國立政治大學資訊管理學系

博士學位論文

指導教授:劉文卿博士

高頻交易研究的演化及新興趨勢
Evolution and Emerging Trends in HFT Research

研究生：李耀中

中華民國一Ｏ八年六月

DOI:10.6814/NCCU201901120

謝 辭

感謝指導教授劉文卿老師，引領學生進入高頻交易的領域，在學生進行博士班研究期間，給予我許多指導與充分的自主研究空間。除了研究工作之外，這幾年間無論學生在生活上遇到各種困境與挫折，劉老師總是提點我以智慧及圓融處事，常讓我在沮喪之餘，又重燃信心。感謝口試委員林我聰教授、蔡瑞煌教授、國立臺北大學邱光輝教授、國立台灣科技大學王有禮教授，以專業的論文審核委員角色，對學生的英文論文提供許多指正與寶貴意見，對於本論文趨於完善更有莫大的幫助，謹此致上萬分謝忱。令人不捨的趙玉教授的辭世，學生在此表示最深的懷念，感謝他在期刊文章上提供學生許多值得思考的方向與協助。感謝華夏科技大學資訊管理系周志隆老師，師承臺大管理學院名譽教授江炯聰，周老師不僅引領吾人進入學術殿堂，更在期刊文章及實驗技術上全力支援，才有吾人的丁點成果。學海無涯，今後自當更努力與精進，才不致辜負所有師長們的諄諄教誨。

感謝學長宗憲的及時雨，延續趙玉教授對學生在期刊文章上的協助。感謝博士班同學們的加油與打氣，馨瑩、泉錫、凱康、學隆，永遠記得一起奮戰的你們！

感謝我目前任教學校同事們的關心、加油與打氣。感謝劉院長、王主任、安副主任等鈞長對我進修的支持與鼓勵及校系務的安排，感謝游系祕、林系祕協助排課暨系務。感謝行政單位同事詩惟、昱容協助提供資料分析之實務經驗。

感謝吾人的人生導師郭蕙玉老師，其父親郭金塔為台北帝大<台灣大學前身>醫學博士。緣自於大學時受教的情誼，郭老師充份實踐父親「敬天愛人」的家訓，對週遭人事不論尊卑，在能力範圍內皆能適時適度在關鍵之處予以援助。臺大出版中心曾出版「運籌帷幄-台大經理人」，郭老師在受訪內容中曾期勉學生「千萬要避免知識的勢利」、「取其白、守其黑」，諸如此類教誨及經驗傳承，讓後輩吾等非常受用，也必將繼續傳遞給下一代。

最後還要謝謝我的家人，感謝父母在我求學過程中全心的奉獻，讓我可以兼顧事業、家庭及進修，整合零碎時間進行研究與撰寫論文。感謝內人、岳父及眾親友，二位市長獎女兒從小協助照顧二位雙胞胎弟弟，爸爸以你們四位的懂事為傲。因為有家人們的一路相伴，讓我在家庭生活上無後顧之憂，得以分心處理學業上的瓶頸。耀中衷心的感謝這一切美好，也願我的師長、同窗、同事、家人、朋友、學生，擁有一切的美善與幸福！

DOI:10.6814/NCCU201901120

Abstract

In this research, the evolution and emerging trends of High Frequency Trading (HFT) research is conducted by examining papers published in the Web of Science (WOS) from 1993 to 2017. A total of 241 papers are included, and 1876 keywords from these articles were extracted and analyzed. For tracing the dynamic changes of the HFT Research, the whole 24 years are further separated into three consecutive periods: 1993-2002, 2003-2012, and 2013-2017. The Ucinet is adopted to get keywords network, or knowledge network, to study the relationship of each research theme. NetDraw is applied to visualize network. The social network analysis (SNA) technique is used to reveal patterns and trends in the research by measuring the association strength of terms representative of relevant publications produced in HFT field. Results indicate that HFT research has been strongly influenced by these keywords: “market”, “prices”, “finance”, “liquidity”, “statistics”, “financial markets”, “stock”, “stochastic”, “model” and “trades” as shown in Table 3, which represent some established research themes. These are major focuses and the bridges connecting to other research themes in HFT. The detailed analysis in “Discussions and implications” provides an overview of evolution and emerging trends in HFT Research. It concludes that “market performance” related keywords, which represent some established research themes, have become the major focus in HFT research. It also changes rapidly to embrace new themes. Especially, this research may make contribution to enlarge research method in that there is no SNA research in HFT research before.

Keywords: High Frequency Trading, HFT, Social network analysis, SNA, emerging trends

DOI:10.6814/NCCU201901120

Table of Contents

DOI:10.6814/NCCU201901120

DOI:10.6814/NCCU201901120

List of Tables

DOI:10.6814/NCCU201901120

1. Introduction

1.1 Background

In 1602, Amsterdam Stock Exchange, world's first stock market, formally began trading in securities. In the beginning of 17th century, Rothschilds started using carrier pigeons to arbitrage prices of the same security by relaying information ahead of their competitors. In 1983 – Bloomberg – world's first computerized system to provide real-time price feed and analytics to Wall Street firms – officially kicked off.

As the stock market has become nearly exclusively electronic, advances in computer technology and automated algorithm trading have speeding the transmission and execution of security transaction orders, and thus establishing High Frequency Trading (HFT)[1].

In 1998, SEC authorized computerized high frequency trading, capable of executing trades 1000 times faster than humans. In 2000, HFT accounted for less than 10% of total equity trading but at the turn of the 21st Century execution time of HFT trades reduced from several seconds to microseconds in 2010 and then further reduced to nanoseconds in 2012.

In 2005, HFT share rose to 35% of total equity trades across the United States. Then in 2010, HFT share in total equity trading increased further to 56%. On May 6th 2010, $1 trillion was wiped off the market as dow plunged 1000 points in a single day because of computer-driven selling of over $4b. HFT firms were blamed for the crash. In 2011, nano trading technology allowed execution of trade in just few nanoseconds. One nanosecond equals one billionth of a second.

In Sep. 2012, the launch of determiner turns social media news into actionable trading signals and helps reporting the latest business news 54 minutes faster than conventional news mediums. In late 2012, the share of HFT increased to 70% of the total US equity trading. During the same year,

DOI:10.6814/NCCU201901120

HFT industry attracted investment worth millions. A custom-made chip developed for HFT allowed trade execution in 0.000000074 second. A special cable worth $300 was built just to save 0.006 second off the transaction time.

In Nov. 2012, FBI began investigation about social media frauds amid its high impact on stock markets. In April 2, 2013, CFTC and SEC banned financial announcements relating to public companies on social media. In April 4, 2013, Bloomberg terminals included live tweets into its data service. April 23, 2013, a false tweet about white house bombing caused Tsunami in financial markets, Dow plunges 1% in just 3 minutes. Later on, a server based in Washington DC archived capability to transmit data to New Jersey at the speed of light thanks to superfast microwave transmission technology

In September 18, 2013, around 2pm, Fed surprised Wall Street by announcing a delay in QE tapering. Assets worth $600m had changed hands in milliseconds before the news reached Chicago. Then on September 2013, Italy became first country to impose tax on HFT. A 0.002% tax was imposed on equity trades closing before 0.5 seconds. In 2013, economists debate risks of HFT trading amid 2010 crash; some economists demanded a complete ban on HFT. In Oct 7, 2016, Pound nosedived more than 800 pips in just few minutes; economists blamed HFT for the cable's mysterious crash[2].

1.1.1 What is High-Frequency Trading (HFT)

High Frequency Trading (HFT) is a type of algorithmic trading in which large volumes of assets are bought and sold automatically at very high speed. HFT is a very popular form of trading. It can generate quick profits with steady win-rate on certain market conditions.

In financial markets, HFT is a type of algorithmic trading characterized by high speeds, high turnover rates, and high order-to-trade ratios that leverages high-frequency financial data and electronic trading tools. While there is no single definition of HFT, among its key attributes are highly

DOI:10.6814/NCCU201901120

sophisticated algorithms, co-location, and very short-term investment horizons. HFT can be viewed as a primary form of algorithmic trading in finance. Specifically, it is the use of sophisticated technological tools and computer algorithms to rapidly trade securities. HFT uses proprietary trading strategies carried out by computers to move in and out of positions in seconds or fractions of a second[3].

HFT is a program trading platform that uses powerful computers to transact a large number of orders in fractions of a second. It uses complex algorithms to analyze multiple markets and execute orders based on market conditions. Typically, the traders with the fastest execution speeds are more profitable than traders with slower execution speeds[4].

1.1.2 How does HFT work?

High frequency trading strategies describe an algorithm that is trading thousands of times a day, to capture inefficiencies in the exchange rate of a currency pair or some other financial instrument. The concept is a relative term, describing how market participants use technology to gain information, and act upon it, in advance of the rest of the market. In essence, high frequency traders are front running your order. If the price of a currency pair is off by even half of a pip, the high frequency trader will attempt to capture this inefficiency.

High frequency traders initially appeared onto the equity market scene. New regulation allowed electronic exchanges to compete with one another, which left the door open for high frequency traders to step in and search for discrepancies in prices. High frequency traders rely on extremely low latencies and use high speed connections in conjunction with trading algorithms to exploit inefficiencies created by these exchanges.

Many HFT strategies revolve around searching for and sniffing out institutional order flows, by going through the multitude of electronic exchanges available to trade securities. These algorithms would detect a trade and attempt to transact the same trade before the order was filled at another electronic exchange. These algorithms are front running many

DOI:10.6814/NCCU201901120

securities orders and are predicated on the idea that speed was of the essence.

1.1.3 Basics of High-Frequency Trading

HFT is complex algorithmic trading in which large numbers of orders are executed within seconds. It adds liquidity to the markets and eliminates small bid-ask spreads.

HFT became commonplace in the markets following the introduction of incentives offered by exchanges for institutions to add liquidity to the markets. By offering small incentives to these market makers, exchanges gain added liquidity, and the institutions that provide the liquidity also see increased profits on every trade they make, on top of their favorable spreads. Although the spreads and incentives amount to a fraction of a cent per transaction, multiplying that by a large number of trades per day amounts to sizable profits for high-frequency traders[5].

HFT became popular when exchanges started to offer incentives for companies to add liquidity to the market. For instance, the New York Stock Exchange (NYSE) has a group of liquidity providers called Supplemental Liquidity Providers (SLPs) that attempts to add competition and liquidity for existing quotes on the exchange. As an incentive to companies, the NYSE pays a fee or rebate for providing said liquidity. In July 2016, the average SLP rebate was $0.0019 for NYSE- and NYSE MKT-listed securities on NYSE. With millions of transactions per day, this results in a large amount of profits. The SLP was introduced following the collapse of Lehman Brothers in 2008, when liquidity was a major concern for investors[6].

1.1.4 What's the importance of speed in trading?

Speed has become so important to the success of a high frequency operation, that these businesses invest enormous sums of money into

DOI:10.6814/NCCU201901120

building their low latency infrastructure. High frequency traders target low latency machinery in an effort to find the fastest computers available. For a high frequency trader, finding the path of least resistance in communication is the key to successful arbitrage. Additionally, the proximity of a high frequency trader's black box to an exchange will reduce or increase the speed at which a transaction is recognized. So, a proximity war, among high frequency firms, has emerged and created competition for real-estate around a physical exchange location, especially in the equity space[7].

1.1.5 High Frequency Trading Strategies

High-frequency trading is quantitative trading that is characterized by short portfolio holding periods. All portfolio-allocation decisions are made by computerized quantitative models. The success of high-frequency trading strategies is largely driven by their ability to simultaneously process large volumes of information, something ordinary human traders cannot do. Specific algorithms are closely guarded by their owners. Many practical algorithms are in fact quite simple arbitrages which could previously have been performed at lower frequency-competition tends to occur through who can execute them the fastest rather than who can create new breakthrough algorithms[8].

There are many strategies employed by high frequency traders to make money; some are quite common, some are more controversial. For instance, some HF traders trade from both sides i.e. they place orders to buy as well as sell using limit orders that are above the current market place (in the case of selling) and slightly below the current market price (in the case of buying). The difference between the two is the profit they pocket. Thus these traders indulge in "market making" only to make profits from the difference between the bid-ask spread. These transactions are carried out by high-speed computers using algorithms. Another way traders make money is by looking for price discrepancies between securities on different exchanges or asset classes. This strategy is called statistical arbitrage, wherein a proprietary trader is on the lookout for temporary inconsistencies in prices across different exchanges. With the help of ultra fast transactions,

DOI:10.6814/NCCU201901120

they capitalize on these minor fluctuations which many don't even get to notice.

The common types of high-frequency trading include several types of market-making, event arbitrage, statistical arbitrage, and latency arbitrage. Most high-frequency trading strategies are not fraudulent, but instead exploit minute deviations from market equilibrium.

1.1.5.1 Market making

According to SEC[9], a "market maker" is a firm that stands ready to buy and sell a particular stock on a regular and continuous basis at a publicly quoted price. You'll most often hear about market makers in the context of the Nasdaq or other "over the counter" (OTC) markets. Market makers that stand ready to buy and sell stocks listed on an exchange, such as the New York Stock Exchange, are called "third market makers." Many OTC stocks have more than one market-maker. Market-makers generally must be ready to buy and sell at least 100 shares of a stock they make a market in. As a result, a large order from an investor may have to be filled by a number of market-makers at potentially different prices.

There can be a significant overlap between a "market maker" and "HFT firm". HFT firms characterize their business as "Market making" - a set of high-frequency trading strategies that involve placing a limit order to sell (or offer) or a buy limit order (or bid) in order to earn the bid-ask spread. By doing so, market makers provide counterpart to incoming market orders. Although the role of market maker was traditionally fulfilled by specialist firms, this class of strategy is now implemented by a large range of investors, thanks to wide adoption of direct market access. As pointed out by empirical studies[10], this renewed competition among liquidity providers causes reduced effective market spreads, and therefore reduced indirect costs for final investors." A crucial distinction is that true market makers don't exit the market at their discretion and are committed not to, where HFT firms are under no similar commitment.

Some high-frequency trading firms use market making as their

DOI:10.6814/NCCU201901120

primary strategy. Automated Trading Desk (ATD), which was bought by Citigroup in July 2007, has been an active market maker, accounting for about 6% of total volume on both the NASDAQ and the New York Stock Exchange. In May 2016, Citadel LLC bought assets of ATD from Citigroup. Building up market making strategies typically involves precise modeling of the target market microstructure[11] together with stochastic control techniques[12].

These strategies appear intimately related to the entry of new electronic venues. Academic study[13] of Chi-X's entry into the European equity market reveals that its launch coincided with a large HFT that made markets using both the incumbent market, NYSE-Euronext, and the new market, Chi-X. The study shows that the new market provided ideal conditions for HFT market-making, low fees (i.e., rebates for quotes that led to execution) and a fast system, yet the HFT was equally active in the incumbent market to offload nonzero positions. New market entry and HFT arrival are further shown to coincide with a significant improvement in liquidity supply[14].

1.1.5.2 Fraud

The Michael Lewis book Flash Boys: A Wall Street Revolt discusses high-frequency trading, including the tactics of spoofing, layering and quote stuffing, which are all now illegal[15]. The book details the rise of high-frequency trading in the US market[16].

1.1.5.3 Ticker tape trading

Much information happens to be unwittingly embedded in market data, such as quotes and volumes. By observing a flow of quotes, computers are capable of extracting information that has not yet crossed the news screens. Since all quote and volume information is public, such strategies are fully compliant with all the applicable laws.

DOI:10.6814/NCCU201901120

Filter trading is one of the more primitive high-frequency trading strategies that involves monitoring large amounts of stocks for significant or unusual price changes or volume activity. This includes trading on announcements, news, or other event criteria. Software would then generate a buy or sell order depending on the nature of the event being looked for[17].

Tick trading often aims to recognize the beginnings of large orders being placed in the market. For example, a large order from a pension fund to buy will take place over several hours or even days, and will cause a rise in price due to increased demand. An arbitrageur can try to spot this happening then buy up the security, then profit from selling back to the pension fund. This strategy has become more difficult since the introduction of dedicated trade execution companies in the 2000s which provide optimal trading for pension and other funds, specifically designed to remove the arbitrage opportunity.

1.1.5.4 Event arbitrage

Certain recurring events generate predictable short-term responses in a selected set of securities[18]. High-frequency traders take advantage of such predictability to generate short-term profits[19].

1.1.5.5 Statistical arbitrage

Another set of high-frequency trading strategies are strategies that exploit predictable temporary deviations from stable statistical relationships among securities. Statistical arbitrage at high frequencies is actively used in all liquid securities, including equities, bonds, futures, foreign exchange, etc. Such strategies may also involve classical arbitrage strategies, such as covered interest rate parity in the foreign exchange market, which gives a relationship between the prices of a domestic bond, a bond denominated in a foreign currency, the spot price of the currency, and the price of a forward contract on the currency. High-frequency trading

DOI:10.6814/NCCU201901120

allows similar arbitrages using models of greater complexity involving many more than four securities.

The TABB Group estimates that annual aggregate profits of high-frequency arbitrage strategies exceeded US$21 billion in 2009[20], although the Purdue study estimates the profits for all high frequency trading were US$5 billion in 2009[21].

1.1.5.6 Index arbitrage

Index arbitrage exploits index tracker funds which are bound to buy and sell large volumes of securities in proportion to their changing weights in indices. If a HFT firm is able to access and process information which predicts these changes before the tracker funds do so, they can buy up securities in advance of the trackers and sell them on to them at a profit.

1.1.5.7 News-based trading

Company news in electronic text format is available from many sources including commercial providers like Bloomberg, public news websites, and Twitter feeds. Automated systems can identify company names, keywords and sometimes semantics to make news-based trades before human traders can process the news.

1.1.5.8 Low-latency strategies

A separate, "naïve" class of high-frequency trading strategies relies exclusively on ultra-low latency direct market access technology. In these strategies, computer scientists rely on speed to gain minuscule advantages in arbitraging price discrepancies in some particular security trading simultaneously on disparate markets.

DOI:10.6814/NCCU201901120

Another aspect of low latency strategy has been the switch from fiber optic to microwave technology for long distance networking. Especially since 2011, there has been a trend to use microwaves to transmit data across key connections such as the one between New York City and Chicago. This is because microwaves travelling in air suffer a less than 1% speed reduction compared to light travelling in a vacuum, whereas with conventional fiber optics light travels over 30% slower[22].

1.1.5.9 Order properties strategies

High-frequency trading strategies may use properties derived from market data feeds to identify orders that are posted at sub-optimal prices. Such orders may offer a profit to their counterparties that high-frequency traders can try to obtain. Examples of these features include the age of an order[23] or the sizes of displayed orders[24]. Tracking important order properties may also allow trading strategies to have a more accurate prediction of the future price of a security.

1.1.6 What algorithms are used in HFT?

There are many algorithms that are common in High Frequency Trading, let's pick up and describe three of them;

1) Pair Trading - Trade two currencies which naturally track each other an example could be Euro and US Dollar, make money when they fall out of line on the idea that they will have to revert back to tracking each other.

2) Volume-Weighted Average Price - VWAP is used to execute large orders at a better average price. It is the ratio of the value traded to the total volume traded over a time period

3) Time-Weighted Average Price - TWAP like VWAP is another sophisticated strategy for buying or selling large blocks of currencies without affecting the price.

DOI:10.6814/NCCU201901120

1.1.7 Benefits and critiques of HFT

There are many advantages associated with HFT. For example, all trades are executed at the best possible prices. Instant and accurate trade order placement is another key advantage of HFT that consequently increases chances of execution at desired levels. Reduced transaction cost is another main pro of HFT. Similarly, simultaneous automated checks can be implemented on multiple market conditions. HFT also reduces risk of manual errors in placing the trades. Last but not the least; HFT reduces possibility of mistakes by human traders based on emotional and psychological behavior.

HFT is a great way to make quick profits. It is highly risky but very effective trading strategy. HFT is executed via robots so there is no risk that you can lose your money due to emotional trading. Robots place all trades on the basis of predefined trading strategies; there is no way they can deviate from predefined trading plan which is a key to success in forex trading.

The major benefit of HFT is that it has improved market liquidity and removed bid-ask spreads that previously would have been too small. This was tested by adding fees on HFT, and as a result, bid-ask spreads increased. One study assessed how Canadian bid-ask spreads changed when the government introduced fees on HFT, and it was found that bid-ask spreads increased by 9%.

HFT is controversial and has been met with some harsh criticism. It has replaced a number of broker-dealers and uses mathematical models and algorithms to make decisions, taking human decision and interaction out of the equation. Decisions happen in milliseconds, and this could result in big market moves without reason. As an example, on May 6, 2010, the Dow Jones Industrial Average (DJIA) suffered its largest intraday point drop ever, declining 1,000 points and dropping 10% in just 20 minutes before rising again. A government investigation blamed a massive order that triggered a sell-off for the crash.

DOI:10.6814/NCCU201901120

Many see high-frequency trading as unethical and an unfair advantage for large firms against smaller institutions and investors. Stock markets are supposed to be fair and a level playing field, which HFT arguably disrupts since the technology can be used for abusive ultra-short-term strategies. High-frequency traders prey on any imbalance between supply and demand, using arbitrage and speed to their advantage. Their traders are not based on fundamental research about the company or its growth prospects, but on opportunities to strike. Though HFT doesn't target anyone in particular, it can cause collateral damage to retail investors, as well as institutional investors like mutual funds that buy and sell in bulk[25].

Another major complaint about HFT is the liquidity provided by HFT is "ghost liquidity," meaning it provides liquidity that is available to the market one second and gone the next, preventing traders from actually being able to trade this liquidity.

Generally speaking, there are three primary criticisms of HFT. The first one is that it allows institutional players to gain an upper hand in trading because they are able to trade in large blocks through the use of algorithms. An additional critique of HFT is that it allows large companies to profit at the expense of the "little guys," or the institutional and retail investors. The third criticism against HFT is that the liquidity produced by this type of trading is momentary. It disappears within seconds, making it impossible for traders to take advantage of it.

1.2 History of high-frequency trading

High-frequency trading has taken place at least since the 1930s, mostly in the form of specialists and pit traders buying and selling positions at the physical location of the exchange, with high-speed telegraph service to other exchanges[26].

The rapid-fire computer-based HFT developed gradually since 1983 after NASDAQ introduced a purely electronic form of trading[27].

DOI:10.6814/NCCU201901120

Although the first research in my database about HFT is at 1993, the exact history of HFT can be traced back at least since 1998. The U.S. Securities and Exchange Commission (SEC) adopted Regulation ATS (Alternative Trading Systems), including electronic exchanges.

After that, SEC's Regulation NMS (National Market System), which was adopted in 2005, further provided strong incentives for trading venues to automate, especially the NYSE, which was the last major floor-based exchange in the U.S.

At the turn of the 21st century, HFT trades had an execution time of several seconds, whereas by 2010 this had decreased to milli- and even microseconds[28].

Until 2010, SEC issued a Concept release[29] seeking public comments on issues such as HFT. SEC admitted, "The term (HFT) is relatively new and is not yet clearly defined."

HFT was not a well-known topic outside the financial sector, until an article[30] published by the New York Times in July 2009 which was one of the first to bring the subject to the public's attention. HFT has radically changed the stock markets.

On September 2, 2013, Italy became the world's first country to introduce a tax specifically targeted at HFT, charging a levy of 0.02% on equity transactions lasting less than 0.5 seconds[31].

Some view the Flash Crash of May 6, 2010 as evidence of the potential harmful effects of HFT. Michael Lewis's book in 2014, FLASH BOYS, even raised the controversy concerning about HFT by pointing out that electronic trading has rigged the market against ordinary investors, particularly in America, but Lewis paid little attention to the market benefits of HFT.

Many academics raised the controversy concerning about HFT[32]. Even SEC Division of Trading and Markets Director Brett Redfearn admitted, "There are a lot of different definitions of HFT." The development of HFT has ignited a heated debate among participants, researchers and regulators

DOI:10.6814/NCCU201901120

about the benefits and concerns related to HFT.

1.2.1 Market growth

In the early 2000s, high-frequency trading still accounted for fewer than 10% of equity orders, but this proportion was soon to begin rapid growth. According to data from the NYSE, trading volume grew by about 164% between 2005 and 2009 for which high-frequency trading might be accounted. As of the first quarter in 2009, total assets under management for hedge funds with high-frequency trading strategies were $141 billion, down about 21% from their peak before the worst of the crises, although most of the largest HFT's are actually LLC's owned by a small number of investors.

The high-frequency strategy was first made popular by Renaissance Technologies who use both HFT and quantitative aspects in their trading. Many high-frequency firms are market makers and provide liquidity to the market which lowers volatility and helps narrow bid-offer spreads, making trading and investing cheaper for other market participants[33].

1.2.2 Market share

In the United States in 2009, high-frequency trading firms represented 2% of the approximately 20,000 firms operating today, but accounted for 73% of all equity orders volume. The major U.S. high-frequency trading firms include Virtu Financial, Tower Research Capital, IMC, Tradebot and Citadel LLC. The Bank of England estimates similar percentages for the 2010 US market share, also suggesting that in Europe HFT accounts for about 40% of equity orders volume and for Asia about 5–10%, with potential for rapid growth. By value, HFT was estimated in 2010 by consultancy Tabb Group to make up 56% of equity trades in the US and 38% in Europe.

As HFT strategies become more widely used, it can be more difficult

DOI:10.6814/NCCU201901120

to deploy them profitably. According to an estimate from Frederi Viens of Purdue University, profits from HFT in the U.S. has been declining from an estimated peak of $5bn in 2009, to about $1.25bn in 2012.

Though the percentage of volume attributed to HFT has fallen in the equity markets, it has remained prevalent in the futures markets. According to a study in 2010 by Aite Group, about a quarter of major global futures volume came from professional high-frequency traders. In 2012, according to a study by the TABB Group, HFT accounted for more than 60 percent of all futures market volume in 2012 on U.S. exchanges.

In 2017, Aldridge and Krawciw[34] estimated that in 2016 HFT on average initiated 10-40% of trading volume in equities, and 10-15% of volume in foreign exchange and commodities. Intraday, however, proportion of HFT may vary from 0% to 100% of short-term trading volume. Previous estimates reporting that HFT accounted for 60-73% of all US equity trading volume, with that number falling to approximately 50% in 2012 were highly inaccurate speculative guesses. High-frequency traders move in and out of short-term positions at high volumes and high speeds aiming to capture sometimes a fraction of a cent in profit on every trade. HFT firms do not consume significant amounts of capital, accumulate positions or hold their portfolios overnight. As a result, HFT has a potential Sharpe ratio (a measure of reward to risk) tens of times higher than traditional buy-and-hold strategies. High-frequency traders typically compete against other HFTs, rather than long-term investors[35]. HFT firms make up the low margins with incredibly high volumes of trades, frequently numbering in the millions.

A substantial body of research argues that HFT and electronic trading pose new types of challenges to the financial system[36]. Algorithmic and high-frequency traders were both found to have contributed to volatility in the Flash Crash of May 6, 2010, when high-frequency liquidity providers rapidly withdrew from the market[37]. Several European countries have proposed curtailing or banning HFT due to concerns about volatility[38].

DOI:10.6814/NCCU201901120

1.3 Definition of high-frequency trading

1.3.1 Metrics for Defining HFT from SEC

The Concept Release from Securities and Exchange Commission (SEC) on Equity Market Structure recognized that HFT is one of the most significant market structure developments in recent years[39]. It noted, for example, that estimates of HFT typically exceeded 50% of total volume in U.S.-listed equities and concluded that, "[b]y any measure, HFT is a dominant component of the current market structure and likely to affect nearly all aspects of its performance."

The Concept Release also noted that the term "HFT" was not clearly defined. To deal with this problem, the Concept Release first generally defined "proprietary firm" as "professional traders acting in a proprietary capacity that generate a large number of trades on a daily basis." These traders could be organized in a variety of ways, including as a proprietary trading firm (which may or may not be a registered broker-dealer and a member of FINRA), as the proprietary trading desk of a multi-service broker-dealer, or as a hedge fund.

Next, the Concept Release identified five other characteristics that often are attributed to HFT:

1) Use of extraordinarily high speed and sophisticated programs for generating, routing, and executing orders.

2) Use of co-location services and individual data feeds offered by exchanges and others to minimize network and other latencies.

3) Very short time-frames for establishing and liquidating positions.

4) Submission of numerous orders that are cancelled shortly after submission.

DOI:10.6814/NCCU201901120

5) Ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions overnight).

1.3.2 HFT definition from NASDAQ.com

Refers to computerized trading using proprietary algorithms. There are two types high frequency trading. Execution trading is when an order (often a large order) is executed via a computerized algorithm. The program is designed to get the best possible price. It may split the order into smaller pieces and execute at different times. The second type of high frequency trading is not executing a set order but looking for small trading opportunities in the market. It is estimated that 50 percent of stock trading volume in the U.S. is currently being driven by computer-backed high frequency trading. Also known as algo or algortihmic trading[40].

1.3.3 HFT definition from Wikipedia

In financial markets, high-frequency trading (HFT) is a type of algorithmic trading characterized by high speeds, high turnover rates, and high order-to-trade ratios that leverages high-frequency financial data and electronic trading tools. While there is no single definition of HFT, among its key attributes are highly sophisticated algorithms, co-location, and very short-term investment horizons. HFT can be viewed as a primary form of algorithmic trading in finance. Specifically, it is the use of sophisticated technological tools and computer algorithms to rapidly trade securities. HFT uses proprietary trading strategies carried out by computers to move in and out of positions in seconds or fractions of a second[41].

1.3.4 HFT definition from Investopedia

High-frequency trading (HFT) is an automated trading platform used

DOI:10.6814/NCCU201901120

by large investment banks, hedge funds and institutional investors that utilizes powerful computers to transact a large number of orders at extremely high speeds. These high-frequency trading platforms allow traders to execute millions of orders and scan multiple markets and exchanges in a matter of seconds, thus giving the institutions that use the platforms a huge advantage in the open market.

The systems use complex algorithms to analyze the markets and are able to spot emerging trends in a fraction of a second. By being able to recognize shifts in the marketplace, the trading systems send hundreds of baskets of stocks out into the marketplace at bid-ask spreads that are advantageous to the traders. By essentially anticipating and beating the trends to the marketplace, institutions that implement high-frequency trading can gain favorable returns on trades they make by essence of their bid-ask spread, resulting in significant profits[42].

1.3.5 Conclusion

Since the Securities and Exchange Commission (SEC) has no formal definition of HFT, it only attributes HFT with certain features listed above. This represents that the researches in this area are still highly diverse regarding issues for this new research agenda.

1.4 Current qualitative research in HFT

Since history of HFT is not so long, the researches in this area are highly diverse regarding issues for this new research agenda. According to a Review article by editor Goldstein (2014)[43], unlike established topics in finance, such as dividend policy, capital structure, or asset pricing, HFT is a new, emerging, and rapidly evolving area for the markets, regulators, and the public.

In more recent years, many attempts have been made to research HFT by a number of scholars. Goldstein, Kumar and Graves (2014)[44] reviewed

DOI:10.6814/NCCU201901120

the works in this area from the empirical and theoretical papers and assigned them into following six categories on the basis of different topics:

1) market performance (Budish, Cramton, & Shim, 2015; Menkveld, 2014; Schwartz & Wu, 2013; Jarnecic & Snape, 2014; Brogaard, Hendershott, & Riordan, 2014; Hendershott & Riordan, 2013; Popper, 2012c; Baron, Brogaard, & Kirilenko, 2012; Jones, 2013a; Hasbrouck & Saar, 2013; Credit Suisse, 2012),

2) strategies and practices (Aldridge, 2013; Kirilenko & Lo, 2013; Laughlin, Aguirre, & Grundfest, 2014),

3) evolution (Rubenstein, 2015; Goldstein, Kumar, & Graves, 2014; Kirilenko & Lo, 2013; Popper, 2012c),

4) speed (Angel, 2014; Laughlin, Aguirre, & Grundfest, 2014; Wissner-Gross & Freer, 2010; Brogaard, Hendershott, Hunt & Ysusi, 2014; Hasbrouck & Saar, 2013),

5) fairness (SEC, 2010; Narang, 2010; Patterson, Strasburg, & Pleven, 2013),

6) regulatory implications (Piwowar, 2013; SEC/CFTC, 2010; Westbrook, 2010).

1.5 Research categories and findings in HFT

Table 1 summarizes findings from Goldstein, Kumar and Graves (2014) by topics about each category which include the current state, topics of debate, and empirical and theoretical researches in HFT.

Table 1 : Research categories and findings in HFT

Category	Authors	Findings by topic

DOI:10.6814/NCCU201901120

• Market performance	• Budish et al. (2015)[45] and Menkveld (2014)[46] and Schwartz & Wu (2013)[47]	• Adverse selection － HFT's "socially wasteful arms race" could disadvantage other ordinary investors and then reduce market quality, as measured by liquidity and price informativeness.
	• Jarnecic & Snape (2014)[48]	• Liquidity － If HFT activity can improve the liquidity of markets?
	Brogaard et al. (2014)[49] and Hendershott & Riordan (2013)[50]	• Market structure － Market structure changes due to the incremental effect of algorithmic trading and HFT.
	• Popper (2012c)[51]	• Transaction costs － As the recent volume of HFT has decreased, the benefits of HFT in reducing trading costs for ordinary investors have stalled.
	• Baron et al. (2012)[52]	• Profitability (of HFT vs. non-HFT) － HFT profits are earned at the expense of other traders. － HFT markets are effectively a "zero sum game".
	• Jones (2013a)[53]	• Volatility － After surveying 30 theoretical and empirical papers on the topic of HFT, Jones (2013a) concludes that HFTs are making markets better.
	• Hasbrouck & Saar (2013)[54]	－ The impact of HFT on market quality and volatility.
	• Credit Suisse (2012)[55]	－ By the findings that long-term volatility in recent years remained within historical norms, while short-term volatility declined, Credit Suisse concludes that markets are "not worse" for the presence of HFT.
		• Market efficiency

DOI:10.6814/NCCU201901120

• Strategies and practices	• Aldridge (2013)[56]	• Algorithmic strategies – Statistical Arbitrage Strategies – Directional Trading Around Events – Automated Market Making – Modeling Information in Order Flow – Latency Arbitrage – Spread Scalping – Rebate Capture – Quote Matching – Layering
	• Aldridge (2013)	• Market Manipulation – Pinging/Sniping/Sniffing/Phishing – Quote Stuffing – Spoofing – Pump-and-Dump – Ignition
	• Kirilenko & Lo (2013)[57]	• Manipulative trading activities – 'Order anticipation' trading strategy (e.g. a "pinging" tactic to discover the price other traders are willing to pay or to discover undisplayed liquidity.)
	• Laughlin et al. (2014)[58]	• Low-latency strategies – Many HFT firms are concerned about transmission speed across geographic distances and utilize strategies that capitalize on their geographic location. • Co-location – The ability to access direct data feeds from exchanges which includes sophisticated order execution algorithms services.
• Evolution	• Rubenstein (2015)[59]	• Trading volume – HFT now accounts for almost 50% of daily stock trades.
	• Goldstein et al. (2014)[60] and Kirilenko & Lo (2013)[61]	• Trading activity – HFT accounted for between 40% and 60% of trading activity across all U.S. financial markets for stocks, options and currencies.

DOI:10.6814/NCCU201901120

	• Popper (2012c)[62]	• Volumes and profits downtrend in US – HFT volume down from 61% in 2009 to 51% in 2012. – HFT profits were estimated at most \$1.25B in 2012, down 35% from 2011 and 74% lower than the peak of about \$4.9B in 2009.
• Speed	• Angel (2014)[63]	• Data transmission – Physical limitations on current trading due to Einstein's theories and related quantum physics to finance.
	• Laughlin et al. (2014)[64] and Wissner-Gross & Freer (2010)[65]	• Data transmission – Techniques to minimize transmission delays and execution latencies and affected price discovery when HFT firms trade securities in different locations around the world.
	• Brogaard et al. (2014)[66] and Hasbrouck & Saar (2013)[67]	• Technology upgrades – Exchanges upgrading for lower latencies
• Fairness	• SEC (2010)[68]	• Market structure – The SEC (2010) concept release directly questions the fairness of the current market structure, HFT, and the use of a variety of HFT tools and strategies. – • Unfair access concerns – The SEC (2010) concept release directly questions if co-location provides HFTs an unfair advantage because of greater resources and sophistication to take advantage of co-location services than other market participants, including long-term investors?
	• Narang (2010)[69]	• Rebate structure – If the current rebate structure based on volume unfairly benefits HFT firms over non-HFT firms?

DOI:10.6814/NCCU201901120

	• Patterson et al. (2013)[70]	• Insider advantages − HFTs are using a hidden facet of the Chicago Mercantile Exchange's computer system to trade on the direction of the futures market before other investors get the same information.
• Regulatory implications	• Piwowar (2013)[71]	• The role of speed − SEC Commissioner called for a comprehensive review of U.S. markets which should examine the role of speed in the markets.
	• SEC/CFTC (2010)	• Market-makers & liquidity providers − Whether HFT market-makers should be subject to regulations that would require them to stay active in volatile markets, rather than deserting the markets en masse and damaging liquidity.
	• Westbrook (2010)[72]	• Concerns − Lawmakers questioned whether the HFT practice is benefiting Wall Street at the expense of individual investors.

1.6 Problems being investigated

There are plenty of findings by topics shown in Table 1 which include the topics of debate, controversial issues, and the current arguments against HFT. However, one drawback of this work by Goldstein, Kumar and Graves (2014)[73] is that those findings about HFT are largely qualitative in nature. Although there have been some review articles concerning HFT, for most of these works, it is not clear if merely depending on review articles would be capable of revealing the developing trends or future orientation of a new research field. Therefore, my general purpose in this research is to provide a quantitative analysis of global empirical and theoretical HFT papers.

DOI:10.6814/NCCU201901120

HFT is an emerging, ever changing and rapidly evolving area with highly interdisciplinary in nature for the markets, regulators, and the public[74]. This diversity may root from the emerging nature of computing technology and its wide appeal as well as unique researcher and practitioner viewpoints. Those findings by topics shown in Table 1 represent the diverse issues and findings in the field of HFT which include the introduction of keywords and ideas and even new concepts about HFT. However, what are the areas of focus in HFT? What about their association with each other? What are the developing trends in current research?

1.7 General approach

Keywords have been generally identified as the words that reflect the research themes of individual publications that concern researchers. Further, keywords network represents relationships of keywords among HFT papers. When two keywords occur in a same article, it is an indication of connection between the themes which they represent. Therefore, a comprehensive network perspective analysis is required to reveal the developing trends or future orientation of possible new research field from HFT.

Social network analysis (SNA), sometimes also referred to as "structural analysis"[75], is a broad strategy for investigating social structures. For measurement, social network analysis (SNA) measures are a vital tool for understanding the behavior of networks and graphs. These algorithms use graph theory to calculate the importance of any given node in a network[76]. When they're well implemented, SNA measures allow the analyst to cut through noisy data and hone into the parts of a network that require further attention.

1.8 Purpose of the Research

In this research, my focus is to construct and analyze keywords network by using the Social network analysis (SNA) techniques which

DOI:10.6814/NCCU201901120

have already been widely applied in many disciplines of science. Specifically, this study will quantitatively analyze existing empirical and theoretical HFT papers to address the following objectives:

1) To construct keywords network from HFT papers published in world leading journals during the period from 1993 to 2017.

2) To investigate the characteristics of keywords network of HFT papers by utilizing social network analysis (SNA) techniques.

3) To find and compare the changes in keywords network of HFT papers over time.

4) To compare the structural analysis of keywords network with those qualitative findings by Goldstein, Kumar and Graves (2014)[77] to reveal the developing trends or future orientation of possible new research field about HFT.

Despite the high growth rate of publications, there have been few attempts to gather systematic data on the global scientific production of research on HFT. These investigations can help researchers to realize the breadth of HFT research and to establish future research directions and to provide an entry point to any academic, regardless of their prior knowledge of the theme. These investigations can also help to get different point of view that emerge from a comprehensive review of existing HFT papers by future researchers before choosing their interested field.

The remainder of this research is organized as follows. Chapter two presents the literature reviews which include social network analysis (SNA) and Bibliometrics as the background knowledge. Chapter three shows the research methods. Chapter four performs the experiment and results in order to report findings and present them in a systematic manner. Chapter five provides discussions and implications. Chapter six offers some concluding remarks.

DOI:10.6814/NCCU201901120

2. Literature review

2.1 Social network analysis (SNA)

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory[78]. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them. These networks are often visualized through sociograms in which nodes are represented as points and ties are represented as lines. Examples of social structures commonly visualized through social network analysis include social media networks, memes spread, information circulation, friendship and acquaintance networks, business networks, social networks, collaboration graphs, kinship, disease transmission, and sexual relationships.

SNA has emerged as a key technique in modern sociology. It has also gained a significant following in anthropology, biology, demography, communication studies, economics, geography, history, information science, organizational studies, political science, social psychology, development studies, sociolinguistics, and computer science and is now commonly available as a consumer tool[79]. SNA has found applications in various academic disciplines, as well as practical applications such as countering money laundering and terrorism.

2.2 Some SNA measures

SNA measures are a vital tool for understanding the behavior of networks and graphs. These algorithms use graph theory to calculate the importance of any given node in a network. When they're well implemented, SNA measures allow the analyst to cut through noisy data and hone into the parts of a network that require further attention. SNA measures at least include measuring degree centrality, betweenness

DOI:10.6814/NCCU201901120

centrality, closeness centrality, EigenCentrality or PageRank for each network quantitatively[80].

2.2.1 SNA measure 1 : Degree Centrality

The degree centrality measure finds nodes with the highest number of links to other nodes in the network. Nodes with a high degree centrality have the best connections to those around them – they might be influential, or just strategically well-placed.

(1) Definition: Degree centrality assigns an importance score based purely on the number of links held by each node.

(2) What it tells us: How many direct, 'one hop' connections each node has to other nodes within the network.

(3) When to use it: For finding very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

(4) A bit more detail: Degree centrality is the simplest measure of node connectivity. Sometimes it's useful to look at in-degree (number of inbound links) and out-degree (number of outbound links) as distinct measures, for example when looking at transactional data or account activity.

2.2.2 SNA measure 2 : Betweenness centrality

Nodes with a high betweenness centrality score are the ones that most frequently act as 'bridges' between other nodes. They form the shortest pathways of communication within the network.

Usually this would indicate important gatekeepers of information between groups.

DOI:10.6814/NCCU201901120

(1) Definition: Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.

(2) What it tells us: This measure shows which nodes act as 'bridges' between nodes in a network. It does this by identifying all the shortest paths and then counting how many times each node falls on one.

(3) When to use it: For finding the individuals who influence the flow around a system.

(4) A bit more detail: Betweenness is useful for analyzing communication dynamics, but should be used with care. A high betweenness count could indicate someone holds authority over, or controls collaboration between, disparate clusters in a network; or indicate they are on the periphery of both clusters.

2.2.3 SNA measure 3 : Closeness centrality

This is the measure that helps you find the nodes that are closest to the other nodes in a network, based on their ability to reach them.

To calculate this, the algorithm finds the shortest path between each node, then assigns each node a score based on the sum of all the paths.

Nodes with a high closeness value have a lower distance to all other nodes. They'd be efficient broadcasters of information.

(1) Definition: This measure scores each node based on their 'closeness' to all other nodes within the network.

(2) What it tells us: This measure calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths.

(3) When to use it: For finding the individuals who are best placed to influence the entire network most quickly.

DOI:10.6814/NCCU201901120

(4) A bit more detail: Closeness centrality can help find good ‘broadcasters’, but in a highly connected network you will often find all nodes have a similar score. What may be more useful is using Closeness to find influencers within a single cluster.

2.2.4 SNA measure 4 : PageRank

PageRank identifies important nodes by assigning each a score based upon its number of incoming links (its ‘indegree’). These links are weighted depending on the relative score of its originating node.

(1) Definition: PageRank is a variant of EigenCentrality, also assigning nodes a score based on their connections, and their connections’ connections. The difference is that PageRank also takes link direction and weight into account – so links can only pass influence in one direction, and pass different amounts of influence.

(2) What it tells us: This measure uncovers nodes whose influence extends beyond their direct connections into the wider network.

(3) When to use it: Because it factors in directionality and connection weight, PageRank can be helpful for understanding citations and authority.

(4) A bit more detail: PageRank is famously one of the ranking algorithms behind the original Google search engine (the ‘Page’ part of its name comes from creator and Google founder, Sergei Brin).

2.2.5 SNA measure 5 : EigenCentrality

Very similar to PageRank, Eigenvector centrality is a measure of influence that takes into account the number of links each node has and the number of links their connections have, and so on throughout the network.

DOI:10.6814/NCCU201901120

(1) Definition: Like degree centrality, EigenCentrality measures a node's influence based on the number of links it has to other nodes within the network. EigenCentrality then goes a step further by also taking into account how well connected a node is, and how many links their connections have, and so on through the network.

(2) What it tells us: By calculating the extended connections of a node, EigenCentrality can identify nodes with influence over the whole network, not just those directly connected to it.

(3) When to use it: EigenCentrality is a good 'all-round' SNA score, handy for understanding human social networks, but also for understanding networks like malware propagation.

(4) A bit more detail: One way to calculate each node's EigenCentrality by converging on an eigenvector using the power iteration method. Learn more.

2.3 Some SNA software packages

Social network analysis software (SNA software) is software which facilitates quantitative or qualitative analysis of social networks, by describing features of a network either through numerical or visual representation. Visual representations of social networks are important to understand network data and convey the result of the analysis. Visualization often also facilitates qualitative interpretation of network data[81]. Followings are two example packages.

2.3.1 KeyLines from Cambridge Intelligence

KeyLines process graph visualization for JavaScript developers. Users can build game-changing graph visualization products that turn connected data into insight. Users can add graph visualization to their applications that work anywhere, as part of any stack[82].

DOI:10.6814/NCCU201901120

2.3.2 UCINET 6 for Windows

UCINET 6 for Windows is a software package for the analysis of social network data. It was developed by Lin Freeman, Martin Everett and Steve Borgatti. It comes with the NetDraw network visualization tool.

2.4 Bibliometrics

Bibliometrics is a helpful and widely used quantitative tool for evaluating the social and scientific importance of a specific discipline during a given period of time. The term bibliometrics has been introduced by Pritchard (1969)[83] who explained it as "the application of mathematical and statistical methods to books and other media of communication"[84]. Bibliometrics are measures of output and indicators of impact. For example, the simplest bibliometric is a count of publications. It is assumed that highly cited articles are important articles, and that reputable authors are read and cited frequently.

Bibliometric studies constitute an effective complement to the opinions and judgments of the experts in each field, providing useful and objective instruments for assessing the results of scientific activity and offering a more realistic view of this activity and its possible evolution and trends[85]. More advanced bibliometrics help researchers to understand the impact of their academic publications within the scope of the worldwide research community[86].

The three prime laws of bibliometrics are Bradford's Law of Scattering, Lotka's Law and Zipf's Law. Each of these is introduced in this section. There are three basic laws in the Bibliometrics which include Bradford’s law, Lotka's law and Zipf's law.

DOI:10.6814/NCCU201901120

2.4.1 Bradford's law

Bradford's law of scattering has been used extensively in the information science literature to describe the dispersion of articles in any scientific field and to identify core journals of serial titles[87]. Bradford found a pattern of how literature in a subject is distributed in journals[88]. He expressed his law of scattering in the following way[89]:

If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as 1:n:n2 . . .

2.4.2 Lotka's law

Lotka's law deals with the frequency distribution of scientific productivity of authors in any given field.

2.4.3 Zipf's law

Zipf's law concerns word frequency in the text. That means, if a word had been used more times, the length of the word will be shorter. Zipf's Law essentially predicts the phenomenon that we use familiar words with high frequency as we write[90].

The frequency of any word is inversely proportional to its rank in the frequency table. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc.

DOI:10.6814/NCCU201901120

2.4.4 Comparison and Summary

Comparison of three basic laws of Bibliometrics is as following Table 2.

Table 2 : Comparison of three basic laws of Bibliometrics

Three basic laws	Major Function
Bradford's law	to identify core journals
Lotka's law	scientific productivity of authors
Zipf's law	word frequency in the text

Since my purpose is to quantitatively analyze existing empirical and theoretical HFT papers to reveal the developmental trends or future orientation of a new research field like HFT, however, none of above fit the need of this study.

2.4.5 Other Bibliometrics methods

Conventional bibliometric methods often evaluate research trends by the publication outputs of countries (Braun et al., 1995)[91], research institutes (Xie et al., 2008)[92], journals (Colman et al., 1995)[93], and research fields (Ugolini et al., 1997)[94], as well as by citation analysis (Li and Ho, 2008)[95].

More information, closer to this research itself, has been introduced which include paper titles (Li et al., 2009b)[96], author keywords (Ugolini et al., 2001)[97], KeyWords Plus (Qin, 2000)[98], and abstracts (Zhang et al., 2010)[99].

2.4.5.1 Publication outputs of paper titles

The title of an article always includes the information that the author

DOI:10.6814/NCCU201901120

would most like to express to the readers, therefore analysis of the word distribution in article titles in different periods was recently used to evaluate research trends (Xie et al., 2008[100]; Zhang et al., 2010[101]).

The words in titles and abstracts will be separated, and then conjunctions and prepositions such as "and", "of", "in", and "on" will be discarded, as they are meaningless for further analysis.

2.4.5.2 Publication outputs of KeyWords Plus

KeyWord Plus is a kind of automatic indexing used in the citation databases produced by ISI. KeyWords Plus supplies additional search terms extracted from the titles of articles cited by authors in their bibliographies and footnotes in the ISI database, and substantially augment title-word and author-keyword indexing (Garfield, 1990)[102].

2.4.5.3 Publication outputs by abstract analysis

Zhang et al. (2010) first used the analysis of single words in abstracts to make specific inferences about the scientific literature and identify the subjective focus and emphasis specified by authors.

DOI:10.6814/NCCU201901120

3. Research methods

3.1 Purpose of the research

The objective of the present work is to describe the scientific output on the latest advances in HFT based on previous research. In this study, a keywords network analysis based on social network analysis (SNA) techniques will be used to describe the latest advances in HFT. Specifically, this study will quantitatively analyze existing empirical and theoretical HFT papers to address the following objectives:

1) To construct keywords network from HFT papers published in world leading journals during the period from 1993 to 2017.

2) To investigate the characteristics of keywords network of HFT papers by utilizing social network analysis (SNA) techniques.

3) To find and compare the changes in keywords network of HFT papers over time.

4) To compare the structural analysis of keywords network with those qualitative findings by Goldstein, Kumar and Graves (2014)[103] to reveal the developing trends or future orientation of possible new research field about HFT.

Findings from these investigations can help researchers to realize the breadth of HFT and to establish future research directions.

3.2 Materials and methods

The data will be based on the online version of Web of Science (WoS, previously known as Web of Knowledge). It is an online subscription-

DOI:10.6814/NCCU201901120

based scientific citation indexing service maintained by Thomson Reuters that provides a comprehensive citation search. It gives access to multiple databases that reference cross-disciplinary research, which allows for in-depth exploration of specialized sub-fields within an academic or scientific discipline[104].

The multidisciplinary coverage of the Web of Science encompasses over 50,000 scholarly books, 12,000 journals and 160,000 conference proceedings (as of September 3, 2014)[105]. The selection is made on the basis of impact evaluations. It comprises open-access journals, spanning multiple academic disciplines. The coverage includes: the sciences, social sciences, arts, and humanities, and goes across disciplines. However, Web of Science does not index all journals, and its coverage in some fields is less complete than in others.

The objective of the present work is to identify the important keywords from the scientific output on the latest advances in HFT, and to describe the characteristics of the keywords network of HFT research. To achieve these goals, we selected the Web of Science (WOS), which includes SCIE and SSCI and A&HCI from the Institute of Scientific Information (ISI) Web of Science databases. WOS is the most important and frequently used source for a broad review of scientific accomplishment in all research fields[106].

We constructed a database composed of keywords from HFT papers published in the WOS during the 24-year period from 1993 to 2017. The keywords were obtained from following two sources: (1) Author Keywords and (2) Keywords Plus in the ISI database[107]. All articles referring to HFT will be assessed according to: author keywords and Keywords Plus.

3.2.1 Author keywords

Keywords will be defined as comma-separated items of one or more words. All keywords, both those reported by authors and those attributed by ISI as well as the words in titles and abstracts, will be identified and separated into 3 eight-year periods, and then they will be further calculated

DOI:10.6814/NCCU201901120

in order to thoroughly and precisely analyze the variations of trends.

The Author Keywords are provided by the original authors. Author keywords analysis offers information about research trends that concern researchers. Bibliometric methods concerning author keywords have only been used in recent years (Chiu and Ho, 2007)[108], and their use in analyzing research trends is rare (Xie et al., 2008; Li et al., 2009a[109]; Zhang et al., 2010).

3.2.2 Keywords Plus

Keywords Plus are extracted from the titles of the cited references by Thomson Reuters. Keywords Plus, generated by an automatic computer algorithm, are words or phrases that appear frequently in the titles of an article's references and not necessarily in the title of the article or as Author Keywords (Garfield, 1990; Garfield & Sher, 1993).

3.2.3 Screening Methodology

The search will be conducted through Web of Science (WOS). There will be some categories of words used to identify the relevant literature. Following categories are some examples:

1) Words that are related to high frequency trading (HFT).

2) Words that are related to program trading.

3) Words that are related to algorithmic trading (Algo Trading).

3.3 Refinement of keywords and keywords databases

Due to different words may represent same or similar ideas and

DOI:10.6814/NCCU201901120

concepts, we standardize the keywords before constructing the keywords network. The basic rule for the refinement of keywords was that all keywords with identical meaning or similar ideas or concepts or even misspelled keywords from different articles will be grouped and considered as a single keyword. This refinement leads to a meaningful keywords database. The example of SNA steps in literature-based research was shown in Fig.1.

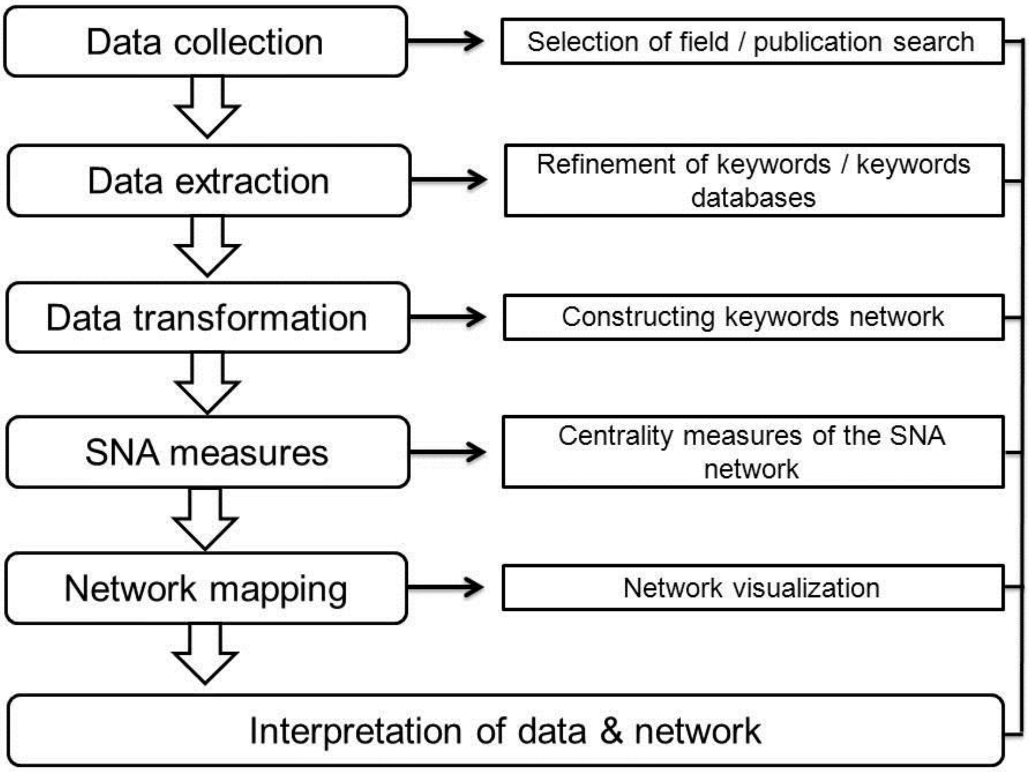


Fig. 1 example of SNA steps in literature-based research

3.4 Constructing keywords network

The construction of keywords network is based on three continuous stages which include following steps:

1) data collection stage

2) data extraction stage

DOI:10.6814/NCCU201901120

3) data transformation stage

During the data extraction stage, core keywords are identified from HFT papers and are changed to a standard form. Then in the data transformation stage, all the refined keywords will be input to the most popular social network research tool, Ucinet 6 for Windows[110] to get keywords network, or knowledge network, to study the relationship of each research theme.

3.5 Centrality measures of the SNA network

Network centrality[111] in the keywords network can measure the degree of relations among keywords. Social network analysis (SNA) measures include measuring degree centrality, betweenness centrality, closeness centrality, EigenCentrality or PageRank for each network quantitatively[112]. In order to understand the characteristics of the overall keywords network in HFT research, we selectively used betweenness centrality measuring to study the relationship of each research theme. Betweenness centrality is the extent to which a node lies on the paths between other nodes. It is measured as the fraction of the shortest paths between all pairs of other nodes in the network containing the node. A keyword that lies between two distinctive research themes can have high betweenness centrality even though it may have a small number of connections to other keywords in each theme[113]. In the keywords network, this represents the importance of a keyword in bridging subsets of keywords.

3.6 Network visualizations

Network visualizations is generally known as network mapping which can be generated from raw network data within Netdraw, a mapping program in Ucinet. NetDraw was applied to visualize network. It helps to obtain a clear sense of connectivity of keyword networks and to illustrate

DOI:10.6814/NCCU201901120

the overall patterns of networks over time. This method enables the researchers to explicitly understand representation of emerging themes.

DOI:10.6814/NCCU201901120

4. The experiment and results

4.1 Keywords network

Fig. 2 shows keywords network by co-occurrence (1993-2017). The nodes are the keywords. The size of nodes can reflect the frequency of keywords. Larger size of node means higher frequency of occurrence of keyword. The lines between two nodes stand for the associations of two keywords, or represent the co-occurrence of these keywords in a paper. The thickness of line indicates the co-occurrence frequency of keyword pairs, or represents the number of times each pair of keywords was mentioned together in papers. The thickness of line is proportional to the closeness of connections between two keywords. The thicker line between two keywords, the closer their relationship is. The more co-occurrence between two keywords, the closer their relationship is. It shows the strength of the connection.

According to Fig.2, we can see that keywords such as “market”, “prices”, “finance”, “liquidity”, “statistics”, “financial markets”, “stock”, “stochastic”, “model” and “trades” became important keywords, which means that they have played an important role in bridging other research themes.

DOI:10.6814/NCCU201901120

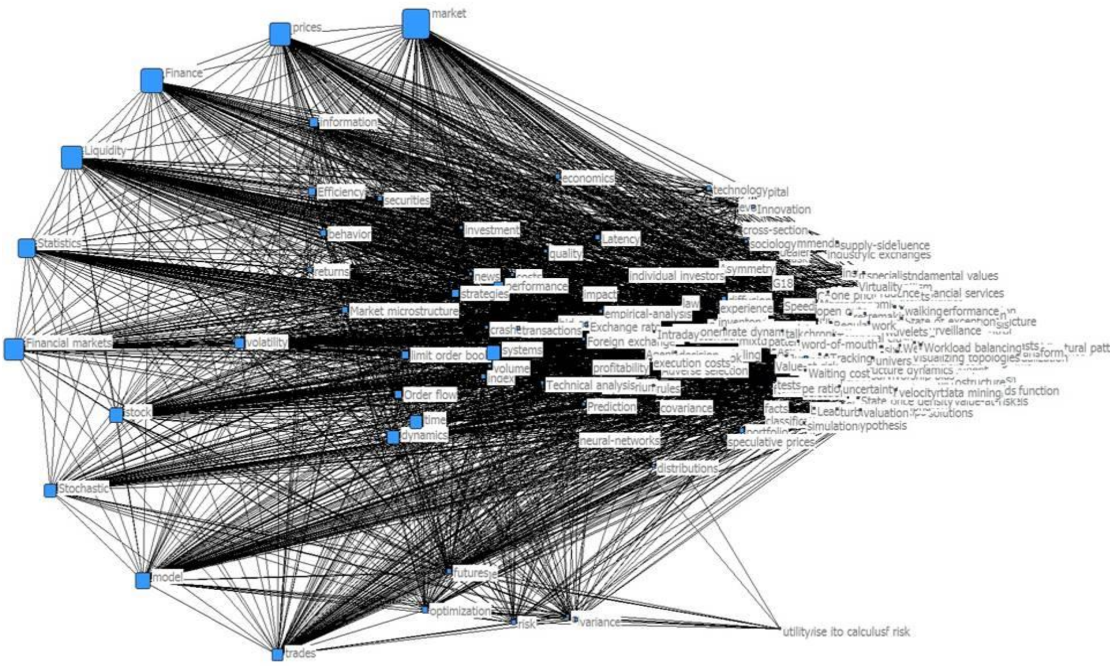


Fig. 2 Keywords network by co-occurrence (1993-2017)

4.2 Betweenness centrality measuring for all period (1993-2017)

Keywords serve as an indicator of the importance of the research themes they represent. The top ten keywords from betweenness centrality measuring for all period (1993-2017) are “market”, “prices”, “finance”, “liquidity”, “statistics”, “financial markets”, “stock”, “stochastic”, “model” and “trades” as shown in Table 3.

The results indicate that these research themes are major focuses and the bridges connecting to other research themes in HFT. These findings show that these research themes attract more attention and have a closer relationship with other research themes in HFT. Notice that keywords like “High Frequency Trading” and “Algorithm(s)” have very broad meanings. Due to this kind of keywords are meaningless for this study, we excluded them from the above analysis.

DOI:10.6814/NCCU201901120

Table 3 : Betweenness centrality measuring for all period (1993-2017)

Table 1:Betweenness centrality measuring for all period (1993-2017)

	1993 - 2017			1993 - 2017	
rank	Keywords		rank	Keywords	
1	High-frequency trading	20972.666	46	costs	479.408
2	Algorithms	14873.707	47	empirical-analysis	428.283
3	market	11343.833	48	quality	425.657
4	prices	9020.214	49	volume	425.474
5	Finance	8998.353	50	variance	420.01
6	Liquidity	8372.251	51	sociology	403.265
7	Statistics	6550.62	52	power	381.97
8	Financial markets	6365.009	53	Technical analysis	378.11
9	stock	5479.472	54	Foreign exchange	343.948
10	Stochastic	5255.224	55	universal portfolios	339.003
11	model	5207.57	56	impact	305.744
12	trades	4817.368	57	Exchange rate	291.833
13	systems	4643.722	58	Prediction	284.327
14	dynamics	4483.216	59	options	283.819
15	time	3957.8	60	equilibrium	253.83
16	volatility	3471.744	61	competition	241.319
17	management	3040.512	62	Agent-based modelling	209.288
18	information	2945.935	63	Innovation	207.791
19	Order flow	2761.069	64	neural-networks	194.051
20	strategies	2683.609	65	individual investors	186.9
21	performance	2651.219	66	covariance	172.821
22	Efficiency	2355.799	67	Content-based	161.023
23	behavior	1895.172	68	bid-ask spread	158.613
24	Market microstructure	1890.566	69	decision	150.691
25	optimization	1622.698	70	sharpe ratio	140.984
26	returns	1461.738	71	evolution	129.907
27	index	1404.703	72	profitability	124.693
28	limit order book	1389.61	73	selection	123.284
29	capital	1313.427	74	Manipulation	117.729
30	portfolio	1312.679	75	turbulence	115.056
31	arbitrage	1276.097	76	execution costs	98.269
32	Latency	1247.931	77	law	94.131
33	futures	1177.844	78	rules	85.183
34	securities	1130.502	79	Lead-lag relationship	82.599
35	technology	1065.972	80	exchange	76.936
36	risk	1020.679	81	Adverse selection	70.177
37	investment	974.707	82	Approximation	68.299
38	economics	806.251	83	experience	63.362
39	news	750.328	84	ask	51.969
40	transactions	713.42	85	Asymmetry	40.406
41	Automation	681.092	86	Codings	40.025
42	diffusion	652.382	87	dealer	39.832
43	distributions	584.501	88	classification	39.252
44	crashes	532.398	89	speculative prices	38.483
45	Online learning	501.978	90	Intraday	38.163

DOI:10.6814/NCCU201901120

4.3 Changes in important keywords over time

How have the important keywords changed over time and what are the recent important keywords? In order to trace dynamic changes of the HFT Research, the whole 24 year was further separated three consecutive periods: 1993-2002, 2003-2012, and 2013-2017. We constructed three keywords networks as shown in Fig. 3 to 5. For the full lists of keywords in these three periods, see Appendix A through Appendix C.

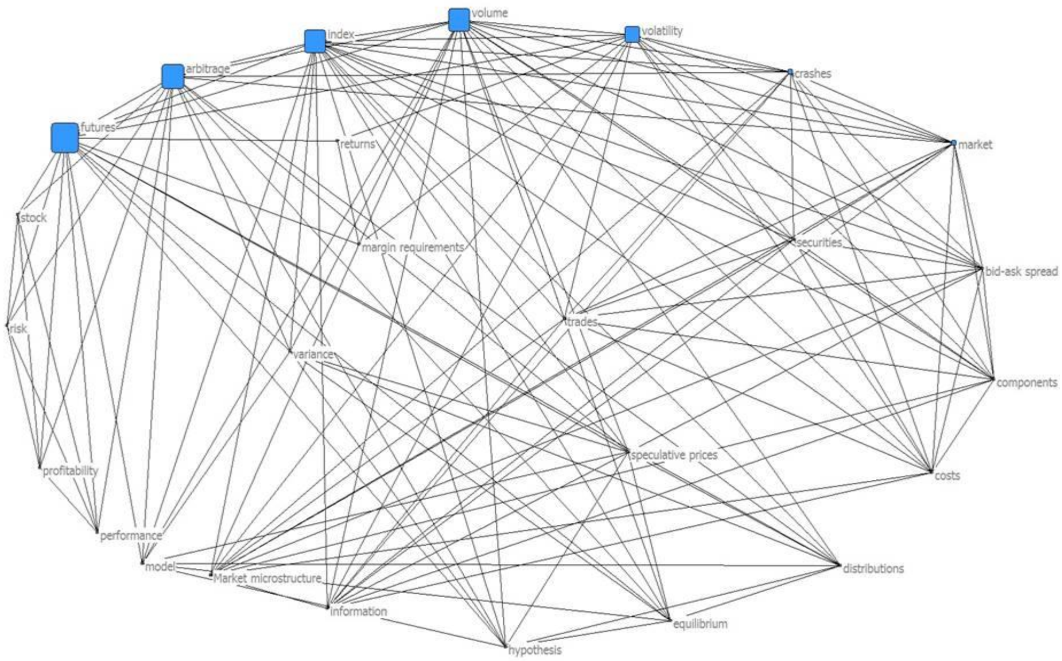


Fig. 3 Keywords network by co-occurrence (1993-2002)

DOI:10.6814/NCCU201901120

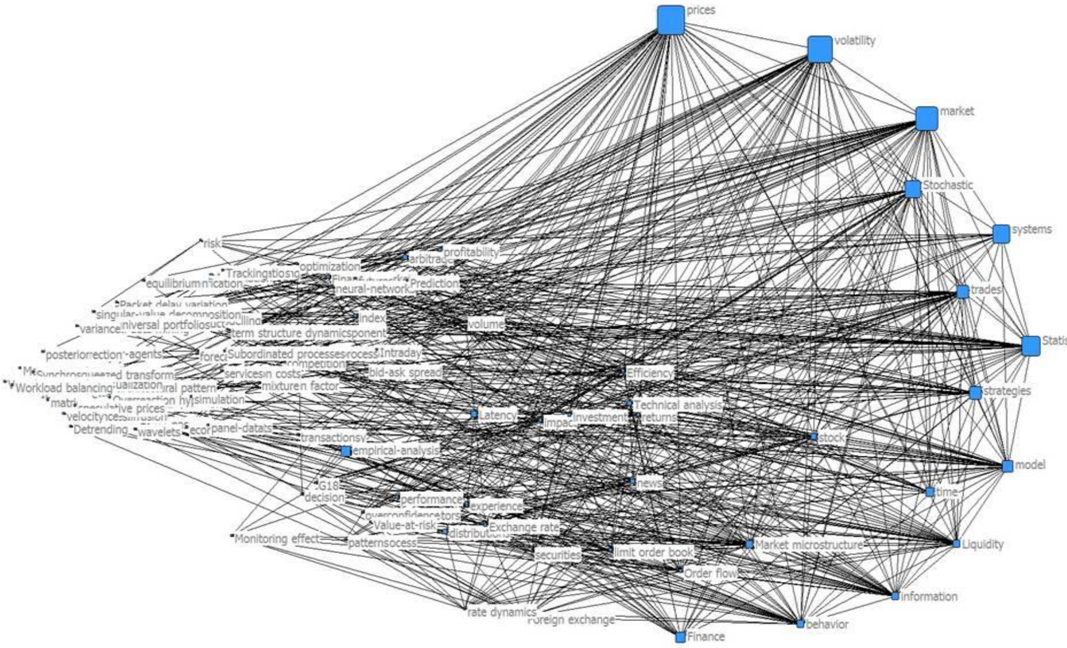


Fig. 4 Keywords network by co-occurrence (2003-2012)

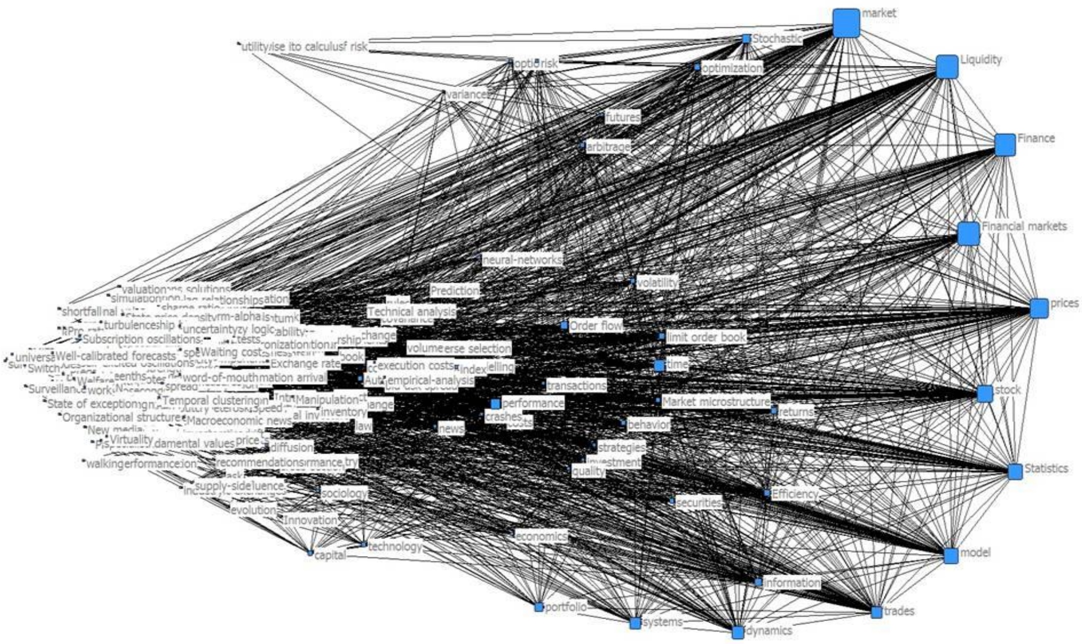


Fig. 5 Keywords network by co-occurrence (2013-2017)

DOI:10.6814/NCCU201901120

5. Discussions and implications

5.1 The findings of the quantitative analysis results

For further showing statistics of keywords network in different time slices, we compared the rank of the important keywords in the three keywords networks constructed as shown in Fig. 6 in order to thoroughly and precisely analyze the variations of trends. Please notice that the important keywords in Fig. 6 are from top ten keywords in Table 3. The top ten keywords from betweenness centrality measuring for all period (1993-2017) are “market”, “prices”, “finance”, “liquidity”, “statistics”, “financial markets”, “stock”, “stochastic”, “model” and “trades” as shown in Table 3.

The findings include following:

1) “Market” revealed to be the most important keyword by betweenness centrality measuring for all three periods, because it has received consistent upward attention.

2) “Stock” even received sharply upward attention since 1993 until 2017.

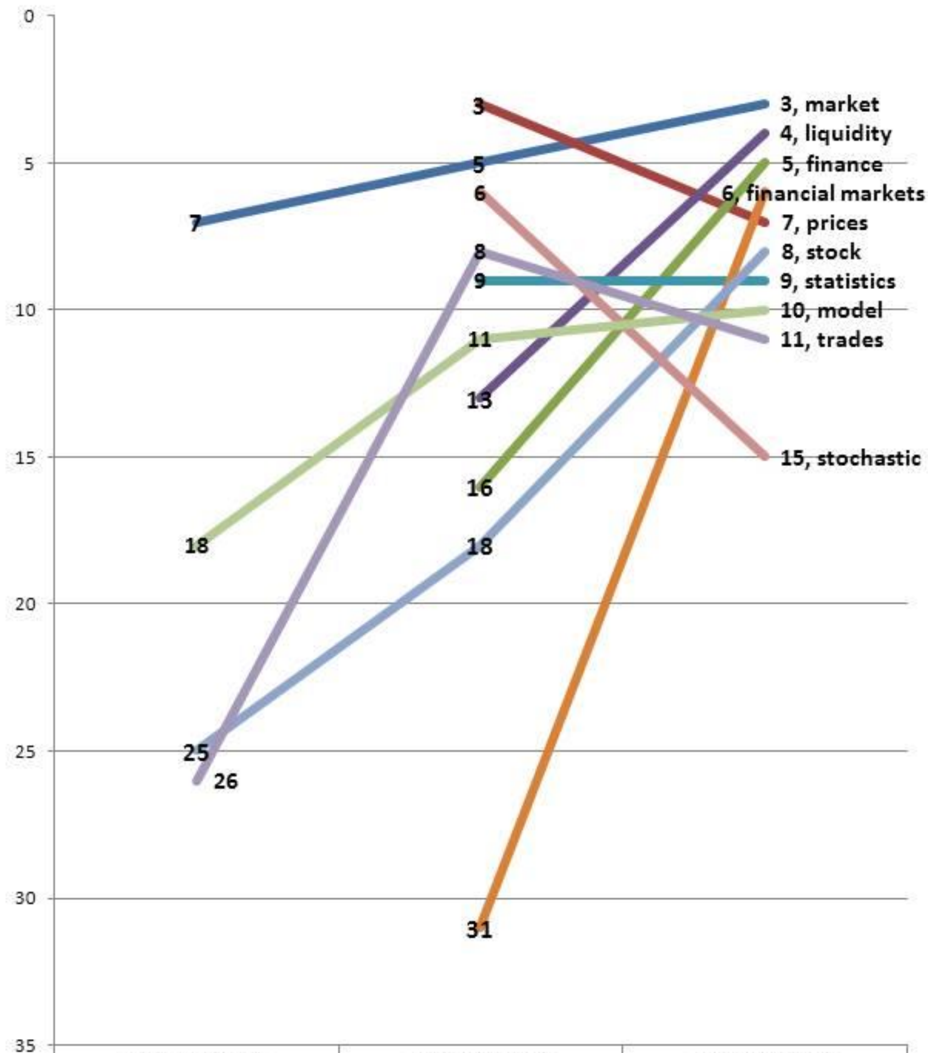
3) “Liquidity” and “finance” and “financial markets” are emerging theme since 2003 year due to they were not appeared in period of 1993 to 2002.

4) “Model” and “trades” were paid growing attention from 1993 through 2012 period, while 2013 to 2017 were not. One possible explanation may be that “model” and “trades” are viewed as common sense already in HFT research until recent years.

5) “Prices” and ”stochastic” emerged since 2003 year, but they were paid less attention from 2013 through 2017 period.

DOI:10.6814/NCCU201901120

The above analysis provides an overview of HFT research and it concludes that "market performance" related keywords, which represent some established research themes, have become the major focus in HFT research. It also changes rapidly to embrace new themes.



	1993 - 2002 rank	2003 - 2012 rank	2013 - 2017 rank
market	7	5	3
prices		3	7
finance		16	5
liquidity		13	4
statistics		9	9
financial markets		31	6
stock	25	18	8
stochastic		6	15
model	18	11	10
trades	26	8	11

Fig. 6 Changes in important keywords over time

DOI:10.6814/NCCU201901120

5.2 Comparison with previous qualitative research in HFT

On the other hand, according to previous chapters, "1.4 Current qualitative research in HFT" and "1.5 Research categories and findings in HFT", many attempts have been made to research HFT by a number of scholars in recent years. Goldstein, Kumar and Graves (2014)[114] reviewed the works in this area from the empirical and theoretical papers and assigned them into following six categories on the basis of different topics:

1) market performance (Budish, Cramton, & Shim, 2015; Menkveld, 2014; Schwartz & Wu, 2013; Jarnecic & Snape, 2014; Brogaard, Hendershott, & Riordan, 2014; Hendershott & Riordan, 2013; Popper, 2012c; Baron, Brogaard, & Kirilenko, 2012; Jones, 2013a; Hasbrouck & Saar, 2013; Credit Suisse, 2012),

2) strategies and practices (Aldridge, 2013; Kirilenko & Lo, 2013; Laughlin, Aguirre, & Grundfest, 2014),

3) evolution (Rubenstein, 2015; Goldstein, Kumar, & Graves, 2014; Kirilenko & Lo, 2013; Popper, 2012c),

4) speed (Angel, 2014; Laughlin, Aguirre, & Grundfest, 2014; Wissner-Gross & Freer, 2010; Brogaard, Hendershott, Hunt & Ysusi, 2014; Hasbrouck & Saar, 2013),

5) fairness (SEC, 2010; Narang, 2010; Patterson, Strasburg, & Pleven, 2013),

6) regulatory implications (Piwowar, 2013; SEC/CFTC, 2010; Westbrook, 2010).

However, those findings about HFT are largely qualitative in nature. what are the areas of focus in HFT? What are the developing trends in current research? Although there have been some review articles concerning HFT, for most of these works, it is not clear if merely depending on review articles would be capable of revealing the developing trends or

DOI:10.6814/NCCU201901120

future orientation of a new research field.

Now, there is agreement with previous work through this research which provides a quantitative analysis of global empirical and theoretical HFT papers. This research further points out that "market performance" related keywords, which represent some established research themes, have become the major focus in HFT research. It changes rapidly to embrace new themes. It can also help to evaluate the need for regulatory intervention or regulatory purposes.

These investigations can help researchers to realize the breadth of HFT research and to establish future research directions and to provide an entry point to any academic, regardless of their prior knowledge of the theme. These investigations can also help to get different point of view that emerge from a comprehensive review of existing HFT papers by future researchers before choosing their interested field.

DOI:10.6814/NCCU201901120

6. Conclusion

6.1 The original problem statement

HFT is an emerging, ever changing and rapidly evolving area with highly interdisciplinary in nature for the markets, regulators, and the public[115]. This diversity may root from the emerging nature of computing technology and its wide appeal as well as unique researcher and practitioner viewpoints. However, current research categories and findings about HFT are largely qualitative in nature. Although there have been some review articles concerning HFT, for most of these works, it is not clear if merely depending on review articles would be capable of revealing the developing trends or future orientation of a new research field. Therefore, my general purpose in this research is to provide a quantitative analysis of global empirical and theoretical HFT papers.

Those findings by topics shown in Table 1 represent the diverse issues and findings in the field of HFT which include the introduction of keywords and ideas and even new concepts about HFT. However, what are the areas of focus in HFT? What are the developing trends in current research?

6.2 Summarization

In this research, we used social network analysis (SNA) technique to give a comprehensive understanding of HFT research during 1993 to 2017. We obtain some clear and reasonable results which can provide useful insights to better understand evolution and emerging trends in HFT research.

The findings of the quantitative analysis indicate that HFT research has been strongly influenced by “market”, “prices”, “finance”, “liquidity”,

DOI:10.6814/NCCU201901120

"statistics", "financial markets", "stock", "stochastic", "model" and "trades", which represent some established research themes. They are major focuses and the bridges connecting to other research themes in HFT.

Comparing with previous qualitative research in HFT, this research further points out that "market performance" related keywords, which represent some established research themes, have become the major focus in HFT research. It also changes rapidly to embrace new themes.

These investigations can help researchers to realize the breadth of HFT research and to establish future research directions and to provide an entry point to any academic, regardless of their prior knowledge of the theme. These investigations can also help to get different point of view that emerge from a comprehensive review of existing HFT papers by future researchers before choosing their interested field.

6.3 Limitations

This research is just a preliminary and still has limitations need to be addressed. The main limitation of SNA technique is that it is just one of the tools that can be used to understand evolution and emerging trends in HFT research. It is just one piece of the puzzle. Subject matter experts are needed to provide a context for the research.

On the other hand, this study tries to explore the evolution and emerging trends in HFT papers published in world leading journals but the Web of Science database may not completely cover the scientific research of HFT.

6.4 Paths for future research

In the future, comparative research with other method in the same HFT field could also be explored because different methods may have very different research emphases which would also be worthy of further

DOI:10.6814/NCCU201901120

exploration to extend HFT research theme.

This study utilizes the advantage of SNA technique and such keywords analysis might be helpful to stimulate further research or identify some fruitful future research opportunities.

DOI:10.6814/NCCU201901120

Reference

[1] Brogaard, J., Carrion, A., Moyaert, T., Riordan, R., Shkilko, A., & Sokolov, K. (2018). High frequency trading and extreme price movements. *Journal of Financial Economics, 128*(2), 253-265.
[2] https://snipethetrade.com/us/high-frequency-trading
[3] https://en.wikipedia.org/wiki/High-frequency_trading
[4] https://www.investopedia.com/terms/h/high-frequency-trading.asp
[5] https://www.investopedia.com/ask/answers/09/high-frequency-trading.asp
[6] https://www.investopedia.com/terms/h/high-frequency-trading.asp
[7] Kumar, P., Goldstein, M., Graves, F., & Borucki, L. (2011). Trading at the speed of Light. *The Brattle Group Finance (2)*.
[8] Aldridge, I. (2013). *High-frequency trading: a practical guide to algorithmic strategies and trading systems* (Vol. 604). John Wiley & Sons. ISBN 978-0-470-56376-2
[9] https://www.sec.gov/fast-answers/answersmktmakerhtm.html
[10] Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity?. *The Journal of Finance*, 66(1), 1-33.
[11] Cartea, A., & Jaimungal, S. (2013). Modelling asset prices for algorithmic and high-frequency trading. *Applied Mathematical Finance*, 20(6), 512-547.
[12] Guéant, O., Lehalle, C. A., & Fernandez-Tapia, J. (2013). Dealing with the inventory risk: a solution to the market making problem. *Mathematics and financial economics*, 7(4), 477-507.
[13] Jovanovic, B., & Menkveld, A. J. (2016). Middlemen in limit order markets. *Available at SSRN 1624329*.
[14] Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of financial Markets*, 16(4), 712-740.
[15] Weil, J. (2014). Weil on Finance: FBI Hops on Michael Lewis Bandwagon. *Bloomberg. URL: https://www.bloomberg.com/opinion/articles/2014-04-01/weil-on-finance-fbi-hops-on-michael-lewis-bandwagon*
[16] Bradford, H. (2014). FBI Investigating High-Frequency Traders: WSJ. *Huffington Post. URL: https://www.huffingtonpost.com/2014/03/31/fbi-high-speedtrading_n_5065622.html (дата звернення: 19.03. 2018).*
[17] https://www.coursehero.com/file/p32cbud/The-World-Of-High-Frequency-Trading-6-Primary-Strategies/
[18] Gregoriou, G. N. (Ed.). (2015). *Handbook of high frequency trading*. Academic Press. p. 28. ISBN 978-0-1280-2362-4.
[19] Bozdog, D., Florescu, I., Khashanah, K., & Wang, J. (2011). Rare Events Analysis for High-Frequency Equity Data. *Wilmott, 2011*(54), 74-81.
[20] Iati, R. (2009). The real story of trading software espionage. *Advanced Trading*.
[21] Cookson, C. (2013). Time is money when it comes to microwaves. *Financial Times, 10*.
[22] Cookson, C. (2013). Time is money when it comes to microwaves. *Financial Times, 10*.
[23] https://mechanicalmarkets.wordpress.com/2015/01/20/identifying-trader-type-pt-2/
[24] https://mechanicalmarkets.wordpress.com/2015/02/09/order-size-in-the-hft-era-identifying-trade-type-pt-3/
[25] https://www.investopedia.com/ask/answers/09/high-frequency-trading.asp

DOI:10.6814/NCCU201901120

[26] Milnor, J. W., & Randall, G. A. (1931). The Newfoundland-Azores high-speed duplex cable. *Transactions of the American Institute of Electrical Engineers*, 50(2), 389-396.
[27] Aldridge, I. (2013). *High-frequency trading: a practical guide to algorithmic strategies and trading systems* (Vol. 604). John Wiley & Sons.
[28] Haldane, A. (2010). Patience and finance, speech to the Oxford China Business Forum. *Richard Murphy*. https://www.bis.org/review/r100909e.pdf
[29] U.S. Securities and Exchange Commission (2010), Part III: Concept release on equity market structure; Proposed Rule, 17 CFR Part 242, *Federal Register* 75(13) 3594–3614 : Jan. 14 https://www.sec.gov/rules/concept/2010/34-61358.pdf
[30] Duhigg, C. (2009), Stock Traders Find Speed Pays, in Milliseconds, *The New York Times*, July 23. http://www.nytimes.com/2009/07/24/business/24trading.html
[31] Stafford, P. (2013). Italy introduces tax on high-speed trade and equity derivatives. *Financial Times*, 1.
[32] Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance, 72*(3), 967-998.
[33] https://en.wikipedia.org/wiki/High-frequency_trading
[34] Aldridge, I., & Krawciw, S. (2017). *Real-time risk: What investors should know about FinTech, high-frequency trading, and flash crashes*. John Wiley & Sons. ISBN 978-1-119-31896-5
[35] Easley, D., De Prado, M. M. L., & O'Hara, M. (2011). The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. *The Journal of Portfolio Management*, 37(2), 118-128.
[36] IOSCO, J. (2011). Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency. *Consultation Report*.
[37] Vuorenmaa, T. A., & Wang, L. (2014). An agent-based model of the flash crash of may 6, 2010, with policy implications. *Available at SSRN 2336772*.
[38] Ross, A., Fitzgibbon, W., & Mathiason, N. (2012). Britain opposes MEPs seeking ban on high-frequency trading. UK fighting efforts to curb high-risk, volatile system, with industry lobby dominating advice given to Treasury. *The Guardian*.
[39] https://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf
[40] https://www.nasdaq.com/investing/glossary/h/high-frequency-trading
[41] https://en.wikipedia.org/wiki/High-frequency_trading
[42] https://www.investopedia.com/ask/answers/09/high-frequency-trading.asp
[43] Goldstein, M. A. (2014), Special Issue on Computerized and High-Frequency Trading: Guest Editor's Note, *The Financial Review*, Vol. 49, Issue 2, 173–175 : May 2014
[44] Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014
[45] Budish, E. B., P. Cramton, & J. J. Shim (2015), The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response, *Chicago Booth Research Paper*, No. 14-03 : Feb. 2015
[46] Menkveld, A. J. (2014), High-Frequency Traders and Market Structure, *The Financial Review*, Vol. 49, Issue 2, 333–344 : May 2014
[47] Schwartz, R. & L.Wu (2013), Equity trading in the fast lane: The Staccato Alternative, *Journal of Portfolio Management*, Vol. 39, No. 3, 3–6 : Spring 2013
[48] Jarnecic, E. & M. Snape (2014), The Provision of Liquidity by High-Frequency Participants, *The Financial Review*, Vol. 49, Issue 2, 371–394 : May 2014
[49] Brogaard, J., T. Hendershott & R. Riordan (2014), High-Frequency Trading and

DOI:10.6814/NCCU201901120

Price Discovery, *The Review of Financial Studies*, Vol. 27, Issue 8, 2267–2306 : May 2014

[50] Hendershott, T. & R. Riordan (2013), Algorithmic Trading and the Market for Liquidity, *Journal of Financial and Quantitative Analysis*, Vol. 48, No. 4, 1001–1024 : Aug. 2013

[51] Popper, N. (2012c), High-speed trading no longer hurtling forward, *The New York Times*, Oct. 14.

[52] Baron, M., J. Brogaard, & A. Kirilenko (2012), The trading profits of high-frequency traders. *Working paper*, University of Washington

[53] Jones, C. (2013a), What do we know about high-frequency trading? *Columbia Business School Research Paper*, No. 13–11

[54] Hasbrouck, J. & G. Saar (2013), Low-latency trading, *Journal of Financial Markets*, Vol. 16, Issue 4, 646–679 : Nov. 2013

[55] Credit Suisse (2012), Who let the bots out? Market quality in a high frequency world https://www.managedfunds.org/wp-content/uploads/2012/11/HFT.pdf

[56] Aldridge, I. (2013), *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems, 2nd Edition* (John Wiley and Sons, Inc., Hoboken, NJ).

[57] Kirilenko, Andrei A., & Andrew W. Lo (2013), Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents, *Journal of Economic Perspectives*, Vol. 27, No. 2, 51-72 : Spring 2013

[58] Laughlin, G., A. Aguirre, & Grundfest (2014), Information Transmission between Financial Markets in Chicago and New York, *The Financial Review*, Vol. 49, Issue 2, 283–312 : May 2014

[59] Rubenstein, Ari. (2015), This High-Speed Trader Says Thanks, Regulators. *The Wall Street Journal*, April 23, C1.

[60] Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014

[61] Kirilenko, Andrei A., & Andrew W. Lo (2013), Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents, *Journal of Economic Perspectives*, Vol. 27, No. 2, 51-72 : Spring 2013

[62] Popper, N. (2012c), High-speed trading no longer hurtling forward, *The New York Times*, Oct. 14.

[63] Angel, J. J. (2014), When Finance Meets Physics: The Impact of the Speed of Light on Financial Markets and Their Regulation, *The Financial Review*, Vol. 49, Issue 2, 271–281 : May 2014

[64] Laughlin, G., A. Aguirre, & Grundfest (2014), Information Transmission between Financial Markets in Chicago and New York, *The Financial Review*, Vol. 49, Issue 2, 283–312 : May 2014

[65] Wissner-Gross, A. D. & C. E. Freer (2010), Relativistic statistical arbitrage, *Physical Review*, E 82, 056104.

[66] Brogaard, J., T. Hendershott, S. Hunt, & C. Ysusi (2014), High-frequency trading and the execution costs of institutional investors, *The Financial Review*, Vol. 49, Issue 2, 345–369 : May 2014

[67] Hasbrouck, J. & G. Saar (2013), Low-latency trading, *Journal of Financial Markets*, Vol. 16, Issue 4, 646–679 : Nov. 2013

[68] U.S. Securities and Exchange Commission (2010), Part III: Concept release on equity market structure; Proposed Rule, 17 CFR Part 242, *Federal Register* 75(13) 3594–3614 : Jan. 14 https://www.sec.gov/rules/concept/2010/34-61358.pdf

DOI:10.6814/NCCU201901120

69 Narang, M. (2010), Tradeworx, Inc. Public commentary on SEC market structure concept release, SEC Comment letter. https://www.sec.gov/comments/s7-02-10/s70210-129.pdf
70 Patterson, S., J. Strasburg, & L. Pleven (2013), High-speed traders exploit loophole, *The Wall Street Journal*, May 1. http://www.wsj.com/articles/SB10001424127887323798104578455032466082920
71 Piwowar, M. (2013), The benefit of hindsight and the promise of foresight: A proposal for a comprehensive review of equity market structure. U.S. Securities and Exchange Commission, Speech in London, England, Dec. 9. http://www.sec.gov/News/Speech/Detail/Speech/1370540470552
72 Westbrook, J. (2010), SEC Vote Shows Scope of High-Frequency Trading Rules (Update2), *Bloomberg*, Jan. 13. http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aiP2T5Sx9Nq4
73 Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014
74 Chung, K. H., & Lee, A. J. (2016). High-frequency Trading: Review of the Literature and Regulatory Initiatives around the World. *Asia-Pacific Journal of Financial Studies, 45*(1), 7-33.
75 Wellman, B., & Berkowitz, S. D. (Eds.). (1988). *Social structures: A network approach* (Vol. 2). (Cambridge : Cambridge University Press, 1988, Archive).
76 https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/
77 Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014
78 Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, 28(6), 441-453.
79 https://en.wikipedia.org/wiki/Social_network_analysis
80 https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/
81 https://en.wikipedia.org/wiki/Social_network_analysis_software
82 https://cambridge-intelligence.com/keylines/
83 Pritchard, A. (1969), Statistical bibliography or bibliometrics? *Journal of Documentation*, 25(4), 348-349 : Dec. 1969
84 https://www.ecoom.be/en/research/bibliometrics
85 García-García, P., López-Muñoz, F., Callejo, J., Martín-Águeda, B., & Álamo, C. (2005). Evolution of Spanish scientific production in international obstetrics and gynecology journals during the period 1986–2002. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 123*(2), 150-156.
86 https://www.surrey.ac.uk/library/documents/sridocs/Quick%20Introduction%20to%20Bibliometrics%20for%20Researchers%20-%20information%20sheet.pdf
87 Goffman, W., & Warren, K. S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures. *Nature, 221*(5187), 1205.
88 https://www.ischool.utexas.edu/~donturn/research/bibweb.html
89 Bradford, S. C. (1934), Sources of information on specific subjects, *Engineering: An Illustrated Weekly Journal*, Vol. 137, 85-86 : Jan. 1934
90 https://www.ischool.utexas.edu/~donturn/research/bibweb.html
91 Braun, T., W. Glanzel, & H. Grupp (1995), The scientometric weight of 50 nations in 27 science areas, 1989–1993. Part I. All fields combined, mathematics,

DOI:10.6814/NCCU201901120

engineering, chemistry and physics, *Scientometrics*, Vol. 33, Issue 3, 263–293 : Jul.–Aug. 1995
[92] Xie, S. D., J. Zhang, & Y. S. Ho (2008), Assessment of world aerosol research trends by bibliometric analysis, *Scientometrics*, Vol. 77, Issue 1, 113–130 : Jul. 2008
[93] Colman, A. M., D. Dhillon, & B. Coulthard (1995), A bibliometric evaluation of the research performance of British university politics departments: Publications in leading journals, *Scientometrics*, Vol. 32, Issue 1, 49–66 : Jan. 1995
[94] Ugolini, D., S. Parodi, L. Santi (1997), Analysis of publication quality in a Cancer Research Institute, *Scientometrics*, Vol. 38, Issue 2, 265–274 : Feb. 1997
[95] Li, Z. & Y. S. Ho (2008), Use of citation per publication as an indicator to evaluate contingent valuation research, *Scientometrics*, Vol. 75, Issue 1, 97–110 : Apr. 2008
[96] Li, L. L., G. H. Ding, N. Feng, M. H. Wang, & Y. S. Ho (2009b), Global stem cell research trend: bibliometric analysis as a tool for mapping of trends from 1991 to 2006, *Scientometrics*, Vol. 80, Issue 1, 39–58 : Jul. 2009
[97] Ugolini, D., M. A. Cimmino, C. Casilli, & G. S. Mela (2001), How the European Union writes about ophthalmology, *Scientometrics*, Vol. 52, Issue 1, 45–58 : Sep. 2001
[98] Qin, J. (2000), Semantic similarities between a keyword database and a controlled vocabulary database: an investigation in the antibiotic resistance literature, *Journal of the American Society for Information Science*, Vol. 51, Issue 2, 166–180 : Jan. 2000
[99] Zhang, G. F., S. D. Xie, & Y. S. Ho (2010), A bibliometric analysis of world volatile organic compounds research trends, *Scientometrics*, Vol. 83, Issue 2, 477–492 : May 2010
[100] Xie, S. D., J. Zhang, & Y. S. Ho (2008), Assessment of world aerosol research trends by bibliometric analysis, *Scientometrics*, Vol. 77, Issue 1, 113–130 : Jul. 2008
[101] Zhang, G. F., S. D. Xie, & Y. S. Ho (2010), A bibliometric analysis of world volatile organic compounds research trends, *Scientometrics*, Vol. 83, Issue 2, 477–492 : May 2010
[102] Garfield, E. (1990) *KeyWords Plus®:* ISI®'s breakthrough retrieval method. Part I. Expanding your searching power on *Current Contents on Diskette®. Current Contents* (32):5-9, 6 August 1990. (Reprinted in: *Essays of an Information Scientist.*) Philadelphia: ISI Press®, 1991. Vol. 13. 295-9.
[103] Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014
[104] Drake, M. & M. Dekker (2003), *Encyclopedia of Library and Information Science, Second Edition*, (CRC Press, New York, N.Y.)
[105] Web of Science (2014), *Thomson Reuters*
[106] Aghaei Chadegani, A., Salehi, H., Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science, 9*(5), 18-26.
[107] Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology, 67*(4), 967-972.
[108] Chiu,W.T. & Y. S. Ho (2007), Bibliometric analysis of tsunami research. *Scientometrics*, Vol. 73, Issue 1, 3–17 : Oct. 2007
[109] Li, J.F., Y. H. Zhang , X. S. Wang & Y. S. Ho (2009a), Bibliometric analysis of atmospheric simulation trends in meteorology and atmospheric science journals, *Croatica Chemica Acta*, Vol. 82 Issue 3, 695–705 : Dec. 2009

DOI:10.6814/NCCU201901120

[110] Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis. Harvard, MA: Analytic Technologies.
[111] Larsen, E. N. (1986). An introduction to structural analysis: The network approach to social research: SD Berkowitz, Toronto: Butterworth, 1982.
[112] https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/
[113] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.
[114] Goldstein, M. A., P. Kumar, & F. C. Graves (2014), Computerized and High-Frequency Trading, *The Financial Review*, Vol. 49, Issue 2, 177–202 : May 2014
[115] Chung, K. H., & Lee, A. J. (2016). High-frequency Trading: Review of the Literature and Regulatory Initiatives around the World. *Asia-Pacific Journal of Financial Studies, 45*(1), 7-33.

DOI:10.6814/NCCU201901120

Appendix A

Betweenness centrality measuring for the first sub-period (1993-2002)

		1993 - 2002	
No.	rank	Keywords	
8	1	futures	91.474
1	2	arbitrage	69.006
11	3	index	68.29
27	4	volume	68.29
26	5	volatility	46.118
5	6	crashes	9.371
14	7	market	9.371
2	8	bid-ask spread	0
3	9	components	0
4	10	costs	0
6	11	distributions	0
7	12	equilibrium	0
9	13	High-frequency trading	0
10	14	hypothesis	0
12	15	information	0
13	16	margin requirements	0
15	17	Market microstructure	0
16	18	model	0
17	19	performance	0
18	20	profitability	0
19	21	returns	0
20	22	risk	0
21	23	securities	0
22	24	speculative prices	0
23	25	stock	0
24	26	trades	0
25	27	variance	0

DOI:10.6814/NCCU201901120

Appendix B

Betweenness centrality measuring for the second sub-period (2003-2012)

		2003 - 2012				2003 - 2012	
No.	rank	Keywords		No.	rank	Keywords	
5	1	Algorithms	2723.209	119	46	risk	21.541
68	2	High-frequency trading	1947.709	44	47	economics	21.067
113	3	prices	1943.382	62	48	futures	19.741
153	4	volatility	1592.97	13	49	bid-ask spread	18.193
86	5	market	1466.067	149	50	Value-at-risk	15.594
129	6	Stochastic	1124.416	75	51	Intraday	12.218
136	7	systems	1108.412	15	52	Boosting	2.133
144	8	trades	844.66	1	53	1st passage	0
128	9	Statistics	844.355	2	54	Active measurement	0
131	10	strategies	776.203	3	55	Adaptive trader-agents	0
92	11	model	714.488	4	56	Agent-based modelling	0
141	12	time	550.082	6	57	amorphous solids	0
84	13	Liquidity	393.819	7	58	anomalous diffusion	0
73	14	information	387.353	8	59	Approximation	0
12	15	behavior	379.06	10	60	Asynchronous data	0
55	16	Finance	330.972	14	61	Binary classification	0
87	17	Market microstructure	305.516	16	62	C33	0
130	18	stock	276.927	17	63	C41	0
41	19	distributions	198.327	18	64	C50	0
137	20	Technical analysis	190.978	19	65	cascades	0
95	21	news	186.846	20	66	choice	0
80	22	Latency	164.73	21	67	classification	0
9	23	arbitrage	163.332	22	68	Cloud computing	0
45	24	Efficiency	135.079	23	69	Codes of conduct	0
46	25	empirical-analysis	134.262	24	70	Codings	0
100	26	Order flow	130.235	25	71	Commodity hardware	0
58	27	Foreign exchange	128.547	26	72	Common factor	0
76	28	investment	128.284	27	73	Commonality	0
118	29	returns	114.102	28	74	competition	0
83	30	limit order book	109.02	29	75	component analysis	0
56	31	Financial markets	97.433	30	76	components	0
107	32	performance	87.288	31	77	continuous double auction	0
71	33	index	56.19	32	78	costs	0
115	34	profitability	56.01	33	79	covariance	0
36	35	decision	55.881	34	80	crashes	0
52	36	experience	54.663	35	81	Data stream processing	0
50	37	Exchange rate	52.486	37	82	Detrending	0
70	38	impact	50.131	38	83	diffusion	0
98	39	optimization	45.131	39	84	disposition	0
121	40	securities	34.477	40	85	Distributed processing	0
116	41	rate dynamics	31.575	42	86	dynamics	0
154	42	volume	30.481	43	87	EaaS	0
11	43	Automation	26.278	47	88	equilibrium	0
112	44	Prediction	26.052	48	89	error-correction	0
94	45	neural-networks	22.656	49	90	evolution	0

DOI:10.6814/NCCU201901120

Appendix C

Betweenness centrality measuring for the third sub-period (2013-2017)

		2013 - 2017				2013 - 2017	
No.	rank	Keywords		No.	rank	Keywords	
163	1	High-frequency trading	15349.117	255	46	options	322.945
10	2	Algorithms	9033.231	277	47	power	320.466
217	3	market	8454.144	248	48	news	298.916
211	4	Liquidity	7070.412	365	49	volume	244.794
143	5	Finance	6733.11	362	50	variance	210.176
144	6	Financial markets	5455.364	246	51	neural-networks	206.152
280	7	prices	5301.453	7	52	Agent-based modelling	185.942
333	8	stock	5137.49	279	53	Prediction	181.464
331	9	Statistics	4706.375	62	54	competition	168.874
234	10	model	3940.322	171	55	impact	158.898
353	11	trades	3479.517	74	56	Content-based	144.998
111	12	dynamics	3275.315	309	57	selection	135.227
346	13	systems	2731.716	203	58	law	107.017
275	14	portfolio	2696.57	216	59	Manipulation	104.194
332	15	Stochastic	2543.021	120	60	empirical-analysis	97.351
215	16	management	2533.632	305	61	rules	95.875
256	17	Order flow	2276.406	83	62	covariance	85.314
352	18	time	2174.244	181	63	Innovation	85.282
180	19	information	2104.56	127	64	equilibrium	79.572
269	20	performance	1684.337	178	65	individual investors	77.577
254	21	optimization	1560.585	134	66	exchange	75.6
117	22	Efficiency	1499.502	33	67	bid-ask spread	74.359
44	23	capital	1204.612	136	68	execution costs	68.596
297	24	returns	1129.319	6	69	Adverse selection	64.104
31	25	behavior	1043.882	315	70	sharpe ratio	58.984
210	26	limit order book	981.42	348	71	Technical analysis	53.715
349	27	technology	929.715	17	72	ask	52.004
218	28	Market microstructure	885.26	135	73	Exchange rate	48.534
175	29	index	855.817	22	74	Asymmetry	45.24
364	30	volatility	833.319	139	75	facts	42.178
14	31	arbitrage	755.86	94	76	dealer	35.465
299	32	risk	752.568	38	77	book	33.13
308	33	securities	745.66	97	78	decision	31.873
191	34	investment	711.843	132	79	evolution	31.071
155	35	futures	635.551	192	80	issues	27.821
354	36	transactions	618.743	327	81	spread	23.048
334	37	strategies	590.264	106	82	discovery	22.219
202	38	Latency	579.702	133	83	Evolutionary computation	20.034
24	39	Automation	509.524	236	84	Momentum	18.852
103	40	diffusion	495.759	50	85	classification	17.174
116	41	economics	456.811	245	86	networks	16.995
285	42	quality	415.741	190	87	Inventory risk	16.062
321	43	sociology	409.809	8	88	aggressiveness	15.185
82	44	costs	401.178	123	89	entropy	14.797
84	45	crashes	334.673	89	90	cross-section	14.076

DOI:10.6814/NCCU201901120