member randomly draws its counterpart from an urn. Thus, the strength of our approach is that we validate the number of interactions occurring between market members with respect to a null model that takes into account the heterogeneity of their activity. Summary statistics of the heterogeneity of the number of transactions of market members are shown in *SI Appendix*, section 4. In *SI Appendix*, section 5 we list the international securities identification number (ISIN) of the investigated stocks together with the average number of transactions observed for each venue and each time period.

Our statistical test works as follows: For each pair of market members (labeled here as A and B) we count the transactions N^i_{AB} occurring between them when they are trading a stock of ISIN code i. We take into account the directionality of the couple (i.e., we consider distinct the aggressor [buyer] role from the counterpart [seller] role) and we apply the test on both the aggressor $\rightarrow$ counterpart (buyer $\rightarrow$ seller) and counterpart $\rightarrow$ aggressor (seller $\rightarrow$ buyer) directions. Then we count the total number of transactions for stock i that we label as N^i, the number of transactions N^i_A in which A is acting as an aggressor (buyer) on i, and the number of transactions in which B is acting as a counterpart (seller) on i (labeled as N^i_B). We then compute the probability of finding a number of transactions larger than or equal to N_{AB} as the result of two random draws, one for market member A acting as an aggressor (buyer) and one for market member B acting as a counterpart (seller). This probability is well approximated by (37, 38)

$$p_1(N^i_{A,B}) = 1 - \sum_{X=0}^{N^i_{A,B}-1} H(X|N^i, N^i_A, N^i_B), \qquad \textbf{[1]}$$

where H is the hypergeometric distribution. We can use the probability $p_1(N^i_{A,B})$ as a P value to test the null hypothesis of random pairing of the directed pair of market members (A, B). We apply the control for the false discovery rate as a correction method for multiple-hypothesis testing (36). The couples that reject the null hypothesis are showing an overexpressed number of transactions, signaling that they are interacting more than expected under our null model. By looking at the other tail of the hypergeometric distribution we are also able to compute the probability of finding a number of transaction smaller than or equal to N^i_{AB},

$$p_2(N^i_{A,B}) = \sum_{X=0}^{N^i_{A,B}} H(X|N^i, N^i_A, N^i_B). \qquad [2]$$

The couples of market members that reject the null hypothesis with this P value show an underexpressed number of transactions, indicating that they are interacting less than expected by the null hypothesis of random pairing.

Networked Markets

In Fig. 1*A* we show the density of statistically validated links, i.e., the ratio between the number of statistically validated links divided by the total number of possible links in the network. The investigated network is the AC network but a similar behavior is also observed for the BS network. Fig. 1 *A*, *Top* refers to overexpressed links whereas Fig. 1 *A*, *Bottom* refers to underexpressed links. The left symbols show results for the LSE in 2004 to 2006 (292 market members were active in this market in the considered period), the center symbols refer to Nasdaq Nordic Stockholm Exchange (XSTO) in 2010 to 2011 (i.e., the Stockholm venue of the Nasdaq OMX with 98 market members active in the considered period), and the right symbols refer to XSTO in 2018 (with 75 market members active). Trading networks were investigated on a monthly basis for the 20 most liquid stocks traded in the market. Fig. 1 *A* shows a progressive increase of the density of overexpressed and underexpressed links as a function of time.

For the investigated venues, time periods, and the 20 most liquid stocks, we compute the time to fill detected in each transaction to track the boost in speed introduced by the diffusion and development of high-frequency trading. The time to fill is a basic indicator providing information about the typical time needed such that a limit order ends up in a transaction. In Fig. 1*B* we show the probability density function of the time to fill for all transactions performed for the 20 most liquid stocks traded at the LSE in 2004 to 2006 (Fig. 1 *B*, *Top Left*), XSTO in 2010 to 2011 (Fig. 1 *B*, *Middle Left*), and XSTO in 2018 (Fig. 1 *B*, *Bottom Left*). In Fig. 1 *B*, *Right* we show the corresponding cumulative distribution functions. Fig. 1*B* shows that markets have significantly decreased the time to fill over the years and that transacted limit orders present a time to fill shorter than 1 s in about 25% of the cases in 2018.

The average value of the time to fill observed for transactions occurring in a given time period is an aggregated information. It is possible to disaggregate this information by considering the time to fill for transactions occurring between each couple of market members where member A is acting as an aggressor and member B is acting as a counterpart for each stock and each trading month. Median times to fill for representative stocks traded at the LSE in 2006, XSTO in 2011, and XSTO 2018 are shown in *SI Appendix*, section 6. When this disaggregation is performed, a large heterogeneity is observed among different pairs of market members. In fact, the range in Fig. 1*B* covers up to 10 orders of magnitude.

We discover that the increase in trading speed, that is documented by the probability density functions and cumulative density functions in Fig. 1*B*, is associated with an increase of the density of over- and underexpressed relationships between market members (Fig. 1*C*). Indeed, scatter plots show the relationship between the median of the time to fill and the average density of overexpressions (underexpressions) for each month and each stock. Different colors highlight different venues in different periods of time. There is a clear relationship between speed and number of over- and underexpressions. The faster the system is, the larger the number of detected over- and underexpressions. Part of the effect we observe might be due to the increased number of transactions observed over the years. In fact, the number of transactions is affecting the power of our statistical test (in the present case the number of transactions of a stock occurred in 1 mo). By analyzing cases where the power of the statistical test is the same, we have verified that our results are only marginally affected by the different power of the test. An example of our control of the power of the test is given in *SI Appendix*, section 7.

Our statistical test is performed by estimating overexpressions and underexpressions in terms of the number of transactions observed between two market members. In general, different transactions have associated different volumes and therefore in principle an overexpressed number of transactions do not necessarily imply an overexpressed amount of exchanged volume. We have therefore tested that overexpressed and underexpressed numbers of transactions also imply overexpressed and underexpressed exchanged volume for the large majority of couples of market members. Details of our test are shown in *SI Appendix*, section 8. For overexpressed links, we have also observed that the average fraction of exchanged volume normalized per

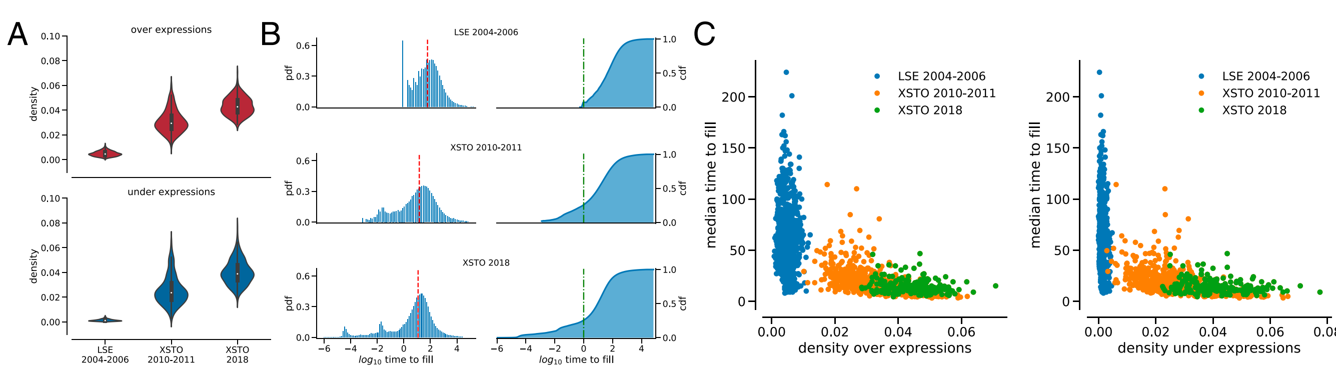

Fig. 1. (*A*) Violin plots of the link density of the overexpressed (*Top*) and underexpressed (*Bottom*) as a function of the set of data for the aggressor → counterpart network. The investigated sets concern monthly overexpressed and underexpressed networks for the 20 most liquid stocks traded at the LSE in 2004 to 2006, at the XSTO in 2010 to 2011, and at the XSTO in 2018. (*B*) Probability density function (*Left*) and cumulative density function (*Right*) of the time to fill for all transactions performed for the 20 most liquid stocks traded at the LSE in 2004 to 2006 (*Top Left* and *Top Right*), XSTO in 2010 to 2011 (*Middle Left* and *Middle Right*), and XSTO in 2018 (*Bottom Left* and *Bottom Right*). In panels of probability density function (*Left*) the red vertical line indicates the median value of time to fill. In panels of cumulative distribution function (*Right*) the green vertical line indicates the fraction of transactions with time to fill shorter than 1 s. (*C*) Scatter plots of the median of the time to fill (in seconds) and the average density of overexpressions (*Left*) or underexpressions (*Right*) on the investigated stocks. Each symbol refers to a specific month and a specific stock. Different colors of the symbol are defined by the venues and periods of time. Blue symbols refer to LSE in 2004 to 2006, orange symbols to XSTO in 2010 to 2011, and green symbols to XSTO in 2018.

number of overexpressed links has steadily increased during the time periods of 2011 and 2018. Details of this investigation are given in *SI Appendix*, section 11.

In the following, to minimize the role of the different power of the test for different stocks and months, we present a detailed analysis of three stocks that approximately traded the same number of transactions in the two venues in 2006, 2011, and 2018. The selected stocks are British Petroleum (BP) for LSE in 2006, Electrolux B (ELUX) for XSTO in 2011, and SKF B (SKF) for XSTO in 2018. BP was the most liquid stock at the LSE in 2006. ELUX and SKF are chosen because they have an average number of monthly transactions similar to those of BP in 2006. In fact, these three stocks had an average number of monthly transactions equal to 103,841, 103,192, and 99,488 in 2006, 2011, and 2018, respectively.

In Table 1 we show the number of directed couples of market members that are validated for a number of months ranging from 1 to 12 (column 1) in the AC and BS networks. We note that at the LSE in 2006 the overexpressed (OE) and underexpressed (UE) observed relationships were poorly persistent over time. In fact, only 2 OE couples were observed for 6 or more months in the BS network and only 4 OE couples in the AC network. UE couples were a bit more persistent, totaling 12 UE couples in the BS network and 13 UE couples in the AC network. It is worth noting that only one couple of market members turns out to be validated in all 12 mo. This occurs for a UE couple in the AC network. The pattern is different for Electrolux in 2011 and SKF in 2018. For Electrolux in 2011, a number of directed pairs of market members present a persistence for 6 or more months. Several pairs present a persistence for 9 or more months and the BS and AC network presents 3 UE and 5 UE pairs, respectively, that are observed for all months of the year. The persistence of the networked relations further increases in 2018. In fact, for SKF in 2018 we detect many OE and UE relationships that are observed for 6 or more and 9 or more months and more than 10 OE or UE pairs are observed for all months of the year.

In Fig. 2*A* we show six panels describing the time persistence of overexpression or underexpression of each aggressor → counterpart couple observed in the AC statistically validated networks of SKF in the 12 mo of 2018. The statistical validations are grouped in three sets. Fig. 2 *A*, *Left* includes couples of market members that are not HFTs, Fig. 2 *A*, *Center* includes couples of one HFT interacting with a non-HFT, and Fig. 2 *A*, *Right* includes couples of two HFTs. For a given month, the presence of a vertical segment in the panels indicates overexpression or underexpression of the number of transactions of a market member couple (labeled along the x axis by a numerical index). Fig. 2 *A*, *Center* shows that the number of trades between HFTs and non-HFTs is highly and persistently overexpressed for a large set of couples whereas the number of transactions between couples of non-HFTs (Fig. 2 *A*, *Bottom Left*) or couples of HFTs (Fig. 2 *A*, *Bottom Right*) is persistently underexpressed. The persistence of underexpression over time is particularly striking for HFT couples. This result shows the strategic ability of HFTs in avoiding transactions among them despite the anonymity of the fast submitted (and canceled) quotes.

Fig. 2*B* shows the most persistent overexpressed aggressor → counterpart pairs detected in 9 or more months of 2018. Fig. 2*C* is a similar plot showing only the underexpressed aggressor → counterpart pairs detected in 9 or more months of 2018. We label the nodes with the tick symbol of the market member. The color of the node is purple when the market member is categorized as performing high-frequency trading. The market members that are not categorized as high-frequency traders are shown as orange nodes. Fig. 2 *B* and *C* shows visually the networked aspect of the AC trading networks. When we look at the underexpressions observed during 9 or more months (Fig. 2*C*), we see that market members described as HFTs (purple nodes) present underexpressed interactions between couples of them. In parallel to this strategic avoidance, we also detect that some market members not performing high-frequency trading also present persistent underexpressed interactions between them (orange nodes). The same type of networked structure is observed in the investigated most liquid stocks both in the AC and in the BS networks. Figures analogous to Fig. 2*A* computed for stocks BP in 2006 and ELUX in 2011 can be found in *SI Appendix*, section 9. For the sake of completeness, in *SI Appendix*, section 10 we report some basic metrics of transaction networks of market members for the same stocks and time periods.

The behavior observed for BP, ELUX, and SKF is not a special one but rather it is representative of most liquid stocks. In Fig. 3 we show the mean value (averaged over the 20 most liquid stocks of each venue) of the Jaccard index computed between statistically validated networks detected at month i and at month $i+g$ of the investigated venue and period. Given a stock k, we consider the statistically validated links of over- and underexpressed relationships at months i and $i+g$. For illustrative purposes, let us consider overexpressed relationships. We label the set of overexpressions for stock k at months i and $i+g$ as $s_o^k(i)$ and

Table 1. Number of directed pairs of market members that are validated in *k* months (column 1)

No.	BP 2006				ELUX 2011				SKF 2018			
of months	BS OE	BS UE	AC OE	AC UE	BS OE	BS UE	AC OE	AC UE	BS OE	BS UE	AC OE	AC UE
1	633	106	496	73	649	423	521	315	332	198	257	176
2	92	16	84	21	173	169	125	110	82	101	81	77
3	25	8	11	10	65	64	63	69	52	54	53	42
4	8	11	3	10	41	40	33	33	34	39	23	35
5	1	5	9	2	35	35	21	21	22	29	19	29
6	1	2	1	5	14	15	20	13	17	19	18	24
7	0	3	0	1	16	8	18	14	14	23	24	7
8	1	2	1	1	16	10	12	6	17	16	13	12
9	0	3	1	1	9	4	9	2	9	11	12	10
10	0	1	1	2	8	8	6	5	15	13	12	10
11	0	1	0	2	4	1	5	3	19	12	10	4
12	0	0	0	1	0	3	0	5	16	10	15	14

The three stocks analyzed are BP, ELUX, and SKF. The years of the analysis are 2006, 2011, and 2018, respectively. The numbers of market members trading each stock in each venue and year were 203, 78, and 64, respectively.

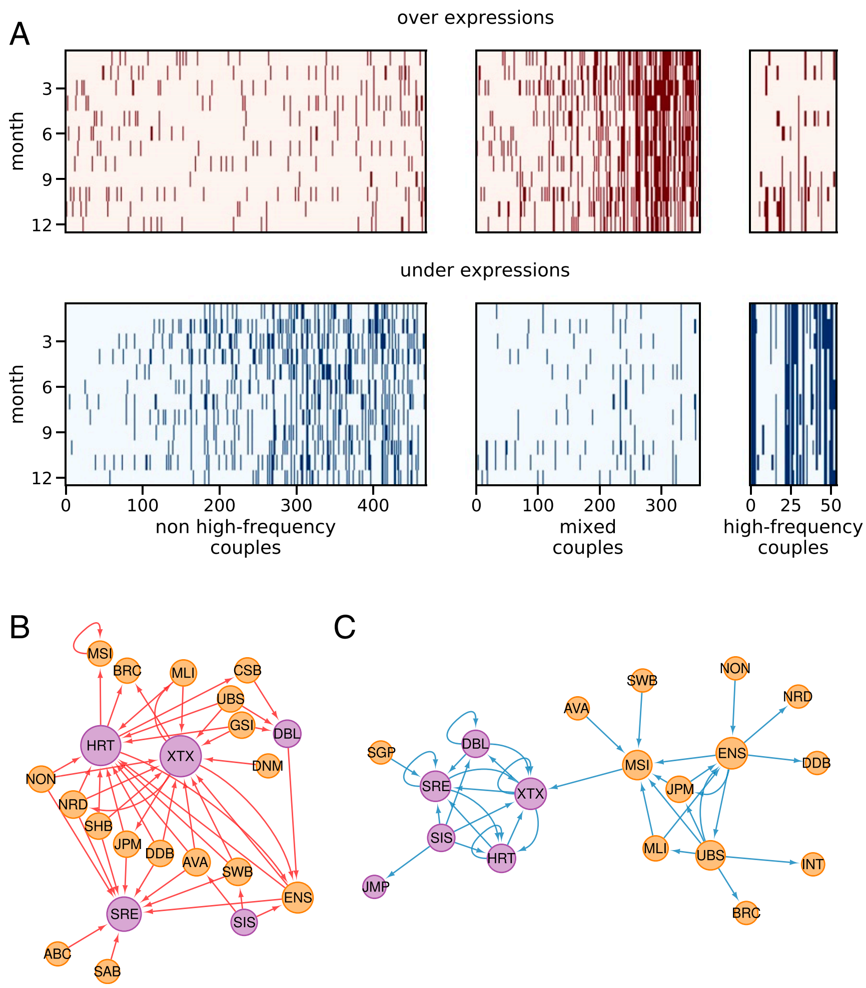

Fig. 2. (*A*) Time persistence of overexpression or underexpression of each aggressor $\rightarrow$ counterpart couple observed in the AC statistically validated networks of SKF in the 12 mo of 2018 (vertical scale of the panels). The statistical validations are grouped in three sets. *Left* sets include couples of market members that are not HFTs, *Center* sets include couples of market members composed of one HFT interacting with a non-HFT, and *Right* sets include couples of two HFTs. For *Top* panels, the presence of a vertical segment indicates overexpression of the number of transactions of the market member couple (labeled by a numerical index) whereas in *Bottom* panels underexpressions are highlighted. *B* shows the overexpressed (red arcs) aggressor $\rightarrow$ counterpart couples detected in 9 or more months of 2018. *C* shows the underexpressed (blue arcs) aggressor $\rightarrow$ counterpart pairs detected in 9 or more months of 2018. In *B* and *C*, we label nodes with the tick symbol of the market member and we use a purple color when the market member is described is a HFT. The market members that are not explicitly described as HFTs are shown as orange nodes. We use trading ID abbreviations to market members' names, which are the same as used by Nasdaq Nordic (see http://www.nasdaqomxnordic.com/membership-list).

$s_o^k(i+g)$, respectively. For stock k the Jaccard index is therefore computed as

$$J_o^k(i,i+g) = |s_o^k(i) \cap s_o^k(i+g)|/|s_o^k(i) \cup s_o^k(i+g)|, \qquad [3]$$

where the symbol $|s|$ indicates the number of elements of the set s. The Jaccard index goes from 0 (no overlap of links between the two networks) to 1 (perfect overlap of all links). An analogous definition can be written for underexpressed statistically validated networks. For each pair of months, the Jaccard index is computed for each stock of the venue and the results obtained are averaged over the 20 stocks of each venue and for all possible month gaps g. Fig. 3 shows the average degree of persistence of overexpressions (Fig. 3, *Left*) and underexpressions (Fig. 3, *Right*) for the AC networks for the LSE venue in 2006 (blue line), the XSTO venue in 2011 (orange line), and the XSTO venue in 2018 (green line). The average persistence is quite limited for the LSE in 2006 (especially for overexpressed links), significantly pronounced for XSTO in 2011, and still more pronounced for XSTO in 2018. A plot of the matrix of the average Jaccard index for each pair of months is shown in *SI Appendix*, section 9.

The increase in the degree of persistence in 2011 and 2018 is evident both in terms of the degree of persistence and in the temporal extension of it. A similar pattern is observed for underexpressions. It is worth noting that the relatively high value of the average Jaccard index observed for underexpressions in LSE 2006 could be due to the fact that statistically validated networks of underexpressed links for this venue are typically composed of few links and Jaccard indexes between networks with few links have associated large digitization noise. The persistence behavior is also quite evident when investigated in a single stock. In *SI Appendix*, section 9 we show the plot of matrices of the Jaccard index for each pair of months for representative stocks BP, ELUX, and SKF and for most traded stocks BP, VOLVO, and H&M. We interpret the observation of these high levels of persistence as a strong empirical evidence of the fact that equity markets have evolved toward a networked state. In the new networked state, some market members preferentially interact with or avoid other market members over periods of time covering up to several months.

Our results show that market venues have evolved from a state where overexpressed and underexpressed market member

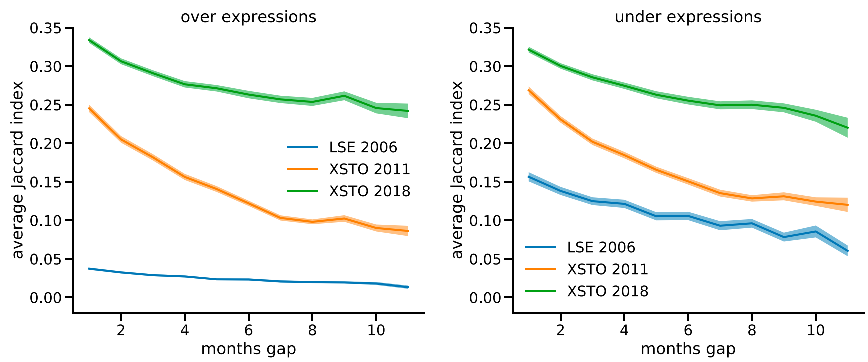

Fig. 3. Average Jaccard index between AC statistically validated networks computed for each month gap. The average is computed on the values obtained for the 20 most liquid stocks traded in each venue for the considered year. *Left* shows overexpressions and *Right* shows underexpressions. Blue line refers to the LSE venue in 2006, orange line refers to the XSTO venue in 2011, and green line to the XSTO venue in 2018. The color band around each line indicates the average value plus and minus the standard deviation of the mean. The average persistence is quite limited for LSE in 2006, significantly pronounced and persistent for XSTO in 2011, and further more pronounced and persistent in XSTO in 2018.

interactions were low in relative number and poorly persistent to a progressive increase of the number and persistency of overexpressed and underexpressed interactions over the years (Figs. 2 and 3 and Table 1). What does this mean for the liquidity provision of market members acting as market makers? To answer this question we investigate how the HFT market members interact with the remaining ones when they act as an aggressor or a counterpart. We first focus on the asymmetry of interaction when a market member acts as aggressor or a counterpart. In *SI Appendix*, Fig. 21 we show the scatter plot of the median time to fill for all pairs of market members trading both as aggressor and as counterpart a given stock during a specific month. The results shown in *SI Appendix* refer to BP, ELUX, VOLVO, SKF, and HM. Our results show that the pattern of pair interactions has dramatically increased its complexity starting from the diffusion of HFT and market fragmentation (i.e., after 2009). In fact, scatter plots in *SI Appendix*, Fig. 21 show a very rich presence of asymmetric interactions between a large number of pairs of market members. For Nasdaq data we can investigate the role of HFTs in these asymmetric interactions. Therefore, for XSTO data, we verify that the median time to fill of transactions occurring when a HFT acts as an aggressor on a non-HFT counterpart is on average different from the median time to fill occurring when the aggressor is a non-HFT and the counterpart is a HFT. Moreover, the median time to fill when a HFT is the aggressor is close to the one observed when both market members are non-HFT. The shortest median time to fill is observed for the transaction occurring between pairs of HFTs. We report these measurements for the representative stocks in *SI Appendix*, Tables 11–13.

Our results are compatible with the hypothesis that HFTs perform transactions with a portfolio of strategies that can be roughly classified as follows: HFTs 1) act as market makers (42), 2) perform proprietary statistical arbitrage, 3) perform predatory trading (43), and 4) execute back run on a detected ongoing large institutional order (16, 44). Asymmetry of interactions was limited at the LSE in 2006, quite pronounced at XSTO in 2011, and very pronounced at XSTO in 2018, testifying to a continuous modification and specialization of the ecological profile of market participants. Our results based on a complex network approach, detailed in *SI Appendix*, section 12, allow us to conclude that, during market activity, HFTs typically switch between acting as market makers providing liquidity to the market and an alternative activity as back runner (and therefore liquidity takers). The ecology of HFT market members is therefore richer than a basic division between liquidity providers and liquidity takers. Several HFT market members are building up a kind of hybrid profile including both aspects and switching between the two on the basis of rapidly accessed and processed market information.

Discussion

Today financial markets are intrinsically different from the financial markets of 20 y ago. The demand for efficiency and competition among market makers requested by regulators and the technological innovations introduced in finance have produced a new market organization where the role of market makers is today taken by specialized companies that are able to access the electronic order book within a few microseconds. In parallel to the continuous decrease of this response time, a proliferation of trading venues has been observed in the largest financial centers of the world, making financial centers highly fragmented (7).

In this work we have analyzed three different market venues with state-of-the-art technological settings at the period of the recording of the data. The first one is the electronic venue of the LSE during 2004 to 2006. At that time, the London equity markets were poorly fragmented and high-frequency traders were still in a phase of rapid technological innovation and growth in terms of share of the transactions executed. Our analysis of this market shows moderate signs of networked structure of the market and it reveals that the persistence of the networked overexpressed or underexpressed relations was poor. For the most traded stock of that period, British Petroleum, we note only one underexpression persistent during all months analyzed in 2006. Most of the overexpressions and underexpressions had a limited persistence over time. The second set of data refers to the Stockholm venue of the Nasdaq OMX during 2010 to 2011. During those years, both fragmentation (12) and high-frequency trading were present in the market. The analysis of these sets of data already shows a pronounced persistence of overexpressed and underexpressed trading relationships. This is quite remarkable when considering that we are able to track the pairwise relationships only in one venue and it suggests that our statistical detection is providing a lower limit of the networked pairwise relationships detected in the full market. We detect the presence of persistent networked relationships between pairs of market members since 2010 and we note that they become enhanced and more persistent in 2018.

This result suggests a few considerations: 1) Today markets are complex sociotechnical institutions operating daily with time scales ranging from microseconds to tens of thousands of seconds (see, for example, the probability density function of time to fill of pairs of market members trading most liquid stocks shown in Fig. 1). This means that time scales covering more than

https://doi.org/10.1073/pnas.2015573118

eight orders of magnitude are present in the market, probably one of the broadest ranges of time scales observed daily in a human institution. 2) Several HFTs are primarily acting as market makers reducing the time to fill of all categories of investors and providing liquidity in normal market conditions. However, we observe robust statistical evidence that the liquidity of HFTs is not provided in a way proportional to the trading interests of investors acting in the market but rather HFTs selectively direct a statistically sizable portion of their provided liquidity to orders submitted by specific market members. This is achieved by processing the accessible information (including book data distributed by different venues) (45–48) in a fast and competitive way and by using this information for designing and performing strategic trading decisions not accessible in practice to a large number of market members.

The networked structure of modern financial markets indicates that competition among market members acting as market makers is suboptimal and that there is room to propose and explore regulatory constraints that could minimize the type and degree of persistence of these networked relationships. Our results show that the debate about the need for a careful investigation of the basic aspect of trading of assets in a contemporary financial market is timely and needed to ensure a fair and efficient trading of financial assets (1, 5, 33, 49, 50).

Data Availability. Data cannot be shared (data are proprietary and can be accessed by contacting the Nasdaq OMX market).

ACKNOWLEDGMENTS. The first part of this work was done in collaboration with Luca Marotta when he was affiliated with University of Palermo. We thank him for discussions and for the work done. F.M. and R.N.M. acknowledge financial support by the Italian Ministry of Education, University and Research, progetto di ricerca di interesse nazionale 2017WZFTZP, Stochastic forecasting in complex systems.

1. M. B. Fox, L. R. Glosten, G. V. Rauterberg, The new stock market: Sense and nonsense. *Duke L. J.* **65**, 191 (2015).
2. Securities_and_Exchange_Commission, Regulation nms. *Fed. Regist.* **70**, 124 (2005).
3. R. S. Miller, G. Shorter, *High Frequency Trading: Overview of Recent Developments* (Congressional Research Service Report, Washington, DC, 2016), **vol. 4**.
4. A. Gerig, High-frequency trading synchronizes prices in financial markets. SSRN [Preprint] (2015). https://ssrn.com/abstract=2173247 (Accessed 1 April 2021)..
5. M. O'Hara, High frequency market microstructure. *J. Financ. Econ.* **116**, 257–270 (2015).
6. N. Johnson *et al.*, Abrupt rise of new machine ecology beyond human response time. *Sci. Rep.* **3**, 2627 (2013).
7. J. P. Pardo-Guerra, *Automating Finance: Infrastructures, Engineers, and the Making of Electronic Markets* (Cambridge University Press, 2019).
8. A. Vespignani, Predicting the behavior of techno-social systems. *Science* **325**, 425–428 (2009).
9. A. Spelta, A. Flori, N. Pecora, S. Buldyrev, F. Pammolli, A behavioral approach to instability pathways in financial markets. *Nat. Commun.* **11**, 1707 (2020).
10. T. Hendershott, C. M. Jones, A. J. Menkveld, Does algorithmic trading improve liquidity? *J. Finance* **66**, 1–33 (2011).
11. R. Litzenberger, J. Castura, R. Gorelick, The impacts of automation and high frequency trading on market quality. *Annu. Rev. Financ. Econ.* **4**, 59–98 (2012).
12. A. J. Menkveld, High frequency trading and the new market makers. *J. Financ. Mark.* **16**, 712–740 (2013).
13. J. Brogaard, T. Hendershott, R. Riordan, High-frequency trading and price discovery. *Rev. Financ. Stud.* **27**, 2267–2306 (2014).
14. A. Kirilenko, A. S. Kyle, M. Samadi, T. Tuzun, The flash crash: High-frequency trading in an electronic market. *J. Finance* **72**, 967–998 (2017).
15. J. Breckenfelder, Competition among high-frequency traders, and market quality. SSRN [Preprint] (2019). https://ssrn.com/abstract=3402867 (Accessed 1 April 2021).
16. V. Van Kervel, A. J. Menkveld, High-frequency trading around large institutional orders. *J. Finance* **74**, 1091–1137 (2019).
17. J. D. Farmer, A. W. Lo, Frontiers of finance: Evolution and efficient markets. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9991–9992 (1999).
18. J. D. Farmer, Market force, ecology and evolution. *Ind. Corp. Change* **11**, 895–953 (2002).
19. A. W. Lo, The adaptive markets hypothesis. *J. Portfolio Manag.* **30**, 15–29 (2004).
20. M. Tumminello, F. Lillo, J. Piilo, R. N. Mantegna, Identification of clusters of investors from their real trading activity in a financial market. *New J. Phys.* **14**, 013041 (2012).
21. F. Musciotto, L. Marotta, J. Piilo, R. N. Mantegna, Long-term ecology of investors in a financial market. *Palgrave Commun.* **4**, 1–12 (2018).
22. A. Flori, F. Pammolli, S. V. Buldyrev, L. Regis, H. E. Stanley, Communities and regularities in the behavior of investment fund managers. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6569–6574 (2019).
23. G. Iori, R. N. Mantegna, "Empirical analyses of networks in finance" in *Handbook of Computational Economics*, C. Hommes, B. Le Baron, Eds. (Elsevier, 2018), vol. 4, pp. 637–685.
24. M. Boss, H. Elsinger, M. Summer, S. Thurner 4, Network topology of the interbank market. *Quant. Finance* **4**, 677–684 (2004).
25. L. Adamic, C. Brunetti, J. H. Harris, A. Kirilenko, Trading networks. *Econom. J.* **20**, S126–S149 (2017).
26. M. Gentile, S. F. Fioravanti, The impact of market fragmentation on European stock exchanges. SSRN [Preprint] (2011). https://ssrn.com/abstract=1997419 (Accessed 1 April 2021).
27. B. Hagströmer, L. Nordén, The diversity of high-frequency traders. *J. Financ. Mark.* **16**, 741–770 (2013).
28. B. Hagströmer, L. Nordén, D. Zhang, How aggressive are high-frequency traders? *Financ. Rev.* **49**, 395–419 (2014).
29. J. Brogaard, B. Hagströmer, L. Nordén, R. Riordan, Trading fast and slow: Colocation and liquidity. *Rev. Financ. Stud.* **28**, 3407–3443 (2015).
30. A. J. Menkveld, M. A. Zoican, Need for speed? Exchange latency and liquidity. *Rev. Financ. Stud.* **30**, 1188–1228 (2017).
31. M. Baron, J. Brogaard, B. Hagströmer, A. Kirilenko, Risk and return in high-frequency trading. *J. Financ. Quant. Anal.* **54**, 993–1024 (2019).
32. *Reference Manual for TotalView - ITCH 4.0* (NASDAQ, 2010).
33. M. Baldauf, J. Mollner, High-frequency trading and market performance. *J. Finance* **75**, 1495–1526 (2020).
34. C. Cao, E. Ghysels, F. Hatheway, Price discovery without trading: Evidence from the NASDAQ preopening. *J. Finance* **55**, 1339–1365 (2000).
35. M. Newman, *Networks: An Introduction* (Oxford University Press, 2010).
36. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
37. M. Tumminello, S. Micciche, F. Lillo, J. Piilo, R. N. Mantegna, Statistically validated networks in bipartite complex systems. *PloS One* **6**, e17994 (2011).
38. V. Hatzopoulos, G. Iori, R. N. Mantegna, S. Miccichè, M. Tumminello, Quantifying preferential trading in the e-mid interbank market. *Quant. Finance* **15**, 693–710 (2015).
39. M. Á. Serrano, M. Boguná, A. Vespignani, Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6483–6488 (2009).
40. F. Radicchi, J. J. Ramasco, S. Fortunato, Information filtering in complex weighted networks. *Phys. Rev.* **83**, 046101 (2011).
41. T. Kobayashi, T. Takaguchi, A. Barrat, The structured backbone of temporal social ties. *Nat. Commun.* **10**, 220 (2019).
42. S. J. Grossman, M. H. Miller, Liquidity and market structure. *J. Finance* **43**, 617–633 (1988).
43. M. K. Brunnermeier, L. H. Pedersen, Predatory trading. *J. Finance* **60**, 1825–1863 (2005).
44. L. Yang, H. Zhu, Back-running: Seeking and hiding fundamental information in order flows. *Rev. Financ. Stud.* **33**, 1484–1533 (2020).
45. T. Foucault, J. Hombert, I. Roşu, News trading and speed. *J. Finance* **71**, 335–382 (2016).
46. C. M. Jones, *Understanding the Market for Us Equity Market Data* (NYSE, 2018).
47. L. R. Glosten, Economics of the stock exchange business: Proprietary market data. SSRN [Preprint] (2020). https://ssrn.com/abstract=3533525 (Accessed 1 April 2021).
48. T. Hendershott, M. Rysman, R. Schwabe, Stock exchanges as platforms for data and trading. SSRN [Preprint] (2020). https://ssrn.com/abstract=3730100 (Accessed 1 April 2021).
49. E. Budish, P. Cramton, J. Shim, The high-frequency trading arms race: Frequent batch auctions as a market design response. *Q. J. Econ.* **130**, 1547–1621 (2015).
50. Y. Ait-Sahalia, M. Saglam, High frequency market making: Implications for liquidity. SSRN [Preprint] (2017). https://ssrn.com/abstract=2908438 (Accessed 1 April 2021).