Asset Pricing and Network Topology

Mr. Harnchai Eng-uthaiwat

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Economics
Faculty of Economics
Chulalongkorn University
Academic Year 2017

การศึกษาราคาของสินทรัพย์ด้วยทฤษฎีการวิเคราะห์แบบโครงข่าย

นายหาญชัย เอ็งอุทัยวัฒน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาเศรษฐศาสตรดุษฎีบัณฑิต
สาขาวิชาเศรษฐศาสตร์
คณะเศรษฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2560

Thesis Title	Asset Pricing and Network Topology
By	Mr. Harnchai Eng-uthaiwat
Field of Study	Economics
Thesis Advisor	Assistant Professor Pongsak Luangaram, Ph.D.
Thesis Co-Advisor	Associate Professor Tanapong Potipiti, Ph.D.

Accepted by the Faculty of Economics, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.... Dean of the Faculty of Economics
(Professor Worawet Suwanrada, Ph.D.)

THESIS COMMITTEE

.... Chairman
(Associate Professor Phornchanok Cumperayot kouwenberg, Ph.D.)

.... Thesis Advisor
(Assistant Professor Pongsak Luangaram, Ph.D.)

.... Thesis Co-Advisor
(Associate Professor Tanapong Potipiti, Ph.D.)

.... Examiner
(Assistant Professor San Sampattavanija, Ph.D.)

.... Examiner
(Assistant Professor Thaisiri Watewai, Ph.D.)

.... External Examiner
(Associate Professor Yuthana Sethapramote, Ph.D.)

หาญชัย เอ็งอุทัยวัฒน์ : การศึกษาราคาของสินทรัพย์ด้วยทฤษฎีการวิเคราะห์แบบโครงข่าย (Asset Pricing and Network Topology) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.พงศ์ศักดิ์ เหลืองอร่าม, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร.ธนะพงษ์ โพธิปิติ, 110 หน้า.

วัตถุประสงค์หลักของวิทยานิพนธ์นี้ คือการนำเสนอหลักฐานเชิงประจักษ์ของความสัมพันธ์ระหว่างราคาของสินทรัพย์และความเชื่อมโยงทางโครงสร้างของสินทรัพย์ (network structure) ในประเทศอเมริกา ตลาดระหว่างประเทศ และประเทศไทย การศึกษาที่ผ่านมาบ่งชี้ว่าโครงสร้างของสินทรัพย์ที่สร้างจากสหสัมพันธ์ของราคาสามารถจัดกลุ่มสินทัรพย์ที่มีความสัมพันธ์ทางเศษรฐศาสตร์ได้อย่างถูกต้อง ซึ่งทำให้โครงสร้างของสินทรัพย์มีประโยชน์ในการเป็นช่องทางการส่งผ่านความเสี่ยงจำเพาะ (idiosyncratic risk) ไปยังสมาชิกอื่นๆ ในระบบ โดยที่ความเชื่อมโยงทางโครงสร้างของสินทรัพย์จะทำหน้าที่กำจัดความเสี่ยงจำเพาะที่จะมีผลต่อระบบโดยรวม หรืออาจจะทำหน้าที่ในการขยายความเสี่ยงจำเพาะไปจนกระทั่งมีผลกับระบบ ดังนั้นความเสี่ยงจำเพาะซึ่งไม่มีความสำคัญในโมเดลการคำนวณราคาของสินทรัพย์ (capital asset pricing model) จึงยังอาจจะมีผลกับราคาของสินทรัพย์ผ่านทางโครงสร้างของสินทรัพย์ งานวิจัยนี้ได้ประยุกต์ใช้แนวคิดการส่งผ่านความเสี่ยงจำเพาะทางโครงสร้างสินทรัพย์กับโมเดลการประเมินราคาของสินทรัพย์ โดยเฉพาะอย่างยิ่งผลกระทบของการเปลี่ยนแปลงลักษณะทางโครงสร้างของสินทรัพย์ต่อผลตอบแทนของสินทรัพย์

งานวิจัยนี้สร้างโครงสร้างของสินทรัพย์โดยใช้เทคนิคการแยกแก่นแท้ของข้อมูลออกจากเส้นเชื่อมความสัมพันธ์ระหว่างผลตอบแทนของสมาชิกของเครือข่าย (Pearson correlation of return) ซึ่งวิธีดังกล่าวมีข้อดีในการลดความสัมพันธ์ที่ซับซ้อนของสมาชิก และเมื่อนำข้อมูลดังกล่าวมาประกอบเป็นเน็ตเวิร์กหรือโครงสร้างของสินทรัพย์ เราจะสามารถวัดคุณลักษณะต่างๆ ทางโครงสร้างของระบบได้ ซึ่งงานวิจัยนี้ศึกษาสองคุณลักษณะกล่าวคือ ลักษณะการเชื่อมต่อของเน็ตเวิร์ก (network topology) และ ระดับความเป็นศูนย์กลางของสมาชิก (centrality) โดยที่ลักษณะการเชื่อมต่อของเน็ตเวิร์กมีหลายรูปแบบตั้งแต่การกระจุกตัวกันของเส้นเชื่อมความสัมพันธ์ในรูปแบบ star network ไปจนถึงการกระจายตัวเป็นสายโซ่ในรูปแบบ chain network ซึ่งแต่ละรูปแบบจะมีผลต่างกันต่อความสามารถในการส่งผ่านความเสี่ยงจำเพาะ อีกคุณลักษณะของเน็ตเวิร์กในงานวิจัยนี้คือระดับความเป็นศูนย์กลางของสมาชิก (centrality) ซึ่งวัดอิทธิผลของสมาชิกในระบบทั้งในด้านการส่งผ่านความเสี่ยงจำเพาะไปยังสมาชิกอื่นๆและความเปราะบางต่อความเสี่ยงจำเพาะที่ถูกส่งผ่านมาในระบบ

ผลการศึกษาที่สำคัญของงานวิจัยนี้คือคุณสมบัติทางโครงสร้างของสินทรัพย์มีผลอย่างมีนัยสำคัญทางสถิติต่อราคาของสินทรัพย์ โดยที่ในการศึกษาดัชนี S&P500 ของประเทศสหรัฐอเมริกาพบว่า ลักษณะการเชื่อมต่อของบริษัทในตลาด (network topology) มีผลต่อผลตอบแทนของตลาดในอนาคต โดยทำหน้าที่เป็นช่องทางการส่งผ่านความเสี่ยงจำเพาะ นอกจากนั้นลักษณะการเชื่อมต่อของเน็ตเวิร์กยังช่วยในการคาดการณ์ผลตอบแทนเชิงลบระดับสูงในตลาดระหว่างประเทศ หลักฐานเชิงประจักษ์สุดท้ายของงานการศึกษานี้คือการศึกษาผลกระทบของเน็ตเวิร์กในตลาดประเทศไทย ซึ่งพบว่าสมาชิกที่มีระดับความเป็นศูนย์กลางสูงในด้านการส่งผ่านความเสี่ยงจำเพาะจะมีผลตอบแทนน้อยกว่าสมาชิกที่มีระดับความเป็นศูนย์กลางต่ำ และสมาชิกที่มีความเปราะบางสูงจะมีผลตอบแทนมากกว่าสมาชิกที่มีความเปราะบางต่ำ นอกจากนี้งานวิจัยนี้ยังพบว่าระดับความเป็นศูนย์กลางของสมาชิกอาจจะเป็นปัจจัยเสี่ยร่วม (common risk factor) ที่มีประโยชน์ในการคำนวณหาผลตอบแทนของสินทรัพย์ในตลาดประเทศไทย

สาขาวิชา เศรษฐศาสตร์	ลายมือชื่อนิสิต ....
ปีการศึกษา 2560	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
	ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

5585904229 : MAJOR ECONOMICS

KEYWORDS: STOCK MARKET NETWORK / NETWORK TOPOLOGY / CENTRALITY / IDIOSYNCRATIC RISK / ASSET PRICING / EXTREME NEGATIVE RETURNS

HARNCHAI ENG-UTHAIWAT: Asset Pricing and Network Topology. ADVISOR: ASST. PROF. PONGSAK LUANGARAM, Ph.D., CO-ADVISOR: ASSOC. PROF. TANAPONG POTIPITI, Ph.D., 110 pp.

This dissertation provides new empirical evidence for the relationship between network structure and asset returns in US, international markets, and Thailand. Previous studies find that a network of return correlations provides the meaningful economic taxonomy of the equity market. This finding makes the network structure a suitable channel through which an idiosyncratic shock propagates. This network feature can eliminate or amplify the idiosyncratic shock on the system-wide level. Therefore, the diversifying argument of the capital asset pricing models is not always true as the idiosyncratic shock becomes more significant when interacting with the network measures. Based on this idiosyncratic shock propagation concept, this dissertation incorporates the measures of interconnectedness and centrality into the asset pricing models.

In this dissertation, a stock network is constructed from the Pearson correlation matrix of stock returns that is filtered by a network algorithm. Unlike the unfiltered matrix, the filtered one contains only the essential information about the interrelationships. More importantly, it enables us to create a refined network of which many network characteristics can be quantified. The important network characteristics used in this dissertation are network topology and stock centrality. The network topology reflects the pattern of interconnections which may be integrated into a star-like network or even dispersing into a chain-like network. Each pattern has different ability to facilitate the idiosyncratic shock propagation. The stock centrality reflects the relative influence of the stock in two directions. The first direction is the stock's ability to influence the other stocks in the network while the other direction is its vulnerability to propagated shocks.

The key finding of this dissertation is that the measures of network structure are statistically significant to explain or predict asset returns. In the US market, I study the stocks listed in S&P500 and find that the network topology, measured by diameter, works together with the idiosyncratic risk, measured by average stock variance, to predict returns on the market portfolio. Furthermore, on the international financial markets, the network measures have power to predict the probability of extreme negative returns when working with the idiosyncratic risk measure which is average volatility of stock market returns. Lastly, in Stock Exchange of Thailand, I find that the portfolios formed by the network criteria earn abnormal returns that cannot be explained by the capital asset pricing model. The high systematic-important firms have lower returns than the low ones. The firms with high fragility level have higher returns than the others. Moreover, the network centrality may be useful in explaining the cross-sectional expected returns in Thailand.

Field of Study:	Economics	Student's Signature	....
Academic Year:	2017	Advisor's Signature	....
		Co-Advisor's Signature	....

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my thesis advisor Assistant Professor Pongsak Luangaram, Ph.D., for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research, publication, and writing of this thesis.

A very special gratitude also goes to my co-advisor Associate Professor Tanapong Potipiti, Ph.D., for his consideration and constructive suggestions during the last six years of my Ph.D. study. The door to his office was always open whenever I ran into trouble or had a question about my research.

Last but by no means least, I would like to thank my wife for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

This accomplishment would not have been possible without them. Thank you.

CONTENTS

Page

THAI ABSTRACT ...iv

ENGLISH ABSTRACT...v

ACKNOWLEDGEMENTS ...vi

CONTENTS...vii

Chapter 1 Introduction ...1

1.1 Background and motivation...1

1.2 Network structure of financial asset returns ...2

1.3 Thesis objectives and structure...4

Chapter 2 Stock market return predictability: Does network topology matter? ...7

2.1 Introduction...8

2.2 Methodology ...12

2.2.1 Network construction ...12

2.2.2 Measuring network topology with diameter ...13

2.2.3 Approximation of risk measures ...15

2.3 Description of data...16

2.4 Empirical results ...17

2.4.1 Simple illustration of network topology in the stock market ...17

2.4.2 Dynamic of network topology and market timing ...19

2.4.3 Predicting excess portfolio returns with network topology...21

2.4.4 Predicting excess portfolio returns with different horizons ...28

2.4.5 Out-of-sample predictability ...29

2.4.6 Predictive regressions with controlled variables ...32

2.4.7 Predictive regressions with alternative network measures...34

2.5 Concluding remarks...38

2.6 Appendix...39

2.6.1 Average stock variance ...39

Chapter 3 Predicting the probability of extreme negative returns: A network approach ...42

Page
3.1 Introduction....43
3.2 Methodology....45
3.2.1 Extreme negative return and contagion....46
3.2.2 Measuring the probability of extreme negative returns....46
3.2.2.1 Percentage with an extreme negative return....47
3.2.2.2 A measure of country-specific risk....47
3.2.2.3 Measures of cross-country linkages....48
3.3 Description of data....50
3.4 Empirical results....53
3.4.1 Explanatory variables over time....53
3.4.2 Extreme value analysis....55
3.4.3 The effect of idiosyncratic risk propagation model....59
3.4.4 The effect of global shocks on the extreme value analysis....62
3.4.5 Alternative network measures....66
3.5 Concluding remarks....69
3.6 Appendix....71
3.6.1 Area under the ROC curve (AUC)....71
Chapter 4 Interconnectedness and equity returns: The case of Thailand....73
4.1 Introduction....74
4.2 Methodology....76
4.2.1 Network structure of the stock market....77
4.2.2 Measures of interconnectedness....79
4.3 Description of Data....80
4.4 Empirical Results....81
4.4.1 CheiRank and PageRank....81
4.4.2 Cross-sectional regressions....83
4.4.3 Relationship between CheiRank/PageRank, size, and book-to-market....86

Page

4.4.4 Relationship between returns, size, book-to-market, CheiRank, and PageRank......88
4.4.5 Portfolio construction for asset pricing tests......90
4.4.6 Regression results for CheiRank-Size/Book-to-market portfolios......92
4.4.7 Regression results for PageRank-Size/Book-to-market portfolios......95
4.5 Conclusion......98
4.6 Appendix......100
4.6.1 PageRank......100
4.6.2 Power Method of PageRank......102
Chapter 5 Conclusion......104
REFERENCES......106
VITA......110

LIST OF TABLES

Table 2.1 Summary statistics for five-stock networks 18

Table 2.2 Summary statistics for predictive regressions 23

Table 2.3 Predicting subsequent excess portfolio returns from January 1990 to December 2014. 27

Table 2.4 Regressions of the excess portfolios returns with different horizons.......... 29

Table 2.5 Forecasting out-of-sample excess portfolio returns.................................... 32

Table 2.6 Regressions of the excess portfolios returns with controlled variables....... 34

Table 2.7 Regressions of the excess portfolios returns with alternative network measures.. 37

Table 2.8 Percentage of idiosyncratic component to the total risk............................. 40

Table 2.9 Percentage of the idiosyncratic component of the total risk in two subsamples. 41

Table 3.1 Data Description of Stock Market Returns 52

Table 3.2 Extreme value analysis – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B 58

Table 3.3 Idiosyncratic risk propagation – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B 61

Table 3.4 Controlling for global shocks – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B 64

Table 3.5 Ranking of the model with average correlation and diameter..................... 68

Table 3.6 Possible outcomes of the binary classification. 71

Table 4.1 Average CheiRank and PageRank 82

Table 4.2 Average slopes from month-by-month cross-sectional regressions. 86

Table 4.3 Average CheiRank and PageRank on size/book-to-market portfolios. 88

Table 4.4 Panel regressions between CheiRank/PageRank, size, and book-to-market 88

Table 4.5 Average monthly excess returns of portfolios sorted on size, book-to-market, CheiRank, and PageRank 89

Table 4.6 Summary statistics for the intersection portfolios. 91

Table 4.7 Regression estimates of excess stock returns on the portfolios formed by CheiRank, size, and book-to-market from January 1990 to December 2014. 94

Table 4.8 Regression estimates of excess stock returns on the portfolios formed by PageRank, size, and book-to-market from January 1990 to December 2014. 97

LIST OF FIGURES

Figure 1.1 An illustration of two PMFG graphs that share the same MST structure....3

Figure 2.1 Examples of five-stock networks....12

Figure 2.2 Histograms of average correlation and diameter for small networks. 18

Figure 2.3 Frequency of network structures classified by the correlation groups. 19

Figure 2.4 Diameter, Average stock variance and S&P500 index over time....21

Figure 3.1 A Simple Network Model with Two Countries....44

Figure 3.2 Illustration of Five-stock Network....50

Figure 3.3 The Global Network of the Full Sample....51

Figure 3.4 Dynamic of Contagion Measures 54

Figure 3.5 The Graphical AUC Performance 69

Figure 3.6 ROC (Receiver Operating Characteristic) space and curve. 72

Figure 4.1 Illustration of partial correlation....78

Figure 4.2 Dynamic of CheiRank and PageRank portfolios....82

Figure 4.3 The PCPG network during the global crisis period in September 2008....83

Figure 4.4 The network diagram with nodes' sizes proportional to PageRank. 100

Figure 4.5 Illustration of PageRank procedure. 101

Chapter 1 Introduction

1.1 Background and motivation

Network theory is originally the study of graph that consists of interrelationships between objects. In the past few decades, the network theory has been successfully applied to many fields of research. For example, in epidemiology, the network theory models how diseases spread, how people respond to the diseases, and how much diseases affect society. In computer science, the network becomes the most important tool to study the communication system, data transmission and internet. In sociology, the network is used to model the social structures such as friends of friends (Jackson & Rogers, 2007), employees' referrals for a job contact network (Montgomery, 1991), and crime network (Calvo-Armengol & Zenou, 2004). Recently, research on networks in the sociology and economics is merged into the field of social economics (or economic sociology). In social economics, the network is used to model the interactions among people, firms or institutions, which in turn influences the various economic behaviors such as the decision to buy or sell goods (Jackson, 2007). The countless ways that network structures affect our well-being make it an interesting and worthwhile area of research. As such, modeling how network structure influences economic activity, in general, and understanding financial interconnectedness, in particular, have been listed as one of the important research agenda for the next decade (Jackson, 2010).

In financial economics, the network theory can be useful in studying systemic risk and financial system. Systemic risk is originally a concept related to bank run and currency crises. However, the recent financial crisis in 2008 has renewed the interest in the systemic risk and even broadened the concept to the entire financial system. Due to the complex nature of the financial system, there is still no consensus to define the concept of systemic risk. Thus, the existing literature has focused on specific mechanisms of the systemic risk such as correlated exposures, feedback behavior, asset bubbles, and contagion. The network theory is useful in modeling some of these mechanisms, in particular, contagion and shock propagation mechanism. Acemoglu et al. (2015), for instance, report that the financial architecture is a source of systemic risk because of its shock propagation and amplification features.

More specific, the key advantage of the network-based models is the ability to explain the propagation mechanism of a complex system in which multiple relationships are taken into account. As a stock market is constituted by a group of listed companies, it is complicated by nature and thus directly affected by interrelationships and shock transmission. The network theory is then suitable to explain some economic activities in the stock market which are otherwise hard to explain by the existing equilibrium models. The global financial crisis in 2007-2008 is an excellent example in this case. The collapse of the Lehman Brother in September 2008 is estimated to cost only 5 billion dollars, but the event actually escalated to the collapse of the whole financial network in the US. The effect of this systemic event even spreads to other countries in the world, including Thailand. In September 2008 when the Lehman Brother collapsed, the Stock Exchange of Thailand dropped 30% in a single month. The striking feature of the event is that the subsequent loss is far greater than the initial damage (Haldane, 2013). Since modeling such an event involves multiple relationships between firms as well as propagation of the firms' shocks, the existing economic models would be exceedingly complex and not tractable. Therefore, this thesis paper incorporates the network concepts into the asset pricing models and provides empirical evidence for the relationship between equity returns and network structure in various equity markets.

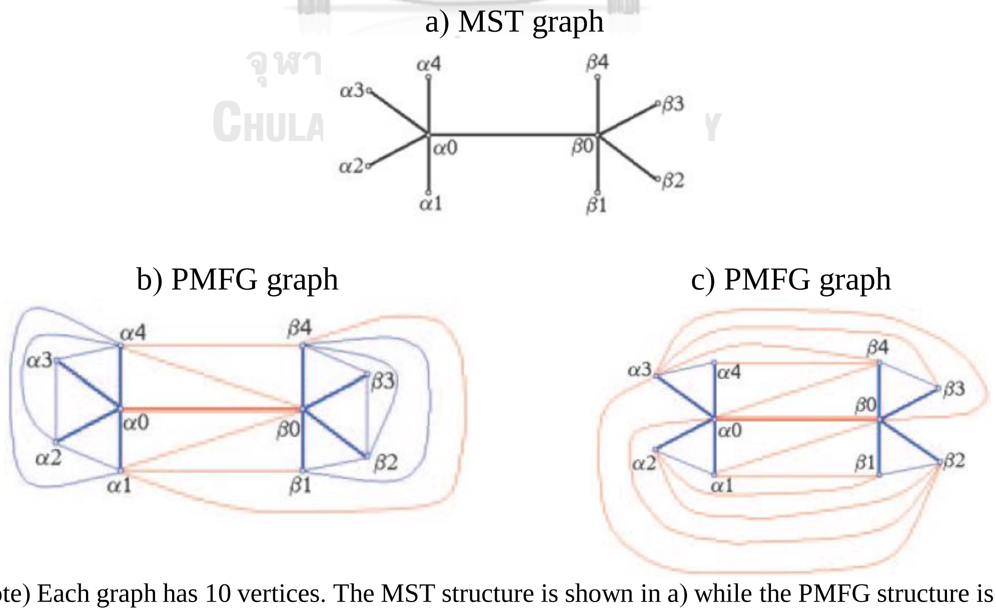
1.2 Network structure of financial asset returns

In modeling the structure of equity portfolios, the typical starting point is the cross-correlation of returns of asset pairs. The average of those simple correlation of returns indicates the co-movement of the stock returns and has been a well-known factor observed in financial markets. The presence of high degree of the average cross-correlation has been empirically documented to be associated with a crisis such as Black Friday in October 1987 (Onnela et al., 2003) and financial crisis in 2008 (Pollet and Wilson, 2010). However, it is known that a simple correlation of return may introduce the spurious correlations or noises in the network structure (Bonanno et al., 2004). Moreover, a financial market consists densely connected structure which is complex and cannot be represented by the average correlation. Therefore, there is a need to filter out the noise and transform the complex structure into a simpler and meaningful network.

Among other filtering approaches, Mantegna (1999) introduces a powerful methodology to extract a minimal set of relevant interactions, called the minimum spanning tree (MST). The MST algorithm essentially retains only a set of the highest correlations of returns that make a connected graph. The complexity of the system is substantially reduced from $n(n-1)/2$ to $n-1$ interactions. Therefore, the MST network will have a hierarchical structure with the essential information of the time-series of the stock returns. The economic justification of this methodology can be expressed by two reasons. First, Onnela et al. (2003) show that the distribution of the MST distance elements retains most of the features of the distribution of the correlation of returns. Specifically, their corresponding moments have high correlation or anticorrelation (above 0.8 in absolute value). Thus, the MST simplifies the complex correlation structure while still retains the relevant information. Another key justification of the MST is its ability to provide a meaningful economic taxonomy for a stock market. Since stocks in the same sector have similar common economic factors that drive stock prices, they should be clustered together in the network. Mantegna (1999) and Onnela et al. (2003) evaluate the economic meaningfulness of grouping stocks in the MST asset tree with a third party reference classification, www.Forbes.com in this case. They find that the MST network taxonomy is well compatible with the reference classification. Both reasons strongly advocate for this filtering approach in stock market analysis.

Nevertheless, the reduction of the fully connected structure to the minimal asset tree is an extreme approach that may lose some valuable information. Tumminello et al. (2005) therefore propose another filtering algorithm, called Planar Maximally Filtered Graph (PMFG). The PMFG is, in fact, an extension of the MST. The PMFG network consists not only the minimal skeleton structure of the MST but also some additional links that form loops or cliques of three or four nodes. As a result, one justification to use the PMFG is that it retains all information and features of the MST. To illustrate, Figure 1.1 shows three graphs with 10 vertices. Figure 1.1a is the MST graph with 9 links. Figure 1.1b and 1.1c show two PMFG graphs with 24 links. We can see that the PMFG graphs contain all links from the MST graph. Furthermore, the additional links of the PMFG allow more variety of the structure than the MST. As shown in Figure 1.1, the two PMFG graphs have a different structure that shares the same MST structure. Also, the additional links of the PMFG graphs enable the feedback loop which is one of the mechanisms that can explain the systemic risk. For these reasons, the thesis will mainly focus on the properties of the PMFG graph.

Figure 1.1 An illustration of two PMFG graphs that share the same MST structure.



(Note) Each graph has 10 vertices. The MST structure is shown in a) while the PMFG structure is shown in b) and c). The source of this figure is from Tumminello et al. (2005)

Recent related papers have documented the empirical evidence of the correlation structure for the stock market analysis. For example, Buccheri et al. (2013) find that the correlation structure between US industry indices presents both slow and fast dynamics. The slow dynamics suggests that the different investment diversification is possible in different periods of time. The timescale is as slow as five years. Also, the fast dynamics is detected on the monthly time scale with the three-month rolling window. The changes in PMFG structure are detected both in the presence of exogenous events (ASIAN crisis in 1977 and the sovereign debt crisis in 2011) and endogenous events (market decline in October 1978 and market crash of October 1987, dot-com in 1999-2001, subprime in 2008-2009). Another example is presented by Birch et al. (2016) who compare three filtering procedures by using the stock returns in DAX 30. The filtering procedure includes MST, PMFG and asset tree. The asset tree contains only the largest n-1 correlations of returns but does not require a connected graph. They find that all resulting structures are useful in providing insights into the growth dynamics of an economy. Specifically, the structures are corresponding to a period of crisis (October-December 2008) and a period of recovery (May-August 2010).

1.3 Thesis objectives and structure

In this thesis paper, I incorporate the network theory into the asset pricing study by using the propagation mechanism. Specifically, the network architecture of a stock market is recreated from the correlation of stock returns in the market. A firm or stock is a node in that network, and the correlation of stocks' returns represents the relationship or link between the stocks. When a firm-specific shock occurs, it could propagate through the network structure to affect the stocks in the system. Therefore, the role of the network structure in the stock market is to facilitate the propagation of the firm-specific shocks. Through the transmission process, the shocks may be diversified away or amplified to affect the entire stock market.

Based on this network concept, this dissertation examines the relationship between equity returns and network structure in various equity markets. The second chapter of the dissertation provides empirical evidence for the return predictability of excess market portfolio returns of stocks. Specifically, I create a series of the stock networks from some largest stocks in S&P500 during 1990 to 2014. The interconnection of the networks is then a proxy for the propagation channel of firm-specific shocks. Among other interconnectedness characteristics, the paper focuses on the network topology or the pattern of interconnections. The dynamic of the network topology indicates that some patterns are more suitable for the shock propagation than the others. I quantify the network topology by the shortest distance between the two farthest nodes, called diameter. A small diameter implies that the stocks are closely related such that a shock can propagate to the others with the short distance. In counterpoint, the network topology of the high diameter implies the long distance that a shock needs to propagate to the others. In consistence with the network concept of the shock propagation, I find that the idiosyncratic shock, measured by the average stock variance, performs better when the diameter is taken into account. The results hold true for both monthly and quarterly intervals. The relationship also remains significant after controlling for variables known to forecast the stock market returns, including TED spread, TERM spread, dividend yield, 3-month Treasury bill, market capitalization, and book-to-market.

The third chapter of the dissertation tests whether or not the network topology can help to identify or predict the probability of extreme negative returns for the international equity markets. Similar to the second chapter, this article relies on the concept of idiosyncratic shock propagation in which the network topology serves as the propagation channel. Instead of the US market, the data of this paper consists of 43 indexes listed on MSCI World and MSCI Emerging markets. A network then represents the interconnections of the international markets. When an extremely negative shock spills over and affects the other countries in the network, contagion occurs. Thus, the key components that can affect the contagion are country-specific shock and propagation channel. The measure of the idiosyncratic risk is average variance calculated as the equal-weighted average of stock markets' variances, using within-month daily data. Two network measures are used to capture two aspects of the propagation channel. The first measure is an average correlation of returns which reflects the strength or width of the propagation channel. The second measure is diameter which captures the distant of the propagation channel. I find that the measures of propagation channel are rather weakly related to the extreme negative returns, while a measure of country-specific risk is significant. However, once the network measures interact with the idiosyncratic risk measure, their ability to predict the probability of extreme negative returns increases significantly.

The fourth chapter of the dissertation provides empirical evidence for the relationship between network centrality and asset returns in the Stock Exchange of Thailand. Unlike the previous two chapters, this chapter focuses on the developing market in which the empirical research on this subject is lacking. Moreover, instead of the global characteristics of the network structure, this chapter uses the local aspects including systematic importance and fragility. A firm is systematically important if it can affect the other firms in the network. CheiRank is a network measure to capture the systematical importance property. On the contrary, a firm is fragile if it gets affected by propagated shocks. PageRank is a network measure to capture the fragility property. The methodology of this chapter is similar to the portfolio mimicking and the asset pricing model of Fama and French (1993). I find that CheiRank has a significant and negative relationship with equity returns, whereas PageRank has a significant and positive relationship. Therefore, the local measures of the network structure may be useful in explaining cross-sectional and time-series expected returns for firms listed on the Stock Exchange of Thailand.

The rest of the dissertation is organized into four chapters. Chapter 2 elaborates the motivation, methodologies, and results of the tests for the relationship between network topology and market portfolio returns on the US market. Chapter 3 explains the methodology and the findings for the network effect on international financial contagion. Chapter 4 provides empirical evidence of the network effect on the Stock Exchange of Thailand. Chapter 5 concludes the thesis.

Chapter 2 Stock market return predictability: Does network topology matter?

Abstract This paper provides new evidence for the predictability of excess market portfolio returns using a network approach. In particular, this article introduces a measure of interconnectedness to capture the interrelationship of returns of 100 largest stocks in S&P500 during 1990- 2014. In the financial network literature, the interconnection of a stock network is often regarded as a channel through which an idiosyncratic shock propagates. The idiosyncratic risk propagation is crucial to the debate over the relationship between idiosyncratic risk and market returns because the idiosyncratic risk is not always diversified away. Rather, the network can sometimes amplify the effect of the idiosyncratic risk to cause aggregate fluctuation. In accordance with this theoretical argument, I empirically show that the network topology, measured by diameter, works together with the idiosyncratic risk, measured by average stock variance, to affect the market portfolio returns. This relationship persists after controlling for well-known variables known to forecast the stock market returns.

Keywords Stock market network, Network topology, Return predictability, Diameter, Idiosyncratic risk, Average stock variance

JEL Classification G12, D85

2.1 Introduction

Are stock market returns predictable? Cochrane (1999) responds, "We once thought that stock and bond returns were essentially unpredictable. Now we recognize that stock and bond returns have a substantial predictable component at long horizons."

Among other predictors of returns, the variance of stock market returns is an intuitive measure of risk and has also been used by many papers to predict the stock market returns. Unfortunately, the relationship between such risk and stock market returns is not straightforward and often found to be insignificant. Pollet and Wilson (2010) empirically show that the stock market variance cannot predict the subsequent quarterly returns on the CRSP value-weighted index. Theoretically, the stock market variance is composed of two components including idiosyncratic and systematic components. The idiosyncratic component is measured by average stock variance, which is essentially the diagonal information of the variance-covariance matrix of the stock returns, while the systematic component is measured by average correlation (AC), representing non-diagonal information of the correlation matrix. Pollet and Wilson demonstrate that the weak relationship between the stock market variance and return is primarily due to the idiosyncratic component. Therefore, the average correlation is a better proxy for the aggregate risk that is statistically significant in predicting excess stock market returns.

In contrast, a number of academic studies have documented the ability of the idiosyncratic risk to explain and predict equity returns in many developed markets. Using CRSP data, Goyal and Santa-Clara (2003) find a significantly positive relationship between the stock market returns and the average stock variance. The authors show that the idiosyncratic component is the major part of the average stock variance while it is diversified away in the stock market variance. This finding advocates the average stock variance as a suitable measure for idiosyncratic risk and in turn raises the importance of idiosyncratic risk on forecasting the stock market returns. Nanisetty et al. (1996) examine the intertemporal capital asset pricing model (ICAPM) that includes idiosyncratic risk premia and market risk premium in the pricing equation. They find that the idiosyncratic risk premium is significant in explaining returns on the size and industry portfolios of equities listed on New York Stock Exchange (NYSE). Drew et al. (2007) examine the relationship between idiosyncratic volatility and stock excess returns for equities listed on the New Zealand Exchange. They use the approach of Fama and French (1993) and find that the idiosyncratic volatility is statistically significant in explaining the cross-section expected returns. Last but not least, Vidal-Garcia et al. (2016) study the effect of the liquidity and idiosyncratic risk factors in the European mutual fund market. They report that both liquidity and idiosyncratic risk factors are relevant to mutual fund performance and robust to the well-known risk factors regarding market, size, valuation, and momentum.

Given the mixed evidence from the existing literature, the relationship between the idiosyncratic risk and market returns remains an open discussion. This paper presents new evidence that the idiosyncratic risk, measured by average stock variance, has forecasting power for market portfolio returns when interacting with a network measure. Specifically, a stock market can be viewed as a complex network in which stocks interact. The interconnection serves as a channel through which idiosyncratic shocks propagate. The idiosyncratic shocks may be either diversified away or amplified throughout the network. Thus, the diversifying argument is not always true, and the idiosyncratic risk can manifest under suitable environments.

The concept of idiosyncratic risk propagation is not new and has been studied in many related fields. For example, Acemoglu et al. (2012) study U.S. intersectoral input-output data and show that microeconomic idiosyncratic shocks can cause aggregate fluctuations in the economy. Diebold and Yilmaz (2014) document idiosyncratic volatility spillover among major U.S. financial institutions. Acemoglu et al. (2015) and Elliott et al. (2014) study network architectures of financial interdependencies and show that different network structures are associated with varying levels of interconnectedness. However, individual shocks can, in some cases, trigger a cascade of failures and escalate into systemic events.

The economic impact of such a mechanism in a complex network can be extraordinary when idiosyncratic shocks are amplified via feedback loops and cascades of failures. As pointed out by Haldane (2013), when Lehman Brothers collapsed in September 2008, the damage was not limited to itself but also spread to other firms in the US market and eventually the global market. The direct cost of the Lehman Brothers' bankruptcy was estimated to be around US$5 billion, but the IMF revised global growth down by more than 5 percent. The striking feature is that the markets can amplify an idiosyncratic shock in such a way that the subsequent loss is far greater than the initial damage.

Given this rationale, I study the relationship between the idiosyncratic risk and stock market returns in the S&P500. The primary objective is to reexamine this relationship when network measures of interconnection are taken into consideration. Following Goyal and Santa-Clara (2003), the idiosyncratic risk is measured by the average stock variance. The measures of interconnection, on the other hand, are not as straightforward and required a number of tasks. To achieve this goal, I first simulated a stock market by a network of stocks. A connection between stocks is measured by Pearson's correlation of stock returns. This setup allows me to capture the interconnectedness of the stock market that functions as a propagation channel of idiosyncratic shocks.

A simple measure of interconnection for a correlation-based network is the mathematic average of all correlations except for the diagonal elements. This measure will be called average correlation (AC). However, I contend that the AC is not a good candidate for the propagation channel of idiosyncratic risk because of two reasons. First, it relates more to systematic risk than idiosyncratic risk. Since the AC implies how strong stock returns are moving together in aggregate level, it is naturally associated with the common risk profile of the stock market. Pollet and Wilson (2010) provide empirical evidence that distinguishes the AC from the idiosyncratic component of stock market variance and establishes it as a measure of the systematic component. Second, from the network theory perspective, the AC is a crude measure for interconnection. It does not give the complete picture of interdependencies nor the full state of the complex system. More specifically, interconnection is a term used for collective relationships or links that form a network. Two important aspects of the interconnection are the strength of those relationships and pattern of the connections. Though the AC is a natural and good measure of interconnection strength by construction, it provides little information about the pattern of interconnection. Throughout the paper, the pattern of interconnection will be referred to as "network topology."

I emphasize the network topology precisely in this study because it provides direct information about the propagation channel, which is essential to the idiosyncratic risk propagation mechanism. To illustrate, Figure 2.1 shows different patterns of interconnection in a simple network. Figure 2.1b represents a simple chain network, and Figure 2.1d depicts a simple star network. In theory, the latter will facilitate shock propagation better than the first because it allows individual shocks to reach throughout the network much more quickly. This indicates that the star-like network has the level of interconnectedness higher than the chain-like network. To measure the network topology, I chose an easy-to-understand network concept, called the "diameter." The diameter is the shortest distance between the two farthest nodes. A small diameter implies that the stocks are closely related such that the network topology becomes more like a star shape, as shown in Figure 2.1d. In counterpoint, the network topology of the high diameter leans towards the chain structure, illustrated in Figure 2.1b.

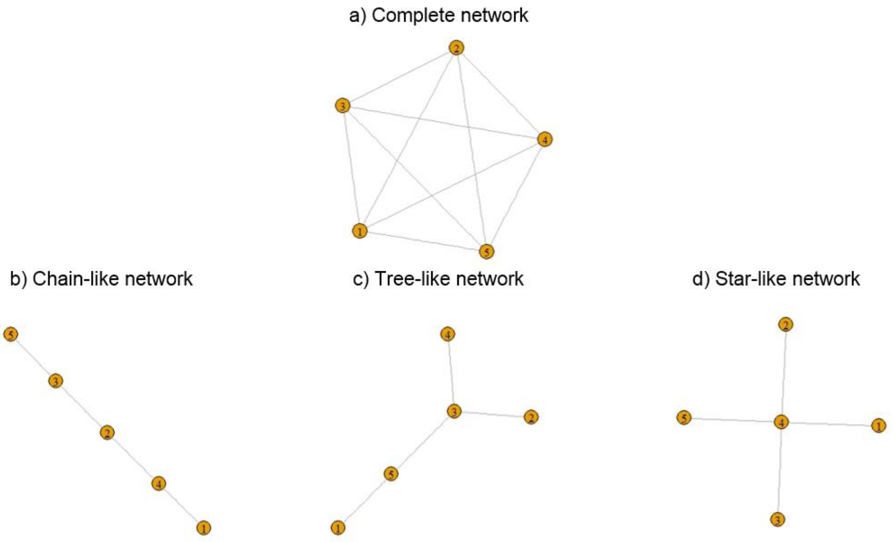
Based on this background, I formulated the main research question as follows: Does the correlation structure have the power to predict the stock market portfolio returns? The paper's hypothesis is that the correlation-based network should have the power to predict the returns or at least help the existing risk factors (average correlation and variance) to predict the future returns. There are two main reasons that support this hypothesis. The first one is that the network structure is constructed from the filtering procedure that retains the essential information and most properties of the correlation matrix. Since the prediction power of the average correlation is empirically proven by

Pollet and Wilson (2010), there is a high possibility that the network structure will be relevant to the future returns. Secondly, many prior works have asserted the compatibility of the network structure and the real economic taxonomy. The change of the network structure would then affect the common economic factors that drive the stock prices. When the structure becomes tightly packed, a common economic factor specific to a sector can affect the stocks in the other sector more easily. Moreover, the shock specific to a firm can efficiently propagate throughout the network and even be amplified by the network feature to cause a system-wide risk.

Results of the paper show that the network topology, measured by diameter, is a potential indicator of idiosyncratic shock propagation channel. The measure of idiosyncratic risk alone is barely able to predict the return with the monthly t-statistic of -1.938 and the quarterly of -1.878. The adjusted R^2 is also very low at 1.4% and 1.9%, respectively. By adding the interaction term between the diameter and average stock variance, I can test the effect of the idiosyncratic risk propagation model on returns. The coefficients of average stock variance and interaction term are both statistically significant at 95% confidence interval. The adjusted R^2 is improved significantly to 3.6% and 8.8% for the monthly and quarterly intervals. The ability to explain the return variation is greatly improved from when the average stock variance is the sole predictor. These results imply that idiosyncratic shocks cannot be entirely rejected due to a diversification argument. Instead, they can sometimes affect the stock market returns under favorable network topologies that serve as propagation channel with an amplification or diversification function. Furthermore, the findings support the diameter as a good propagation channel for idiosyncratic shocks.

The rest of the paper is organized into four sections. Section 2.2 explains the methodologies applied in constructing the stock networks and generating measures of network topology and idiosyncratic risk. Section 2.3 presents the description of data and summary statistics. Section 2.4 reports the empirical results from time-series regressions as well as alternative specifications for robustness checks. Section 2.5 concludes the paper.

Figure 2.1 Examples of five-stock networks.



(Note) This figure illustrates the possible network structure of a system consisting of five stocks. The complete network in a) is the representation of a structure with all possible relationships. The minimum-spanning-tree filtering procedure (MST) compresses the complete network into three structures, including Chain-like network in b), Tree-like network in c) and Star-like network in d)

2.2 Methodology

Assessing the impact of network topology on the relationship between market portfolio returns and idiosyncratic risk requires the definitions of several related concepts and parameters. I will begin with explaining necessary concepts for the construction of a stock network and the filtering procedure that allows us to observe the dynamics of interconnection structures. I will then introduce the concept of diameter for capturing the network topology of the filtered network. Lastly, I will define the idiosyncratic risk measure and related variables used for the predictive regressions on portfolio returns.

2.2.1 Network construction

To simulate the large and complex system of a stock market, I carried out three steps: i) define the node, ii) define the links between a pair of nodes, and iii) eliminate the unimportant links to capture the essential structure of the network.

In this paper, a node represents a stock. A link or relationship between nodes is defined as a correlation between the stock returns. Although the relationship can be various, I choose to work with the simple correlation of stock returns for two reasons. First, it is widely used in the financial network literature (See for example Mantegna (1999), Tumminello et al. (2005), and Engle and Kelly (2012)). Second, a correlation of returns contains all information about the stock relationship, including investor expectations that are otherwise difficult to measure or obtain.

The last step of the network construction is to apply a filtering procedure to control complexity and yet maintain the essence of the stock interrelationship. More specifically, a correlation matrix without the diagonal elements is, in fact, a fully connected network. The total number of the network connections is n(n-1)/2 correlations for n stocks. For instance, when n is 5, there are ten correlations as shown in Figure 2.1a. If n is 100, the number of links becomes 4,950. As n grows, the network becomes complicated and hard to deduce any patterns of interconnection. To reduce the complexity while maintaining the minimal-yet-meaningful structure, Mantegna (1999) introduces a filtering procedure, called Minimum Spanning Tree (MST). The algorithm starts by ordering the correlations from high to low. Next, the highest correlation is picked first, followed by the next highest correlation as long as the graph is connected without a loop or cycle. If an additional link does not satisfy the condition, the algorithm skips to the next link. The total number of links then reduces to n-1. Figures 2.1b-2.1d illustrate simple MST networks for five stocks. We can see that the number of connections decreases from 10 in the complete network to 4 in the MST networks.

However, the reduction to a minimum network structure is an extreme approach with a large amount of information lost. Tumminello et al. (2005) therefore propose another filtering algorithm, called Planar Maximally Filtered Graph (PMFG). The PMFG is very similar to the MST, except for a more relaxed network constraint. That is, the PMFG keeps adding links as long as the graph can still be drawn on a 2-D surface without link crossing. This constraint is called the planarity condition. Consequently, the PMFG network contains all of the MST links and some additional links that form loops or cliques of three or four nodes. If a graph contains a subgraph K5 (a complete graph on five vertices) or K3,3 (a complete bipartite graph on six vertices), it is not planar. The maximum number of links is 3(n-2), which is much higher than that with the MST. For instance, when n is 5, the MST network has four edges, and the PMFG has 6. When n is 100, the MST graph contains 99, and the PMFG is 294. In short, I choose to work with the MST networks for basic illustration and the PMFG networks for the main results due to the additional valuable information.

2.2.2 Measuring network topology with diameter

Since a Pearson's correlation matrix is a complete network with all non-diagonal pairs connected, measuring a network topology from such a network is not feasible. The only network measure of the complete network is the average correlation, which mainly captures the strength of interconnectedness. Consequently, to extract a pattern of interconnection called network topology, I apply the PMFG filtering algorithm to compress the complex network into a smaller one that contains the essence of the interrelationship. Different network measures then can be computed from this kind of the network, such as diameter, degree distribution, clustering coefficient, average path length, and centrality measures. Interested readers can consult Jackson (2008) for a more detailed explanation.

In this paper, I measure the topology of the filtered network with a network concept, called "diameter." If the shortest path or geodesic path of each pair of nodes is the lowest number of links between the two nodes, the diameter is the largest of all geodesic paths. This definition makes the diameter an easy-to-understand and intuitive measure for the network topology. To illustrate, I create a fully connected graph for five stocks 1-5 as shown in Figure 2.1a. By applying the MST algorithm, the complete network can be compressed to one of the three structures, including star, tree, or chain network. Figure 2.1b is a chain-like structure that has the longest diameter of 4 as measured by the number of links of the largest geodesic path. Figure 2.1c is a tree-like structure with a diameter of 3. Figure 2.1d is a star-like network with the shortest diameter of 2. This exercise clearly shows us that the large diameter indicates the chain-like network and vice versa.

At this point, I measure the pattern and strength of interconnection by the diameter and average correlation, respectively. The important question is whether or not the diameter can provide additional insights into the stock returns. To achieve this goal, I simulate a series of equally-weighted MST networks of five stock returns' correlations on a quarterly basis. I choose to illustrate the small networks because they provide better conceptual illustration and visualization for the analysis. Furthermore, the five-stock network is the smallest and simplest network that allows the diameter to accurately specify the shapes of network structures as shown in Figures 2.1b-2.1d. Additionally, I can control the effect of average correlation on the network by making all links attach to a given node with equal weight. The average correlation is assigned into three groups because the average correlation is continuous and the diameter is discrete in value. Therefore, if the network topology, measured by diameter, provides additional market information to the average correlation, I should see the variety of diameters in each correlation group. The result is presented in Section 2.4.1 and suggests that the diameter holds some unique information about stock returns which is not captured by the average correlation.

2.2.3 Approximation of risk measures

Following Goyal and Santa-Clara (2003), I calculate the monthly variance of an asset by using within-month daily returns, as expressed in Equation (2.1).

$$V_{a,t} = \sum_{d=1}^{D_t} r_{a,d}^2 + 2\sum_{d=2}^{D_t} r_{a,d} r_{a,d-1}, \tag{2.1}$$

where $r_{a,d}$ is the return of asset a on day d. D_t is the number of the trading days in month t. According to Goyal and Santa-Clara (2003), the second term on the right-hand side is used for the autocorrelation adjustment in daily returns. If asset a is a portfolio P, $r_{p,t}$ is the portfolio return and $V_{p,t}$ is a measure of the portfolio risk.

If asset a is a stock i, $V_{i,t}$ is a monthly variance of stock i. Goyal and Santa-Clara (2003) propose that the equally-weighted average of $V_{i,t}$ is a good approximation of idiosyncratic risk of the stock market portfolio. Specifically, the equally-weighted average variance can be decomposed into systematic and idiosyncratic components. The effect of idiosyncratic risk constitutes the majority part of the variance while systematic risk is negligible. Ultimately, Goyal and Santa-Clara show that this risk measure has forecasting power for market returns. However, Bali et al. (2005) argue that small stocks and liquidity premiums drive the predictability of such a measure. Rather, the value-weighted measure of idiosyncratic risk can mitigate the problem and is more natural for the predictability of market returns. Thus, in this paper, I use the value-weighted average stock variance to measure the idiosyncratic risk as in Equation (2.2).

$$AV_t = \sum_{i=1}^{N_t} w_{i,t} V_{i,t}, \tag{2.2}$$

where the weights for stock i, $w_{i,t}$, are the market capitalization of stock i at the last trading day in period t divided by the market capitalization of the entire market portfolio. I assume that the weights are constant in period t. N_t is the number of stocks used in the calculation of period t.

Lastly, following Pollet and Wilson (2010), average correlation is estimated as the value-weight mathematic average of the correlations as in Equation (2.3).

$$AC_t = \sum_{i=1}^{N_t} \sum_{k \neq i} w_{i,t}\, w_{k,t} \rho_{ik,t}, \tag{2.3}$$

where the weights for stock i, $w_{i,t}$, are the market capitalization of stock i at the last trading day in period t divided by the market capitalization of the entire market portfolio. $\rho_{ik,t}$ is the Pearson's correlation between stock i and k. N_t is the number of stocks used in the calculation of period t. While the average correlation is a measure of stock market risk as documented by Pollet and Wilson, it is also a measure of interconnection in the network literature.

2.3 Description of data

I compute the measures of network topology and risk using a set of 100 stocks listed in the S&P500. The sample begins in January 1990, ends in December 2014 and is collected from Bloomberg Terminal. The daily stock return, $r_{i,t}$, is generated from the difference in log prices between two consecutive days, as shown in Equation (2.4). When returns are not available due to a holiday or other reasons, I use the number from the last-known period.

$$r_{i,t} = \ln(p_{i,t}) - \ln(p_{i,t-1}) \tag{2.4}$$

The list of 100 stocks is determined by market capitalization and return availability. Specifically, at the end of each quarter, I sort all stocks in the S&P500 by market capitalization. A stock with missing return data is eliminated. Finally, I select 100 largest stocks in the remaining list. This data set is then used to compute the correlation matrix, network measures, and risk measures.

For the analytical purpose, the main results are reported on both monthly and quarterly intervals. In the monthly analysis, a network is created using the 3-month rolling sample from month t-2 to t. Thus, a series of diameter is estimated from 300 rolling-sample PMFG networks. It should be noted that I use a rolling sample instead of the within-month sample because the latter is rather too short for revealing the stocks' interrelationships. Without going into detail, the results from the within-month data still support the paper's hypothesis but are much weaker. On the other hand, the risk measures are slightly in favor of the within-period sample.

Monthly risk measures are calculated from the within-month daily returns of the 100 selected stocks as shown in Equation (2.1), (2.2) and (2.3). The excess market portfolio return is the log return on the value-weighted market portfolio over the 3-month Treasury Bill. I use the portfolio returns instead of the S&P500 index return to avoid potential biases from smaller stocks not included in the sample.

In the quarterly analysis, PMFG networks and diameters are created using the within-quarter data, which provide 100 networks and associated diameters, in total. The risk measures and portfolio returns are also constructed using the within-quarter data.

2.4 Empirical results

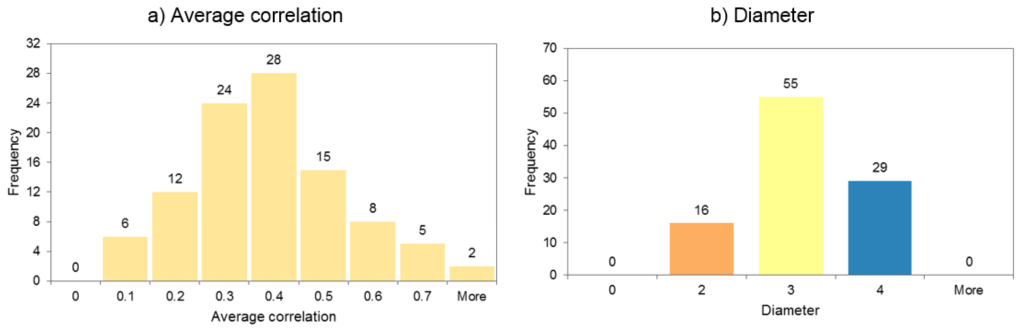
2.4.1 Simple illustration of network topology in the stock market

In the introduction, I postulate that the average correlation (AC) captures one aspect of interconnectedness and the diameter capture another, network topology in particular. This section provides a simple illustration of this statement by comparing diameters and average correlations associated with the portfolio returns. The result demonstrates that the diameter is informational about the asset returns after controlling for the AC.

For illustration, I simulate a series of small MST networks with five stocks on the quarterly interval. Specifically, out of 1,086 listed firms in the sample period, five stocks are selected, based on market capitalization and return availability. The five largest stocks in size include XOM (Exxon Mobile), GE (General Electric), MSFT (Microsoft), WMT (Wal-Mart), and PFE (Pfizer). The network topology is then measured by the diameter, whereas the commonality among asset returns is captured by the average correlation.

Figure 2.2 and Table 2.1 present the histogram and summary statistics of average correlation (AC_5) and diameter during the 100 periods of the sample. The average AC_5 of five stocks is quite high at 0.34, given that the five stocks belong to different industries. The positive sign also indicates that the stock prices usually move together in the same direction. The average diameter (DIA_5) is 3.13. Out of 100 periods, the tree-like network with the diameter of 3 appears the most frequent at 55 periods, followed by the chain at 29, and then the star at 16. This result indicates that the network topology of the portfolio can change over time and, thus, has some implication on the portfolio returns. The question is whether or not the diameter can have an influence on the returns in addition to the AC, which is simpler and proved to be significant by Pollet and Wilson (2010).

Figure 2.2 Histograms of average correlation and diameter for small networks.



(Note) Figure 2a represents the frequency of average correlation, which is defined as the cross-sectional average of the Pearson's correlation of daily returns of the selected five stocks. Figure 2b represents the frequency of diameter, which is defined as the largest geodesic path of a five-stock network

Table 2.1 Summary statistics for five-stock networks

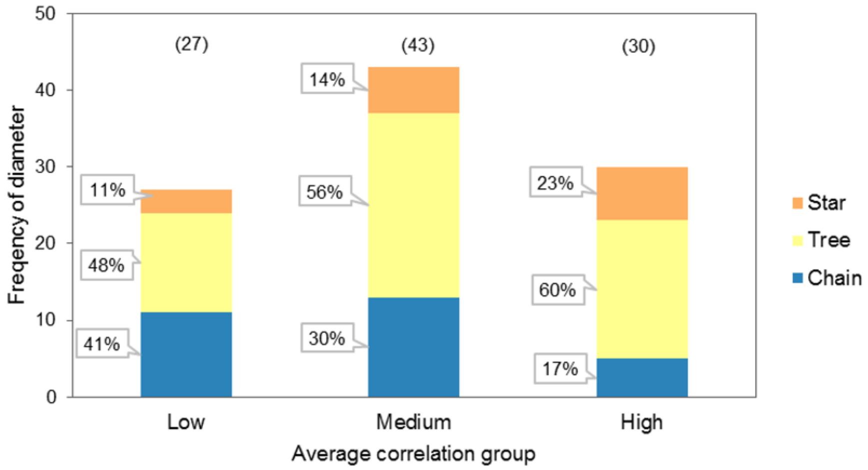
Average correlation			*Diameter*		
Correlation Group	**AC_5**	**N**	**Network Topology**	**DIA_5**	**N**
Low (< 0.25)	0.16	27	Chain	4	29
Medium (0.25-0.4)	0.33	43	Tree	3	55
High (>0.4)	0.53	30	Star	2	16
Full sample	0.34	100	Full sample	3.13	100

(Note) Table 2.1 reports the summary statistic of average correlation (AC_5) and diameter (DIA_5) from the five-stock networks. Average correlation is classified into three groups: Low, Medium, and High. Diameter represents three structure of the network topology, including Chain, Tree or Star. AC_5 is the cross-sectional average of average correlation for the corresponding groups. DIA_5 is average diameter for the corresponding network structures. N is the number of periods that the portfolio returns belong to the corresponding groups or network structures

Since it is difficult to compare the continuous variable (average correlation) with the discontinued variable (diameter), I classify the average correlations into three groups: Low, Medium, and High. The histogram in Figure 2.2a shows that the cutoffs at 0.25 and 0.4 ensure the sufficient sample size in each group. Table 2.1 reports that the sample sizes for the low, medium and high correlation groups are 27, 43 and 30, respectively, with the average correlation of 0.16, 0.33, and 0.53. Next, the correlation groups are mapped with the diameter as shown in Figure 2.3. All three types of network structures can appear in each AC group. The low correlation group is composed of 41% chain structure, 48% tree structure and 11% star structure. The medium group consists of 30% chain structure, 56% tree network and 14% star structure. The high group comprises 23% chain structure, 60% tree structure and 17% star structure. The results suggest that the diameter can change under a relatively stable condition of the AC. In other words, the diameter bears some new insights of interconnectedness and stock market return not reflected in the average correlation.

These observations and conceptual understandings serve as my motivation for using the diameter as the measure of network topology to predict market portfolio returns. The next section will provide time-series analysis of the diameter and the portfolio returns of the full sample.

Figure 2.3 Frequency of network structures classified by the correlation groups.



(Note) The average correlation is the cross-sectional average of the Pearson's correlation of daily returns of the selected five stocks. The Low/Medium/High groups consist of the quarterly periods with the average correlation below 0.25, between 0.25-0.4 and above 0.4, respectively. The frequency of diameter is the number of diameters that are matched to each correlation group. Star on the top of the candle represents the number of networks with a diameter of 2. Tree in the middle of the candle is the number of networks with a diameter of 3. Chain in the bottom is the number of networks with a diameter of 4

2.4.2 Dynamic of network topology and market timing

In this section, the goal is to see if the network topology as measured by the diameter shows a certain trend with the stock market movement. To better reflect the actual market, I compute a diameter from the PMFG 100-stock network in each quarter. I then plot the diameter over time along with the portfolio returns as shown in Figure 2.4.

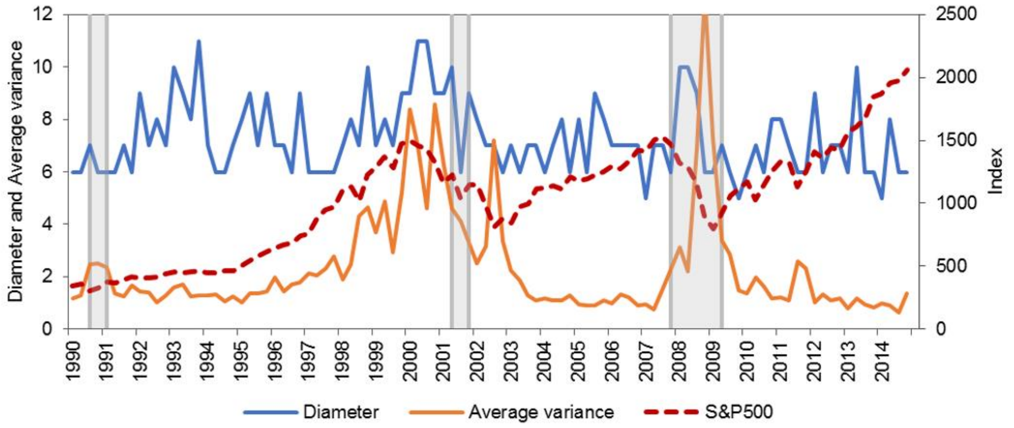
The very first thing we can see from the graph is that the network structures lean towards the star-like shape. For a network with 100 members, the most extreme star and chain networks will have the diameter of 2 and 99, respectively. However, the diameters of the stock networks are closer to the star-like network with a range from 5 to 11. This result implies that, in general, one stock can reach the other stocks in the network with a short distance.

Although the diameter does not exhibit a pronounced upward or downward trend, it does show that the stock market structure becomes more star-like in the second half of the sample period. In the first half, the diameter ranges from 6 to 11, while it goes from 5 to 10 in the second half. One possible explanation is that the stock market becomes more integrated during the 2000s possibly due to advances in computer and network technologies and financial innovations such as CDO and CDS.

In addition, Figure 2.4 illustrates the relationship of diameter and the major market timings. The shaded areas represent three periods of recession as defined by the NBER, including July 1990 – March 1991, March 2001 – November 2001, and December 2007 – June 2009. In these contraction periods, the diameter has volatile movement with a sharp decline in value to 6. For example, before the global market sell-off in September 2008, the diameter started to fall from a peak in 2008 Q2 with the diameter of 10 and reached the bottom in 2008 Q4 with the diameter of 6.

Similar to Kaya (2014), the jumps in the network topology can be seen as a necessary condition for market sell-off but not a sufficient condition given that the measure assigns some false indications. For example, during 1997, the diameter remains very low at 6 possibly due to the fear of contagion from the East-Asia crisis. However, the star-like network does not trigger the sell-off in the US market. This event can be partly explained when the idiosyncratic risk, measured by the average stock variance, is taken into account. In particular, since the diameter reflects the propagation channel of the idiosyncratic risk, analyzing the individual effect of diameter might not be accurate and sometimes cause the false indications. Figure 2.4 shows that the diameter will be more relevant to the market timings when coupling with the average stock variance. During the Mexican peso crisis (1993-1994), the East-Asia crisis (1997) and the dot-com crisis (2000), the diameters drop sharply from the peak, but the average stock variance does not change much. As a result, the effect of diameter on those crisis periods is rather limited. Conversely, during the last two recessions, both diameter and average stock variance work together to cause a disturbance in the stock market and result in intermittent drops of the stock market index.

Figure 2.4 Diameter, Average stock variance and S&P500 index over time.



(Note) The figure presents the dynamic of the diameter, average stock variance and S&P500 index over time from 1990 Q1 to 2014 Q4. The vertical axis on the left shows the value of diameter and the percentage of average stock variance. The vertical axis on the right shows the S&P500 price index. Diameter is calculated as the longest distance of all geodesic paths in a given network. Average stock variance is the value-weighted average of stock variances. The shaded gray areas represent periods of contraction according to NBER

2.4.3 Predicting excess portfolio returns with network topology

This section reports the predictability performance of the diameter, how it relates to the market portfolio returns and its role as a channel of idiosyncratic risk propagation. To assess the relationship between returns and explanatory factors, I estimate the regressions and report in Table 2.3. The dependent variable is excess market portfolio returns at period *t+ 1*. The explanatory variables are average correlation, diameter, average variance, stock market portfolio variance, and two interaction terms at period *t*.

Table 2.2 shows the summary statistics of the related variables used in the predictive regressions. Panel A and B report the statistics for the monthly and quarterly intervals, respectively. The excess monthly and quarterly portfolio returns (r_p) are averaged at 0.007 and 0.021 with the standard deviation of 0.041 and 0.075. Compared to the excess returns on S&P500 (r_{sp500}), my market portfolio returns are somewhat more stable. For instance, the SD of monthly r_p is 0.041 while that of monthly r_{sp500} is 0.043. The skewness of monthly r_p is -0.567 whereas that of monthly r_{sp500} is -0.800. The kurtosis of monthly r_p is 4.045 whereas that of monthly r_{sp500} is 4.662. This slight difference between r_p and r_{sp500} is mainly because my sample consists 100 largest stocks in S&P500 and in turn encounters less effect from smaller stocks.

The average correlation *(AC)* is the value-weighted average of correlations of 100 stocks. The *AC* is moderate at 0.308 for both monthly and quarterly data. That is, the stock prices tend to move in the same direction with the moderate strength. The average variance (*AV*) is the value-weighted cross-sectional average of stock variances and represents the idiosyncratic risk level of the stock market. The mean of monthly *AV* at 0.008 is four times higher than that of the portfolio variance (V_p) at 0.002, while the quarterly *AV* at 0.023 is about three times greater than the V_p at 0.007. This evidence indicates that the market risk itself is much smaller than the total risk measured by *AV*. The idiosyncratic risk, therefore, represents a significant proportion of the total risk and its fluctuation tends to be greater at a higher frequency. The diameter *(DIA)* measures an aspect of the PMFG network that enables us to see the channel through which shocks propagate. The means of *DIA* are very small at 7.323 and 7.290 for monthly and quarterly samples. That is, it takes only seven links on average for one stock to affect the other stocks. The effect of an idiosyncratic shock could then be amplified with just seven links and present a threat to the whole stock market.

The first-order autocorrelation (AR1) of each variable is also reported in Table 2.2. The AR1 of r_p are 0.006 and 0.044 for the monthly and quarterly samples. The AR1 of rs_{p500} are 0.061 and 0.070 for the monthly and quarterly samples. Clearly, both r_p and r_{sp500} are not persistent, and their lagged variables are not likely to have the power to predict the future returns. On the other hand, *DIA*, *AC*, *AV*, and V_p are all relatively persistent with the AR1 of 0.467, 0.601, 0.705, and 0.473 for the monthly sample, 0.263, 0.633, 0.745, and 0.363 for the quarterly sample.

The Augmented Dickey-Fuller statistics are also reported for the unit-root test. To account for one year of information, I used 12 lags and 4 lags for the monthly and quarterly samples. For the monthly statistics, all variables except for the *AC* rejected the null hypothesis of a unit root at 5% critical value. For the quarterly statistics, the *AC* and *AV* exhibit an evidence of a unit root while the others are not.

The correlation matrix in Table 2.2 reports the correlations between returns and independent factors. *AV*, $V_{p,}$ and *AC* are negatively correlated with the contemporaneous portfolio returns, whereas *DIA* is virtually not correlated with r_p. This result suggests that there is a difference between *AC* and *DIA*, regarding the market return information. Also, *AC* and *DIA* are somewhat negatively correlated at -0.389 and -0.483 for the monthly and quarterly sample. This negative relationship indicates that periods of high average correlation are not necessarily the same as periods of low diameter. This result confirms the previous findings that diameter holds some different information from average correlation. Lastly, when compared to the conventional risk measures, the average correlation is more related to systematic risk while the diameter leans towards the idiosyncratic risk.

Table 2.2 Summary statistics for predictive regressions

	Panel A: Monthly Statistics						*Panel B: Quarterly Statistics*					
	r_p	**r_{sp500}**	**DIA**	**AC**	**AV**	**V_p**	**r_p**	**r_{sp500}**	**DIA**	**AC**	**AV**	**V_p**
mean	0.007	0.003	7.323	0.308	0.008	0.002	0.021	0.010	7.290	0.308	0.023	0.007
min	-0.158	-0.186	5.000	0.028	0.001	0.000	-0.190	-0.256	5.000	0.078	0.006	0.001
max	0.106	0.102	11.000	0.741	0.067	0.040	0.208	0.179	11.000	0.669	0.128	0.073
SD	0.041	0.043	1.311	0.139	0.008	0.004	0.075	0.080	1.431	0.121	0.020	0.009
Skew	-0.567	-0.800	0.781	0.542	3.358	5.930	-0.409	-0.791	0.813	0.593	2.470	4.870
Kurt	4.045	4.662	3.053	3.127	18.104	51.957	3.402	3.844	2.876	3.433	10.114	34.265
AR1	0.006	0.061	0.467	0.601	0.705	0.473	0.044	0.070	0.263	0.633	0.745	0.363
ADF	-3.991	-4.137	-3.348	-2.191	-3.048	-3.684	-3.719	-3.820	-3.675	-2.076	-2.664	-3.186
Correlation Matrix							**Correlation Matrix**					
r_p	1.000						1.000					
r_{sp500}	0.984	1.000					0.983	1.000				
DIA	0.006	-0.001	1.000				-0.009	-0.027	1.000			
AC	-0.235	-0.252	-0.389	1.000			-0.276	-0.307	-0.483	1.000		
AV	-0.217	-0.288	0.198	0.198	1.000		-0.358	-0.454	0.239	0.199	1.000	
V_p	-0.260	-0.324	-0.038	0.430	0.852	1.000	-0.404	-0.488	-0.057	0.486	0.844	1.000

(Note) Summary statistics of the 100-stock portfolio are reported in Panel A for monthly sample and in Panel B for the quarterly sample. The sample period is January 1990 to December 2014 (300 monthly observations and 100 quarterly observations). r_p is the log value-weighted portfolio return minus the 3-month Treasury bill. r_{sp500} is the log return of the S&P500 index minus the 3-month Treasury bill. AC is the value-weighted cross-sectional average of the Pearson's correlations of the 100 stocks. DIA is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. AV is the measure of idiosyncratic risk and calculated as the value-weighted cross-sectional average of stock variances. Vp is the portfolio variance. Skew is the skewness, Kurt is the kurtosis, and AR1 is the first-order autocorrelation. ADF is the Augmented Dickey-Fuller statistic calculated with a constant and 12 lags for the monthly sample and 4 lags for the quarterly sample. The critical values for rejection of ADF statistics at five percent levels are -2.879 for the monthly sample and -2.894 for the quarterly sample

The predictive regression results with the full sample period are presented in Table 2.3 (the second quarter of 1990 to the fourth quarter of 2014). Panel A reports the results for the monthly sample, and Panel B reports the results for the quarterly sample. Newey-West t-statistics with the maximum of 6 lags are reported in the brackets. The first model in column 1 indicates that the average correlation (AC_t) has a positive relationship with subsequent portfolio returns ($r_{p,t+1}$). However, the relationship is not robust with the quarterly t-statistic of 1.071 and adjusted R^2 of 0.018. The monthly statistics are even worse with the monthly t-statistic of 0.697 and adjusted R^2 of 0.000. These results contradict the evidence of Pollet and Wilson (2010) that report the significant coefficients of the average correlation. This suggests that the average correlation is not robust to the change in the sample period, at least in this study.

The second specification in column 2 shows that the diameter (DIA_t) is a strong predictor of the subsequent excess portfolio returns, with the robust t-statistics of -1.951 and -3.600 for monthly and quarterly analysis. The negative sign of diameter coefficient indicates the inverse relationship between the diameter and the subsequent portfolio returns, which is fairly intuitive, bearing in mind that the low diameter is an indicator for the star-like network and the high diameter is an indicator for the chain-like network. Since the star-like network allows shocks to propagate more easily than the chain-like network, the low-diameter network should be more fragile and thus demand greater compensation than the chain-like network with higher diameter. The following results reveal that the higher return is actually to compensate for higher idiosyncratic risk amplified by the diameter's functionality.

Compared to the average correlation, the diameter appears to reflect more desirable properties of interconnectedness for predicting portfolio returns. In addition to the evidence above, one standard deviation of the diameter also accounts for the subsequent excess portfolio returns almost twice. If the diameter increases by one standard deviation (1.431), the quarterly returns will decrease by 2.2%, which represents 29.3% of a standard deviation of the portfolio return. On the other hand, when the average correlation increases by one standard deviation (0.121), the quarterly return increases by 1.3%.

The third specification in column 3 includes both average correlation and diameter in the linear regression. With the monthly adjusted R^2 of 0.006 and the quarterly R^2 of 0.069, the model does not add much value in explaining the variation of the subsequent portfolio returns, in comparison to the second specification. Moreover, the power of the model mostly comes from the diameter, as the quarterly t-statistics of the average correlation and diameter are 0.190 and -2.278, respectively.

The models in column 4 and 5 show the predictive regression results of the average stock variance (AV_t) and portfolio variance ($V_{p,t}$). Consistent with Goyal and Santa-Clara (2003) and Pollet and Wilson (2010), the $V_{p,t}$ has a negative relationship with subsequent portfolio returns($r_{p,t+1}$) and insignificant coefficients with the monthly t-statistic of -0.675 and the quarterly t-statistic of -0.023. Similar to Goyal and Santa-Clara, I find a significant relationship between AV_t and $r_{p,t+1}$ but with weaker t-statistics. Nonetheless, the coefficients' sign appears to be negative instead of positive like the Goyal and Santa-Clara's result, possibly due to the difference in data and sample period.

The models in column 6 and 7 test the paper's main hypothesis that the network topology serves as a channel through which idiosyncratic shocks propagate to affect the portfolio returns. In the specification 6, I regress the portfolio returns ($r_{p,t+1}$) on the average stock variance (AV_t) and the interaction term between diameter and average stock variance ($AV_t xDIA_t$). The Newey-West t-statistics of AV_t and $AV_t xDIA_t$ are 2.122 and -3.329 for the monthly data, and 2.308 and -3.628 for the quarterly data. The significant coefficients indicate that the diameter and idiosyncratic shocks, measured by AV_t, work together to affect the portfolio returns. For robustness check, I include the diameter in the specification 7. The interaction term is robust in both monthly and quarterly sample with the monthly t-statistic of -3.231 and the quarterly t-statistic of -1.910, whereas the coefficients of diameter are insignificant. This evidence supports the hypothesis that the diameter works through the idiosyncratic risk to affect the portfolio returns.

The model in column 8 of Table 2.3 assesses the effect of the average correlation (AC_t) in idiosyncratic shock propagation mechanism. It appears that the AC_t does not have a significant effect on the mechanism as the coefficients of the interaction term ($AV_t xDIA_t$) are insignificant with the monthly t-statistic of -0.625 and the quarterly t-statistic of -0.620.

Comparing Panel A to Panel B, I observed the different regressive results of average stock variance (AV_t), average correlation (AC_t) and diameter (DIA_t). In column 4, the t-statistics of AV_t coefficient estimates are -1.938 in Panel A and -1.878 in Panel B. That is, the AV_t is slightly more significant in the short run than in the medium run. This result is expected because the AV_t captures idiosyncratic risk which is likely to fade away or cancel out in the long run. While the effect of the idiosyncratic risk decreases in the longer term, the influence of the systematic risk, measured by AC_t, increases. In column 1, the t-statistic and adjusted $R^2$of AC_t are 1.071 and 1.8% for Panel B whereas they are only 0.697 and 0% for Panel A. This evidence indicates that AC_t has a stronger effect in the longer term. In column 2, the t-statistic and adjusted $R^2$of DIA_t are -3.600 and 7.7% for Panel B whereas they are -1.951 and 1% for Panel A. As a measure of the propagation channel, the DIA_t outperforms the AC_t in both short and medium runs. Additionally, the diameter demonstrates the ability of return prediction in the medium run better than in the short term. This evidence suggests that the propagation channel, captured by the DIA_t, is associated with permanent relationships of stocks more than temporary ones. This property is also supported by the statistics in Table 2.1. Specifically, the diameters in both panels of Table 2.1 have the similar means and do not vary much. However, the AR1 of the diameter is much more persistent in the short run than in the medium run. This finding indicates that the diameter takes time to change and thus does not perform well in the data with high frequency.

Lastly, the role of diameter on the idiosyncratic risk propagation model is revealed in column 6. While the idiosyncratic risk measure has a negative linear relationship with returns, the diameter makes the relationship nonlinear. The graph of the AV_t in the model 6 is a bell shape for both Panel A and B. For example, in Panel A, the predicted returns are estimated as "$0.012 + 2.439 * AV_t - 0.401 * AV_t x DIA_t$." The lower region of the DIA_t (5,6) gives the positive slope. The higher region of the DIA_t (>= 7) gives the negative slope. These results imply that the idiosyncratic risk can transmit to the others when the diameter is low. On the other hand, when the diameter is high, the effect of idiosyncratic risk is lessened, possibly due to diversification. In short, the specification 6 provides some evidence of the shock propagation and how the diameter facilitates the propagation of idiosyncratic shocks.

In summary, the findings from the OLS predictive regressions show that the network topology, measured by the diameter, can affect the stock market returns by serving as the propagation channel for the idiosyncratic risk, measured by average stock variance. The rest of Section 4 provides the analysis of various specifications to check the robustness of the idiosyncratic risk propagation model.

Table 2.3 Predicting subsequent excess portfolio returns from January 1990 to December 2014.

The dependent variable is the monthly excess portfolio returns at *t+1* in Panel A and quarterly excess portfolio returns at *t+1* in Panel B

Panel A: Monthly

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Constant	0.002	0.033	0.031	0.013	0.008	0.012	0.000	0.002
	[0.300]	[2.588]	[1.405]	[4.868]	[3.775]	[3.960]	[-0.005]	[0.295]
AC_t	0.017		0.005					0.036**
	[0.697]		[0.170]					[2.022]
DIA_t		-0.004*	-0.003				0.002	
		[-1.951]	[-1.474]				[0.834]	
AV_t				-0.683*		2.439**	3.127**	-0.394
				[-1.938]		[2.122]	[2.202]	[-0.614]
$V_{p,t}$					-0.458			
					[-0.675]			
$AV_t xDIA_t$						-0.401**	-0.496**	
						[-3.329]	[-3.231]	
$AV_t xAC_t$								-0.926
								[-0.625]
Adj R^2	0.000	0.010	0.006	0.014	-0.002	0.036	0.034	0.016
N	299	299	299	299	299	299	299	299

Panel B: Quarterly

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Constant	-0.010	0.135	0.123	0.037	0.022	0.033	0.077	-0.010
	[-0.347]	[4.354]	[1.698]	[4.571]	[2.675]	[3.205]	[1.891]	[-0.337]
AC_t	0.104		0.021					0.161*
	[1.071]		[0.190]					[1.844]
DIA_t		-0.016**	-0.015**				-0.006	
		[-3.600]	[-2.278]				[-1.173]	
AV_t				-0.627*		2.610**	1.668	-0.449
				[-1.878]		[2.308]	[1.288]	[-0.732]
$V_{p,t}$					-0.022			
					[-0.023]			
$AV_t xDIA_t$						-0.405**	-0.274*	
						[-3.628]	[-1.910]	
$AV_t xAC_t$								-0.898
								[-0.620]
Adj R^2	0.018	0.077	0.069	0.019	-0.010	0.088	0.083	0.045
N	99	99	99	99	99	99	99	99

(Note) Table 2.3 presents the results of one-period ahead predictive regressions of the excess value-weighted portfolio returns on lagged variables. AC_t is the value-weighted cross-sectional average of the Pearson's correlations of the 100 stocks. DIA_t is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. AV_t is the measure of idiosyncratic risk and calculated as the value-weighted cross-sectional average of stock variances. $V_{p,t}$ is portfolio variance. $AV_t xDIA_t$ is the interaction term between average stock variance and diameter. $AV_t xAC_t$ is the interaction term between average stock variance and average correlation. N is the number of observations. Newey-West t-statistics with the maximum of 6 lags are reported in the brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

2.4.4 Predicting excess portfolio returns with different horizons

This subsection tests how long the diameter and the idiosyncratic risk propagation model can influence the portfolio returns. To achieve this goal, I regressed the independent variables at time t on the annualized k-month excess portfolio returns such that $r_{p,t+k} = \left(\frac{12}{k}\right) * \sum_{j=1}^{k}(r_{p,t+j} - r_{f,t+j})$. For instance, the dependent variable of the 3-month model is the sum of log portfolio returns minus the 3-month Treasury Bill returns from $t+1$ to $t+3$, whereas the 24-month model uses the sum of log excess portfolio returns from $t+1$ to $t+24$. These sum returns are then annualized by the factor $12/k$, which enables me to compare the coefficients across all horizons.

Table 2.4 presents the results of the predictive regressions on the annualized k-month excess portfolio returns. In Panel A, I use the monthly diameter to predict the portfolio returns at different frequencies. As expected, the coefficients of the diameter decline monotonically from the 3-month model to the 48-month model. The regressions also suggest that the diameter can hold the information about the portfolio returns up to 24 months with the t-statistic of -1.847 and adjusted R^2 of 6%. After this point, the adjusted R^2 statistics decline and the coefficients are insignificant. Consistent with the results in Section 2.4.3, the negative regression coefficients across all time horizons suggest that the diameter has an inverse relationship with the portfolio returns.

In Panel B, I test the robustness of the idiosyncratic risk propagation concept at different frequencies by regressing the average stock variance (AV_t), diameter (DIA_t) and the interaction ($AV_t x DIA_t$). The coefficients of AV_t and the interaction decrease as the returns are extended into the future. The adjusted R^2 increases to a peak at 24 months, which accounts for 19.5% of the variation in returns. Then, the adjusted R^2 decreases monotonically. Similar to the specification of Panel A, the result suggests that the specification of Panel B has information relevant up to two-year portfolio returns. Moreover, during the two years, the diameter coefficients are insignificant while the coefficients of AV_t and the interaction are significant. These findings support my hypothesis that the diameter works through the idiosyncratic risk.

It should be noted that the adjusted R^2 is artificially high because of the nature of annualized accumulated returns. When k increases, the standard deviation of the returns will decrease, and in turn the monthly independent variables will be able to capture a better proportion of the return variation. While this kind of test is useful to indicate a trend in the future, one should be careful in using the R^2 statistics to infer the model fit.

Table 2.4 Regressions of the excess portfolios returns with different horizons.

Panel A: $r_{p,t+k} = a_{t+k} + bDIA_t + e_{t+k}$

Horizon (months)	**1**	**3**	**6**	**9**	**12**	**18**	**24**	**30**	**36**	**42**	**48**
Constant	0.401	0.456	0.440	0.415	0.357	0.272	0.263	0.226	0.179	0.147	0.119
	[2.588]	[4.141]	[3.907]	[4.368]	[3.959]	[2.980]	[2.795]	[2.716]	[2.610]	[2.443]	[1.978]
DIA_t	-0.043*	-0.050**	-0.048**	-0.045**	-0.037**	-0.025**	-0.024*	-0.019	-0.013	-0.009	-0.005
	[-1.951]	[-3.237]	[-2.934]	[-3.232]	[-2.908]	[-2.015]	[-1.847]	[-1.643]	[-1.340]	[-1.031]	[-0.592]
Adj R^2	0.010	0.051	0.087	0.109	0.092	0.055	0.060	0.046	0.023	0.010	0.001
N	299	297	294	291	288	282	276	270	264	258	252

Panel B: $r_{p,t+k} = a_{t+k} + b_1AV_t + b_2DIA_t + b_3AV_txDIA_t + e_{t+k}$

Horizon (months)	**1**	**3**	**6**	**9**	**12**	**18**	**24**	**30**	**36**	**42**	**48**
Constant	-0.001	0.319	0.194	0.192	0.138	0.036	0.005	-0.003	-0.015	-0.020	-0.043
	[-0.005]	[2.248]	[1.385]	[1.704]	[1.431]	[0.391]	[0.051]	[-0.045]	[-0.222]	[-0.312]	[-0.658]
AV_t	37.523**	10.155	25.232**	23.287**	22.424**	23.275**	25.129**	21.739**	17.764**	15.318**	14.590**
	[2.202]	[0.691]	[2.386]	[3.044]	[3.268]	[4.329]	[4.219]	[5.183]	[4.640]	[4.458]	[4.375]
DIA_t	0.020	-0.026	-0.012	-0.012	-0.005	0.010	0.014	0.015	0.017*	0.017*	0.020**
	[0.834]	[-1.437]	[-0.613]	[-0.756]	[-0.345]	[0.814]	[1.146]	[1.471]	[1.745]	[1.761]	[2.121]
AV_txDIA_t	-5.958**	-1.993	-3.664**	-3.310**	-3.221**	-3.414**	-3.665**	-3.249**	-2.715**	-2.346**	-2.237**
	[-3.231]	[-1.255]	[-3.077]	[-3.577]	[-3.769]	[-4.819]	[-4.745]	[-5.304]	[-4.935]	[-4.743]	[-4.762]
Adj R^2	0.034	0.068	0.127	0.154	0.149	0.153	0.195	0.187	0.148	0.121	0.117
N	299	297	294	291	288	282	276	270	264	258	252

(Note) Table 2.4 presents the results of one-period ahead predictive regressions of the excess value-weighted portfolio returns on lagged variables. The dependent variable is the sum of k-month excess value-weighted portfolio returns. AV_t is the measure of idiosyncratic risk and calculated as the value-weighted average of stock variances. DIA_t is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. AV_txDIA_t is the interaction term between average stock variance and diameter. N is the number of observations. Newey-West t-statistics with the maximum of 6 lags are reported in the brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

2.4.5 Out-of-sample predictability

The objective of this subsection is to assess whether or not forecasting power of the idiosyncratic risk propagation model still holds in the out-of-sample exercises. Following McCracken (2007), I use the one-period-ahead forecasts with a recursive scheme. Specifically, the sample is divided into two sets for initial and evaluation periods. I denote R as the number of initial periods and T as the total sample periods. The number of evaluation periods, P, is then $T – R +1$. Starting from period $t = R$, the parameters are estimated and used to forecast the excess portfolio returns at $t+1$. The process continues until the last sample period T. Under the recursive scheme, the OLS estimation will use all available information from the first period to t.

The common approach to evaluate the forecasting accuracy is to compare the out-of-sample forecasts of the restricted and unrestricted models. The null hypothesis is equal to the forecasting accuracy or the forecast encompassing the restricted and unrestricted models. In this paper, the restricted model is autoregressive of order 1 or AR(1) of the portfolio returns as shown in Equation (2.5). The unrestricted models are the AR(1) plus the first lag of the studied variables. Equation (2.6) provides the forecasts of the predictors (x_t), including average correlation, diameter, and average stock variance. Equation (2.7) provides the forecasts of the idiosyncratic risk propagation model.

$$r_{p,t+1} = b_{1,1} + b_{1,2}r_{p,t} + u_{1,t+1} \tag{2.5}$$

$$r_{p,t+1} = b_{2,1} + b_{2,2}r_{p,t} + b_{2,3}x_t + u_{2,t+1} \tag{2.6}$$

$$r_{p,t+1} = b_{2,1} + b_{2,2}r_{p,t} + b_{2,3}AV_t + b_{2,4}DIA_t + b_{2,5}AVDIA_t + u_{2,t+1} \tag{2.7}$$

To test the null hypothesis, I use four measures of forecasting accuracy, including the out-of-sample R2, dRMSE, ENC-NEW, and MSE-F. The out-of-sample R2 is computed from 1 minus the ratio of the mean squared forecasted errors over the mean squared errors from the restricted model. The dRMSE is the difference between the root mean squared of the restricted model, and the root mean squared of the unrestricted model. The positive R2 and dRMSE mean that the forecast accuracy of the unrestricted model is superior to that of the restricted model in predicting the subsequent excess portfolio returns. ENC-NEW is a formal test of forecast encompassing applied to 1-step ahead prediction and tests the null hypothesis that the restricted model forecast encompasses the unrestricted model. Clark and McCracken (2001) show that the asymptotic distribution of ENC-NEW statistics depend on the ratio of the evaluation period and initial period ($\pi =$ P/R) and provide the asymptotic critical values accordingly. MSE-F is the F-type test of out-of-sample predictive ability concerning loss function. McCracken (2007) shows that the asymptotic distribution of MSE-F statistics also depends on π. To determine the significance of ENC-NEW and MSE-F, I use the asymptotic critical values available on Clark and McCracken (2001) and McCracken (2007), respectively.

Table 2.5 reports statistics of the out-of-sample tests as described above. According to Hansen and Timmermann (2012), the forecasting performance depends on how the data set is split. I, therefore, consider two forecasting periods. For the monthly sample in Panel A, the long out-of-sample forecast begins in January 2000 (P = 180), and the corresponding R and π are 120 and 1.5. The short one starts in September 2006 (P = 100), and the corresponding R and π are 200 and 0.5. For the quarterly sample in Panel B, the long out-of-sample forecast starts at 2000 Q1 (P = 60), and the corresponding R and π are 60 and 1.5. The short one begins in 2006 Q2 (P = 35), and the corresponding R and π are 65 and 0.54.

The findings for the long evaluation period ($\pi = 1.5$) provide strong evidence of out-of-sample predictability of average correlation, diameter and average stock variance in comparison to the lagged portfolio returns. They all have positive out-of-sample R^2 and dRMSE. The ENC-NEW and MSE-F statistics are also significant at 95% confidence interval. The idiosyncratic risk propagation model as in Equation (2.7) shows the strongest forecasting power with the monthly out-of-sample R^2 of 10.2% and the monthly dRMSE of 6.7%.

The findings for the short evaluation period, on the other hand, provide mixed evidence of the out-of-sample forecasting ability. For a monthly interval, only average stock variance can increase the predictive power of the lagged returns, whereas average correlation and diameter cannot. For the quarterly interval, all predictors, except for average correlation, are significant at 95% confidence interval. Similar to the long-period forecasting exercise, the idiosyncratic risk propagation model exhibits the strongest out-of-sample ability for both monthly and quarterly samples.

In short, the diameter is sufficiently robust to the quarterly sample exercises but not in the case of monthly sample exercises. The idiosyncratic risk propagation model, on the other hand, is robust to the out-of-sample tests in both intervals. The findings strongly support the hypothesis that network topology, measured by diameter, can help idiosyncratic risk, measured by average stock variance, to predict the subsequent portfolio returns.

Table 2.5 Forecasting out-of-sample excess portfolio returns.

Panel A: Monthly

	Average correlation		**Diameter**		**Average stock variance**		**AV-DIA model**	
	π = 1.5	π = 0.5	π = 1.5	π = 0.5	π = 1.5	π = 0.5	π = 1.5	π = 0.5
Out-of-sample R^2	1.160	-0.224	2.601	-0.072	4.159	4.622	10.205	6.714
dRMSE	0.025	-0.005	0.055	-0.001	0.089	0.098	0.221	0.143
ENC-NEW	2.029**	0.253	3.801**	0.655	6.112**	4.006**	16.434**	4.578**
MSE-F	2.112**	-0.223	4.807**	-0.071	7.811**	4.846**	20.457**	7.197**

Panel B: Quarterly

	Average correlation		**Diameter**		**Average stock variance**		**AV-DIA model**	
	π = 1.5	π = 0.54	π = 1.5	π = 0.54	π = 1.5	π = 0.54	π = 1.5	π = 0.54
Out-of-sample R^2	4.037	-1.449	12.104	8.978	5.386	5.627	18.146	16.429
dRMSE	0.161	-0.055	0.493	0.347	0.215	0.215	0.751	0.648
ENC-NEW	3.031**	0.490	6.892**	3.300**	3.200**	1.908**	10.852**	4.554**
MSE-F	2.524**	-0.500	8.262**	3.452**	3.415**	2.087**	13.302**	6.881**

(Note) Table 2.5 presents the results of one-period ahead forecasts with the recursive scheme. The dependent variable is the excess value-weighted portfolio returns. The first three columns compare the restricted model with the unrestricted model in Equation (2.6). Average correlation is the value-weighted average of the Pearson's correlations of the 100 stocks. Diameter is the longest distance of all geodesic paths in a given network. Average stock variance is the value-weighted average of stock variances. AV-DIA model is the idiosyncratic risk propagation model as in Equation (2.7). Measures of forecasting accuracy or forecasting encompassing are out-of-sample R^2, dRMSE, ENC-NEW and MSE-F as described in Section 4.5. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

2.4.6 Predictive regressions with controlled variables

Table 2.6 reports the regression estimates of diameter, average stock variance and the interaction term when various predictors are presented. The goal is to ensure that the primary results are unaffected and still robust in the presence of the controlled variables. The first controlled factor is the lagged excess portfolio return (r_p). The second (r_f) is the 3-month Treasury Bill. Yield is the dividends relative to the S&P500 index price (*d/p*). TED spread is retrieved from Federal Reserve Bank of St. Louis and calculated as the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. Term spread is also retrieved from Federal Reserve Bank of St. Louis and calculated as the difference between 10-year Treasury Constant Maturity and 3-month Treasury Constant Maturity. $Size_t$ is defined as the market capitalization of the S&P500 index. Book to market (BM_t) is the ratio of book value to the market capitalization of the S&P500 index.

Specification 1 is the same as the regression in model 7 of Table 2.3 implying the significant effect of the idiosyncratic risk propagation model on portfolio returns. Specifications 2 to 8 regress the idiosyncratic risk propagation model with each predictor. Across the models, the coefficients of diameter are constantly insignificant while the coefficients of average stock variance and interaction term are significant. Lagged excess portfolio returns, Risk-free rates, Dividend yield, TED spread, Size and BM have insignificant coefficients and do not improve the adjusted R^2 of the based model. Only Term spread is significant with the t-statistic of -2.038 and the adjusted R^2 of 4.5%.

Specification 9 controls all seven predictors. The diameter coefficient remains insignificant. The coefficients of the average stock variance (AV_t) and the interaction term ($AV_t xDIA_t$) are relatively stable and significant with the Newey-West t-statistics of 2.354 and -2.984, respectively. Compared to the first specification, the adjusted R^2 improves from 3.4% to 4.6%. Term spread and Size appear to contain distinct information about the portfolio returns in addition to the idiosyncratic risk propagation model. These results indicate that the changes in the future returns are associated with not only some global shocks but also some sufficient idiosyncratic shocks interacting with network topology. This evidence confirms the previous findings of the importance and role of the network topology in the market portfolio returns.

Table 2.6 Regressions of the excess portfolios returns with controlled variables. The dependent variable is the monthly excess portfolio returns at *t+1*

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Constant	0.000	0.000	-0.003	0.002	0.000	0.005	0.007	0.007	0.077
	[-0.005]	[0.008]	[-0.232]	[0.136]	[0.000]	[0.326]	[0.414]	[0.310]	[2.084]
DIA_t	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.001
	[0.834]	[0.826]	[0.850]	[0.869]	[0.835]	[1.149]	[0.616]	[0.806]	[0.351]
AV_t	3.127**	3.099**	3.388**	3.289**	3.139**	3.929**	2.865**	3.370**	2.856**
	[2.202]	[2.175]	[2.512]	[2.531]	[2.497]	[2.826]	[2.018]	[2.635]	[2.354]
$AV_t xDIA_t$	-0.496**	-0.494**	-0.533**	-0.520**	-0.497**	-0.612**	-0.460**	-0.532**	-0.463**
	[-3.231]	[-3.187]	[-3.629]	[-3.605]	[-3.356]	[-3.929]	[-2.973]	[-3.721]	[-2.984]
$r_{p,t}$		-0.009							-0.020
		[-0.144]							[-0.327]
$r_{f,t}$			1.370						-6.190
			[1.083]						[-1.441]
$Yield_t$				-0.143					-0.541
				[-0.352]					[-0.452]
TED_t					0.000				-0.003
					[-0.025]				[-0.213]
$Term_t$						-0.004**			-0.012**
						[-2.038]			[-2.689]
$Size_t$							0.000		0.000**
							[-1.110]		[-2.349]
BM_t								-0.017	0.012
								[-0.480]	[0.173]
Adj R^2	0.034	0.031	0.035	0.032	0.031	0.045	0.033	0.032	0.046
N	299	299	299	299	299	299	299	299	299

(Note) Table 2.6 presents the results of one-period ahead predictive regressions of the excess value-weighted portfolio returns on lagged variables. DIA_t is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. AV_t is the measure of idiosyncratic risk and calculated as the value-weighted average of stock variances. $AV_t xDIA_t$ is the interaction term between average stock variance and diameter. "$r_{f,t}$" is the 3-month Treasury Bill. $Yield_t$ is the dividend to price ratio of the S&P500 index. TED_t is the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. $Term_t$ is the spread between 10-year Treasury Constant Maturity and 3-month Treasury Constant Maturity. $Size_t$ is the market capitalization of the S&P500 index at the end of the month. BM_t is the book-to-market ratio of the S&P500 index at the end of the month. N is the number of observations. Newey-West t-statistics with the maximum of 6 lags are reported in the brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

2.4.7 Predictive regressions with alternative network measures

This last subsection presents the regression estimations of the alternative network measures. The objective is to assess if the implication of interconnectedness still holds when alternative network measures are applied instead of the diameter. The alternative measures are average closeness centrality, average eccentricity centrality, average eigenvalue centrality, and average KNN centrality.

Table 2.7 reports the regression estimates of the alternative network measures in comparison to the diameter. The first two models in the table are the same as the model 2 and 7 in Panel A of Table 2.3. The specification 1 estimates the effect of the idiosyncratic risk, measured by the average stock variance, on the market returns. The specification 2 tests the effect of diameter and its role as the propagation channel of the firm-specific shocks. The main implication of this model is that the diameter reflects the network topology of the portfolio, which in turn serves as the propagation channel of idiosyncratic risk. Without going into detail, average correlation is another measure of interconnectedness which appears to be insignificant in the propagation mechanism as mentioned in Section 2.4.3.

Specification 3 uses the average closeness centrality as the measure of interconnectedness rather than the diameter. In network theory, closeness measures how easily a node can reach the other nodes. A node with high closeness centrality means that it takes shorter distance to reach all other nodes in the network. Since I use the equal-weighted PMFG network to calculate the network measures, changes in the closeness centrality are informative about the network topology although not as direct and straightforward as the diameter. High average cross-sectional closeness centrality indicates that the stocks are closed and the stock network would lean to the star-like configuration. Similar to the diameter model, the closeness coefficient is not significant, and the interaction between average stock variance and closeness is significant with the t-statistic of 1.750 and adjusted R^2 of 2%.

Specification 4 uses the average eccentricity centrality as the measure of interconnectedness. The eccentricity of a node is its shortest distance to the farthest node in the network. Similar to diameter and closeness, the average eccentricity centrality can reflect network topology of a stock market. Its meaning is much like the diameter. That is, low eccentricity would indicate the star-like structure, and high eccentricity is more chain-like. The regression coefficient of the eccentricity is negative but insignificant. The significant interaction term with the t-statistic of -2.635 suggests that the eccentricity works through the idiosyncratic risk to affect the subsequent portfolio returns.

Specification 5 employs the concept of eigenvector centrality, which measures the relative influence of a node in a network. The node attached to the high-scoring nodes will have a higher level of eigenvector centrality than a connection to the low-scoring nodes. This definition makes the eigenvector centrality closer to the average correlation and less informative about the network topology. Similar to the other network measures, the coefficient of the individual eigenvector is not significant with the t-statistic of 0.845. Unlike the diameter, the eigenvector centrality does not work with the idiosyncratic risk to affect the market with the interaction-term t-statistic of 0.173 and adjusted R^2 of 1.5%.

Lastly, Specification 6 tests the role of KNN centrality for the idiosyncratic risk propagation mechanism. KNN stands for K-nearest neighbors and is calculated as the average nearest neighbor degree of the given nodes with K distance. For simplicity, I use only adjacent neighbors and make K equal to 1. In the setting of this paper, high average KNN centrality means that this node tends to connect to the high degree nodes rather than low degree nodes. The network topology implication of KNN is mixed because the high KNN nodes might occur only in part of the network and sometimes result in a chain-like structure. Moreover, the regression estimation indicates that the KNN does not fit the role of the propagation channel with insignificant coefficients and the adjusted R^2 of 1.1%.

In summary, the evidence for interconnectedness in the idiosyncratic risk propagation mechanism is not unanimous. Closeness and Eccentricity are statistically significant while AC, Eigenvalue, and KNN are not. These findings suggest that not all aspects of the interconnectedness are relevant to the portfolio returns, and diameter is the best among other tested network measures, at least in my sample.

Table 2.7 Regressions of the excess portfolios returns with alternative network measures.

The dependent variable is the monthly excess portfolio returns at *t+1*

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	0.013	0.000	0.007	0.007	-0.075	-0.009
	4.868	-0.005	0.314	0.426	-0.730	-0.460
Average stock variance	-0.683*	3.127**	-3.518**	2.638*	-2.541	-0.763
(AV)	[-1.938]	[2.202]	[-2.246]	[1.757]	[-0.233]	[-0.423]
Diameter		0.002				
		[0.834]				
Closeness			0.015			
			[0.214]			
Eccentricity				0.001		
				[0.323]		
Eigenvalue					1.147	
					[0.845]	
KNN						0.002
						[1.180]
(AV)x(Diameter)		-0.496**				
		[-3.231]				
(AV)x(Closeness)			9.701*			
			[1.750]			
(AV)x(Eccentricity)				-0.544**		
				[-2.635]		
(AV)x(Eigenvalue)					25.343	
					[0.173]	
(AV)x(KNN)						0.014
						[0.079]
Adj R^2	0.014	0.034	0.020	0.030	0.015	0.011
N	299	299	299	299	299	299

(Note) Table 2.7 presents the results of one-period ahead predictive regressions of the excess value-weighted portfolio returns on lagged variables. Average stock variance is the measure of idiosyncratic risk and calculated as the value-weighted average of stock variances. Diameter is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. Closeness, Eccentricity, Eigenvalue, and KNN are network measures as defined in Section 4.7. The interaction terms are the interaction between average stock variance and the network measures. N is the number of observations. Newey-West t-statistics with the maximum of 6 lags are reported in the brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

2.5 Concluding remarks

The relationship between idiosyncratic risk and market portfolio returns has been supported by mixed evidence from the existing literature. Goyal and Santa-Clara (2003), in particular, advocate the significant forecasting power of idiosyncratic risk, measured by average stock variance, while Pollet and Wilson (2010) provide evidence against this relationship. The typical rationale against this relationship is that idiosyncratic shocks are diversified away in the market portfolio and thus should not affect the market returns. However, from the network perspective, the diversification argument is not always true. The idiosyncratic shocks can sometimes propagate to the other stocks and in turn affect the aggregate fluctuation of the portfolio returns. Thus, the network structure serves as a propagation channel of the idiosyncratic risk with amplification or diversification functionality.

Although a number of network measures are studied in existing literature, I have focused on a simple network concept, called "diameter." This network measure is particularly useful in capturing the network topology or pattern of connections among the stocks. I conjecture that the network topology has a favorable impact on the relationship between idiosyncratic risk and portfolio returns. The empirical evidence suggests that the diameter is a good indicator for interconnectedness that functions as the channel through which idiosyncratic risk propagates. In particular, the findings strongly support the interaction of diameter and average stock variance to predict the subsequent excess portfolio returns. The interaction term between the two predictors is statistically significant in both monthly and quarterly samples. It is, also, robust to the different horizons of the returns and out-of-sample exercises. Moreover, the effect of the idiosyncratic risk propagation model is not affected by the well-known predictors such as the dividend yield, TED spread, Term spread, market capitalization and book-to-market ratio.

This work opens up many interesting areas of research regarding the relationship between network structure and asset pricing models. As mentioned earlier, the diameter is just one of many network measures, and the network topology is just one of several network characteristics. The other network measures with unique network information may be complementary or even substitute the diameter. Furthermore, the homogeneous and static behavior of the network members is assumed in this paper. Accordingly, incorporation of agent-based modeling and allowance of network evolution over time are promising areas for future research. The caveat, though, is that this kind of work will inevitably complicate the subject further in comparison to this study.

2.6 Appendix

2.6.1 Average stock variance

This appendix provides some additional tests of the average stock variance. In section 2.2.3, the average stock variance is used as a proxy of idiosyncratic component of the stock market. The underlying reason is pointed out by Goyal and Santa-Clara (2003). The average stock variance, as shown in Equation (2.2), is a direct measure of the total risk that can be decomposed into systematic and idiosyncratic components. Since the systematic component is arguably much smaller than the idiosyncratic component, the average stock variance is virtually an approximation of idiosyncratic risk. This section reports additional evidence on this issue. Table 2.8 shows the percentage of the idiosyncratic component to the total risk measured by the average stock variance. In the first two rows, I directly use the data of average stock variance and the market portfolio variance from the original sources (Goyal and Santa-Clara 2003; Pollet and Wilson 2010). In the third and fourth rows, the systematic component is the stock market portfolio's variance in accordance with Equation (2.1). The idiosyncratic component of the first four rows is estimated from the subtraction between total risk and systematic component. In the last two rows, I estimate the idiosyncratic component from the variance of residual of the one-factor model whereby the market return is the sole factor.

In Table 2.8, we can see that the overall idiosyncratic component is approximately 70% or above. Goyal and Santa-Clara get the higher number at 94.06% because they use most stocks in CRSP to calculate the parameters. Moreover, they use the equal-weighted approach to construct the variable, while this thesis and Pollet and Wilson use the value-weighted approach. These two factors tend to decrease the weight of the systematic component and increase the weight of the idiosyncratic component. Pollet and Wilson (2010) get the second highest at 77.79%, which is a bit higher than my data. This is expected because my data comprises of the largest 100 stocks rather than the largest 500 stocks in Pollet and Wilson (2010).

Table 2.8 Percentage of idiosyncratic component to the total risk.

			Total risk	Systematic component	Idiosyncratic component	
			Average stock variance	Market portfolio variance	Estimated/Residual variance	
Goyal and Santa-Clara	Monthly 1962-1999	All CRSP	0.029	0.002	0.027	**94.06%**
Pollet and Wilson	Quarterly 1963-2006	Largest 500 CRSP	0.022	0.005	0.017	**77.79%**
Thesis	Monthly 1990-2014	Largest 100 SP	0.008	0.002	0.006	**70.00%**
Thesis	Quarterly 1990-2014	Largest 100 SP	0.023	0.007	0.016	**69.87%**
Thesis	Monthly residual variance	Largest 100 SP	0.008	-	0.006	**70.65%**
Thesis	Quarterly residual variance	Largest 100 SP	0.023	-	0.017	**73.40%**

(Note) The data in first two rows is directly retrieved from Goyal and Santa-Clara (2003) and Pollet and Wilson (2010). In third and forth rows, the average stock variance is calculated from Equation (2.2). The systematic component is the stock market portfolio's variance in accordance with Equation (2.1). The idiosyncratic component of the first four rows is estimated from the subtraction between total risk and systematic component. The systematic component is the stock market portfolio's variance in accordance with Equation (2.1). The idiosyncratic component is estimated from the subtraction between total risk and systematic component. In the last two rows, I estimate the idiosyncratic component from the variance of residual of the one-factor model whereby the market return is the sole factor.

Nonetheless, the systematic component is the embedded part of the average stock variance and may take over the total risk in some periods. In particular, the crisis period is known to have a high degree of the global shocks that are considered the systematic risk. Therefore, there is a need to test the periods of crisis and non-crisis separately. Similar to the section 2.4.2, I define the crisis period as the contraction period in accordance to NBER, including July 1990 – March 1991, March 2001 – November 2001, and December 2007 – June 2009. Table 2.9 reports the percentage of the idiosyncratic component to the total risk in two subsamples. Panel A uses the data from the non-contraction period. Panel B uses the data from the contraction period. As expected, the portion of the idiosyncratic component in the non-contraction period is in general higher than that in the contraction period. In the average stock variance approach, the number is about 71% in the non-contraction period while it is about 64% in the contraction period. Similarly, in the one-factor approach, the number is about 75% in the non-contraction period while it is about 60% in the contraction period. Thus, regarding the idiosyncratic risk information, the average stock variance approach is better than the one-factor approach. All in all, the average stock variance in the thesis is a good proxy for the idiosyncratic component. Nevertheless, since the portion of the systematic risk is fairly large in the contraction period, it will be interesting to see the use of different idiosyncratic component measures in the future research.

Table 2.9 Percentage of the idiosyncratic component of the total risk in two subsamples.

		Total risk	**Systematic component**	**Idiosyncratic component**	
Panel A Non-contraction periods					
		Average stock variance	Market portfolio variance	Estimated/Residual variance	
Average variance approach	monthly	0.007	0.002	0.005	**71.61%**
Average variance approach	quarterly	0.020	0.006	0.015	**71.50%**
One-factor approach	monthly	0.007	-	0.005	**74.11%**
One-factor approach	quarterly	0.020		0.016	**76.91%**
Panel B: Contraction periods					
		Total risk	**Systematic component**	**Idiosyncratic component**	
		Average stock variance	Market portfolio variance	Estimated/Residual variance	
Average variance approach	monthly	0.015	0.005	0.010	**64.85%**
Average variance approach	quarterly	0.044	0.016	0.029	**64.83%**
One-factor approach	monthly	0.015	-	0.009	**59.88%**
One-factor approach	quarterly	0.044	-	0.028	**62.56%**

(Note) This table shows the percentage of the idiosyncratic component in two subsamples. Panel A uses the data from the non-contraction period. Panel B uses the data from the contraction period according to NBER. In the first two rows, the average stock variance is calculated from Equation (2.2). The systematic component is the stock market portfolio's variance in accordance with Equation (2.1). The idiosyncratic component of the first four rows is estimated from the subtraction between total risk and systematic component. The systematic component is the stock market portfolio's variance in accordance with Equation (2.1). The idiosyncratic component is estimated from the subtraction between total risk and systematic component. In the last two rows, I estimate the idiosyncratic component from the variance of residual of the one-factor model whereby the market return is the sole factor

Chapter 3 Predicting the probability of extreme negative returns: A network approach

Abstract This paper proposes a network model to predict the probability of the extreme negative returns of global stock markets during 2000 to 2015. The extreme negative return is defined as the bottom five percent of the country's return distribution. In the network model, the global market can be demonstrated in a large network where countries are connected by some kinds of relationships. A country-specific shock then propagates through the cross-country linkages to the other countries, which may result in the extreme negative situation in the respective countries. This paper focuses on the properties of the propagation channel and studies the role of the network structure in determining the probability of the extreme negative returns. I find that the network measures themselves have the weak ability to identify or predict the extreme situations. Rather, the results suggest that the network measures help a measure of country-specific shocks to improve the ability to predict the probability of the extreme negative returns.

Keywords: Extreme negative returns, Financial network, Diameter, Country-specific shocks, Extreme value analysis.

JEL classifications: D85, F36, G15

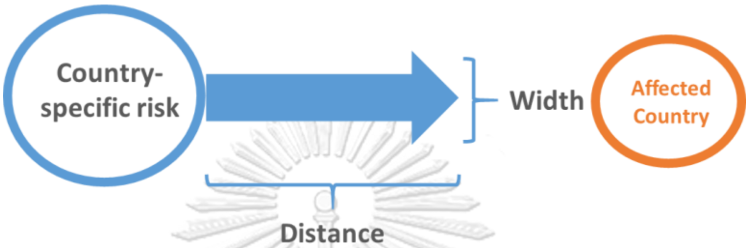
3.1 Introduction

The US subprime mortgage crisis in 2007-2008 is the extremely negative event that affects not only the US but also other countries in the world. Similarly, the Greek government-debt crisis is another event where the downfall of one country can potentially threaten the stability of other financial markets. The common interesting feature of both systemic events is the phenomenon where an extremely negative shock can spillover and affect other countries. This phenomenon is called “contagion.” The primary objective of this paper is to study the probability of the extreme negative returns that are associated with contagion and in particular examine the factors that could identify and predict the probability of the extreme negative returns.

To begin with, in the context of the international equity market, the extreme negative returns is used to study contagion. In fact, there are many ways to define and study contagion in the existing literature. For example, Masson (1999) considers an event as contagion if the market co-movement is not explained by global shock and linkages through normal trade and economic relationships. Boyer et al. (2006) define contagion as the excess correlation between stock markets. Forbes and Warnock (2012) classify contagion if a country-specific shock causes changes in another country’s gross capital inflows or outflows. Forbes (2012) provides a thorough summary of contagion and discussed several definitions of contagion. She suggests that a concept of the extreme negative returns is suitable for the broad definition of contagion. Specifically, the contagion can be broadly defined as the transmission of an extreme negative shock in one country to others through numerous real and financial channels. This broad definition is particularly appropriate for the use of citizens and policymakers because it is quite straightforward and can be measured in real time. Due to this reason, I am motivated to study the relationship between extreme negative returns and network structure which is the essential element of contation.

According to the broad definition of contagion, both US and Greek crises are naturally classified as contagion because the extreme negative shocks are actually transferred and cause extreme negative events in the other countries. Figure 3.1 demonstrates the contagion process in which a country-specific shock travels through the propagation channel to the affected country. Forbes (2012) explains that a contagion may occur when one or more countries experience extreme negative returns. The coincidence of countries with extreme negative events (ENR_{all}) is, then, naturally associated with contagion. Forbes (2012) provides empirical evidence that this measure of contagion is robust and statistically significant in identifying the probability of the extreme negative returns. The intuition for this measure is straightforward. Contagion is considered occurring if several stock markets experience extreme negative returns at the same time. The high percentage of countries facing extreme negative returns means that the extremely negative shocks do hit those countries. The affected countries would then have the high probability of the extreme negative returns. If the extreme negative events are not caused by global shocks, then they are likely to result from contagion. After controlling the global shocks, Forbes finds that the ENR_{all} is still significant and even dominates the effect of global shocks. In this paper, I also reach the same conclusion that ENR_{all} is the robust measure for measuring the probability of the extreme negative returns in the contemporary periods.

Figure 3.1 A Simple Network Model with Two Countries



(Note) This figure demonstrates the contagion components of a simple network model, including the country-specific risk, width or strength of propagation channel, distance of propagation channel and the affected country

However, while the coincidence of extreme negative returns (ENR_{all}) is a natural indicator for the probability of each country's extreme negative return in the same period, it may not be the case for the prediction. Specifically, the extreme negative returns in one period may not be a good predictor of the probability of the extreme negative returns in the next period. Instead, this article proposes that the underlying structure that allows the transmission of the extreme negative shocks is more relevant to the future return situations. From Figure 3.1, one can easily see that the ENR_{all} is not the only factor that can affect contagion. The country-specific shock and the propagation channel are two components that come before the ENR_{all}. Without the sufficient degree of both factors, the possibility of contagion is likely to decrease substantially. In the existing literature, the shock and propagation channel are two important ingredients of contagion. For example, Elliott et al. (2014) report that the large idiosyncratic shock can result in the initial failure, which in turn triggers cascades of failures through the propagation channel in the financial architecture. Acemoglu et al. (2015) study different network architectures of financial interdependencies. They show that different network structures are associated with different levels of interconnectedness, some of which can facilitate the shock propagation and even turn it into a systemic event. Likewise, Bisias et al. (2012) emphasize that the network approach can explain how a systemic event unfolds and usually be regarded as an early warning of the systemic event.

This paper focuses on the measures of idiosyncratic risk and propagation channel in addition to the percentage of countries with extreme negative returns on the stock markets (ENR_{all}). The measure of idiosyncratic risk is straightforward and calculated as the equal-weighted average of stock markets' variances, using within-month daily data. To quantify the properties of the propagation channel, I use the techniques from the network theory. Specifically, I view the global market as a network whereby stock markets are connected by their correlation of returns. In a simplest two-countries setting, a country-specific shock can propagate to another via the channel of return correlation between two stock markets. Thus, the probability of future extreme negative return will depend on country-specific shock, strength/width of the propagation channel and distance of the channel, as shown in Figure 3.1.

With this background in mind, I construct the main research question of the paper as follows: Does the network structure have an influence on the extreme negative returns in the immediately subsequent period? I expect that the measure of the network structure will have the power to predict the extreme negative returns. In particular, the underlying structure of the global market serves as the channel through which shocks propagate. Since the procedure of shock propagation and amplification is a part of the buildup of the systemic risk, the network approach is forward-looking and useful in tracking and monitoring threats. The network measures should then have some degree of forecast power. I find that the measures of propagation channel are rather weak to identify and predict the probability of extreme negative returns, while a measure of country-specific risk is significant. However, once the network measures interact with the idiosyncratic risk measure, their predictive power increases significantly. As for the identification of the contemporary extreme negative returns, the interactions are also significant, but they do not significantly improve the performance of the idiosyncratic risk measure.

The contribution of this paper is two-fold. First, the percentage of the sample with extreme negative returns (ENR_{all}) is only moderately significant to predict the probability of the extreme negative returns. More importantly, I apply the network approach to the extreme value analysis. I find that the network measures can improve the performance of idiosyncratic risk measure to predict the probability of the extreme negative returns. The rest of the paper is divided into five sections. In Section 3.2, I will explain the methodologies applied in constructing the stock networks and generating related variables. Section 3.3 presents the description of data and some basic statistics. Section 3.4 reports the empirical results from logistic regressions. Section 3.5 concludes the paper.

3.2 Methodology

To achieve the main goal of the study of the relationship between network structure and extreme negative return, it is necessary to define the extreme negative return, the relevant context and measures. Section 3.2.1 explains the extreme negative returns and its significance in contagion. Afterward, Section 3.2.2 explains an approach to study the probability of extreme negative returns and the methods to construct the studied measures.

3.2.1 Extreme negative return and contagion

Following Forbes (2012), I define an extreme negative return as the bottom 5% of that country's monthly return distribution. Apparently, the extreme negative return and contagion are closely related for several reasons. First, according to Forbes (2012), the contagion is defined as the transmission of an extreme negative shock in one country to one another or more countries. If many countries experience the extremely negative events in a given period, there is a high possibility of contagion. This version of contagion does not concern how the shock transmits nor what channel it employs. Rather, the occurrence of contagion depends on the number of countries affected by the shock. Second, the extreme negative shock can also affect the other countries without the actual propagation. General citizens do not care how the shocks are propagated to them and simply care about its effect on them. The citizens, for example, may fear that negative events in other countries would go out-of-control and affect them. Thus, without the actual propagation, the shock would still affect other markets. Last but not least, although there are many ways to define contagion, policymakers and governments may find it difficult to identify various forms of contagion promptly. This board form of contagion is straightforward and flexible enough to satisfy the needs of policymakers and governments. Thus, Forbes (2012) believe that the measure of contagion is significant in explaining the extreme negative returns.

3.2.2 Measuring the probability of extreme negative returns

In this paper, the extreme value analysis is used for three reasons. First, the extreme value approach directly tests the probability of extreme negative returns in the tail distribution of returns. Second, the extreme value analysis is robust to different assumptions about return distribution. Since this approach focuses on the behavior of the tail distribution, it does not matter if the whole distribution is normal or asymmetry. Last but not least, contagion is often not limited to crisis periods. With the suitable environment, an extreme negative event can sometimes occur in the non-crisis periods. Nonetheless, this extreme value approach has two disadvantages. The sample of extreme returns is inevitably small. To mitigate the problem, I apply the panel data analysis to the extreme value approach. Another problem is that extreme returns in multiple markets may not result from contagion measure or network measures but global shocks. To investigate this issue, I also include some global shock measures in the analysis.

Following Forbes (2012), I examine the possibility of the coincidence of extreme negative returns by using the binomial logistic regression. The outcome variable is the dichotomous variable of extreme negative returns, $ENR_{i,t}$. A country experiences an extremely negative period when its stock market experiences an extreme negative return. The $ENR_{i,t}$ is defined as the bottom 5% of that country's monthly return distribution. $ENR_{i,t}$ is 1 if country *i* is experiencing an extremely negative return in month *t*. $ENR_{i,t}$ is 0 if otherwise. The predictor variables of interest are the percentage of countries with extreme negative returns, $ENR_{all,t}$, a measure of country-specific risk, a measure of propagation channel strength and a measure of propagation channel distance. The first measure is a benchmark which is directly borrowed from Forbes' paper. The other measures are the extension of the based model and therefore the contribution of this article.

3.2.2.1 Percentage with an extreme negative return

According to Forbes (2012), ENR_{all} is the natural factor to identify the contagion and therefore will be served as the benchmark factor. The rationale behind this measure is that it reflects the impact of the idiosyncratic shocks on other countries. If an idiosyncratic shock affects only a respective country and not the others, the number of countries facing extreme negative returns should not exceed 5% of the sample in each period. However, the fact that some periods have high ENR_{all} indicates that the idiosyncratic shocks could propagate to the other countries and results in extreme negative returns of` multiple countries. Therefore, without taking into account the spillover channels whatsoever, the ENR_{all} can display the impact of idiosyncratic shocks and be a good indicator for contagion identification.

3.2.2.2 A measure of country-specific risk

Following Goyal and Santa-Clara (2003), I construct a measure of country-specific risk from the equal-weighted average of the members' variance in a portfolio, as expressed in Equation (3.1).

$$AV_t = \frac{1}{N_t}\sum_{i=1}^{N_t} V_{i,t}, \tag{3.1}$$

$$V_{i,t} = \sum_{d=1}^{D_t} r_{i,d}^2 + 2\sum_{d=2}^{D_t} r_{i,d}r_{i,d-1}, \tag{3.2}$$

where AV_t is the measure of country-specific risk in month *t*. N_t is the number of stocks used in the calculation of period *t*. $V_{i,t}$ is the monthly variance of the stock market of the respective country i in month t. The formula of $V_{i,t}$ is shown in Equation (3.2) and is calculated from the daily returns within month t. $r_{i,d}$ is the return of the stock market i on day d. D_t is the number of the trading day in month t. The second term on the right-hand side is used for the autocorrelation adjustment in daily returns.

According to Goyal and Santa-Clara (2003), AV_t is originally the measure of total risk of a portfolio and can be decomposed into the idiosyncratic component and systematic component. Because the systematic component is much smaller than the idiosyncratic component, the AV_t is virtually an approximation of idiosyncratic risk.

3.2.2.3 Measures of cross-country linkages

Measures of cross-country linkages are relatively different from the ENR_{all}, as they primarily focus on the cross-country linkages and transmission of shocks. Specifically, the cross-country linkages serve as channels through which idiosyncratic shocks propagate. Changes in these links would directly influence the extent to which the idiosyncratic shocks affect members in a network. In this paper, I use two measures of such linkages, including cross-market correlation and diameter. The first one is used to capture the strength of the propagation channel, whereas the latter is used to estimate the distance that the shocks would travel.

Cross-market correlation is calculated as the equal-weighted average of correlations of stock market returns using a 52-week rolling sample of 43 market returns. I use weekly return instead of daily returns to avoid the non-synchronous timing among the countries. By this construction, the average correlation implies how strong the market returns are moving together in aggregate level. In terms of cross-country linkages, this implication of the average correlation is directly associated with the strength of propagation channel. If the average correlation is high, a country-specific shock can easily affect the counterparties. However, the average correlation does not capture how the shock transmits from one country to the rest of the network. In addition, the cross-correlation is a crude measure of contagion with some limitations as discussed by Forbes and Rigobon (2002), Ang and Chen (2002) and Forbes (2012). Equation (3.3) show the formula for the average correlation.

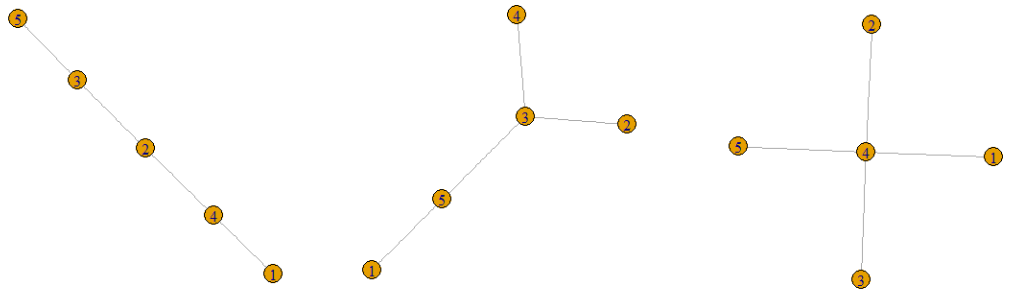
$$AC_t = \frac{1}{n}\sum_{i=1}^{N}\sum_{k\neq i}\rho_{ik,t} \quad , \tag{3.3}$$

where $\rho_{ik,t}$ is the Pearson's correlation between stock i and k. N is the number of stocks used in the calculation, equal to 43. n is the number of correlation excluding the diagonal elements. In this case, n is 1,806. One distinct advantage of the average correlation is that it is also an indicator of systematic risk. Pollet and Wilson (2010) empirically show that it is better than many measures of systematic risk to predict the subsequent US stock market returns. Thus, the average correlation is complementary to the measure of country-specific risk (AV_t).

Diameter is a network concept that can reveal how the linkages are connected and formed a network structure. By knowing the pattern of the network structure, we can quantify the aspect of propagation channels that allow idiosyncratic shocks to reach all members of a network. The standard definition of the diameter is the longest distance of all geodesic paths in a given network. The geodesic path is the lowest number of links between two nodes. Consequently, the diameter is an intuitive measure for the distance of propagation channel. To illustrate, I generate a simple network of 5 markets as shown in Figure 3.2. Figure 3.2a is a chain-like structure that has the longest diameter of 4 as measured by the number of links of the largest geodesic path. Figure 3.2b is a tree-like structure with a diameter of 3. Figure 3.2c is a star-like network with the shortest diameter of 2. The average correlation is just an average number of links and clearly not sufficient to capture the network pattern. The diameter, on the other hand, provides direct information about the pattern and thus the distance of propagation channel.

Unlike the average correlation, the calculation of the diameter is not as straightforward and required incorporation from the field of the network theory. I carry out this task in three steps. First, I estimate a matrix of correlation of returns, just like when I construct the average correlation. A correlation matrix is constructed from 52-week rolling returns and updated at the end of the month. Creating a network from this full correlation matrix is possible. However, the network will be so noisy and complicated that it is impossible to see any pattern of connections as shown in Figure 3.3a. Second, to compress such a complex network, the correlation matrix is filtered by an algorithm, called Planar Maximally Filtered Graph (PMFG). This network algorithm is introduced by Tumminello et al. (2005). The algorithm starts by ordering the correlations from high to low. Then, the highest correlation is picked first, followed by the next highest correlation as long as the graph can be drawn on a 2-D surface without link crossing. If an additional link does not satisfy the planarity constraint, the link is removed, and the process continues to the next link. The total number of connections in a network is reduced from $n(n-1)/2$ to $3(n-2)$. The filtered network then contains the essence of stock interrelationships which forms the network topology as shown in Figure 3.3b. The noisy and weak relationships in Figure 3.3a would then be eliminated and in fact transformed into long-distance paths in Figure 3.3b. Lastly, the diameter is calculated by counting the number of links between the two furthest nodes in Figure 3.3b. The network structure is updated at the end of each month from January 2000 to December 2015. The sample has a total of 192 monthly networks as well as the list of respective diameters.

Figure 3.2 Illustration of Five-stock Network



a) Chain-like network b) Tree-like network c) Star-like network

(Note) This figure illustrates the possible patterns of the connected network with the minimal number of links. There are three patterns, including Chain-like network in a), Tree-like network in b) and Star-like network in c).

3.3 Description of data

In order to study contagion of the global market, I work with the stock market returns of 43 countries. The data includes 22 developed markets of MSCI World and 21 emerging markets of MSCI Emerging markets. The developed markets include Canada (can), United States (usa), Austria (atr), Belgium (bel), Denmark (den), Finland (fin), France (fra), Germany (ger), Ireland (ire), Italy (ita), Netherland (net), Norway (nor), Portugal (por), Spain (spn), Sweden (swe), Switzerland (swz), United Kingdom (uk), Australia (aus), Hongkong (hk), Japan (jpn), Newzealand (nz), Singapore (sgp). The emerging markets are Brazil (bra), Chile (chl), Columbia (col), Mexico (mex), Peru (per), Czech (cz), Egypt (egp), Greece (gre), Hungary (hun), Poland (pol), Russia (rus), South Africa (sfr), Turkey (tur), China (chn), India (ind), Indonesia (idn), South Korea (kor), Malaysia (may), Philippines (phl), Taiwan (tai), Thailand (tha). The daily stock returns are generated from the difference in log prices between two consecutive days, collected from Bloomberg Terminal. Monthly returns are calculated from the within-month daily returns. The sample period is January 2000 to December 2015.

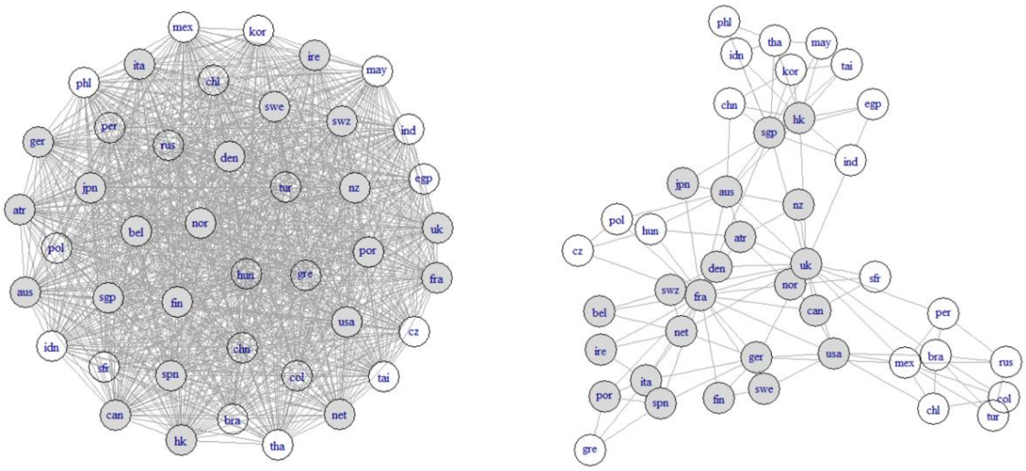
Table 3.1 reports the summary statistics of the 43 market on a monthly basis. The average monthly return of the MSCI world is very low at 0.0001, whereas that of the MSCI Emerging markets is much higher at 0.0038. On the other hand, the average standard deviation of the MSCI world markets at 0.056 is lower than that of the MSCI emerging countries at 0.0773. As expected, the developed markets are more stable than the emerging, but the return on investment is also lower during the sample period.

Figure 3.3b depicts the global network of 43 markets during the sample period. This network is constructed from the correlations of weekly returns and then filtered by the PMFG algorithm. Grey nodes represent stock markets listed in the MSCI world.

White nodes are stock markets listed in the MSCI emerging market. The network provides us with some useful information that cannot be seen from the standard statistics. Firstly, the markets tend to cluster around each other in the same region. For example, the markets in Asia are very close to each other. An idiosyncratic shock takes only 1 or 2 links to reach the others. Likewise, the markets in South America tend to cluster around each other with Brazil at the center. Thus, a country-specific shock would first transmit to the countries in the same region and then propagate through the other regions of the world. Furthermore, the developed markets in gray are at the center while the emerging markets tend to be on the outer part of the network. That is, shocks from developed markets, in general, will travel across the network faster than those of emerging markets. For instance, a shock from the USA can affect countries in Europe and South America simultaneously. A shock from Greece, on the other hand, only has a direct effect on countries in Europe. Nevertheless, if the shock from Greece is large enough, it can travel throughout the network and cause a systemic event.

Based on these observations, the network structure can be regarded as propagation channel through which idiosyncratic shocks propagate. The network measures are then closely related to the contagion of the idiosyncratic shocks. The next section provides some empirical evidence for the relationship between the extreme negative returns and network measures.

Figure 3.3 The Global Network of the Full Sample



a) Correlation-based complete network   b) Correlation-based PMFG network

(Note) This figure depicts two networks of 43 countries using the weekly returns from January 2000 to December 2015. A node represents a country and a link represents a correlation of returns between two countries' stock market returns. The gray nodes are developed markets, and the white nodes are emerging markets. Figure 3.3a depicts a complete network constructed from the full correlation matrix. Figure 3.3b depicts a PMFG network constructed from
the PMFG-filtered correlation matrix.

Table 3.1 Data Description of Stock Market Returns

MSCI World						
	mean	**min**	**max**	**SD**	**Skew**	**Kurt**
can	0.0023	-0.1827	0.1098	0.0431	-0.9569	5.3745
usa	0.0016	-0.1893	0.1029	0.0444	-0.7060	4.3977
atr	0.0001	-0.3651	0.1759	0.0688	-1.5217	8.9239
bel	0.0003	-0.3527	0.1391	0.0602	-2.0281	10.8566
den	0.0073	-0.1963	0.1699	0.0544	-0.6857	4.6972
fin	-0.0046	-0.3711	0.2776	0.0856	-0.6265	5.6669
fra	-0.0007	-0.1741	0.1201	0.0519	-0.5763	3.6462
ger	0.0001	-0.2867	0.1796	0.0632	-0.8971	5.5605
ire	-0.0035	-0.2576	0.1519	0.0658	-0.8747	4.3763
ita	-0.0033	-0.1672	0.1715	0.0578	-0.3660	3.3919
net	-0.0001	-0.2039	0.1233	0.0552	-1.0049	4.8832
nor	0.0029	-0.2760	0.1253	0.0638	-1.2148	6.3943
por	-0.0052	-0.2228	0.1236	0.0556	-0.7603	4.7598
spn	-0.0006	-0.1923	0.1616	0.0608	-0.4157	3.8996
swe	0.0011	-0.2067	0.2025	0.0639	-0.5039	4.5523
swz	0.0010	-0.1324	0.1036	0.0407	-0.6499	3.4962
uk	-0.0004	-0.1398	0.0847	0.0411	-0.6825	3.6877
aus	0.0029	-0.1166	0.0808	0.0383	-0.6310	3.1808
hk	0.0019	-0.2434	0.1576	0.0621	-0.6068	4.5301
jpn	-0.0004	-0.2365	0.1192	0.0527	-0.5326	4.4692
nz	0.0002	-0.1575	0.1006	0.0419	-0.5694	3.8747
sgp	-0.0003	-0.3079	0.1917	0.0599	-1.1258	7.5177
Average	0.0001	-0.2263	0.1442	0.0560	-0.8153	5.0972

MSCI Emerging Markets						
	mean	**min**	**max**	**SD**	**Skew**	**Kurt**
bra	0.0008	-0.3908	0.2498	0.1032	-0.6291	4.4144
chl	0.0044	-0.1463	0.1458	0.0446	0.0072	3.4206
col	0.0096	-0.3362	0.2201	0.0883	-0.4182	3.6517
mex	0.0054	-0.3671	0.1548	0.0695	-0.9564	6.4564
per	0.0081	-0.4470	0.2368	0.0865	-0.7753	6.1492
cz	0.0044	-0.2633	0.2064	0.0660	-0.3462	4.6833
egp	0.0092	-0.3745	0.3113	0.0943	-0.2262	4.1714
gre	-0.0183	-0.4474	0.2225	0.1016	-0.8166	4.9496
hun	0.0023	-0.4086	0.1790	0.0793	-0.8689	6.1298
pol	-0.0001	-0.2729	0.1932	0.0692	-0.2092	3.9757
rus	0.0031	-0.4350	0.3189	0.1074	-0.5558	4.5502
sfr	0.0085	-0.1803	0.1317	0.0501	-0.3122	3.6287
tur	0.0074	-0.4122	0.4453	0.1076	-0.0234	4.8937
chn	0.0030	-0.2601	0.1730	0.0807	-0.6878	3.9101
ind	0.0081	-0.2847	0.2520	0.0737	-0.5076	4.4275
idn	0.0094	-0.3643	0.1917	0.0772	-0.7295	5.5040
kor	0.0044	-0.2355	0.2363	0.0697	-0.2100	4.0430
may	0.0036	-0.1632	0.1381	0.0464	-0.3573	4.3300
phl	0.0044	-0.2398	0.1552	0.0633	-0.3504	3.7911
tai	-0.0012	-0.2454	0.2348	0.0680	-0.1327	4.3178
tha	0.0038	-0.3661	0.2621	0.0773	-0.7076	6.4905
Average	0.0038	-0.3162	0.2218	0.0773	-0.4673	4.6614

(Note) The data consists of 22 markets in the MSCI World index and 21 markets in the MSCI Emerging markets index. The statistics are calculated monthly using within-month daily returns. “SD” is standard deviation, “Skew” is the skewness, and “Kurt” is the kurtosis.

3.4 Empirical results

3.4.1 Explanatory variables over time

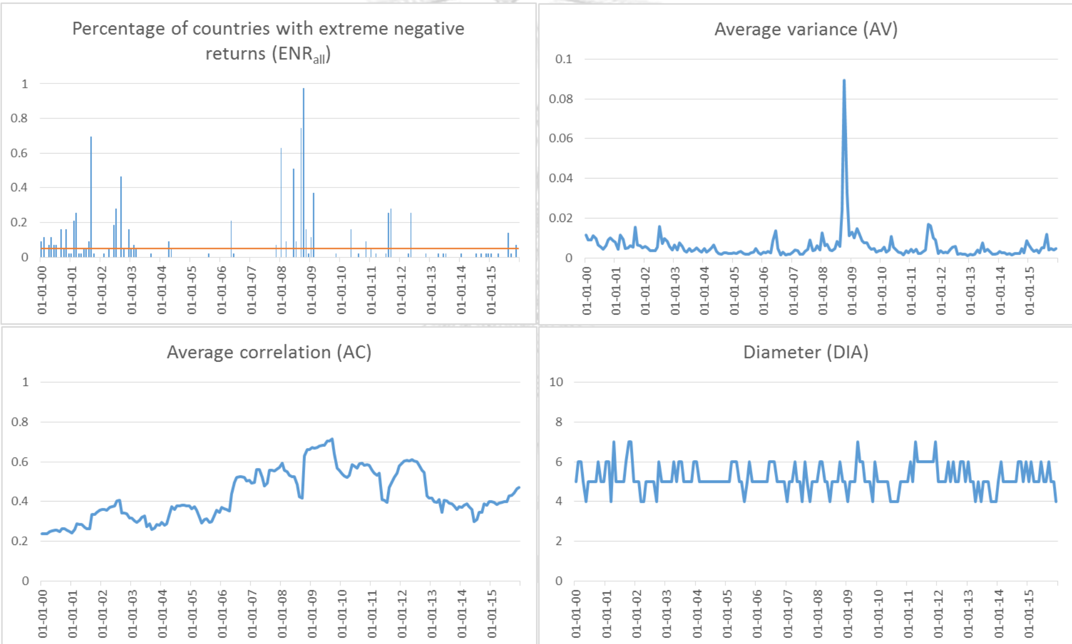
To identify and predict the contagion, I focus on four explanatory variables, including the percentage of the sample with extreme negative returns (ENR_{all}), average correlation (*AC*), average variance (*AV*) and diameter (*DIA*). The ENR_{all} is the original measure in Forbes' model, and the remaining variables are the extension of the model.

Figure 3.4 depicts monthly dynamic of the four variables over time. The bar chart of ENR_{all} reports the percentage of countries with an extremely negative return in each month *t*. The straight line at 5% is the threshold for classifying the extreme negative returns. If there is no contagion, the ENR_{all} should not exceed 5%. Instead, the ENR_{all} chart clearly shows that the extreme negative returns are not evenly distributed. For instance, in September 2008, the ENR_{all} reached 74%. Clearly, not only did the collapse of Lehman's Brothers affect the USA but also caused the transmission of an extreme negative shock throughout the international markets. In this regard, the evidence is consistent with Forbes (2012) and presents the ENR_{all} as a potential identifier of contagion.

AV graph shows the dynamic of the equal-weighted average variance of the stock markets' returns. The graph indicates that the country-specific risk shows no trend but clear bursts. The spikes are corresponding to the extreme events in the sample period, especially the financial crisis in 2008. For this reason, *AV* is also a potential candidate for measuring extreme negative returns.

The graphs of *AC* and *DIA* depict the measures of cross-country linkages over time. Unlike ENR_{all} and *AV*, they are constructing by using the 52-week rolling sample and are updated every month. Thus, in theory, they should be more capable of capturing a trend in international markets. As expected, the *AC* graph demonstrates a certain trend for the global market. It has an upward trend from 2000 to 2009 and then a downward trend after 2009. Other things equal, if the *AC* is high, it is easier for an idiosyncratic shock to affect many markets and even the entire network. For example, the financial crisis in 2008-2009 was associated with the high level of *AC*, and a country-specific shock can easily affect the other countries. On the other hand, the *DIA* graph shows no trend but some small spikes throughout the sample period. Its fluctuation pattern even somewhat resembles the behavior of ENR_{all} and *AV*, albeit lesser degree. Therefore, it is clear that *DIA* captures the different aspect of propagation channel from the *AC*. Furthermore, among the other tested measures, the *DIA* is the only one that has discrete values ranging from 4 to 7. The lowest diameter at 4 means that it takes only four steps for a country-specific shock to reach all members of the network. The largest diameter at 7 indicates that the shock would take more time and effort to affect the whole network. In general, the network's diameter will be at 5 and 6 about 60% and 25% of the time. When the diameter reaches 4, an idiosyncratic shock would propagate easier, and the likelihood of contagion should be higher. However, the *DIA* graph of Figure 3.4 shows that the diameter of 4 includes the periods of both normal and crisis time. The reason is that the diameter is not directly related to the market returns. Theoretically, it serves as a measure of propagation channels through which a shock propagates. As a result, to assess the effect of the diameter more accurately, I incorporate a measure of idiosyncratic risk into the analysis of diameter in Section 3.4.3.

Figure 3.4 Dynamic of Contagion Measures



(Note) The figure presents the dynamic of ENR_{all}, AV, AC, and DIA over time from January 2000 to December 2015. "ENR_{all}" is the percentage of countries that experience extreme negative returns in each month. "AV" is the measure of idiosyncratic risk, calculated as the equal-weighted cross-sectional average of stock variances. "AC" is the equally-weighted cross-sectional average of the Pearson's correlations of the 43 markets. "DIA" is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network

3.4.2 Extreme value analysis

Based on Forbes (2012), this paper uses an approach of extreme-value analysis to explain and predict the probability of extreme negative returns. I define this change as the bottom 5% of that market's monthly return distribution, $ENR_{i,t}$. The extreme negative returns can be caused by a number of factors, including contagion measures, idiosyncratic shocks, and global shocks. Forbes (2012) proposes one contagion measure, called the percentage of the extreme negative returns. More specifically, the extreme negative return may be caused by the transmission of the extreme negative shock from the other countries. As a result, the infected stock market will encounter an abnormal change in returns and sometimes result in the extreme negative return. In this paper, I extend Forbes' model by incorporating the measures from the network model as mentioned in Section 3.2.2 and 3.2.3. To assess the effect of the explanatory factors, I estimate the conditional probability of extreme negative returns by using the logistic regression in Equation (3.4).

$$Prob\left(ENR_{i,t} = 1\right) = F(const + b_1 ENR_{all,t} + b_2 AV_t + b_3 AC_t + b_4 DIA_t), \quad (3.4)$$

ENR_{it} is the binary outcome and equal to 1 if country i has an extremely negative return in month t. $ENR_{all,t}$ is the percentage of countries with extremely negative returns in month t. Forbes (2012) provided empirical evidence of its ability to identify financial contagion. Thus, $ENR_{all,t}$ will serve as the benchmark and control factor of the other models. AV_t is the average variance in month t and represents an idiosyncratic risk of the international market. AC_t is the average correlation in month t. It captures the co-movement of market returns as well as the strength of cross-country linkages. DIA_t is the diameter of a global network in month t. The diameter quantifies a pattern of connections and distance of propagation channel into a categorical number. The lower the number is, the closer the network topology is to a star-like network, which has a short distance of the propagation channel. The high diameter, on the other hand, indicates the chain-like network that has a long distance of the propagation channel

Table 3.2 reports regression results of the full sample from January 2000 to December 2015. Panel A and B provide the regression results of the cross-country extreme returns in month t and $t+1$, respectively. The main purpose of Panel A is to address the research question regarding the identification of the extreme negative returns. The goal of Panel B is to study the effect of the independent variables on the prediction of the extreme negative returns. Z-statistics are reported in brackets and McFadden's pseudo R^2 is also provided in Table 3.2. Another important estimation is an area under a ROC curve (AUC). The AUC measures the ability of the model to classify binary outcomes correctly. The basic concept of AUC is to calculate the area under the ROC curve (Receiver Operating Characteristic). This curve is a graphical plot that illustrates the performance of a binary classifier system. The construction of the graph relies on the actual outcomes and predicted outcomes. The actual binary outcome, $ENR_{i,t}$, is 1 if a market's return is extremely negative and 0 otherwise. The predicted outcomes are calculated from the probability provided by the logistic regression in Equation (3.4). If the fitted probability exceeds a certain threshold, the respective outcome will be 1. With the actual and predicted outcomes, we can determine the statistics of true positive, true negative, false positive and false negative. The false positive rate (FPR) is the X-axis in the ROC space, and the true positive rate (TPR) is the Y-axis. As the discrimination threshold is varied from 0 to 1, we will get a set of coordination based on the FPR and TPR. When plotted on the ROC space, the ROC curve is formed and usually concave downward. If the curve is closed to the upper vertical axis, the AUC will be closed to 1. The tested model with AUC of 1 represents the perfect test. If the graph is the diagonal line, the *AUC* is 0.5, and the model is indifferent from a constant model.

In Panel A, the constant model is reported to have McFadden's pseudo R^2 of 0% and AUC of 0.500. These statistics will be the benchmark for the tested model afterward. In column 2, the percentage of markets with extreme negative returns (ENR_{all}) has a strongly positive coefficient with a z-statistic of 27.681. Its R^2 is also very high at 37.8%. Thus, when the number of countries experiencing a distress period increases, the probability of extreme negative returns increases as well. Additionally, with the AUC of 94.6%, the ENR_{all} is extremely efficient to identify the extreme negative events.

In column 3, the average stock market variance (AV) is statistically significant to explain the extreme negative returns, with a z-statistic of 19.797 and R^2 of 20.7%. The average variance also has a high level of AUC at 85.0%, which is much greater than the constant model. This indicates that a period with high idiosyncratic risk is likely to result in extreme negative returns. Nevertheless, compared to the ENR_{all}, the AV is somewhat less effective.

In column 4 and 5, the coefficients of the network measures are all insignificant. Their R^2 is statistically zero, and their AUC around 50% is practically indifferent from the constant model. This finding implies that the individual network measures are not directly related to the extreme negative returns in the contemporary period.

Lastly, when all factors are included in column 6, the McFadden's R^2 and AUC are 38.7% and 93.6%, respectively. Since the numbers are very similar to the ENR_{all} model in column 2, the result strongly supports the dominant effect of the ENR_{all} on the identification of the extreme negative returns.

Panel B shows that the constant model in column 1 has the R^2 of 0% and the AUC of 50%. Both ENR_{all} and *AV* in column 2 and 3 are statistically significant in explaining the extreme negative returns at month *t+1*. The coefficients of both ENR_{all} and *AV* are significant and positive with the z-statistics of 12.914 and 8.817. Clearly, both measures can not only identify but also predict the extreme negative events. However, the ability of both models to predict the outcomes at month *t+1* is much lower than that the ability to identify the contemporary outcomes in Panel A. The AUC of ENR_{all} is 62.3%, and the AUC of *AV* is 68% in Panel B.

In column 3 and 4 of Panel B, I find that the coefficients of *AC* and *DIA* are statistically significant with the z-statistics of -4.856 and -3.331, respectively. Nevertheless, the values of AUC are still low at 54.8% for *AC* and 54.1% for *DIA*. The R^2 is also very low and in fact almost indifferent from the constant model. Therefore, the individual effects of the tested network measures are not sufficient to identify and predict the probability of the extreme negative returns.

Table 3.2 Extreme value analysis – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B

Panel A: Dependent variable is the dummy variable $ENR_{i,t}$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-2.939	-4.078	-4.294	-3.242	-2.614	-2.605
	[-58.340]	[-49.235]	[-45.391]	[-10.935]	[-6.613]	[-4.463]
% with extreme negative		8.699**				7.909**
returns, $ENR_{all,t}$		[27.681]				[20.960]
Average variance,			177.997**			56.590**
AV_t			[19.797]			[4.593]
Average correlation,				2.584		-9.847**
AC_t				[1.042]		[-3.382]
Diameter,					-0.063	-0.128
DIA_t					[-0.827]	-[1.312]
R-squared	0.000	0.378	0.207	0.000	0.000	0.387
Area under ROC curve	0.500	0.946	0.850	0.528	0.500	0.936
Observations	8,256	8,256	8,256	8,256	8,256	8,256

Panel B: Dependent variable is the dummy variable $ENR_{i,t+1}$						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-2.944	-3.177	-3.157	-1.612	-1.597	-0.051
	[-58.160]	[-55.000]	[-53.974]	[-5.893]	[-3.957]	[-0.102]
% with extreme negative		2.842**				2.893**
returns, $ENR_{all,t}$		[12.914]				[8.345]
Average variance,			31.319**			4.356
AV_t			[8.817]			[0.735]
Average correlation,				-11.629**		-14.226**
AC_t				[-4.856]		[-5.789]
Diameter,					-0.262**	-0.299**
DIA_t					[-3.331]	[-3.629]
R-squared	0.000	0.039	0.018	0.007	0.003	0.054
Area under ROC curve	0.500	0.623	0.68	0.548	0.541	0.651
Observations	8,213	8,213	8,213	8,213	8,213	8,213

(Note) This table presents the results of logistic regressions of the dummy variable of extreme negative returns on contagion measures in Equation (3.4). "$ENR_{all,t}$" is the percentage of countries with extreme negative returns on month t. "AV_t" is the measure of idiosyncratic risk, calculated as the equal-weighted cross-sectional average of stock variances. "AC_t" is the equally-weighted cross-sectional average of the Pearson's correlations of the 43 markets. "DIA_t" is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. "R-squared" is McFadden's pseudo R-squared. "Area under ROC curve" is the statistic to determine the ability to classify contagion. Z-statistics are reported in brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

3.4.3 The effect of idiosyncratic risk propagation model

This section uses the concept of idiosyncratic risk propagation to identify and predict the probability of the extreme negative returns. In Section 3.4.1, we learn that the network measures are inefficient to measure the extreme negative returns in both contemporary and subsequent periods. Rather, I conjecture that the network structure serves as channels through which idiosyncratic shocks propagate. Without the incorporation of these shocks, the network measures by themselves are not sufficient to measure the probability of the extreme negative returns. To capture this risk propagation concept, I extend the Equation (3.4) by adding the interaction terms between the average variance and network measures, as shown in Equation (3.5).

$$Prob(ENR_{i,t} = 1) = F\begin{pmatrix} const + b_1 ENR_{all,t} + b_2 AV_t + b_3 AC_t + b_4 DIA_t \\ +n_1 AV_t * AC_t + n_2 AV_t * DIA_t \end{pmatrix} \quad (3.5)$$

$ENR_{i,t}$ is a dummy variable equal to 1 if the stock market of country i has an extremely negative return in month t. $ENR_{all,t}$ is the percentage of the sample with extremely negative returns in month t. AV_t is the average variance that measures idiosyncratic risk of the international market in month t. AC_t is the average correlation that measures the strength of propagation channel. DIA_t is the diameter of a network. The diameter reveals a pattern of connections that determines paths of idiosyncratic shock propagation. The effects of idiosyncratic risk propagation are captured by the interaction terms between average variance and network measures. AV_t*AC_t represents the effect of idiosyncratic shocks when interacting with the strength of propagation channel. AV_t*DIA_t represents the effect of idiosyncratic shocks when interacting with diameter. Table 3.3 provides empirical evidence for measuring the current extreme negative returns in Panel A and the future extreme negative returns in Panel B. The first two models in Table 3.3 assess how well each of the measures of cross-country linkages can facilitate the transmission of idiosyncratic shocks.

In column 1 and 2 of Panel A, the AUC of the *AC* and *DIA* models are 83.1% and 84.1%, which are lower than the single factor model of *AV* in Table 3.2. Even if both AC and *DIA* are included in column 4 and 5, the AUC remains at around 85%. That is, the network measures are not capable of improving the probability of the extreme negative returns of the average variance. In column 6, the regression includes all tested factors and interaction terms as well as the benchmark factor, ENR_{all}. I find that the ENR_{all} is indeed a dominant factor for measuring the probability of the extreme negative returns in the contemporary period.

In Panel B, I find that the average correlation and the diameter are both capable of helping the average variance to predict the probability of the extreme negative returns. For reference, the base model is the single factor model with AV_t as a sole predictor. The regression estimates of AV_t in column 2 of Table 3.2 are 1.8% for McFadden's pseudo R^2 and 68% for AUC. In column 1, I extend the base model by including AC_t and AV_t*AC_t. This purpose is to test the role of AC_t as a propagation channel strength of the country-specific shock, AV_t. The interaction term, AV_t*AC_t, is statistically significant with the z-statistic of -10.052. The McFadden's R^2 is 5.9%, and the AUC is 69.7%. Both statistics are clearly higher than those of the one-factor *AV* model. The result advocates the role of average correlation, *AC*, in helping the average variance, *AV*, to predict the probability of the extreme negative returns.

In column 2, I extend the base model by including DIA_t and AV_t*DIA_t. The goal is to assess the role of diameter on the propagation of the country-specific shock. The interaction term, AV_t*DIA_t, is statistically significant with the z-statistic of -9.135. The McFadden's pseudo R^2 is 4.9%, and the AUC is 71.1%. Similar to the average correlation, the diameter also helps the average variance to predict the probability of the extreme negative returns. Moreover, the negative sign of AV_t*DIA_t indicates that when the distance is short, the shock tends to cause the extreme negative return in the subsequent period.

In column 3, I extend the individual AV_t model by including AV_t*AC_t and AV_t*DIA_t. This model enables me to test the strength and distance of propagation channel simultaneously. The coefficients of both interaction terms are significant at 95% confidence interval. The McFadden's R^2 increases to 7.2% and the AUC increases to 72.1%. These results suggest that the two aspects of propagation channel are complementary to each other in improving the performance of the idiosyncratic shocks.

In column 4, I examine the general model that includes individual factors and interactions. That is, I extend the previous model in column 3 by adding the individual factors, AC_t and DIA_t. The coefficient estimates of AC_t and DIA_t are both insignificant at 95% confidence interval. Compared to the specification 3, the AUC is slightly improved to 73% and the R^2 is roughly unchanged. This finding advocates that the properties of propagation channel do not directly affect the transmission of an idiosyncratic shock. Instead, they serve as a channel through which the shock propagates.

Lastly, I added the benchmark factor, $ENR_{all,t}$, to the prediction models as shown in column 5 and 6. The statistics are very similar to the specifications 3 and 4 from which the $ENR_{all,t}$ is excluded. The coefficients of AV_t and the interaction terms are all statistically significant. The addition of $ENR_{all,t}$ also has no significant impact on the R^2 and AUC. Thus, the relationship between the idiosyncratic risk propagation model and the extreme negative returns is quite robust to the percentage of countries with extreme negative returns.

Table 3.3 Idiosyncratic risk propagation – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B

	Panel A: Dependent variable is the dummy variable $ENR_{i,t}$					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-9.059	-0.600	-4.620	-5.461	-4.462	-4.308
	[-13.129]	[-0.670]	[-43.440]	[-4.305]	[-37.633]	[-3.833]
% with extreme negative					7.694**	7.990**
returns, $ENR_{all,t}$					[20.789]	[19.523]
Average variance,	861.372**	-35.329	620.000**	678.022**	285.944**	260.024**
AV_t	[12.793]	[-0.357]	[11.110]	[5.247]	[4.330]	[2.184]
Average correlation,	35.715**			34.594**		-8.720
AC_t	[6.639]			[6.321]		[-1.500]
Diameter,		-0.730**		-0.682**		0.177
DIA_t		[-4.090]		[-3.536]		[0.995]
AV_t*AC_t	-5162.084**		-2238.741**	-5076.218**	-866.389**	-76.234
	[-10.275]		[-10.203]	[-10.096]	[-2.947]	[-0.129]
AV_t*DIA_t		42.558**	-27.129**	34.375*	-22.142**	-37.768**
		[2.189]	[-2.976]	[1.699]	[-2.183]	[-2.046]
R-squared	0.25	0.214	0.239	0.258	0.387	0.388
Area under ROC curve	0.831	0.841	0.851	0.84	0.941	0.937
Observations	8,256	8,256	8,256	8,256	8,256	8,256

	Panel B: Dependent variable is the dummy variable $ENR_{i,t+1}$					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-5.633	-4.485	-3.773	-4.835	-3.698	-4.283
	[-11.050]	[-7.993]	[-46.136]	[-6.771]	[-44.010]	[-5.884]
% with extreme negative					1.386**	1.350**
returns, $ENR_{all,t}$					3.588	[3.257]
Average variance,	571.730**	508.905**	660.499**	725.201**	575.772**	612.620**
AV_t	[10.755]	[9.906]	[14.506]	[11.599]	[11.016]	[8.527]
Average correlation,	16.236**			6.196		0.963
AC_t	[4.074]			[1.426]		[0.208]
Diameter,		0.188*		0.057		0.088
DIA_t		[1.795]		[0.499]		[0.756]
AV_t*AC_t	-3847.213**		-2025.117**	-2514.667**	-1786.502**	-1834.018**
	[-10.052]		[-9.049]	[-5.801]	[-7.395]	[-3.776]
AV_t*DIA_t		-81.258**	-57.745**	-57.438**	-52.033**	-57.283**
		[-9.135]	[-7.968]	[-5.198]	[-6.794]	[-5.032]
R-squared	0.059	0.049	0.072	0.073	0.076	0.076
Area under ROC curve	0.697	0.711	0.721	0.730	0.719	0.721
Observations	8,213	8,213	8,213	8,213	8,213	8,213

(Note) This table presents the results of logistic regressions of the dummy variable of extreme negative returns on contagion measures and interaction terms in Equation (3.5). "$ENR_{all,t}$" is the percentage of countries with extreme negative returns on month t. "AV_t" is the measure of idiosyncratic risk, calculated as the equal-weighted cross-sectional average of stock variances. "AC_t" is the equally-weighted cross-sectional average of the Pearson's correlations of the 43 markets in month t. "DIA_t" is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. "AV_t*AC_t" is the interaction term between AV_t and AC_t. "AV_t*DIA_t" is the interaction term between AV_t and DIA_t. "R^2" is McFadden's pseudo R^2. "AUC" is the area under ROC curve. Z-statistics are reported in brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval.

3.4.4 The effect of global shocks on the extreme value analysis

In this paper, the possibility of contagion depends on the number of countries that experience a distressed period. In Section 3.4.2 and 3.4.3, I find that the percentage of countries with extreme negative returns, ENR_{all}, is the most suitable to identify contagion, while the models of idiosyncratic risk propagation are useful for predicting contagion. However, it is also possible that the extremely negative returns in multiple markets are caused by global shocks. To test whether or not the main results are unaffected from the global shocks, I introduced three possible global shocks into the regression models. The first one is Bloomberg Commodity Index ($BCOM_t$) which reflects commodity futures price movement in month *t*. The second is TED_t spread, calculated as the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. The last one is $US10b_t$ or 10-year Treasury constant maturity US government bond in month *t*. The data of TED_t and $US10b_t$ is obtained from Federal Reserve Bank of St. Louis.

Table 3.4 presents the estimates from the logistic regressions after controlling for the global shocks. Excluding the controlled variables, the six specifications are the same as those in Section 3.4.3. In Panel A, the goal is to estimate the conditional probability that a country has an extreme negative return in month *t*. In column 1-4, the average variance and interaction terms all remain significant after controlling for the global shocks. Moreover, the estimates of R^2 and AUC are all marginally different from the models without the global shocks in Table 3.3. This finding indicates that the ability to identify contagion of the network-related terms is robust to the global shocks. In column 5 and 6, when the ENR_{all} is included, the R^2 and AUC also remain roughly unchanged from those in Table 3.4. The ENR_{all} is once again the most dominant factor for measure the probability of the extreme negative returns in the same period.

In Panel B, the main objective is to estimate the conditional probability of extreme negative returns in month $t+1$. AV_t is statistically significant in all specifications. That is, the idiosyncratic risk is indeed important for predicting contagion in a subsequent period. In column 1,2 and 3, the coefficients of AV_t*AC_t and

$AV_t * DIA_t$ are all statistically significant at 95% confident interval. Thus, after controlling for the global shocks, the average correlation and diameter can facilitate and amplify the effect of idiosyncratic shocks. The most contribution of the global shocks are the increases in McFadden's pseudo R^2 of all six models. They increase around 3% or more across all tested models. For instance, the R^2 of the two-interaction model in column 3 increases from 7.2% in Table 3.3 to 10.1% in Table 3.4. Among the global shocks, I find that this contribution comes from TED spread. The TED spread is the only global shock with strongly significant coefficients across all six models.

However, compared to the specifications without the global shocks, the AUC remain roughly unchanged. Since the AUC is an important figure for discrimination of extreme negative returns, the joint coincidence of extreme negative returns is then closely related to the idiosyncratic shocks and network structure. This relationship persists even after the global shocks are included in the tests.

Table 3.4 Controlling for global shocks – Regression results for the current extreme negative turns in Panel A and the future the extreme negative returns in Panel B

	Panel A: Dependent variable is the dummy variable $ENR_{i,t}$					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-8.925	-1.08	-4.126	-5.663	-4.929	-4.750
	[-11.030]	[-1.142]	[-19.390]	[-4.191]	[-17.365]	[-3.733]
% with extreme negative					8.180**	8.454**
returns, $ENR_{all,t}$					[20.381]	[19.533]
Average variance,	858.172**	-45.416	752.703**	751.054**	275.805**	283.099**
AV_t	[12.474]	[-0.415]	[11.103]	[5.540]	[3.466]	[2.348]
Average correlation,	35.043**			34.323**		-8.724
AC_t	[6.188]			[5.972]		[-1.396]
Diameter,		-0.793**		-0.625**		0.194
DIA_t		[-4.208]		[-3.194]		[1.065]
AV_t*AC_t	-5127.664**		-2783.248**	-5063.660**	-124.75	555.926
	[-10.103]		[-9.210]	[-10.009]	[-0.307]	[0.914]
AV_t*DIA_t		43.436**	-36.214**	20.829	-38.222**	-56.704**
		[2.075]	[-3.813]	[0.990]	[-3.423]	[-2.948]
Commodity index,	-2.153*	-4.701**	-2.610**	-2.868**	-4.488**	-4.292**
$BCOM_t$	[-1.761[	[-3.798]	[-2.050]	[-2.268]	[-3.057]	[-2.873]
TED spread,	-13.522	-13.488	-28.205**	-24.255*	-60.474**	-63.970**
TED_t	[-1.071]	[-1.058]	[-2.082]	[-1.763]	[-3.799]	[-3.968]
10-year bond rate,	-0.138	22.298**	-13.790**	0.461	18.016**	14.562*
$US10b_t$	[-0.022]	[4.748]	[-2.333]	[0.073]	[2.378]	[1.809]
R-squared	0.252	0.224	0.244	0.26	0.394	0.395
Area under ROC curve	0.829	0.847	0.85	0.839	0.931	0.929
Observations	8,256	8,256	8,256	8,256	8,256	8,256

	Panel B: Dependent variable is the dummy variable $ENR_{i,t+1}$					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	-3.429	-4.214	-3.806	-2.521	-3.836	-2.503
	[-5.190]	[-6.749]	[-18.842]	[-2.933]	[-18.694]	[-2.918]
% with extreme negative returns, $ENR_{all,t}$					0.762*	0.792*
					[1.735]	[1.729]
Average variance, AV_t	287.348**	175.916**	413.700**	334.852**	370.867**	288.488**
	[4.649]	[2.807]	[6.985]	[4.297]	[5.722]	[3.483]
Average correlation, AC_t	-2.193			-3.019		-4.747
	[-0.477]			[-0.642]		[-0.990]
Diameter, DIA_t		-0.038		-0.16		-0.125
		[-0.342]		[-1.348]		[-1.036]
AV_t*AC_t	-2050.792**		-2010.734**	-1886.074**	-1828.448**	-1560.355**
	[-4.728]		[-6.626]	[-4.060]	[-5.607]	[-3.087]
AV_t*DIA_t		-30.689**	-22.005**	-11.487	-20.427**	-12.777
		[-2.925]	[-2.598]	[-0.930]	[-2.362]	[-1.017]
Commodity index, $BCOM_t$	1.541	-0.053	1.261	1.193	1.382	1.358
	[1.320]	[-0.045]	[1.071]	[1.001]	[1.172]	[1.134]
TED spread, TED_t	127.326**	112.379**	113.013**	118.870**	109.574**	115.682**
	[11.420]	[10.258]	[9.950]	[9.820]	[9.425]	[9.399]
10-year bond rate, $US10b_t$	-7.853	17.988**	-5.121	-7.511	-2.946	-5.863
	[-1.314]	[3.945]	[-0.911]	[-1.250]	[-0.507]	[-0.955]
R-squared	0.099	0.088	0.101	0.102	0.102	0.103
Area under ROC curve	0.695	0.697	0.719	0.709	0.724	0.711
Observations	8,213	8,213	8,213	8,213	8,213	8,213

(Note) This table presents the results of logistic regressions of the dummy variable of extreme negative returns on contagion measures, interaction terms, and global shocks. "$ENR_{all,t}$" is the percentage of countries with extreme negative returns on month t. "AV_t" is the measure of idiosyncratic risk, calculated as the equal-weighted cross-sectional average of stock variances. "AC_t" is the equally-weighted cross-sectional average of the Pearson's correlations of the 43 markets in month t. "DIA_t" is the diameter of the correlation-based PMFG network and calculated as the longest distance of all geodesic paths in a given network. "AV_t*AC_t" is the interaction term between AV_t and AC_t. "AV_t*DIA_t" is the interaction term between AV_t and DIA_t. "$BCOM_t$" is Bloomberg Commodity Index and calculated from commodity futures price movement. "TED_t" is the spread between 3-month LIBOR based on US dollars and 3-month Treasury Bill. "$US10b_t$" is 10-year Treasury constant maturity US government bond. "R^2" is McFadden's pseudo R^2. "AUC" is the area under ROC curve. Z-statistics are reported in brackets. ** is significant at 95% confidence interval. * is significant at 90% confidence interval

3.4.5 Alternative network measures

In this paper, I compress the complex interrelationships of the international stock markets into a network model. Even though the network structure greatly simplifies the complex relationships, it is still impossible to use a single measure to describe the whole network. In the network, one country can directly affect the adjacent countries as well as indirectly affect the countries in the further part of the network. Its shock can also reach one country or many countries through the propagation channel of the network. Therefore, the network literature produces many network measures to quantify the properties of network structure. Among other measures, I use the average correlation to capture the strength of the propagation channel and the diameter to capture the distance of propagation channel. This section compares the ability to classify the contagion between the regression model in Equation (3.6) and the other specifications in which alternative network measures are presented.

$$Prob(ENR_{i,t+1} = 1) = F\begin{pmatrix} const + b_1AV_t + b_2AC_t + b_3DIA_t \\ +n_1AV_t * AC_t + n_2AV_t * DIA_t \end{pmatrix} \qquad (3.6)$$

Other than the diameter, the distance of propagation channel can be measured by the average shortest path and the average of eccentricity. In a network, there are several paths that can take one country to reach another country. These paths have different distances, and the shortest one between the two countries is called the shortest path. The average of all possible shortest paths is then a measure of the distance of the propagation channel. A shock can propagate faster in a network with a small average shortest path than the large one. Another measure of the distance is the eccentricity. An eccentricity of a country is the shortest distance of that country to its furthest country in the network. The average of the countries' eccentricity is then comparable to the radius of the network. Its interpretation is very similar to the diameter. The low average eccentricity would indicate the star-like structure and let a shock to propagate faster. On the other hand, the network with the high eccentricity is lean towards the chain-like structure.

Another important property of the network is the relative importance of a node. Measures of this property can give different weights to each node in a network. A node with the high relative importance would have a significant influence on the whole market. In particular, when an extreme negative shock hits the important country, there is a high possibility that the other countries will be affected by the shock. In this paper, I quantify this network characteristic with three network measures. The first measure is the eigenvector centrality. This measure calculates the relative importance from the eigenvector and eigenvalue of the adjacent matrix that records the connections in a network. The country attached to the high-scoring nodes will have a higher level of eigenvector centrality than a connection to the low-scoring nodes. The high average eigenvector centrality would mean that there are many high-influence countries in the network. Another measure of the relative importance is the degree centrality. The degree of a node in a network is the number of links attached to it. The high number of links indicates the high level of degree centrality. In this paper, I calculate the degree centrality from the PMFG network which has a fixed number of links in every month. As a result, the average degree centrality will be constant over time. In the extreme value analysis, this measure will be customized to each country in each month, while the other network measures are the same for every country in each month.

The last measure of the relative importance is the KNN, which stands for K-nearest neighbors. The nearest neighbor, in this paper, is defined as the adjacent countries and thus make K equal to 1. The KNN centrality of a node is then calculated as the average degree of its adjacent neighbors. The high average KNN centrality indicates that there are many countries with high relative influence. Nevertheless, these centrality measures are not as direct as the diameter in determining the network topology or shape. The nodes with high relative importance can be in the center or any part of the network, and thus sometimes result in a star-like structure and sometimes appear to be a chain-like structure.

The goal in this section is to assess the performance of average correlation and diameter as opposed to the alternative network measures in predicting contagion. I achieve this goal by using the brute-force approach in the extreme value analysis. Specifically, I regress the extreme negative returns, $ENR_{i,t+1}$, on all possible combination of the independent variables. The tested factors include the percentage of countries with negative return ($ENR_{all,t}$), average market variance (AV_t), average correlation (AC_t), diameter (DIA_t), average shortest path (ASP_t), average eccentricity ($ECCEN_t$), average eigenvector centrality ($EIGEN_t$), average KNN centrality (KNN_t), and degree centrality ($DEG_{i,t}$). Additionally, I also include the interaction between the average variance and network measures. The total number of the independent variables is sixteen and results in 65,536 combinations. Please note that I do not report the results of the extreme value analysis at month t because the $ENR_{all,t}$ dominates all other variables.

Table 3.5 reports the AUC performance of the model in Equation (3.6) in comparison to all other possible models. I conduct the test for both in-sample and out-of-sample analysis. For the in-sample regressions, I regress $ENR_{i,t+1}$ on the independent variables by using the full sample observations. The AUC of the model is 73%, and its rank is 3,056 out of 65,536 models. The model's ranking is very high in the top 5th percentile. Figure 3.5a depicts the AUC performance of all possible combinations whereby the vertical axis is the AUC scores, and the horizontal axis is the number of independents variables ranging from one to sixteen. Our interested model is the red dot coordinated at five independent variables in the x-axis and 0.73 AUC in the y-axis. The tested model ranks very high within its group of five variables. Also, the rank is still high even if the model is compared to the specifications with more explanatory variables.

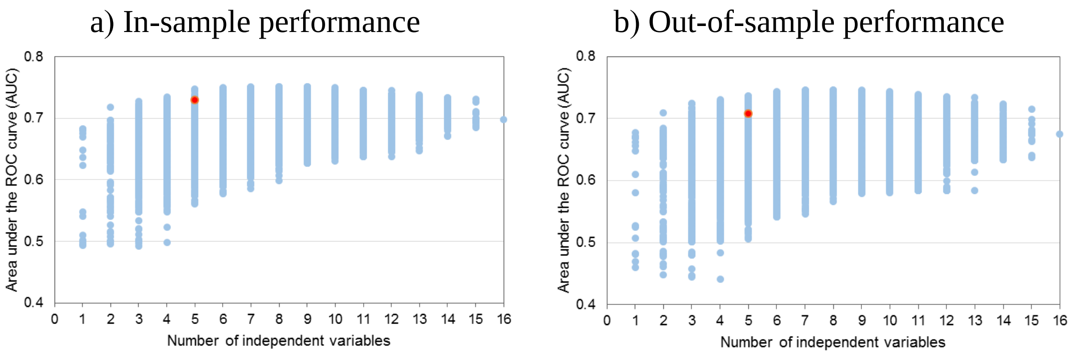
For the out-of-sample regressions, the sample is randomly divided into the initial and evaluation periods. The initial period has 6,555 observations or 80% of the total sample, while the evaluation period has 1,658 observations or 20% of the full sample. To conduct the out-of-sample test, I first estimate the coefficients and odd-ratios from the initial sample by using the extreme value analysis approach. Then, the AUC estimation will use these parameters to classify the binary outcomes of the evaluation periods. The out-of-sample AUC of Equation (3.6) is 70.7% and ranks 6,391 of all possible combinations. The model's ranking is relatively high in the top 10th percentile. Figure 3.5b depicts the graphical ranking of all specifications. The tested model in Equation (3.6) is the red dot coordinated at five independent variables in the x-axis and 0.707 AUC in the y-axis. Similar to the in-sample analysis, the interested model is located in the high region of the out-of-sample AUC. The results from both in-sample and out-of-sample tests suggest that the average correlation and diameter are suitable to serve as the measures of the propagation channel in comparison to the alternative network measures.

Table 3.5 Ranking of the model with average correlation and diameter

	In-sample performance	Out-of-sample performance
Pseudo R-squared	0.073	0.076
AUC	0.730	0.707
AUC ranking	3,056	6,391
AUC percentile	Top 5^{th}	Top 10^{th}
Total number of combinations	65,536	65,536
Observations in each regression	8,213	1,658

(Note) This table presents the ranking of the model in Equation (3.6) among all possible combinations of nine explanatory factors and seven interaction terms. Nine individual factors are the percentage of countries with negative return ($ENR_{all,t}$), average market variance (AV_t), average correlation (AC_t), diameter (DIA_t), average shortest path (ASP_t), average eccentricity ($ECCEN_t$), average eigenvector centrality ($EIGEN_t$), average KNN centrality (KNN_t), and degree centrality ($DEG_{i,t}$). Seven interaction terms are $AV_t xAC_t$, $AV_t xDIA_t$, $AV_t xASP_t$, $AV_t xECCEN_t$, $AV_t xEIGEN_t$, $AV_t xKNN_t$, and $AV_t xDEG_t$. There are 65,536 combinations from the sixteen variables

Figure 3.5 The Graphical AUC Performance



(Note) This figure depicts the AUC performance of all tested models to predict the binary outcome of the extreme negative returns in month t+1. The vertical axis is the area under the curve (AUC), and the horizontal axis is the number of independent variables. The red dot is the performance of the model in Equation (3.6) which has five variables, including average stock variance, average correlation, diameter, the interaction of average variance and average correlation, and the interaction of average variance and diameter

3.5 Concluding remarks

Contagion can cause an idiosyncratic shock of one country to propagate to numerous countries. The effect of the idiosyncratic shock is not always diversified away as the conventional asset pricing models claim. Instead, the country-specific shock can sometimes propagate and amplify to cause a systemic event such as the US financial crisis in 2007-2008 and the Greek government-debt crisis in 2011. This paper focuses on the network structure that plays a crucial role in international financial contagion. In particular, the network measure is used to determine the probability of the extreme negative returns, which is defined as the bottom 5% of a country's return distribution. The network variables are tested in both contemporary periods at month *t* and future period at month *t+1*.

I find that the percentage of countries with extreme negative returns is the most dominant factor to measure the probability of the extreme negative returns in month t. It also has some power to predict the probability of the extreme negative returns but is not as good as the factors in the idiosyncratic risk propagation model, including the average variance, the network measures, and their interactions. The average variance represents the country-specific risk of the global market. The average correlation and diameter are the network measures that reflect some aspects of cross-country linkages. The interactions represent the effects of the propagation of idiosyncratic shocks through the network linkages. I find that the interactions increase the ability to predict the probability of the extreme negative returns.

This article opens up an interesting research topic. Specifically, Bisias et al. (2012) pointed out that network models help us to understand more about how systemic events unfold. This paper supports this argument and provides empirical evidence that network measures are indeed relevant to the probability of the extreme negative returns. In particular, the network structure serves as a channel through which an idiosyncratic shock propagates. This process could then have more or less effect on the subsequent extreme negative situation of countries in the network. However, this paper touches just some aspects of the network theory. It does not explain as to why and how networks change from one structure to another. The study of this network transition naturally gives us some new information and more fundamental knowledge about how financial contagion and systemic events occur.

3.6 Appendix

3.6.1 Area under the ROC curve (AUC)

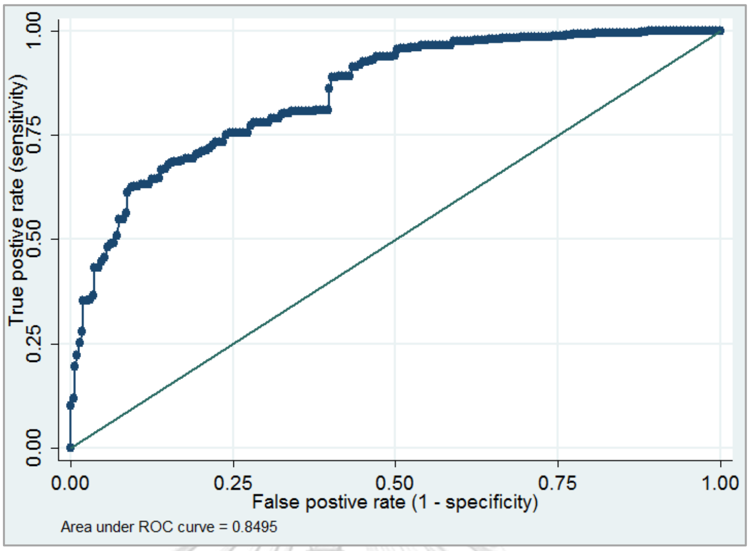
The AUC measures the ability of the model to classify binary outcomes correctly. The basic concept of AUC is to calculate the area under the ROC curve (Receiver Operating Characteristic) as shown in Figure 3.6. The construction of the ROC curve relies on the actual outcomes and predicted outcomes. The actual binary outcome, $ENR_{i,t}$, is 1 if a market's return is extremely negative and 0 otherwise. The predicted outcomes are calculated from the probability provided by the logistic regression in Equation (3.4). If the fitted probability exceeds a certain threshold, the respective outcome will be 1. With the actual and predicted outcomes, we can determine the statistics of true positive, true negative, false positive and false negative, as shown in Table 3.6. The outcome is true positive if both actual and predicted outcomes are 1. The outcome is true negative if the actual outcome is 0 and the predicted outcomes are 1. The outcome is false positive if the actual outcome is 1 and the predicted outcomes are 0. The outcome is false negative if both actual and predicted outcomes are 0.

Table 3.6 Possible outcomes of the binary classification.

		Actual outcome	
		True (1)	False (0)
Predicted outcome	Positive (1)	True positive	False positive
	Negative (0)	True negative	False negative

The ROC space is defined by the true positive rate (TPR) on the Y-axis and the false positive rate (FPR) on the X-axis. The TPR is equivalent to sensitivity and calculated as the number of true positive outcomes divided by the number of true outcomes. The FPR is equal to 1-specificity and calculated as the number of false positive outcomes divided by the number of true outcomes. As the discrimination threshold is varied from 0 to 1, we will get a set of coordination based on the FPR and TPR. When plotted on the ROC space, the ROC curve is formed and usually concave downward. If the curve is closed to the upper vertical axis, the AUC will be closed to 1. The tested model with AUC of 1 represents the perfect test. In other words, the predicted outcomes of the model are perfectly equal to the actual outcomes. If the graph is the diagonal line, the AUC is 0.5, and the model is indifferent from a constant model without any predictors.

Figure 3.6 ROC (Receiver Operating Characteristic) space and curve.



(Note) This figure shows an illustration of the ROC space. The vertical axis is the true positive rate or sensitivity. The horizontal axis is the false positive rate or 1-specificity

Chapter 4 Interconnectedness and equity returns: The case of Thailand

Abstract Interconnectedness has been regarded as a key driver of the global financial crisis in 2007-2008. However, empirical research that directly tests the relationship between interconnectedness and equity returns is essentially lacking. This paper aims to contribute to the field by providing empirical evidence in the Stock Exchange of Thailand. By incorporating techniques from the network theory, I quantify two characteristics of interconnectedness: systematical importance and fragility. The first measures the ability to spread shock while the latter measures the vulnerability to incoming shocks. I find evidence of the positive and significant relationship between systematical importance and stock returns and the negative and significant relationship for fragility. The measure of systematic importance, in particular, can capture cross-sectional variation in stock returns, while the well-known risk factors such as market risk, size and book-to-market cannot.

Keywords Interconnectedness, Systematical importance, Fragility, Asset pricing, CheiRank, PageRank

JEL Classification G12, D85

4.1 Introduction

Due to the wide-spread and exceptional damage to many financial markets, the 2007-2008 financial crisis has been subject to considerable research effort in the past decade. The previous studies report that interconnectedness is a key driver of the crisis instead of the present common risk factors such as firm's size. Specifically, the interconnectedness between firms forms a network structure that facilitates the propagation of firm-specific shocks to cause cascades of failures, financial contagion, and ultimately a systemic event. In this kind of network, the highly interconnected firms tend to have a significant influence on the market because they can swiftly transmit their own shocks to multiple firms or be vulnerable to propagated shocks. Therefore, the interconnectedness can induce the system-wide risk and be considered as a common risk factor for the equity market.

Although a number of studies have focused on modeling the interconnectedness and studying its relationship with the systemic risk, little attention has been paid to the effects of interconnection on equity returns. Empirical research on this subject is essentially lacking, and most of them focus on the developed markets and international indices. The primary purpose of this paper is to fill this gap by providing empirical evidence for the relationship between interconnectedness and equity returns in the stock exchange of Thailand.

The previous study that examines the relationship between an interconnectedness measure and equity returns is "Eccentricity in Asset Management" of Kaya (2014). This paper focuses on returns on international equities such as stock indices, bonds, commodities, sectors, and industries. The author simulates a network of the global equities whose connection is calculated from the joint distribution of two equities' returns. The centrality of assets is then estimated from this network. The key finding of this paper is that the assets located towards the center area of the network tend to have higher returns than the other assets in the network. Buraschi and Porchia (2012) provide another empirical evidence that is consistent with Kaya's conclusion. Unlike Kaya (2014), Buraschi and Porchia focus on the US stock market instead of the global equities. Specifically, they study a network of US-listed companies whose values depend on other firms' dividend states. A firm, which is likely to affect the others, would have a high degree of active connectivity (DC) and thus be more central to the network. The paper reports that the CAPM will hold if firms have homogeneous connections and DC, which can be observed in a symmetric network such as a disconnected network and a complete undirected network. On the other hand, the CAPM is not valid in an asymmetric network that has heterogeneous connections and DC. That is, a firm with a high degree of connectivity tends to earn higher expected returns, which cannot be explained by the market risk of the CAPM. Furthermore, the author reports that the average slope of the DC is positive and significant in the test of

Fama-Macbeth (1992). In other words, the central stocks tend to gain higher expected return than the peripheral stocks. Similarly, Ahern (2013) and Chen (2014) also report the same result in the cross-sectional regression. Ahern (2013) finds that central industries in US tend to have greater exposure to sectoral shocks. Therefore they are more risky economically and demanded higher returns as a compensation. Chen (2014) finds that central stocks in a volatility network have higher returns than the rest. The primary reason is because idiosyncratic shocks cause the affected stocks to move together. The co-movements result in a volatility-based network, of which central stocks have the highest degree of co-movement. Thus, investors demand higher returns for stocks that are more central.

Although these works seem to report in the significantly positive relationship between network position and returns, they mainly focus on the US market. Instead of the developed market, this paper focuses on analyzing the relationship in the context of the developing market that is Stock Exchange of Thailand (SET). This paper reexamines the relationship between interconnectedness measures and asset returns. I aim to enrich the empirical evidence regarding this relationship with two primary research questions: 1) Does network structure of stocks matter in explaining cross-sectional expected returns in SET? 2) Can the network structure proxy for sensitivity to common risk factors in expected returns?

The network structure can affect stocks in two ways. First, a stock uses the network structure as a channel to transmit its shocks. Second, a stock gets affected by the propagated shocks. To simulate a network with these properties, I use a different approach from the previous literature such as Kaya (2014) and Buraschi and Porchia (2012) to simulate a network of selected firms listed in SET. Following Kenett et al. (2010), I create a directional network of stocks whose relation is calculated from the partial correlation of returns. The main advantage of the partial correlation is that the targeted relationship is free from the influence of the third parties and thus mitigate the problem of spurious correlations. Moreover, because the partial correlation between A and B (A→B) is not equal to the reciprocal (B→A), the network is naturally directional. As a result, I can directly measure two characteristics of the interconnectedness, including systematic importance and fragility. A firm is systematically important if it can affect the other firms in the network. On the contrary, a firm is fragile if it gets affected by propagated shocks. I use the concept of CheiRank to capture the systematic importance and PageRank to capture the fragility. To the best of the author’s knowledge, this paper is the first study to investigate the relation of both characteristics of the interconnectedness to stock returns. This is also the first article that explores this matter in Stock Exchange of Thailand.

Similar to the previous literature, this paper reports the significant relationship between interconnectedness measures and equity returns. In particular, I do agree with Chen (2014) that the network structure serves as a local channel through which idiosyncratic shocks spread, even though our scope is different. Chen (2014) focuses on the network risk induced by idiosyncratic shocks and thus constructs the network from the covariance of residuals from the three-factor model of Fama and French (1993). On the other hand, I am interested in any factors that can cause the co-movements between stocks, rather than the idiosyncratic shocks. Specifically, I use the approach of Kenett et al. (2010) to construct a network of correlations of returns. The resulted network clusters the stocks with the same common economic factors together and thus provide the meaningful economic taxonomy of the market. The change in the network structure would also affect the transmission of the common economic factors and in turn the common risk factor in expected returns.

Moreover, I find that that the network structure helps to explain cross-sectional returns on stocks in the Stock Exchange of Thailand. Following Fama and French (1992), I conduct the Fama-Macbeth cross-sectional tests by regressing the individual excess stock returns on market beta, size, book-to-market, CheiRank and PageRank. The measure of systematic importance, CheiRank, has a significant and positive relationship with equity returns. On the other hand, the measure of fragility, PageRank, has a negative relationship but not significant. In addition, when stocks are sorted into three portfolios by their CheiRank or PageRank, the sign of the relationship is confirmed. The highly systematically-important portfolio earn 0.18% on a monthly basis, which is about 3.7 times lower than the low group. The highly fragile stocks earn returns twice as much as the low group.

The rest of the paper is organized into four sections. Section 4.2 explains the methodologies to simulate networks of firms in SET50 and to quantify the characteristics of interconnectedness. Section 4.3 presents the description of data. Section 4.4 reports the empirical results for the relationship between interconnectedness measures and equity returns. Section 4.5 concludes the paper.

4.2 Methodology

This paper focuses on two important characteristics of the interconnectedness to explain the cross-section of stock returns. The first aspect is a firm's ability to affect the others, which is usually called "systematic importance" in the existing literature. The second one, on the other hand, is a firm's vulnerability or fragility to propagated shocks from the others. To measure both characteristics of the interconnectedness, I proceed in two steps. I first construct a network architecture of the stock market which allows us to see the interconnection pattern of the individual stocks. Then, I introduce the concepts of PageRank and CheiRank to quantify the interconnectedness properties.

4.2.1 Network structure of the stock market

To simulate the meaningful network from such a complex system as the stock market, I follow the innovative approach proposed by Kenett et al. (2010). Prior to their work, many papers use the Pearson correlation coefficient to create a network and investigate stock relationships. However, the correlation between two stocks may be spurious due to the effect of the third party. For example, A may be highly correlated with B possibly due to C. To eliminate this problem, Kenett et al. (2010) introduce a concept of partial correlation to simulate a network, instead of the simple Pearson correlation. The partial correlation coefficient is built on the observed Pearson correlation on which the effects of the different stocks get eliminated. For example, if I want to quantify the effect of Z on X (Z→X), the first step to create a partial correlation is to estimate the correlation of returns between two stocks, $\rho(X, Y1)$. The second step is to eliminate the effect of the third party Z as follows:

$$\rho(X, Y1{:}\, Z) = \frac{\rho(X, Y1) - \rho(X, Z)\rho(Y1, Z)}{\sqrt{[1 - \rho^2(X, Z)][1 - \rho^2(Y1, Z)]}} \quad (4.1)$$

$\rho(X, Y1{:}\, Z)$ is the partial correlation between X and Y1 of which the effect of Z is eliminated. $\rho(X, Z)$ is Pearson correlation coefficient between X and Z. $\rho(Y1, Z)$ is Pearson correlation coefficient between Y1 and Z. The upper relationship in Figure 4.1a illustrates the removal of Z in the second step.

The third step is to keep only the effect of Z on the relationship of X and Y and therefore create another partial correlation d(X,Y1:Z) as follows:

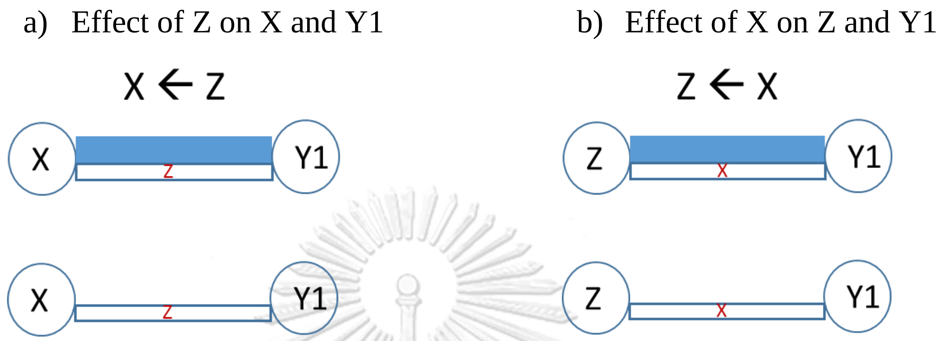
$$d(X, Y1{:}\, Z) = \rho(X, Y1) - \rho(X, Y1{:}\, Z) \quad (4.2)$$

The lower picture of F4.1a illustrates the remaining effect of Z on the relationship of X and Y1 in the third step. Then, we repeat the step two and three by replacing Y1 with Y2, Y3, Y4 and so on. The fourth step is to calculate the equal-weighted average influence of Z on X as follows:

$$d(X{:}\, Z) = \frac{1}{N}\sum_{i=1}^{N} d(X, Y_i{:}\, Z)\,; Y \neq X, Z \quad (4.3)$$

It is important to notice that d(X:Z) is not equal to d(Z:X). As shown in Figure 4.1a and 4.1b, taking the effect of Z out of the X-related relationship is not equal to the opposite. Retaining only the larger one allows us to assign the direction between X and Z. For instance, if d(X,Z) > d(Z,X), the direction in the network will be Z→X, which summarizes the influence of Z on the correlations between X and all the other elements in the system. This direction is the prime advantage of this network construction approach among the other network construction methods. It is also crucial to note that the partial correlation is not intended for a causality measure but a mean to understand the correlation-based architecture of a stock market.

Figure 4.1 Illustration of partial correlation.



(Note) This figure illustrates the process of partial correlation. Figure a) shows the effect of Z on the relationship of X and Y1. Figure b) shows the effect of X on the relationship of Z and Y1. The upper picture shows the result of Equation (4.1). The lower picture depicts the result of Equation (4.2)

After iterating the previous steps for all non-diagonal elements, the partial correlation matrix will have N*(N-1)*(N-2)/2 partial correlation interactions, whereby N is the number of stocks in the system. If N is 50, the total number of partial correlations will be 58,800. When constructing a network with these correlations, the network would be extremely complicated for the estimation of the individual interconnectedness. Therefore, the last step is to filter out the partial correlation matrix and retain the essential information about the interrelationships. In the correlation-based network, three algorithms are usually used to reduce the links: Threshold method, Minimum Spanning Tree (MST), and Planar Maximally Filtered Graph (PMFG). The Threshold method reduces links by using thresholds. For example, the relationship that has correlation below a certain threshold (i.e., 0.7) will be dropped out of the network. The advantage of this method is that it can control the desired amount of information in the network by changing the threshold levels. The other two methods reduce the number of links by using topological constraints as described in graph theory. As discussed in the paper "Hierarchical Structure in Financial Markets" (Mantegna, 1999), the MST approach provides a hierarchical tree of the stocks which reflects the memberships in the real sectors and sub-sectors classified by the Forbes. Then, in an attempt to incorporate more relevant information into the network, Tumminello et al. (2005) introduce a new algorithm called PMFG. This PMFG network has the same hierarchical tree as the MST and some additional structures of loops and cliques which satisfy the planarity condition. Additionally, the PMFG contains enough information to make some analysis of the causal relationships of the stocks. Kenett et al. (2010)

further develop the PMFG into the PCPG (Partial Correlation Planar Maximally Filtered Graph). The PCPG adds directions into relationships of stocks so that we will know the effect of one stock on another. Unlike the MST which gives a symmetric correlation matrix, the PCPG use partial correlation to create an asymmetry correlation matrix which in turn provides a directional network of the stock markets. In this paper, I use the PCPG approach to create a network because it gives directional relationships which are crucial to quantify the characteristics of the interconnectedness. The PCPG network also simplifies the description of the system greatly by reducing the number of interactions to 3(N-2).

4.2.2 Measures of interconnectedness

This paper considers two measures of interconnectedness, including systematic importance and fragility. The systematic importance reflects the influence of a particular firm on the network. The fragility indicates the vulnerability of the firm to propagated shocks. The simplest network measure for both characteristics of the stock interconnectedness is degree centrality, which counts the adjacent links to a firm in the network. The high number of outgoing links indicates the high level of degree centrality associated with systematic importance, while the high number of incoming links indicates the high level of degree centrality associated with fragility.

Although the degree centrality is a simple and easy-to-understand concept, it only takes into account the adjacent neighbors and omits the information beyond that point. Sergey Brin and Larry Page (1998) introduce another concept, called Google PageRank, which incorporates the entire network into the calculation of a firm's centrality measure and assigns the relative importance within the set. The Google PageRank matrix is calculated from the eigenvector of the incoming link matrix with the maximum real eigenvalue = 1. The PageRank is used initially to assign the likelihood of each webpage being visited in corresponding to the searching words. The webpage with highest PageRank probability will rank first. The calculation of CheiRank is similar to PageRank. One difference, however, is that CheiRank uses the eigenvector of the outgoing link matrix instead of the incoming one. Therefore, PageRank reflects fragility while CheiRank indicates the systematic importance of a member of the system.

Due to the success of the PageRank in the Google website business, many papers apply the PageRank concept to many fields of research. For example, Ermann and Shepelyansky (2011) create the world trade network of which trade flows are classified by the PageRank and CheiRank algorithm. Dungey et al. (2012) use a methodology based on the Google PageRank algorithm to measure the systemic risk and rank systemically important financial institutions (SIFIs) for listed companies in the S&P500. They report that the systemic risk, measured by the PageRank, are relevant. It increases before the 2008 crisis, peaks at the collapse of Lehman Brothers, and decline afterward. Then, the systemic risk picks up again in response to the European sovereign debt crisis. Furthermore, they also find out that the financial sector is systemically important in the market, supporting new regulations (Basel III) of the Basel Committee on Banking Supervision.

In this paper, I apply the PageRank as the measure of a firm's fragility and the CheiRank as the measure of a firm's systematic importance. Both measures together explain some aspects of the firm's interconnectedness.

4.3 Description of Data

In each month, I calculate the related variables from the daily returns of stocks listed on the Stock Exchange of Thailand (SET50). A stock's daily return is calculated from the difference in log prices between two consecutive days as follows:

$$r_{i,t} = \ln(p_{i,t}) - \ln(p_{i,t-1}) \tag{4.4}$$

The sample period is January 2000 – December 2014 or 180 months in total. Our short sample period is inevitable due to the severe missing-data problem of stock prices and risk-free rate before 2000. The data is primarily retrieved from Bloomberg Terminal. The risk-free rate primarily comes from the daily rate of the 1-month T-Bill, retrieved from the Bank of Thailand website. The monthly risk-free rate used throughout the dissertation is calculated by summing the daily rate in a given month. The missing values of the risk-free rate in the 2000 and the first two months of 2001 are proxied by short-term government bond yield minus one percent. I obtain the proxy from the available data in 02/2001 – 12/2001. Specifically, the one percent difference for the proxy is the average difference of the available 1-month T-bill and the short-term government bond in 2001 that expires in August 2003 (LB038A). Similarly, I apply the estimation of the risk-free rate for the year 2000 by using the government bond that will expire in one and a half year from January 2000 (LB026A). Additionally, I use the monthly returns based on the historical daily prices to conduct the analysis.

I decide to use only the stocks in the SET50 because I want to mitigate the liquidity effect as much as possible while I still have enough stocks to conduct the reliable analysis. Interested readers can see more detail on the liquidity effect from Jegadeesh and Subrahmanyam (1993). The SET50 consists of 50 largest and liquidating stocks which are selected by the Index Advisory Committee. The index also updates twice a year on the first trading day of January and July to adjust for any changes in the market. Thus, the list of 50 stocks will change every six months to reflect the actual SET50 index accurately.

4.4 Empirical Results

4.4.1 CheiRank and PageRank

This section examines the characteristics of CheiRank and PageRank for the stocks in SET50 (See the detail explanation in Appendix 4.6). The measures are estimated from the rolling PCPG network on a monthly basis. In each month t, a PCPG network is constructed from the daily returns from month t-6 to t-1 by using the approach in Section 4.3. Then, CheiRank and PageRank are assigned to individual stocks in the network. Lastly, the stocks are classified into three groups (Low, Medium, High) in accordance with CheiRank or PageRank. The group of low CheiRank(PageRank) consists of stocks with its ranking below the 30th percentile. The group of medium CheiRank(PageRank) consists of stocks with its ranking between the 30th and 70th percentile. The group of high CheiRank(PageRank) consists of stocks with its ranking above the 70th percentile.

Table 4.1 shows the equal-weighted average of CheiRank(PageRank) for each group. The average CheiRank(PageRank) for the high ranking portfolio is 0.0472(0.0295), whereas the average CheiRank(PageRank) for the high ranking portfolio is much lower at 0.076(0.0134). The statistics indicate that the relative influence of the portfolios is differentiated clearly in the sample. On the other hand, the standard deviation within each group is very small. The highest standard deviation belongs to the high ranking portfolio at 0.16% for CheiRank and 0.13% for PageRank. The standard deviations of the other portfolio are about half of the highest number.

Figure 4.2 depicts the dynamic of CheiRank and PageRank portfolios. The graph has the similar implication to the statistics in Table 4.1. The difference of the relative importance between the groups is noticeable while the stock behavior within each group is relatively close. The noticeable difference between CheiRank and PageRank is the distribution of the ranking values. The ability to affect the other stocks tend to be mostly concentrated in the high CheiRank portfolios while the ability in the medium and low groups is somewhat indifferent. On the other hand, the fragility across the three PageRank portfolio is differentiated quite well.

To further investigate the PCPG network and the interconnectedness measures, I simulate a PCPG network in September 2008 when the financial crisis unfolded after the Lehman's bankruptcy. Figure 4.3 shows this particular network of which the node size indicates CheiRank level, the outgoing arrow represents its influence over the others, and the incoming arrow indicates the source of the propagated shock. Moreover, the systemically-important firms tend to be more central whereas the fragile firms tend to be distant from the center. In this sample network, KBANK and BBL apparently have the highest level of CheiRank and are central to the network. Due to these properties, they play a major role in spreading shocks to the other firms in the market. This description is closely matched the actual activities. The shocks from the global crisis propagate from the US through the banking channel to affect the other countries, including Thailand. As a consequence, the stocks in the banking sector are the first to receive the damage and in turn spread the damage to the other firms.

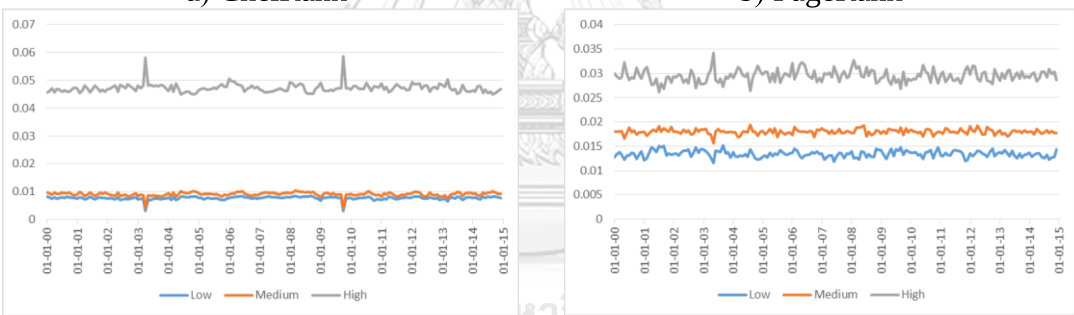
Table 4.1 Average CheiRank and PageRank

	Mean			Standard deviation		
	Low	Medium	High	Low	Medium	High
CheiRank	0.0076	0.0091	0.0472	0.0006	0.0008	0.0016
PageRank	0.0134	0.0180	0.0295	0.0007	0.0005	0.0013
Number of stocks	15	20	15	15	20	15

(Note) Table 4.1 reports some statistics of CheiRank and PageRank. Stocks are classified into three groups (Low, Medium, and High) in accordance with their CheiRank and PageRank. Mean is the equal-weighted average of CheiRank and PageRank within each group. The number of stocks in each group is also reported in the table.

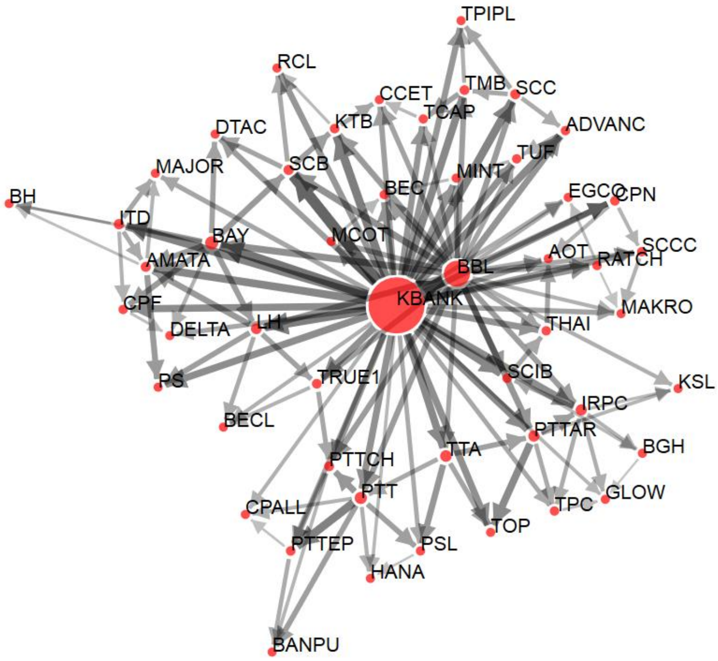
Figure 4.2 Dynamic of CheiRank and PageRank portfolios.

a) CheiRank b) PageRank



(Note) The figure presents the dynamic of the CheiRank portfolios in (a) and PageRank portfolios in (b) over time from January 1990 to December 2014. The vertical axis shows the equal-weighted average of CheiRank or PageRank. CheiRank is calculated from the outgoing link matrix and represents the relative influence of a firm by its ability to transmit its shocks to the others. PageRank is calculated from the incoming link matrix and represents the relative influence of a firm by its vulnerability to propagated shocks.

Figure 4.3 The PCPG network during the global crisis period in September 2008.



(Note) This figure shows an example of a PCPG network constructed from the daily returns of stocks listed in Stock Exchange of Thailand.

4.4.2 Cross-sectional regressions

The primary objective of this section is to test whether or not the network measures can explain the return differences among the stocks in SET50. Following the methodology of Fama and MacBeth (1973), I conduct the cross-sectional regressions on a monthly basis. As shown in Equation (4.5), the dependent variable is the return of an individual stock over the risk-free rate. The explanatory variables are market beta (MKTB), size (market capitalization or ME), book-to-market (BM), CheiRank (CR), and PageRank (PR). All independent variables except for the market beta are transformed by natural logarithm.

$$R_{i,t} - R_{f,t} = const_t + b_{beta,t}\left(MKTB_{i,t}\right) + b_{size,t}\, ln(ME)_{i,t} + b_{bm,t}\, ln(BM)_{i,t} + b_{cr,t} ln(CR * 1000)_{i,t} + b_{pr,t} ln(PR * 1000)_{i,t} + e_{i,t} \quad (4.5)$$

The first independent variable, $MKTB_{i,t}$, is the market risk or market beta of stock i, in month t. It is calculated from the one-factor regression of the excess daily returns on the individual stock from month t-6 to t-1. Thus, the $MKTB_{i,t}$ is considered the pre-ranking beta for individual stocks. This market beta is different from the one used in Fama and French (1992) who use portfolios to estimate the market beta. The reason they do that is that the portfolio approach reduces idiosyncratic volatility and allow more precise estimates of the market beta. Although this statement should not be ignored, I use individual stocks to estimate the market risk due to the data limitation. If I were to follow the portfolio approach, I would have only six portfolios, each of which has 8 members on average. That is, I will have only six market betas to conduct the Fama-MacBeth regressions. On the other hand, the individual stock approach allows me to have 50 different numbers in each Fama-MacBeth regression. Therefore, the paper uses the individual stock approach to estimate the market risk premium. Nonetheless, in the analysis, one should be careful with the effect of idiosyncratic volatility.

ME is a stock's market capitalization. BM is a stock's book value to market capitalization. Both ME and BM are updated twice a year at the beginning of January and July of each year t. CR is a stock's CheiRank that implies the degree of systematic importance. PR is a stock's PageRank that indicates the degree of fragility. The calculation of CR and PR is explained in Appendix 4.6. In addition, both CR and PR are multiplied by 1,000 before taking the natural logarithm in order to make them remain positive. In Section 4.4.1, we can see that the values of CR and PR are very low. The low CR portfolio, for example, is 0.0076 on average. As a result, it is needed to multiply by 1,000 to make the natural logarithm in the positive range.

Table 4.2 reports the average slopes of the monthly cross-sectional regressions in Equation (4.5). The average slope is the average of a factor's estimated coefficients over 180 months. The t-statistic in the blanket is the average slope divided by its time series standard error. The one-factor regressions are reported in Model 1-5. Similar to the existing literature, the market risk premium is not significant in explaining the cross-sectional returns. While the size and book-to-market factors are well-known to be significant in developed markets, their effect on the developing countries is mixed. In this paper, I find that the size and BM factors are not significant in explaining the cross-sectional returns. Nonetheless, the signs of the factors are consistent with the existing literature. Size has the negative relationship with the returns. That is, big stocks tend to earn lower returns than the small stocks. Book-to-market has the positive relationship with returns. This result indicates that the value stocks have higher returns.

Out of five one-factor regressions, CheiRank's average slope is solely significant. Its negative coefficient implies that the high CheiRank stocks earn lower returns than the low CheiRank stocks. The possible reason may be their relative systematic importance which gives them multiple connections that can send transmitted shocks and their own shocks to the other stocks. For instance, in Figure 4.3, KBANK and BBL are two stocks with the highest CheiRank. They are systematically important because they have many outward links that can affect the others. When a shock hits KBANK and BBL, it will be quickly diversified away by the network. As a result, investors would demand less compensation than the low CheiRank stock. On the other hand, the PageRank is positively related to the stock returns, but not statistically significant. That is, a transmitted shock is highly likely to find its way to the high

PageRank stocks which has no way to diversify shocks. In Figure 4.3, such stocks are BEC and GLOW who has the highest PageRank among others. It is natural that investors would demand high compensation for such fragile stocks.

In Model 6, the stock returns are regressed on market risk and CheiRank. Interestingly, the coefficient of CheiRank becomes insignificant. There are two explanations for this result. First, it could be that the CheiRank captures some undiversified risk that is also captured by the market risk premium. Another reason is the effect of idiosyncratic volatility on the pre-ranking beta of the individual stocks. Since the network structure serves as the channel for transmission of idiosyncratic shocks, the network measures, and idiosyncratic volatility is relatively correlated.

Model 7 regresses the stock returns on market risk and PageRank. Model 8 regresses the stock returns on size and book-to-market. In Model 10, the individual stock returns are regressed on the three factors of Fama and French (1993), including market beta, size, and book-to-market. The three cross-sectional tests confirm that market risk, size, book-to-market, and PageRank are not statistically significant in explaining cross-sectional returns.

Model 9 regresses the stock returns on CheiRank and PageRank. The coefficient of the CheiRank is still significant at the 90% confidence level. Compared to the one-factor Model 4, the effect of CheiRank decreases from -0.357 to -0.268. This result is expected because PageRank is somewhat reciprocal to CheiRank and both factors are highly correlated. In Model 11, the market beta is added, and the regression becomes a three-factor model. The coefficient of CheiRank becomes even lower at -0.104.

In Model 12 (13), I regress the stock returns on three factors, including, size, book-to-market and CheiRank (PageRank). The goal is to see whether or not the network measures are cross-sectionally affected by the size and book-to-market. The coefficient of CheiRank at -0.336 does not change much from the one-factor Model 4 at -0.357. Likewise, The coefficient of PageRank at 0.623 does not change much from the one-factor Model 5 at 0.592. The results infer that CheiRank (PageRank)'s ability to explain stock returns do not come from size and book-to-market.

Table 4.2 Average slopes from month-by-month cross-sectional regressions.

	Model	**const**	**MKTB**	**ln(ME)**	**ln(BM)**	**ln(CR*1000)**	**ln(PR*1000)**
One factor	1	0.984**	-0.638				
		[2.157]	[-1.150]				
	2	1.750		-0.058			
		[0.434]		[-0.377]			
	3	0.381			0.207		
		[0.605]			[1.102]		
	4	1.342**				-0.357**	
		[2.205]				[-2.333]	
	5	-1.278					0.592
		[-0.898]					[1.595]
Two factors	6	1.296**	-0.488			-0.188	
		[2.362]	[-0.829]			[-1.296]	
	7	-0.480	-0.422				0.433
		[-0.413]	[-0.727]				[1.273]
	8	-0.555		0.040	0.194		
		[-0.136]		[0.252]	[1.023]		
	9	0.447				-0.268*	0.229
		[0.297]				[-1.716]	[0.606]
Three factors	10	-2.901	-0.782	0.162	0.244		
		[-0.782]	[-1.365]	[1.102]	[1.288]		
	11	0.066	-0.355			-0.104	0.308
		[0.048]	[-0.592]			[-0.677]	[0.851]
	12	-0.580		0.076	0.254	-0.336**	
		[-0.143]		[0.473]	[1.360]	[-2.062]	
	13	-3.463		0.086	0.254		0.623
		[-0.751]		[0.540]	[1.353]		[1.640]
Five factors	14	-4.785	-0.512	0.212	0.291	-0.118	0.255
		[-1.118]	[-0.848]	[1.395]	[1.507]	[-0.781]	[0.646]

(Note) The table reports average slopes and t-statistics of the monthly cross-sectional regressions from January 2000 to December 2014. In each month, individual excess stock returns are regressed on the independent variables, including market beta (MKTB), size (market capitalization or ME), book-to-market (BM), CheiRank (CR), and PageRank (PR). MKTB is the pre-ranking beta of an individual stock. ln(ME) is a stock's market capitalization. ln(BM) is a stock's book value to market capitalization. Both ln(ME) and ln(BM) are updated twice a year at the beginning of January and July of each year t. ln(CR*1000) is a stock's CheiRank that implies the degree of systematic importance. ln(PR*1000) is a stock's PageRank that indicates the degree of fragility

4.4.3 Relationship between CheiRank/PageRank, size, and book-to-market

Since the study of Fama and French (1993), size and book-to-market have been two well-known common risk factors in addition to the market risk. It becomes a common practice to use size and book-to-market as a benchmark or controlled factors in asset pricing tests. Therefore, this paper also investigates the relationship between CheiRank/PageRank, size, and book-to-market.

To begin with, size is measured by market capitalization, retrieved from Bloomberg. Book-to-market (BM) is an acronym for book-value to market equity of stock. Both size and BM are updated twice a year at the beginning of January and July of each year t. Please note that the variables are updated semiannually because the SET50 constitution changes twice a year. Using these data, I sort the stocks into three portfolios by size or book-to-market. The breakpoints are the 30th and 70th percentile. Therefore, the top 30% are large size stocks (B) or high BM stocks (H). The middle 40% are medium size stocks (M) or medium BM stocks (M). The lower 30% are small stocks (S) or low BM stocks (L). Table 4.3 reports the average CheiRank and average PageRank of portfolios ranked on size or book-to-market (BM). Since the big stocks have the highest CheiRank among others, they are systemically important to the stock market. Although the big stocks tend to be less fragile than the rest, the average PageRank is not much different from the small and medium stocks. Moreover, I also find evidence of the relationship between book-to-market and CheiRank. The average CheiRank does increase a little across the BM portfolios. On the other hand, the relationship between book-to-market and PageRank is rather weak as the average PageRank of the three BM portfolios is very close.

To further investigate this matter, I conduct the panel regression test of CheiRank/PageRank on size and book-to-market as follows:

$$ln(CR*1000)_{i,t} = const + b_1 size_{i,t} + e_{i,t} \tag{4.6}$$

$$ln(PR*1000)_{i,t} = const + b_2 size_{i,t} + e_{i,t} \tag{4.7}$$

$$ln(CR*1000)_{i,t} = const + b_3 BM_{i,t} + e_{i,t} \tag{4.8}$$

$$ln(PR*1000)_{i,t} = const + b_4 BM_{i,t} + e_{i,t} \tag{4.9}$$

$ln(CR*1000)_{i,t}$ is the CheiRank of firm *i* in month *t*. $ln(PR*1000)_{i,t}$ is the PageRank of firm *i* in month *t*. Both CheiRank and PageRank are calculated from the PCPG network as described in Section 4.4.1. $size_{i,t}$ is the equal-weighted average of daily sizes of firm *i* within month *t*. $BM_{i,t}$ is the equal-weighted average of daily book-to-market of firm *i* within month *t*. Table 4.4 reports the regression estimates of four specifications above. In Panel A, the relationship between CheiRank and average firm size is significantly positive with t-statistics of 16.981 and adjusted R^2 of 0.031. This result is consistent with the findings from Table 4.3 whereby large stocks tend to be more systematically important. The relationship between PageRank and average firm size is significantly negative with t-statistics of -9.937 and adjusted R^2 of 0.011. That is, the small stocks tend to be more fragile than the large stock. Similarly, the relationship between CheiRank (PageRank) and book-to-market is statistically significant and positive (negative). However, the adjusted R^2 is very low at 0.005 and 0.002 for CheiRank and PageRank, respectively. Additionally, the sizes of coefficients are month lower than those in Panel A. This finding confirms our earlier observation in Table 4.3.

Table 4.3 Average CheiRank and PageRank on size/book-to-market portfolios.

	Portfolios ranked on size			**Portfolios ranked on book-to-market**		
	Small	Medium	Big	Low	Medium	High
Average CheiRank	0.0151	0.0158	0.0307	0.0145	0.0238	0.0205
Average PageRank	0.0203	0.0209	0.0186	0.0207	0.0198	0.0197
Number of stocks	15	20	15	15	20	15

(Note) Table 4.3 presents the equal-weighted average of CheiRank and PageRank on portfolios formed by size and book-to-market. Small (Low) portfolio consists of stocks with market capitalization (book-to-market) below 30th percentile. Medium portfolio consists of stocks with market capitalization (book-to-market) between 30th and 70th percentile. Big (High) portfolio consists of stocks with market capitalization (book-to-market) above 70th percentile.

Table 4.4 Panel regressions between CheiRank/PageRank, size, and book-to-market.

Panel A: Explanatory factor is average firm size

	const	b	t(const)	t(b)	Adj R^2
ln(CR*1000)	-0.091	0.105**	-0.600	16.981	0.031
ln(PR*1000)	3.613	-0.028**	53.187	-9.937	0.011

Panel B: Explanatory factor is average book-to-market

	const	b	t(const)	t(b)	Adj R^2
ln(CR*1000)	2.525	0.065**	245.744	6.764	0.005
ln(PR*1000)	2.925	-0.020**	641.208	-4.702	0.002

(Note) Panel A presents the estimation of panel regressions of CheiRank and PageRank on firms' size as shown in Equation (4.6) and (4.7). Panel B presents the estimation of panel regressions of CheiRank and PageRank on book-to-market as shown in Equation (4.8) and (4.9). CheiRank is calculated from the outgoing link matrix and represents the relative influence of a firm by its ability to transmit its shocks to the others. PageRank is calculated from the incoming link matrix and represents the relative influence of a firm by its vulnerability to propagated shocks. Size is market capitalization. Book-to-market is the ratio of book value to market equity. ** is significant at 95% confidence interval.

4.4.4 Relationship between returns, size, book-to-market, CheiRank, and PageRank

This section explores the characteristics of monthly portfolio returns ranked on size, book-to-market, CheiRank, and PageRank. Similar to the previous sections, I sort the excess stock returns into three portfolios with the breakpoints at the 30th and 70th percentile. Table 4.5 reports the excess returns and the standard deviation for each portfolio. The excess returns are the value-weighted portfolio returns minus the risk-free rate, and the standard deviation is calculated accordingly. In consistency with Fama and French (1993), I find that the small stocks have higher returns than the big stocks. The average portfolio return on the small stocks is 0.58% with the standard deviation of 9.95%, while the average portfolio return on the big stocks is 0.43% with the standard deviation of 7.65%. However, the general trend of returns sorted on size does not conform to their finding. The portfolio return of the medium size stocks that should be higher than the big size is apparently lower in my data. This issue is possibly caused by the specific nature of the data set which focuses the emerging market (Thailand) instead of the developed market (US).

Book-to-market, on the other hand, has a clear increasing trend in portfolio returns. Consistent with the existing literature, the value stocks tend to have higher returns than the growth stocks. The average portfolio return on the high book-to-market stocks is 0.67% with the standard deviation of 9.10%, while the average portfolio return on the low book-to-market stocks is 0.25% with the standard deviation of 7.34%.

Both CheiRank and PageRank have a noticeable decreasing or increasing trend on average portfolio returns, respectively. The average portfolio return on the high CheiRank stocks is 0.18% with the standard deviation of 9.09%, while the average portfolio return on the low CheiRank stocks is 0.68% with the standard deviation of 6.88%. It is clear that the high CheiRank stocks earn considerably lower returns than the low CheiRank stocks. On the contrary, The average portfolio return on the high PageRank stocks is 0.57% with the standard deviation of 7.28%, while the average portfolio return on the low PageRank stocks is 0.32% with the standard deviation of 8.83%. That is, the low PageRank stocks earn much lower returns than the high PageRank stocks. The findings on CheiRank and PageRank are indeed complementary as they measure two opposite characteristics of the interconnectedness. In other words, the stock returns are negatively related to their relative systemic importance and positively related to their relative fragility in the system.

Table 4.5 Average monthly excess returns of portfolios sorted on size, book-to-market, CheiRank, and PageRank.

	Average monthly excess returns			**Standard deviations**		
	Small/Low	Medium	Big/High	Small/Low	Medium	Big/High
Size	0.0058	0.0032	0.0043	0.0995	0.0801	0.0765
Book-to-market	0.0025	0.0059	0.0067	0.0734	0.0834	0.0910
CheiRank	0.0068	0.0040	0.0018	0.0688	0.0781	0.0909
PageRank	0.0032	0.0048	0.0057	0.0883	0.0784	0.0728

(Note) Table 4.5 presents the average monthly excess returns and standard deviations of portfolio formed on size, book-to-market, CheiRank and PageRank. Size is market capitalization. Book-to-market is the ratio of book value to market equity. CheiRank is calculated from the outgoing link matrix and represents the relative influence of a firm by its ability to transmit its shocks to the others. PageRank is calculated from the incoming link matrix and represents the relative influence of a firm by its vulnerability to propagated shocks. The breakpoints for the classification are 30th and 70th percentiles

4.4.5 Portfolio construction for asset pricing tests

The primary purpose of this section is to study the relationship between returns and interconnectedness measures after controlling for size and book-to-market. Following the portfolio construction approach of Fama and French (1993), I form six intersection portfolios based on two risk factors. For example, 2x3 sorts on size and CheiRank are S/CRL, S/CRM, S/CRH, B/CRL, B/CRM, and B/CRH. S stands for small size, B stands for big size. CRL, CRM, and CRH represent low, medium, high CheiRank, respectively. Thus, S/CRL is a portfolio of small firms with low CheiRank. S/CRM is a portfolio of small firms with medium CheiRank. S/CRH is a portfolio of small firms with high CheiRank. Similarly, B/CRL, B/CRM, and B/CRH are portfolios of big firms with low, medium, high CheiRank, respectively.

In addition, I also form 2x3 sorts on size and PageRank, 2x3 sorts on book-to-market and CheiRank, 2x3 sorts on book-to-market and PageRank. Six portfolios formed on size and PageRank are S/PRL, S/PRM, S/PRH, B/PRL, B/PRM, and B/PRH. The portfolios formed on book-to-market and CheiRank are L/CRL, L/CRM, L/CRH, H/CRL, H/CRM, and H/CRH. The portfolios formed on book-to-market and PageRank are L/PRL, L/PRM, L/PRH, H/PRL, H/PRM, and H/PRH. L stands for low book-to-market, H stands for high book-to-market. PRL, PRM, and PRH represent low, medium, high PageRank, respectively. Please note that three categories of CheiRank/PageRank are used to form the intersection portfolios because the main attention of this paper is the interconnectedness. The breakpoints are the 30^{th} and 70^{th} percentile, as usual. Due to limited data, I use only two categories of size and book-to-market to rank the stocks in this analysis. The stocks above the 50^{th} percentile are considered as big size (B) or high book-to-market (H). The stocks below the 50^{th} percentile are considered as small size (B) or low book-to-market (H).

Table 4.6 reports the mean excess returns and standard deviations on the intersection portfolios. Panel A shows the statistics of portfolios formed by size and CheiRank as well as book-to-market and CheiRank. The decreasing mean excess returns across the CheiRank portfolios clearly support the negative relationship between return and CheiRank. This relationship is noticeable in the big stocks and particularly strong in the small stocks. The mean excess return of S/CRL is 0.0124 while the mean of S/CRH is -0.0074. In Section 4.4.3, I find the significant relationship between book-to-market and CheiRank, but the adjusted R^2 is very low. The statistics in Panel A of Table 4.6 provide one explanation for this matter. The mean excess returns of the low book-to-market stocks decline along with the increase in CheiRank. The mean excess returns of the high book-to-market stocks, on the contrary, have a reverse pattern in medium and high CheiRank portfolios. These findings suggest that the weak relationship between book-to-market and CheiRank is primarily due to the inconsistent behavior of the value stocks in the network.

Panel B shows the statistics of portfolios formed by size and PageRank as well as book-to-market and PageRank. The mean excess returns on the small stocks increase noticeably as the stocks become more fragile. The mean excess return of S/PRL is 0.0010 while the mean of S/CRH is 0.0115. The returns on the big stocks also increase marginally with the PageRank. Thus, the findings confirm the positive relationship between size and PageRank. Because of this marginal increase in returns of the big stocks, the size has the weaker relationship with PageRank than CheiRank. Furthermore, the results show that the weak relationship of the high book-to-market stocks is the cause of the high book-to-market portfolio. In the low book-to-market region, the mean returns of the PageRank portfolio monotonically increase from 0.0028 to 0.0030 to 0.0064. On the other hand, the mean returns of the PageRank portfolio with high book-to-market have no trend. The medium PageRank has the highest return at 0.0092 while the others have much lower returns.

Table 4.6 Summary statistics for the intersection portfolios.

Panel A

	CheiRank portfolios					
Size	CRL=Low	CRM=Medium	CRH=High	CRL=Low	CRM=Medium	CRH=High
	Mean excess returns			Standard deviations		
S=Small	0.0124	0.0087	-0.0074	0.0877	0.0967	0.1096
B=Big	0.0043	0.0033	0.0023	0.0715	0.0779	0.0913
Book-to-market						
L=Low	0.0057	0.0046	-0.0003	0.0694	0.0834	0.0905
H=High	0.0130	0.0026	0.0037	0.0901	0.0834	0.0995

Panel B

	PageRank portfolios					
Size	PRL=Low	PRM=Medium	PRH=High	PRL=Low	PRM=Medium	PRH=High
	Mean excess returns			Standard deviations		
S=Small	0.0010	0.0049	0.0115	0.1019	0.0923	0.0949
B=Big	0.0032	0.0047	0.0047	0.0882	0.0799	0.0739
Book-to-market						
L=Low	0.0028	0.0030	0.0064	0.0891	0.0821	0.0762
H=High	0.0039	0.0092	0.0040	0.0960	0.0896	0.0921

(Note) Panel A presents the mean excess returns and standard deviations formed by size, book-to-market and CheiRank. The six intersection portfolios formed by size and CheiRank are S/CRL, S/CRM, S/CRH, B/CRL, B/CRM, and B/CRH. S/CRL(B/CRL) is a portfolio of small (big) firms with low CheiRank. S/CRM(B/CRM) is a portfolio of small (big) firms with medium CheiRank. S/CRH(B/CRH) is a portfolio of small (big) firms with high CheiRank. The six intersection portfolios formed by book-to-market and CheiRank are L/CRL, L/CRM, L/CRH, H/CRL, H/CRM, and H/CRH. L/CRL(H/CRL) is a portfolio of low (high) book-to-market firms with low CheiRank. L/CRM(H/CRM) is a portfolio of low (high) book-to-market firms with medium CheiRank. L/CRH(H/CRH) is a portfolio of low (high) book-to-market firms with high CheiRank. Panel B reports the mean excess returns and standard deviations formed by size, book-to-market, and PageRank. The six portfolios formed by size and PageRank are S/PRL, S/PRM, S/PRH, B/PRL, B/PRM, and B/PRH. The six portfolios formed by book-to-market and PageRank are L/PRL, L/PRM, L/PRH, H/PRL, H/PRM, and H/PRH.

4.4.6 Regression results for CheiRank-Size/Book-to-market portfolios

This section examines the relationship between CheiRank, size, and book-to-market (BM) by using the asset pricing models similar to Fama and French (1993). The primary goal is to test the performance of CheiRank in explaining the time-series of expected portfolio returns. My regression models take the following forms:

$$R_{p,t} - R_{f,t} = const_t + b_p\left(R_{m,t} - R_{f,t}\right) + s_p SMB_t + c_p CLSMCHS_t + e_{p,t} \quad (4.10)$$

$$R_{p,t} - R_{f,t} = const_t + b_p\left(R_{m,t} - R_{f,t}\right) + h_p HML_t + c_p CLBMCHB_t + e_{p,t} \quad (4.11)$$

$R_{p,t} - R_{f,t}$ is the excess returns on portfolio p. $R_{m,t}$ is the market returns and calculated from the value-weighted average of all stock returns in month t. SMB_t is the difference between the equal-weight average of returns on small stocks (S/CRL, S/CRM, and S/CRH) and the equal-weight average of returns on big stocks (B/CRL, B/CRM, and B/CRH). HML_t is the difference between the equal-weight average of returns on value stocks (H/CRL, H/CRM, and H/CRH) and the equal-weight average of returns on growth stocks (L/CRL, L/CRM, and L/CRH). $CLSMCHS_t$ in Equation (4.10) is the difference between the equal-weight average of returns on low CheiRank stocks (S/CRL and B/CRL) and the equal-weight average of returns on high CheiRank stocks (S/CRH and B/CRH). $CLBMCHB_t$ in Equation (4.11) is the difference between the equal-weight average of returns on low CheiRank stocks (L/CRL and H/CRL) and the equal-weight average of returns on high CheiRank stocks (L/CRH and H/CRH).

Table 4.7 reports the regression estimates for the multifactor models. In Panel A, the parameters for size-CheiRank portfolios have been estimated. The market risk factor ($R_{m,t} - R_{f,t}$) is positive and highly significant at 5% level for all portfolios. Thus, the increase in 1 unit of market risk would demand higher compensation on equity returns across the market. However, since the slopes are marginally different across portfolios formed on size and CheiRank, it is obvious that the market risk factor alone cannot explain the differences in portfolio returns. SMB_t is negative and highly significant at 5% level for the big stocks, except for B/CRL. Also, since the coefficients monotonically decrease with the CheiRank, the size factor is partly responsible for the difference in CheiRank portfolio returns with large stocks. On the other hand, SMB_t is positive and highly significant at 5% level for all three portfolios of small stocks. The positive sign of the coefficients tells us that the size effect is needed to be compensated for the small stocks. An interesting point here is that the coefficient of SMB_t on S/CRL is lower than that on S/CRH. This does not conform to the previous findings, which indicate that the stock returns decrease as CheiRank or systematic important level increases. Therefore, the effect of SMB_t on systematic importance is noticeable for big stocks and unclear for the small stocks. It seems that when a small stock is systematic important in the network, it is priced higher than usual. This matter can be an interesting topic for the future research.

$CLSMCHS_t$ in Equation (4.10) is highly significant at 5% level for all portfolios except for B/CRM. The coefficients of CheiRank is positive for S/CRL and S/CRM portfolios but becomes positive for the S/CRH portfolio. The coefficients also decrease monotonically with the CheiRank. The result is also similar for the big portfolios. Thus, $CLSMCHS_t$ can effectively explain the difference in CheiRank portfolio returns, regardless of the size. Furthermore, the CheiRank factor appears to have an only marginal effect on the return difference of small and big portfolios because the coefficients on small and big portfolios are almost indifferent.

In Panel B, the parameters for BM-CheiRank portfolios have been estimated. The market risk factor ($R_{m,t} - R_{f,t}$) is positive and highly significant at 5% level for all portfolios. Thus, the market risk demands some compensation for these portfolios. However, similar to the size effect, the market factor alone is not sufficient to explain the discrepancy in portfolio returns because the coefficients are marginally different. HML_t is negative and highly significant at 5% level for the growth stocks. On the other hand, HML_t is positive and highly significant at 5% level for all three portfolios of value stocks. This stark difference in growth and value portfolios advocate the existence of the book-to-market anomaly in returns of SET50 stocks. Furthermore, the book-to-market factor is somewhat related to the CheiRank as we observe the monotonic increase (decrease) of the slopes for low (high) book-to-market factor.

$CLBMCHB_t$ in Equation (4.11) is highly significant at 5% level for all portfolios except for the medium CheiRank portfolios. The coefficient for L/CRM is significant at 5% level while the coefficient for H/CRM is insignificant. Similar to the CheiRank factor formed by size, the low CheiRank portfolios demand much higher demand for return compensation than the high CheiRank. In fact, the slope of the high CheiRank portfolios is even negative and thus make the portfolio returns very low. Therefore, the effect of CheiRank is significant for portfolios formed on book-to-market and CheiRank.

All in all, CheiRank is a significant risk factor that can affect the time-series returns on equities. The stocks with high CheiRank tend to be systematic important in the network and earn lower expected returns than the other stocks.

Table 4.7 Regression estimates of excess stock returns on the portfolios formed by CheiRank, size, and book-to-market from January 1990 to December 2014.

Panel A						
	CheiRank portfolios					
Size	CRL=Low	CRM=Medium	CRH=High	CRL=Low	CRM=Medium	CRH=High
		const			*t stat (const)*	
S=Small	0.0005	0.0020	-0.0075	0.3045	0.8625	-2.9850
B=Big	-0.0060	-0.0010	0.0020	-2.4085	-0.4906	1.2080
		b			*t stat (b)*	
S=Small	1.0339**	1.0394**	0.9178**	42.2276	31.7073	25.3763
B=Big	0.9447**	0.9857**	1.0608**	26.2833	33.2940	44.4096
		s			*t stat (s)*	
S=Small	0.8616**	0.8231**	0.9835**	21.9827	15.6850	16.9861
B=Big	-0.0419	-0.1262**	-0.1638**	-0.7287	-2.6627	-4.2830
		c			*t stat (c)*	
S=Small	0.5946**	0.1317**	-0.4537**	15.6779	2.5941	-8.0989
B=Big	0.5951**	0.0340	-0.3566**	10.6900	0.7421	-9.6372
		Adj R^2				
S=Small	0.9369	0.9070	0.9118			
B=Big	0.7953	0.8832	0.9447			

Panel B						
	CheiRank portfolios					
Book-to-market	CRL=Low	CRM=Medium	CRH=High	CRL=Low	CRM=Medium	CRH=High
		const			*t stat (const)*	
L=Low	-0.0006	0.0005	-0.0010	-0.2973	0.1962	-0.5479
H=High	0.0009	-0.0034	0.0013	0.4634	-1.4740	0.8095
		b			*t stat (b)*	
L=Low	0.9356**	1.0678**	1.0098**	29.8029	31.0952	37.5762
H=High	1.1021**	0.8832**	1.0278**	37.6367	26.1174	43.8895
		h			*t stat (h)*	
L=Low	-0.4526**	-0.3602**	-0.1647**	-8.2466	-5.9997	-3.5053
H=High	0.7897**	0.7310**	0.5018**	15.4260	12.3636	12.2559
		c			*t stat (c)*	
L=Low	0.5046**	0.1146**	-0.3825**	11.5065	2.3896	-10.1868
H=High	0.6673**	0.0151	-0.4456**	16.3113	0.3206	-13.6194
		Adj R^2				
L=Low	0.8338	0.8623	0.9284			
H=High	0.9143	0.8666	0.9550			

(Note) Panel A presents the regression estimates of excess portfolio returns on market risk, SMB, and CLSMCHS, as shown in Equation (4.9). The six intersection portfolios formed by size and CheiRank are S/CRL, S/CRM, S/CRH, B/CRL, B/CRM, and B/CRH. S/CRL(B/CRL) is a portfolio of small (big) firms with low CheiRank. S/CRM(B/CRM) is a portfolio of small (big) firms with medium CheiRank. S/CRH(B/CRH) is a portfolio of small (big) firms with high CheiRank. Panel B presents the regression estimates of excess portfolio returns on market risk, HML, and CLBMCHB, as shown in Equation (4.10). The six intersection portfolios formed by book-to-market and CheiRank are L/CRL, L/CRM, L/CRH, H/CRL, H/CRM, and H/CRH. L/CRL(H/CRL) is a portfolio of low (high) book-to-market firms with low CheiRank. L/CRM(H/CRM) is a portfolio of low (high) book-to-market firms with medium CheiRank. L/CRH(H/CRH) is a portfolio of low (high) book-to-market firms with high CheiRank. ** is significant at 95% confidence interval. * is significant at 90% confidence interval.

4.4.7 Regression results for PageRank-Size/Book-to-market portfolios

In this section, I explore the relationship between PageRank, size, and book-to-market (BM) by using the asset pricing models similar to Fama and French (1993). The primary goal is to test the performance of PageRank in explaining the cross-section of expected portfolio returns. The regression models take the following forms:

$$R_{p,t} - R_{f,t} = const_t + b_p\left(R_{m,t} - R_{f,t}\right) + s_p SMB_t + c_p PHSMPLS_t + e_{p,t} \quad (4.12)$$

$$R_{p,t} - R_{f,t} = const_t + b_p\left(R_{m,t} - R_{f,t}\right) + h_p HML_t + c_p PHBMPLB_t + e_{p,t} \quad (4.13)$$

$R_{p,t} - R_{f,t}$ is the excess returns on portfolio *p*. $R_{m,t}$ is the market returns and calculated from the value-weighted average of all stock returns in month *t*. SMB_t is the difference between the equal-weight average of returns on small stocks (S/PRL, S/PRM, and S/PRH) and the equal-weight average of returns on big stocks (B/PRL, B/PRM, and B/PRH). HML_t is the difference between the equal-weight average of returns on value stocks (H/PRL, H/PRM, and H/PRH) and the equal-weight average of returns on growth stocks (L/PRL, L/PRM, and L/PRH). $PHSMPLS_t$ in Equation (4.12) is the difference between the equal-weight average of returns on high PageRank stocks (S/PRH and B/PRH) and the equal-weight average of returns on low PageRank stocks (S/PRL and B/PRL). $PHBMPLB_t$ in Equation (4.13) is the difference between the equal-weight average of returns on high PageRank stocks (L/PRH and H/PRH) and the equal-weight average of returns on low PageRank stocks (L/PRL and H/PRL).

Table 4.8 reports the regression estimates for the multifactor models. In Panel A, the parameters for size-PageRank portfolios have been estimated. The PageRank-related results for the market factor and size factor are similar to those of PageRank in the previous section. The market risk factor ($R_{m,t} - R_{f,t}$) is positive and highly significant at 5% level for all portfolios. Therefore, the increase in 1 unit of market risk would demand higher compensation on equity returns across the market. Nevertheless, due to the marginal increase or decrease on the slopes, the stock returns need some new risk factors to explain the variation in returns in addition to the market risk.

SMB_t is negative and highly significant at 5% level for the big stocks, except for B/PRH. On the other hand, SMB_t is positive and highly significant at 5% level for all three portfolios of small stocks. Since the coefficients monotonically decrease with the PageRank, the size factor is partly responsible for the difference in PageRank portfolio returns in the small size domain. The vast difference in coefficients for small and big portfolios supports the existence of the sizes anomaly in returns of SET50 stocks. Furthermore, I observe that the size factor goes against the general trend of the PageRank portfolios in the small portfolios. The coefficient of SMB_t on S/PRL is higher than that on S/PRH. This does not conform to the previous findings, which indicate that the stock returns increase as PageRank or fragility level increases. On the contrary, the pattern in the big portfolios is consistent with the PageRank. B/ PRH has higher coefficient than B/PRL.

$PHSMPLS_t$ in Equation (4.12) is highly significant at 5% level for all portfolios except for B/CRM which is significant at 5% level. The slope of PageRank is positive for the S/PRL portfolio but becomes positive for the S/PRM and S/PRH portfolios. The coefficients also increase monotonically with the PageRank value. The result is similar for the big portfolios. Therefore, $PHSMPLS_t$ can effectively explain the difference in PageRank portfolio returns, regardless of the size. Moreover, the PageRank factor appears to have a noticeable effect on the return difference of both small and big portfolios because the PageRank coefficients on small and big portfolios are relatively different. This finding suggests that the fragility effect appears to be more consistent than the effect of systemically importance.

In Panel B, the parameters for BM-PageRank portfolios have been estimated. The market risk factor ($R_{m,t} – R_{f,t}$) is positive and highly significant at 5% level for all portfolios. Thus, the market risk demands some compensation for these portfolios. However, similar to the results in Panel A, the market factor alone is not sufficient to explain the difference in portfolio returns. HML_t is negative and highly significant at 5% level for the growth stocks. On the other hand, HML_t is positive and highly significant at 5% level for all three portfolios of value stocks. This empirical evidence supports the existence of the book-to-market anomaly in returns of SET50 stocks. Furthermore, the relationship between book-to-market and PageRank is found for the high BM stocks but is unclear for the low BM stocks.

$PHBMPLB_t$ in Equation (4.13) is highly significant at 5% level for all portfolios. Similar to the PageRank factor formed on size, the high PageRank portfolios demand much higher demand for return compensation than the low PageRank. In fact, the slope of the low PageRank portfolios is even negative and thus make the portfolio returns very low. Therefore, the effect of PageRank is significant for portfolios formed on book-to-market and PageRank.

In short, PageRank is a significant risk factor that can affect the returns on equities. The stocks with high PageRank tend to be fragile and earn higher expected returns than the other stocks.

Table 4.8 Regression estimates of excess stock returns on the portfolios formed by PageRank, size, and book-to-market from January 1990 to December 2014.

Panel A

PageRank portfolios

Size	PRL=Low	PRM=Medium	PRH=High	PRL=Low	PRM=Medium	PRH=High
		const			*t stat (const)*	
S=Small	-0.0011	-0.0014	0.0016	-0.4572	-0.6829	1.0022
B=Big	0.0009	0.0001	-0.0018	0.5742	0.0442	-0.7385
		b			*t stat (b)*	
S=Small	0.9250**	0.9858**	1.0993**	29.8474	34.6437	50.0939
B=Big	1.0799**	1.0246**	0.9056**	52.7574	43.1127	27.3996
		s			*t stat (s)*	
S=Small	0.9517**	0.8730**	0.8551**	17.7267	17.7105	22.4933
B=Big	-0.1330**	-0.1507**	-0.0365	-3.7511	-3.6610	-0.6369
		p			*t stat (p)*	
S=Small	-0.5465**	0.1422**	0.6622**	-9.6203	2.7263	16.4614
B=Big	-0.3165**	0.0995**	0.4748**	-8.4342	2.2848	7.8376
		Adj R^2				
S=Small	0.9114	0.9089	0.9488			
B=Big	0.9484	0.9152	0.8083			

Panel B

PageRank portfolios

Book-to-market	PRL=Low	PRM=Medium	PRH=High	PRL=Low	PRM=Medium	PRH=High
		const			*t stat (const)*	
L=Low	-0.0007	-0.0007	0.0018	-0.3622	-0.3238	0.9203
H=High	-0.0004	0.0038	-0.0029	-0.2513	1.6309	-1.6219
		b			*t stat (b)*	
L=Low	1.0389**	1.0349**	1.0103**	39.4961	34.0324	36.0322
H=High	1.0564**	0.9428**	1.0849**	43.2013	28.4475	42.4278
		h			*t stat (h)*	
L=Low	-0.0987**	-0.4134**	-0.4394**	-2.1159	-7.6627	-8.8333
H=High	0.4422**	0.8235**	0.7829**	10.1936	14.0057	17.2568
		p			*t stat (p)*	
L=Low	-0.3735**	0.0396	0.6187**	-8.2273	0.7551	12.7843
H=High	-0.3946**	0.0661	0.6133**	-9.3497	1.1559	13.8956
		Adj R^2				
L=Low	0.9222	0.8776	0.8793			
H=High	0.9421	0.8780	0.9313			

(Note) Panel A presents the regression estimates of excess portfolio returns on market risk, SMB, and PHSMPLS, as shown in Equation (4.11). The six intersection portfolios formed by size and PageRank are S/PRL,S/PRM, S/PRH, B/PRL, B/PRM, and B/PRH. S/PRL(B/PRL) is a portfolio of small (big) firms with low PageRank. S/PRM(B/PRM) is a portfolio of small (big) firms with medium PageRank. S/PRH(B/PRH) is a portfolio of small (big) firms with high PageRank. Panel B presents the regression estimates of excess portfolio returns on market risk, HML, and PHBMPLB, as shown in Equation (4.12). The six intersection portfolios formed by book-to-market and PageRank are L/PRL,L/PRM, L/PRH, H/PRL, H/PRM, and H/PRH. L/PRL(H/PRL) is a portfolio of low (high) book-to-market firms with low PageRank. L/PRM(H/PRM) is a portfolio of low (high) book-to-market firms with medium PageRank. L/PRH(H/PRH) is a portfolio of low (high) book-to-market firms with high PageRank. ** is significant at 99% confidence interval. * is significant at 95% confidence interval.

4.5 Conclusion

In this paper, I examine the relationship between two interconnectedness measures and returns on listed firms in SET50. The first measure is CheiRank which reflects the level of systematic importance. A highly systematic-important firm can efficiently propagate its shocks to the other firms in the network. Another measure is PageRank which quantifies the level of fragility. A firm is fragile if it is vulnerable to propagated shocks. I find that the systematically-important firms (high CheiRank) earn lower returns while the fragile stocks (high PageRank) earns higher returns than the remaining stocks. I also find evidence of a positive (negative) relationship between size and CheiRank (PageRank). The big firms, for instance, tend to be systematically important in the network of stocks in SET50. Also, I find a significant relationship between book-to-market and the interconnectedness measures. However, the relationship with book-to-market is much weaker than with size.

By conducting the cross-sectional regressions similar to Fama and MacBeth (1973), I find that size, book-to-market, and market risk are well-known risk factors, but they are all insignificant in my data. More importantly, I find that CheiRank is significant in explaining cross-sectional returns among individual stocks, whereas PageRank is not. One possible explanation of the CheiRank factor is that the property of systematical importance induces different behavior of stocks in the network. High systematic importance means that a transmitted shock can be efficiently transferred to other stocks via a network structure. On the other hand, low systematically important stocks do not have the ability to transfer and diversify away the shock.

Following the approach of Fama and French (1993), I examine whether or not, CheiRank and PageRank are common risk factors in returns in addition to the well-known risk factors such as market risk, size, and book-to-market. The regression estimates suggest the significant relationship between interconnectedness and stock returns. The slope of the CheiRank factor diminishes and even turns negative as the CheiRank increases. Therefore, in the context of PCPG networks, the high CheiRank stocks may be useful in distributing their own risk and demand lower risk premium than the low CheiRank stocks. On the other hand, The slope of the PageRank factor increases as the PageRank increases. That is, investors demand compensation for the high PageRank firm for the vulnerability to propagated risk.

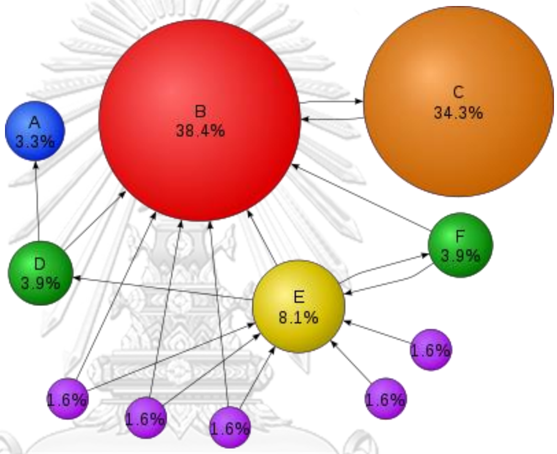
If the systematic importance means that a firm is close to central to the network, the paper's principal findings seem to conflict with the existing literature. Kaya (2014) and Buraschi and Porchia (2012) report the positive relationship between interconnectedness measures and returns, but this paper finds the opposite is true. This matter may be caused by the difference in methodology. It could be the different nature of developing and developed markets. Regardless of the reasons, the discrepancy between our results indicates that empirical research is essentially lacking to conclude the relationship between interconnectedness and equity returns. Considerable research effort is required to establish the relationship. The long-term goal for this field of research could be to find underlying economic state variables that produce variation in returns related to the interconnectedness measures.

4.6 Appendix

4.6.1 PageRank

PageRank is a link analysis algorithm and invented by Larry Page and Sergey Brin in 1996. In fact, Google web search engine uses PageRank as one of many factors to determine ranking websites. As illustrated by the picture below, PageRank is the probability that each webpage, called node, will be reached and thus the sum of all PageRank values is equal to 1.

Figure 4.4 The network diagram with nodes' sizes proportional to PageRank.



PageRank probability is calculated by the following equation. PR(u) is the probability that node u will be reached. Bu is the number of nodes that point to u. PR(v) is the probability that node v will be reached. L(v) is the number of outbound links of node v. The d damping factor is the probability at each page the "random surfer" will get bored and request another random page. In other words, people will get bored and bored at each step with the probability d, and eventually, they stop surfing or switch to a new random page. From previous studies, it is generally assumed that the d damping factor will be set at 0.85. Lastly, N is the number of nodes in the network.

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in Bu} \frac{PR(v)}{L(v)}$$

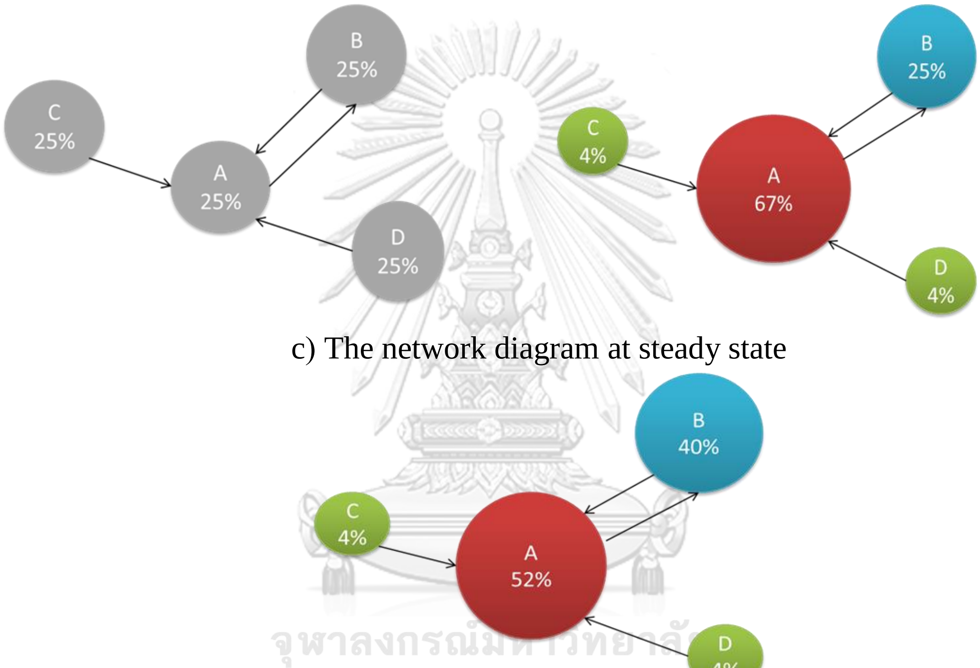
To illustrate, suppose that there are four nodes in the network, including A, B, C, and D as shown in the following diagram. The only links in the system are from B, C, D to A and A to B. Node A will have a PageRank of 0.67, given that the initial value of each node is 0.25. The first term on the RHS means that a random surfer moves to a different webpage by some means other than selecting a link in the existing webpage with the probability 1 – d and weight 1/N for each of webpages.

$$PR(A) = \frac{1 - 0.85}{4} + 0.85 * (\frac{PR(B)}{1} + \frac{PR(C)}{1} + \frac{PR(D)}{1})$$

Figure 4.5 Illustration of PageRank procedure.

a) The initial network diagram with the initial value of 0.25

b) The network diagram after running the first iteration of PageRank algorithm



In general, if d is between 0 and 1, the network will eventually converge to a fixpoint. To prove this statement, let's put the above PageRank equation into a matrix form:

$$\boldsymbol{R}(t+1) = \frac{1-d}{N}\mathbf{1} + d\boldsymbol{M}\boldsymbol{R}(t)$$

$$M_{ij} = \begin{cases} 1/L(p_j) & \text{if j links to i} \\ 0 & \text{otherwise} \end{cases}$$

The process will continue until **R(t+1)** – **R(t)** < Ɛ. Thus, the convergence point will be:

$$\boldsymbol{R} = \frac{1-d}{N}\mathbf{1} + d\boldsymbol{MR}$$

$$\boldsymbol{R} = \frac{1-d}{N}\mathbf{1} * (\boldsymbol{I} - d\boldsymbol{M})^{-\mathbf{1}}$$

The ***R*** matrix represents the final PageRank if the network converges.

As for the network in our example, it requires 10 iterations to reach the steady state, whereby **R'** = (0.52, 0.40, 0.04, 0.04). This steady state of the network is depicted in figure 2c. Nevertheless, in some cases, the network just fails to converge. For example, there is a dangling node that has no outbound links.

4.6.2 Power Method of PageRank

Another method to compute the PageRank is the power method which is originally used by Google. The advantage of this method is dangling node fix. Similar to the previous method, the power method begins with an adjacency matrix **H** of the network (PCPG network in this paper). The element in row i and column j of **H** is H_{ij} = $1/l_i$, whereby l_i is the number of outbound links from i. Otherwise, H_{ij} = 0.

$$H_{ij} = \begin{cases} 1/l_i & \text{if i links to j} \\ 0 & \text{otherwise} \end{cases}$$

But due to the problem of dangling, the system might fail to converge. So, the new matrix S is created to deal with this problem.

$$\mathbf{S = H + dw}$$

d = a vector contained 1 if the node is dangling and 0 if otherwise. **w** is usually a uniform row vector. The example of the matrix S is as follows.

$$
S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}
$$

$$
= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.
$$

In this case, node #4 is a dangling node because there is no outbound link from the node. Therefore, a uniform vector of equal weight is added to the original matrix **H**. The use of the uniform vector implies that all nodes have an equal probability to be reached.

The next step is to create the Google matrix **G** = d**S** + (1-d)**1v**. d is a damping factor as usual, 1 is the column vector of one, and v is a personalization vector, usually a uniform vector of 1/N. Since $\lambda = 1$ is not a repeated eigenvalue of **G** and is greater in magnitude than any other eigenvalue of **G**, we can quickly find the exact solution to the eigensystem, e**G** = e, whereby e is the eigenvector of **G** with the eigenvalue of 1.

Given the starting vector $e^{(0)} = (1/N)\mathbf{1}$, the power method calculates the iterates as follows:

$$e^{(k)} = e^{(k-1)}\mathbf{G}, \text{ where } k = 1,2,\ldots$$
$$e^{(k)} = e^{(0)}\mathbf{G}^{(k)}$$

The process continues until some convergence criterion is satisfied. The final left eigenvector e is called the PageRank vector.

In summary, the PageRank process begins by assigning an initial value to each node in the network, equal to 1/N. Then, the nodes' values are adjusted by the adjacency matrix of the PCPG network. The iteration continues until the steady state of the nodes' values is reached. From the viewpoint of economics, node A in figure 2b has the highest number of inbound links and thus is the most fragile node in the system. The intuition is that when the value of any node in the network drop, node A will have the highest chance to be affected. We can say that node A is more fragile than any other node.

Chapter 5 Conclusion

The effects of the network structure in the stock markets have been documented by researchers and practitioners, especially since the global financial crisis in 2007-2008. One of the prominent effects of the network is the ability to amplify idiosyncratic risk and cause a systemic event. This striking feature directly challenges the traditional asset pricing model that assumes no idiosyncratic risk in a well-diversified portfolio. From the network theory's perspective, the diversifying argument is not always valid as the idiosyncratic shock is not entirely diversified away in the network scheme and sometimes causes the system-wide event. Therefore, equity returns may demand compensation for the idiosyncratic risk and network factors in addition to the market risk.

This dissertation provides empirical evidence for the relationship between equity returns and network structure. In the US market, I observe the dynamic pattern of the network structure over time and capture the network topology by the diameter of the network. I find that the measure of network topology can predict the subsequent US market returns in both monthly and quarterly interval. When I include the measure of idiosyncratic risk in the predictive regression, the diameter becomes insignificant. Thus, the finding suggests that the network topology, measured by the diameter, can affect the stock market returns by serving as the propagation channel for the idiosyncratic risk, measured by average stock variance. Similarly, in the international equity markets, the individual network measures are weak to explain and predict the probability of the extreme negative returns for counties across the world. Rather, once the network measures interact with the idiosyncratic risk measure, the ability to predict the probability of the extreme negative returns increases significantly. Lastly, in Thailand, I examine the effect of CheiRank and PageRank on the cross-sectional stock returns. CheiRank reflects the systematical-important aspect of stock while PageRank reflects the fragility property of the stock. The Fama-MacBeth regressions indicate that both network measures can capture cross-sectional variation in the stock returns. Following Fama and French (1993), I create the portfolios sorted by CheiRank, PageRank, size, and book-to-market. The common risk factors are also estimated accordingly. I find that the network factors may be useful in explaining the time-series of expected portfolio returns and thus be considered potential common risk factors.

Although the network application is very advanced in some fields such as computer and internet, it is still early in financial economics. Theoretical models and empirical evidence are essentially lacking. This dissertation contributes to the field by providing empirical evidence for the relationship between network structure and equity returns. However, this paper leaves many open questions. First and foremost, since the data is limited to US, Thailand, and some international indices, the future research can reexamine the relationship in the context of other countries. Secondly, numerous studies, including this paper, use the correlation of returns as a link or relationship between members. However, none has established the connection between the correlation of returns and actual relationships such as trades and balance sheet. Last but not least, diameter, CheiRank, and PageRank only reflect some characteristics of the network structure. The other network measures with different information may provide more insights into the relationship between equity returns and network structure.

REFERENCES

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica, 80*(5), 1977-2016. doi:10.3982/ECTA9623

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic Risk and Stability in Financial Networks. *American Economic Review, 105*(2), 564-608. doi:10.1257/aer.20130456

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of financial economics, 63*(3), 443-494. doi:10.1016/s0304-405x(02)00068-5

Bali, T. G., Cakici, N., Yan, X. S., & Zhang, Z. H. E. (2005). Does Idiosyncratic Risk Really Matter? *The Journal of Finance, 60*(2), 905-929. doi:10.1111/j.1540-6261.2005.00750.x

Birch, J., Pantelous, A. A., & Soramäki, K. (2016). Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns. *Computational Economics, 47*(4), 501-525. doi:10.1007/s10614-015-9481-z

Bisias, D., Flood, M., Lo, A. W., & Valavanis, S. (2012). A Survey of Systemic Risk Analytics. *Annual Review of Financial Economics, 4*(1), 255-296. doi:10.1146/annurev-financial-110311-101754

Bonanno, G., Caldarelli, G., Lillo, F., Miccich�, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B - Condensed Matter, 38*(2), 363-371. doi:10.1140/epjb/e2004-00129-6

Boyer, B. H., Kumagai, T., & Yuan, K. (2006). How Do Crises Spread? Evidence from Accessible and Inaccessible Stock Indices. *The Journal of Finance, 61*(2), 957-1003. doi:10.1111/j.1540-6261.2006.00860.x

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*(1-7), 107-117. doi:10.1016/s0169-7552(98)00110-x

Buccheri, G., Marmi, S., & Mantegna, R. N. (2013). Evolution of Correlation Structure of Industrial Indices of US Equity Markets. *SSRN Electronic Journal*. doi:10.2139/ssrn.2282385

Buraschi, A., & Porchia, P. (2012). Dynamic Networks and Asset Pricing. *AFA 2013 San Diego Meetings Paper*. doi:10.2139/ssrn.2024483

Calvo-Armengol, A., & Zenou, Y. (2004). Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior*. *International Economic Review, 45*(3), 939-958. doi:10.1111/j.0020-6598.2004.00292.x

Chen, Y.-A. (2014). Network Risk and Cross-Section of Expected Stock Returns. *SSRN Electronic Journal*. doi:10.2139/ssrn.2572593

Cochrane, J. H. (1999). New Facts in Finance. *Economic Perspectives, Federal Reserve Bank of Chicago, 23*(3), 36-58.

Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics, 182*(1), 119-134. doi:10.1016/j.jeconom.2014.04.012

Drew, M. E., Marsden, A., & Veeraraghavan, M. (2007). Does Idiosyncratic Volatility Matter? New Zealand Evidence. *Review of Pacific Basin Financial Markets and Policies, 10*(03), 289-308. doi:10.1142/s0219091507001070

Dungey, M., Luciani, M., & Veredas, D. (2012). Googling SIFIs. *Technical report, University of Tasmania. Working Paper*. doi:10.2139/ssrn.2166504

Elliott, M., Golub, B., & Jackson, M. O. (2014). Financial Networks and Contagion. *American Economic Review, 104*(10), 3115-3153. doi:10.1257/aer.104.10.3115

Engle, R., & Kelly, B. (2012). Dynamic Equicorrelation. *Journal of Business & Economic Statistics, 30*(2), 212-228. doi:10.1080/07350015.2011.652048

Ermann, L., & Shepelyansky, D. L. (2011). Google Matrix of the World Trade Network. *Acta Physica Polonica A, 120*(6A), A-158-A-171. doi:10.12693/APhysPolA.120.A-158

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance, 47*(2), 427-465. doi:10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56. doi:10.1016/0304-405x(93)90023-5

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy, 81*(3), 607-636. doi:10.1086/260061

Forbes, K. J. (2012). Identifying Contagion. *NBER Working Paper, 18465*. doi:10.3386/w18465

Forbes, K. J., & Rigobon, R. (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements. *The Journal of Finance, 57*(5), 2223-2261. doi:10.1111/0022-1082.00494

Forbes, K. J., & Warnock, F. E. (2012). Capital flow waves: Surges, stops, flight, and retrenchment. *Journal of International Economics, 88*(2), 235-251. doi:10.1016/j.jinteco.2012.03.006

Goyal, A., & Santa-Clara, P. (2003). Idiosyncratic Risk Matters! *The Journal of Finance, 58*(3), 975-1007. doi:10.1111/1540-6261.00555

Haldane, A. G. (2013). Rethinking the financial network. In *Fragile Stabilität – stabile Fragilität* (pp. 243-278).

Jackson, M. O. (Writer) & J. E. Rauch (Director). (2007). The Study of Social Networks In Economics. In J. E. Rauch (Producer), *The Missing Links: Formation and Decay of Economic Networks*. New York: Russell Sage Foundation.

Jackson, M. O. (2008). *Social and Economic Networks*. New Jersey: Princeton University Press.

Jackson, M. O. (2010). Research Opportunities in the Study of Social and Economic Networks. *National Science Foundation/Directorate for Social, Behavior & Economic Sciences (NSF/SBE)*.

Jackson, M. O., & Rogers, B. W. (2007). Meeting Strangers and Friends of Friends: How Random Are Social Networks? *American Economic Review, 97*(3), 890-915. doi:10.1257/aer.97.3.890

Jegadeesh, N., & Subrahmanyam, A. (1993). Liquidity Effects of the Introduction of the S&P 500 Index Futures Contract on the Underlying Stocks. *The Journal of Business, 66*(2), 171-187. doi:10.1086/296600

Kaya, H. (2014). Eccentricity in asset management. *The Journal of Network Theory in Finance, 1*(1), 1-32. doi:10.21314/jntf.2015.003

Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., & Ben-Jacob, E. (2010). Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *PLoS ONE, 5*(12). doi:10.1371/journal.pone.0015032

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B, 11*(1), 193-197. doi:10.1007/s100510050929

Masson, P. R. (1999). Multiple Equilibria, Contagion, and the Emerging Market Crises. *Working Paper of the International Monetary Fund, 99*(164).

McCracken, M. W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics, 140*(2), 719-752. doi:10.1016/j.jeconom.2006.07.020

Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis *The American Economic Review, 81*(5), 1408-1418.

Nanisetty, P., Bharati, R., & Gupta, M. (1996). The effect of time-varying covariances on asset risk premia. *Review of Quantitative Finance and Accounting, 7*(2), 205-220. doi:10.1007/bf00243979

Onnela, J. P., Chakraborti, A., Kaski, K., Kertész, J., & Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E, 68*(5). doi:10.1103/PhysRevE.68.056110

Pollet, J. M., & Wilson, M. (2010). Average correlation and stock market returns. *Journal of financial economics, 96*(3), 364-380. doi:10.1016/j.jfineco.2010.02.011

Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences, 102*(30), 10421-10426. doi:10.1073/pnas.0500298102

Vidal-García, J., Vidal, M., & Nguyen, D. K. (2016). Do liquidity and idiosyncratic risk matter? Evidence from the European mutual fund market. *Review of Quantitative Finance and Accounting, 47*(2), 213-247. doi:10.1007/s11156-014-0488-7

APPENDIX

VITA

Harnchai Eng-Uthaiwat is a student in the Ph.D. program in Economics, Faculty of Economics, Chulalongkorn University, Thailand. Harnchai's research interests include networking, financial markets, financial crisis, systemic risk, and asset pricing. He particularly specializes in networking of financial markets, which he formalizes in his Ph.D. dissertation.

Born in Thailand, Harnchai attended and graduated from Chulalongkorn University with an undergraduate degree in Computer Engineering in 2005. Afterwards, he received the master degree in business administration (MBA) from the University of Wisconsin-La Crosse in 2008. Before attending the Ph.D. program, he worked for a leading banking company as an assistant relationship manager who evaluated the clients' credit risk profiles and serviced related business activities. Then, he had an opportunity to work for a research firm in Thailand which provides marketing research service to clients in various fields such as agriculture, energy, real-estate, banking, and government sectors. The research life inspired Harnchai to progress his study in the higher learning and attend the Ph.D. program in Economics at Chulalongkorn University from 2012 to present.

In his thesis, Harnchai presented new empirical evidence for the relationship between network structure and equity returns in US, international markets, and Thailand. This work contributed to his first publication in Review of Quantitative Finance and Accounting. In 2017, he was also a recipient of a Best Paper Award at the conference of Society of Interdisciplinary Business Research in recognition of outstanding contribution to interdisciplinary research.