Financial Conduct Authority
University College London

Influencers in Dynamic Financial Networks

Isobel Seabrook

First supervisor: Fabio Caccioli
Second supervisor: Paolo Barucca

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Computer Science

Declaration

I, Isobel Emily Alice Seabrook confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Any views expressed in this thesis are solely those of mine and so cannot be taken to represent those of the Financial Conduct Authority.

Impact statement

This research has been approached from the perspective of developing techniques for application in financial conduct regulation, but it has application also within academia and industry. Inside academia, my work has an impact on the research area of dynamic financial networks, providing insights into the use of network spectra to find dynamically important nodes and edges in networks, along with proposing potential generative models for dynamic networks, which is an active area of research. I find that structural importance relates to the dynamics of financial networks, which is a useful insight complementing research exploring the dynamics and stability of financial networks.

While my thesis focuses on financial networks, the methods I have developed are not limited to finance, meaning that my work provides a contribution to the field of network science in general. In particular, the methods I put forward are useful for the identification of important nodes and edges in any networked system, and the generative models I developed offer the potential for application to any system where bursty patterns are present.

Outside academia, my research impacts predominantly in financial regulation, where my methods could be used to help identify and reduce consumer harm in the financial markets. For instance, the generative models for transaction sequences I propose could be further developed for use as null models in anomaly detection systems to help flag instances of mis-reporting and market abuse. My methods to measure the structural importance of nodes and edges could be used to target a supervisory response to the market participants who pose the greatest risk to the overall system. My research also has the potential to impact on policy design, since my methods could be used to help target new policies to the most impactful nodes in a financial network, or to riskier financial instruments. Further to this, my methods for analysis of complex regulatory datasets will benefit data practitioners working in financial regulation. Although I have considered applications specifically relevant to the UK financial conduct regulator, the Financial Conduct Authority, this research has the potential to benefit other national and international regulators with similar remits. For example, the Bank of England has published

several research papers looking at incorporating network analytics [1, 2, 3]. In addition, the Bank of International Settlements and the Monetary Authority Singapore have recently launched project ellipse, which has made use of network-based techniques to demonstrate how exposures could be mapped, indicating possible systemic risks to the banking system [4].

The benefits of my research will be realised firstly through engagement with data practitioners at the Financial Conduct Authority to establish the route to application for my methods. This will involve educating practitioners and supervisors alike to help them understand the benefits and insights that can be gained from my methods. I hope this will also involve future collaboration with academics to further develop and test advanced methods for application to financial regulation, opening up significant opportunities to benefit society as a whole.

Abstract

To monitor risk in temporal financial networks, an understanding of how individual behaviours affect the temporal evolution of networks is needed. This is typically achieved using centrality and importance metrics, which rank nodes in terms of their position in the network. This approach works well for static networks, that do not change over time, but does not consider the dynamics of the network. In addition to this, current methods are often unable to capture the complex, often sparse and disconnected structures of financial transaction networks. This thesis addresses these gaps by considering importance from a dynamical perspective, first by using spectral perturbations to derive measures of importance for nodes and edges, then adapting these methods to incorporate a structural awareness. I complement these methods with a generative model for transaction networks that captures how individual behaviours give rise to the key properties of these networks, offering new methods to add to the regulatory toolkit. My contributions are made across three studies which complement each other in their findings.

Study 1:

- I define a structural importance metric for the edges of a network, based on perturbing the adjacency matrix and observing the resultant change in its largest eigenvalues.
- I combine this with a model of network evolution where this metric controls the scale and probabilities of subsequent edge changes. This allows me to consider how edge importance relates to subsequent edge behaviour.
- I use this model alongside an exercise to predict subsequent change from edge importance. Using this I demonstrate how the model parameters are related to the capability of predicting whether an edge will change from its importance.

Study 2:

- I extend my measure of edge importance to measure the importance of nodes, and to capture complex community structures through the use of additional components of the eigenspectrum.
- While computed from a static network, my measure of node importance outperforms other centrality measures as a predictor of nodes subsequently transacting. This implies that static representations of temporal networks can contain information about their dynamics.

Study 3:

- I contrast the snapshot based methods used in the first two studies by modelling the dynamic of transactions between counterparties using both univariate and multivariate Hawkes processes, which capture the non-linear 'bursty' behaviour of transaction sequences.
- I find that the frequency of transactions between counterparties increases the likelihood of them to transact in the future, and that univariate and multivariate Hawkes processes show promise as generative models for transaction sequences.
- Hawkes processes also perform well when used to model buys and sells through a central clearing counterparty when considered as a bivariate process, but not as well when these are modelled as individual univariate processes. This indicates that mutual excitation between buys and sells is present in these markets.

The observations presented in this thesis provide new insights into the behaviour of equities markets, which until now have mainly been studied via price information. The metrics I propose offer a new potential to identify important traders and transactions in complex trading networks. The models I propose provide a null model over which a user could detect outlying transactions and could also be used to generate synthetic data for sharing purposes.

Contents

Chapter 1

Introduction

1.1 Motivation

The financial markets are one of the most notorious examples of a complex system. Although there is no concise definition of a complex system, they tend to consist of multiple elements interacting in an apparently disordered way, out of which is generated a robust order [5]. Financial markets influence, and are influenced by a myriad of endogenous and exogenous effects, and they are hugely important in supporting global society. In particular, stock trading plays an important role in the economy, influencing and responding to external factors and providing investors the opportunity to have a share in the profits of publicly-traded companies.

Because of their significance, financial markets are tightly monitored and regulated, both through qualitative means as well as quantitative monitoring of a vast array of different data streams. They are regulated from a prudential perspective to ensure that markets function well in terms of both integrity and competition, and to reduce the risks of impacts being passed through the system to the detriment of consumers. They are also regulated from a conduct perspective, to avoid the harms of market manipulation, market abuse and money laundering. In the UK, the Financial Conduct Authority is the regulatory body responsible for supervision, enforcement and development of policy in relation to the conduct of financial firms, encapsulated in their three operational objectives [6]:

- To secure an appropriate degree of protection for consumers.
- To protect and enhance the integrity of the UK financial system.
- To promote effective competition in the interests of consumers.

To meet these objectives effectively, it is important that regulators and policy makers understand how these systems function, and have the tools they need to model and monitor behaviours. One way to represent these systems is through the use of network theory to study the interactions between market participants [7, 8], which is the focus of this thesis. I study stock trading systems as dynamic networks to understand influential relationships, entities and processes in the context of financial conduct regulation, with a particular focus on transaction reports for stock trading. I am making a strong assumption in modelling financial systems as dynamic networks, as in reality the behaviours of these systems are influenced by much more than just the relationships between participants. Despite this, in this thesis I am able to demonstrate the power of dynamic networks in delivering significant insights into the behaviours of these systems. Although my methods have not yet been applied in a regulatory context, this thesis lays down significant groundwork to add network theory to the regulatory toolkit.

This thesis focuses on the development of methods to tackle two challenges in data driven financial regulation. First of all, regulating and developing policies for financial markets requires the ability to identify important players and to understand how their actions affect other market participants and the evolution of the system as a whole. My research explores this both conceptually through the development of models that relate structural importance to network dynamics, and analytically through application to a range of different datasets, including transactions conducted on UK equities markets. Secondly, models which are able to generate realistic transaction sequences are of great interest, due to the usage and sharing constraints for market sensitive data. They are also of interest because of their utility in detecting outliers. Finally, they have the potential to deliver insights into the link between microscopic observations obtained from transaction reports showing an individual market participant's trading activity, and macroscopic market evolution and structural change. Although attempts have been made to infer the generative processes from observed prices of financial instruments, there is a lack of methods to generate the transaction sequences themselves, which I explore in my research in chapter 5.

1.2 Research Objectives

Given these overarching focus topics for this research, my objectives are:

- To provide a means of quantifying the structural importance of nodes and edges in a network.
- To understand how the importance of edges or nodes in transaction networks relates to the global network structure and how this relates to network evolution.
- To develop a generative model for transaction networks that is able to reproduce the key properties of these networks and be robust to application over different aggregation scales.
- To use these methods to uncover insights from datasets consisting of the transactions executed on the UK equities markets.

1.3 Research experiments & findings

The above objectives are addressed through experimental studies which have been published during the course of development of this thesis. Note that to address the first objective, I make a distinction between methods at the level of edges in chapter 3 and at the level of nodes in chapter 4, resulting in three experimental studies:

1. **Evaluation of edge importance**: In chapter 3 I will introduce a concept that sits at the core of this thesis, in the use of spectral derivatives to assess the structural influence of edges in a network. I will further explore the relationship between spectral edge importance and subsequent evolution of the network. Specifically, I will propose a Markov model for temporal network evolution, parameterised by the extent to which spectral edge importance indicates resultant changes. Through this I will demonstrate that element-wise derivatives of the network eigenspectrum provide a useful indicator of structural importance and will show that in some cases this provides an indication of subsequent edge changes. The published version of this chapter can be found at [9].
2. **Evaluation of node importance**: Building on the ideas in my first study, In chapter 4 I will present a measure for spectral node importance, to understand who the key players are in terms

of importance. I will show how multiple components of the eigenspectrum can be used to capture importance of nodes which accounts for community structure, and will demonstrate how the importance of nodes when defined in this way is predictive of subsequent node presence. The published version of this chapter can be found at [10].

3. **Hawkes processes for modelling the bursty behaviour of financial networks**: In chapter 5 I will first of all demonstrate key temporal and cross-sectional properties of financial transaction networks, before proposing a generative model which makes use of the Hawkes process – a temporal point process in which the occurrence of an event increases the probability of occurrence of another event – to generate timestamps with properties observed in the real transaction networks. I will consider several different formulations of Hawkes processes and will also introduce methods to select edges to transact at each of the generated timestamps according to their historical importance where this is required. I will demonstrate that multivariate processes show some promise as generative systems for stock network systems as a whole as they are able to reproduce some key cross-sectional properties observed. I will also demonstrate that bivariate models show strong performance in reproducing the temporal behaviour of transaction sequences of buys and sells occuring through a single hub organisation. This chapter is currently in review for publication, and the pre-print can be found at [11].

During the course of the development of this thesis, I have also considered additional avenues of research which, although I do not cover in depth in this thesis, also contribute to the understanding of stock markets and their impact and response to shocks experienced by groups of stocks [12].

1.4 Scientific contributions

This thesis consists of original research and presents contributions to science through demonstrating that network eigenvalue derivatives can be used as a fundamental measure of structural importance in networks, and that this can be applied across multiple applications of network science. It further demonstrates how this relates to predictability and stability in the context of temporal financial networks, with the hope that these methods can be used practically by policy makers in understanding market stability. It contributes to existing literature by enhancing the understanding of the effect of aggregation scales on predictability in financial networks. It develops a generative model which is able to reproduce the real bursty behaviours which are seen in high frequency traded financial instruments. It provides insights into the applicability of techniques from network science that have not been applied in the finance industry. These will help to build the tool-kit of researchers in collaboration with industry experts when analysing regulatory data. On top of this, it presents observations of the key properties and behaviours of equity transactions that have previously been unavailable for study.

1.5 Thesis structure

This thesis first presents the contextual background that motivates the research and related literature explored in the process of developing the methods. It then presents the research experiments outlined in section 1.3. I start chapter 3 by delivering the core concept of this thesis, defining a metric for edge importance measuring the impact on market structure of financial market participants and their actions. I then link this to network dynamics throughout the thesis by considering complementary approaches that can be separated into discriminative modelling and generative modelling [13]. The first of these approaches separates observed data into distinct classes, often referred to as conditional modelling. The second models the data distribution itself rather than the decision boundary, and can be used to generate

new data points. I initially start by exploring time series of static snapshots of financial networks using a discriminative approach, exploring the prediction of changes between successive snapshots. This is followed by the proposal of generative models for the sequences of transactions themselves, which complements the discriminative approaches and allows for more granular properties of the systems to be observed. My conclusions are then drawn considering insights from all of the research experiments conducted, before presenting a summary of the contributions and suggestions of future research to build on these. Figure 1.1 shows the thesis structure from this point onward.

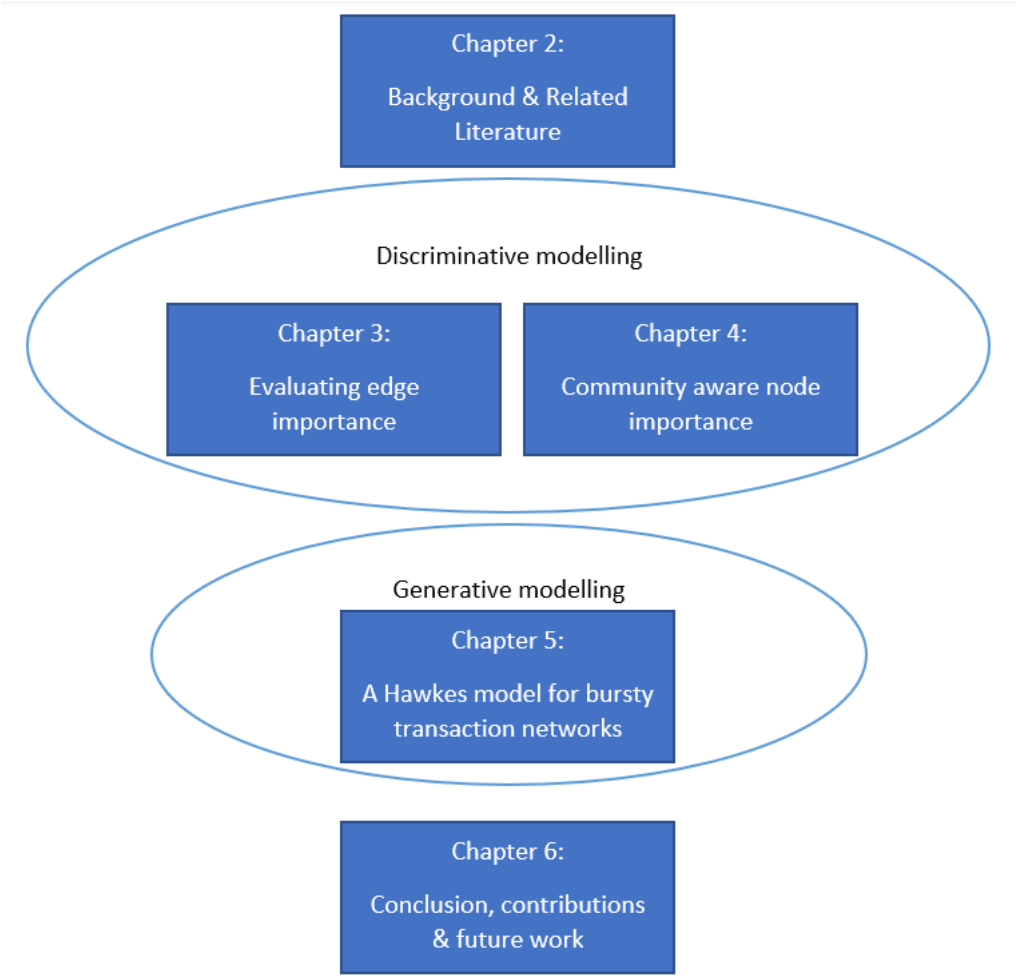


Figure 1.1: Thesis structure

Chapter 2

Background & Related Literature

This section covers the core theory and methods available for the study of transaction networks, in particular the methods that are chosen for use in this thesis. I will first discuss the various methods that exist for the analysis of financial markets, before focusing on the specific methods used in this thesis. I highlight throughout this section where these methods will be applied in further chapters, and the reasons that certain methods are not explored further. I then continue to a thorough review of related literature, in which I discuss methods and applications presented by other researchers, whilst highlighting how my research contrasts and complements the existing literature.

2.1 Background

As noted in the introduction, there are a number of different approaches to the analysis of financial market behaviour. Methods are motivated by the potential benefits they offer, focusing typically and historically on the development of tools for the individual investor. There have also been more recent developments of tools for regulation and monitoring of markets in terms of their stability and dynamics, to help minimise the risks of large scale market turmoil and financial crises.

Considering the individual investor, researchers have predominantly focused on the statistical analysis, modelling and prediction of asset prices and returns. There has been a particular focus on risk management [14, 15, 16], notably on the prediction of loan default [17]. Within this, there has been a significant focus of efforts in the peer-to-peer lending space, driven by the availability of data and the credit risk that individual investors are subjected to [18, 19, 20, 21]. Other researchers have focused on the pricing of financial instruments [22, 23, 24], for example those considering options pricing [25, 26]. In addition to the focus on analysis of prices and returns, some researchers have considered analyses at a lower level of granularity in an attempt to understand the behaviours which shape the financial markets [26, 27, 28, 29, 30, 31]. Complementing these are studies which consider the generative processes of financial markets [32, 33, 34, 35], which offer huge potential for experimentation with data that would otherwise be unavailable. Finally, a number of researchers have focused their efforts on portfolio optimisation [36, 37, 38], due to the financial benefits that can be gained from a high performing investment portfolio.

From the perspective of regulation and monitoring of financial markets, the 2008 crisis prompted new research considering the inter-connectedness of the system, and understanding the mechanisms that can lead to the breakdown of this system [39, 40]. Methods from risk management such as stress testing have been reconsidered from the perspective of the financial network [41]. Further to this, techniques ranging from physics [7] to ecology [42, 43] have been used to explore financial stability [39], and recent

developments in machine learning have been exploited to unearth non-linear relationships from more challenging regulatory datasets [44, 45, 46]. My research has been directed and informed by progress in these areas, however it differs in that I approach things from a conduct perspective, which motivates my aims presented in the introduction.

2.2 Methods

This thesis at its core is a study of the application of network theory to financial systems. I chose this as my focus due to the potential of network-based approaches to provide insights into how financial systems behave, and how an individual can have an impact on the system. A network is simply a collection of points, referred to as nodes, joined together with lines, known as edges. An edge in a network can either carry a value referred to as its weight, or the network can be unweighted. Mathematically, a network can be described by a matrix known as the adjacency matrix, in which the nth row, m^{th} column entry of the matrix contains the weight of the edge between the nodes n and m, or a 1 in the presence of an edge if the network is unweighted. In the case of an undirected network, the edges in a network do not have a direction and the adjacency matrix is symmetric. When the network is directed and the edges have a direction, the adjacency matrix is asymmetric and the $(n, m)^{th}$ entry may not equal the $(m, n)^{th}$ entry.

As an example, consider the simple undirected network presented in figure 2.1. The adjacency matrix for this network is

$$\begin{pmatrix} 0 & 0.6 & 0.2 & 0.3 & 0 & 0 \\ 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0.1 & 0.7 & 0.9 \\ 0.3 & 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0 & 0 \end{pmatrix} \tag{2.1}$$

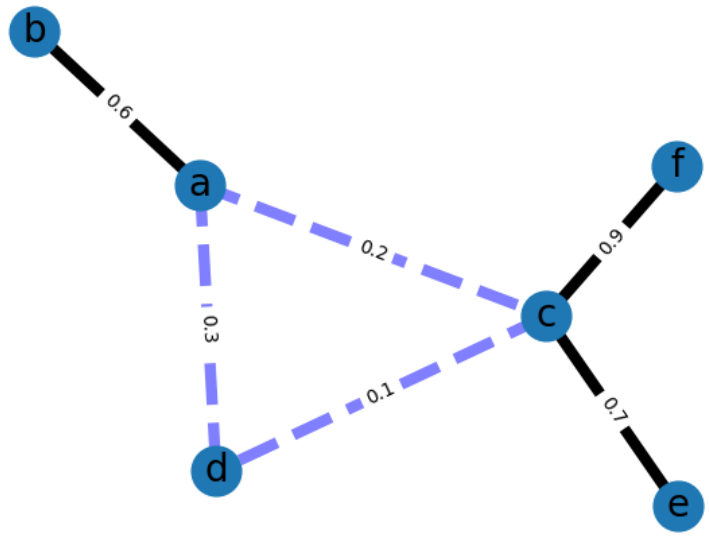


Figure 2.1: Simple weighted, undirected network, image found at [47]

2.2.1 Static networks

There exist a number of ways to quantify the properties networks exhibit. Here I start with the most local measures that describe individual nodes, and I present successively more complex measures which eventually capture the global behaviour of the network. In particular, I present centrality based methods, which are closely related to the methods I develop to address my first aim of this research, to understand how importance relates to network structure and evolution.

Degree

The simplest property describing nodes in networks is the degree. For an undirected network, the degree k_j of node j is the number of edges attached to it [48],

$$k_j = \sum_i A_{ij}. \tag{2.2}$$

Centrality

The next set of measures to note are centrality measures, which rank nodes according to their position in the network. These are of particular relevance to this thesis, as nodes that are more central in networks have a larger potential to impact neighbouring nodes and often play a more influential role in the evolution of the network. One of the most commonly used measures of centrality is eigenvector centrality, which scores nodes based on the concept that having well connected neighbours makes a node more influential. It does this by giving each node a score proportional to the sum of the scores of its neighbours[48]. Eigenvector centrality can be calculated as follows from the adjacency matrix. For node i in a network with adjacency matrix $\mathbf{A}$, the eigenvector centrality is defined to be proportional to the sum of the centralities of i's neighbours:

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j. \tag{2.3}$$

This can be rearranged and written in vector notation as the eigenvector equation:

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}, \tag{2.4}$$

with the eigenvector centrality being the eigenvector associated with the largest eigenvalue of the adjacency matrix. Other commonly used measures of node centrality include closeness centrality, which is calculated as the average shortest path from the node in question to every other node in the network [49], and betweenness centrality, which captures how much a given node is in between others [50]. Betweenness centrality is also the most commonly used edge centrality measure, with edge betweenness centrality defined as the number of the shortest paths that go through an edge [51].

In this thesis, I define static measures of node and edge importance by considering the change to the network spectrum between successive snapshots. Although my measures of importance are static measures, I find that they have some predictive power in predicting the subsequent behaviour, which is a property not demonstrated by other static measures. This finding is significant since it provides a link between static and dynamic representations of networks.

Rich club index

The rich club phenomenon in networks is characterised when the hub nodes with high degrees are on average more intensely connected than the nodes with smaller degree [52]. This can be measured using the rich club coefficient at each degree k, which is the fraction of of the actual and potential number of edges among the set of nodes with degree higher than k:

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}, \tag{2.5}$$

where $N_{>k}$ is the number of nodes with degree larger than k and $E_{>k}$ is the number of edges among those nodes.

Assortativity

Assortative mixing in networks is the tendency for nodes to connect to other nodes that are like them [53]. In this thesis I consider degree assortative mixing, which is a network level property quantified in terms of the quantity e_{ij}, which is the fraction of edges in a network that connect a vertex of degree i to one of degree j. The assortativity coefficient then quantifies the level of assortative mixing, given by:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{\mathrm{Tr}(\mathbf{e}) - ||\mathbf{e}^2||}{1 - ||\mathbf{e}^2||}, \tag{2.6}$$

where $\mathbf{e}$ is the matrix whose elements are e_{ij}, $a_i = \sum_j e_{ij}$ and $b_i = \sum_j e_{ji}$ are the fraction of each type of end of an edge that is attached to vertices of type i, and $||\mathbf{x}||$ means the sum of all elements of the matrix $\mathbf{x}$ [54].

Assortativity is a useful quantity to explore to help understand how nodes having different 'roles' in networks might relate to the overall evolution and stability of the system. In particular, it has been shown that disassortative mixing enhances the stability of banking networks [55]. Later on in this thesis in chapter 4, I observe disassortative mixing in transaction networks of equity stocks, and also find that nodes with a larger value of structural influence are less likely to subsequently transact which is suggestive of an enhanced stability.

Reciprocity

Reciprocity is defined for a directed network as the ratio of the number of edges pointing in both directions to the total number of edges in the network G

$$r = \frac{|(u,v) \in G||(v,u) \in G|}{|(u,v) \in G|}, \tag{2.7}$$

similarly, for a single node, reciprocity is the ratio of the number of edges in both directions to the total number of edges attached to the node in question [56].

In application to address the aims of this thesis, reciprocity is a useful quantity to explore across time since it indicates how prevalent reciprocal relationships are in the networks considered. This helps to provide insights into the role of a historical relationship between two participants increasing the likelihood of them to transact. This is of particular relevance in chapter 5, in which I observe strong performance of a generative model for transaction sequences parameterised by the historical prevalence of trades between counterparties. This highlights the role of historical relationships between participants in trading decisions.

2.2.2 Temporal networks

Traditionally, network analytics has focused on static representations of networks, either looking at single snapshots in time, or considering a projection of the time dimension onto a static view by aggregating the links in a time window. In doing so, some, or all, of the temporal information about the network is lost.

However, recently, there have been developments in modelling systems as temporal networks, which are of particular use to address the aims of this thesis since they help capture the evolution of the system, and they offer opportunities to gain an understanding into the generative processes present. A common approach is to represent the system with a contact sequence (i, j, t), where i and j constitute the vertex set V at time t. This representation also allows for edges which take time to traverse, or contracts completed after a duration δt by representing the contact sequence as $(i, j, t, \delta t)$ [57]. Since I am considering systems where the transactions are instantaneous I am not interested in transmission time for edges, and since I am considering applications where time is discretised, I formally define a temporal network $G_t^w(t_{min}, t_{max})$ as in [58] as the ordered sequence of networks

$$G_{t_{min}}, G_{t_{min}+w}, ..., G_{t_{max}}, \tag{2.8}$$

where w is the size of the time aggregation (e.g. daily). Element A_{ij}^s of the adjacency matrix between nodes i, j at time s is non-zero if and only if there exists a link between i and j in G_t, $t \leq s \leq t + w$.

As in the static case, temporal node, community and network level measures can provide useful insights into the behaviour of the system. Several static measures have received recent attention to extend for use in temporal networks, for example the extension to the static rich club metric presented in Pedreschi et. al. [59], which captures the tendency of well-connected nodes to form simultaneous and stable structures in a temporal network. The measure quantifies whether static rich club patterns correspond to a structure that actually existed at some instant. Many of the centrality measures from static network analytics are also trivial to generalise for temporal networks. For example, for the calculation of closeness centrality, the degree is simply replaced with the latency [57], which is defined to be the shortest time within which i can reach j and gives an indication of how fast information can spread across a network [60].

An example making use of temporal centrality metrics in the literature is found in Tang et. al. [61], who consider extensions to centrality and efficiency metrics based on temporal shortest paths. They show that the temporal metrics are able to provide information about the structure of time varying networks that classical methods are unable to capture [62]. However, for my research, the concept of latency is less meaningful because the transactions I analyse are sparse and there often is not a path between two nodes.

Other methods attempt to generalise spectral measures of centrality like eigenvector centrality, and a number of similar measures, for which the network would need to be represented as a tensor [63], making the calculation of these measures non-trivial. In this thesis, instead of considering fully temporal spectral measures of importance, I instead propose static measures and show that these can provide an indication of subsequent network activity, to address the first aim of this thesis. My approach results in interpretable methods which are usable in financial regulation. I also consider the underlying generative processes of transaction sequences that give rise to the properties I observe in aggregate and in network snapshots, the theory of which will be outlined in the next section.

When representing a dynamic network as static snapshots across time, to capture the dynamic behaviour a probabilistic approach can be taken. It is common to consider the probability for the network to move between snapshots in an vector autoregressive sense, i.e. the future network configuration being dependent on the past network configuration. When doing so, several researchers make the Markov

assumption [64, 26], or in other words for a system X which evolves in time t, the state of the system depends on the previous state only, meaning that the process has no memory and the probability P of the observed network at time t_n obeys:

$$P(X^{t_n}|X^{t_{n-1}}, ..., X^{t_0}) = P(X^{t_n}|X^{t_{n-1}}). \tag{2.9}$$

As well as being dependent on the past network configuration, attributes of nodes (and edges) can have an influence over the state of the subsequent snapshot. This approach is often termed a 'fitness model' where each node has a fitness which controls its propensity to subsequently form edges [65]. Fitness models have most commonly been considered in a static sense to explore the network formation process, however can also be considered in a dynamic sense as demonstrated by Mazzarisi et. al. [66], who combine the concept of Markovian persistence with time evolving node fitnesses in a hidden Markov chain formulation for the temporal network:

$$\begin{cases} \mathcal{P}(\mathbf{\Theta}^t|\mathbf{\Theta}^{t-1}, \phi) = h(\mathbf{\Theta}^t, \mathbf{\Theta}^{t-1}, \phi) \\ \mathcal{P}(\mathbf{A}^t|\mathbf{A}^{t-1}, \mathbf{\Theta}^t, \beta) = g(\mathbf{A}^t, \mathbf{A}^{t-1}, \mathbf{\Theta}^t, \beta). \end{cases} \tag{2.10}$$

Here the node fitnesses $\mathbf{\Theta}$ themselves have a dynamics determined by a one-step transition probability $h(.)$ and $g(.)$ represents the likelihood for the network snapshot at time t given the information about the previous network snapshot, as well as the fitnesses. θ and ϕ represent static parameters. Other researchers have considered how static models of network formation such as the Watts-Strogatz, Exponential Random Graph and Stochastic Block Models extend to dynamic networks (see section 'Longitudinal modelling' in the review presented by Lee et. al. [67], alongside examples in [68, 69, 70, 71]). The approach taken in this thesis in chapter 3 is similar to a dynamic fitness model, however instead of considering a node level fitness variable, I instead control the propensity for changes with an edge level quantity. This helps me to address my aim of understanding how importance relates to network evolution at the most granular level.

In this thesis, I consider dynamic networks as both snapshot networks and as temporal point processes. Up to this point, the majority of the methods discussed have been focused on snapshot networks, which do not allow the true, complex dynamics of the system to be assessed. In this next section, I consider the use of temporal point processes to observe the dynamic behaviour of transactions.

2.2.3 Temporal point processes

Temporal point processes (TPPs) are probabilistic generative models for continuous-time event sequences [72]. TPPs are key to understanding processes in many applications, including earthquake modelling, crime analysis, infectious disease diffusion forecasting and transaction modelling [73].

Temporal point processes have been used extensively in modelling financial transactions, for example by Barcy et. al. [74], who model arrival times of buy and sell orders. The estimation of temporal point processes is well established, both following a parametric approach as in my research and also non-parametrically, using for instance the approach recently proposed in Dalmasso et. al. [75], which makes use of a Variational Autoencoder, applying convolutions to transaction times across interacting sellers and buyers and making use of a recurrent model to encode the sequential structure. Whilst their approach is able to capture some properties of the real transaction network, they were unable to reproduce the overall degree distributions. Also, black box approaches like this are unable to draw any further insights from the model parameters or performance. In contrast, I make use of fully parametric methods in this thesis as these allow interpretation of the parameter values themselves in providing information about the underlying behaviours of transaction sequences.

In this thesis, I use TPPs to address my aim of developing generative models for transaction networks, since they do not require aggregation as is the case with snapshot networks, which can result in behavioural information being lost. Here, I focus on two means of generating temporal sequences: the Poisson process and the Hawkes process.

Poisson processes

The Poisson process is one of the most widely used point process models, used in scenarios where the occurrence of events appears to have a constant rate but is stochastic. Formally, the Poisson process is a renewal process in which the inter-arrival intervals have an exponential distribution, with a density for each X_i of

$$f_X(x) = \lambda e^{-\lambda x}, \tag{2.11}$$

for $x > 0$ [76]. Here λ represents the mean number of events occurring in a given time interval. A key feature of the Poisson process is that it is memoryless, which is a disadvantage when considering financial transactions, in which the market responds to activity.

The Poisson process has been widely applied across finance [77, 78, 79] and other applications, and in combination with network methods [80, 81, 82, 83, 84]. They are used in scenarios where events occur at a constant rate. However, Poisson processes do not allow for any memory within the process and are not able to reproduce many of the properties I observe in this thesis for transaction sequences, such as burstiness - the tendency for transactions to arrive in bursts of activity. Hawkes processes as an alternative offer additional flexibility to capture these properties as they are based on a counting process in which the intensity function depends on all previously occurred events.

Hawkes processes

Hawkes processes [85] are point processes defined by the following conditional intensity function:

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{N} \sum_{t_k^j < t} \phi_{ij}(t - t_k^j), \tag{2.12}$$

where N is the number of dimensions of the process, μ_i is the background rate of the dimension i, sometimes referred to as the baseline, t_k^j are the timestamps of all events of dimension j and ϕ_{ij} are the excitation kernels which capture the extent and mechanism by which events excite future events. Here I have specified a full multidimensional excitation function, which allows for both self-excitation of events within a single dimension i of a multidimensional Hawkes process, and mutual excitation, in which events in one dimension i excite events in another dimension j. The presence of excitation in Hawkes processes has been shown to give rise to burstiness in the generated point process [86].

In the literature, the Hawkes kernels usually take the form of an exponential or power law [87]. The question of which kernel to choose has previously been an active topic of discussion, starting with Filimonov and Sornette [88] finding a good fit of the Hawkes exponential kernel for mid-price changes of E-mini S&P 500 futures contracts, which was later challenged by Hardiman et. al. [89]. Their challenge builds on the observations of Bacry et. al. [90], who introduce a non-parametric approach to estimate the Hawkes kernel finding power-law kernels for Bund and Dax futures. Hardiman et. al. also find that power law kernels are most suited when using Hawkes processes to model the same price changes considered by Filimonov and Sornette, in particular challenging their observation of endogeneity in the markets increasing with time. Their work highlights the importance of the choice of kernel for

studying the closeness of markets to criticality, since sub-criticality can be observed through choosing an exponential kernel to fit a critical, power law Hawkes Process. Although these discussions indicate that the best choice of kernel for financial datasets is likely to be the power law kernel, In chapter 5, I consider exponential excitation functions, since they are practically more suitable with significantly lower time complexity for estimation via a maximum likelihood approach ($\mathcal{O}(N)$ as opposed to $\mathcal{O}(N^2)$ for power law kernels [74]). I find that the exponential kernel performs well for my generative models, but highlight here that exploring the use of power law kernels, in particular improving run-times of available estimation methods for these would be a worthwhile direction for future work.

Specifically, the kernel used in this study ϕ_{ij} takes the form

$$\phi_{ij}(t) = \alpha^{ij}\beta e^{-\beta t}1_{t>0}, \tag{2.13}$$

where α^{ij} is the kernel intensity governing the extent to which excitatory behaviour dominates over the background process and β is the kernel decay governing the timescale over which a transaction influences future transactions. The baseline parameter μ_i can either be constant or can be time varying in nature. While estimation of the process in the case of a time varying baseline has been tackled by Chen et. al. [91], who take a non-parametric approach to estimate the process, for simplicity and following observations of stationarity in the properties of the process, I opt to consider a constant baseline when making use of Hawkes processes in chapter 5 of this thesis.

When using the Hawkes process as a univariate model for a single transaction sequence, $N = 1$ and the Hawkes process reduces to

$$\lambda(t) = \mu + \sum_{t_k<t} \phi(t - t_k). \tag{2.14}$$

Written as a probabilistic model which defines the probability of a transaction at time t given the history of the process up to time t, $\mathcal{H}_t$,

$$\mathcal{P}(N(t+h) - N(t) = 1|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)h + o(h), \tag{2.15}$$

where $N(t)$ is the counting process, or in other words the cumulative count of transactions, and h is an infinitesimally small time interval [92]. The history $\mathcal{H}_t$ is the time-ordered sequence of previous events.

When considering the univariate Hawkes process for generating transaction networks, the cross sectional information is not inherently generated, so later in chapter 5 I present different specifications to probabilistically select the edge to change. The probability of a given link appearing at a given time is then

$$P(N_i(t+h) - N_i(t) = 1|\mathcal{H}_{i,t}) = P(N(t+h) - N(t) = 1|\mathcal{H}_t)p_i, \tag{2.16}$$

where $N_i(t)$ is a counting process for edge i, $\mathcal{H}_{i,t}$ is the history for edge i, and p_i is the probability that edge i is selected.

In the multivariate case, the Hawkes process inherently generates the cross sectional information, so the probability of edge i appearing at time t is specified similarly to equation 2.15:

$$\mathcal{P}(N_i(t+h) - N_i(t) = 1|\mathcal{H}_t) = \lambda_i(t|\mathcal{H}_t)h + o(h). \tag{2.17}$$

Equations 2.16 and 2.17 specify the generative model for the univariate and multivariate Hawkes transaction network models respectively.

Hawkes processes have been extensively applied in finance (for detailed reviews of a large range of financial applications, see [74, 93]), although there are relatively few examples in the literature where the Hawkes process is combined with network models. One such example is presented by Linderman et. al. [84], who developed a probabilistic model that combines the mutually-exciting point process with random network models to produce an implicit network, for use cases such as trades on a stock market

which are likely to cause subsequent activity on stocks in related industries. They apply their methods to discover the latent structure underlying financial markets, which is of use to reveal interpretable patterns of interaction and to provide insight into the stability of markets. Another relevant application of Hawkes Processes is in the modelling of Limit Order Books with several examples in the literature [94, 95, 96, 97], showing particular promise in modelling price jumps [98]. One such example was presented recently in Morariu-Patrichi et. al. [99], who propose an extension to the Hawkes process to allow them to account for the state of the Limit Order Book, meaning that the effects of price, volumes, bid-ask spread or other properties of the order book have influence on the arrival rate of orders. Following a similar approach for the trade executions themselves would be an interesting next step for my research. For my research, I focus on the use of the original Hawkes process for simplicity and interpretability. Given their utility in modelling processes in which events trigger future events, Hawkes processes are also prevalent in non-financial applications, for example in disease modelling [100, 101] and social networks [102, 103].

2.2.4 Datasets for studying financial systems

The financial markets present an excellent opportunity to study complex systems with the availability of large, high frequency datasets and significant industry knowledge to extract insights from the data. Here I briefly summarise the most commonly used types of data that exist, along with their availability:

- End of day prices of stocks

  Data on the final prices of stocks when stock markets conclude are widely available through several data suppliers, so have received significant attention from researchers. The use of networks has been a particular focus, where these are constructed from matrices of the correlations between the time series of different stock prices (see [104] for a review). These networks allow an assessment of how different stocks, and also how different market segments, affect each other. However, as these datasets only present an aggregate view of the markets, they do not allow for a study of how individual's behaviours in the markets affect other participants, or the markets as a whole.

- Transaction datasets

  In contrast to end of day prices, transaction level data for markets allows the study of how individual market participants and their actions affect other market participants and the overall system. This means they are of particular interest with respect to addressing the aims of this thesis. However, data at transaction level is rarely available to researchers due to its sensitivity (see however [84, 105, 106, 107, 108]). There have been recent developments in generating synthetic transactions to allow for data sharing [109].

- Order book data

  Transaction level data shows the transactions that are executed. However, further insight can be gained by studying all of the buy and sell orders, as this provides an additional view of the intentions of market participants [110]. Order book data was not available for the research presented in this thesis.

Since the majority of the research presented in this thesis focuses on transaction reports for individual equities reported to the Financial Conduct Authority under MIFID II regulation, presented here is an overview of these datasets. These datasets were available in their raw transaction form, containing information on price, volume, transaction time and anonymised identities of market participants. The transactions are reported to the nearest microsecond, providing a highly detailed view of the behaviour of these markets at the lowest level of granularity. For my research presented in chapters 3 and 4, I select at random three equity instruments from the energy sector and consider the giant component

Dataset	Time period	# of transactions	# of traders
Equity-1	01/01/18 to 01/01/20	3,496	723
Equity-2	01/01/18 to 01/01/20	1,540	323
Equity-3	03/06/19 to 05/11/19	21,742	3854
FTSE-A	08/12/20	106,929	5,433
FTSE-B	08/12/20	43,246	1,051
FTSE-C	08/12/20	66,595	5,429
FTSE-D	08/12/20	105,070	3,041
FTSE-E	08/12/20	16,544	457

Table 2.1: Details of the datasets of transaction reports used in this thesis, collected under MIFID II regulation.

networks of these. These were traded sufficiently frequently to be able to observe meaningful trading networks on a daily basis. I chose to study networks of transactions for stocks on energy companies to allow a comparison between instruments with expected similar properties and trading frequencies. The first two instruments were traded less frequently than the third as can be seen in table 2.1, with one focusing on oil and gas exploration and production and the second focusing on renewable and alternative energy. The third instrument, another oil and gas production stock, shows a much higher volume of transactions so is considered over a shorter time window. Due to the sensitivity of the data, these have been referred to as Equity networks 1, 2 and 3 throughout this thesis. Later on in this thesis, I also consider significantly higher frequency traded financial instruments as these are most suited for the inference of the transaction sequence generation methods I consider. Specifically, I look at five FTSE 100 constituent instruments focusing on a single day of trading and refer to these as FTSE A-E. Although these instruments are FTSE 100 stocks which will be dominated by high frequency trading strategies, for the day considered they vary significantly in the number of transactions, with the smallest being 16,000 transactions (on average approximately one transaction every 2 seconds) and the largest 107,000 transactions (on average approximately one transaction every 0.2 seconds). Table 2.1 contains the high level properties of these datasets.

Open-source datasets

When considering techniques to study market data, significant insight can also be gained by applying methods to datasets from other fields with similar behaviours. In the research presented in this thesis, I also make use of datasets relating to global trade as well as social network data.

The social network data I consider in chapter 3 is a dataset of private messages sent on an online social network at the University of California. An edge (u, v, t) means that user u sent a private message to user v at time t. As this network is unweighted, the weights of all of the edges have been set to 1. I then aggregate the network into daily snapshots, in which the edge weight is the number of times that edge is active during that day.

The first global trade dataset considered in chapter 3 tracks bilateral trade flows between states from 1870-2014, describing import and export data in current U.S. dollars for pairs of sovereign states [111]. This dataset is interesting not just due to its relevance to my focus on financial markets, but also due to an observed growth across time, apart from in two time periods corresponding to the First and Second World Wars. The second considered in chapter 4 was the Financial Services segment of the Balanced Trade in Services (BaTIS) dataset, a complete and consistent trade in services matrix created by the OECD and WTO, covering the more recent period from 2000 to 2019. Full details of the compilation methodologies for this can be found at [112].

2.3 Related relevant studies

In this section I present a review of related relevant literature. Where there is literature that is specific to a single area of my research, I present it within the relevant chapter. As I present the literature, I also highlight how my research is motivated by other's findings, and how my insights differ and complement these.

2.3.1 Related static network studies

First, here I highlight relevant literature which considers the effect an individual node or edge can have on network structure. This depends not only on the scale of its activity, but also on its position within the network and the activity of neighbouring nodes and edges. Understanding these interrelations remains one of the key challenges in network science.

Focusing first on nodes, structural node importance has gained a large amount of recent attention due to its relevance in use cases across a wide range of fields [113]. Methods have predominantly focused on network spectra, in order to elicit structural information from the network adjacency matrix. This includes numerous studies of epidemic processes, in which it is intuitive that the removal of a node that acts as a bridge between communities can be used to stem the spread of a disease, leading to significant effort being taken to understand the influences of community structure on epidemic spreading [113, 114, 115]. Similar applications include preventing network-based attacks [116, 117] and understanding and actioning on the spread of gossip in society [118]. This idea of network resilience is often approached from the angle of percolation theory, in which the percolation threshold governing the appearance of a giant component is related to the leading eigenvalue of the adjacency matrix [119, 120]. An alternative lens is taken by Wang et. al. [121], who make use of the observation that the spectrum of the adjacency matrix gives an indication of community structure. In noting that for a network with c strong communities, the c largest eigenvalues of the adjacency matrix are significantly larger than the others, they follow a perturbation based approach to define node importance as the relative change in the c largest eigenvalues upon the node's removal. Along with the structural information contained within the network spectrum, several researchers look to use the spectrum because it contains information on how processes behave on networks, for instance the stability of spreading processes on social networks [122, 123], or financial shocks on inter-bank networks [1].

Similar to Wang et. al., Lü et. al. [124] propose a universal structural consistency index for a network based on perturbing the adjacency matrix. They demonstrate that this index is a good index for link predictability. Restrepo et. al. use the same approach to define the dynamical importance of network nodes and edges, instead motivated by the observed relationship between the network leading eigenvalue and dynamical network processes [125]. My work considers the same central concept as these of applying edge based perturbations to the adjacency matrix and focusing on the change in the leading eigenvalue. However, my approaches differ in that I propose the use of this concept as static indicator that can be monitored in networks with an evolving structure, as opposed to a measure capturing the effects of node or edge removal on dynamical processes taking place on networks.

Although the bulk of the attention has focused on importance of nodes in networks, Helander et. al. [126] propose a method for characterising the relative importance of an edge, which they refer to as edge gravity. Edge gravity measures how often an edge occurs in any possible network path. They show that important edges are not necessarily adjacent to nodes of importance as identified by standard centrality metrics and that high centrality nodes often have their centrality over-represented by being adjacent to 'edges to nowhere'. Similar path-based methods include the BCC_{MOD} (Betweenness Centrality and Clique Model) proposed by [127], which weights the importance of the two nodes forming the endpoints of the edge with the number of cliques containing the edge. Their method outperforms several methods

including Jaccard coefficient [128] and betweenness centrality [129] in identifying critical edges both in network connectivity and spreading dynamic. Both these path-based approaches and the approach I consider in this thesis show strong connections to existing centrality measures; as shown in section 3.3, an approximation to the network eigenvalue derivative is proportional to the product of the constituent nodes' centralities. In addition, my research focuses on the temporal behaviour of the network in relation to structural importance.

Also key to this thesis is the concept of communities in networks and how individuals positions within networks with complex communities can be appropriately accounted for. Many algorithms for community detection have been developed over the past two decades, and several comprehensive reviews of these methods exist (see for example [130, 131]). However, no single method is found to outperform on all types of networks, with different algorithms presenting different pros and cons depending on the characteristics of the networks being considered [132]. Of relevance for the development of my methods is that several methods of community detection rely on the identification of nodes or edges with high centrality, for example the Girvan-Newman algorithm, which follows an iterative approach to removal of edges with a high betweenness centrality [133]. Fortunato et. al. take a similar approach to remove edges with the highest information centrality [134]. In the financial literature, Chan-Lau et. al. [135] explore both community and centrality methods to identify nodes that are systemically 'too interconnected to fail' or 'too important to fail'. They note that the two complement each other for assessing systemic risk in financial networks. It is also worth noting how both centrality and community detection can be intuitively considered using the concept of a random walk, with a number of methods for community detection, and also for finding central nodes, being defined from the perspective of a random walk and correspondingly computed using the network eigenspectrum. For example, spectral partitioning, a widely used method for finding communities in networks, can be interpreted as trying to find a partition of the network such that a random walk will stay long within the same cluster and rarely move between clusters [136]. Eigenvector centrality also relates to a random walk of infinite length, in which each node is chosen uniformly at random from the set of neighbours of the current node. Specifically, the eigenvector centrality of a node is proportional to the frequency with which a node is visited during such a walk [137]. When using eigenvector centrality in networks with complex structures care must be taken to account for disconnected communities as the measure makes use of the leading eigencomponent only.

When multiple disconnected components are present in a network, the adjacency matrix can be written in block diagonal form with the eigenspectrum decomposing into the spectra of the individual blocks. This means that the leading eigencomponent of the full network will be the leading component of the largest block, meaning that nodes in a smaller disconnected community will have an eigenvector centrality of 0, even if they play a central role within their community. Katz centrality is a widely used method which accounts for this by adding a free centrality to each node[138]. Other methods such as those presented in Anguzu et. al. [139] simply calculate the eigenvector centralities of the components separately and weight these appropriately. However, in this thesis, I consider whether the other components of the eigenspectrum can be used in addition to the leading component in order to account for community structure.

To establish if I can use the components of the eigenspectrum to account for community structure, I now explore how other researchers have used different parts of eigenspectra to understand network structure. Much of the research in this area has focused on spectral partitioning methods, which make use of the eigenvector corresponding to the second smallest eigenvalue of the Laplacian, also known as the Fiedler vector, to partition networks [140]. These methods make use of the difference between the coordinates of the Fiedler vector, which provides information about the distance between nodes [141]. However, as is noted in Newman et. al. [142], these methods are still limited to just one part of the spectrum and fail in the detection of community structure when many communities are present. Newman et. al. instead gives methods for detecting communities and presents a new idea of

'community centrality', by making the observation that modularity can be expressed in terms of the eigenvalues and eigenvectors of the modularity matrix. They take an approach similar to that used in spectral partitioning to maximise the modularity benefit function and show that the eigenvalues of the modularity matrix relate to the community structure. They further show that negative eigenvalues can be used to indicate bipartivity, as well as presenting methods to evaluate network correlations, such as assortativity, using the modularity matrix.

2.3.2 Related dynamic studies

To understand how networks evolve across time, many researchers have focused on studying mechanisms for network growth and defining network models to understand the origin of observed properties of real networks [143, 66, 144, 145, 146]. These include the Barabasi-Albert model [147], which demonstrates that scale-free degree distributions observed in real networks can be explained by the presence of growth and preferential attachment in the network evolution. Falkenberg et. al. [148] present a simple adaptation to the Barabasi-Albert model, in which new nodes attach to nodes in the existing network in proportion to the number of nodes one or two steps from the target node. This results in an implicit time dependence, which arises from a node's attractiveness being dependent on its local environment which changes as the network evolves. Central to their model is the idea that network structure and temporal evolution are inherently linked. However their model is limited to the influence of the local environment. Others focus on considering temporal networks as multi-layer networks, in which one can account for the fact that connectivity patterns in different layers can depend on each other. Bazzi et. al. [149] propose a model which explicitly incorporates a user-specified dependency between layers that is flexible enough to allow complex inter-layer relationships such as dependencies between a layer and all layers that follow, incorporating memory effects into the model. A handful of studies have attempted to link overall network structure to temporal evolution, such as Peixoto et. al. [150], who suggest a dynamical variation of the degree-corrected stochastic block model that is capable of finding meaningful large-scale temporal structures in real-world systems and predict their temporal evolution. Their method works with both discrete and continuous time representations, making it versatile to a range of applications. Watts et. al. [151] consider semi-random 'small world' networks and show that the dynamics are an explicit function of the network structure. They also find an enhanced propagation speed for small world networks.

As briefly introduced in the methods section, a common and general framework for network growth is the fitness model, in which each node has associated with it a time independent 'fitness' which represents its propensity to attract links, as proposed by Barabasí and Bianconi [65] and further emphasised in [152]. They find that different fitnesses result in multi-scaling in the dynamic evolution, or in other words that the time dependence of a node's connectivity depends on the fitness. Attempts have been made to build on this model in order to understand the origins of network dynamics, such as a recent study by Kobayashi et. al. [153]. They find that population and activity dynamics are sufficient to explain two types of scaling empirically observed in real networks. However, their methods do not explicitly allow for different roles to be captured within a network, by assuming a uniform distribution of fitness parameters.

Given my aim in chapter 4 to develop methods that help in the understanding of node behaviour in a dynamical setting, it is also relevant to explore literature that links network eigenspectra to the dynamics of networks. From the perspective of network community structure, the concept of 'dynamical influence' is explored from the angle of the network's eigenspectrum by Clark et. al. [154], who present methods to find 'Communities of Dynamical Influence' by investigating the relationships between a system's most dominant eigenvectors. The concept of 'temporal centrality' has also been a recent interest of many researchers, with the majority of literature focusing on defining temporal random walks in order to generalise static measures of centrality which are based around the concept of a random

walk [155, 156, 157, 158, 159, 160]. This allows the production of measures which respect the ordering of events in temporal networks and take into account the temporal distance between events. These methods have recently been applied in a temporal context, such as Zhao et. al. [161], who make use of temporal centrality to select peripheral stocks to construct risk diversified portfolios with high return and low risk. Taylor et. al. [162] approach things a little differently, presenting a method to extend eigenvector centrality based methods to temporal networks by coupling centrality matrices for different temporal layers into a supra-centrality matrix, allowing them to calculate both the joint centrality for node i at time t, as well as marginal and conditional centralities. This allows for the study of the node (or temporal layer) centralities separately and analysis of the centrality trajectory across time. In a similar vein to my research, Kim et. al. [163] focus on centrality prediction in dynamic networks, first finding that node centrality is predictable in the context of human social behaviour, before presenting several prediction functions that are suited for different applications. My findings that node and edge importance is predictive of future presence in financial networks complement their findings. By studying how static measures of importance relate to future activity, my work is a step towards connecting the static properties to the dynamics of the network.

2.3.3 Studies exploring network structure, stability and systemic risk

Increasing complexity and stability are inextricably linked, with works as early as May's investigations into ecosystems with increasing biodiversity highlighting the relationship [164]. In the context of financial markets, although market integration and diversification are widely believed to play a stabilising role [165, 166], Bardoscia et. al. [167] demonstrate that two factors of increasing complexity, namely increasing the number of institutions (nodes) and contracts (edges) in an inter-bank network can drive the system to instability. Similarly, Markose et. al. [168] present the idea of institutions being 'too interconnected to fail' through an exploration of the structure of the US CDS market. They consider an empirical network constructed from market shares and make use of the May-Wigner condition for stability [1] in comparison to a random network. They show that although the CDS structure shows better outcomes than a random network when subject to shocks, the demise of any one big player will bring down other big players. Caccioli et. al. [169] showed in a theoretical exploration that uncontrolled proliferation of financial instruments can lead to large instability in markets. They suggest potential interventions such as the introduction of a Tobin tax [170], which is shown by Bianconi et. al. to have a stabilising effect [171]. Related to this, Brock et. al. [172] use 'arrow securities' as a proxy for more complicated hedging instruments and found that these incentivise construction of larger positions, resulting in a reinforcement effect due to large gains/losses as a result of being on the 'right' or 'wrong' side of the market. They showed that this is associated with greater instability and also that the primary bifurcation parameter, marking the onset of instability, occurs earlier when there are more arrow securities.

In contrast to the majority of the data-centric financial literature which focuses on inter-bank trading, Bardoscia et. al. [1] analysed UK Trade Repository data, which includes all transactions occurring through a Central Counterparty clearing house (CCP) in the UK. Considering a snapshot of the open positions on a single day for interest rate derivatives, FX derivatives and credit default swaps as a three layered network, they compare a ranking derived from centrality measures to a ranking derived from modelling the network's response to liquidity contagion, looking at how shocks propagate across the network and translate into payment deficiencies across the different markets. The model considers the stress faced by an institution - the difference between all payments it is required to make and all payment inflows from counterparties, and allows stress to spill over between the layers. They found that centrality measures can be used as a proxy for the vulnerability of financial institutions.

[1] The May-Wigner condition for stability is a critical threshold below which any random network has a high probability of stability and is defined as $D < \frac{1}{Ns^2}$ where D is the network diameter, N is the number of nodes and s is the strength of average interactions between nodes

Many works in the financial literature focus on node specific influence on stability. For example, Battiston et. al. [173] define a node ranking coined DebtRank, which takes recursively into account the impact of distress of an initial node across the whole network. Their measure amounts to the fraction of the total economic value in the network that is potentially affected by the distress or default of a specific node. They apply their method to a network of loans from the Federal Reserve to financial institutions between 2008 and 2010, enriched with equity investment relations, and find a strongly connected core of 22 institutions which all become too systemically important to fail at the 2008 crisis peak. They demonstrate the effectiveness of their node ranking in comparison to other centrality measures and find that it was the only measure to deliver a clear response well before the crisis peak. However, their method specifically considers the case of distress propagation and does not explicitly measure how an individual node or edge affects the structure of the network in general. Barucca et. al. [174] investigate whether a change to a few selected banks in the network of the e-MID[2] market can affect the large scale structure of a network through node removal or degree mutation, and also through comparison of the network structure that results to the original.

Although this thesis focuses on importance from the perspective of individual market participants, considering networks at the level of stocks or market segments, and ranking nodes according to how much they impact others in the system is also of value. In [12], we make use of end of day prices for stocks making up the FTSE 100 and 250 indices over the 15 year period from January 2005 to August 2020, which captures both the 2008 financial crisis and also the initial market shocks experienced in response to the COVID-19 pandemic. By applying an information filtering method to infer the dependency structure from the inverse covariance matrix of the price time series, followed by the application of a sparse probabilistic elliptical model we were able to quantify the 'impact' and 'response' of the system to market shocks. In doing this we observe that central sectors are more likely to be affected by shocks and that diversified stocks are more likely to impact the rest of the market when experiencing shocks. We also observe different behaviour in times of crisis than in periods of relative market calm, in which we see the differences in the sector responses to the shock becoming closer together, suggesting that markets 'behave as one' in times of crisis. This work demonstrates an alternative approach to establishing the importance of nodes in complex systems when we do not have physical observations of the relationships between nodes, so is a useful addition to the toolkit of techniques studied in this thesis.

2.3.4 Generative models for financial transactions

The search for generative models for Market data has received significant interest from the academic community, mainly focusing on time series of stock prices due to the availability of this data. Although many techniques have been developed in an attempt to infer the generative processes that result in the observed prices [175, 23, 176, 177], due to data sharing constraints there has been less focus on the microscopic behaviours of financial systems through the study of the transactions themselves. The temporal dimension has been studied by observing the arrival times of orders in limit order books [94, 95, 96, 97, 98, 99], electronic records of the outstanding orders in individual stocks, as this gives a view of the supply and demand in the market. However, these usually lack counterparty information, so do not allow for a consideration of how the relationships between market participants give rise to the temporal dynamics observed. On occasion, some researchers have had access to counterparty information [84], which has enabled insights such as those presented in Musciotto et. al., who provide evidence of networked structure, defined as market participants having statistically validated preferential trading relationships [105]. Other examples have focused on analysing and classifying the different trading strategies present [106, 107, 108]. However, these studies do not consider the network structure in conjunction with the temporal dynamics of markets. My research in chapter 5 contributes in this space

[2] e-MID is the Italian electronic market for interbank deposits, a platform for trading unsecured money-market deposits.

since I seek models that are able to reproduce cross sectional properties observed in the counterparty information of transactions, in combination with temporal properties observed through the transaction timestamps.

Chapter 3

Evaluating edge importance

3.1 Chapter overview

This chapter focuses on understanding how individual edges affect the structure of networks and how this relates to network stability and evolution. The goals of this chapter are as follows:

- To derive a measure of edge importance which accounts for the structure of the network.
- To explore how this measure of structural edge importance relates to the network evolution.
- To demonstrate how this measure of importance can be used in analysis of real transaction datasets and other network datasets.

To achieve these goals, I explore the use of network spectra to define a measure of edge importance. I then propose a model for temporal network evolution, before exploring the relationship between edge importance and subsequent evolution by assessing the predictability of subsequent edge changes given individual edge importances. I explore these both through the use of toy networks and also in application to transaction datasets and a social network dataset.

3.2 Introduction

Understanding how individual edges in a network influence its structure and evolution is important in a range of applications. Considering financial networks, network structure has implications for financial stability [178], market efficiency [179] and consumer safety [180]. Identification of players to monitor more closely is of paramount importance to regulators and policy makers, with many attributing the severity of the 2008 crisis to systemic flaws in the banking ecosystem [181].

This chapter contributes by defining a measure for edge importance which incorporates information about network structure. It then demonstrates how this measure can be used to identify important relationships in networks which have the potential to influence the subsequent network structure. I start by defining a measure for structural edge importance l_e [1] and propose a model for network evolution in which an edge's importance can be indicative of future edge changes. I show that l_e values are higher for edges which appear to play a more important structural role, and that subsequent changes occurring in the real networks analysed depend to some extent on the value of l_e.

[1] I use l_e as shorthand for l_{ij}

In the following sections, I address two questions: Can the extent to which an edge affects the overall network structure be quantified, and does this provide information on the network's temporal evolution? The literature reviewed in section 2.1 shows that network structural information can be gained from the network spectra, both from the observation that the threshold for the appearance of a giant component in a network relates to the leading eigenvalue, and in that the number of communities can be determined from the number of well separated eigenvalues. The literature also demonstrates that the leading eigenvalue provides an indication of stability in terms of dynamical processes occurring on the network. I want to understand edge importance in terms of network structure and stability, so I thus look to capture both of these in my analysis through considering the derivatives of the network's leading eigenvalue with respect to individual edges. I present evidence that a measure based on this could be a useful indicator in understanding temporal changes in network structure and present the results of its application to five real networks. My main results demonstrate that my measure l_e, calculated as the element-wise derivative of the leading eigenvalue, can be predictive of subsequent edge changes for five different networks analysed. I further show that predictability can be related to the specific realisation of two parameters α and ρ in the network evolution model in which edges change with probability αl_e^ρ. This has potential implications for stability, as a system experiencing more changes to edges of structural dominance could see a reinforcing effect, leading to an unstable system. These methods could be useful in classifying transaction systems to inform regulation activities and policy making. I further show that the scale of resultant transactions can be related to the realisation of two additional parameters β and γ, again with potential stability implications.

3.3 Methodology

3.3.1 Central concept - eigenvalue derivatives as a measure of importance

For a given network $G_t(V, E)$ with adjacency matrix $\mathbf{A}^t$, the eigenspectrum of $\mathbf{A}^t$ is the set of eigenvalues λ[2] that satisfy the equation

$$\mathbf{A^t x} = \lambda \mathbf{x}. \tag{3.1}$$

By observing changes in the eigenspectrum of a network, insight can be gained into structural changes. As I am looking at network snapshots across time, I have a 'time series' of networks and can consider the change in the leading eigenvalue between successive time snapshots,

$$\Delta\lambda = \lambda(\mathbf{A}^{(t+1)}) - \lambda(\mathbf{A}^{(t)}) \approx \sum_{ij} \frac{\partial \lambda}{\partial A_{ij}} \Delta A_{ij}, \tag{3.2}$$

where I have made a first order approximation and the derivative is with respect to the (i, j)th entry of the matrix, as opposed to the entire matrix. Here $\mathbf{A}$ refers to the adjacency matrix, λ refers to the leading eigenvalue of the adjacency matrix and ΔA_{ij} refers to the relative element-wise difference between the two network snapshots, or in other words the change for the individual edge between i and j between the two snapshots.

The two parts of equation 3.2 can be seen as a playoff between the potential for an edge to influence the structure ($\frac{\partial \lambda}{\partial A_{ij}}$) and the actual change in the network structure (ΔA_{ij}). Later in this chapter I experiment with synthetic networks to assess the extent to which my derivation below, which makes approximations and assumptions, captures the true behaviour. The first term, $\frac{\partial \lambda}{\partial A_{ij}}$, measures the sensitivity of the eigenvalue to changes in an individual edge, which I refer to as the structural

[2] Note that earlier in this thesis λ is used to represent the intensity of temporal point processes, here it represents the eigenvalue. Both uses are the standard used in the literature.

importance of an edge and denote by l_e. I now derive equation 3.19 for l_e for the undirected case by taking a perturbation theory approach. Although not explicitly explored in this chapter, I also derive an equivalent for the directed case for completeness.

3.3.2 Undirected case

Consider a perturbation $\epsilon\mathbf{B}$ to the adjacency matrix $\mathbf{A}$:

$$\mathbf{A} \to \mathbf{A} + \epsilon\mathbf{B}, \tag{3.3}$$

where ϵ is an infinitesimally small number. The resulting first order changes to the leading eigenvalue λ and the associated eigenvector $|\lambda\rangle$[3] are:

$$\lambda \approx \lambda_0 + \epsilon\lambda_1 \tag{3.4}$$

$$|\lambda\rangle \approx |\lambda\rangle_0 + \epsilon\,|\lambda\rangle_1\,, \tag{3.5}$$

where λ_0 and λ_1 are the first two terms of a Taylor expansion in of λ, and similarly $|\lambda_0\rangle$ and $|\lambda_1\rangle$ are the first two terms of a Taylor expansion of the eigenvector $|\lambda\rangle$. Substituting these into the eigenvalue equation

$$(\mathbf{A} + \epsilon\mathbf{B})(|\lambda\rangle_0 + \epsilon\,|\lambda\rangle_1) + ... = (\lambda_0 + \epsilon\lambda_1 + ...)(|\lambda\rangle_0 + \epsilon\,|\lambda\rangle_1 + ...), \tag{3.6}$$

and considering terms up to 1st order in ϵ

$$\mathbf{A}\,|\lambda\rangle_0 + \epsilon\mathbf{B}\,|\lambda\rangle_0 + \epsilon\mathbf{A}\,|\lambda\rangle_1 = \lambda_0\,|\lambda\rangle_0 + \epsilon\lambda_1\,|\lambda\rangle_0 + \epsilon\lambda_0\,|\lambda\rangle_1\,. \tag{3.7}$$

Then consider each of the terms in ϵ^n separately,

$$\epsilon_0 : \mathbf{A}\,|\lambda\rangle_0 = \lambda_0\,|\lambda\rangle_0 \tag{3.8}$$

$$\epsilon_1 : \mathbf{B}\,|\lambda\rangle_0 + \mathbf{A}\,|\lambda\rangle_1 = \lambda_1\,|\lambda\rangle_0 + \lambda_0\,|\lambda\rangle_1\,. \tag{3.9}$$

By multiplying the equation for ϵ^1 by the left eigenvector ${}_0\langle\lambda|$ and making use of the Hermitian properties of $\mathbf{A}$ such that ${}_0\langle\lambda|\mathbf{A} = \lambda_0\,{}_0\langle\lambda|$, I find

$${}_0\langle\lambda|\mathbf{B}|\lambda\rangle_0 = \lambda_1\,{}_0\langle\lambda|\lambda\rangle_0\,. \tag{3.10}$$

Since the derivative is with respect to a specific edge, the perturbation corresponds to the single component of the row/column corresponding to that edge, i.e. where $B_{ij} = A_{ij}$ if i or j are the row/column being changed, zero otherwise:

$$B_{ij} = \begin{cases} A_{ij} & \text{if } i = p \text{ and } j = q \text{ or } j = p \text{ and } i = q \\ 0 & \text{otherwise.} \end{cases}$$

Then, expanding the indices,

$$\sum_{ij} x_{0i}B_{ij}x_{0j} = \sum_{ij} x_{0i}A_{ij}x_{0j}\delta_{pq} + \sum_{ij} x_{0i}A_{ij}x_{0j}\delta_{qp} = 2\sum_{ij} x_{0i}A_{ij}x_{0j}, \tag{3.11}$$

where $x_{0,i}$ is the ith component of the eigenvector $|\lambda\rangle$ corresponding to the leading eigenvalue of $\mathbf{A}$. Here I have re-labelled the indices for the second term and have evaluated the δ's. This brings me to my result:

[3] Note that I have switched to Dirac notation here for conciseness, $|\lambda\rangle \equiv \mathbf{x}$

$$l_e = \frac{\partial \lambda}{\partial A_{ij}} = x_{0,i} x_{0,j}. \tag{3.12}$$

3.3.3 Directed case

For the directed case, as before, considering terms in ϵ^n

$$\epsilon_1 : \mathbf{B} \left|\lambda\right\rangle_1 + \mathbf{A} \left|\lambda\right\rangle_0 = \lambda_1 \left|\lambda\right\rangle_0 + \lambda_0 \left|\lambda\right\rangle_1 . \tag{3.13}$$

Here the Hermitian properties of the matrix $\mathbf{A}$ cannot be used, as for directed networks $\mathbf{A}$ is generally not symmetric. I consider instead the matrix $\mathbf{M} = \mathbf{A}\mathbf{A}^{\mathbf{T}}$ and perturbation $\mathbf{M} \to \mathbf{M} + \epsilon \mathbf{C}$ and use the symmetric result from this. This is useful since the singular values of matrix $\mathbf{A}$ are defined as the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathbf{T}}$, such that:

$$\partial \lambda^M = 2 s^A \partial s^A, \tag{3.14}$$

where λ^M is the leading eigenvalue of $\mathbf{M}$ and s^A is the leading singular value of $\mathbf{A}$. I can then make use of my result above for the symmetric matrix,

$${}_0\langle \lambda^M | \mathbf{C} | \lambda^M \rangle_0 = \lambda_1^M, \tag{3.15}$$

where ${}_0\langle \lambda^M|$ and $|\lambda^M\rangle_0$ are the left and right eigenvectors of $\mathbf{M}$. For the directed case, the perturbation is changing just a single element of the adjacency matrix independently, i.e.

$$C_{ij} = \Big\{ M_{ij} \text{ if } i = p \text{ and } j = q, \quad 0 \text{ otherwise.} \tag{3.16}$$

Then, expanding the indices,

$$\sum_{ij} x_{0i}^M C_{ij} x_{0j}^M = \sum_{ij} x_{0i}^M M_{ij} x_{0j}^M \delta_{pq} = \sum_{ij} x_{0i}^M M_{ij} x_{0j}^M, \tag{3.17}$$

leading to the result

$$\frac{\partial s^A}{\partial M_{ij}} = \frac{x_{0,i}^M x_{0,j}^M}{2 s^A}. \tag{3.18}$$

where $x_{0,i}^M$ is the ith component of the eigenvector corresponding to the leading eigenvalue of $\mathbf{M}$. In both the directed and undirected case above, it is worth noting that the derivations can be generalised to allow new links to be added or removed but new nodes cannot be added or removed.

To summarise, in both the undirected and directed cases, my measures of structural edge importance are proportional to the product of the eigenvector centralities of the two nodes involved in the edge:

$$l_e = \frac{\partial \lambda}{\partial A_{ij}} = 2 x_{0,i} x_{0,j} \tag{3.19}$$

$$\frac{\partial s^A}{\partial M_{ij}} = \frac{x_{0,i}^M x_{0,j}^M}{2 s^A}. \tag{3.20}$$

My equations are defined in terms of the eigenvector corresponding to the largest eigenvalue, which

usually has non-zero values only for the largest connected component of a network. For this reason, in this chapter I restrict myself to exploring the giant component of the networks. Generalising these to allow for disconnected components is considered in the following chapter when focusing on node importance. I note here that my approach is general in that l_e can be computed for all networks, weighted or unweighted, directed or undirected, as differentiability of the spectrum is ensured whenever the adjacency matrix is real and symmetric. The perturbative approach is valid in the case of small, isolated perturbations, which I further explore in section 3.4.1.

The relationship between l_e and subsequent edge changes can be captured by observing the distributions of $P(\Delta A_{ij} = 0 | \ln(l_e))$[4] and the joint probability $P(\Delta A_{ij}, l_e)$, which I explore in detail in the results sections 3.5.2 and 3.4.3. My findings from these are compared to my model for the temporal evolution of networks, which I propose in section 3.3.4, to assess the extent to which my model captures the true behaviour observed.

The second term of equation 3.2 considers the changes that subsequently occur in response to the value of l_e. This is of significance from a stability perspective; edges that are structurally important could cause a system to become unstable by changing frequently or by a large amount. Conversely, they may also act to stabilise a system if it begins to move towards a regime of instability. This can be explored by assuming that the evolution of the temporal network is Markovian as discussed in the methods section. I consider this first of all in the proposal of a model for network evolution, parameterised by the extent to which l_e is indicative of the propensity of an edge to change and the scale of the resultant changes. I further assess the predictability of changes from the value of l_e through the use of a logistic regression classifier and relate the performance of this to the model parameters.

3.3.4 Model for network evolution

In order to understand the relation between structural importance and stability of a network over time, I need a model that captures two behaviours. The first of these is that the value of l_e is indicative of the probability for an edge to change, and the second is that the size of a resultant change can be related to l_e.

I thus propose a model in which I control the extent to which l_e influences a subsequent edge change, both in probability of occurrence and resultant scale. Specifically, I propose a model in which the network evolution exhibits the Markovian property as in [66, 150]:

$$A_{ij}^{t+1} = \mathcal{V}_{ij}^t A_{ij}^t \mathcal{U}_{ij}^t + (1 - \mathcal{V}_{ij}^t) A_{ij}^t, \tag{3.21}$$

where $\mathcal{V}_{ij}^t$ follows a Bernoulli distribution $\mathcal{B}(\alpha(l_e)^\rho)$ and $\mathcal{U}_{ij}^t$ is the distribution of edge changes which I choose to take as $\mathcal{U}_{ij}^t = \mathcal{N}(\mu = 0, \sigma = \beta l_e^\gamma)$. Here I introduce four parameters. The first two control the probability of an edge to change - ρ which controls the level to which the value of l_e influences the probability for an edge to change, and α scales $\mathcal{V}_{ij}^t$ to ensure that it is a valid probability. A positive value for ρ indicates that more important edges are more likely to change and a negative ρ would indicate the opposite. The second two parameters control the scale of the resulting changes: β controls the width of the distribution of edge changes, and γ controls the level to which l_e influences the variance of the edge change distribution. The simplicity of this model means that it is unable to account for edges appearing and disappearing in the network. This could be addressed in future research.

[4] Here I have conditioned on the logarithm to make the relationships more visually interpretable.

3.3.5 Parameter estimation in real networks

Assuming that the data evolves according to the model in equation 3.21, observations from real networks can be used to estimate the most likely values of the parameters α, ρ, β and γ from the data.

Estimation of ρ and α

I assume in this section that the networks I consider can be described by a model in which the probability of an edge changing is given by

$$P(\Delta) = \theta_e = \begin{cases} 0 & \alpha l_e^\rho \leq 0 \\ \alpha l_e^\rho & 0 < \alpha l_e^\rho < 1 \\ 1 & \alpha l_e^\rho \geq 1 \end{cases}$$

The maximum likelihood estimate of the probabilities $\theta = (\theta_0, \theta_1, ..., \theta_n)$ then follows the same procedure as in the case of a (potentially biased) coin toss - given a sample of changes k_e, the likelihood of observing these changes given θ is

$$L(k_1, k_2, ...k_n|\theta) = \prod_e f(k_e|\theta_e), \tag{3.22}$$

where $f(k_e|\theta_e)$ follows the Bernoulli distribution $\theta_e^{k_e}(1-\theta_e)^{1-k_e}$, where k_e is the observed outcome of edge e. Taking the logarithm of this, the log-likelihood is given by

$$\ln(L(\mathbf{k}|\theta)) = \sum_e^N k_e \ln(\theta_e) + (1-k_e)\ln(1-\theta_e). \tag{3.23}$$

Since $P(\Delta)$ is constrained to be a probability, to estimate the parameters which result in the maximum likelihood, I need to minimise the negative log-likelihood with respect to multiple inequality constraints:

$$0 \leq \alpha l_e^\rho \leq 1, \tag{3.24}$$

where there is one inequality constraint for each l_e. To do this, I make use of the Karush-Kuhn-Tucker conditions [182] and numerical optimisation, to find the optimal saddle point which maximises L with whilst satisfying these constraints. In practice, numerical optimisation of the log-likelihood in equation 3.23 was used to estimate α and ρ.

Estimation of β and γ

For the case of the distribution of edge changes drawn from a Gaussian distribution with μ=0 and $\sigma = \beta l_e^\gamma$, the log-likelihood is given by

$$\ln(L) = \sum_e^N \ln\left(\frac{1}{\sqrt{2\pi}\beta l_e^\gamma}\right) \exp\left(\frac{-(\Delta A_e^{rel})^2}{2\beta^2 l_e^{2\gamma}}\right), \tag{3.25}$$

where ΔA_e^{rel} refers to the observed relative change of edge e. Differentiating with respect to β,

$$\beta = \sqrt{\frac{1}{N}\sum_e^N \frac{(\Delta A_e^{rel})^2}{l_e^{2\gamma}}}, \tag{3.26}$$

from which I recover the expected standard deviation for a Gaussian in the case of γ=0.

Differentiating with respect to γ,

$$\frac{\partial \ln(L)}{\partial \gamma} = \sum_{e}^{N} -\ln(l_e) + \frac{\partial}{\partial \gamma} \frac{(\Delta A_e^{rel})^2}{2\beta^2} \exp\left(-2\gamma \ln\left(l_e\right)\right), \tag{3.27}$$

which when set to 0 results in

$$\sum_{e}^{N} \ln(l_e) \left(1 + \frac{(\Delta A_e^{rel})^2 ln(l_e)}{\beta^2 l_e^{2\gamma}}\right). \tag{3.28}$$

Substituting 3.26 for β and solving numerically allows me to produce an estimate for γ.

Structural influence and network predictability

Depending on the values of the parameters for a given dataset, the observed values of l_e might be expected to be predictive of subsequent change. Specifically, since ρ controls the relationship between l_e and the propensity for an edge to change, a high value of ρ would suggest that l_e would be more predictive of future change. Similarly for α, within the constraints for αl_e^{ρ} to give the probability of an edge to change, a larger α will increase the distance between change probabilities for edges with different l_e, thus also strengthening the relationship between the value of l_e and the propensity for an edge to change. In order to evaluate these effects, I make use of a logistic regression for classification of edges into changing vs. unchanging from the values of l_e. I compare the results to a null model consisting of the average over multiple trials in which edges randomly change with probability equal to the fraction of observed changes. The data is split into training and test sets in a stratified manner, with 20% used to test the model on unseen data. To account for any class imbalance in the datasets, I make use of random over-sampling, in which observations from the minority class are duplicated. The predictions are compared according to balanced accuracy, defined as the average of recall obtained on each class [183], and area under curve (AUC) scores for both receiver operating characteristic (ROC) curves [184] and precision recall (PR) curves [185] which show the true positive/false positive rates or precision/recall respectively for a range of probability thresholds for the logistic regression.

3.4 Results - synthetic networks

3.4.1 Validation of method using toy networks

Here I assess the extent to which the approximations made in calculating l_e hold. I do this by approximating the change in eigenvalues as the coefficient weighted sum of the edge weight changes, $\Delta\lambda = \sum_e l_e \Delta A_{ij}$, and comparing the gradient of this to the value of l_e. My derivation of l_e makes the simplification in assuming that edge changes occur independently of each other. My first test thus considers the case of an individual edge changing at each timestep, and I consider perturbations applied to a barbell network, to observe the effects of network structure, a ring network, to observe the effects of weight with structural equivalence, and a Erdős–Rényi (ER) network as a baseline. The results in figures 3.1, 3.2 and 3.3 show the line of constant l_e, overlaid with the observed ΔA_{ij} and corresponding $\Delta\lambda$ values.

I observe here that the linear approximation generally holds for relative edge changes less than $\Delta A_{ij} = 0.05$. I also see for the barbell network that l_e captures the structural role of the edges, with edges in the cliques having higher values of l_e than those in the bridge. For the ring network, I observe a poorer fit for edges with low values of l_e, and the larger l_e edges tend to be adjacent to edges with

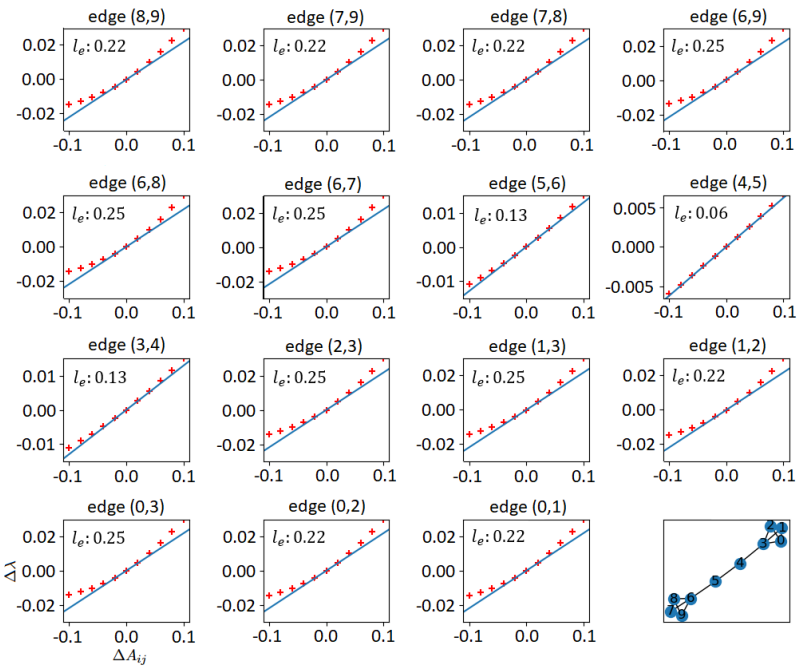


Figure 3.1: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Barbell network, with equal initial weights.

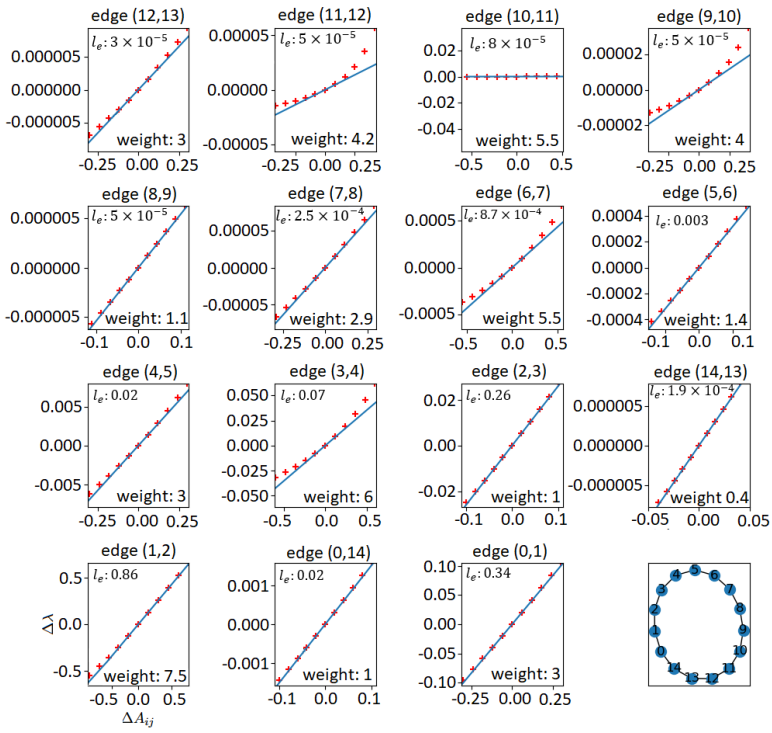


Figure 3.2: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Ring network with each edge independently assigned a random integer between 1 and 10.

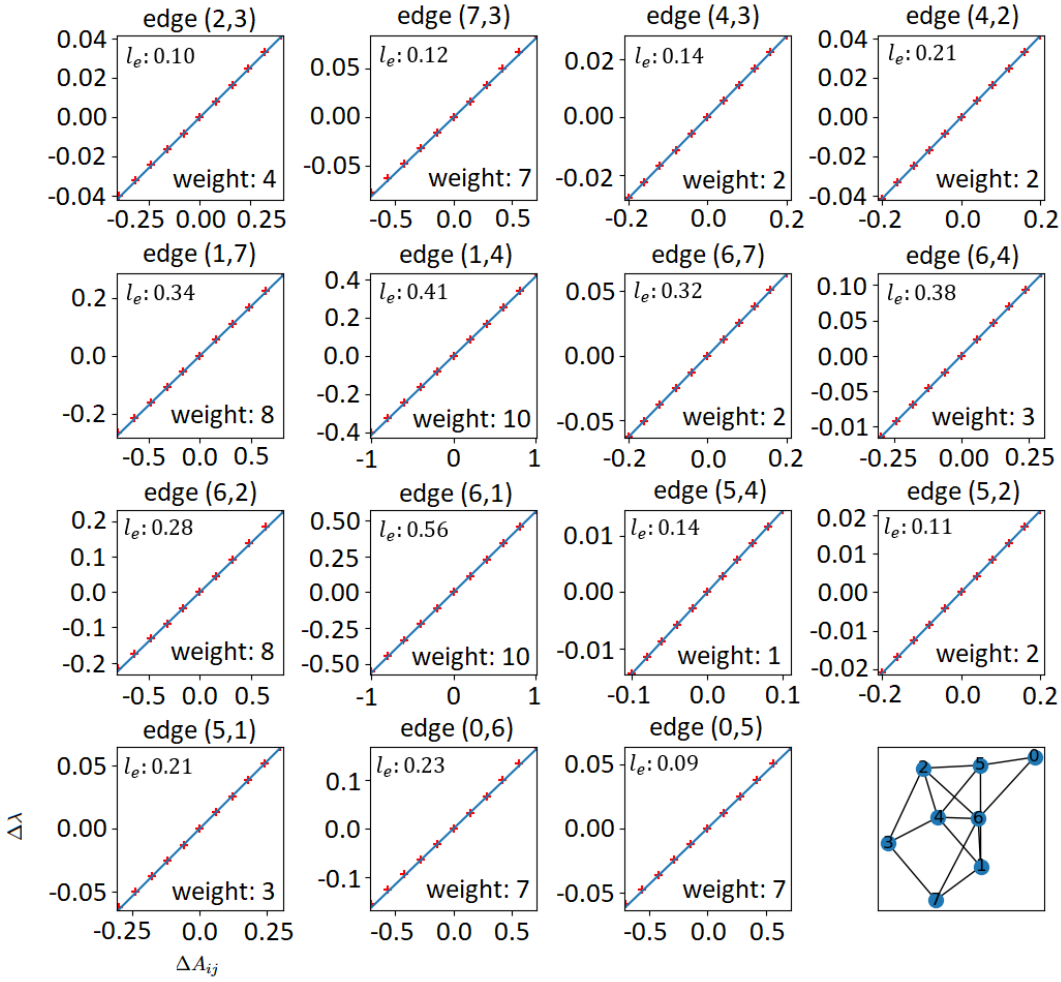


Figure 3.3: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Erdős–Rényi network with each edge independently assigned a random integer between 1 and 10.

similar l_e values. Although the edge with the largest weight also has the largest value of l_e, in general there does not appear to be a simple relationship between edge weight, or weight of neighbouring edges, and the value of l_e. Similar observations are made for the weighted ER network, with the lowest l_e values observed for more peripheral edges, and the two edges with the largest weights also having the highest l_e values. Further results for the case of a weighted barbell, and unweighted ring and random networks are shown in figures 5 - 7 in the appendix.

Results for the case of two edges changing are also shown in the appendix in figures 8-13. In these for the barbell network better fit is observed for higher values of l_e. For the ring networks and random networks, I observe that my model performs well if the observed edge has a larger value of l_e than the other changing edge, but performs poorly when the value of l_e is smaller. The case of complete structural equivalence and equal weights in the unweighted ring network shows good performance for all edges.

The breakdown of the method when there are multiple changes occurring between snapshots suggests that my approximation for l_e may be better suited to a continuous or pseudo-continuous representation of a temporal network, which can be seen as the limit of a discrete temporal network in which each snapshot captures an individual edge change occurring at an infinitesimally different time to the neighbouring snapshot changes.

3.4.2 Relationship between edge importance and the presence of edge changes

I now explore the role of the parameters α and ρ by observing the effect of varying the parameters on the distributions of the values of l_e for changing vs. non-changing edges, $P(\Delta A_{ij} = 0 | \ln(l_e))$. I consider this for synthetic data generated according to my model in equation 3.21, first keeping ρ fixed and varying α, then fixing α and varying ρ.

Model with varying α

Figures 3.4 and 3.5 show the resulting distributions for varying values of α. These show that an increase in α results in a decrease in the probability of an edge to remain unchanged for all values of l_e. Also, For larger values of α, the rate of increase of change probability with l_e is slightly larger.

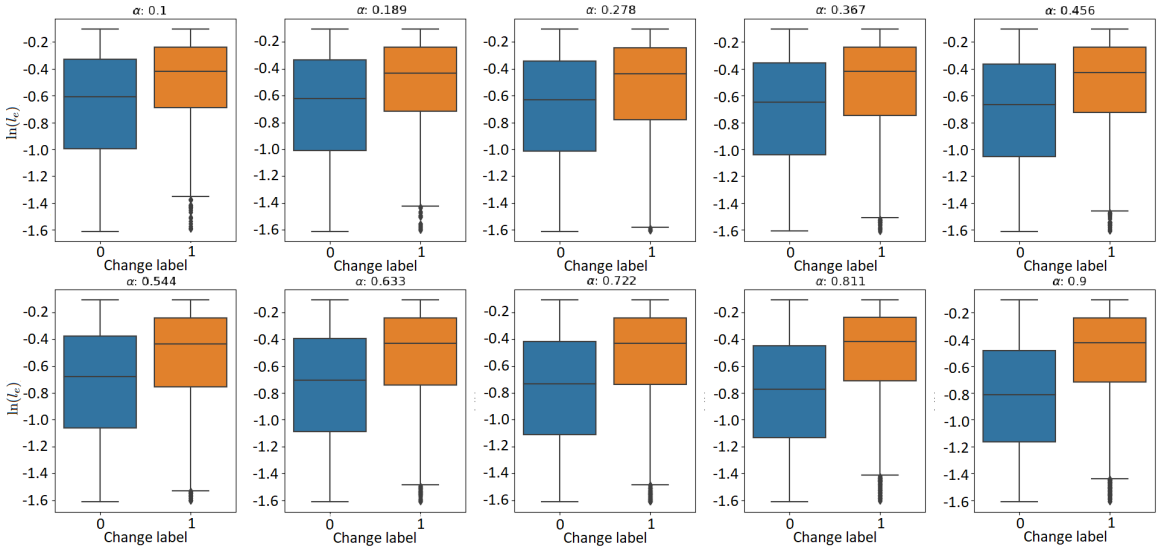


Figure 3.4: Distributions of l_e for edge changes vs. no changes, when varying α.

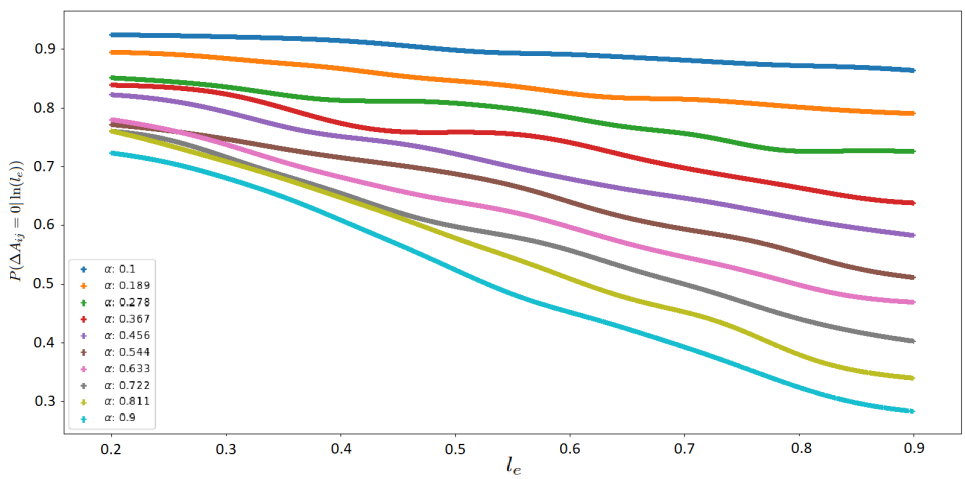


Figure 3.5: $P(\Delta A_{ij} = 0 | \ln(l_e))$ as a function of $\ln(l_e)$ for $0.1 < \alpha < 1$

Model with varying ρ

Figures 3.6 and 3.7 show the resulting distributions for varying values of ρ. These show that for increasing ρ, the probability of observing no change increases for increasing l_e, the probability decreases for a given ρ, at a rate that shows a significant dependence on ρ.

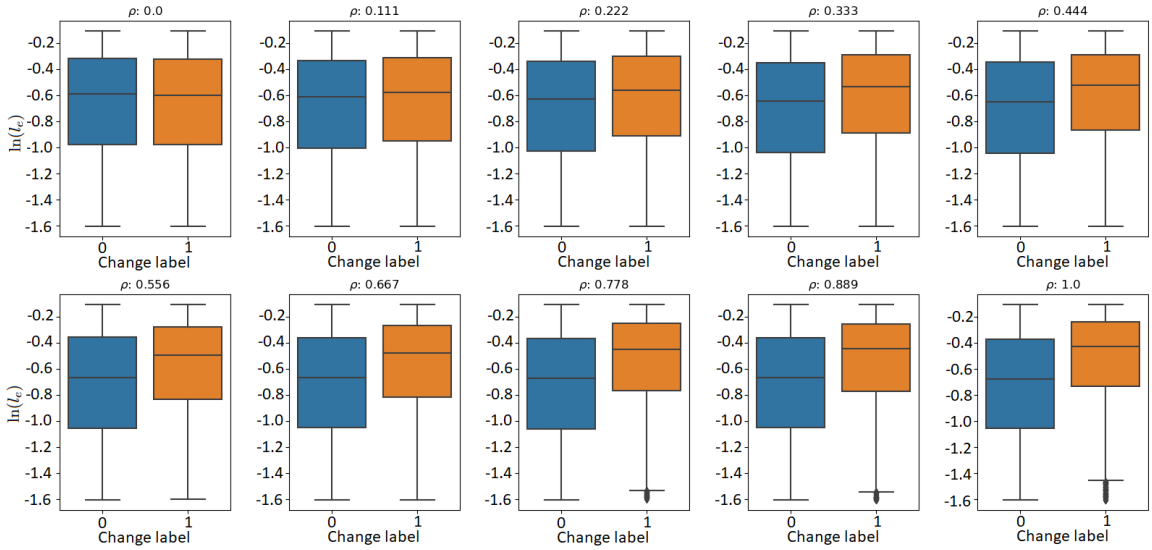


Figure 3.6: Distributions of l_e value for the case of edge changes vs. no changes.

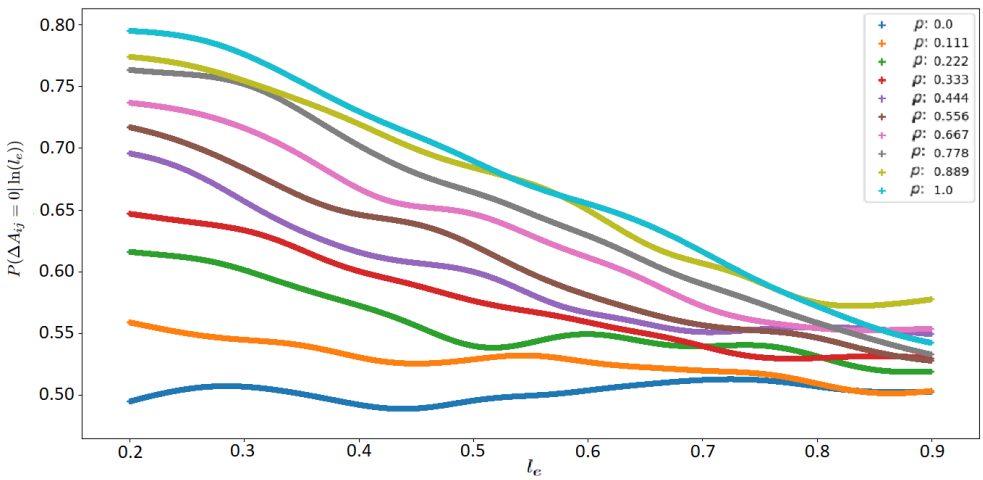


Figure 3.7: $P(\Delta A_{ij} = 0 | \ln(l_e))$ as a function of $\ln(l_e)$ for $0 < \rho < 1.0$.

3.4.3 Relationship between edge importance and size of weight changes

I now consider if the value of l_e is observed to have an effect on the scale of subsequent edge changes.

Variation of γ

Figure 3.8 shows the distributions of $P(\ln(1+\Delta A_{ij}), l_e)$ for a range of values of γ. This shows that for positive γ, the width of the distribution widens for larger l_e. For negative γ, the opposite behaviour is observed, in that the width of the distribution becomes narrower for larger l_e.

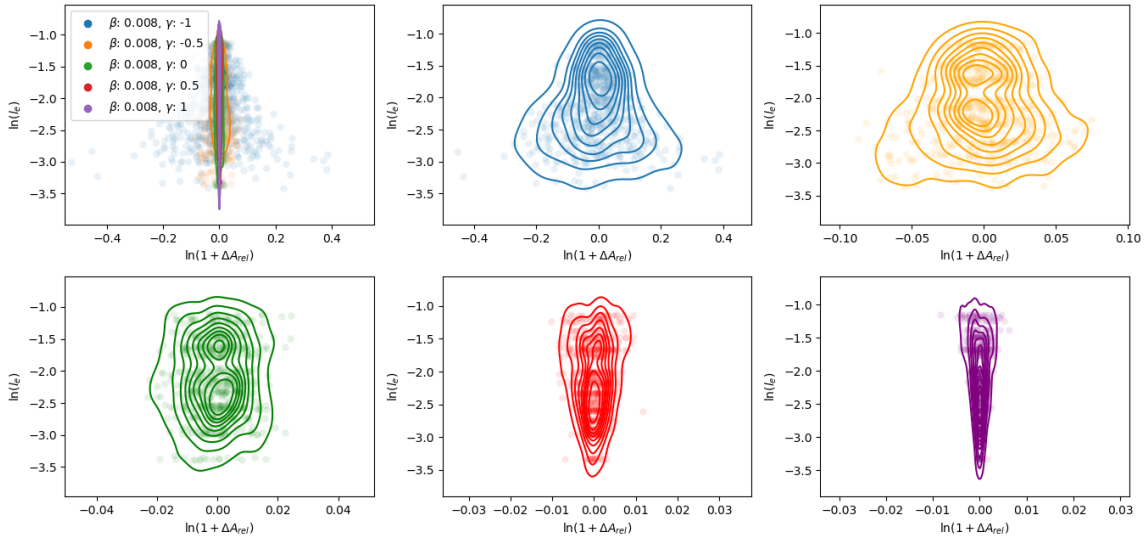


Figure 3.8: Distributions of $P(\ln(1+\Delta A_{ij}), \ln(l_e)$ for fixed $\beta = 0.008$, $-1 < \gamma < 1$. Underlying observations of $\ln(l_e)$ and $\ln(1+\Delta A_{ij})$ represented by the dots underlying these.

Variation of β

Figure 3.9 shows the distributions of $P(\ln(1+\Delta A_{ij}), l_e)$ for a range of values of β. This shows that as β increases, the width of the distributions increase.

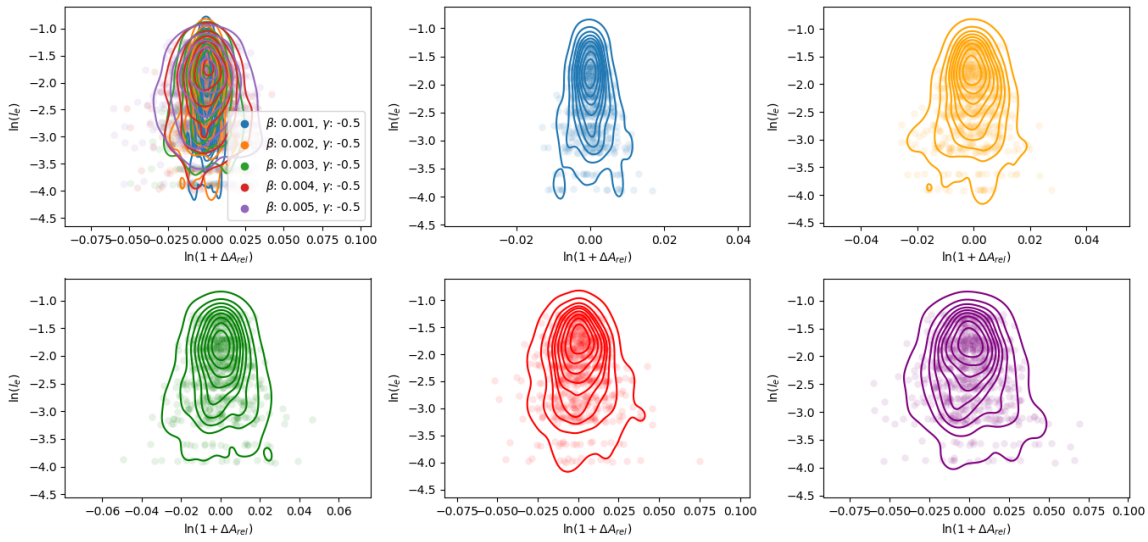


Figure 3.9: Distributions of $P(\ln(1+\Delta A_{ij}), \ln(l_e)$ for fixed $\gamma = -0.5$, $0.001 < \beta < 0.005$. Underlying observations of $\ln(l_e)$ and $\ln(1+\Delta A_{ij})$ represented by the dots underlying these.

Dataset	Corr(l_e, ΔA)	Corr(l_e, EBC)	Corr(l_e, $k_i k_j$)	Corr(l_e, $S_i S_j$)
Bilateral Trade	-0.061	-0.397	0.352	0.786
College Message	-0.169	0.035	0.581	0.434
Equity-1	-0.104	0.135	0.717	0.320
Equity-2	-0.047	0.166	0.580	0.265
Equity-3	-0.010	0.041	0.923	0.763

Table 3.1: Spearman's rank correlations for l_e with the rank by edge weight, edge betweenness centrality (EBC) and product of nodes' degrees and strengths.

3.5 Results - real networks

3.5.1 Static observations in real networks

The above application to synthetic networks demonstrates that my model behaves as expected, with networks with a large ρ (and α) being more predictable. Now I explore the performance of my structural influence metric and model through the application to five real datasets. Firstly, given that my research has been motivated by a need to monitor risks in a financial setting, I consider a network of country level bilateral trade [111] and three different capital markets transaction datasets reported under MIFID II regulations. However, my methods can be applied more generally to any temporal networks. Due to the availability and high volume of research conducted into social networks (see [118]), I also consider a network of messages sent between College students [186]. A full description of these can be found in section 2.2.4.

In order to understand the usefulness of l_e as a metric for structural importance, I first examine the edges that rank the highest according to their values of l_e for the bilateral trade dataset, since the historical context of international trade can give an indication of which edges I might expect to be 'important'. For the bilateral trade dataset, the largest value of l_e is observed for the edge between Portugal and Spain in 1872, and considering the sum across all time for Greece and Turkey. These are examples of edges with both nodes having large eigenvector centrality; edges involving only one central node are seen to have lower values of l_e. This means that inter-European edges almost exclusively make up the top 100 ranked edges, whereas the lowest ranked l_e edges occur when one, or both, of the nodes have very low centrality scores. Similarly, for the other datasets, the highest values of l_e are also observed for edges involving nodes with high eigenvector centrality. In general, the rankings of l_e are observed in table 3.1 to be uncorrelated with the rankings of edges according to their betweenness centrality, or their mean value of ΔA_{ij}, but do for some cases correlate with the product of the participating node's degrees and strengths.

As the equity datasets contain large numbers of edges (the smallest contained 2785 edges), I cannot fully explore all of the individual observed values of l_e as for the toy networks. Instead, I consider the probabilities of observing values of l_e by making use of Kernel Density Estimation to estimate the probability density functions from the data.

Figure 3.10 shows the estimated probability density functions of the logarithm of the value of l_e. These show that, in all networks, the values observed for l_e tend to be very small. Omitting the tails of the distributions for diminishingly small values of l_e, a similarity in the values of l_e is observed across the 3 equity datasets. Although across all 5 datasets analysed, the distribution is found to be approximately lognormal, the social network shows a much broader distribution of l_e. The peak of the distribution for the college messaging dataset is also much lower, observed at approximately $\ln(l_e) = -8.8$, whereas the bilateral trade dataset shows a peak at -3.3 and the equity datasets at -3, -2.5 and -4.2.

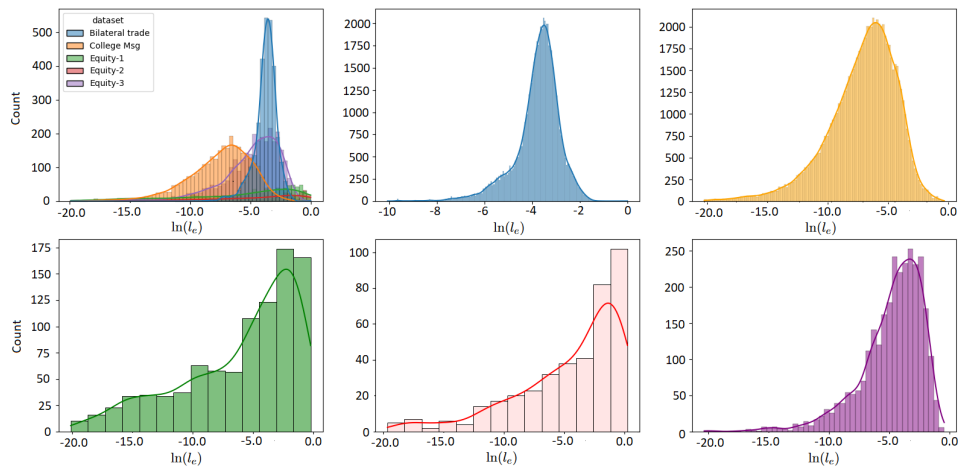


Figure 3.10: Probability distribution of the values of $\ln(l_e)$ for different networks.

3.5.2 Relationship between edge importance and the presence of edge changes

I now address the central concept of the relationship of l_e observed for the real networks and the probability of an edge to change. Figure 3.11 shows the distributions of the $\ln(l_e)$ values observed for non-changing edges in comparison to changing edges. These show that in all cases, there is a shift in the mean value of $\ln(l_e)$ towards higher values for edges which do change, which would be suggestive of a positive ρ parameter, and potentially the ability to predict the presence of changes given the value of l_e[5]. The smallest shifts are observed for the bilateral trade dataset and Equity-3, which show negligible differences in the mean and quartiles of the values of l_e for changes and no changes, suggesting that predictability of changes might not be expected from the values of l_e in these cases. In all cases, the differences in the mean values of l_e for change vs. no change is significant, with a two-sided t-test showing $p < 0.05$ for all datasets.

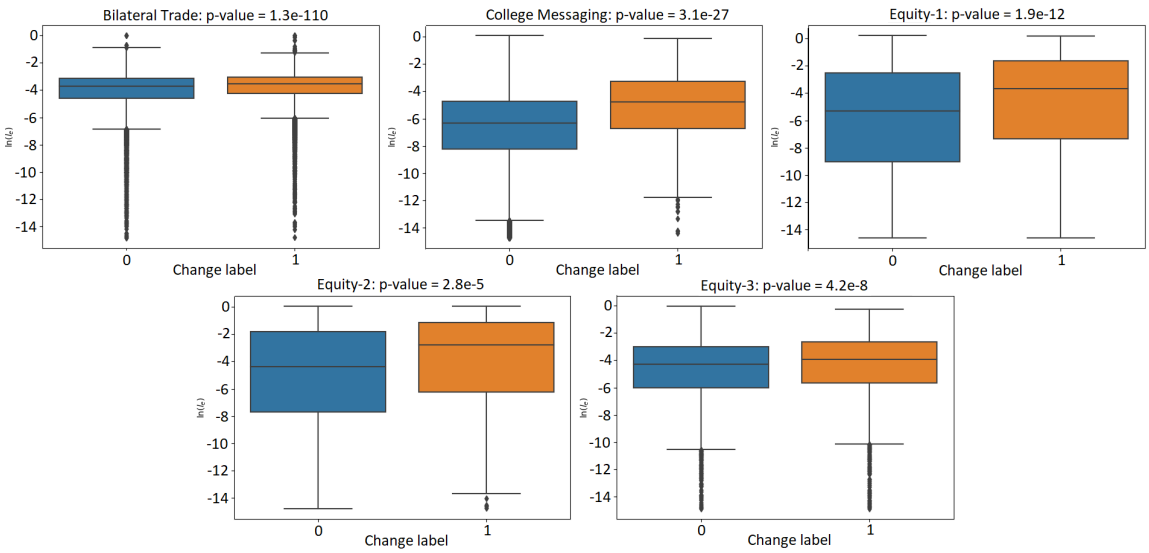


Figure 3.11: Boxplots showing the distribution of l_e values observed according to the presence or absence of an edge subsequently changing.

[5]In appendix section .1.1 I show how predictability relates to the values of the parameters α and ρ.

To further understand how the value of l_e relates to the probability for edges to change, I consider the distributions of $P(\Delta A_{ij} = 0|l_e)$ as shown in figure 3.12. Here a decreasing probability of $\Delta A_{ij} = 0$ is observed for the bulk of the distribution for increasing l_e for the bilateral trade and Equity-3 datasets. However, the rarely observed edges with $l_e > 0.3$ for these datasets show larger probabilities to remain unchanged. A slight initial decrease is also observed for Equity-1 and 2 datasets however the relationship is clearly non-linear for large l_e. The college messaging dataset shows a much larger probability in general for edges to remain unchanged and shows a very slight decrease in probability to remain unchanged for very small l_e values but is dominated by noise for $l_e > 0.05$.

Referring back to section 3.4.2, I considered the ideal cases of linear positive, neutral and negative relationships between l_e and the probability of edge changes. In reality, as shown in figure 3.12, things are more complex, with different relationships apparent for different l_e ranges. In particular, for edges with lower values of l_e, the negative relationship between the value of l_e and the probability of an edge to remain unchanged suggests that a parameterisation of my model with positive value of ρ would be effective in capturing the behaviour of the bulk of the network. However, changes to the small handful of edges with the largest values of l_e are less likely. These observations could suggest that there are a few structurally important edges with consistent trading patterns which act to stabilise a system which would otherwise move towards a regime of instability.

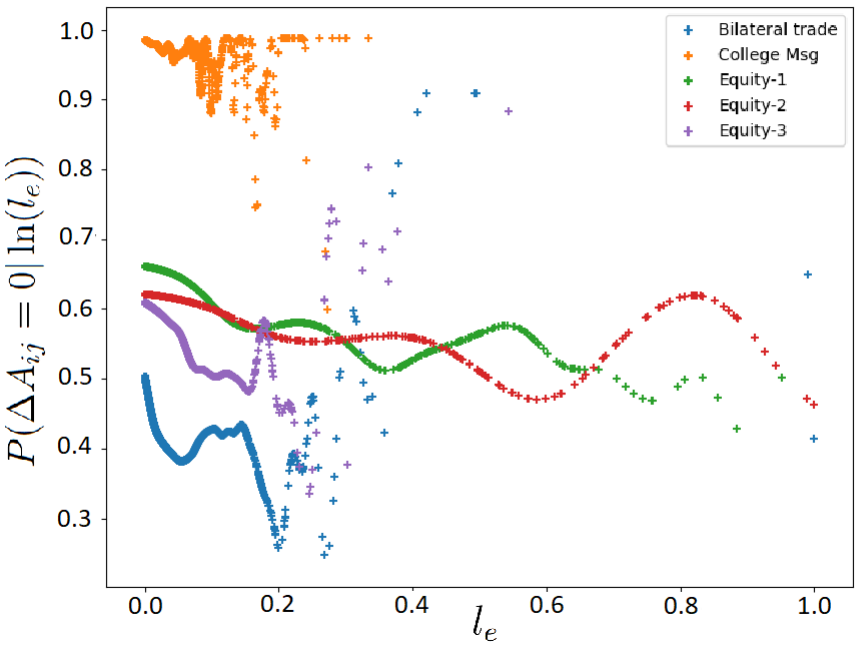


Figure 3.12: $P(\Delta A_{ij} = 0|\ln(l_e))$ as a function of $\ln(l_e)$ for the 5 real datasets.

Estimation of α and ρ from data

In table 3.2, I present the values of α and ρ estimated for my 5 different datasets. The errors on these estimations are given by the inverse Hessian of the log-likelihood, which is found by numerical approximation. In comparison with figures 3.5 and 3.12, the ordering of the estimated value of α appears to agree with the positions of the college messaging dataset and the equity datasets. The parameter ρ appears to correspond with the overall gradients observed in figure 3.12 for the bulk of the distributions observed for low values of l_e. These observations suggest that my model is mostly capturing

Dataset	Estimated α	Estimated ρ
Bilateral trade	$0.783 \pm 1.08 \times 10^{-4}$	$0.072 \pm 1.21 \times 10^{-5}$
College messaging	$0.033 \pm 3.03 \times 10^{-6}$	$0.270 \pm 2.76 \times 10^{-6}$
Equity-1	$0.392 \pm 2.74 \times 10^{-4}$	$0.030 \pm 2.82 \times 10^{-5}$
Equity-2	$0.401 \pm 3.13 \times 10^{-5}$	$0.016 \pm 4.93 \times 10^{-5}$
Equity-3	$0.465 \pm 2.57 \times 10^{-4}$	$0.036 \pm 4.21 \times 10^{-5}$

Table 3.2: Estimated α and ρ for the 5 real datasets. Standard errors have been calculated from the inverse Hessian of the log-likelihood.

the imbalance of observed changes in the parameter ρ and the overall average change probability for each dataset in the parameter α.

Figure 3.13 shows the bulk of the distributions for $P(\Delta A_{ij} = 0 | \ln(l_e))$ for my 5 datasets, in comparison to the equivalent generated from my model for network evolution which I remind the reader is:

$$A_{ij}^{t+1} = \mathcal{V}_{ij}^t A_{ij}^t \mathcal{U}_{ij}^t + (1 - \mathcal{V}_{ij}^t) A_{ij}^t. \tag{3.29}$$

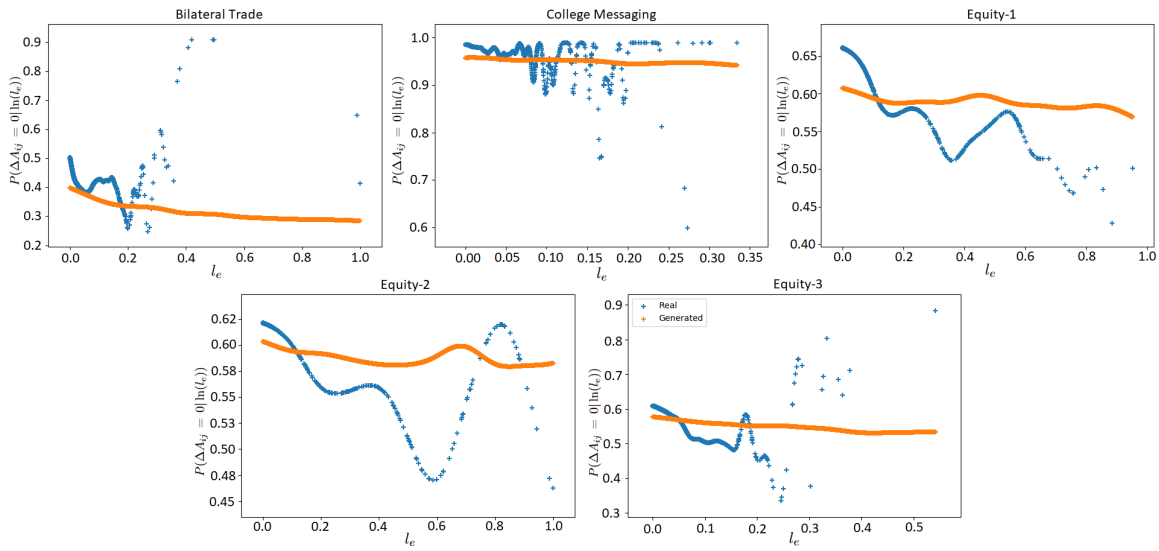


Figure 3.13: $P(\Delta A_{ij} = 0 | \ln(l_e))$ as a function of $\ln(l_e)$ for the 5 real datasets, overlaid with the distributions for data generated according to the model in 3.29

This shows that the dataset generated according to the parameters estimated appears to show a reasonable agreement to the actual distribution and differences here can be attributed to the differences in the initial network conditions.

3.5.3 Relationship between edge importance and size of weight changes

I now explore the distributions of $P(\ln(1 + \Delta A_{ij}), \ln(l_e))$, as considered for synthetic data in section 3.4.3, for the case of edges that do change, i.e. $\Delta A_{ij} \neq 0$ for the five real networks. These can be seen in figure 3.14. Note that ΔA_{ij} refers to the relative change in the value of the edge weight from t_0 to t_1, which takes values in the interval $[-1, \infty]$, and l_e is measured at time t_0. Infinite values for ΔA_{ij}, corresponding to the case of a new edge appearing, were observed but are not captured in the plots. The prominence of these across the different datasets are 4.7% of the bilateral trade dataset, 0.086% of

Dataset	Estimated β	Estimated γ
Bilateral trade	$1.11 \times 10^{-4} \pm 8.93 \times 10^{-18}$	$1.93 \pm 1.10 \times 10^{-27}$
College messaging	$7.77 \times 10^{-5} \pm 1.10 \times 10^{-16}$	$1.15 \pm 3.07 \times 10^{-27}$
Equity-1	$1.51 \times 10^{-5} \pm 4.55 \times 10^{-24}$	$1.35 \pm 2.55 \times 10^{-36}$
Equity-2	$1.42 \times 10^{-4} \pm 1.36 \times 10^{-26}$	$1.32 \pm 5.81 \times 10^{-37}$
Equity-3	$1.63 \times 10^{-4} \pm 2.20 \times 10^{-26}$	$1.42 \pm 1.43 \times 10^{-36}$

Table 3.3: Estimated β and γ for the 5 real different datasets

the college messaging dataset, 0.012%, 0% and 0.0028% of the equity datasets[6]. A slight widening of the distributions is observed for larger values of l_e for Equity-1 and 2 datasets and to a larger extent for the third equity dataset. The bilateral trade dataset shows initial widening as l_e increases, but narrows again for the largest l_e edges. The college messaging dataset shows two distinct peaks, corresponding to changes in edge weight of ± 1, which are over-represented in this dataset as it is unweighted and the edge weight solely represents the count of interactions in the time window of consideration. The slight widening for larger l_e for all datasets is suggestive of a positive relationship between the value of l_e and the variance of the distribution of subsequent edge changes.

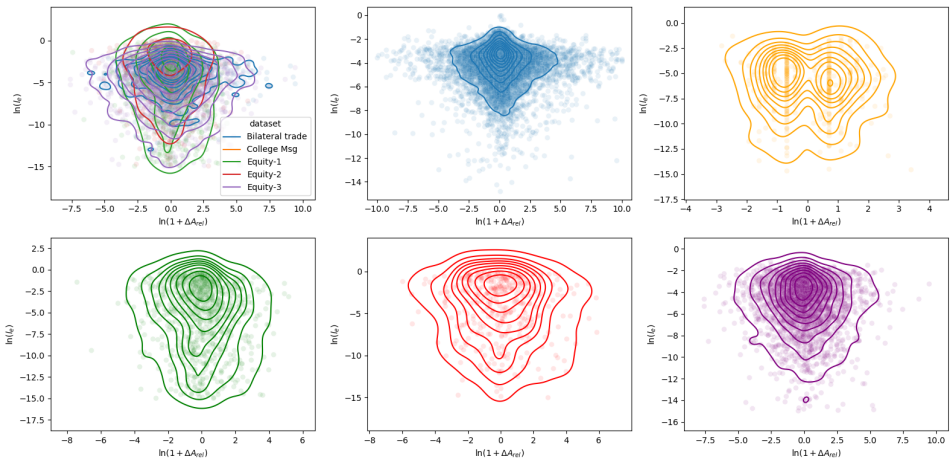


Figure 3.14: Contours showing the distributions of $P(\ln(1 + \Delta A), \ln(l_e))$ for the 5 real datasets. Underlying observations of $\ln(l_e)$ and $\ln(1 + \Delta A_{ij})$ represented by the dots underlying these.

Estimation for β and γ from data

All 5 datasets show positive values of γ, suggestive of a relationship between the width of the distribution of edge changes and the value of l_e. The dataset with the highest value for γ, the bilateral trade dataset dataset, also shows the largest level of bias towards larger change distribution width for higher l_e in figure 3.14. Correspondingly, the lowest γ value is seen for the college messaging dataset, which shows the least bias towards larger changes occurring for larger values of l_e. The values for β are similar across the 5 datasets and all relatively low. It is difficult to draw conclusions from these, as the behaviours controlled by the two parameters cannot be separated and observed alone in the distributions in figure 3.14.

[6]The prominence of new edges has been significantly reduced by focusing on the giant component

Dataset	Balanced accuracy	ROC AUC	Precision-Recall AUC
Bilateral trade	0.542 (0.5 ± 0.0011)	0.554 (0.5 ± 0.0013)	0.628 (0.595 ± 0.0021)
College messaging	0.623 (0.5 ± 0.0027)	0.678 (0.5 ± 0.0029)	0.017 (0.005 ± 0.0045)
Equity-1	0.568 (0.5 ± 0.0039)	0.576 (0.5 ± 0.0046)	0.365 (0.313 ± 0.0062)
Equity-2	0.566 (0.5 ± 0.0061)	0.579 (0.5 ± 0.0073)	0.430 (0.351 ± 0.010)
Equity-3	0.527 (0.5 ± 0.0025)	0.542 (0.5 ± 0.0045)	0.424 (0.381 ± 0.0030)

Table 3.4: Values of AUC scores for ROC and PR curves. Numbers in brackets represent the score and confidence intervals achieved by a model which randomly predicts 1 or 0 in proportion to the dataset prior, averaged over 100 trials.

3.5.4 Edge change predictability

Given the non-zero estimated values of the parameters α and ρ in section 3.5.2, it is natural to assess the performance of using the value of l_e to predict a subsequent change. Figure 3.15 shows the ROC Curves for the 5 different datasets, and the PR curve can be found in figure 16 in the appendix.

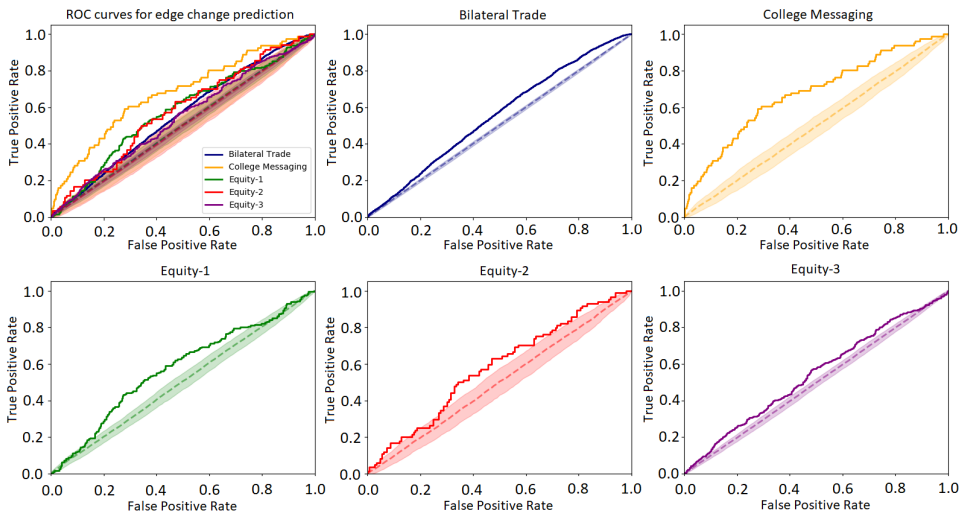


Figure 3.15: ROC curves for a logistic regression classifier making use of $\ln(l_e)$ to predict $\Delta A_{ij} = 1$. The dashed lines and shaded areas represent the mean 95% confidence intervals for the dummy model.

All datasets are seen to perform slightly better than the dummy model, with better performance seen for the college messaging dataset and Equity-1 and 2, which also show larger differences in the distribution of l_e across change vs. no change in figure 3.6. Poorer performance is seen for the bilateral trade and Equity-3 datasets, which show similar shaped distributions in figure 3.12 with an initial steep decrease in probability to remain unchanged for increasing l_e. However, this trend appears to reverse for $l_e > 0.3$. These datasets also show little difference in the distribution of values observed in figure 3.6 and are found to have low values of ρ. Although the college messaging dataset shows the best performance, particularly in the left hand side of the ROC curve, this is driven by the significant class imbalance with only 5% of the observations showing a non-zero ΔA_{ij}, as opposed to the bilateral trade dataset which shows a 20% proportion of non-zero changes and the Equities 1, 2 and 3 datasets which show 68%, 61% and and 63% proportions of non-zero changes respectively. This is also reflected in the PR AUC score for the college messaging dataset being close to the upper margin of error for the null model.

3.5.5 Predictability & aggregation scale

It is worth noting that the behaviour of the networks I have been considering in this chapter is partially dependent on the aggregation scale used in constructing the snapshot networks. Figure 3.16 shows network snapshots at three different aggregation scales across a single afternoon of trading[7] for the Equity-1 dataset. These show that aggregation has an affect on the density of each snapshot, with the fourth snapshot for the hourly aggregation containing very few edges. These also demonstrate how the non-Poissonian nature of this dataset adds an additional consideration, as the latest snapshots appear very similar for all three aggregations, as although there is activity spread across the entire time frame, the bulk of the activity happens after 7pm, reflecting trading strategies which place orders on or after market close, in anticipation of a price movement the next day.

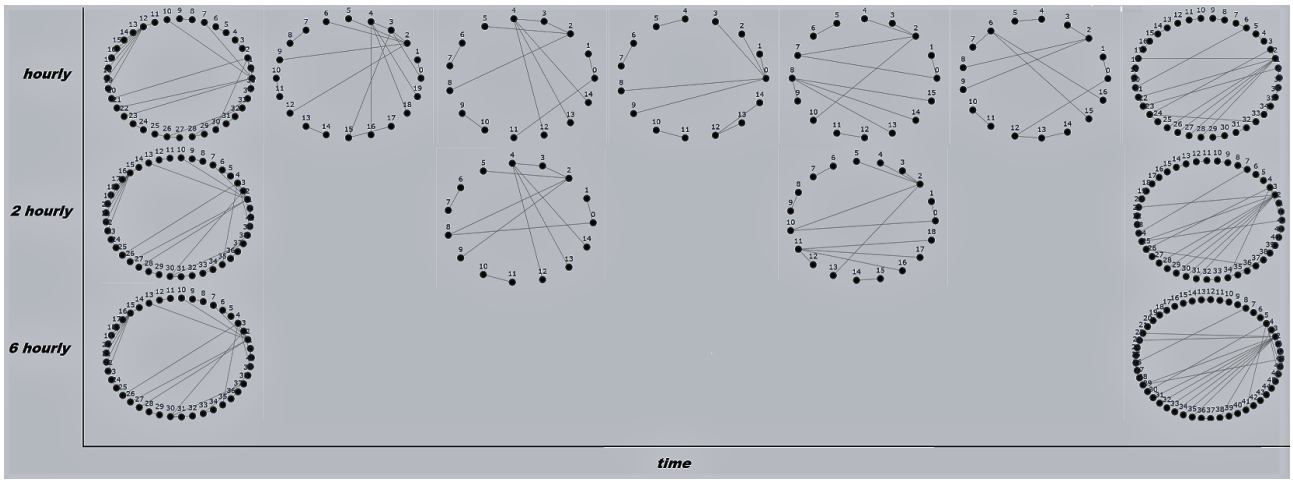


Figure 3.16: hourly, two hourly and 6 hourly aggregated networks across an afternoon (1pm -8pm) for the Equity-1 dataset

Given these differences in the network structure at different aggregation scales and observations presented in Bandi et. al. [187], who note that predictability is aggregation scale specific, I now explore how the predictability of edge changes varies with aggregation scale. For this, I consider a simple experiment where my methods presented in the previous sections are applied across a range of different aggregations, and the performance statistics across the different aggregations are compared. Figure 3.17 and 17 and 18 in the appendix show the improvement of the three classification performance metrics in comparison to the null model. These show that larger aggregation scales produce lower predictability and that the predictability varies with aggregation, displaying a 'sweet spot' in which the predictability is maximised at a given aggregation scale. However, also note that my methods are unsuitable for low levels of aggregation, due to the sparsity of the network at these scales. This presents an interesting trade-off with the results presented in section 3.4.1, in which I demonstrated that the measure l_e itself is most valid in the case of isolated edge changes.

3.6 Summary & next steps

The ability to understand how microscopic changes in networks affect the macroscopic evolution across time is one of the key challenges in dynamic network analysis. In this chapter I have begun to explore the use of derivatives of network spectra to capture this. I have derived a measure of edge based structural influence, l_e, and explored the extent to which the value is indicative of future changes. I first

[7] A single afternoon of trading was chosen here to allow for networks of a reasonable size to analyse visually

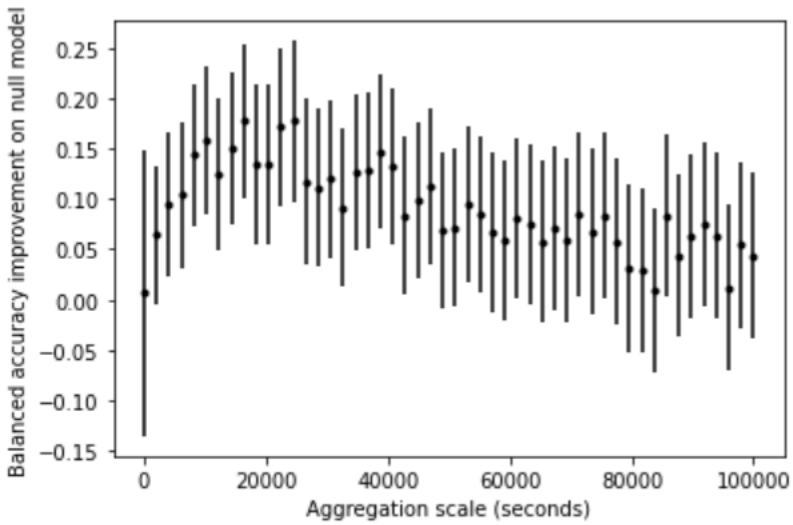


Figure 3.17: Average improvement and confidence intervals of balanced accuracy from null model for aggregation scales ranging from 50 seconds to 27 hours, calculated over 100 trials.

of all demonstrated that for small and isolated perturbations applied to the network, the eigenvalue derivative in equation 3.2 is able to capture the effect of an individual edge on the eigenvalue. However, I observed the relationship breaks down for multiple changes happening during the same time snapshot, suggesting that the measure may be more suited to a continuous or pseudo-continuous representation of network evolution, in which each time snapshot contains a single edge change.

Considering the 5 real datasets, I observed lognormal distributions of the values of l_e, indicating structural influence dominated by a small handful of edges. I proposed a model in which the probability for an edge to change is given by αl_e^{ρ}. This model allows the user to control the extent to which l_e dictates the propensity for an edge to change and also controls the scale of a subsequent change. Focusing on the former, I observed similarities in the shapes of the distributions of $P(\Delta A_{ij} = 0, \ln(l_e))$ when generating synthetic networks according to this model and those observed in the data. Also, the values observed for α and ρ are suggestive of a relationship between the value of l_e and the subsequent presence of change. In using l_e in a logistic regression classifier to predict change, I observed that l_e is slightly predictive of change in all cases, but only marginally so for the case of the bilateral trade and Equity-3 datasets. This corresponds with my observations of small values of ρ for these datasets, along with similar, non-linear distribution shapes for the probability of no change for increasing l_e. These observations indicate that the static structural importance can be indicative of the presence of a subsequent change. It is likely that the results of the prediction exercise are impacted by the class imbalance of the datasets considered. In the next chapter, I will be conducting a similar prediction exercise for node importance for which this class imbalance is exacerbated as across all of its edges with individual nodes being highly unlikely to remain unchanged between snapshots. In order to reach a more balanced dataset, instead of looking to predict nodes which remain unchanged, I will instead look to predict whether or not nodes transact in the next time snapshot.

I note here that α and ρ themselves are useful parameters that could be used to classify networks according to their growth stability. A large value of α would be an indicator for larger levels of overall network activity. A network with very large ρ would be characterised by changes occurring to the edges with the largest l_e, conversely, a network with very small ρ would see changes distributed across all edges, regardless of the value of l_e. In the context of financial markets, these contrasting situations would require different approaches, and ρ could be used by policy makers to inform which asset classes should be monitored as a whole (for the case of small ρ) or following an approach targeting those edges with the highest l_e.

When considering the parameters which control the extent to which importance influences the scale of the resultant edge changes, I observe a slight widening of the distribution for larger l_e for all datasets. This is suggestive of a positive relationship between the value of l_e and the variance of the distribution of subsequent edge changes, which is supported by observations of positive γ for all datasets. As above, this observation could be used by policy makers since a larger γ would motivate a monitoring of larger l_e edges due to their potential to make larger changes.

My model doesn't account for edges appearing and disappearing in the network and assumes that edge changes are independent of each other. For the first limitation, note that edge appearance and disappearance would be unlikely to heavily influence the behaviour of the Equity networks, as I observed very low percentages (0.012%, 0% and 0.0028%) of new edges appearing [8], but for the other two networks this behaviour is more prominent at 4.7% for the bilateral trade network and 0.086% for the college messaging network. On the second point, I noted in my exploration of toy networks that the ability of using the eigenvalue derivative to quantify importance breaks down for multiple edge changes present, and in section 3.5.5, I demonstrated the how predictability initially improves with aggregation scale before deteriorating. This presents a trade-off between improved capability of l_e for the quasi-continuous limit in which each time snapshot contains a single edge change and improved predictability for larger aggregation scales. This motivates the research I present in section 5, in which I consider instead modelling transaction networks as temporal point processes as opposed to snapshot networks. It is also worth noting that although using raw transaction data gives the lowest granularity view of the data, my work has not considered the higher order effects of trading behaviour on price. Such an effect results in the influence of edges reaching disconnected components, which cannot be captured by my methods, so in the next chapter I consider generalising my methods to allow for networks with disconnected components.

[8] This is largely due to the preprocessing steps applied to the network, since I am only considering the giant component.

Chapter 4

Node importance

4.1 Chapter summary

This chapter expands on the work presented in chapter 3 to adapt my methods to the importance of nodes in such a way that allows for potentially disconnected components present in the network. The main goals of this chapter are as follows:

1. To extend my spectral edge importance method to instead consider nodes.
2. To improve my measures of spectral importance to allow for disconnected components in a network.
3. To assess the predictability of a node's behaviour given its importance.

To achieve these goals, I first of all build on my measure of edge importance to derive an equivalent measure for the importance of nodes, before demonstrating how making use of further components of the eigenspectrum in addition to the leading component allows my measure of importance to become 'community aware'. I then assess the predictability of subsequent node presence from the value of node importance, along with measures of centrality and other node level quantities. I demonstrate how my measure of importance is able to outperform other measures in predicting subsequent node presence in equity transaction networks.

4.2 Introduction

Following my exploration of edge importance in chapter 3, the natural next step is to consider node importance. From the perspective of financial regulation, being able to identify important nodes is vital to ensure that these market participants are adequately monitored to minimise the risk they pose to the overall system. Important nodes can be identified by considering concepts such as 'centrality', for which there are a number of measures that rank nodes according to their position in the network [121, 126].

Markets are often characterised by a wide range of different participant behaviours, manifesting in transaction networks displaying complex structures with both communities and dominant nodes, and wide ranges of transaction values and trading frequencies [188, 189]. For a measure of node importance to provide useful insight to policy makers, it needs to account for these complexities. Furthermore, it should provide information on how the network would react to changes in the node's activity. For this

reason, in this study I derive a measure that can be calculated from a static snapshot of a temporal network, but which considers how a change in a node's strength would affect the subsequent structure of the network, which I characterise in terms of the eigenvectors and eigenvalues of the network's weighted adjacency matrix as in chapter 3. I show that my measure of importance can be used for networks with complex community structures and high heterogeneity of nodes' strengths. I demonstrate that it provides an indication of importance in financial transaction networks, where a key concern is the impact an individual would have on the system if they become unable to continue their current level of participation [190].

To bring me a step closer to understanding real networks and their stability, I also pose the question of whether important nodes are more or less likely to continue their market participation in a dynamical setting. I address this through the application of my methods to the same networks analysed in chapter 3, formed from transactions of individual equities traded on the UK capital markets. I consider daily snapshots of transaction networks and perform a classification exercise using a logistic regression model to predict which nodes will appear in the next snapshot given the historical behaviour of the network. This model is a probabilistic model which describes the probability of a node to transact at the next timestamp, so is contributing to the growing body of research exploring the temporal aspects of financial networks. My results show that the measure of node importance I propose in this chapter can predict nodes being present in the next time snapshot better than other importance measures, including two widely used measures of centrality and the frequency of a node's previous transactions.

These results indicate that in the context of these equity networks, defining 'importance' in terms of how a change occurring will affect the subsequent network structure whilst accounting for communities and disconnected components provides useful insights into the role of network structure in the evolution of these networks. This highlights the importance of additional research in this area to further understand how network structure relates to stability, particularly in the context of financial networks.

Taking into account examples in the literature reviewed in section 2.1 of the entire network spectrum and its relevance to network community structure and dynamics, I first provide a measure of importance for nodes based on the spectrum. Then I look at whether this measure is predictive of nodes being present in the subsequent snapshot in the context of equity networks, to understand whether I would expect important nodes in these networks to show lower or higher activity. This in turn will help build an understanding of the roles that nodes of differing importance play in establishing the stability of these systems as a whole. A key thing to highlight is the simplicity of my methods - both in their use of the spectrum of the adjacency matrix itself and in the use of snapshots to capture temporal information. Moreover, the results I now present are significant and meaningful despite this simplicity, suggesting that I have uncovered fundamental findings about the behaviour and evolution of financial networks.

4.3 Proposed method

4.3.1 Defining structural node importance

As shown in chapter 3, I can make use of the derivative of a network's leading eigenvalue with respect to adjacency matrix components as a measure of edge importance:

$$l_e = \frac{\partial \lambda}{\partial A_{ij}} = 2x_{0,i}x_{0,j}, \tag{4.1}$$

where $e \equiv ij$ denotes each edge, λ refers to the leading eigenvalue, A_{ij} is the ijth component of the (weighted) adjacency matrix, and $x_{0,i}$ is the ith component of the eigenvector corresponding to the leading eigenvalue. This was derived considering small perturbations to the adjacency matrix. Through

application of the chain rule, measures for structural node importance can be derived based on the derivative of the eigenvalue of the adjacency matrix with respect to an individual node's strength, where a node's strength S_i is the sum of the weights attached to that node[1]. I do this below for undirected and directed networks respectively[2].

Undirected case

To derive the equivalent to equation 4.1 for node importance, I can again consider perturbations to the adjacency matrix to find the derivative with respect to node strength, $\frac{\partial \lambda}{\partial S_i}$. However, in contrast to chapter 3, my perturbation now consists of adding a fixed amount to each node's strength S_i:

$$S_i \to S_i + \epsilon. \tag{4.2}$$

For this to occur, the change to a node's strength is distributed across its edges. So now, if I apply a perturbation $\epsilon \mathbf{V}$ to the adjacency matrix,

$$A_{ij} \to A_{ij} + \epsilon V_{ij}, \tag{4.3}$$

where

$$V_{ij} = \begin{cases} \frac{A_{ij}}{S_k} & \text{if } i = k \text{ or } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

The perturbation approach then proceeds as follows: First, consider a perturbation to the adjacency matrix $\mathbf{A}$:

$$\mathbf{A} \to \mathbf{A} + \epsilon \mathbf{V}, \tag{4.5}$$

and the resulting first order changes to the leading eigenvalue λ and the associated eigenvector $|\lambda\rangle$ [3]:

$$\lambda \approx \lambda_0 + \epsilon \lambda_1 \tag{4.6}$$

$$|\lambda\rangle \approx |\lambda\rangle_0 + \epsilon \, |\lambda\rangle_1 \, . \tag{4.7}$$

Substituting these into the eigenvalue equation

$$\begin{aligned} &(\mathbf{A} + \epsilon \mathbf{V})(|\lambda\rangle_0 + \epsilon \, |\lambda\rangle_1) \\ &= (\lambda_0 + \epsilon \lambda_1 + ...)(|\lambda\rangle_0 + \epsilon \, |\lambda\rangle_1 + ...), \end{aligned} \tag{4.8}$$

and considering terms up to 1st order in ϵ

$$\begin{aligned} &\mathbf{A} \, |\lambda\rangle_0 + \epsilon \mathbf{V} \, |\lambda\rangle_0 + \epsilon \mathbf{A} \, |\lambda\rangle_1 \\ &= \lambda_0 \, |\lambda\rangle_0 + \epsilon \lambda_1 \, |\lambda\rangle_0 + \epsilon \lambda_0 \, |\lambda\rangle_1 \, . \end{aligned} \tag{4.9}$$

Then consider each of the terms in ϵ^n separately,

$$\epsilon_0 : \mathbf{A} \, |\lambda\rangle_0 = \lambda_0 \, |\lambda\rangle_0 \tag{4.10}$$

$$\epsilon_1 : \mathbf{V} \, |\lambda\rangle_0 + \mathbf{A} \, |\lambda\rangle_1 = \lambda_1 \, |\lambda\rangle_0 + \lambda_0 \, |\lambda\rangle_1 \, . \tag{4.11}$$

Multiplying the equation for ϵ^1 by the left eigenvector ${}_0\langle\lambda|$ and making use of the Hermitian properties

[1] In an unweighted network, node strength is equivalent to node degree.
[2] I consider the undirected case only when applying in section 4.4.
[3] Note that I have switched to Dirac notation for conciseness for the rest of the derivation.

of $\mathbf{A}$ such that ${}_0\langle\lambda|\mathbf{A} = \lambda_0\,{}_0\langle\lambda|$, results in

$$ {}_0\langle\lambda|\mathbf{V}|\lambda\rangle_0 = \lambda_1\,{}_0\langle\lambda|\lambda\rangle_0\,. \tag{4.12} $$

Expanding the indices of this and considering the specific perturbation in equation 4.4,

$$ \begin{aligned} &\sum_{ij} x_{0,i} V_{ij} x_{0,j} \\ &= \sum_{ij} x_{0,i} \frac{A_{ij}}{S_k} x_{0,j} \delta_{ik} + \sum_{ij} x_{0,i} \frac{A_{ij}}{S_k} x_{0,j} \delta_{kj} \\ &= \frac{2}{S_k} \sum_j x_{0,k} A_{kj} x_{0,j}. \end{aligned} \tag{4.13} $$

Here $x_{0,i}$ is the ith component of the eigenvector $|\lambda\rangle$ corresponding to the leading eigenvalue of $\mathbf{A}$ and I have evaluated the δ terms and relabelled the indices. From this, I find the derivative of the eigenvalue with respect to node strength:

$$ \frac{\partial\lambda}{\partial S_i} = \frac{\partial\lambda}{\partial(\sum_j A_{ij})} = \frac{2}{S_i} \sum_j x_{0,i} A_{ij} x_{0,j}. \tag{4.14} $$

Directed case

In the case of a directed network $\mathbf{A}$, the perturbations to the matrix either correspond to changes to in strength or out strength, and I do not need to perturb the matrix symmetrically. Further to this, in contrast to the above, $\mathbf{A}$ is not Hermitian and so I cannot use that $\mathbf{x}^T\mathbf{A} = \lambda\mathbf{x}^T$. However, the matrix product $\mathbf{M} = \mathbf{A}\mathbf{A}^\mathbf{T}$ is symmetric and Hermitian.

The edge level result for the directed case from chapter 3 is

$$ \frac{\partial s^A}{\partial M_{ij}} = \frac{x^M_{0,i} x^M_{0,j}}{2s^A}, \tag{4.15} $$

where $x^M_{0,i}$ refers to the ith component of the eigenvector of $\mathbf{M}$ corresponding to the leading eigenvalue of $\mathbf{M}$, which is also also known as the singular value of $\mathbf{A}$, s^A. I can again relate to the strength by considering a Taylor expansion of the matrix $\mathbf{A}$

$$ A_{ij} = A^0_{ij} + \epsilon A^1_{ij} + \epsilon^2 A^2_{ij}, \tag{4.16} $$

which means that to 1st order,

$$ M_{ij} = \sum_k (A^0_{ik} + \epsilon A^0_{ik})(A^0_{jk} + \epsilon A^0_{jk}) \tag{4.17} $$

$$ = M^0_{ij} + 2\epsilon M^0_{ij}, \tag{4.18} $$

so

$$ \frac{\partial M_{ij}}{\partial\epsilon} = 2M_{ij}, \tag{4.19} $$

which gives my result when applying the chain rule as above

$$\frac{\partial s^A}{\partial S_i} = \frac{1}{S_i s^A} \sum_j x_{0,k} x_{0,j} M_{ij}. \tag{4.20}$$

To summarise my final results of these derivations, I present equations 4.21 and 4.22 for undirected and directed networks respectively:

$$m_i = \frac{\partial \lambda}{\partial S_i} \equiv \frac{2}{S_i} \sum_j x_{0,i} x_{0,j} \tag{4.21}$$

$$m_i = \frac{\partial s^A}{\partial S_i} \equiv \frac{1}{S_i s^A} \sum_j x_{0,i} x_{0,j} M_{ij}. \tag{4.22}$$

Here m_i denotes the importance of node i. I note here that my measure of node importance is, by design, inversely proportional to node strength. Although this is in contrast to measures of centrality, here I am defining importance by considering an individual node experiencing a fixed size change to its strength, meaning that a more connected node will distribute its change across more edges, having a smaller effect on each of its neighbours individually. An alternative definition of importance could consider fixed changes to each edge, effectively producing the inverse of my defined measure. However, for my application to financial transaction networks, it is important to understand the scenario in which a participant in the market experiences a decrease in its available inventory and how this impact will propagate to its neighbours. A well connected node in a network will have the option of spreading this impact across multiple trading relationships, whereas a poorly connected node will present a larger risk to its counterparties. For this reason, in this chapter, I define importance from the perspective of fixed changes to node strength. In both the directed and undirected case, it is also worth noting that the derivations can be generalised to allow new links to be added/removed, but new nodes cannot be added or removed.

Extension of node importance method

Although in chapter 3 I considered only perturbations to the leading eigenvalue and its associated eigenvector, equations 4.21 and 4.22 are relevant for any of the single components of the eigenspectrum. Later in section 4.4 I demonstrate how different parts of the networks considered in this chapter relate to different parts of the eigenspectrum and propose that my methods can be made 'structurally aware' through the use of multiple components of the eigenspectrum. First, I note that care must be taken in identifying the relevant eigenvector from the eigenspectrum of the network.

Toy network exploration of network spectrum

Here I briefly explore whether the use of multiple components of network spectra can be used to capture different structures in networks through the use of a toy network. I consider a barbell network with two unevenly sized cliques joined by a bridge, shown in figure 4.1, in order to observe how the different components of the eigenspectra are relevant for the different communities present in this network. Table 4.1 shows the eigenvector values corresponding to the 3 positive eigenvalues of the adjacency matrix. If I consider the nodes in the largest clique (top right in figure 4.1, nodes 6 to 10), I see that the largest eigenvector components are seen for the eigenvector corresponding to the leading eigenvalue (eigenvector 1). Considering nodes 0 to 3 (in the bottom left clique), I see that the largest magnitude eigenvector components are seen for eigenvector 2. The nodes in the bar (nodes 4 and 5) both show the largest component for eigenvalue 3. I further support these observations through the use of a k-means

Node	Eigenvector 1	Eigenvector 2	Eigenvector 3
0	0.006	-0.478	-0.159
1	0.006	-0.478	-0.159
2	0.006	-0.478	-0.159
3	0.013	-0.524	0.121
4	0.033	-0.189	0.629
5	0.122	-0.060	0.658
6	0.463	0.002	0.187
7	0.439	0.016	-0.106
8	0.439	0.016	-0.106
9	0.439	0.016	-0.106
10	0.439	0.016	-0.106

Table 4.1: Eigenvector values corresponding to the three positive eigenvalues for the barbell network displayed in figure 4.1. Nodes 0 to 3 are the nodes within the top left clique, nodes 4 and 5 make up the bridge and nodes 6 to 10 are the nodes within the bottom right clique.

clustering, applied to the three positive eigenvectors, which resulted in the clustering of the nodes shown by the different colours in figure 4.1, demonstrating that the different eigenvectors have relevance for the different communities present in the network.

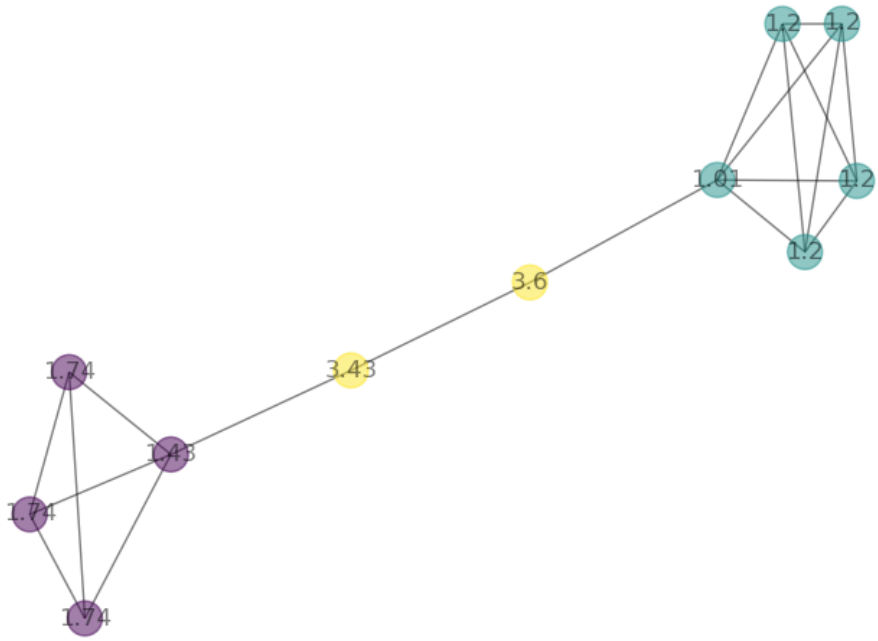


Figure 4.1: Barbell network, nodes coloured by result of k-means run on the eigenvectors corresponding to the positive eigenvalues of the adjacency matrix. Nodes are labelled by the value of the measure m_b.

I have shown through this example that the n'th largest community is found to correspond to the n'th largest eigenvalue and its eigenvector and that the magnitude of the components of this eigenvector for the given community will be larger than the components for the other eigenvectors. So I expect that by taking the largest magnitude eigenvector components corresponding to the nodes in the community as the 'correct' eigenvector components to represent the nodes in that community, my measure will be 'community aware'. To assess this, I propose extending my structural importance metric to make use of the spectrum in one of four ways:

1. Only make use of the leading eigenvalue and its associated eigenvector in equation 4.21. This measure is expected to perform well when there are no communities present. I will refer to this as m_a.

2. Identify, for each node, the eigenvector with the largest magnitude component for that node and the eigenvalue associated with this and use these to compute equation 4.21 for each node. I will refer to this as m_b.

3. To understand whether node importance has meaningful contributions from all parts of the spectrum, consider importance as the sum of equation 4.21 for all eigencomponents. I will refer to this as m_c.

4. Consider as for m_c, but only make use of the part of the eigenspectrum with positive eigenvalues. I will refer to this as m_d.

The nodes in figure 4.1 are also labelled with the individual m_b node importances. This shows that the nodes making up the bridge, which themselves have very few connections but connect the communities, are the most important, and the nodes in the larger clique are the least important. This suggests that the measure m_b is performing as expected, as a fixed change to a node's strength when the node is present in the bridge would have a larger impact on the rest of the network than for a node in a clique.

It is worth noting that m_b is not suitable for use on random networks, as in this case, no single entry of the eigenvector would be relevant for each node, so there is no guarantee that the eigenvector with the largest entry for each node is the correct part of the spectrum for that node. This restricts my method to the application of networks which are known to have a non-random structure.

4.3.2 Node importance and network evolution

It is intuitive to explore how importance in a static sense relates to future activity in a network, since an importance measure is only useful in practice if it is able to provide actionable information. This complements several studies which have used node or edge importance to explain structural changes in real network systems [125, 191, 192], and in particular financial systems [173, 174, 1, 168].

When considering the use of my measures of importance in understanding real networks and their stability, I can consider whether important nodes are more or less likely to be present in the subsequent snapshot given their current importance. To do this, I make use of logistic regression to predict subsequent node presence from historical feature vectors. The logistic regression I consider can be interpreted as a probabilistic model that gives the probability that a node subsequently transacts given historical properties:

$$P(Y = 1|\mathbf{X} = \mathbf{x}) = \frac{\exp\left(\boldsymbol{\alpha} + \boldsymbol{\beta}^T \boldsymbol{x}\right)}{1 + \exp\left(\boldsymbol{\alpha} + \boldsymbol{\beta}^T \boldsymbol{x}\right)}, \tag{4.23}$$

where Y is my target variable taking a binary value per node, $\mathbf{X}$ is a vector of feature values, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the regression model coefficients. The feature vectors consist of my node importance measures, along with eigenvector and pagerank centralities as benchmark measures, and other node level attributes, degree and community. These features are calculated as averages over all the previous time periods in the data available prior to the snapshot in question, to answer the question 'given what I know about the network up to today, what do I know about tomorrow?'. I also include a further feature of the number of times that a node has been present in the network prior to the snapshot as a benchmark to compare my measure to. For my first experiment the target variable for the classifier is a binary variable indicating whether a node that is present in the current snapshot is also present in the next snapshot. Since I observe fairly high levels of class imbalance across the datasets considered

[4], I apply a random over-sampling strategy to correct this for all three datasets. I make use of 5-fold cross validation [5] to select the best classifier and its associated parameter values[6]. This not only allows me to assess whether my measures of node importance are predictive of subsequent node activity, but it also provides me with the means of comparing the different entries of the feature vector through their feature importances. To do this, I make use of permutation importance, which calculates the increase in the model's prediction error after permuting the feature [193]. Since this measure is only able to capture the importance of features in a global sense without accounting for the role of a feature in individual predictions, I also make use of Shapley values, which use concepts from co-operative game theory to explain the additional importance of each variable for each individual observation [194]. More specifically, I make use of the SHapley Additive exPlanations (SHAP) approach, which quantifies the contribution each feature in a Machine Learning model makes to the prediction of individual observations [195]. When using SHAP to explain the probability of a linear logistic regression model, I note that strong interaction effects would affect the performance since the model is not additive in the probability space. To account for this, I use SHAP to explain the log-odds of the model, as there is a linear relationship between the model's inputs and this output [195][7]. When evaluating the overall performance of the classifiers, I make use of precision, recall and the ROC AUC.

I benchmark the results against a null model consisting of the average over 100 trials in which edges are randomly present with probability equal to the fraction of observed edges. For this benchmark, the confidence intervals were determined empirically by discarding the top and bottom 5% of the precision and recall scores, and a 'coin toss' approach was taken to calculate the confidence intervals for the model precision and recall thresholds such that the probability of observed outcome is higher than some threshold α. The method for this can be found in the appendix. In practice, the Central Limit Theorem can be used when there is a sufficiently large number of samples, since the sum of random variables will closely follow a normal distribution, allowing me to make use of the confidence intervals of a normal distribution. When comparing to the null model results in this chapter, I also present empirically calculated confidence intervals and average precision and recall for my null model, which randomly predicts 1 or 0 in proportion to the dataset prior. The confidence intervals for the ROC AUC were calculated using a bootstrap approach, with 1000 iterations of random sampling with replacement from the training dataset.

4.4 Applications

4.4.1 Application to individual equity stocks

The bulk of the results presented in this chapter focus on the identification of important nodes in transaction networks of the same three different equity stocks traded on the UK capital markets considered in chapter 3 and explored in section 2.2.4. Later in this chapter I also include an application of my methods to an open source dataset of inter-country trades in financial services, created by the OECD and WTO [112].

[4] Equity-1 shows 1499 of 2063 present in the subsequent snapshot, Equity-2 shows 724 of 880 present, and Equity-3 shows 1803 of 2237 present

[5] For this, I split my data into training and test sets, with a 40-40-20 train, validation, test split. I split the data whilst keeping the ordering of time, so that the model is not trained on data from the future.

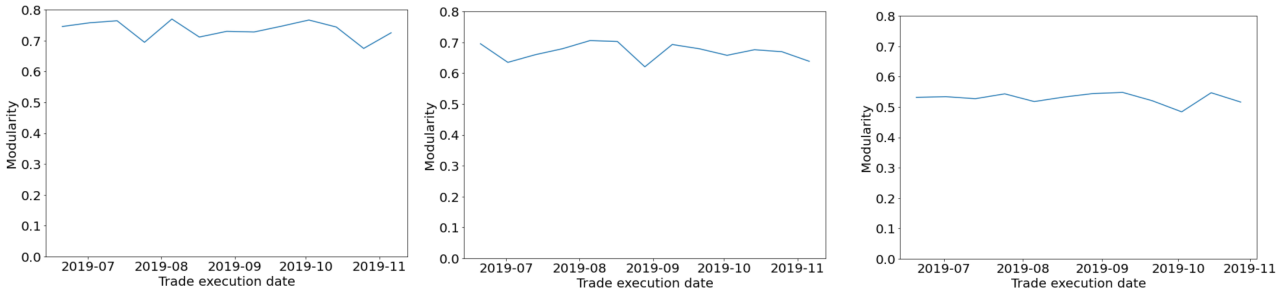
[6] Both logistic regression and random forest classifiers were considered to allow for potentially non-linear relationships. In practice, the logistic regression model consistently performed the best, so the results presented in this chapter are for logistic regression classifiers.

[7] The results of using SHAP can be found in the appendix in figure 19

Exploration of network community structure

First, to build my understanding of how the equity transaction networks evolve, I explore how the community structure varies across time. I also explore how the distributions of various node level measures differ for nodes that do appear in subsequent snapshots to nodes that don't. The former is considered since I am proposing a measure of node importance that is able to capture community structure, so I first need to verify that the networks I am considering consistently show a community structure. The latter provides me with an indication of which node level measures I might expect to provide information on network evolution, and also to help indicate whether different classes of nodes are more, or less likely to subsequently transact.

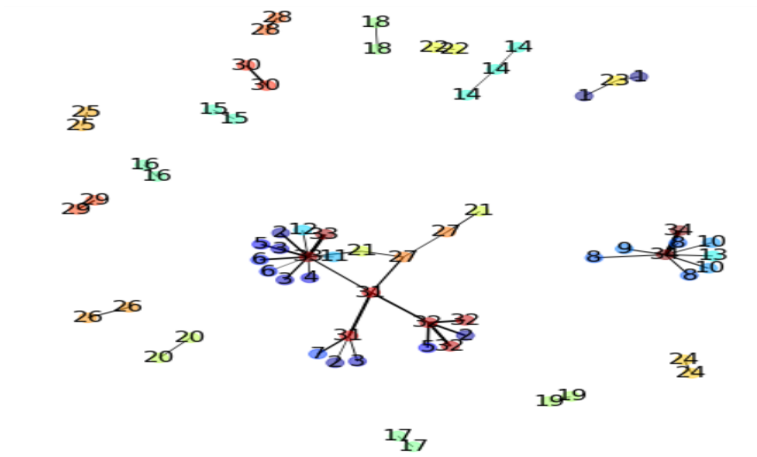
Figure 4.2 examines how the networks evolve across time, by considering the variation in the modularity (the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random [196]). Aggregating on a half-monthly basis to reduce noise corresponding to days of very low trading activity at weekends and public holidays, I see that all three networks have a largely static modularity, with all showing similar average modularity of around 0.5-0.7. This is suggestive of a meaningful community structure that does not vary significantly across the observation period. The network with the lowest modularity is Equity-3, which is in agreement with what is seen when visually exploring these networks in figures 4.3a-4.3c, as this network consists of one large connected component, with the smaller disconnected components observed for the other two networks not present in this dataset.



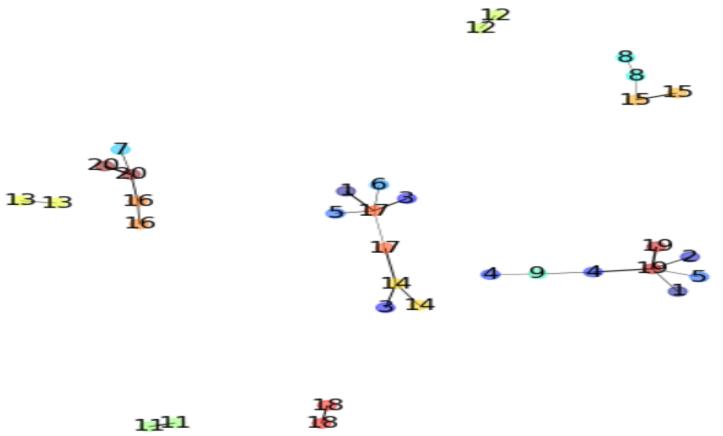
(a) Equity-1 (b) Equity-2 (c) Equity-3

Figure 4.2: Modularity across time for the three different equities networks, for the full observation period aggregated on a half monthly basis.

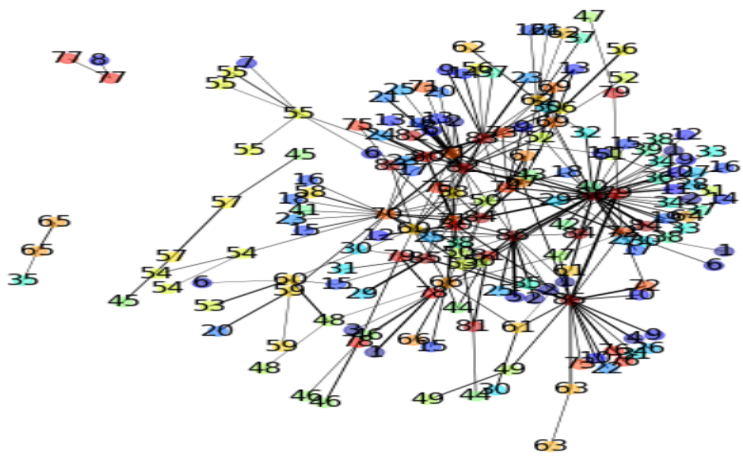
I also consider how the different components of the eigenspectra relate to the communities of the networks. I do this by exploring how nodes rank by eigenvalue if I select for each node the eigenvalue with the largest magnitude eigenvector component for that node. In figures 4.3a, 4.3b and 4.3c, the nodes are coloured and numbered by the rank of the eigenvalue that is selected (rank 1 corresponds to the largest eigenvalue). These show in all networks that nodes within small communities often, but not exclusively, select the same eigenvalue, and that nodes playing similar roles within the network show similar ranks for their eigenvalue. Where nodes in the same connected component select different eigenvalues, hub nodes select higher ranked eigenvalues. This suggests that if I make use of my measure of importance in equation 4.21 whilst selecting the most relevant eigencomponent, this would assign a larger importance to these nodes. However, this will be partially counteracted by the inverse strength factor in my structural importance measure, which makes sense since a node in a small but well connected community has few direct neighbours to spread the impact of a change in strength between but will have a high reachability to other nodes overall.



(a) Equity-1

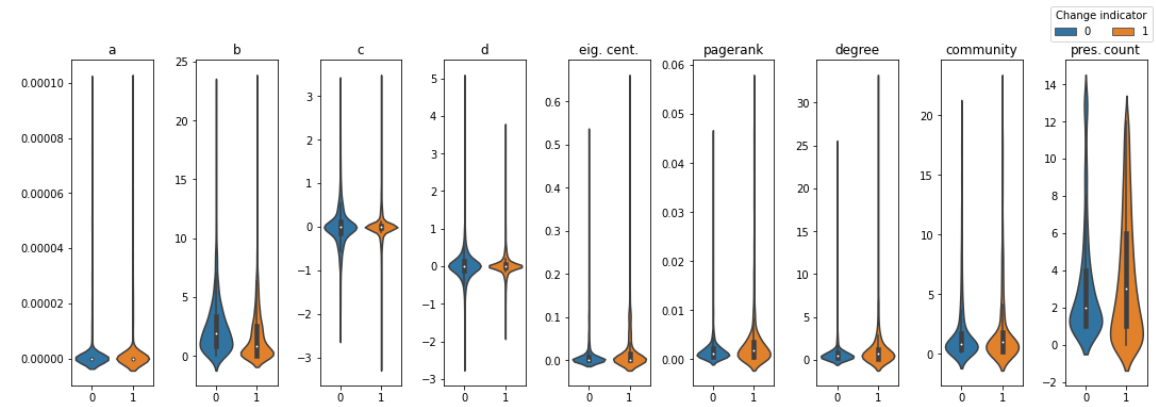


(b) Equity-2

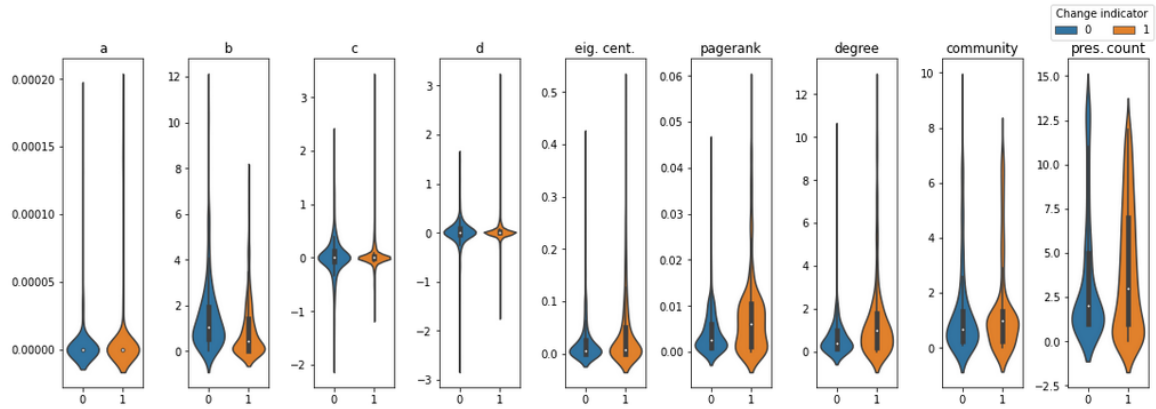


(c) Equity-3

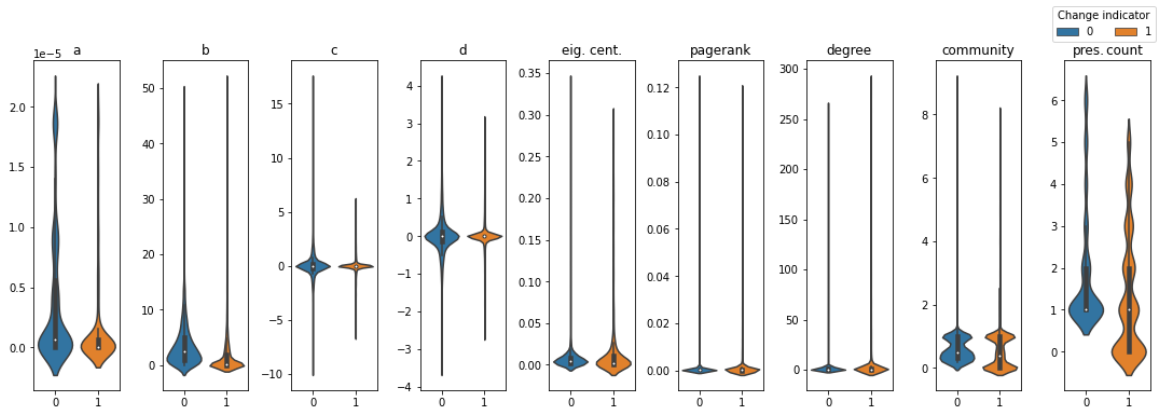
Figure 4.3: Initial snapshot networks for the three equities datasets, colours and numbers representing the ranking of the eigenvalue that corresponds to the eigenvector with the largest magnitude for each node.



(a) Equity-1



(b) Equity-2



(c) Equity-3

Figure 4.4: Distributions of the different node importance measures across nodes which are subsequently present in comparison to those that are subsequently absent.

Figure 4.4 shows the distributions of the values of m_{a-d} computed according to the 4 different eigenvalue inclusion schemes along with the two benchmark measures of node importance, pagerank and eigenvector centrality and also degree, community label and the number of times a node has been present in the historical data (presence count). I use violin plots to present the distributions, which show the kernel density estimated distribution plotted on top of a boxplot showing the mean and interquartile range. The plots are split by whether or not nodes are subsequently present in the network. I see here that m_b, which selects the relevant eigenvalue component for each node, visually shows the largest difference in the distribution mean for present nodes in comparison with absent nodes across the three datasets.

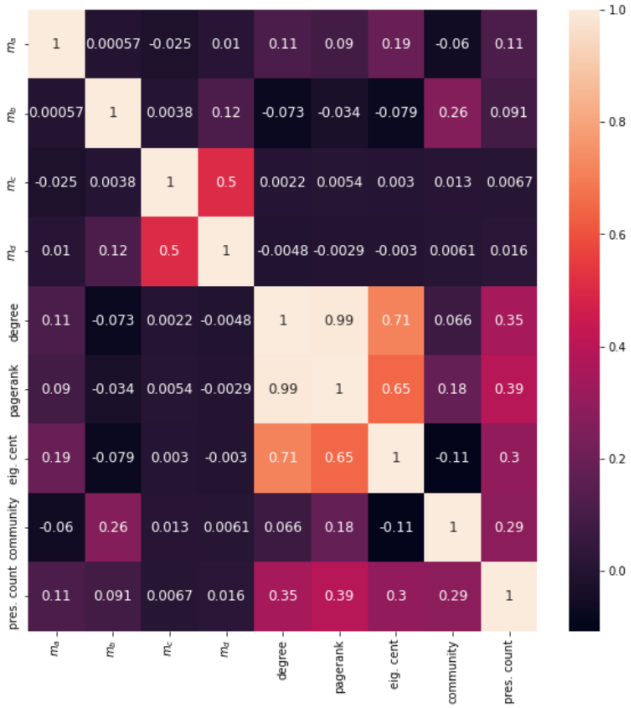
Measure	Equity-1	Equity-2	Equity-3
m_a	3.23×10^{-11}	8.17×10^{-2}	5.82×10^{-63}
m_b	1.34×10^{-4}	3.96×10^{-13}	2.42×10^{-135}
m_c	1.18×10^{-1}	8.47×10^{-1}	3.62×10^{-1}
m_d	5.09×10^{-1}	8.23×10^{-1}	4.62×10^{-1}
Community	1.76×10^{-21}	1.03×10^{-22}	2.57×10^{-2}
Degree	9.53×10^{-49}	2.36×10^{-13}	4.62×10^{-10}
Eig. cent.	4.73×10^{-37}	42.99×10^{-8}	9.89×10^{-12}
Pagerank	3.51×10^{-58}	1.56×10^{-14}	5.25×10^{-5}
Pres. count	3.72×10^{-53}	1.22×10^{-2}	9.78×10^{-52}

Table 4.2: p-values for a two-sided t-test for the differences in the mean values for nodes which are present and nodes which are absent for each of the different node level measures, for the equity transaction datasets.

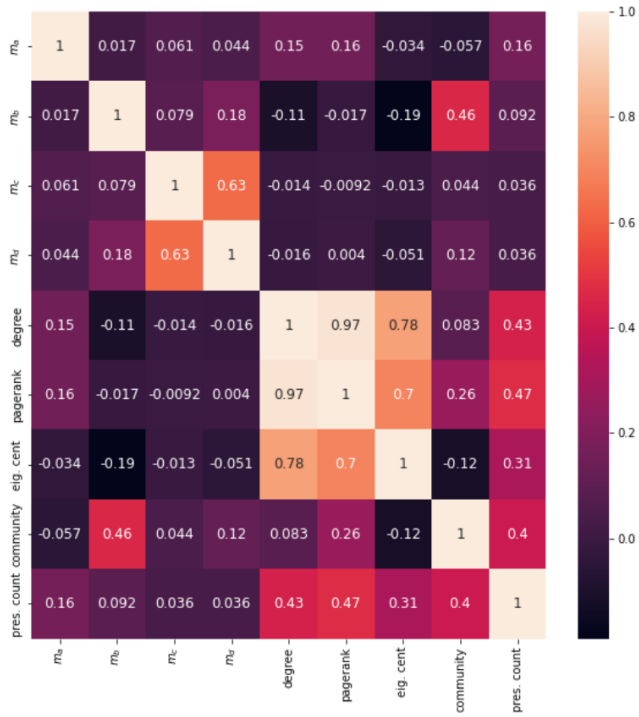
I also observe that nodes that are subsequently present are observed with smaller values of m_b, in contrast to eigenvector centrality and pagerank, which both show changing nodes having slightly larger values. Table 4.2 shows the p-values for a two-sided t-test for the differences in the mean values for presence vs. absence of nodes for each of the different measures. I see that m_a m_c, and m_d do not show $p < 0.001$ [8] for all datasets. This non-significant difference in the mean values for changing vs. unchanging nodes suggests that I would not expect these measures to be predictive of subsequent node presence. On the other hand, m_b, community, degree, eigenvector centrality, pagerank and presence count all show significant differences in the mean values for all datasets, making these measures better candidates for prediction of subsequent node presence.

In order to assess the similarities between the different measures, and also to ensure that any predictive model is not impacted by large correlations between the features, I consider the Pearson correlations between the rankings of nodes according to the different measures, shown in figure 4.5. In general across all three datasets, I see that the measures m_a, m_b, community and presence count show no significant correlations with any other measures. For Equity-1 and Equity-2, m_c and m_d are moderately correlated with each other, which is expected since the two measures differ only in their use of the part of the spectra with negative eigenvalues for which the eigenvector components will be small. For Equity-1, high correlations were observed between degree and both pagerank and eigenvector centrality, so degree was not included in the feature vector for the classifier for this dataset. For the other two datasets, high correlations were observed between pagerank, degree, and eigenvector centrality, so both pagerank and degree were not included in the feature vector for the classifiers for these datasets. These large correlations are indicative of the dominance of hub nodes in these networks.

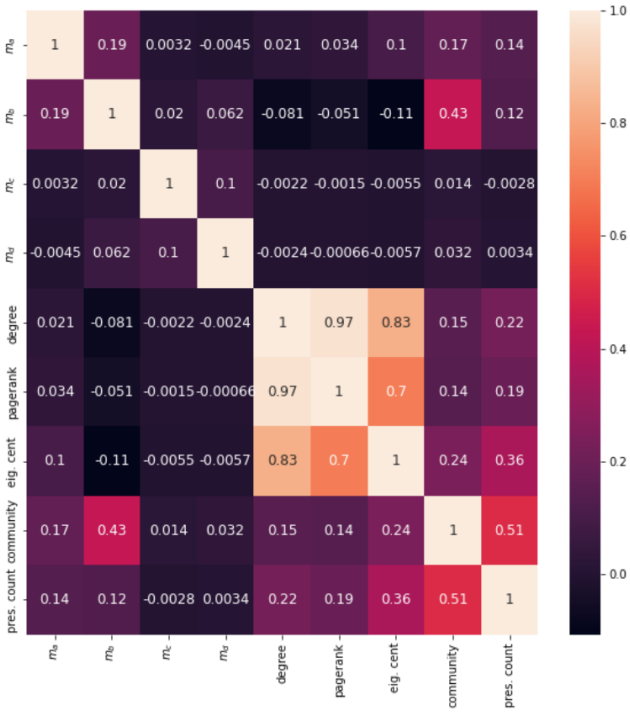
[8] Since I am comparing 9 metrics simultaneously, I have applied a Bonferroni correction to the p-value threshold.



(a) Equity-1



(b) Equity-2



(c) Equity-3

Figure 4.5: Pearson correlations between the different node level features, for the different equity transaction datasets, with both colours and labels representing to the correlation value.

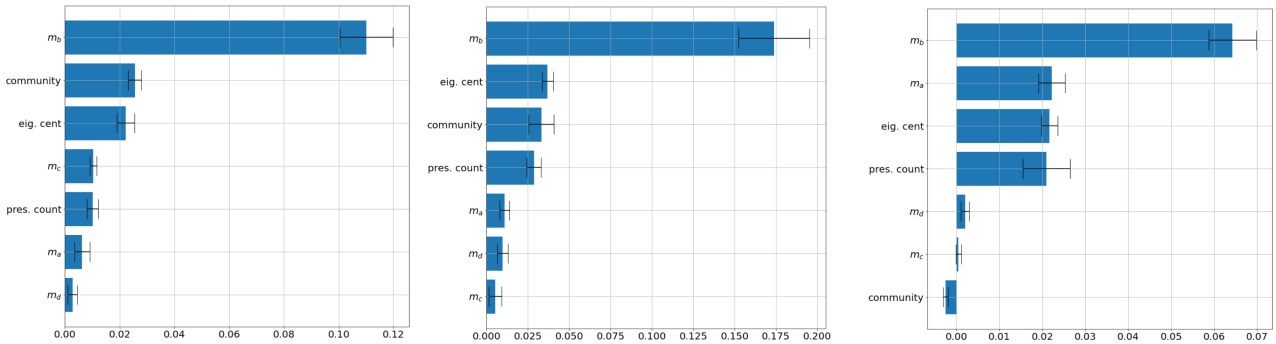
Prediction experiments

I now present the results of the prediction experiments described in section 4.3.2. First, considering the prediction of subsequent node presence from the node level features, the precision and recall of the classifiers when applied to test sets are shown in table 4.3, alongside the performance of the null model. For all three datasets, the prediction showed reasonable precision and recall, which in all cases showed no overlap in the 95% confidence intervals with the null model.

Measure	Equity-1	Equity-2	Equity-3
Precision	0.83	0.79	0.73
CI	(0.78,0.87)	(0.72,0.85)	(0.70,0.76)
P (N.M.)	0.54	0.60	0.57
CI	(0.47,0.59)	(0.52,0.69)	(0.53,0.62)
Recall	0.66	0.81	0.72
CI	(0.61,0.71)	(0.75,0.88)	(0.69,0.75)
R (N.M.)	0.49	0.51	0.50
CI	(0.29,0.60)	(0.21, 0.75)	(0.28,0.69)
AUC	0.66	0.61	0.76
CI	(0.64, 0.67)	(0.61, 0.63)	(0.75, 0.76)
AUC (N.M.)	0.5	0.5	0.5
CI	(0.47, 0.54)	(0.44, 0.55)	(0.47, 0.54)

Table 4.3: Precision, recall and ROC AUC for the classification model for the 3 different datasets, presented alongside the same average precision and recall for the null model (N.M.) trials. The brackets denote the 95% confidence intervals (CI).

Figure 4.6 shows the permutation importance for the different features. I observe that the measure m_b



(a) Equity-1 (b) Equity-2 (c) Equity-1

Figure 4.6: Logistic regression permutation importance for the different node level features. The error bars represent the standard deviation of the importances across the different trees that make up the model.

is by far the most important feature in the prediction across all three datasets considered. Although for Equity-3, m_a and eigenvector centrality are moderately important, as would be expected in a network with a single connected component, along with the presence count, none of the other node level measures are consistently important across all 3 datasets. I also make similar observations in the appendix using SHAP values to assess feature importance in figure 19. In figure 4.7, which shows the coefficients of the model, I observe that m_b, community and presence count are the only features that show a consistent sign and approximate size of the parameter, and also p-values of $p < 0.001$. I

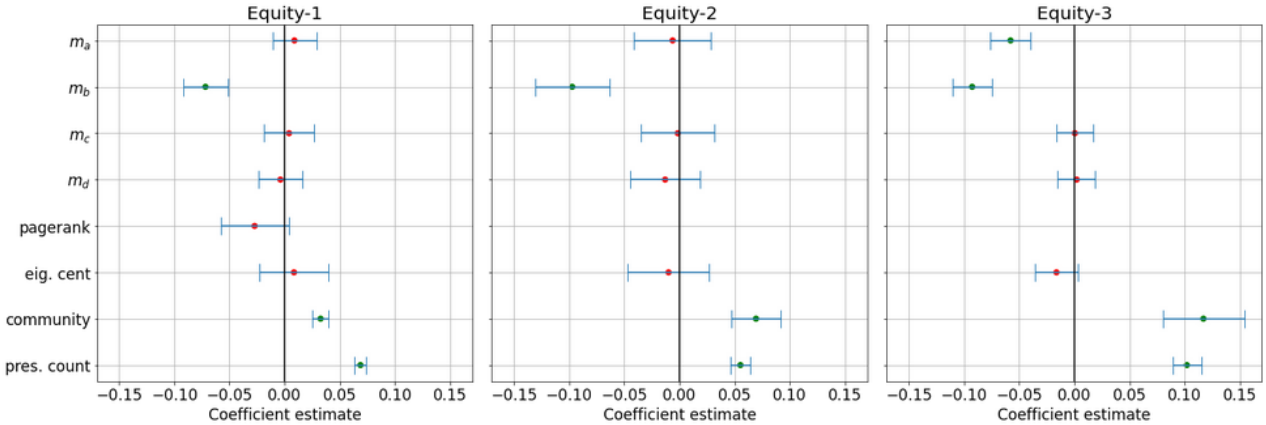


Figure 4.7: Classification model coefficients, with 95% confidence intervals indicated by the error bars. If a p-value is less than 0.001, it is coloured green, otherwise red.

also note that m_b has the largest magnitude coefficient[9] for both Equity-1 and Equity-2, which is in agreement with the feature importances in figure 4.6. If the parameter m_b is used as the only feature in the model, I observe that the coefficient is consistently negative and significant, which indicates that nodes that are more important are less likely to subsequently transact.

When considering the exercise of predicting whether or not nodes are subsequently present from the different measures, in order to validate that the measure m_b is the most predictive, I also re-run the model with the measure m_b removed. The results for this are found in table 4.4.

Measure	Equity-1	Equity-2	Equity-3
Precision	0.77	0.76	0.69
CI	(0.73,0.83)	(0.70, 0.83)	(0.66,0.73)
P (N.M.)	0.67	0.70	0.57
CI	(0.62,0.71)	(0.63, 0.83)	(0.53,0.59)
Recall	0.40	0.5	0.69
CI	(0.36,0.44)	(0.43, 0.57)	(0.66,0.72)
R (N.M.)	0.52	0.51	0.49
CI	(0.27, 0.75)	(0.31, 0.74)	(0.38,0.64)
AUC	0.53	0.60	0.71
CI	(0.50, 0.57)	(0.53, 0.63)	(0.71, 0.72)
AUC (N.M.)	0.50	0.51	0.50
CI	(0.47, 0.54)	(0.45, 0.57)	(0.47, 0.52)

Table 4.4: Precision and recall for the logistic regression for the 3 different datasets, presented alongside the same average precision and recall for the null model (N.M.) trials, when considering the prediction without the feature m_b. The brackets denote the 95% confidence intervals (CI).

I observe that although there is a drop in all precision scores, this only brings the model performance within the confidence interval range of the null model for the Equity-2 dataset. The recall and ROC AUC also drop for all datasets, falling for both the Equity-1 and Equity-2 datasets to be within the confidence interval range of the null model. As expected, the model still retains some performance since some of the information captured by the measure m_b is also captured by the other features, as shown by the correlations between the features. In this case, for Equity-1 and Equity-2, degree was found to be the most important of the remaining features, and for Equity-3, m_a was the most important. The

[9]The features are standardised prior to use in the model, which allows for size comparison of the coefficients.

drop in model performance was least prominent for Equity-3, which is as expected due to m_b providing additional information in networks with disconnected components, which is not the case for Equity-3.

For nodes which show a persistence in transacting between snapshots, I also considered whether the sign of the change to nodes' strength is predictable from the different node level features, and the results of this are shown in table 4.5. For this, I observe precision and ROC AUC scores for all three datasets which are not within the confidence intervals of the null model. However, the recall is never outside of the confidence intervals of the null model, so I cannot conclude that the sign of a change is predictable from my chosen node level features. In all cases, the 'presence count' feature was the most important feature in predicting sign change.

Measure	Equity-1	Equity-2	Equity-3
Precision	0.80	0.84	0.93
CI	(0.75,0.84)	(0.77,0.90)	(0.91,0.95)
P (N.M.)	0.68	0.69	0.75
CI	(0.62,0.73)	(0.60,0.76)	(0.71,0.80)
Recall	0.66	0.68	0.75
CI	(0.61,0.71)	(0.61,0.76)	(0.71,0.78)
R (N.M.)	0.49	0.51	0.49
CI	(0.31,0.67)	(0.38, 0.65)	(0.27,0.71)
AUC	0.76	0.74	0.85
CI	(0.74, 0.76)	(0.73, 0.76)	(0.84, 0.85)
AUC (N.M.)	0.50	0.49	0.50
CI	(0.45, 0.54)	(0.41, 0.56)	(0.47, 0.54)

Table 4.5: Precision and recall for the logistic regression predicting the sign of the change in strength for the 3 different datasets, presented alongside the same average precision and recall for the null model (N.M.) trials, when considering the predicting the sign of the subsequent change in strength to a node. The brackets denote the 95% confidence intervals (CI).

Finally, I further consider whether the value of the change in strength is predictable from the node level features, by considering a regression of the features onto the value of the relative change in strength. The results for this are found in table 4.6.

Measure	Equity-1	Equity-2	Equity-3
R^2 score	0.10	0.11	0.10
CI	(0.08,0.16)	(0.08,0.19)	(0.08,0.15)
R^2 (N.M.)	0.12	0.11	0.11
CI	(0.10,0.13)	(0.10,0.13)	(0.10,0.12)

Table 4.6: Coefficient of determination R^2 for a linear regression with endogenous value of the relative change in node strength, exogenous variables the node level features as used in the classification exercise. This is compared to a null model in which the relative change in node strength is randomly shuffled in 100 trials, and the average and 95% confidence intervals (CI) reported.

For all three datasets, the confidence intervals for the R^2 score of the regression overlap with those for a null model in which node strength is randomly shuffled, and that the R^2 values are higher for the null model. This means that I can conclude that the change in strength is not predictable in these networks. When looking at the coefficients of the regression model in figure 4.8, only eigenvector centrality and community show consistent sign of the coefficient across all three datasets, only presence count shows a

high significance for the coefficients across the three datasets, and in general many of the coefficient values are close to 0.

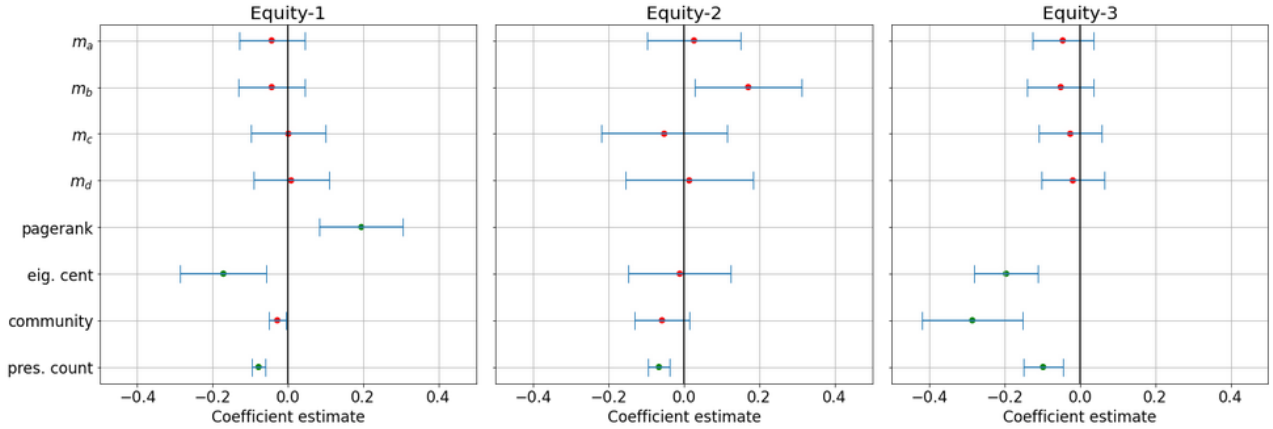


Figure 4.8: Regression model coefficients, with 95% confidence intervals indicated by the error bars. If a p-value is less than 0.001, it is coloured green, otherwise red.

4.4.2 Application to open source data: BaTIS dataset

Since the equity transaction datasets explored above are not publicly available, here I present the results of applying my methods to the Financial Services segment of the BaTIS dataset. In contrast to the equity datasets, the BaTIS dataset has a natural persistence of activity (generally, countries that trade with each other continue to do so year on year) so instead of looking to predict whether or not a node will be present in the subsequent snapshot, I instead look to predict whether or not a node will show a change in strength in the subsequent snapshot[10]. Figure 4.9 shows the distributions of the different node level metrics for nodes that do subsequently change, and those that don't [11]. I also observe how

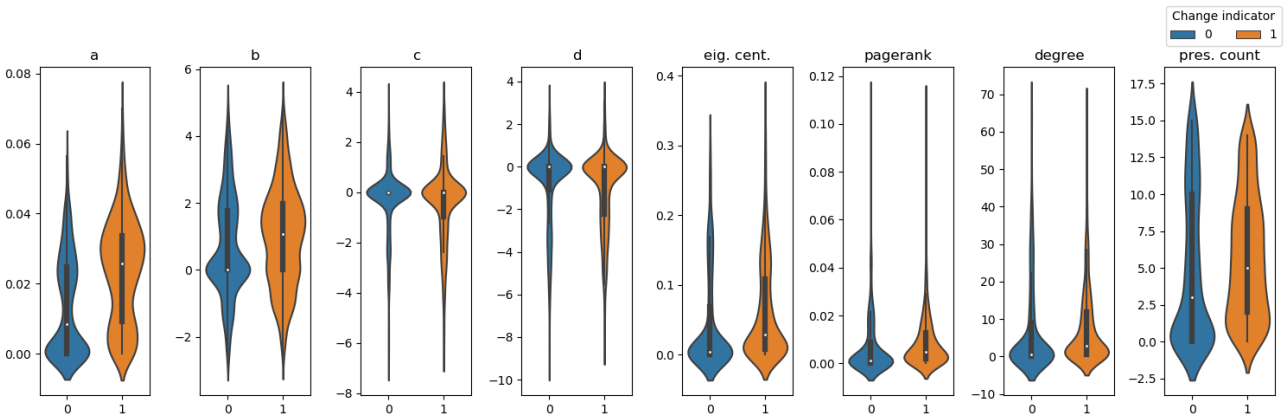


Figure 4.9: Distributions of the different node importance measures across edges which do subsequently change in comparison to those that don't, for the BaTIS trade dataset.

selection of the maximum eigencomponent for each node manifests itself in this dataset in figure 4.10, which shows the initial snapshot of the network with nodes coloured by the rank of the most relevant eigencomponent for that node.

[10]Specifically, I define a significant change to a node's strength as a change of more than 5% between snapshots.
[11]Note that a log transform has been applied to m_b, m_c and m_d due to these features spanning a few orders of magnitude

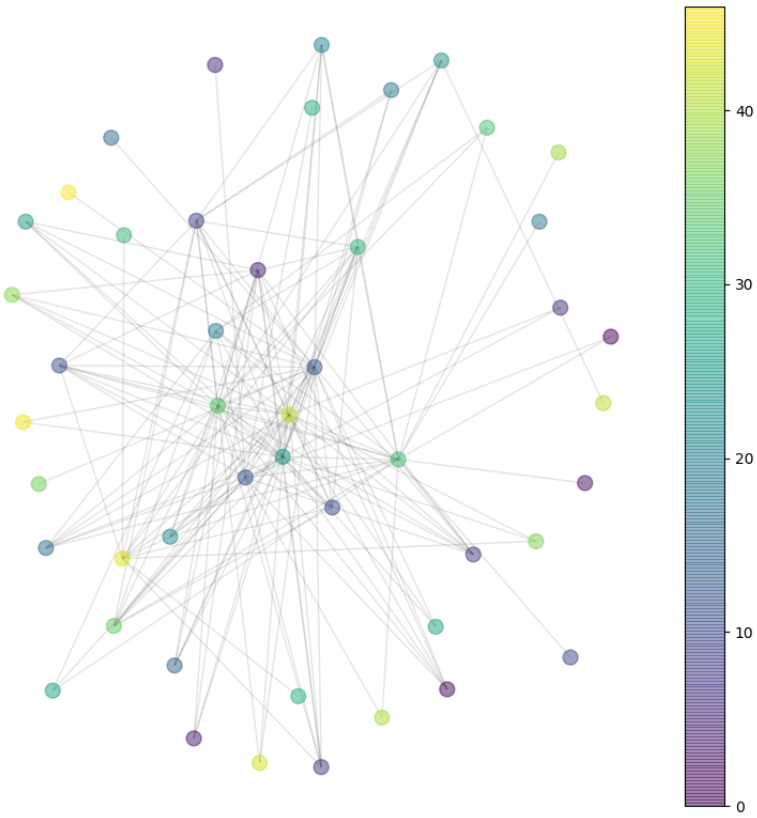


Figure 4.10: Network showing the initial snapshot for the BaTIS dataset, colours and numbers representing the ranking of the eigenvalue that localises to a given node.

In figure 4.11 I present the modularity across time for this dataset, and I see that in comparison to the equity transaction datasets, the BaTIS dataset shows a much lower average modularity, and unlike the equity datasets which showed a stable modularity across time, I see a decreasing trend for this dataset.

Figure 4.11: Modularity across time for the BaTIS dataset, for the full observation period.

This is in agreement with what I see in figure 4.10, in which I observe that the network has just one community, with a small number of peripheral nodes. These peripheral nodes tend to show lower eigenvalue rankings in comparison to the more densely connected core. Looking at figure 4.9, I see that the measures m_a, m_b, eigenvector centrality and presence count show clear differences in the

Measure	BaTIS dataset
m_a	2.62×10^{-47}
m_b	6.34×10^{-6}
m_c	5.83×10^{-2}
m_d	1.71×10^{-3}
Degree	2.19×10^{-1}
Eig. cent.	3.51×10^{-7}
Pagerank	2.79×10^{-4}
Pres. count	1.46×10^{-4}

Table 4.7: p-values for a two-sided t-test for the differences in the mean values for nodes which change and nodes which don't, for each of the different node level measures, for the BaTIS dataset

distributions for changing vs. unchanging nodes. The p-values for a two-sided t-test for the differences in the mean values for change vs. no change for each of the different measures are shown in table 4.7. Here I see that m_a shows the most significant difference in the means, but m_c and degree have p-values >0.001, suggesting that any difference I can visually observe for these variables is not unlikely to have occurred by chance.

Now, prior to considering the role of node importance in change predictability for the BaTIS dataset, I first of all explore the correlations between the different measures used in the predictor, shown in figure 4.12.

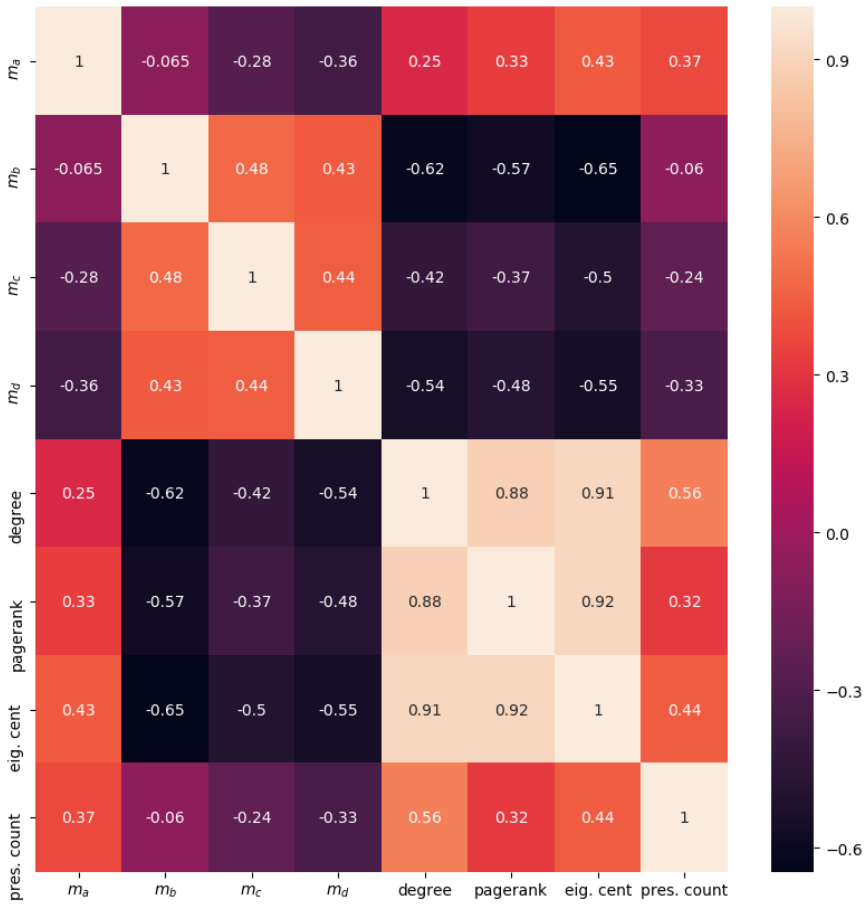


Figure 4.12: Pearson correlations between the different node level features for the BaTIS dataset, with both colours and labels representing to the correlation value.

Here, I see large correlations between degree, and the two centrality measures pagerank and eigenvector centrality, suggesting that these measures are only capturing the node degree as an indicator for importance. The measures m_a, m_b, m_c and m_d show no significant correlations with the other measures, although there is a reasonable negative correlation between m_b and degree, eigenvector centrality and pagerank, which is as expected from my definition of structural importance, particularly in the case when there is only one community in the network.

Prediction experiments

As before, I now present the results of the prediction experiments described in section 4.3.2 for the BaTIS dataset. For the prediction of whether or not a node changes strength given the node level features, the classifier shows good performance in table 4.8 on the test set, with a precision of 0.70, recall of 0.65 and a ROC AUC of 0.71. I observe that although the precision and ROC AUC are better

Measure	BaTIS dataset
Precision	0.70 (0.63,0.76)
Precision (null model)	0.52 (0.45,0.59)
Recall	0.65 (0.59,0.72)
Recall (null model)	0.52 (0.36,0.66)
ROC AUC	0.71 (0.69, 0.71)
AUC (null model)	0.50 (0.46, 0.54)

Table 4.8: Precision and recall for the logistic regression for the BaTIS dataset, presented alongside the average precision and recall for the hull model. The brackets denote the 95% confidence intervals.

than the null model, the recall confidence intervals overlap with those of the dummy model so I cannot conclude that subsequent change to node strength is predictable from the node level features used for this dataset.

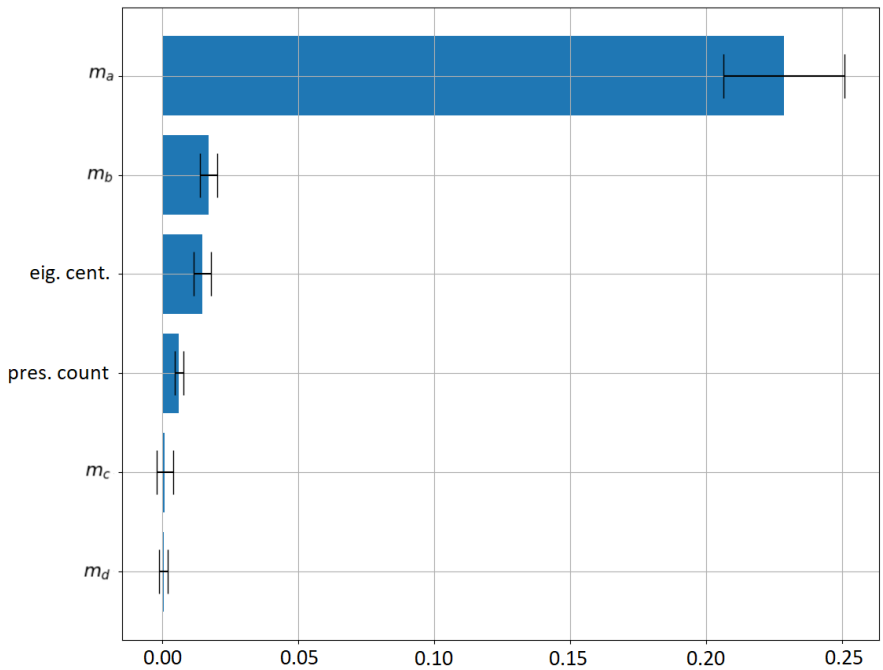


Figure 4.13: Logistic regression permutation importance for the different importance measures, for the BaTIS dataset. The error bars represent the standard deviation of the importances across the different trees that make up the model.

Measure	BaTIS dataset
Precision	0.80 (0.72,0.89)
Precision (null model)	0.66 (0.58,0.74)
Recall	0.65 (0.57,0.74)
Recall (null model)	0.51 (0.39,0.64)
AUC	0.69 (0.65, 0.73)
AUC (null model)	0.50 (0.47, 0.52)

Table 4.9: Precision and recall for the logistic regression for the BaTIS dataset, presented alongside the same average precision and recall for the Null model trials, when considering predicting a change in sign for node strength. The brackets denote the 95% confidence intervals.

Considering the feature importances in figure 4.13, in contrast to the equities transaction networks, I now see that the feature m_a shows the largest importance, followed by m_b. This is as expected given the connected nature of the network, meaning that the measure m_a is not impacted by disconnected components, as the leading eigenvalue will be relevant for all nodes. I see similar results when looking at the SHAP values in figure 20 in the appendix. However, here I observe the most important feature to be m_b closely followed by m_a. When looking at the coefficients in figure 4.14, only eigenvector centrality does not show a significant coefficient, and only m_d and presence count show positive coefficients suggesting that higher values of these features would be indicative of nodes being more likely to change, whereas the other coefficients would show larger values for nodes that are less likely to subsequently change.

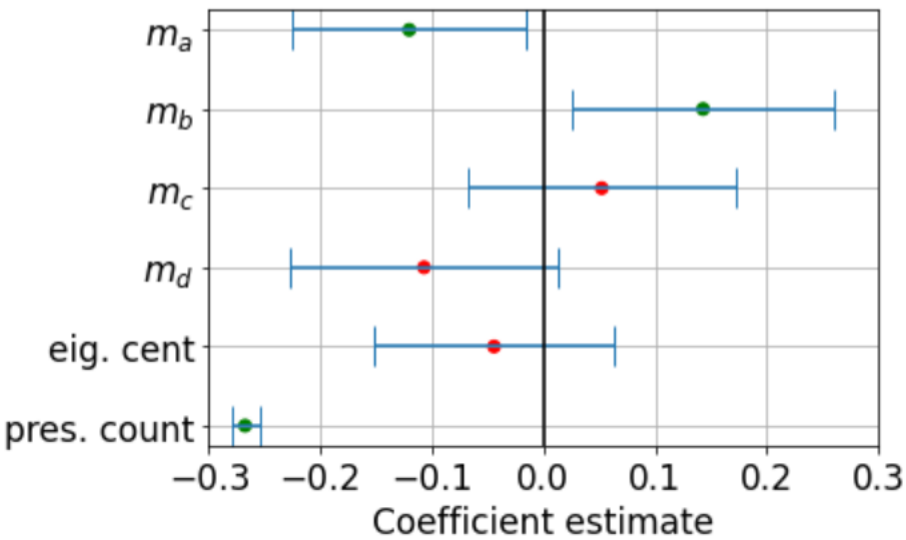


Figure 4.14: Classification model coefficients, with 95% confidence intervals indicated by the error bars. If a p-value is less than 0.05, it is coloured green, otherwise red.

I now consider the predictability of a change in sign of the strength, as considered for the equities networks. For this dataset, The results for this are shown in table 4.9. I see similar results to the equities data, with good performance of the model in terms of precision and ROC AUC. However, poorer performance is observed for the recall showing overlap with the confidence intervals of the null model, meaning that again I cannot conclude that sign change is predictable from my node level features. For the case of sign prediction, the presence count was found to be the most important feature, closely followed by m_a.

Finally, I consider the predictability of the value of the change in strength through the use of a linear regression model. The results for this are shown in table 4.10. In contrast to the equities datasets,

Measure	BaTIS dataset
R^2 score	0.69
CI	(0.66,0.72)
R^2 (N.M.)	0.64
CI	(0.64,0.66)

Table 4.10: Coefficient of determination R^2 for a linear regression with endogenous value of the relative change in node strength, exogenous variables the node level features as used in the classification exercise, for the BaTIS dataset. This is compared to a null model in which the relative change in node strength is randomly shuffled in 100 trials, and the average and 95% confidence intervals (CI) reported.

the coefficient of determination (R^2) suggests that the model has a reasonable prediction capability, and although the confidence intervals for the dummy model are close, there is no overlap between these suggesting that this result is significant, and I can conclude that the size of a change to a node's strength is predictable from my node level features for this dataset. The model coefficients for each of the features are shown in figure 4.15. Here I see that the significant coefficients are m_a, m_b and presence count, and of these only m_b has a positive coefficient. This suggests that nodes with higher values of m_b are more likely to show larger relative changes. Negative coefficients are seen for m_a and presence count suggesting that nodes with lower values of these are more likely to show larger relative changes.

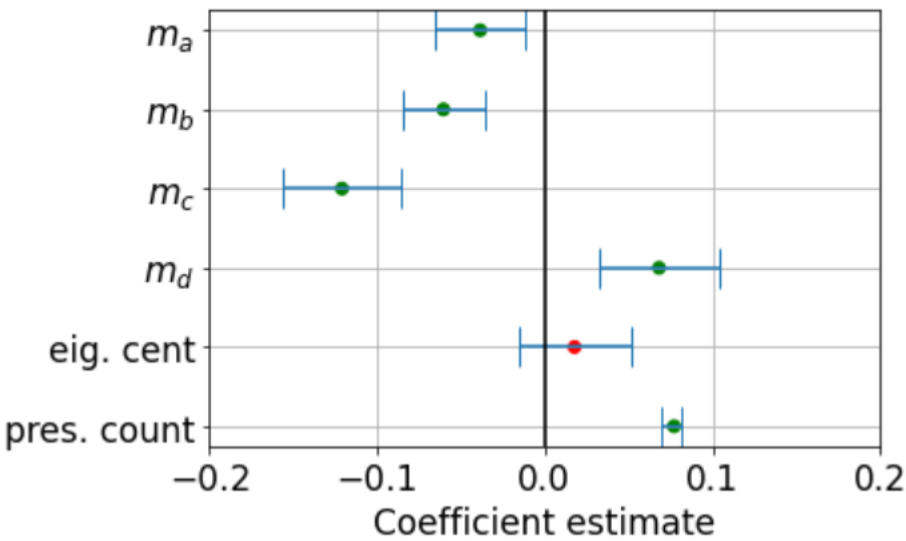


Figure 4.15: Regression model coefficients, with 95% confidence intervals indicated by the error bars. If a p-value is less than 0.05, it is coloured green, otherwise red.

4.5 Discussion

Through consideration of the full network spectrum, I present different ways of considering node importance in networks, with the aim of accounting for both community structure and hub nodes, which are key characteristics of many systems including financial networks. Motivated by several examples in the literature, to achieve this I demonstrate and make use of the most relevant eigenspectra components in order to capture 'community aware' node importance. The result of this is a measure which when applied to financial transaction networks, tells me how much a change to an individual node's strength, which in the context of equity transaction networks is their available funds or product, will impact the rest of the network. This sets my measure apart from centrality measures, as is it is defined in a

temporal sense considering how the network will respond to changes.

By incorporating more than just the leading eigenpair, my measure is able to capture node importance in the context of complex structures, which makes my methods particularly suited for studying equity transaction networks. For these networks, which all show complex structures with both disconnected communities and 'hub' nodes, I compare my measures of node importance to two commonly used centrality measures and also to degree, community label and the number of times a node has appeared historically. This allows me to demonstrate that my measure is not simply acting as a proxy for these key node properties.

When exploring whether static node importance is able to predict the presence of nodes in subsequent snapshots given features derived from the network history, I see that the measure m_b, which makes use of the eigenvector with the largest magnitude for each node, is the most important in determining the prediction for all three equities datasets. Not only do my results demonstrate that m_b is a useful indicator of node importance in a static sense for networks with complex structures, they also provide evidence of the nodes in these equity networks having an evolution which depends on their importance when defined in this way. The latter of these observations is a useful insight for policy makers, as it motivates taking into account the full structure of these networks in determining which nodes to monitor more closely for their effects on the system. It also provides insights into the evolutionary properties of these networks which is interesting from a macro-economical perspective - I observe that more structurally important nodes are less likely to subsequently transact, and given that these nodes tend to show positions in the network in which an impact to their strength would be spread across few counterparties, the observation of these nodes showing less frequent changes relates to the overarching stability of these networks.

Interestingly, if I compare my results on the equity datasets to application to a denser network of global trades in financial services (BaTIS), I observe poorer performance in the prediction of the presence or absence of changes. Also, I also no longer see m_b as the dominating feature in predicting the sign of any change. These results suggest that structural importance could be a unique property of sparse transaction networks. However, the value of a change to a node's strength for this dataset is predictable, with several node level features including m_b being significant predictors.

These observations contribute to the growing body of studies that provide insights into the evolution and stability of financial networks, for example the observations of Bardoscia et. al. [167] that market practices that contribute to cyclical patterns tend to amplify distress. Further relevant to my work is Haldane et. al. [197], in which it is noted that up until the 2008 crisis, the global financial system appeared to be self-regulating and self-repairing despite experiencing several exogenous shocks. However, in the crisis, enduring stress in the money markets was observed due to the interdependence of banks, who rationally sought to protect themselves from infection from other banks by hoarding liquidity. The findings I present in this chapter provide a novel insight into the evolutionary behaviour of transaction networks for the capital markets, which I hope will motivate further research into the links between structural importance, network evolution and how these relate to market stability constraints.

Chapter 5

A Hawkes model for bursty transaction networks

5.1 Chapter summary

This chapter focuses on proposing generative models for the transaction times of high frequency traded equities instruments that are able to reproduce key properties of equity transaction datasets.

To summarise the main goals of this chapter:

1. To assess the key cross sectional and temporal properties of equity transactions that a model should reproduce.
2. To propose models for the generation of transaction sequences.
3. To demonstrate the capability of these models in reproducing the observed cross sectional and temporal properties of equity transaction datasets.

To achieve these goals, I first of all present an exploration of the key properties of transaction reports for five randomly selected FTSE 100 equities. I then present several formulations of Hawkes processes as potential models for the generation of the transaction times of these datasets, before presenting a statistical demonstration of the reproducibility of key cross sectional and temporal properties of these datasets.

The main results of this chapter are as follows:

1. Models which produce transaction sequences specific for each pair of counterparties – which I will refer to as 'edge level' throughout this chapter – show a stronger performance than those where a single Hawkes process is used to generate events independently of the pair of counterparties involved.
2. Persistent relationships between counterparties has influence on key cross-sectional and temporal properties of transaction systems.
3. Multidimensional Hawkes processes, whilst modelling the temporal dimension, are inherently able to reproduce cross sectional properties for the transaction systems considered.
4. Strong performance is observed when applying a model which captures mutual excitation between buy and sell sequences to trades executed via a single hub.

5.2 Introduction

The generation of synthetic transaction level data is of great importance as it allows financial institutions and regulators to find solutions for technical problems through the sharing of data. From a conduct regulation perspective, stock markets require constant monitoring, to avoid the harms of market manipulation, market abuse, and money laundering. Regulators often have access to highly detailed transaction level data for trades of regulated financial instruments. However, the usage and sharing of this data is highly restricted due to its sensitive contents. This means that the generation of synthetic transaction level data with realistic properties is of great importance as these could be shared as a surrogate for real data, allowing for instance to outsource the development and training of advanced machine learning models. Regulators have already begun to seek solutions for this, for example through running hackathons and techsprints to collaborate across regulated firms, start-ups, academia, and professional services to develop high quality, synthetic financial datasets [198]. Synthetic financial data generation has also been the focus of financial services firms; a detailed discussion can be found in Assefa et. al. [32].

In this chapter, I propose generative models for transaction times of highly frequently traded equity instruments. In particular, I do not just consider transaction sequences in a univariate sense for a single dimensional transaction sequence, but also explore the inter-relations between market participants. To this end, I consider the system as a dynamic network. More specifically, I model the systems using multidimensional Hawkes processes, where each pair of counterparties, or each edge using the terminology from network science, is a singe dimension of the Hawkes process. I can then make use of either independent univariate Hawkes processes modelling each dimension independently, allowing me to evaluate the effects of self-excitation at the level of independent transaction sequences for individual edges, or multivariate Hawkes processes, allowing me to evaluate mutual excitation between the edge transaction sequences.

I make observations of both the burstiness of transaction sequences to evaluate the temporal dimension of transaction sequences, as well as several cross-sectional, network-based properties of the systems considered and look for models capable of reproducing these. Properties such as the degree distribution have been a core focus of research in financial networks [199, 200, 55] which often display fat tailed distributions due to the large range of different market participant types. Higher order properties, capturing the relationships between nodes, have also received attention, such as reciprocity [200, 201] and assortativity [202, 203], as well as the rich club coefficient [204, 205] which measures the extent to which well connected nodes connect to each other. I make observations of all of these in datasets of transaction reports for five FTSE 100 stocks (a general exploration of these datasets is found in the appendix section .1.3). I make my observations for three different subsets of these systems, exploring their properties at the level of all the visible trades for a given instrument, trades through a single trading venue, and trades executed off exchange[1], to help build an understanding of how different trading mechanisms may be driving the properties I observe. I then make observations of the same set of properties in data generated by the generative models I propose, which allows me to assess their capability to reproduce the behaviours of these systems that are driven by the relationships between market participants.

5.3 Properties of real transaction systems

When evaluating the performance of a generative model, a set of target properties is needed that the model should be able to reproduce. Here I briefly explore the key temporal and cross sectional,

[1] referred to as 'full', 'single venue' and 'off exchange' throughout this chapter

Data	All	Single venue	Off venue	Single CCP
A	0.35	0.24	0.25	0.27
B	0.48	0.32	0.23	0.39
C	0.37	0.22	0.27	0.28
D	0.30	0.20	0.14	0.20
E	0.46	0.33	0.18	0.35

Table 5.1: Burstiness of the 5 equities considered in this chapter, considering the entire dataset, only trades on the venue with the most trading activity, only off venue trades and only trades through the main clearing house.

network-based properties of the five transaction datasets considered in this chapter. In appendix section .1.4 I also present additional results for a further ten instruments to evidence the robustness of my results.

First I focus on reproducing burstiness as a key temporal property of transaction sequences. Burstiness can be directly derived from the sequence of inter-transaction times, by comparing to the statistical properties of the inter-transaction times for a Poissonian sequence. Starting from the coefficient of variation of the inter-contact times ($\frac{\sigma_\tau}{\mu_\tau}$), which equals 1 for a Poissonian sequence, the burstiness can be defined as

$$B = \frac{\frac{\sigma_\tau}{\mu_\tau} - 1}{\frac{\sigma_\tau}{\mu_\tau} + 1} = \frac{\sigma_\tau - \mu_\tau}{\sigma_\tau + \mu_\tau}, \tag{5.1}$$

where σ_τ is the standard deviation of the inter-contact times and μ_τ is the mean. $B = 1$ indicates a very bursty sequence, $B = 0$ a Poissonian sequence and $B = -1$ an entirely periodic sequence [57].

For all 5 instruments considered in this chapter, for which I consider all transactions in a day, in table 5.1 I present the burstiness of the transaction sequences. Here, the burstiness is calculated for all trades for each instrument, along with trades executed only through the largest trading venue, and those executed off exchange. I refer to these filtered datasets as data subsets throughout the rest of this chapter. Similar levels of burstiness are observed across all datasets, with instrument D consistently showing the lowest burstiness, and of all the subsets, trades through a single CCP showing the largest burstinesses. Figure 5.1 shows the network level burstiness for instrument A for illustration, computed across rolling windows containing 200 transactions, with the window size chosen as an optimal size in containing enough data points for each window for all datasets. I also present the same plots for instruments

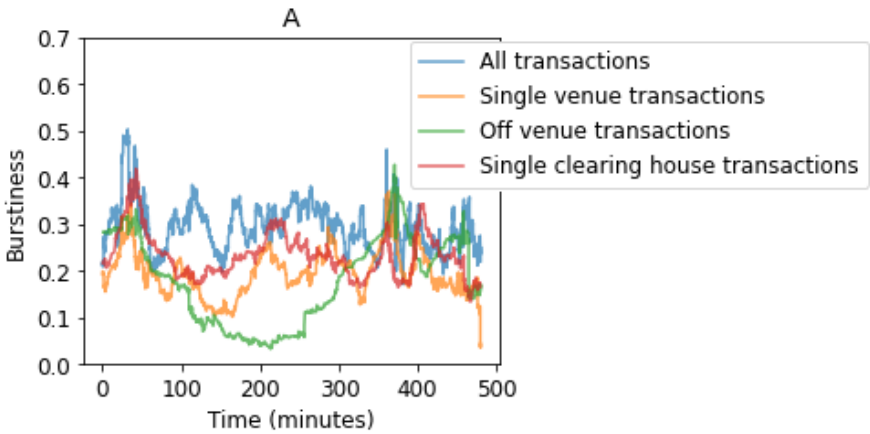
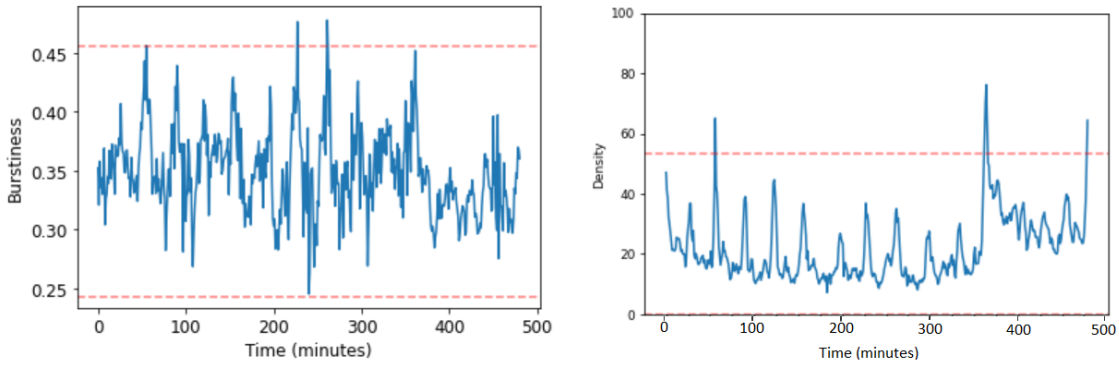


Figure 5.1: Network level burstiness across rolling windows of length 200 transactions, for instrument A for illustration.

B-E in the appendix in figure 23, demonstrating that all instruments have a similar burstiness for the full transaction set and in general the off venue subset shows the lowest burstiness, and in some cases shows a visible dip in the burstiness around midday. Applying an Augmented Dickey-Fuller test for stationarity [206] to the burstiness calculated in a rolling window of 200 transactions, the null hypothesis of a unit root present (which would indicate non-stationarity) is rejected at 5% significance for all except one of the transaction sequences.



(a) Burstiness (b) Density

Figure 5.2: Rolling burstiness and density, averaged across four different days for one of the datasets considered. The dashed lines represent 3 standard deviations from the mean of the time series. The burstiness is computed with a window of 200 transactions and the density over a 10 minute time window.

In addition to the single day explorations of these datasets, I also look to see if there are any consistent patterns across multiple days in the burstiness, to understand whether the burstiness observed is affected by regular events e.g. different markets opening. Figure 5.2 shows the rolling burstiness in comparison to the rolling density for one of the datasets, averaged over four different trading days. Here I observe that, although there are a few significant peaks in the burstiness occurring around the middle of the day, the density shows significantly more prevalent peaks representing variations in market activity. Further to this, there is a clearly observable rise in the density in the later third of the day, which is not reflected in the burstiness suggesting that the burstiness can be treated as a noisy but fixed quantity across a day's trading.

Since I am seeking a generative model that is able to reproduce the complex relational properties between counterparties, I also explore the cross sectional, network-based properties observable for the transaction when aggregated to obtain network snapshots. Here I consider higher and higher order properties allowing me to study the cross-sectional behaviour of the systems studied first at the node level through the degree distribution, then the level of neighbours through assortativity and beyond through network level reciprocity and rich club distributions:

- **Degree distribution** - the fraction of nodes in the network with degree k

$$P(k) = \frac{N_k}{N}, \tag{5.2}$$

  where N_k is the number of nodes with degree k and N is the total number of nodes

  Figure 5.3a shows the cumulative degree distributions for instrument A, with those for the other instruments in the appendix in figure 27. Here I observe similar degree distributions for the different subsets, showing fat tails in which the majority of nodes have a low degree with a small number of hub nodes having very high degree.

- **Assortativity** - As outlined in section 2.1, assortativity is given by:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{\mathrm{Tr}(\mathbf{e}) - ||\mathbf{e}^2||}{1 - ||\mathbf{e}^2||}. \tag{5.3}$$

  Figure 5.3b shows the assortativity computed in a rolling 200 transaction window for the different subsets of instrument A for illustration, and the same for the other instruments can be found in the appendix in figure 28. I observe here that the assortativity is consistently the most negative for the single venue subset, indicating disassortative relationships in which nodes connect to other nodes with very different degrees to them. The off exchange subsets shows the least negative values, which is as expected as this subset will contain a lower prevalence of hub nodes.

- **Reciprocity** - Again as outlined in section 2.1, network level reciprocity is given by

$$r = \frac{|(u,v) \in G||(v,u) \in G|}{|(u,v) \in G|}, \tag{5.4}$$

  with reciprocity for a single node being the ratio of the number of edges in both directions to the total number of edges attached to the node in question [56].

  Figure 5.3c shows the reciprocity again computed in rolling windows of 200 transactions for instrument A, and the same for the other instruments can be found in the appendix in figure 29. Here I see a large level of variation across the different subsets, with the off-venue subset showing a lower reciprocity in general, and significant variation over the day.

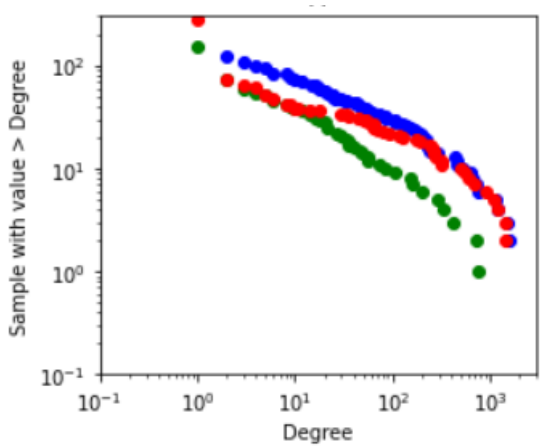
- **Rich club index** - I measure the rich club phenomenon using the rich club coefficient at each degree k, which is the fraction of of the actual and potential number of edges among the set of nodes with degree higher than k:

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}, \tag{5.5}$$

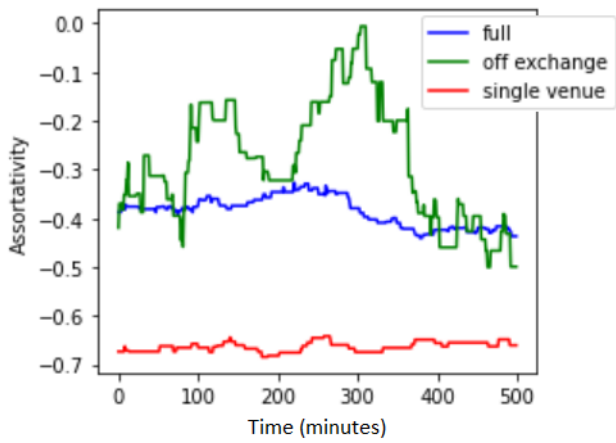
  where N_k is the number of nodes with degree larger than k and E_k is the number of edges among those nodes.

  Figure 5.3d shows the rich club coefficient distributions for 10 minute snapshots of these networks for the 5 instruments, presented here for instrument A for illustration. I observe here that the nodes with the largest degree have low values of the rich club indices, which is driven by the presence of disconnected star sub-networks containing a single hub node connected to a number of other nodes that are only connected to the hub. This is also reflected in the rich club distributions of the other datasets presented in the appendix in figure 30, and the single venue subsets in figure 31 but less so for the off-venue subset in figure 32.

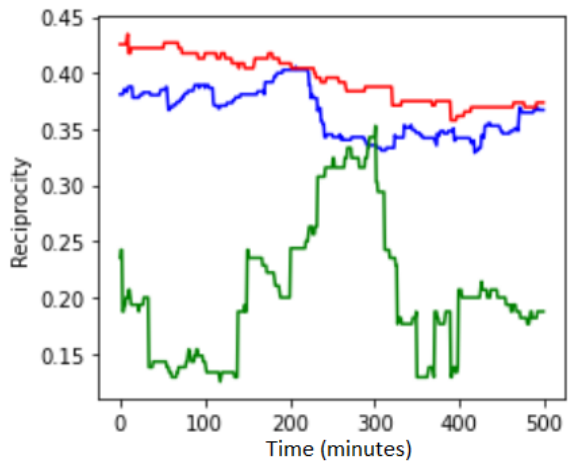
To summarise, I observe in all 5 datasets burstiness appears to be an inherent property of the transaction sequences, that isn't constrained to a particular type of trading activity. Across all transactions, the burstiness shows stability across time even when the underlying transaction density varies. I observe consistency in the degree distributions, but variation across the different subsets and different datasets for the assortativity, reciprocity and rich club distributions. Given their ability to produce bursty sequences, in the next section I introduce Hawkes processes as generative models for transaction sequences. I consider different approaches making use of both univariate and multivariate Hawkes processes in single and multiple dimensions, to establish the extent to which self- and mutual exciting behaviour could be responsible for the burstiness I observe. I further evaluate the models by exploring their ability to reproduce the cross sectional, network-based properties of the transaction sequences when aggregated, finding that for the case of multidimensional models, these properties are inherently reproduced by the temporal point process.



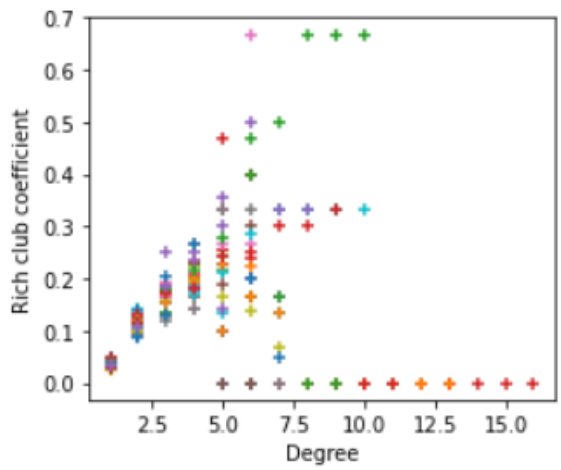
(a) Cumulative degree distribution computed over the entire time period and plotted on a log-log scale.



(b) Assortativity computed across time in rolling windows of 200 transactions for the different subsets.



(c) Reciprocity computed across time in rolling windows of 200 transactions for the different subsets.



(d) Rich club coefficients for the degrees of nodes for the 5 different datasets. The colours represent snapshot networks aggregated every 10 minutes.

Figure 5.3: Cross sectional network properties for instrument A.

5.4 Proposed generative models

When considering transactions for a single instrument that is traded on multiple venues, or through a single venue for which trades are cleared by multiple clearing houses, or considering off exchange transactions, I treat these systems as dynamic networks. For the transactions making up these dynamic networks, there are a number of different excitation mechanisms which may result in a trade. There may be some level of mutual excitation, in which the presence of a trade between two individuals excites future trades between other pairs of individuals. There may also be excitation on the level of an individual pair of traders, where the presence of a trade between them excites (or inhibits) the presence of a trade between them in the future. There will also be trades that are controlled by exogenous factors, such as news events. Further to this, I can also not rule out the possibility of an inherent excitation in the transaction sequence as a whole, regardless of who the participants are. In order to get a step closer to understanding the generative process underlying these networks, I compare a number of different approaches using the Hawkes process to simulate transactions.

Hawkes type	Granularity	Dataset
Univariate	Overall transaction sequence	All venues
		Single venue
		Off venue
Univariate	Edge level	All venues
		Single venue
		Off venue
Multivariate	Edge level	All venues
		Single venue
		Off venue
Univariate	Buys/sells	Single CCP
Bivariate	Buys/sells	Single CCP

Table 5.2: Summary of different applications of Hawkes Processes considered in this chapter

I also make use of multidimensional, multivariate Hawkes processes to generate transaction sequences composed of sub-sequences per pair of counterparties, or in other words for each edge if using terminology from network science. In doing so, my model will capture both self- and mutual excitation within and between transaction sequences for each edge. I can also make use of multiple univariate Hawkes processes to generate transaction sequences for each edge independently, so that my models have only self-excitation for each edge. Both of these approaches thus inherently produce the information on counterparties involved in each trade. However, if I use a univariate Hawkes process to generate a single transaction sequence for the whole system, to model self-excitation of the transaction sequence as a whole, I am not inherently producing information on which counterparties are transacting. This means that I also need a method to select which edge will transact at each point in the sequence. So in this case, the model I propose consists of two parts - timestamp generation and edge selection. I do this by selecting the edge for each transaction according to the frequency of their historical transactions and compare this to simply selecting edges randomly as a null model.

Table 5.2 summarises the different ways I make use of the Hawkes process and the dataset subsets I apply these to. I further compare these models to models using Poisson processes, considering both a Poisson process for a one dimensional transaction sequence, and individual Poisson processes multidimensionally for each edge.

Parameter estimation

A univariate Hawkes process is controlled by three parameters, the baseline intensity and the kernel intensity and decay, which can be estimated using maximum likelihood estimation. Full details of the Hawkes process, and the meaning of the parameters, can be found in section 2.2.3. The log-likelihood can be expressed in this case as [207]

$$\mathcal{L} = \log \frac{\prod_{i=1}^{n} \lambda(t_i)}{\exp \int_0^T \lambda(t)dt} = \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \lambda(t)dt. \tag{5.6}$$

However, considering the multivariate Hawkes process with mutual excitation between edges, the number of parameters to be estimated increases as $\mathcal{O}(2n + n^2)$, as each edge will have its own baseline intensity and kernel decay parameters, as well as cross terms within the kernel matrix to capture the influence of each edge level process on every other process. This means that even for a modest number of processes, a standard maximum likelihood approach is computationally infeasible [2]. Instead, I make use of the

[2] On a 2nd generation Intel Xeon Platinum 8000 series processor, the parameter estimation for a system with 200 cross

ADM4 method developed by Zhou et. al. [207], using the implementation provided by the tick package [208], to estimate the μ_i and α_{ij} parameters for a process with a given decay parameter β. I then use a brute force approach to find the value for β which produces a sequence with a burstiness closest to the real sequence. The ADM4 method first of all makes use of the sparsity and low-rank properties common in many network systems and applies sparse and low-rank regularisation to the likelihood function. Since the resulting likelihood is non-differentiable and difficult to optimise, it then combines the idea of alternating direction method of multipliers and majorisation minimisation to convert the optimisation problem to several sub-problems that are much easier to solve. The performance of this method is demonstrated in [207], in application to both simulation and real-world datasets. A drawback of this method is that it is only implemented with a constant kernel decay across all processes, whereas I know from my exploration later in this chapter that the individual processes are heterogeneous in terms of their burstiness. Fast estimation of a multivariate Hawkes with the flexibility of a varying kernel decay would thus be a useful area for future research efforts.

5.5 Results

5.5.1 Single dimension model - univariate Hawkes process with counterparty selection

First, I present the results of fitting a univariate Hawkes process to the full transaction sequence, and using the fitted process to generate new transactions. Looking at the burstinesses in table 5.3

Data	Real data	Poisson	Hawkes
A (full)	0.40	0 (-0.01, 0.01)	0.30 (0.27, 0.33)
A (single venue)	0.30	0 (-0.02, 0.02)	0.13 (0.11,0.14)
A (off exchange)	0.29	0 (-0.02, 0.02)	0.40 (0.34, 0.46)
B (full)	0.54	0 (-0.02, 0.02)	0.27 (0.24, 0.29)
B (single venue)	0.39	0 (-0.01, 0.01)	0.41 (0.35, 0.46)
B (off exchange)	0.31	0 (-0.02, 0.02)	0.08 (0.06, 0.11)
C (full)	0.44	0 (-0.01,0.01)	0.51 (0.48, 0.56)
C (single venue)	0.32	0 (-0.01, 0.01)	0.37 (0.34, 0.40)
C (off exchange)	0.32	0 (-0.02, 0.02)	0.78 (0.48, 0.94)
D (full)	0.34	0 (-0.01,0.01)	0.07 (0.06, 0.08)
D (single venue)	0.26	0 (-0.02, 0.02)	0.21 (0.01, 0.39)
D (off exchange)	0.19	0 (-0.02, 0.02)	0.75 (0.37, 0.96)
E (full)	0.55	0 (-0.02, 0.02)	0.41 (0.38, 0.43)
E (single venue)	0.45	0 (-0.02, 0.02)	0.54 (0.48, 0.60)
E (off exchange)	0.27	0 (-0.02, 0.02)	0.53 (0.43, 0.64)

Table 5.3: Burstiness of the timestamps generated by univariate Hawkes and Poisson processes, in comparison to the real timestamps

(visualised in figure 36 in the appendix), I immediately see that although the Hawkes process produces sequences with a positive burstiness closer to the real data than a Poisson, I rarely see overlap of the the 95% confidence intervals [3] with the burstiness observed in the real dataset. This suggests that self-excitation alone is not able to explain the burstiness observed. I further explore the parameter values in the appendix in table 2, for which I observe counter-intuitive values which again suggest a poor fit of the univariate Hawkes process to the transaction sequences as a whole. This suggests

terms took over 8 hours to complete.

[3] the confidence intervals were obtained by running the simulations 1000 times

Data	Random	Frequency
A (full)	2.99×10^{-20}	0.38
A (single venue)	1.88×10^{-20}	0.73
A (off exchange)	3.01×10^{-6}	0.06
B (full)	4.20×10^{-20}	0.36
B (single venue)	9.39×10^{-18}	0.030
B (off exchange)	8.63×10^{-7}	0.01
C (full)	1.74×10^{-33}	0.41
C (single venue)	3.53×10^{-24}	0.001
C (off exchange)	8.21×10^{-5}	0.01
D (full)	2.41×10^{-31}	0.035
D (single venue)	$2.44 \times 10^{-}13$	0.35
D (off exchange)	7.04×10^{-8}	0.04
E (full)	4.87×10^{-36}	0.018
E (single venue)	1.67×10^{-24}	0.18
E (off exchange)	9.50×10^{-9}	1.02×10^{-4}

Table 5.4: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the distributions of the Rich Club index over different degrees for the two different edge selection methods in comparison to the real data, for the 5 datasets considered.

that mutual excitation, and self-excitation specific to individual edges, could be a relevant generative property of transaction sequences across the different subsets.

As noted in section 5.4, fitting a univariate Hawkes process for a 1-dimensional transaction sequence does not inherently generate information on the counterparties involved in each transaction, so here I also present the results of selecting the counterparties to transact according to their frequency of transactions in the training data, and compare this to a null model of randomly selecting them. I evaluate the performance of these methods by assessing their ability to reproduce the cross sectional, network-based properties explored in section 5.3 for the real datasets.

The first property I consider is the distribution of rich club coefficients at different node degrees. Figure 37 in the appendix presents an example of these distributions for instrument A for illustration, and shows clear visual similarity between the distributions for the real data with the frequency based method of edge selection. In order to quantitatively assess the similarities of these distributions to those of the real data, I make use of a two dimensional two sided Kolmogorov-Smirnov test [209], the results of which are presented in table 5.4. Here, small p-values mean that the two samples are significantly different, with the null hypothesis being that the two distributions are the same. I see that for frequency based selection, considering the full datasets, four of the datasets show a p-value of > 0.05, meaning that for these the null hypothesis is not rejected and the distributions are not significantly different from the rich club distributions of the real data. For the dataset for which the null hypothesis is rejected for frequency based edge selection, the p-values are still much higher than for the random selection, suggesting some level of similarity between these distributions and the rich club distribution of the real data. Similar results are seen for the single venue subsets, where three of the datasets have p-values > 0.05. However, for off-exchange trading the null hypothesis is rejected at 5% confidence, suggesting that transactions executed off exchange are less dependent on the prevalence of historical trades between the two counterparties than the frequency based selection models. The null hypothesis is as expected rejected for all of the datasets for my null model of randomly selecting counterparties for each transaction.

I also consider in a similar vein the degree distributions, again finding that frequency based edge selection produces distributions not significantly different from the real data in all cases. The results

for this can be found in table 3 the appendix.

Next I consider the reciprocity computed over a rolling window for all subsets, presented in figure 5.4 for instrument A for illustration, with the results for instruments B-E presented in the appendix in figure 39. I clearly see that the frequency based method produces a similar reciprocity to the real data, but the null model randomly selecting counterparties is unable to reproduce the reciprocity. The reciprocity of the real data regularly falls within the 95% confidence intervals of the frequency based selection method, however this is less than 95% of the time for four of the five datasets. I make similar observations for the assortativity across time, which I again present in the appendix in figure 40.

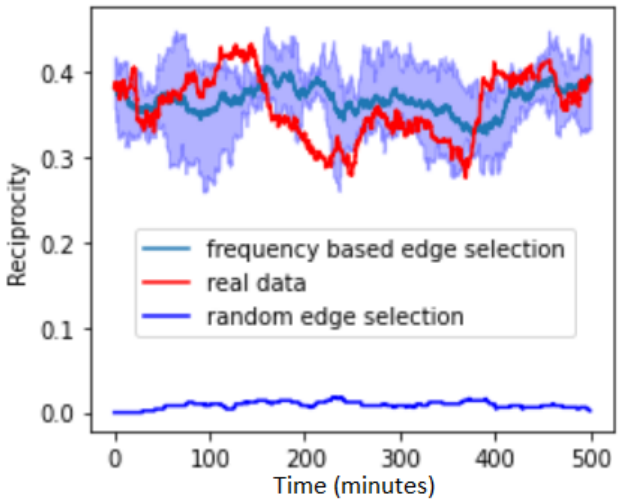


Figure 5.4: Reciprocity computed in rolling windows of 200 transactions for the 5 different datasets, presented for the two methods of edge selection along with the real data, for instrument A for illustration.

To summarise these results for univariate Hawkes processes, I find that these processes are unable to reproduce the burstiness observed for all the datasets considered, which suggests that self-excitation at the level of the overall transaction sequence does not explain the burstiness. However, the frequency based method of selecting the counterparties for each transaction is able to reproduce the network-based, cross sectional properties that I observe for the real data, suggesting that preferential trading relationships play a meaningful part in shaping the structure of these markets.

5.5.2 Multidimensional models

I now look instead at fitting multidimensional Hawkes processes, considering my transaction sequences as multidimensional systems with a dimension for each edge, to assess the extent to which mutual excitation between edges, and self-excitation of each edge independent of other edges to explain the temporal and cross-sectional properties I have observed. It is worth noting that one of the drawbacks of using Hawkes processes multidimensionally is that several sequences are observed to contain only a single transaction, so when considering multidimensional Hawkes processes, I restrict to sequences with more than 10 transactions. This roughly halves the number of counterparties for all of the datasets considered. I compare multidimensional, multivariate Hawkes processes to multidimensional, univariate Hawkes processes where the process is fitted to each dimension independently, to disentangle the effects of self- and mutual excitation.

Table 5.5 shows the burstiness of the simulated transactions for the different datasets across the different subsets, in comparison to the real burstinesses. These are also presented visually in the appendix

	True	Univ. HP	Multiv. HP
A (full)	0.45	**0.48(0.41, 0.56)**	0.35 (0.33, 0.40)
A (single venue)	0.21	**0.15(0.07, 0.25)**	**0.24(0.21, 0.26)**
A (off exchange)	0.18	**0.18(0.09, 0.30)**	0.26 (0.21, 0.30)
B (full)	0.60	0.43 (0.34, 0.58)	0.41 (0.39, 0.46)
B (single venue)	0.16	**0.16(0.08, 0.26)**	0.28 (0.25, 0.31)
B (off exchange)	0.10	**0.12(0.02, 0.23)**	0.17 (0.13, 0.21)
C (full)	0.54	**0.54(0.49, 0.60)**	**0.44(0.35, 0.62)**
C (single venue)	0.37	0.20 (0.10, 0.33)	0.19 (0.16, 0.22)
C (off exchange)	0.23	**0.16(0.07, 0.26)**	**0.24(0.20, 0.29)**
D (full)	0.43	0.26 (0.20, 0.37)	0.58 (0.53, 0.66)
D (single venue)	0.34	0.22 (0.13, 0.32)	0.20 (0.18, 0.22)
D (off exchange)	0.20	**0.15(0.07, 0.23)**	**0.18(0.15, 0.22)**
E (full)	0.56	**0.55(0.50, 0.59)**	0.34 (0.31, 0.38)
E (single venue)	0.18	**0.13(0.05, 0.23)**	0.28 (0.25, 0.31)
E (off exchange)	0.10	**0.17(0.07, 0.28)**	0.14 (0.12, 0.17)

Table 5.5: Burstiness of the timestamps generated by edge level Hawkes processes, in comparison to the real timestamps. Results shown in bold are the cases where the true burstiness falls within the 95% confidence intervals of the simulation results.

in figure 41. As before, I consider a Poisson process as a benchmark which I do not present as this process as expected always has a burstiness of 0. I observe that the univariate approach has the best performance, showing the true burstiness falling within the 95% confidence intervals in all cases for the off exchange data subsets and 3 of the 5 datasets for the full and single venue subsets. This suggests that much of the burstiness of the real data can be attributed to self-excitation at the edge level, or in other words, that the presence of a trading relationship between two counterparties increases the likelihood of future trading relationships. This is in agreement with my results from the previous section where selecting counterparties to transact based on their historical transaction frequency showed good performance in reproducing the cross sectional properties of the datasets. Both observations suggest that preferential trading is a key property shaping the behaviour of these systems. The multivariate approach shows the true burstiness falling outside the confidence intervals for the majority of the trials. However, the method itself produces results with tighter confidence intervals around the mean, suggesting a higher level of stability in the resulting simulations.

In contrast to the unidimensional case, when considering multidimensional Hawkes processes, the models inherently produce the counterparties involved in the trade. I now explore the cross-sectional properties of these, restricting to the full datasets as the network structure diminishes when considering any smaller subsets when removing counterparties with limited activity.

Starting with observations of the rich club coefficient distributions, I observe similarities of both the approaches, for which the null hypothesis of differences between the distributions is rejected for both models, for all datasets. The same can be said for the degree distributions, which also show the null hypothesis rejected in all cases. Plots of the rich club and degree distributions for the instrument A can be found in the appendix in figures 42 and 43.

Next, considering the reciprocity and assortativity computed over rolling windows of 200 transactions, I observe that both the assortativity and reciprocity of the sequences generated using a multivariate Hawkes are similar in value and variance to the real process. This can be seen visually in figures 44 and 45 in the appendix. Although there is a visually high similarity of both reciprocity and assortativity for the multidimensional Hawkes process sequences in comparison to the real data, in making use of a Levene test [210] to assess whether the sequences have equal variance and in making use of a one-sided

	True	univ. HP	univ. edge HP	multiv. HP
A (full)	0.45	0.30 (0.27, 0.33)	**0.48(0.41, 0.56)**	0.35 (0.33, 0.40)
A (single venue)	0.21	0.13 (0.11, 0.14)	**0.15(0.07, 0.25)**	**0.24(0.21, 0.26)**
A (off exchange)	0.18	0.40 (0.34, 0.46)	**0.18(0.09, 0.30)**	0.26 (0.21, 0.30)
B (full)	0.60	0.27 (0.24, 0.29)	0.43 (0.34, 0.58)	0.41 (0.39, 0.46)
B (single venue)	0.16	0.41 (0.35, 0.46)	**0.16(0.08, 0.26)**	0.28 (0.25, 0.31)
B (off exchange)	0.10	**0.08(0.06, 0.11)**	**0.12(0.02, 0.23)**	0.17 (0.13, 0.21)
C (full)	0.54	**0.51(0.48, 0.56)**	**0.54(0.49, 0.60)**	**0.44(0.35, 0.62)**
C (single venue)	0.37	**0.37(0.34, 0.40)**	0.20 (0.10, 0.33)	0.19 (0.16, 0.22)
C (off exchange)	0.23	0.78 (0.48, 0.94)	**0.16(0.07, 0.26)**	**0.24(0.20, 0.29)**
D (full)	0.43	0.07 (0.06, 0.08)	0.26 (0.20, 0.37)	0.58 (0.53, 0.66)
D (single venue)	0.34	**0.21(0.01, 0.39)**	0.22 (0.13, 0.32)	0.20 (0.18, 0.22)
D (off exchange)	0.20	0.75 (0.37, 0.96)	**0.15(0.07, 0.23)**	**0.18(0.15, 0.22)**
E (full)	0.56	0.41 (0.38, 0.43)	**0.55(0.50, 0.59)**	0.34 (0.31, 0.38)
E (single venue)	0.18	0.54 (0.48, 0.60)	**0.13(0.05, 0.23)**	0.28 (0.25, 0.31)
E (off exchange)	0.10	0.53 (0.43, 0.64)	**0.17(0.07, 0.28)**	0.14 (0.12, 0.17)

Table 5.6: Burstiness of the timestamps generated by unidimensional, multidimensional, univariate and multidimensional, multivariate Hawkes processes, in comparison to the real timestamps. Results shown in bold are the cases where the true burstiness falls within the 95% confidence intervals of the simulation results.

t-test to assess whether the mean of the difference between the simulated and actual reciprocity and assortativity is significantly different from 0, in all cases the p-values were <0.5, meaning that I reject the null hypothesis in both cases and validate the similarity between the sequences.

To summarise the performance of both unidimensional and multidimensional modelling approaches, I present the burstinesses of all of the Hawkes approaches in a single table in table 5.6. Here I can see that the univariate Hawkes process applied to multiple dimensions independently shows the strongest overall performance, although this is heavily dependent on the wider confidence intervals of this model.

5.5.3 Bivariate Hawkes process for buy/sell trades

When considering trades occurring through a central clearing party, from a network perspective the system is a network consisting of a single star. This means the insights I can gain into the behaviour of the system by modelling it as a dynamic network are limited. Instead, I re-frame the modelling of the system to consider the buy and sell trade executions as separate sequences and allow for mutual excitation between the buys and sells. This is similar in approach to the recently published methods in [211] who make use of a bivariate marked Hawkes process to model aggressive market order arrivals. As before, I make use of Maximum Likelihood estimation to fit the parameters of the Hawkes process to the transaction sequences. Since I am considering a bivariate Hawkes process, there are 8 parameters to estimate with a 2×2 matrix for the adjacency α, two decay parameters β, and two baseline intensity parameters μ_i.

The estimated parameters for the bivariate Hawkes model can be found in the appendix in table 8, and the resulting burstiness of the overall sequences, as well as the buy and sell sequences separately, are shown in table 5.7. Considering the burstiness, I see that in the majority of cases, the true burstiness falls within the confidence intervals of the simulations for both the full transaction sequences as well as the buys and sells themselves. For the case of dataset D, which shows a true burstiness overall outside of the confidence intervals of the simulation, if I look at the rolling burstiness of the buys and sells separately in figure 34 in the appendix, I see that the burstiness of the sells drops during the latter

Data	Real data	Univariate	Bivariate
A (all)	0.30	0.09 (0.05, 0.13)	**0.26** (0.16,0.39)
A (buys)	0.19	**0.14** (0.08, 0.21)	**0.26** (0.16, 0.37)
A (sells)	0.24	**0.17** (0.08, 0.28)	**0.21** (0.09, 0.37)
B	0.45	0.12 (0.05, 0.22)	**0.49** (0.20, 0.72)
B (buys)	0.35	0.20 (0.10 0.33)	**0.36** (0.01, 0.67)
B (sells)	0.41	0.21 (0.09, 0.34)	**0.51** (0.22, 0.74)
C	0.34	0.13 (0.08, 0.20)	**0.27** (0.10, 0.46)
C (buys)	0.27	0.18 (0.11, 0.25)	**0.35** (0.22, 0.49)
C (sells)	0.28	**0.21** (0.11, 0.33)	**0.27** (0.07, 0.49)
D	0.24	0.09 (0.05, 0.15)	0.12 (0.05, 0.21)
D (buys)	0.20	**0.17** (0.08, 0.31)	**0.15** (0.05, 0.29)
D (sells)	0.18	**0.15** (0.09, 0.24)	**0.12** (0.05, 0.24)
E	0.49	0.12 (0.06, 0.19)	**0.42** (0.09, 0.77)
E (buys)	0.41	0.19 (0.10, 0.32)	**0.34** (0.28, 0.77)
E (sells)	0.43	0.21 (0.11, 0.34)	**0.37** (0.09, 0.67)

Table 5.7: Burstiness of timestamps generated with a bivariate Hawkes process.

part of the day, which explains why my model is unable to reproduce the overall burstiness. I can also compare to fitting Hawkes processes to the buys and sells separately, also shown in table 5.7, to help me understand how much the observed burstiness is driven by self excitation as opposed to mutual excitation. I see here that although the confidence intervals of the buys and sells contains the true burstiness in many cases, the simulated transaction sequences as a whole have a consistently lower burstiness than the real transaction sequences. This means that some of the burstiness of this system is likely to be driven by mutual excitation between buy trades and sell trades.

5.6 Summary

In this chapter, I have explored the use of Hawkes processes as generative models for transaction sequences for FTSE 100 stocks. This is challenging since these markets are highly heterogeneous, with many different methods of trading giving rise to different behaviours, which is reflected in the observations I make of differing properties across the 5 different instruments considered, and across different subsets of the data for transactions that are executed on the dominant exchange, and only those that are executed off exchange in comparison to all trades. My task has thus been to assess the capability of Hawkes process models to reproduce these properties with a flexibility allowing for this variation, in particular across both temporal and cross sectional properties simultaneously.

When modelling the transaction sequences as a single dimension using a univariate Hawkes process, I observed counter-intuitive parameter values and although the model was able to reproduce sequences with a burstiness closer to the real data than the benchmark of a Poisson process, in most cases this model underestimated the burstiness, suggesting that self-excitation at the level of the overall transaction sequence does not accurately represent the true generative process of these transaction sequence. In selecting which pair of counterparties should transact at each point in the transaction sequence, selecting based on their historical trading frequency from the training set produced cross-sectional properties that were the most similar to those of the real data, suggesting that historical counterparty relationships influences the probability that a pair of counterparties will transact in the future.

When considering multidimensional approaches, the approach making use of a univariate process for each pair of counterparties (each edge) was able to reproduce the burstiness of the real dataset

much more often than the mutlivariate approach could. However, the multivariate approach showed tighter 95% confidence intervals for the burstiness, suggesting that this model was more stable. The better performance of the univariate edge level model is in agreement with the strong performance of the frequency based edge selection, since both suggest that historical counterparty relationships are influential in the presence of a future relationship. However, the multivariate approach performed slightly better than the multidimensional, univariate model in reproducing the cross sectional properties of the transaction networks, suggesting that improvements to this model, for example allowing the kernel decay to vary across the different edges, or considering other choices of kernel such as power law kernels, and establishing methods to compute the likelihood for these, would be a promising avenue for further research.

Finally, I observed strong performance in applying a bivariate model to buys and sells for trades occurring through a single central clearing counterparty, with the burstiness reproduced in the majority of cases. I also observed a better performance of the bivariate approach in comparison to a univariate Hawkes process for buys and sells separately suggesting that mutual excitation between the buys and the sells contributes to the overall burstiness. Further exploration into the performance of edge level modelling in conjunction with the bivariate model would be an interesting avenue of further exploration.

Since I am proposing generative methods of transaction sequences, it is worth discussing the potential uses of simulated transaction sequences. In an ideal world in which transaction reporting contains no errors, duplicate reports or nuances which cause deviations from the generative model, these methods would be valuable in anomaly detection, in flagging transactions which deviate too far from the model, or alternatively systems which show differing aggregate properties, allowing consideration of how these properties constitute risk. For example, I can see that all 5 systems considered in this chapter show similar burstinesses across time and similar degree distributions. The presence of burstiness shows the market response to activity, which drives price formation, and the stability of prices is integral to the good function of markets as a whole. The degree distributions I observe demonstrate fat tails in which the majority of participants have a low degree with a small number of hub nodes with very large degrees, usually central clearing houses which play a large part in mitigating the risks of highly interconnected networks and the propagation of risk. In the appendix section .1.3 I present an example of the use of the multidimensional, univariate Hawkes process model to detect anomalies in the reciprocity across time. A similar approach could be taken for other properties of the transaction systems, depending on the types of anomalies that are of interest. It is also worth noting that a number of problems encountered in finance involve extreme class imbalances, for example fraud detection. A generative model for transactions could be used as an oversampling method to generate artificial data, to help alleviate class imbalance, as is explored in Hung et. al. [212], who make use of Generative Adversarial Networks to assist on the classification of credit card fraudulent transactions, which are highly imbalanced with only 0.17% of transactions of the positive (fraudulent) class. Assessing the capability of my Hawkes based generative models in this application would be an interesting next step for this research.

Chapter 6

Conclusion and future work

6.1 Summary & Contributions

In this thesis, I have used techniques from network theory to assess how the trading behaviour of individual market participants influences the evolution of the markets they participate in. I have focused on networks constructed from transaction reports of equities traded on the UK capital markets. Understanding how these systems function is of importance to regulators, who are responsible for ensuring that markets function well and for minimising risks passed to consumers. To do this, they need to be able to identify important market participants, and to understand how participant behaviours can influence change in the systems as a whole. This thesis has contributed methods which deliver these capabilities, and has also provided novel observations into the behaviour of these systems. For some of these observations, I have found behaviours which are generally observed for financial systems. For example, all financial networks I have considered have a high reciprocity exceeding 55% indicating a high prevalence of repetitive relationships common in other financial systems. Other properties observed are specific to the types of market considered, such as the prevalence of trades cleared through central counterparty clearing houses, which act as hub nodes for instruments traded on major trading venues. I have produced methods that account for and are able to reproduce these properties, and which provide significant insights into the dynamics of equity trading networks.

My first and central contribution was a novel way to measure the potential impact nodes and edges could have on the global structure of networks if they were to change. This underpins the rest of my research, particularly when studying snapshot networks for selected equity instruments using discriminative methods. In chapter 3 I presented a derivation for a measure of edge importance. This is based on considering the change to the network leading eigenvalue, which can be approximated and decomposed into the potential for an edge to influence the structure, given by the derivative of the leading eigenvalue with respect to an individual edge, and the actual change in the network structure. The former captures the sensitivity of the eigenvalue to changes to an individual edge, making it a natural candidate for an edge importance measure. I took a perturbation theory approach to derive a computable form of this derivative for both undirected and directed networks. I validated my derivation by perturbing individual edges to observe the resultant change to the eigenvalue, and compared this to the value of importance. To explore the temporal behaviour of my measure of importance, I contributed a Markov chain model to describe the dynamics of snapshot networks, parameterised by the extent to which edge importance controls the scale and probability of edge changes between snapshots. I explored this model through simulations to understand how varying the parameters affects the distributions of the values of edge importance for changing vs. non-changing edges. I also applied this model to real data, drawing insights again from the importance distributions alongside using a maximum likelihood approach to estimate the parameters from the data. In using importance to predict edge changes, I found it to be slightly

predictive for all datasets considered, though only marginally so for two of these datasets. I showed that this predictive performance can be linked to the parameters of my dynamical model, allowing me to suggest that the parameters of the model have practical application in classification of financial networks. Other observations made in chapter 3 include edge importance being dominated by a small handful of participants with changes to these edges being less likely. This observation suggests that there are a few structurally important edges which could act to stabilise a system which might otherwise move towards a regime of instability. I also observed positive values for the the parameter controlling the extent to which importance influences the scale of resultant edge changes, which suggests that more important edges are more likely to make larger changes when they do change. Finally, through an experiment to demonstrate how predictability varies with aggregation scale, I found that predictability initially improves with increasing aggregation scale before deteriorating. This presents a trade-off between improved performance of my importance measure for the quasi-continuous limit in which each time snapshot contains a single edge change and improved predictability for larger aggregation scales. This contributed to my motivation to model transaction networks as temporal point processes as opposed to snapshot networks in chapter 5.

When considering instead node importance in chapter 4, I observed that if the most relevant eigenspectra component is chosen for each node, my measure of importance becomes 'community aware'. I supported this through experiments with a toy barbell network in which I show how different parts of the eigenspectrum show relevance to different communities in networks. I found that my importance measure which incorporates this property of the eigenspectrum showed the largest difference in the distribution mean for present nodes in comparison with absent nodes, suggesting its capability as a discriminator for subsequent activity. I then confirmed this by finding that it differs from and outperforms other importance measures in predicting subsequent transactions. I observed that more structurally important nodes are less likely to subsequently transact. This combined with their tendency to show positions in these networks in which an impact would be spread across few counterparties has implications for the stability of these networks. Interestingly, I did not observe correspondingly strong results for my importance measure in prediction experiments for a dataset of global trade. This suggests that structural importance could be a unique property of sparse transaction networks. I complemented my main results with predictability experiments for sign changes and transaction values. I observed in both cases a lack of predictability from node importance for the equity datasets considered, however predictability was observed when considering instead the global trade dataset. A potential explanation for the differences in these results could be the quality of the transaction reports for the equities datasets.

Although my measures of node and edge importance are defined in a static sense, they provide novel insights into the evolution of the financial markets considered. They consider how the network will respond to changes. This, combined with observations of a relationship between importance and the probability of future changes implies that static representations of networks can contain information about their dynamics. My methods thus have particular potential for application in financial regulation, as regulators could be directed towards market participants who have the largest potential to impact the markets. They also show potential in application to a social network and datasets of global trade which are significantly denser, suggesting that my methods could be applicable in other fields.

My findings linking static importance to the dynamics of financial networks motivated me to consider the generative processes producing the systems observed. I focused specifically on FTSE 100 financial instruments given their high levels of trading activity. I observed that these display bursty characteristics, consistent across multiple days and appears at a similar level for buy trades and sell trades. They also showed dominance of hub nodes as in the non-FTSE systems, with 30-40% of all trades intermediated by the same Central Clearing Counterparty. I also observed different types of market being used, with lit markets dominating but dark and off market trading still showing a significant presence. The generative models for transaction networks that I presented reproduced many of the key cross-sectional

and temporal properties of these systems. I specifically focused on using Hawkes processes for these, which model both self- and mutual excitation in which past transactions influence the presence of future transactions. I found that using Hawkes processes in a univariate sense, such that each transaction excites future transactions regardless of the counterparties making the trade, underestimated the burstiness of the transaction sequences. This suggests that self-excitation at the level of all transactions does not accurately represent the true generative process. However, when considering applying a univariate Hawkes process at the level of individual edges, I found that the burstiness can be reproduced. I also found that the univariate approach, which considers self-excitation without mutual excitation between edges, outperforms a multivariate approach with mutual excitation. When considering the cross sectional properties of the systems, I first observed that selecting edges to transact based on their historical transaction frequency was able to reproduce the static network structure I observe in real systems. Alongside this, I found that the multivariate approach to using Hawkes processes also captures the cross sectional properties well. This complements my observation that historical counterparty relationships are influential in the presence of a future relationship. Finally, I observed strong performance of a bivariate model for buy transactions vs. sell transactions, which suggests that mutual excitation between buys and sells contributes to the overall burstiness of the system.

Overall, my observations that structural importance relates to network evolution and my evidence on the suitability of Hawkes processes as generative models for these systems helps build an understanding of the fragility of these financial systems with respect to participant behaviour. These will be of particular use for data practitioners working with transaction data, providing models and methods for which the parameters help the user to understand how the systems behave. My research will help inform null models for detection of outliers and for the generation of data to test algorithms and transaction monitoring systems. I have also presented a number of complementary observations of the datasets studied and have contributed in the development of methods to clean, aggregate and analyse these datasets. These contribute to the academic community who have an interest in the granular behaviour of equities markets, as well as the Financial Conduct Authority who make use of these datasets for market monitoring and other regulatory use cases.

6.2 Further work

This thesis presents several avenues for future research. Focusing on chapters 3 and 4, my methods of edge and node importance have provided novel insights into how the structural importance of network constituents relates to their subsequent change, and there are several natural next steps to extend these. First of all, it would be interesting to assess the subsequent impact on the network structure of any changes, as it may be that a change to an influential node or edge could act to destabilise a system or could move the system towards a state of stability. It would also be interesting to consider the longer term effects as opposed to solely considering the next time snapshot.

Extensions could also be considered to allow for more exogenous influences to be taken into account, for example the effects of individual transactions on the market price of an instrument, which will in turn have an affect on the probability of others to transact. This would help to connect my research to several studies presented in econometric literature, for example the methods presented in [213]. In this paper the authors explore networks of traders of the S&P 500 Stock index futures contracts and show how network variables preempt financial variables such as volume, duration and market liquidity measures, demonstrating the potential for trading networks in assessing liquidity supply and price formation influencing trading strategies. In relation to this, an interesting avenue for further work would be to compare the predictability observed to that obtained using correlation networks to analyse the more widely available data on stock prices, in a similar way to the comparison presented in [214], who demonstrate that a correlation based approach in combination with methods to analyse direct

exposures provides a useful tool for assessing systemic risk. In their scenario of a limited dataset of direct exposures, the predictive power of correlation based approaches is significantly better than the approach making use of direct exposures.

Additional work could also include considering whether the nodes or edges that are changing in these networks do so persistently, as this would allow me to gain deeper insights into the evolutionary behaviour of these networks. Finally, further analysis could be conducted to assess the effectiveness of my measures as indicators for risk, as so far I understand that my importance measures bear some relationship to how the network subsequently changes, but I have not considered in this thesis the resultant changes of nodes or edges with high values of either metric and how these have an effect on the rest of the network in terms of risk and stability.

I also note some limitations presented by the datasets considered. In particular, note that the equity transaction datasets are sparse and may contain outlier values due to reporting errors. A useful further exploration would be to develop an equivalent to the use of filtering methods from random matrix theory, which are widely used in identifying the relevant structure in correlation networks [215, 216, 217, 218, 219, 220, 221]. This could help counteract the lack of predictability of the value of a node's change in strength for the equity datasets observed in chapter 4, as one explanation for this could be the quality of the transaction reports or the methods of preprocessing applied to the data prior to my analysis. I also note that while care has been taken to account for duplicate trade reporting by identifying trades occurring at the same time and quantity between matching counterparties, reporting quality issues may result in trades being duplicated with differences. Further work is needed to fully account for this.

It is also worth noting some areas where alternatives to the modelling approaches in this thesis could be taken. For example, a limitation of both the edge change model I propose in chapter 3 and also the use of Hawkes processes at the edge level in chapter 5, is that these do not allow for edges to appear and disappear. Understanding how these appearances and disappearances can be captured in a model for network growth would be highly beneficial for future work. Further to this, my experiments making use of logistic regression to predict changes in equity networks can be considered as a probabilistic model for the network dynamics. This complements approaches found in literature on econometric network models, where much of the focus is on identification of models for macroeconomic time series given the wealth of price time series available for analysis. Approaches often make use of Vector Autoregressive models, which relate current observations of a variable with past observations of itself and past observations of other variables in the system [222]. Studies such as [223] have demonstrated how these models have good predictive accuracy and offer a good representation of linkages between economic sectors, making them a useful tool for assessing systemic risk. An interesting further area of development for my methods would be to consider autoregressive approaches as an alternative prediction method. Finally, I note that my work in the first two chapters focused on static snapshots of networks, meaning that although I am capturing the temporal behaviour of these to some extent, the aggregation scale used will have an impact on what I am able to observe. My research in chapter 5 partly addresses this by considering the generative processes that give rise to the transactions I observe, but an interesting avenue for further research would be to consider how the metrics proposed can analytically incorporate the full temporal information. Overall, by researching further into the links between structural importance, network evolution and how these relate to market stability constraints would be an exciting and impactful further avenue to take this work.

In chapter 5, the poorer performance of the multivariate model in reproducing the temporal properties of the datasets, combined with the observation that this approach performed better than the univariate edge model in reproducing the cross sectional properties of the transaction networks, suggest that improvements to this model, for example improving the runtime of estimation of processes with power law decay kernels, and also allowing the kernel decay to vary across the different edges, would be promising avenues for further research. Further exploration into the performance of edge level modelling

in conjunction with the bivariate model would be an interesting avenue of further exploration.

Although in chapter 5 I have considered application of different models to different subsets of the trading ecosystems studied, there are several trading mechanisms that I did not explore in terms of their influence on the systems' properties. For example, in markets such as these which are heavily dominated by trades which are centrally cleared, in order to provide clearing members with increased opportunities to net their orders and to give a reduction in outstanding gross exposures in the system, interoperability agreements exist which allow CCPs to link between counterparties allowing cross-system execution of transactions. These links introduce a direct channel of contagion between these critical nodes in the financial system, so in order for my methods to fully capture the macroprudential risks in these systems, these links would need to be included. Further to this, I have considered application of Hawkes processes to trade executions only so further research into how the dynamics of the order book influences the generative processes of these executions would be useful to explore in addition to the other potential avenues suggested.

Finally, significant work can now be undertaken to implement the methods presented in this thesis to the monitoring of transaction networks in a regulatory setting. My metrics for node and edge importance can be compared to existing methods used to rank transactors for monitoring purposes, with the potential to be incorporated as an early indicator for potential harm. The models I propose making use of Hawkes processes can be used to generate transaction sequences, which can act as a null model to detect unusual transactions within the observed sequences. To be able to do this, an in-depth exploration into how different types of unusual or harmful transactions manifest in the temporal and cross-sectional properties that can be observed is needed. Generative models also open up the opportunity to share synthetic transactions with the academic community and industry for further research into the behaviour of capital markets transaction sequences. An intuitive next step to achieve this goal would be to apply my methods to a large number of networks, both to further verify my observations and would be a useful tool for classifying networks according to their evolutionary properties.

Bibliography

[1] M. Bardoscia, G. Bianconi, and G. Ferrara. Multiplex network analysis of the UK OTC derivatives market. *Bank of England Working Papers*, 2018.

[2] A. Kotlicki, A. Austin, D. Humphry, H. Burnett, P. Ridgill, and S. Smith. Network analysis of the UK reinsurance market. *Bank of England Working Papers*, 2022.

[3] G. Covi and X. Gu. Interbank network and banks' credit supply. *Bank of England Working Papers*, 2022.

[4] BIS/MAS - project ellipse. `https://www.bis.org/about/bisih/topics/suptech_regtech/ellipse.htm`. Accessed: 2022-12-20.

[5] J. Ladyman, J. Lambert, and K. Wiesner. What is a complex system? *European Journal for Philosophy of Science*, 3, 06 2013.

[6] Financial Conduct Authority. FCA business plan 2019/20, 2019.

[7] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli. The physics of financial networks. *Nature Reviews Physics*, page 490–507, 03 2021.

[8] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White. Economic networks: The new challenges. *Science*, 325(5939):422–425, 2009.

[9] I. Seabrook, P. Barucca, and F. Caccioli. Evaluating structural edge importance in temporal networks. *EPJ Data Science*, 10(1), 2021.

[10] I. Seabrook, P. Barucca, and F. Caccioli. Structural importance and evolution: An application to financial transaction networks. *Physica A: Statistical Mechanics and its Applications*, 607:128203, 2022.

[11] I. Seabrook, P. Barucca, and F. Caccioli. Modelling equity transaction networks as bursty processes, 2022.

[12] I. Seabrook, F. Caccioli, and T. Aste. Quantifying impact and response in markets using information filtering networks. *Journal of Physics: Complexity*, 3(2):025004, may 2022.

[13] Y. N. Wu, R. Gao, T. Han, and S. C. Zhu. A tale of three probabilistic families: Discriminative, descriptive and generative models, 2018.

[14] D. Brigo and A. Capponi. Bilateral counterparty risk valuation with stochastic dynamical models and application to credit default swaps, 2008.

[15] D. Brigo, A. Dalessandro, M. Neugebauer, and F. Triki. A stochastic processes toolkit for risk management. *arXiv.org, Quantitative Finance Papers*, 12 2008.

[16] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606, 2010.

[17] E. Sariev and G. Germano. Bayesian regularized artificial neural networks for the estimation of the probability of default. *Quantitative Finance*, 20(2):311–328, 2020.

[18] J. Turiel and T. Aste. Peer-to-peer loan acceptance and default prediction with artificial intelligence. *Royal Society Open Science*, 7:191649, 06 2020.

[19] C. Serrano-Cinca, B. Gutierrez-Nieto, and L. Lopez-Palacios. Determinants of default in p2p lending. *PLOS ONE*, 10(10):1–22, 10 2015.

[20] C. Canfield. Determinants of default in p2p lending: the mexican case. *Independent Journal of Management & Production*, 9:001, 03 2018.

[21] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu. Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics*, 47(1):54–70, 2015.

[22] J. M. Hutchinson, A. W. Lo, and T. Poggio. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The journal of Finance*, 49(3):851–889, 1994.

[23] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[24] J. Sirignano and R. Cont. Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, 19(9):1449–1459, 2019.

[25] C. E. Phelan, D. Marazzina, G. Fusai, and G. Germano. Hilbert transform, spectral filters and option pricing. *Annals of Operations Research*, 282(1-2):273–298, may 2018.

[26] R. Cont and A. de Larrard. Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

[27] R. Cont. Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine*, 28(5):16–25, 2011.

[28] R. Cont and . KUKanov. Optimal order placement in limit order markets. *Quantitative Finance*, 17(1):21–39, 2017.

[29] N. Cohen, T. Balch, and M. Veloso. Trading via image classification. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2020. Association for Computing Machinery.

[30] M. Veloso, T. Balch, D. Borrajo, P. Reddy, and S. Shah. Artificial intelligence research in finance: discussion and examples. *Oxford Review of Economic Policy*, 37(3):564–584, 09 2021.

[31] M. Mahfouz, T. Balch, M. Veloso, and D. Mandic. Learning to classify and imitate trading agents in continuous double auction markets. In *Proceedings of the Second ACM International Conference on AI in Finance*, ICAIF '21, New York, NY, USA, 2021. Association for Computing Machinery.

[32] S. Assefa, D. Dervovic, M. Mahfouz, T. Balch, P. Reddy, and M. Veloso. Generating synthetic data in finance: opportunities, challenges and pitfalls. *NeurIPS -Workshop on AI in Financial Services: Data, Fairness, Explainability, Trustworthiness and Privacy. Vancouver, Canada.*, 2019.

[33] D. Borrajo, M. Veloso, and S. Shah. Simulating and classifying behavior in adversarial environments based on action-state traces: An application to money laundering. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2020. Association for Computing Machinery.

[34] L. Ardon, N. Vadori, T. Spooner, M. Xu, J. Vann, and S. Ganesh. Towards a fully rl-based market simulator. In *Proceedings of the Second ACM International Conference on AI in Finance*, nov 2021.

[35] L. de Castro, J. Chen, and A. Polychroniadou. Cryptocredit. In *Proceedings of the First ACM International Conference on AI in Finance*, oct 2020.

[36] D. Byrd, S. Bajaj, and T. Hybinette Balch. Fund asset inference using machine learning methods: What's in that portfolio? In *The Journal of Financial Data Science*, 2019.

[37] F. Pozzi, T. Di Matteo, and T. Aste. Spread of risk across financial markets: Better to invest in the peripheries. *Scientific reports*, 3:1665, 04 2013.

[38] P. F. Procacci and T. Aste. Portfolio optimization with sparse multivariate modelling, 2021.

[39] F. Caccioli, P. Barrucca, and T. Kobayashi. Network models of financial systemic risk: A review. *Journal of Computational Social Science*, 2018.

[40] P. Glasserman and H. P. Young. Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831, September 2016.

[41] R. Bookstaber, J. Cetina, G. Feldberg, M. Flood, and P. Glasserman. Stress tests to promote financial stability: Assessing progress and looking to the future. *Journal of Risk Management in Financial Institutions*, 2014.

[42] F. Musciotto, L. Marotta, J. Piilo, and R. Mantegna. Long-term ecology of investors in a financial market. *Palgrave Communications*, 4:92, 07 2018.

[43] R. May, S. Levin, and G. Sugihara. Complex systems: Ecology for bankers. *Nature*, 451:893–5, 03 2008.

[44] C. Chakraborty and A. Joseph. Machine learning at central banks. Bank of England working papers 674, Bank of England, September 2017.

[45] K. Bluwstein, M. Buckmann, A. Joseph, M. Kang, S. Kapadia, and O. Simsek. Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach. Bank of England working papers 848, Bank of England, January 2020.

[46] S. Doerr, L. Gambacorta, and J. M. Serena Garralda. Big data and machine learning in central banking. BIS Working Papers 930, Bank for International Settlements, March 2021.

[47] Weighted graph - networkx 2.8. `https://networkx.org/documentation/stable/auto_examples/drawing/plot_weighted_graph.htmlhttps://networkx.org/documentation/stable/auto_examples/drawing/plot_weighted_graph.html`. Accessed: 17th July 2022.

[48] M. Newman. *Networks: an introduction.* Oxford University Press, isbn: 978-0-19-920665-0, 2010.

[49] J. Golbeck. Chapter 3 - network structure and measures. In J. Golbeck, editor, *Analyzing the Social Web*, pages 25–44. Morgan Kaufmann, Boston, 2013.

[50] C. Perez and R. Germon. Chapter 7 - graph creation and analysis for linking actors: Application to social data. In R. Layton and P. A. Watters, editors, *Automating Open Source Intelligence*, pages 103–129. Syngress, Boston, 2016.

[51] L. Lu and M. Zhang. *Edge Betweenness Centrality*, pages 647–648. Springer New York, New York, NY, 2013.

[52] J. J. McAuley, L. da Fontoura Costa, and T. S. Caetano. Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, 91(8):084103, Aug 2007.

[53] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20), oct 2002.

[54] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2), Feb 2003.

[55] F. Caccioli, T. A. Catanach, and J. D. Farmer. Heterogeneity, correlations and financial contagion. *Advances in Complex Systems*, 15(supp02):1250058, 2012.

[56] D. Garlaschelli and M. I. Loffredo. Patterns of link reciprocity in directed networks. *Physical Review Letters*, 93(26), Dec 2004.

[57] P. Holme and J. Saramaki. Temporal networks. *Phys. Rep. 519, 97-125 (2012). arXiv:1108.1780 [nlin.AO]*, 2010.

[58] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia. Analysing information flows and key mediators through temporal centrality metrics. *SNS '10: Proceedings of the 3rd Workshop on Social Network Systems Article No.: 3 Pages 1–6 https://doi.org/10.1145/1852658.1852661*, 2010.

[59] N. Pedreschi, D. Battaglia, and A. Barrat. The temporal rich club phenomenon. *Nature Physics*, 2022.

[60] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. *Journal of Computer and System Sciences*, 64(4):820–842, 2002.

[61] J. Tang, I. Leontiadis, S. Scellato, V. Nicosia, C. Mascolo, M. Musolesi, and V. Latora. Applications of temporal graph metrics to real-world networks. *Temporal Networks*, page 135–159, 2013.

[62] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia. Analysing information flows and key mediators through temporal centrality metrics. In *Proceedings of the 3rd Workshop on Social Network Systems*, SNS '10, New York, NY, USA, 2010. Association for Computing Machinery.

[63] L. Lv, K. Zhang, T. Zhang, X. Li, Q. Sun, L. Zhang, and W. Xue. Eigenvector-based centralities for multilayer temporal networks under the framework of tensor computation. *Expert Systems with Applications*, 184:115471, 2021.

[64] E. Paxinou, D. Kalles, C. T. Panagiotakopoulos, and V. S. Verykios. Analyzing sequence data with Markov chain models in scientific experiments. *SN Computer Science*, 2(5):385, Jul 2021.

[65] G. Bianconi and A.L. Barabasi. Competition and multiscaling in evolving networks. *Europhysics Letters. 54 (4): 436–442. arXiv:cond-mat/0011029*, 2001.

[66] P. Mazzarisi, P. Barucca, F. Lillo, and D. Tantari. A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. *European Journal of Operational Research*, 281(1):50–65, 2020.

[67] C. Lee and D.J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1), dec 2019.

[68] M. Ludkin, I. Eckley, and P. Neal. Dynamic stochastic block models: parameter estimation and detection of changes in community structure. *Statistics and Computing*, 28, 11 2018.

[69] K. S. Xu. Stochastic block transition models for dynamic networks. *arXiv pre-print*, 2014.

[70] K. Shang, B. Yang, J. M. Moore, Q. Ji, and M. Small. Growing networks with communities: A distributive link model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(4):041101, 2020.

[71] D. Zambon, D. Grattarola, L. Livi, and C. Alippi. Autoregressive models for sequences of graphs. In *Autoregressive Models for Sequences of Graphs*, pages 1–8, 07 2019.

[72] O. Shchur, A. C. Turkmen, T. Januschowski, and S. Gunnemann. Neural temporal point processes: A review. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4585–4593. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Survey Track.

[73] J. Yan, H. Xu, and L. Li. Modeling and applications for temporal point processes. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &; Data Mining*, KDD '19, page 3227–3228, New York, NY, USA, 2019. Association for Computing Machinery.

[74] E. Bacry, I. Mastromatteo, and J.F. Muzy. Hawkes processes in finance, 2015.

[75] N. Dalmasso, R. E. Tillman, P. Reddy, and M. Veloso. Payvae: A generative model for financial transactions. *AAAI 2021 Workshop on Knowledge Discovery from Unstructured Data in Financial Services Workshop*, 2021.

[76] MIT. *Discrete Stochastic Processes*. MIT OpenCourseWare, 2011.

[77] D. Ilalan. A poisson process with random intensity for modeling financial stability. *The Spanish Review of Financial Economics*, 2015.

[78] A. Kreinin. *Correlated Poisson Processes and Their Applications in Financial Modeling*, pages 191–232. Wiley, 04 2016.

[79] E. Scalas. Mixtures of compound poisson processes as models of tick-by-tick financial data. *Chaos, Solitons & Fractals*, 34(1):33–40, 2007. In Search of a Theory of Complexity.

[80] S. Rajaram, T. Graepel, and R. Herbrich. Poisson-networks: A model for structured point processes. *AISTATS 2005 - Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, 01 2005.

[81] T. C. Brown and P. K. Pollett. Poisson approximations for telecommunications networks (vol 32, pg 348, 1991). *Journal of the Australian Mathematical Society Series B-Applied Mathematics*, 36(01):132–132, 1994.

[82] R. Houssou, J. Bovay, and S. Robert. Adaptive financial fraud detection in imbalanced data with time-varying poisson processes. *Journal of Financial Risk Management*, 08:286–304, 01 2019.

[83] M. Feng, H. Qu, Z. Yi, X. Xie, and J. Kurths. Evolving scale-free networks by poisson process: Modeling and degree distribution. *IEEE Transactions on Cybernetics*, 46(5):1144–1155, 2016.

[84] S. Linderman and R. Adams. Discovering latent network structure in point process data, 22–24 Jun 2014.

[85] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[86] N. Masuda, T. Takaguchi, N. Sato, and K. Yano. Self-exciting point process modeling of conversation event sequences. In *Understanding Complex Systems*, pages 245–264. Springer Berlin Heidelberg, 2013.

[87] C. Zhang. Modeling high frequency data using hawkes processes with power-law kernels. *Procedia Computer Science*, 80:762–771, 2016. International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA.

[88] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5), may 2012.

[89] Stephen Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: A hawkes process analysis. *Physics of Condensed Matter*, 86, 02 2013.

[90] E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5), may 2012.

[91] F. Chen and P. Hall. Inference for a nonstationary self-exciting point process with an application in ultra-high frequency financial data modeling. *Journal of Applied Probability*, 50, 12 2013.

[92] P. J. Laub, T. Taimre, and P. K. Pollett. Hawkes processes, 2015.

[93] A.G. Hawkes. Hawkes processes and their applications to finance: a review. *Quantitative Finance*, 18(2):193–198, 2018.

[94] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10(1):1–25, 2007.

[95] M. Rambaldi, E. Bacry, and F. Lillo. The role of volume in order book dynamics: a multivariate hawkes process analysis. *Quantitative Finance*, 17(7):999–1020, Dec 2016.

[96] A. Cartea, S. Jaimungal, and J. Ricci. Algorithmic trading, stochastic control, and mutually exciting processes. *SIAM Review*, 60:673–703, 01 2018.

[97] L. Xiaofei and A. Frederic. High-dimensional hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. *Quantitative Finance*, 18(2):249–264, 2018.

[98] E. Bacry, A. Iuga, M. Lasnier, and C. A. Lehalle. Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 01(02):1550009, 2015.

[99] M. Morariu-Patrichi and M.S. Pakkanen. State-dependent hawkes processes and their application to limit order book modelling. *Quantitative Finance*, 2021.

[100] H. Unwin, I. Routledge, S. Flaxman, M. Rizoiu, S. Lai, J. Cohen, D. Weiss, S. Mishra, and S. Bhatt. Using hawkes processes to model imported and local malaria cases in near-elimination settings. *PLOS Computational Biology*, 17:e1008830, 04 2021.

[101] E. Choi, N. Du, R. Chen, L. Song, and J. Sun. Constructing disease network and temporal progression model via context-sensitive hawkes process. In *2015 IEEE International Conference on Data Mining*, pages 721–726, 2015.

[102] J. R. Zipkin, F. P. Schoenberg, K. Coronges, and A. L. Bertozzi. Point-process models of social network interactions: Parameter estimation and missing data recovery. *European Journal of Applied Mathematics*, 27(3):502–529, 2016.

[103] J. Pinto, T. Chahed, and E. Altman. Trend detection in social networks using hawkes processes. *6th International Workshop on Mining and Analyzing Social Networks for Decision Support (MSNDS 2015) in conjunction with IEEE/ACM ASONAM 2015*, 08 2015.

[104] V. KUKreti, H. K. Pharasi, P. Gupta, and S. Kumar. A perspective on correlation-based financial networks and entropy measures. *Frontiers in Physics*, 8, 2020.

[105] F. Musciotto, J. Piilo, and R.N. Mantegna. High-frequency trading and networked markets. *Proceedings of the National Academy of Sciences*, 118(26):e2015573118, 2021.

[106] V. Van Kervel and A. J. Menkveld. High-frequency trading around large institutional orders. *The Journal of Finance*, 74(3):1091–1137, 2019.

[107] B. Hagströmer and L. Nordén. The diversity of high-frequency traders. *Journal of Financial Markets*, 16(4):741–770, 2013. High-Frequency Trading.

[108] B. Hagströmer, L. Nordén, and D. Zhang. How aggressive are high-frequency traders? *Financial Review*, 49(2):395–419, 2014.

[109] S. A. Assefa, D. Dervovic, M. Mahfouz, R. E. Tillman, P. Reddy, and M. Veloso. Generating synthetic data in finance: Opportunities, challenges and pitfalls. In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2020. Association for Computing Machinery.

[110] B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, 50(5):1655–1689, 1995.

[111] K. Barbieri and Omar K. Correlates of war project trade data set codebook, version 4.0, 2016.

[112] A. Liberatore and S. Wettstein. The OECD-WTO balanced trade in services database (bpm6 edition), 2021.

[113] X. Ren and L. Lu. Review of ranking nodes in complex networks. *Chinese Science Bulletin 59. 1175. 10.1360/972013-1280.*, 2014.

[114] C. Stegehuis, R. van der Hofstad, and J. van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Sci Rep 6, 29748 https://doi.org/10.1038/srep29748*, 2016.

[115] G. Ren and X. Wang. Epidemic spreading in time-varying community networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(2):023116, 2014.

[116] Y.C. Lai, A.E. Motter, and T. Nishikawa. Attacks and cascades in complex networks. *In: Ben-Naim E., Frauenfelder H., Toroczkai Z. (eds) Complex Networks. Lecture Notes in Physics, vol 650. Springer, Berlin, Heidelberg*, 2004.

[117] H. Xu, J. Zhang, J Yang, and L. Lun. Identifying important nodes in complex networks based on multiattribute evaluation. *Mathematical Problems in Engineering, vol. Article ID 8268436, https://doi.org/10.1155/2018/8268436*, 2018.

[118] Y. Moreno, M. Nekovee, and A. F. Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E*, 69:066130, Jun 2004.

[119] J.G. Restrepo, E. Ott, and B.R. Hunt. Weighted percolation on directed networks. *Physical Review letters E71,036151*, 2005.

[120] B. Bollobas, C. Borgs, J. Chayes, and O. Riordan. Percolation on dense graph sequences. *The Annals of Probability*, 38(1):150–183, 2010.

[121] Y. Wang, Z. Di, and Y. Fan. Identifying and characterizing nodes important to community structure using the spectrum of the graph. *PLoS ONE 6(11) e27418.*, 2011.

[122] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. *22nd International Symposium on Reliable Distributed Systems. Proceedings.*, 2003.

[123] S. Pei and H.A. Makse. Spreading dynamics in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013.

[124] E. Stanley, L. Lu, L. Pan, T. Zhou, and Y. Zhang. Toward link predictability of complex networks. *112 (8) 2325-2330; DOI: 10.1073/pnas.1424644112*, 2015.

[125] J. Restrepo, E. Ott, and B. Hunt. Characterizing the dynamical importance of network nodes and links. *Physical review letters*, 97:094102, 10 2006.

[126] M.E. Helander and S. McAllister. The gravity of an edge. *Applied Network Science 3, 7*, 2018.

[127] E. Yu, D. Chen, and J. Zhao. Identifying critical edges in complex networks. *Scientific Reports 8, 14469 https://doi.org/10.1038/s41598-018-32631-8*, 2018.

[128] P Jaccard. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50, 1912.

[129] L.C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[130] M. A. Javed, M.S. Younis, S. Latif, J. Qadir, and A. Baig. Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108:87–111, 2018.

[131] Z. Yang, R. Algesheimer, and C. Tessone. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 6, 08 2016.

[132] M. Plantié and M. Crampes. Survey on social community detection. In *Social media retrieval*, pages 65–85. Springer, 2013.

[133] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[134] S. Fortunato, V. Latora, and M. Marchiori. Method to find community structures based on information centrality. *Physical Review E*, 70:056104, Nov 2004.

[135] J. Chan-Lau. Systemic centrality and systemic communities in financial networks. *Quantitative Finance and Economics*, 2:468–496, 06 2018.

[136] U. Luxburg. A tutorial on spectral clustering. *Kluwer Academic Publishers*, 17(4):395–416, 2007.

[137] E. Estrada. Generalized walks-based centrality measures for complex biological networks. *Journal of Theoretical Biology*, 263(4):556, 2010.

[138] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, Mar 1953.

[139] A. Collins, C. Engströmb, J. Magero, H. Kasumbaa, S. Silvestrov, and B. Abola. Eigenvector centrality and uniform dominant eigenvalue of graph components, 2021.

[140] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[141] V. Nr and P. Fjallstrom. Algorithms for graph partitioning: A survey. *Linkoping Electronic Articles in Computer and Information Science*, 3, 10 1998.

[142] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), Sep 2006.

[143] T. Aste, W. Shaw, and T. Di Matteo. Correlation structure and dynamics in volatile markets. *New Journal of Physics*, 2010.

[144] P. Barucca and F. Lillo. Disentangling bipartite and core-periphery structure in financial markets. *Chaos, solitons and fractals*, 2015.

[145] B. Barucca, M. Bardoscia, F. Caccioli, M. D'Errico, G. Visentin, S. Battiston, and G. Caldarelli. Network valuation in financial systems. *Mathematical Finance.30:1181–1204.*, 2020.

[146] K. Soramäki, M. Bech, J. Arnold, R. Glass, and W. Beyeler. The topology of interbank payment flows. *Physica A 379 (2007) 317–333*, 2007.

[147] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[148] M. Falkenberg, J. Lee, S. Amano, K. Ogawa, K. Yano, Y. Miyake, T. Evans, and K. Christensen. Identifying time dependence in network growth. *Physical Review Research 2, 023352*, 2020.

[149] M. Bazzi, L. G. S. Jeub, A. Arenas, S. D. Howison, and M. A. Porter. A framework for the construction of generative models for mesoscale structure in multilayer networks. *Physical Review Research*, 2:023100, Apr 2020.

[150] T. P. Peixoto and M. Rosvall. Modelling sequences and temporal networks with dynamic community structures. *Nature Communications 8, 582, https://doi.org/10.1038/s41467-017-00148-9*, 2017.

[151] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature 393, 440–442 https://doi.org/10.1038/30918*, 1998.

[152] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89:258702, Dec 2002.

[153] T. Kobayashi and M. Génois. Two types of densification scaling in the evolution of temporal networks. *Physical Review E 102, 052302 2020*, 2020.

[154] R. Clark, G. Punzo, and M. Macdonald. Network communities of dynamical influence. *Scientific Reports*, 9(1), Nov 2019.

[155] A. Alsayed and D. Higham. Betweenness in time dependent networks. *Chaos, Solitons & Fractals*, 72:35–48, 2015. Multiplex Networks: Structure, Dynamics and Applications.

[156] E. Estrada. Communicability in temporal networks. *Physical Review E*, 88:042811, Oct 2013.

[157] P. Grindrod and D. Higham. A dynamical systems view of network centrality. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 470, 05 2014.

[158] P. Grindrod, M. Parsons, D. Higham, and E. Estrada. Communicability across evolving networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 83:046120, 04 2011.

[159] R. K. Pan and J. Saramaki. Path lengths, correlations, and centrality in temporal networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 84:016105, 07 2011.

[160] L. E. C. Rocha and N. Masuda. Random walk centrality for temporal networks. *New Journal of Physics*, 16(6):063023, jun 2014.

[161] L. Zhao, G. J. Wang, M. Wang, W. Bao, W. Li, and H. E. Stanley. Stock market as temporal network. *Physica A: Statistical Mechanics and its Applications*, 506:1104–1112, 2018.

[162] D. Taylor, S. Myers, A. Clauset, M. Porter, and P. Mucha. Eigenvector-based centrality measures for temporal networks. *Multiscale Modeling & Simulation*, 15, 07 2015.

[163] H. Kim, J. Tang, R. Anderson, and C. Mascolo. Centrality prediction in dynamic human contact networks. *Computer Networks*, 56(3):983–996, 2012. (1) Complex Dynamic Networks (2) P2P Network Measurement.

[164] R. M. May. Will a large complex system be stable? *NATURE VOL. 238 AUGUST 18 1972*, 1972.

[165] F. Restoy. Market integration: the role of regulation. *Speech by Chairman of the Financial Stability Institute, Bank for International Settlements, at the IIF Market fragmentation roundtable, Washington DC, United States, 10 April 2019.*, 2019.

[166] P. A. Samuelson. General proof that diversification pays. *The Journal of Financial and Quantitative Analysis, 2(1):pp. 1–13*, 1967.

[167] M. Bardoscia, S. Battiston, F. Caccioli, and G. Caldarelli. Pathways towards instability in financial networks. *Nature Communications 8:14416 | DOI: 10.1038/ncomms14416*, 2016.

[168] S. Markose, S. Giansante, M. Gatkowski, and A. R. Shaghaghi. Too interconnected to fail: Financial contagion and systemic risk in network model of cds and other credit enhancement obligations of us banks. *Comisef Working Papers Series WPS-033*, 2010.

[169] F. Caccioli, M. Marsili, and P. Vivo. Eroding market stability by proliferation of financial instruments. *The European Physical Journal B*, 71(4):467–479, 2009.

[170] J. Tobin. The new economics one decade older. *The Eliot Janeway Lectures on Historical Economics in Honour of Joseph Schumpeter 1972 (Princeton University Press, Princeton, US);Eastern Economic Journal IV153-159*, 1978.

[171] G. Bianconi, T. Galla, M. Marsili, and P. Pin. Effects of tobin taxes in minority game markets. *Journal of Economic Behavior and Organization, Volume 70, Issues 1–2*, 2009.

[172] W.A. Brock, C.H. Hommes, and Wagener F.O.O. More hedging instruments may destabilize markets. *Journal of Economic Dynamics & Control 33 1912–1928*, 2009.

[173] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli. Debtrank: Too central to fail?financialnetworks, the fed and systemic risk. *Scientific Reports 2,541; DOI:10.1038/srep00541*, 2012.

[174] P. Barucca and F. Lillo. The organization of the interbank network and how ecb unconventional measures affected the e-mid overnight market. *Computational Management Science (2017), arxiv: 1511.08068*, 2017.

[175] J. D. Turiel, P. Barucca, and T. Aste. Simplicial persistence of financial markets: filtering, generative processes and portfolio risk, 2020.

[176] T. Nayak, S. Sharma, Y. Butala, K. Dasgupta, P. Goyal, and N. Ganguly. A generative approach for financial causality extraction. In *Companion Proceedings of the Web Conference 2022*, WWW '22, page 576–578, New York, NY, USA, 2022. Association for Computing Machinery.

[177] A. Todd, P. Beling, B. Scherer, and S. Yang. Agent-based financial markets: A review of the methodology and domain. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–5, 12 2016.

[178] E. Nier, J. Yang, T. Yorulmazer, and A. Alentorn. Network models and financial stability. *Bank of England Working Paper no. 346*, 2008.

[179] T.C. Silva, S. M. Guerra, B. M. Tabak, and R.C. de Castro Miranda. Financial networks, bank efficiency and risk-taking. *Banco Central do Brasil Working Paper n. 428.*, 2016.

[180] H. J. Allen. Financial stability regulation as indirect investor/consumer protection regulation: Implications for regulatory mandates and structure. *Legal Studies Research Paper Series Research Paper 17-5*, 2017.

[181] A. Haldane and R. May. Systemic risk in banking ecosystems. *Nature 469, 351–355 https://doi.org/10.1038/nature09659*, 2011.

[182] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proceedings of the Seconds Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1950.

[183] A. Fernndez, S. Garca, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera. *Learning from Imbalanced Data Sets*. Springer Publishing Company, Incorporated, 1st edition, 2018.

[184] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.

[185] V. Raghavan, P. Bollmann, and G. S. Jung. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Trans. Inf. Syst.*, 7(3):205–229, jul 1989.

[186] P. Panzarasa, T. Opsahl, and K. M. Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community., 2009.

[187] F.M. Bandi, B. Perron, A. Tamoni, and C. Tebaldi. The scale of predictability. *Journal of Econometrics*, 208(1):120 – 140, 2019. Special Issue on Financial Engineering and Risk Management.

[188] X. Fang, B. Wang, L. Liu, and L. Song. Heterogeneous traders, the leverage effect and volatility of the chinese p2p market. *Journal of Management Science and Engineering*, 3(1):39–57, 2018.

[189] K. Khashanah and T. Alsulaiman. Network theory and behavioral finance in a heterogeneous market environment. *Complexity*, 21, 09 2016.

[190] S. Bikhchandani and S. Sharma. Herd behavior in financial markets. *IMF Staff Papers*, 47(3):279–310, 2000.

[191] H. Kim, J. Tang, R. Anderson, and C. Mascolo. Centrality prediction in dynamic human contact networks. *Computational Networks*, 56(3):983–996, 2012.

[192] F. Takes and E. Heemskerk. Centrality in the global network of corporate control. *Social Network Analysis and Mining*, 6, 10 2016.

[193] A. Fisher, C. Rudin, and F. Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019.

[194] P. Giudici and E. Raffinetti. Shapley-lorenz explainable artificial intelligence. *Expert Systems with Applications*, 167:114104, 2021.

[195] S. M. Lundberg and S. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[196] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, Feb 2004.

[197] A. G. Haldane. *Rethinking the financial network - Fragile Stabilität – stabile Fragilität*, pages 243–278. Springer Fachmedien Wiesbaden, Wiesbaden, 2013.

[198] FCA datasprint kernel description. `https://www.fca.org.{UK}/firms/innovation/digital-sandbox-pilot-datasprint`. Accessed: 2021-03-05.

[199] V. Boginski, S. Butenko, and P. M. Pardalos. Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, 48(2):431–443, 2005.

[200] J. Engel, A. Pagano, and M. Scherer. Reconstructing the topology of financial networks from degree distributions and reciprocity. *Journal of Multivariate Analysis*, 172:210–222, 2019. Dependence Models.

[201] T. C. Johnson. Reciprocity as a foundation of financial economics. *Journal of Business Ethics*, 131(1):43–67, 2015.

[202] T. R. Hurd, J. P. Gleeson, and S. Melnik. A framework for analyzing contagion in assortative banking networks. *PLOS ONE*, 12, 02 2017.

[203] D. Fricke, K. Finger, and T. Lux. On assortative and disassortative mixing in scale-free networks: The case of interbank credit networks. In *Kiel Working Paper, No. 1830, Kiel Institute for the World Economy (IfW), Kie*, 2013.

[204] B. Tao, H. Dai, J. Wu, I. Ho, Z. Zheng, and C. F. Cheang. Complex network analysis of the bitcoin transaction network. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(3):1009–1013, 2022.

[205] J. Alstott, P. Panzarasa, M. Rubinov, E.T. Bullmore, and P. Vertes. A unifying framework for measuring weighted rich clubs. *Scientific Reports*, 4, 2014.

[206] Y. W. Cheung and K. S. Lai. Lag order and critical values of the augmented dickey-fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280, 1995.

[207] K. Zhou, H. Zha, and L. Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In Carlos M. Carvalho and Pradeep Ravikumar, editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pages 641–649, Scottsdale, Arizona, USA, 2013. PMLR.

[208] E. Bacry, M. Bompaire, S. Gaiffas, and S. Poulsen. tick: a python library for statistical learning, with a particular emphasis on time-dependent modeling. *Journal of Machine Learning Research*, 18(214):1–5, 2018.

[209] F. J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[210] M. B. Brown and A.B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

[211] H. Xu and W. Zhou. Modeling aggressive market order placements with hawkes factor models. *PLOS ONE*, 15(1):1–12, 01 2020.

[212] H. Ba. Improving detection of credit card fraudulent transactions using generative adversarial networks. *ArXiv*, abs/1907.03355, 2019.

[213] L. Adamic, C. Brunetti, J. H. Harris, and A. Kirilenko. Trading networks. *Econometrics Journal*, 20(3):126–149, October 2017.

[214] P. Giudici, P. Sarlin, and A. Spelta. The multivariate nature of systemic risk: Direct and common exposures. *SSRN Electronic Journal*, 01 2016.

[215] M. Tumminello, F. Lillo, and R. Mantegna. Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75:40–58, 09 2009.

dataset	connectivity	density	reciprocity	assortativity
Equity - 1	30%	12.99%	67.72%	-23.19%
Equity - 2	39.2%	42.14%	75.03%	-25.97%
Equity - 3	34.58%	13.2%	57.66%	-36.84%

Table 1: Network statistics for the three Equity datasets

[216] X. Guo, H. Zhang, and T. Tian. Development of stock correlation networks using mutual information and financial big data. *PLOS ONE*, 13(4):1–16, 04 2018.

[217] W. Barfuss, G. P. Massara, T. Di Matteo, and T. Aste. Parsimonious modeling with information filtering networks. *Physical Review E*, 94(6), dec 2016.

[218] T. Aste. Topological regularization with information filtering networks. *Information Sciences*, 608:655–669, 2022.

[219] J. Daly, M. Crane, and H. J. Ruskin. Random matrix theory filters in portfolio optimisation: A stability and risk assessment. *Physica A: Statistical Mechanics and its Applications*, 387(16):4248–4260, 2008.

[220] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426, 2005.

[221] G. P. Massara, T. Di Matteo, and T. Aste. Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, 5(2):161–178, 06 2016.

[222] H. Lutkepohl. Vector autoregressive models. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pages 139–164. Edward Elgar Publishing, 2013.

[223] D.F. Ahelegbey, M. Billio, and R. Casarin. Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics*, 31(2):357–386, 2016.

[224] J. Karel and K. Martin. Matching algorithms of international exchanges, 2007.

[225] European Central Bank. Financial stability review, 2015.

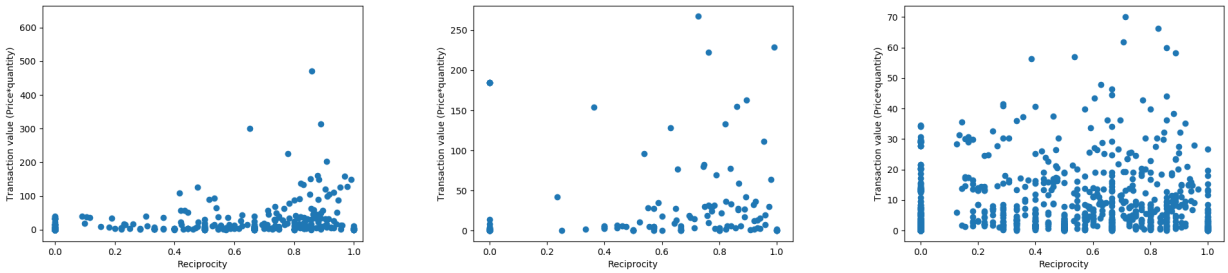
.1 Appendix

.1.1 Additional results for chapter 3

Summary statistics for the Equity 1-3 datasets

Here I present an exploration of the three datasets considered in chapter 3, to highlight some of their key properties. I aggregate the transactions daily, ignore non-trading days such as weekends and holidays and construct networks with the market participants as the network nodes and edges representing the net value traded between two market participants. All statistics are based on the networks following the removal of nodes which appear on less than 5 days in the sample, which I classed as 'inactive'.

Table 1 shows key network level statistics for these three datasets. These show that all three networks have similar connectivities, but Equity-3 is significantly denser than the other two and shows a higher level of reciprocity which would be expected for a denser network. All datasets have a similar level of assortativity, which is negative, a reflection of the prevalence of a small number of hub nodes.



(a) Equity-1 (b) Equity-2 (c) Equity-3

Figure 1: Reciprocity vs. price for the Equity networks

Figure 1 demonstrates that large transaction values are more likely to have a high reciprocity for the first and second dataset. However the same cannot be said for the third Equity dataset.

The evolution of high level network statistics are shown in figures 2,3 and 4. These show that the networks fluctuate around a relatively stable mean, with no obvious level of growth or decay across the time period.

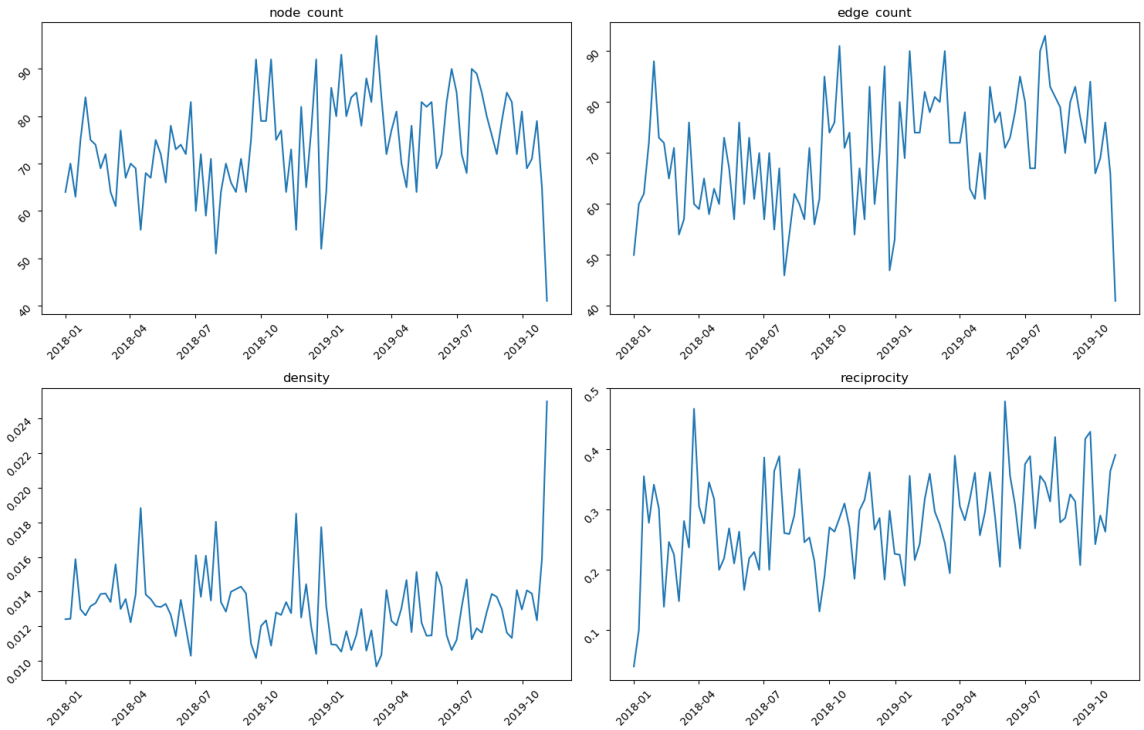


Figure 2: Daily counts of nodes and edges, density and reciprocity across the entire investigation period for Equity-1

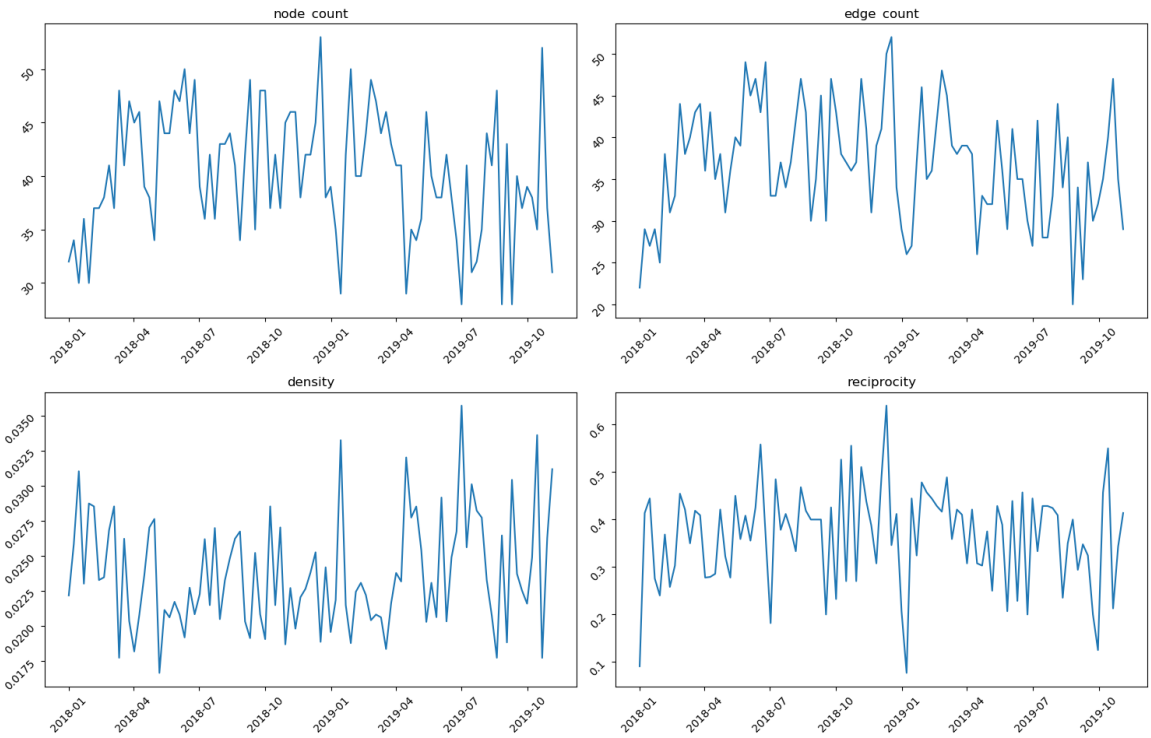


Figure 3: Daily counts of nodes and edges, density and reciprocity across the entire investigation period for Equity-2

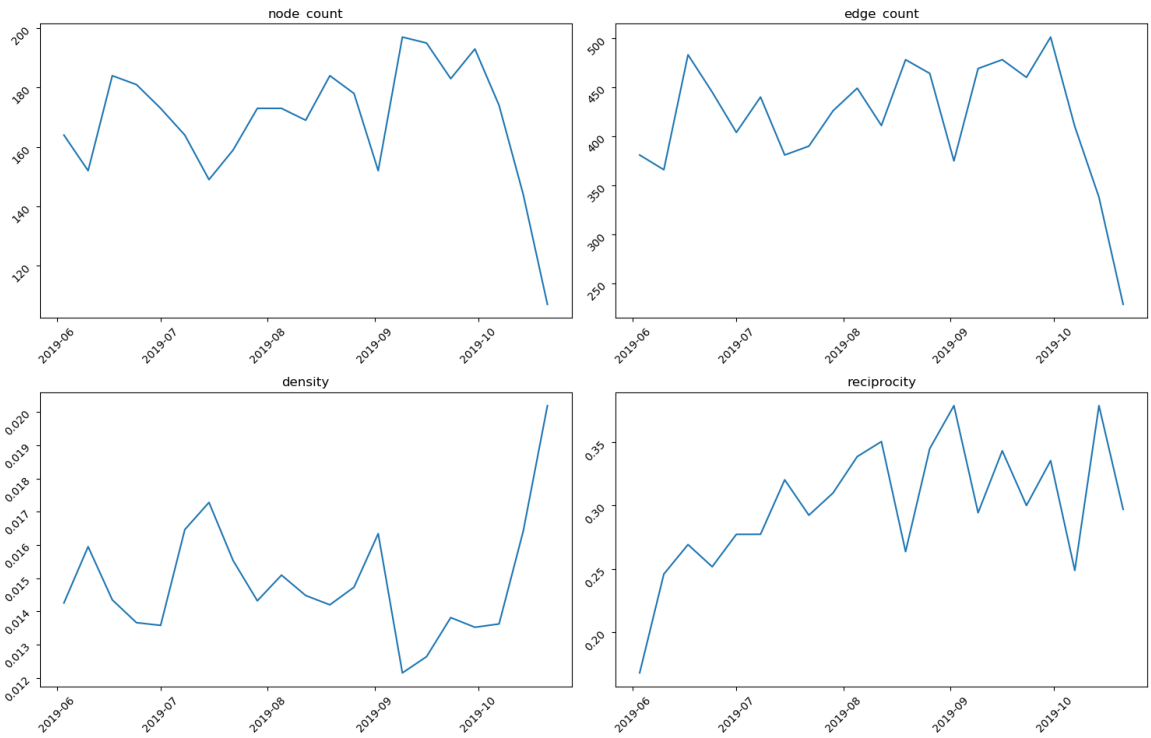


Figure 4: Daily counts of nodes and edges, density and reciprocity across the entire investigation period for Equity-3

Perturbing nodes in toy networks

The plots shown in figures 5 - 7 show additional results for the case of perturbing single edges for weighted barbell networks, and unweighted ring and Erdős–Rényi networks.

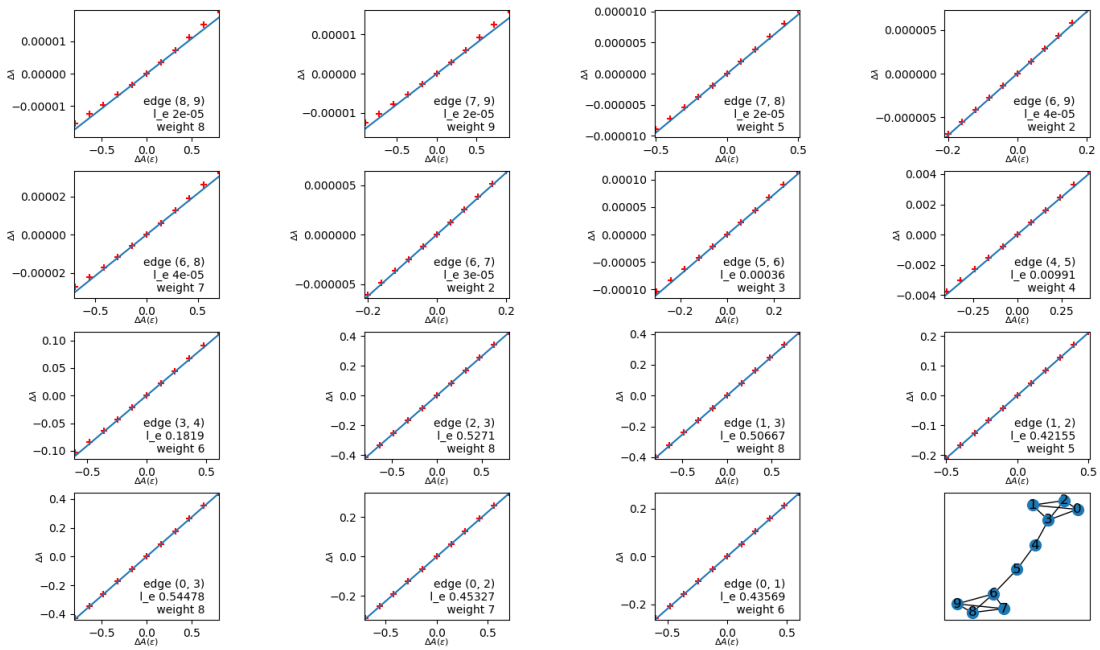


Figure 5: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Weighted barbell network

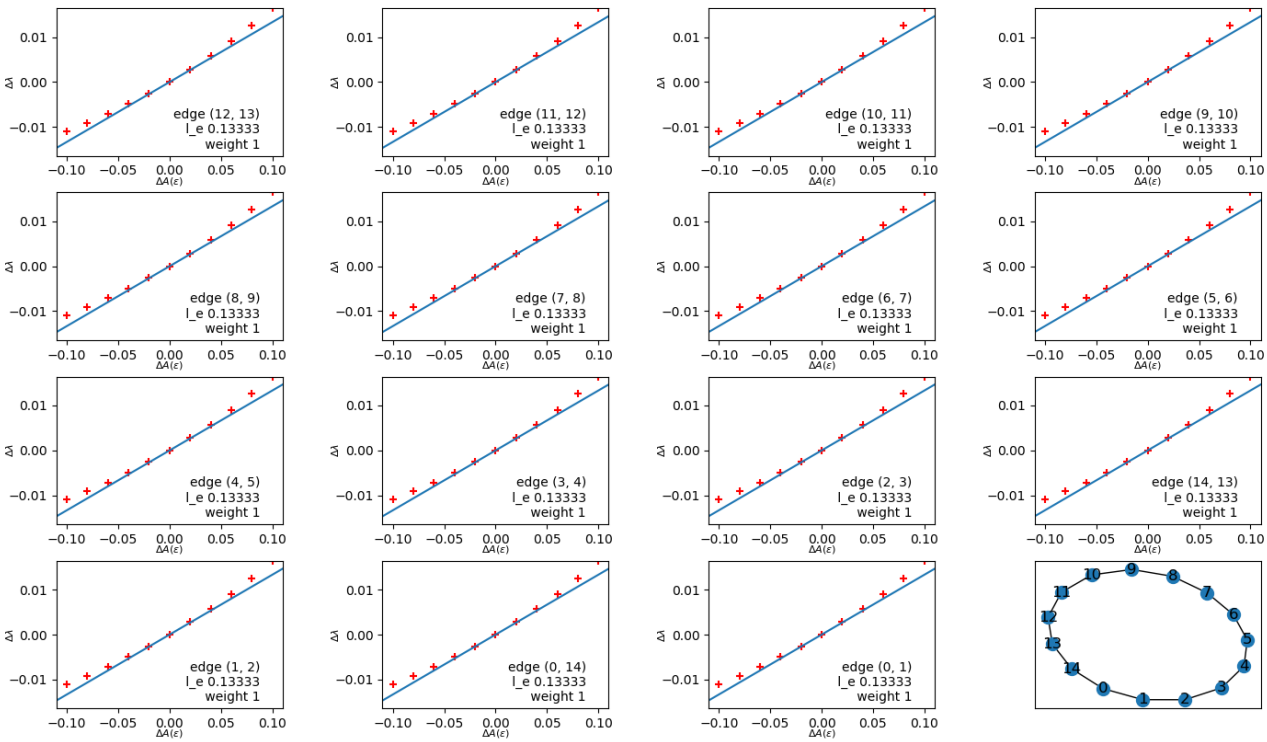


Figure 6: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Unweighted ring network.

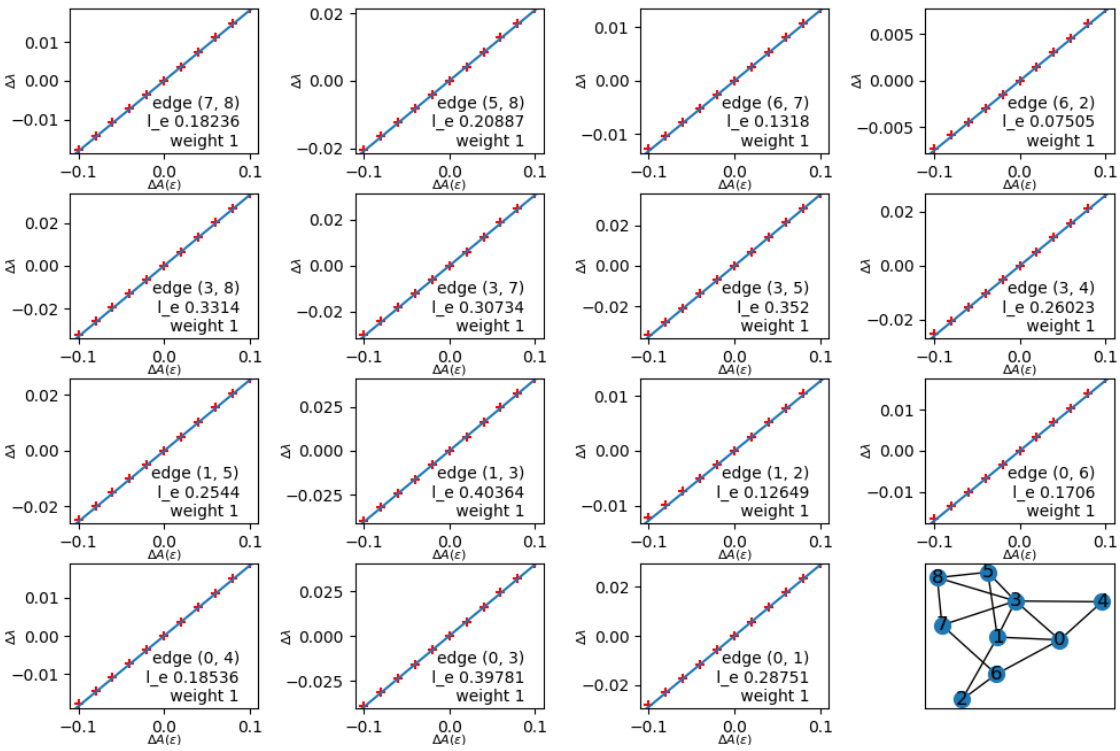


Figure 7: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e. Unweighted Erdős–Rényi network

In figures 8 - 13 I present the results of changing two individual edges, and observing the resulting change in λ for the range of perturbations applied. I overlay this with a line of constant l_e, to assess the performance of my approximation.

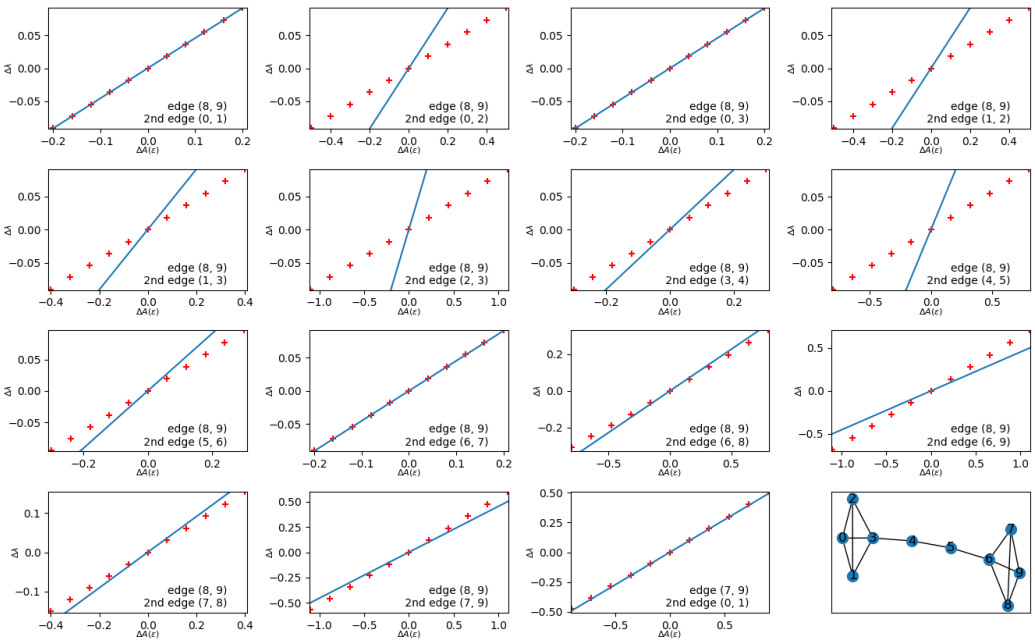


Figure 8: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Barbell network, with equal initial weights

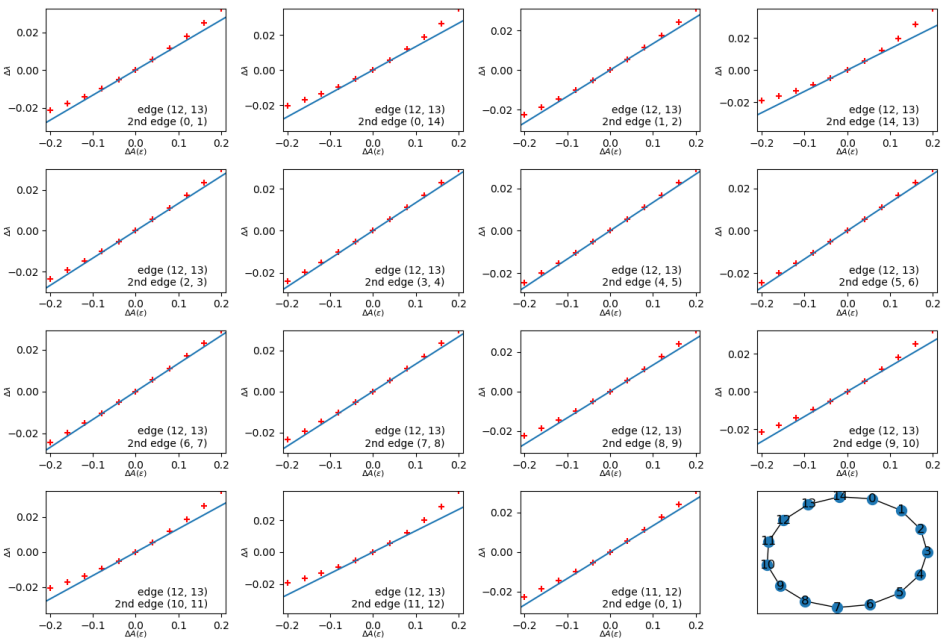


Figure 9: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Ring network with randomly assigned weights.

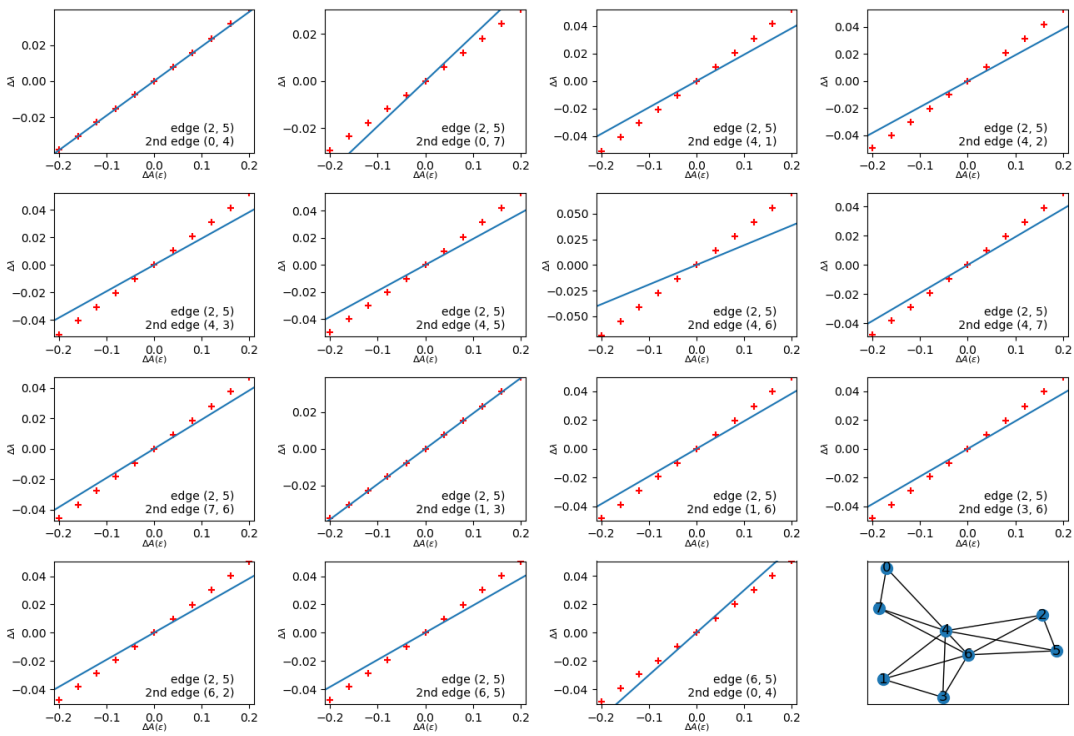


Figure 10: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Erdős–Rényi network with randomly assigned weights

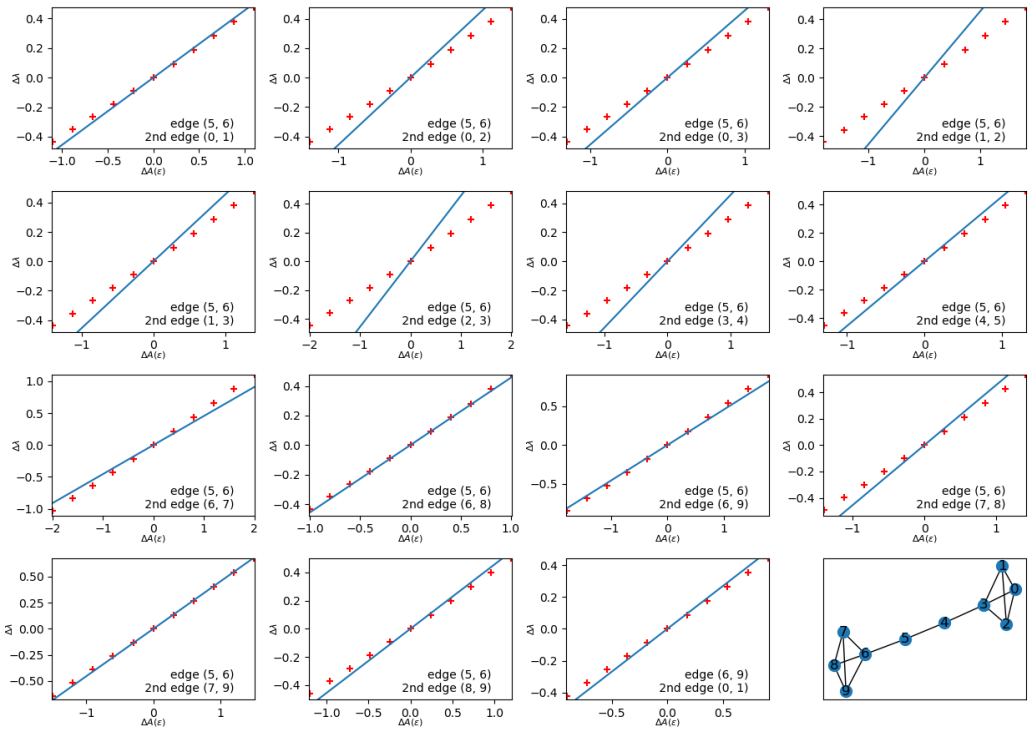


Figure 11: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Barbell network, with randomly assigned weights.

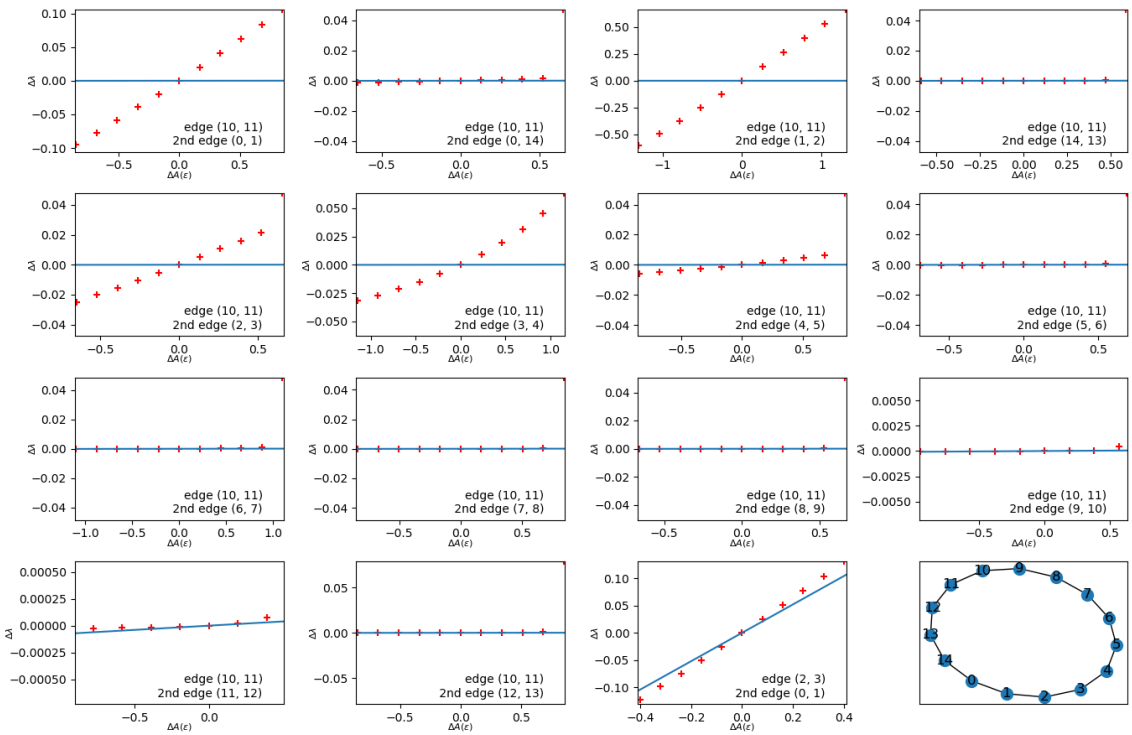


Figure 12: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Ring network with equal weights.

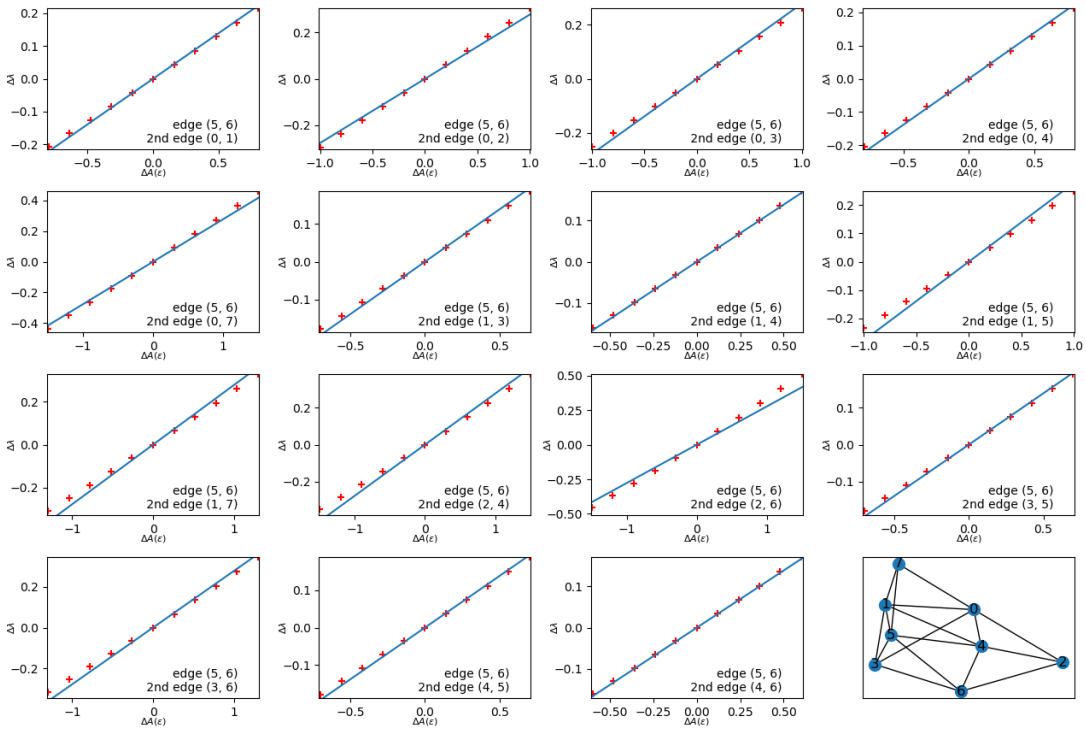


Figure 13: Scatter plot of perturbations ΔA_{ij} and the resulting $\Delta\lambda$, compared to line of constant l_e, for the case of two edges changing. The plots consider one of the two perturbed edges. Erdős–Rényi network with equal weights

Predictability improvement with α and ρ

As detailed in section 3.3.5, here I apply a logistic regression classifier with a single feature l_e to datasets with varying α and ρ. Figures 14 and 15 show the improvement in the test set PR AUC for increasing values of each parameter. These show that increasing both parameters improves the predictability of changes given the value of l_e, consistent with my observations of the rate of increase of change probability being positively correlated with both α and ρ.

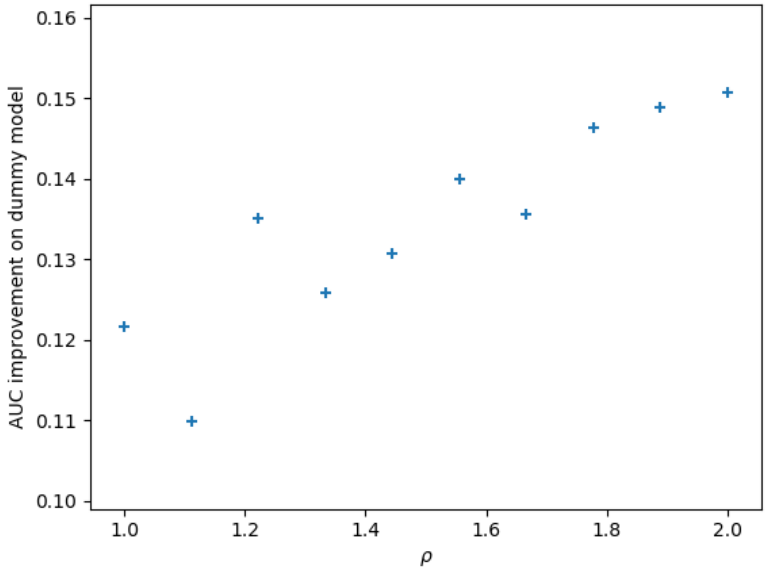


Figure 14: Model prediction PR AUC score improvement with ρ

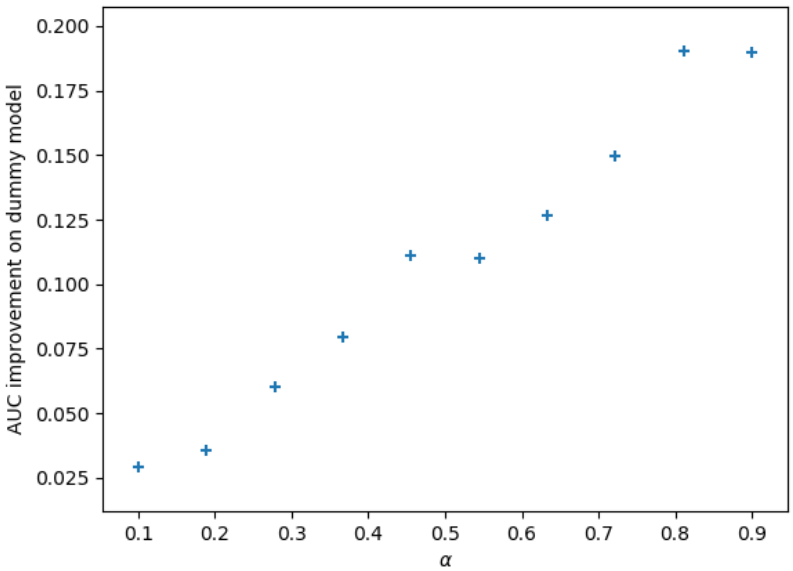


Figure 15: Model prediction PR AUC score with α

Prediction of edge changes

Figure 16 shows the PR curves, to complement the ROC curve for the prediction of edge changes presented in chapter 3.

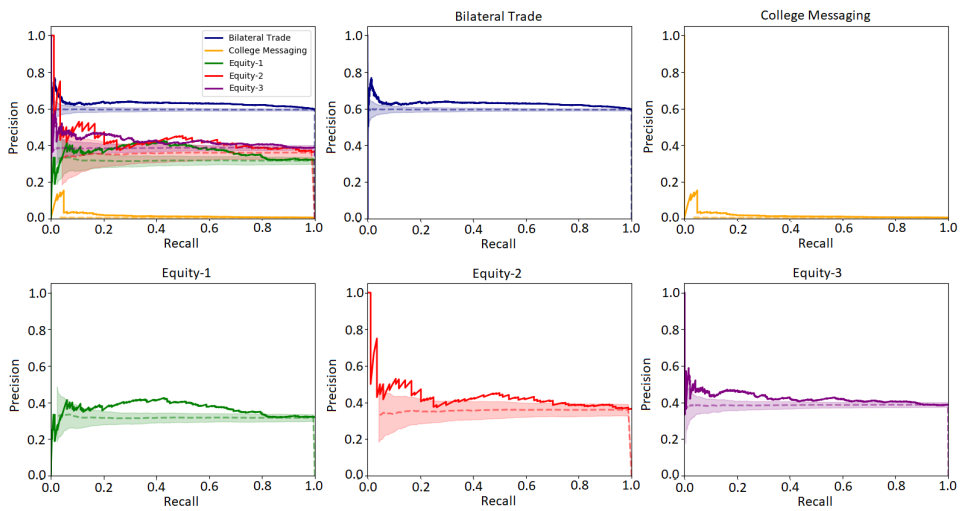


Figure 16: PR curves for a logistic regression classifier making use of $\ln(l_e)$ to predict $\Delta A_{ij} = 1$. The dashed lines represent the results for a stratified random allocation of labels.

Figures 17 and 18 show the improvement on the null model of the precision-recall and ROC AUC scores with increasing aggregation scale, to complement figure 3.17 in chapter 3.

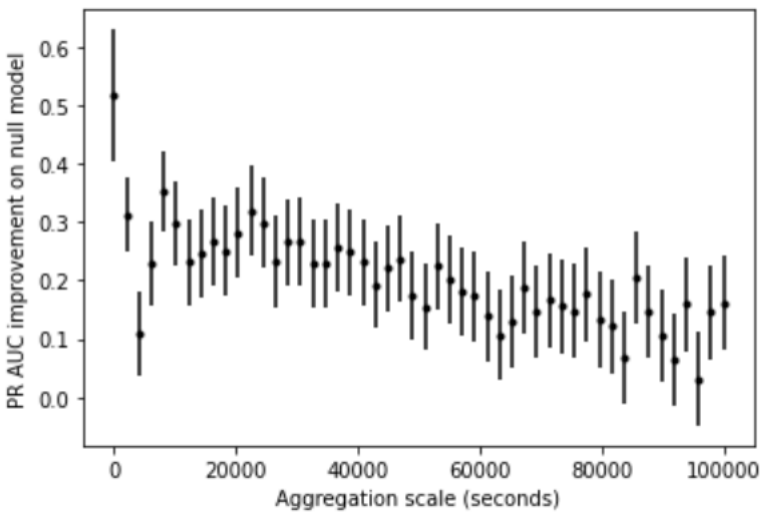


Figure 17: Improvement of PR AUC from null model for aggregation scales ranging from 50 seconds to 27 hours

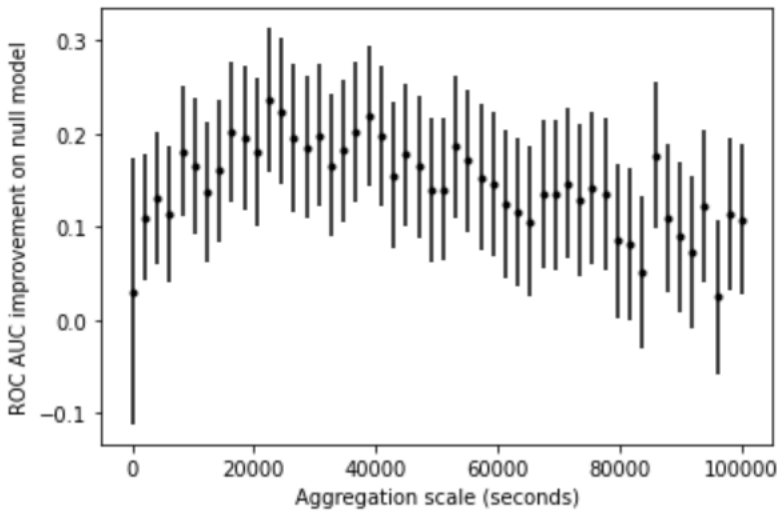


Figure 18: Improvement of ROC AUC from null model for aggregation scales ranging from 50 seconds to 27 hours

.1.2 Additional results for chapter 4

A coin-toss approach to estimating confidence intervals

In order to estimate the confidence intervals for precision and recall presented in this chapter 4, I can take a coin-toss approach to estimate the precision/recall thresholds such that the probability of observed outcome is higher than some threshold α. The first step of this approach is to note that precision is the probability of ground truth positive from all positive predictions, and recall is the probability of a ground truth positive from all correct predictions. I can then make use of the binomial distribution probability mass function to estimate the probability of the observed outcome depending on the chance of a 'positive flip' (the value of precision or recall):

$$Pr(k = TP; n, p) = \frac{n!}{TP!(n - TP)!} p^{TP} (1 - p)^{(n-TP)}, \tag{1}$$

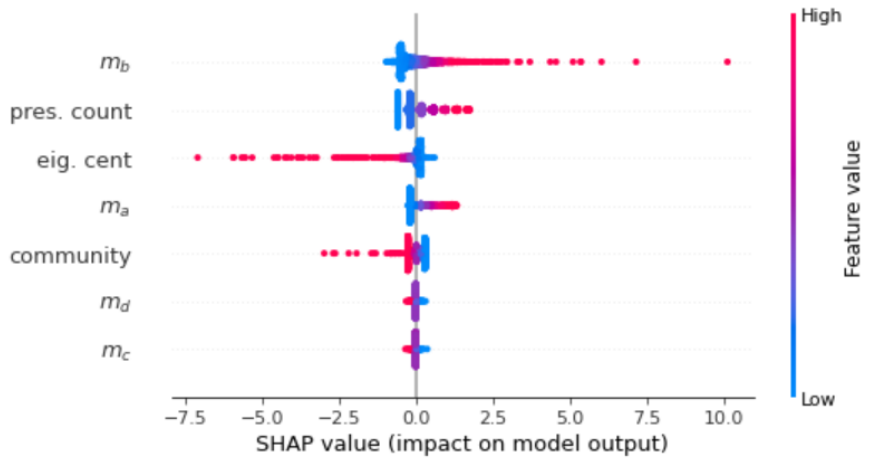
where k, the number of successes, is equal to the number of true positives (TP) in both cases, n is the number of true positives and false positives (TP+FP) for the case of precision and the number of true positives and false negatives (TP+FN) for the case of recall and p is the probability of success. Taking the cumulative of this and inverting for p allows me to find the confidence intervals for the precision and recall.

Prediction of node presence

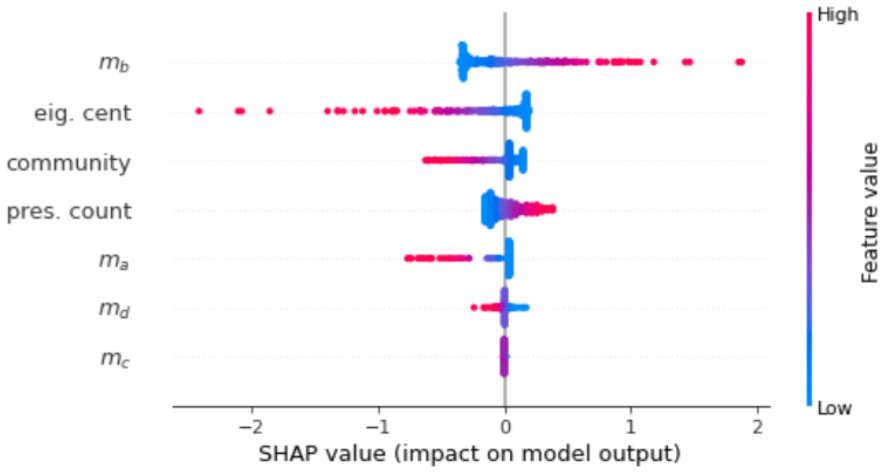
In figure 4.6 in chapter 4 I present evidence of the importance of my measure m_b in predicting subsequent node presence, using permutation importance to measure the importance of the model features. This is consistent with observations from figure 19, in which each dot represents a single SHAP explanation of the log-odds for a single observation by the feature the row of the plot corresponds to.

The features are ordered by the mean absolute value of the SHAP value for each feature. I observe that m_b is also the most important feature on average for all three datasets, and that nodes with a high m_b have a lower chance of of being subsequently present.

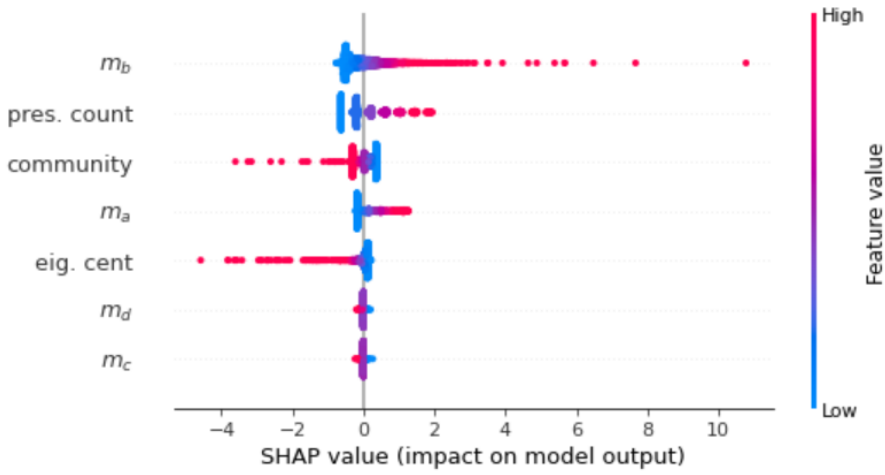
Figure 20 shows the equivalent SHAP plot for the BaTIS dataset.



(a) Equity-1



(b) Equity-2



(c) Equity-3

Figure 19: SHAP feature importance for the different node importance measures. Each dot represents a single SHAP explanation of the log-odds for a single observation by the feature the row of the plot corresponds to.

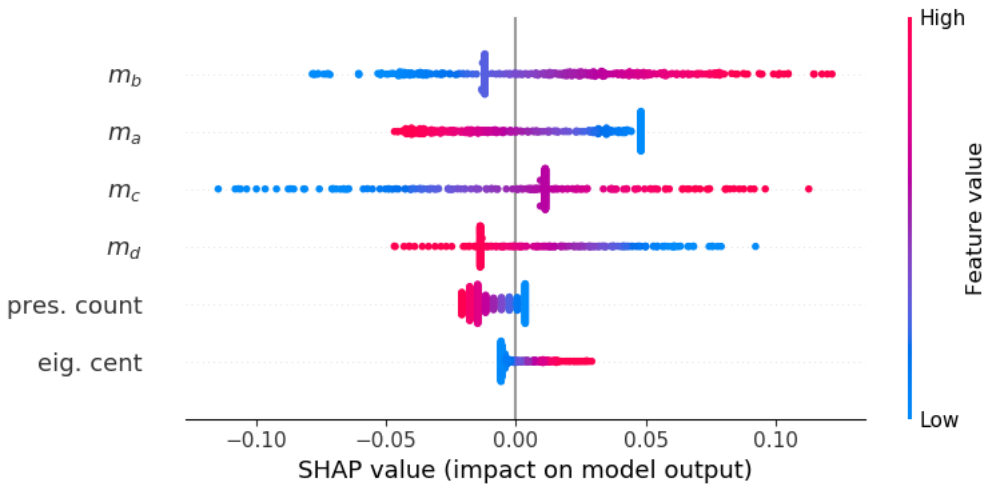


Figure 20: SHAP feature importance for the different node measures, for the BaTIS dataset. Each dot represents a single SHAP explanation of the log-odds for a single observation by the feature the row of the plot corresponds to.

.1.3 Additional results for chapter 5

Data exploration

In chapter 5, I make use of transaction reports containing the details of 5 randomly selected FTSE 100 constituent instruments, reported under MIFID II regulations. Although all of the instruments are FTSE 100 stocks which will be dominated by high frequency trading strategies, for the day considered they vary significantly in the number of transactions, with the smallest being 16,000 transactions (on average approximately one transaction every 2 seconds) and the largest 110,000 transactions (on average approximately one transaction every 0.2 seconds).

As the majority of trading in these liquid instruments is through market exchanges, one of the prevalent properties of these systems is the dominance of central institutions, usually central counterparty clearing houses (CCPs). A CCP is an institution that sits between parties involved in a transaction, to absorb any counterparty credit risk. For the instruments in question, all show $30-40\%$ of trades intermediated by the same central clearing house. For most major venues order book trades are cleared through a single CCP for each instrument. However, with the existence of interoperability agreements which give clearing members the opportunity for netting across CCPs covered by these agreements, it is not uncommon to see trades cleared through several CCPs for the same instrument[1]. Despite this, I observe a large proportion of trades through one venue, as can be seen in figure 21, and a large proportion of trades made by a single counterparty in figure 22. I thus consider application of Hawkes processes in

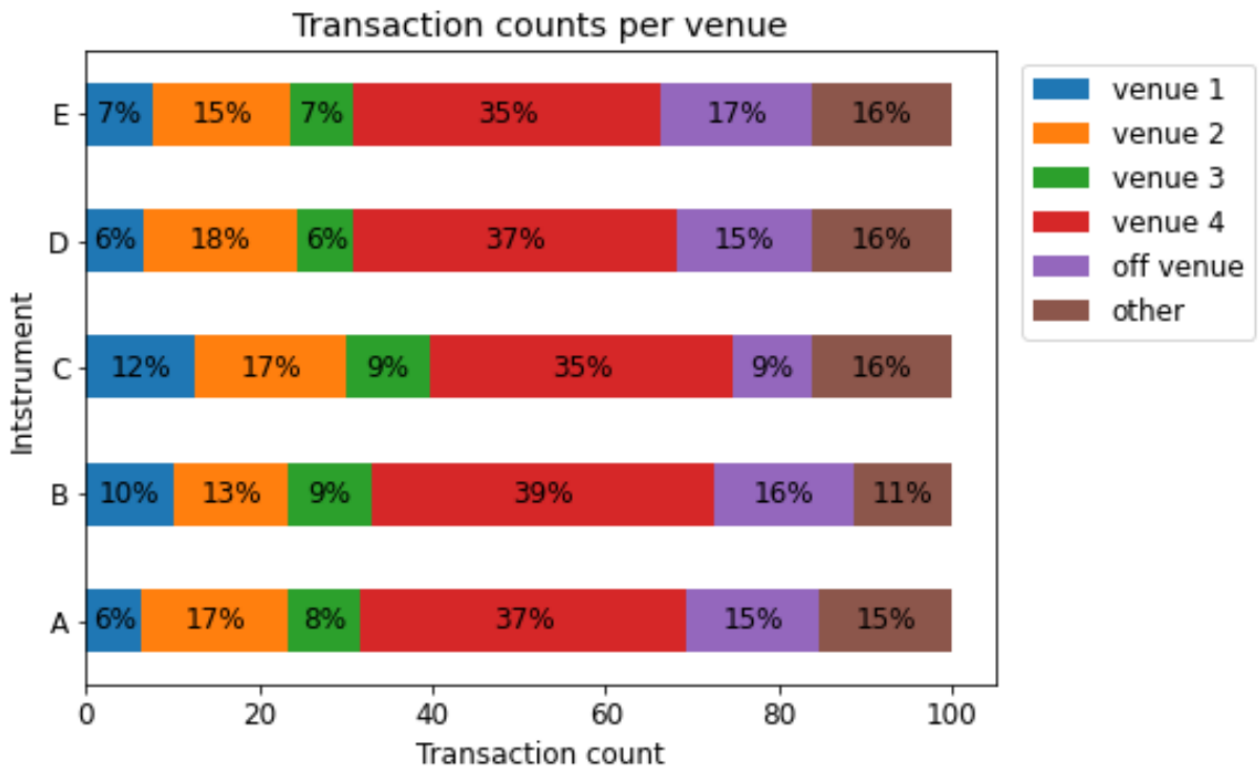


Figure 21: Transaction counts by venue, showing the top 4 venues and off venue trades separately and the remaining venues aggregated ('other').

various different ways, to different subsets of these systems: I consider application to all the visible

[1]This is not the situation for many other asset classes, for example ICE futures contracts will be cleared through ICE only.

trades for these instruments, trades through a single exchange, trades executed off exchange[2] and trades only with the largest participant, a prominent clearing house. It is theoretically possible to disintermediate the central clearing house, but most of the trading in these cases will be on central limit order books, the buyers and sellers either side of the clearing house will not deliberately trade with each other, but are matched according to the order matching system in place [224]. Also, by trading via the clearing house, the risk of the trade is absorbed by the clearing house. For these reasons, I decide not to disintermediate the central clearing house as it represents a true node in the transaction network.

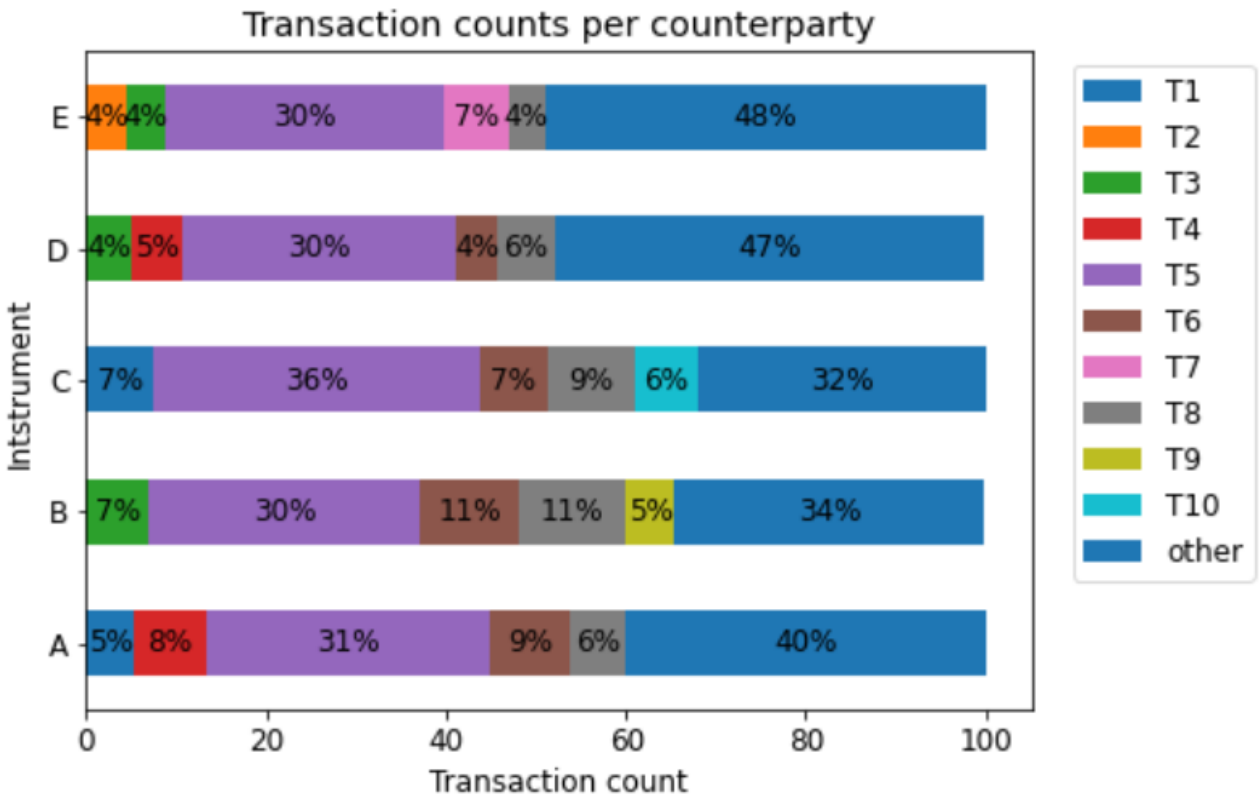


Figure 22: Transaction counts by counterparty, showing the top 5 counterparties for each instrument separately and the remaining counterparties aggregated ('other').

Of the trades operated through the major venues, I also observe different types of market - for example, the second and third largest venues for all instruments considered in chapter 5 are both examples of dark order book trading venues[3]. These are venues in which there is a lack of transparency of the order book, allowing large institutional investors to place trades while minimising the impact the markets [225]. On the other hand, the fourth largest is a lit order book venue, where the full book of orders is visible to all market participants. Different trading mechanisms would be expected for lit vs. dark order books, as the former allows the market price to respond more readily than the latter.

Additional observations of temporal properties

In chapter 5, I make observations of the burstiness of these datasets across time for a single instrument for illustration. Here I supplement these with additional observations for other instruments in figure 23,

[2]Off exchange trades of exchange listed securities occur for a number of reasons, for example they may occur between different arms of a large institution in different jurisdictions in order to move client money between these different arms

[3]The top four venues are the London Stock Exchange, the LSEG's Turquoise Lit venue, and the BATs Europe and Chi-X Europe venues operated by Cboe Europe.

alongside observations of transaction density and burstiness at edge level.

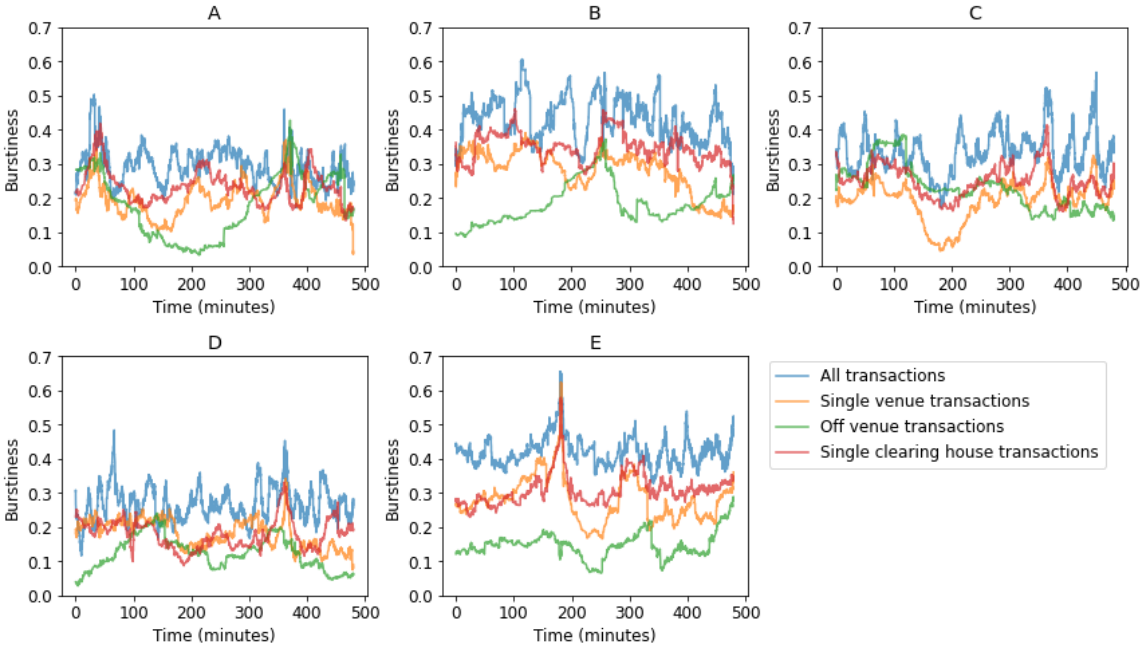


Figure 23: Network level burstiness across rolling windows of length 200 transactions, for the different data subsets.

Figure 24 shows the rolling transaction density (number of transactions in a given time window) for 10 minute time windows, with the window size chosen optimally to contain enough data points in each window for all datasets. I observe that unlike the burstiness, the transaction density is in general not stationary and in some cases shows a visible trend.

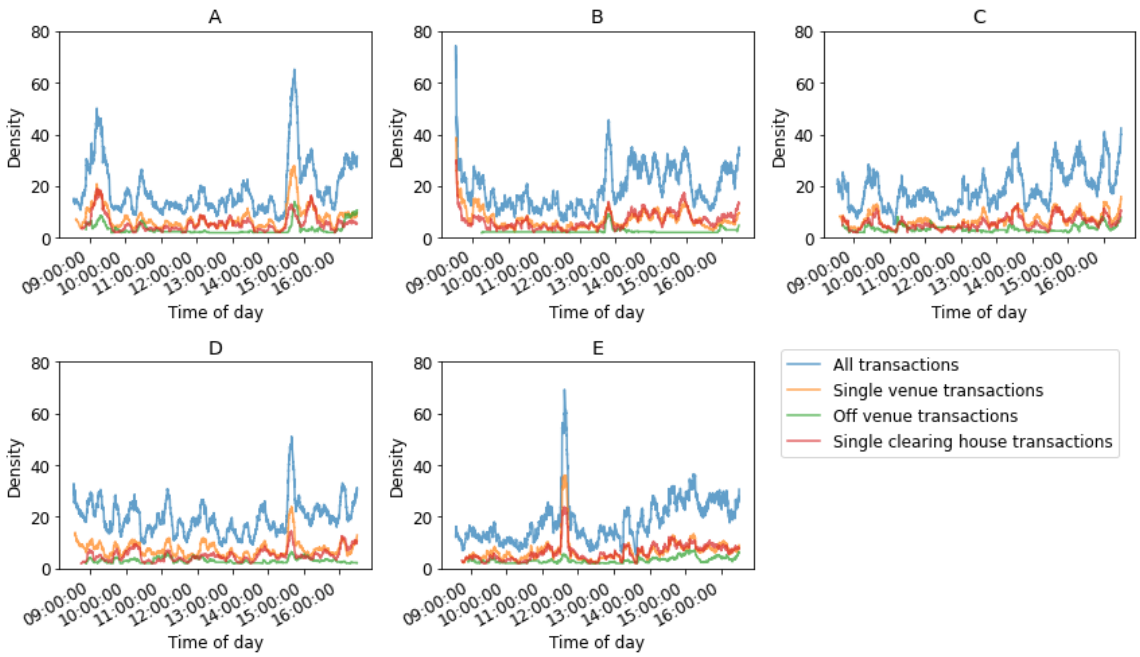


Figure 24: Network level density (number of transactions in a given time window) across rolling 10 minute windows

When considering individual edge burstiness, I observe that the individual edges tend to show a lower burstiness, with average edge burstinesses between 0.1 and 0.3 for the five datasets, in comparison to the network as a whole which shows average burstinesses between 0.3 and 0.5. This suggests that mutual-excitation is a contributing factor to the burstiness. In figure 25 I observe a significant variation in the edge level burstiness, with some edges showing periodic behaviour (negative burstiness). The subsets containing only trades with the most active central clearing house show the largest average burstiness across the edges.

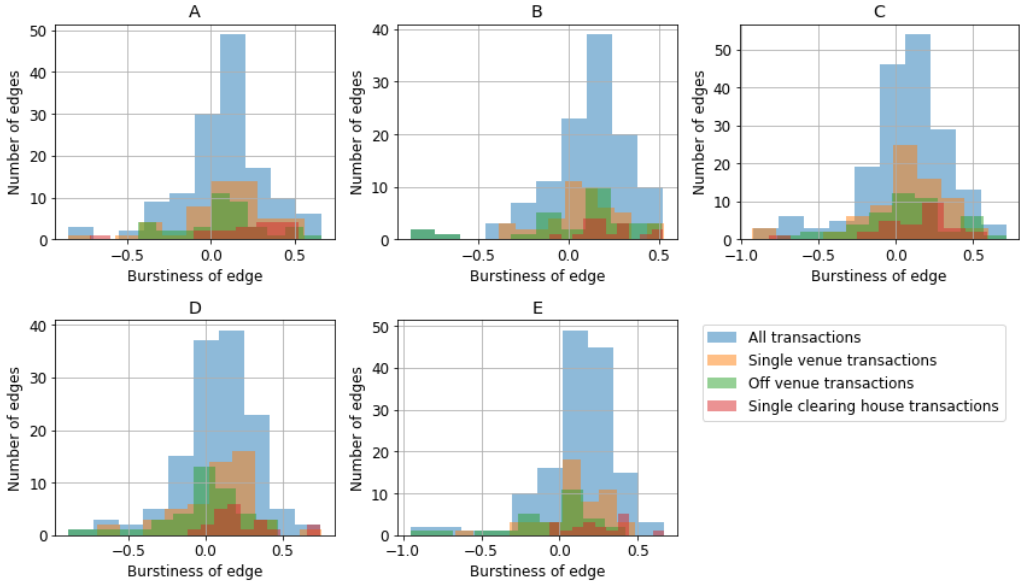


Figure 25: Histograms of the burstiness of individual edges across the entire observation period, for the 5 different datasets for their different subsets

It is worth noting that a key property of these systems is the presence of multiple transactions occurring between the same counterparties at the same time or very close together. This is likely to be as a result of multiple limit orders being placed at the same price, matching with a single larger order in the other direction. In order to understand the extent to which the observed burstiness in these

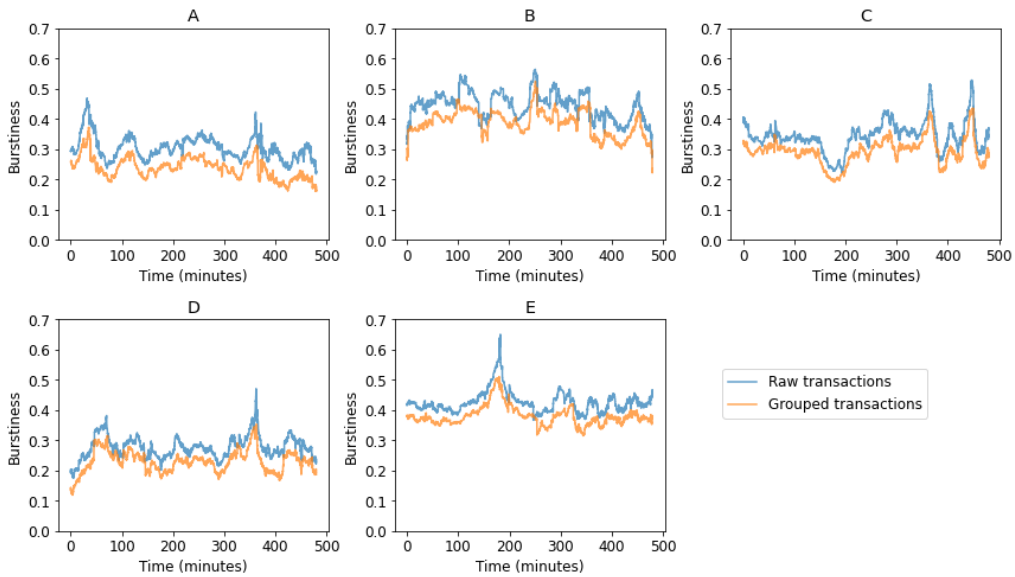


Figure 26: Rolling burstiness in windows containing 200 transactions, of all transactions for the instruments considered, both for the raw transaction timestamps and grouped timestamps.

systems is explained by these trades alone, I apply my methods both to the observed timestamps and to timestamps grouped at an edge level, considering transactions as part of the same larger order if the transactions in a group have inter-trade times of less than 1 second. Figure 26 shows the rolling burstiness produced as in figure 23 for all transactions, both for the observed transactions, and for transactions grouped by searching for bursts of transactions which have an inter-transaction time of less than 1 second. This shows that grouping the transactions produces a sequence with a slightly lower burstiness but does not entirely remove the burstiness of these sequences.

Additional observations of cross sectional properties

In chapter 5, I also make observations of cross-sectional, network-based properties. Figure 27 shows additional results for the cumulative degree distributions for additional datasets, demonstrating similar behaviour across all datasets.

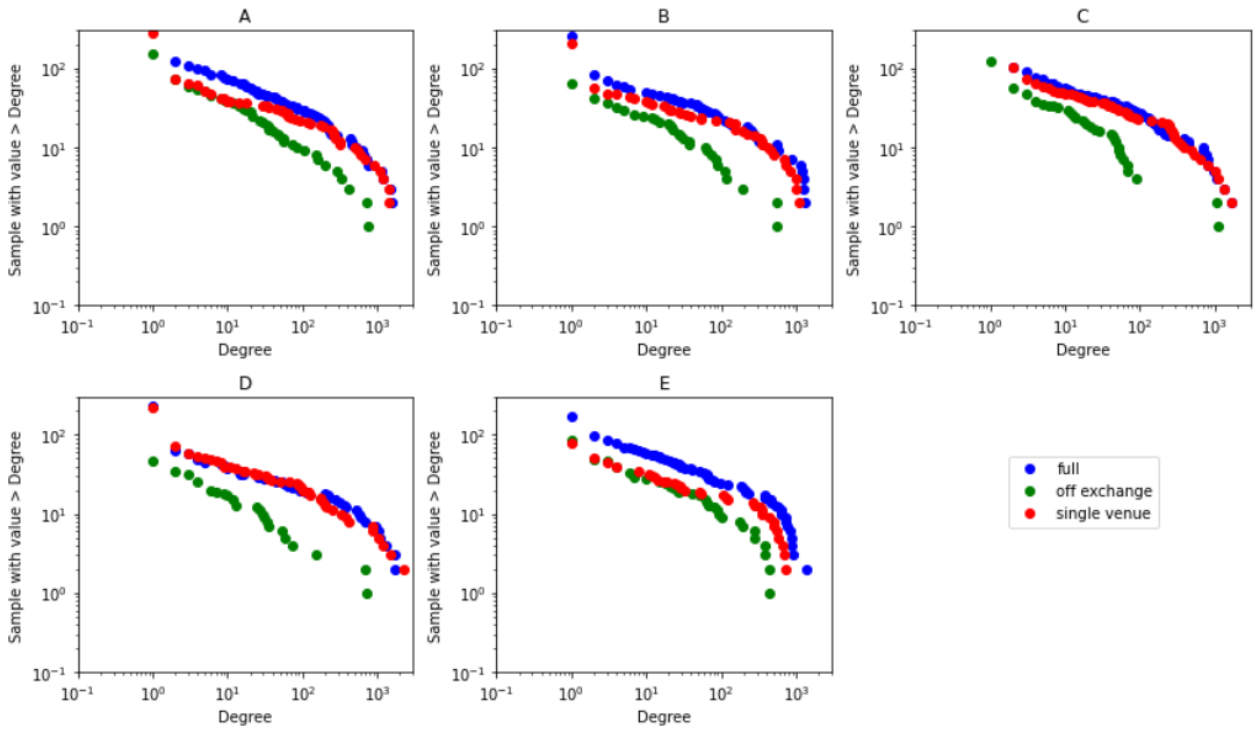


Figure 27: Cumulative degree distributions for the 5 different datasets for the different subsets considered, computed over the entire time period and plotted on a log-log scale.

Figures 28 and 29 show the assortativity and reciprocity across time for all 5 datasets. I observe similar levels of assortativity for all datasets, but a reasonable level of variation across the 5 different datasets and subsets for reciprocity, with some instruments showing larger values, indicating a higher prevalence of participants acting as both buyers and sellers. In particular, over half of the counterparty relationships for the single venue subsets of instruments D and E appear in both directions, whereas roughly a quarter do for instrument C.

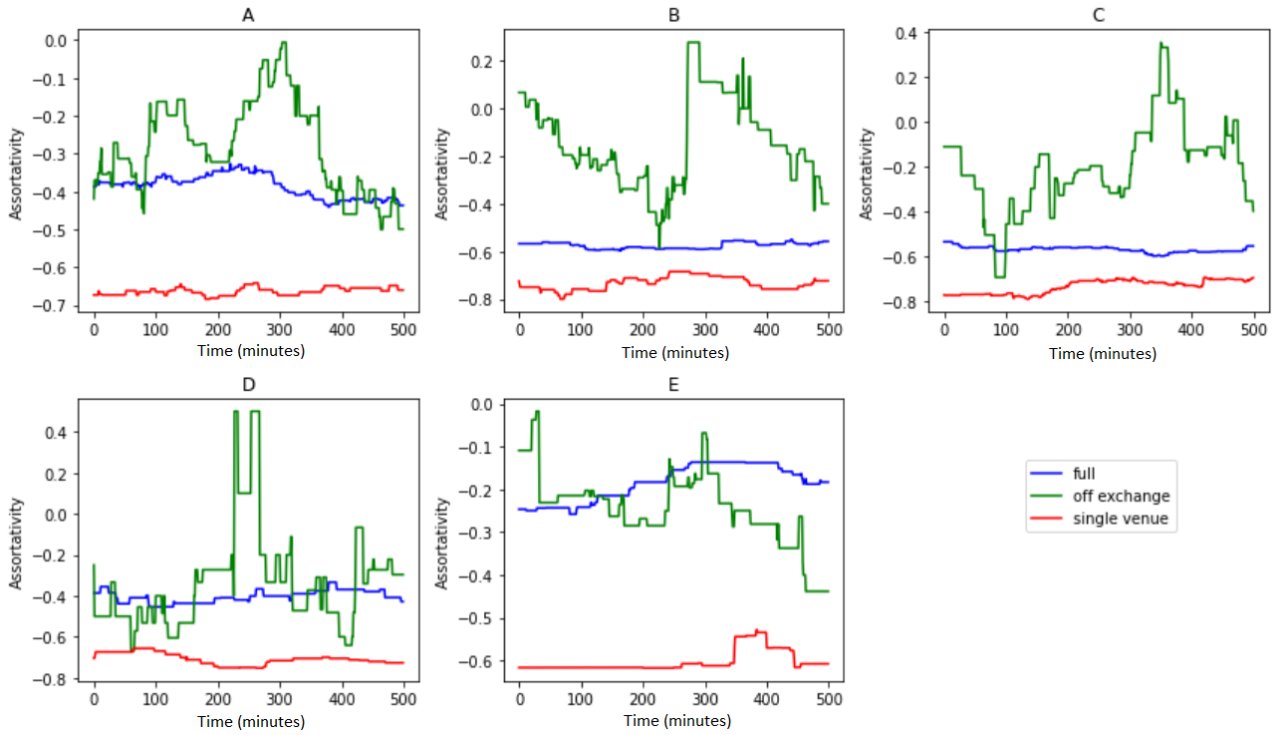


Figure 28: Assortativity computed across time in rolling windows of 200 transactions for the 5 different datasets, for the different subsets.

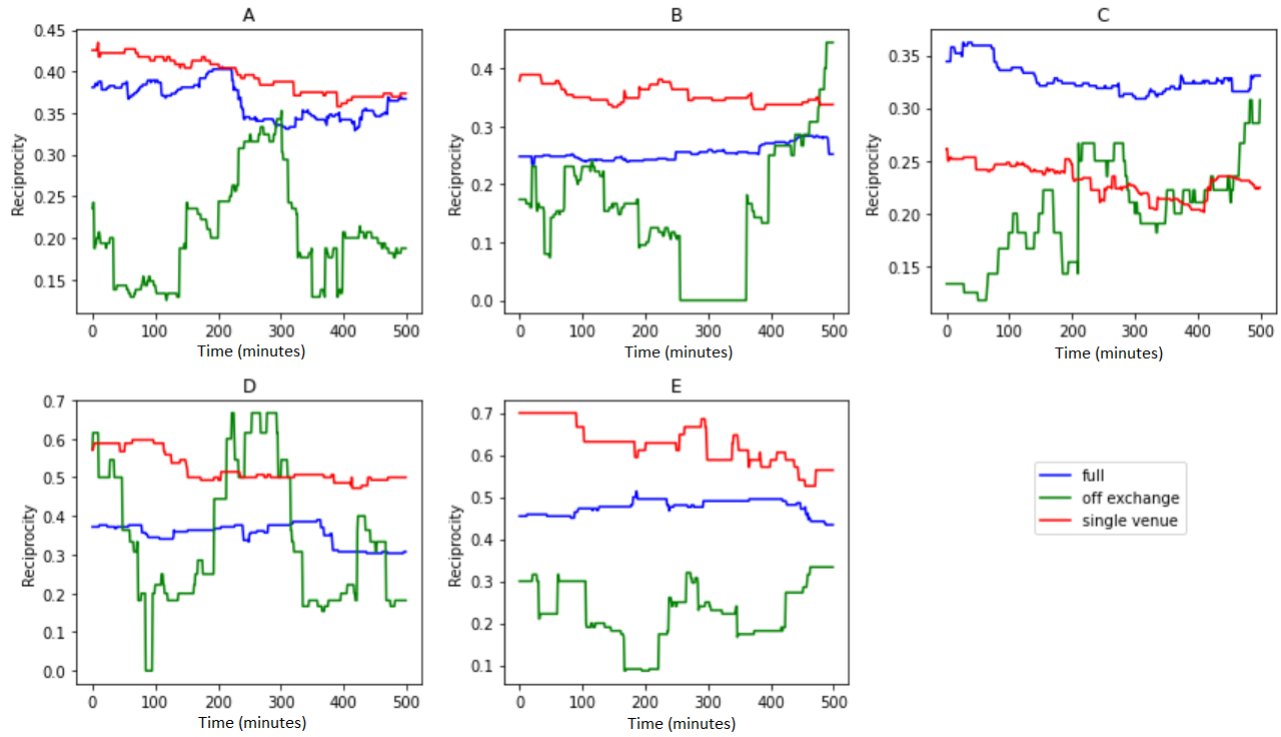


Figure 29: Reciprocity computed across time in rolling windows of 200 transactions for the 5 different datasets, for the different subsets.

Figures 30, 32 and 31 show the distributions of rich club indices for the data subsets for the full datasets, trades through a single variable and trades executed off exchange. For the full datasets, I observe similarity in the distributions across all 5 datasets, with the nodes with the largest degree showing low values of the rich club indices, driven by the presence of disconnected star subnetworks. Although the single variable and off exchange subsets are limited by the sparsity of these datasets, here I observe that trading through a single venue shows the similar behaviour of nodes with high degree showing very low rich club coefficients, which is not the case for the off exchange dataset.

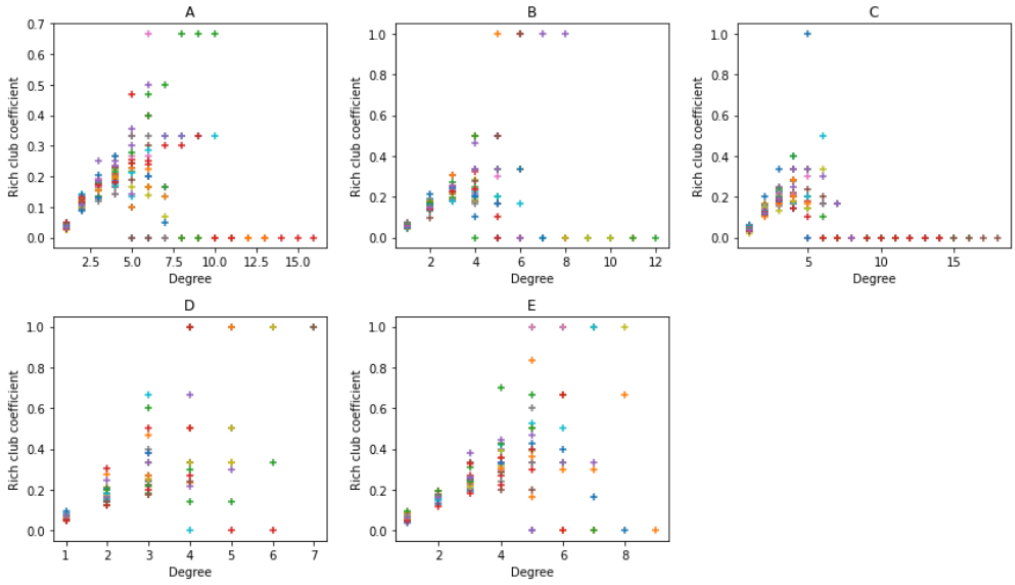


Figure 30: Rich club coefficients for the degrees of nodes for the 5 different datasets. The colours represent snapshot networks aggregated every 10 minutes.

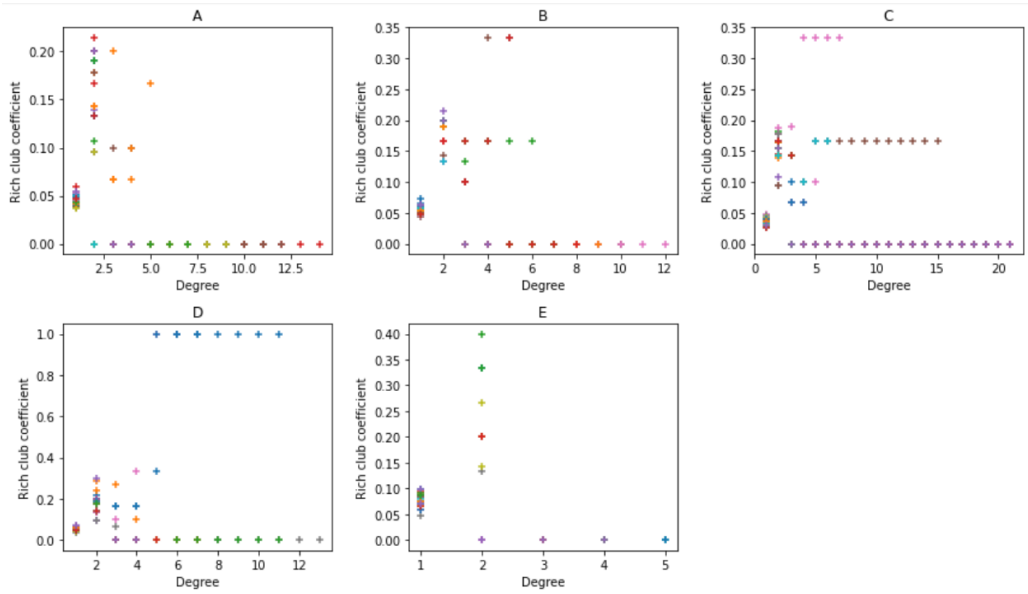


Figure 31: Rich club coefficients for the degrees of nodes for the 5 different datasets, for trades executed on a single venue. The colours represent snapshot networks aggregated every 10 mintues.

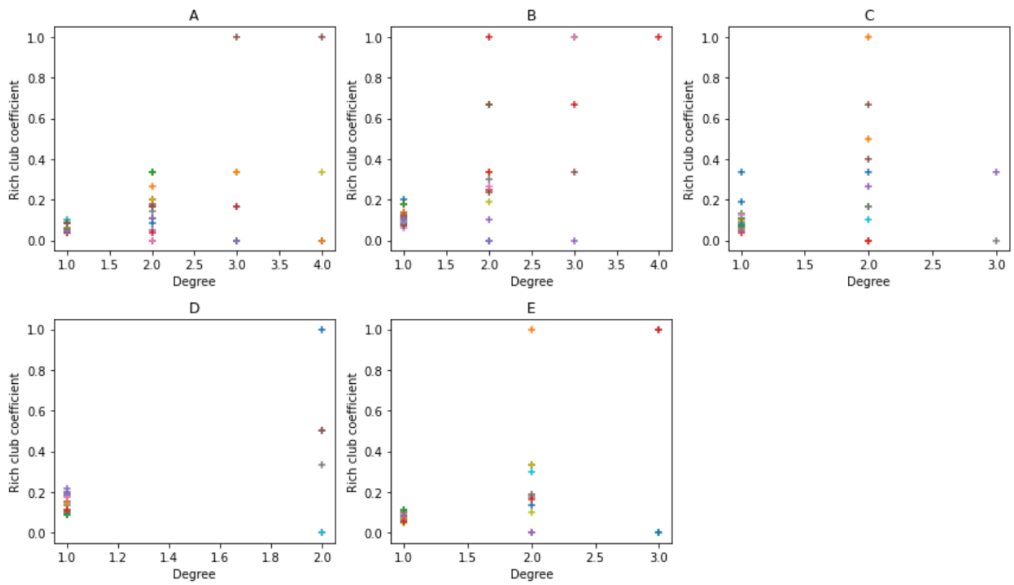


Figure 32: Rich club coefficients for the degrees of nodes for the 5 different datasets, for trades executed off exchange. The colours represent snapshot networks aggregated every 10 minutes.

Assortativity and degree distribution constraints

Care must be taken in interpreting the degree distribution and assortativity results in the chapter 5 for the case of frequency based edge selection, as the method of sampling the edges to change probabilistically based on their frequency essentially fixes the degree distribution for a large enough sample size when the degree distribution is temporally stable. The values that the assortativity can take when the degree distribution is fixed will be constrained, since the degree distribution constrains the configuration of the network, which places bounds on the values that the assortativity can take. To understand the extent to which this explains the observed assortativity, I conduct an experiment in which the edges of the network are randomly re-wired to destroy any correlations between neighbours while preserving the degree distribution. The results of the assortativity across a rolling window a tenth of the length of the transaction sequence for both the real network and the rewired are shown in figure 33. I can see from this that although some of the assortativity is preserved when rewiring,

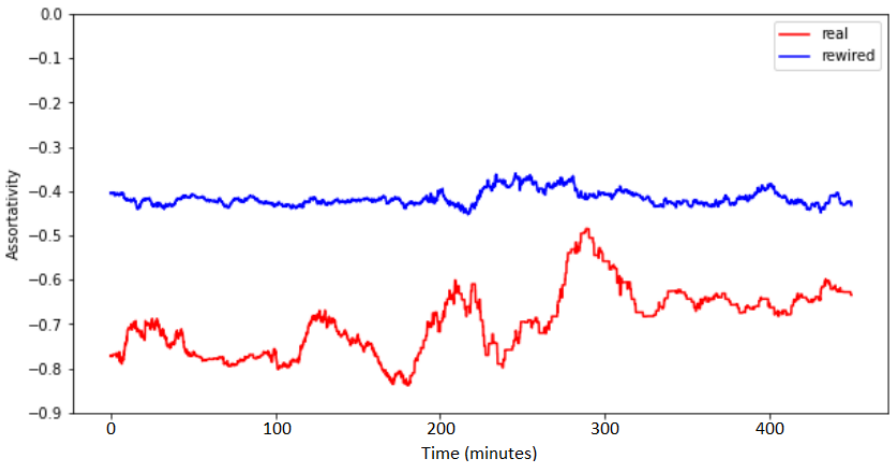


Figure 33: Assortativity computed across time in rolling windows of 200 transactions for dataset A, in comparison to the assortativity computed for dataset A with edge rewiring applied N times, with N the number of edges in the network

not all of the observed assortativity in the real data is explained by the constraints imposed by the selection method. In general, fixing the degree distribution means that I would expect the cross sectional properties to be reproduced, so this should be taken into consideration when interpreting the results of the frequency based selection of edges, as well as the multidimensional model making use of univariate Hawkes processes.

Finally, I note that in the case of the subset of trades executed through a single clearing house, I can alternatively consider the transaction sequences of buys and sells separately. Figure 34 shows the rolling burstinesses of the buys and sells separately for trades executed through the most active clearing house. I can see from these that the burstinesses of the buys and sells are at a similar level, although there are periods for certain instruments where the two transaction sequences differ in their burstinesses. Visually, there also appears to be some influence of higher periods of burstiness of buys influencing a higher burstiness of sells.

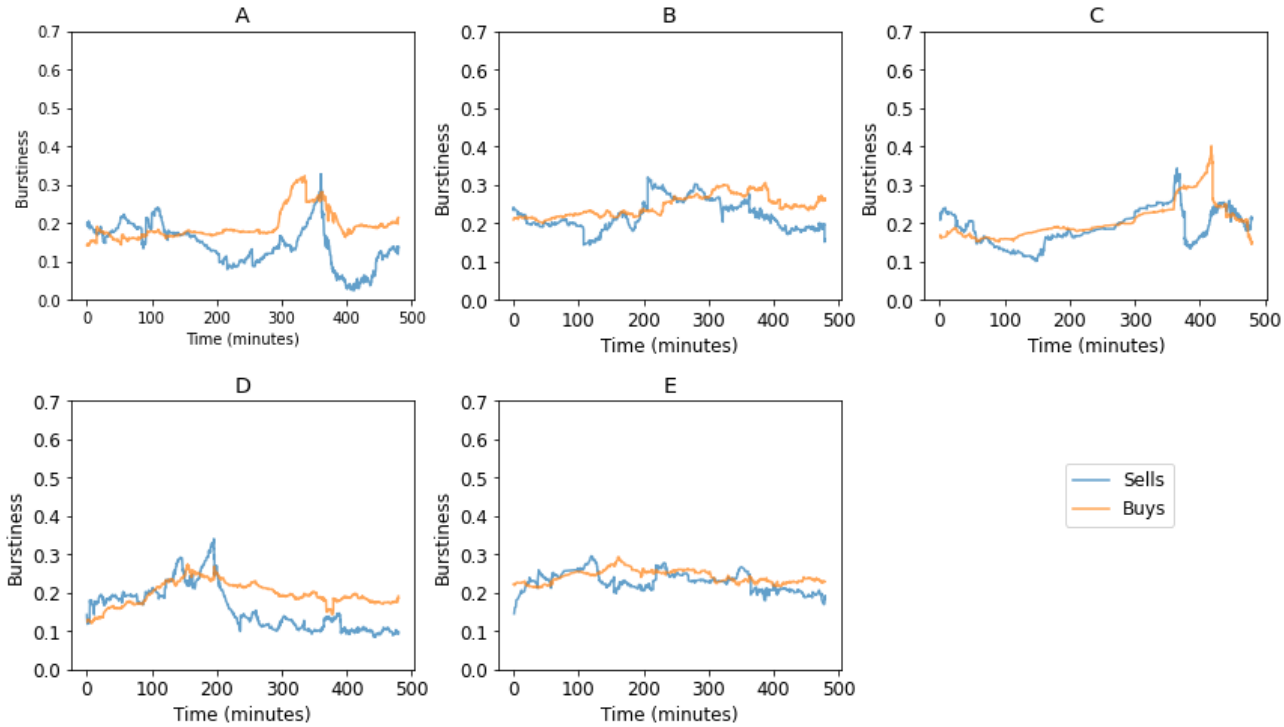


Figure 34: Rolling burstiness over a window of 200 transactions for the 5 different instruments, considering transaction sequences of buys and sells separately for transactions executed through the main clearing house.

Data	Baseline int.	Kernel int.	Kernel dec.
A (full)	0.45± 0.08**	0.55 ± 0.01***	0.62 ± 0.06**
A (single venue)	0.99 ± 0.16	0.48 ± 0.01***	0.92± NA
A (off exchange)	0.23 ± 0.19	0.56 ± 0.02***	0.66 ± 0.40
B (full)	0.09 ± 0.02**	0.35 ± 0.01***	$0.63 \pm NA$
B (single venue)	$0.41 \pm NA$	0.72 ± 0.01***	0.77 ± 0.10*
B (off exchange)	2.84 ± 0.58	0.75 ± 0.03***	0.80 ± 0.45
C (full)	0.12 ± 0.04*	0.36 ± 0.01**	$0.38 \pm NA$
C (single venue)	0.42 ± 0.10**	1.05 ± 0.02***	$1.33 \pm NA$
C (off exchange)	0.03 ± 0.03	1.29 ± 0.05***	1.30 ± 0.23*
D (full)	0.90 ± 0.04**	0.36 ± 0.01***	$1.59 \pm NA$
D (single venue)	0.01 ± 0.01	0.27 ± 0.01**	$0.63 \pm NA$
D (off exchange)	0.01 ± 0.01	$0.93 \pm NA$	$0.94 \pm NA$
E (full)	0.23 ± 0.06**	0.56 ± 0.01**	$0.66 \pm NA$
E (single venue)	0.10 ± 0.10	0.52 ± 0.10	0.57 ± 0.11
E (off exchange)	0.04 ± 0.04	1.31 ± 0.04***	1.54 ± 0.20*

Table 2: Parameters for the univariate Hawkes model estimated using a Maximum Likelihood approach for the chosen datasets, across the different subsystems considered as dynamic networks. P-values indicated as follows: ***<0.001, **<0.01, *<0.05. P-values and standard error calculated by taking the inverse of the Hessian, so are not calculated when the Hessian is non-invertible.

Additional model results - univariate Hawkes processes

The parameters estimated using a maximum likelihood approach to fit the three parameters of a univariate process to the full transaction sequence are shown in table 2. I observe a reasonable amount of variation across the parameter values for the different datasets for all three subsets considered, as demonstrated in figure 35. For the baseline intensity and kernel decay parameters, there is no consistent difference in the parameter value for the different data subsets across all of the datasets considered. However, the Kernel intensity appears to show the largest values for the off venue subset and smaller values for the full dataset. Revisiting my earlier observation that the burstiness is highest for the full dataset and lowest for off venue transactions, this appears contradictory, suggesting that the model is not able to accurately estimate the parameters in the univariate case.

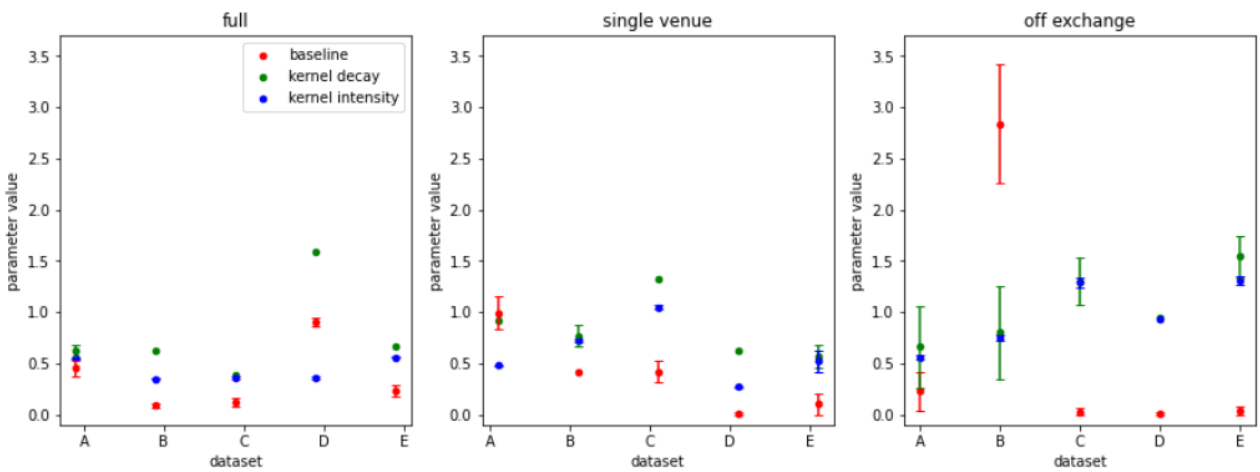


Figure 35: Parameter values for the 5 different datasets, across the three subsets considered. Error bars (where present) represent the standard error calculated by taking the inverse of the Hessian. The standard error is not calculated when the Hessian is non-invertible.

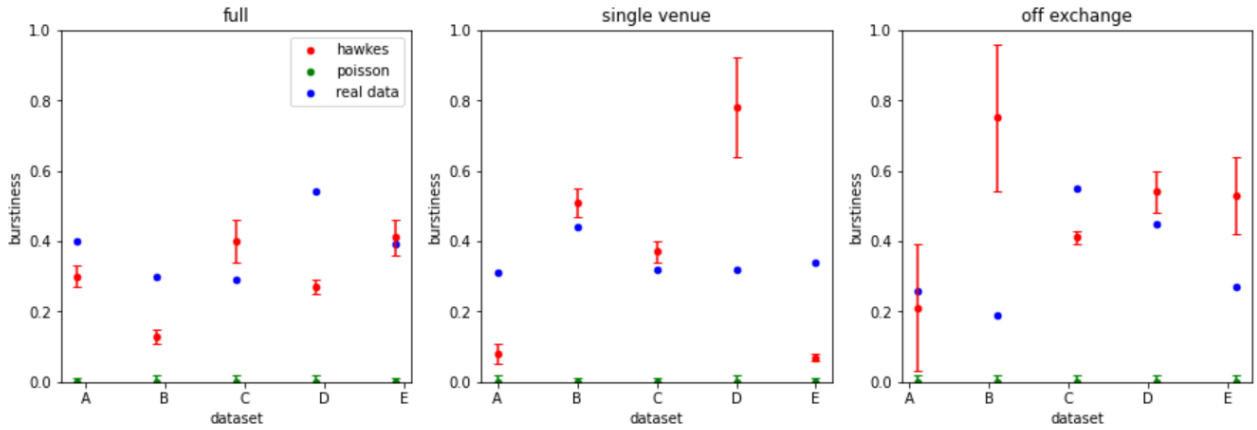


Figure 36: Burstiness of the simulated timestamps using the parameters in table 35, for the 5 different datasets across the different subsets, in comparison to the burstiness of the real data and a univariate Poisson process fitted to the real data.

Figure 36 visualises the burstiness of sequences generated by Hawkes processes using the parameters in table 2, that I also present in table 5.3.

In chapter 5, when using a univariate Hawkes process, I consider two different methods to select the edge to change at each time step. To evaluate the performance of these models, I compare their ability to reproduce several properties of the real data. Here I present additional observations of these properties. The first property I consider is the rich club coefficients at different node degrees. Figure 37 shows the rich club coefficients at different node degrees for snapshots of networks every 10 minutes, for all transactions for instrument A for illustration. I observe that for the real data, the nodes with the

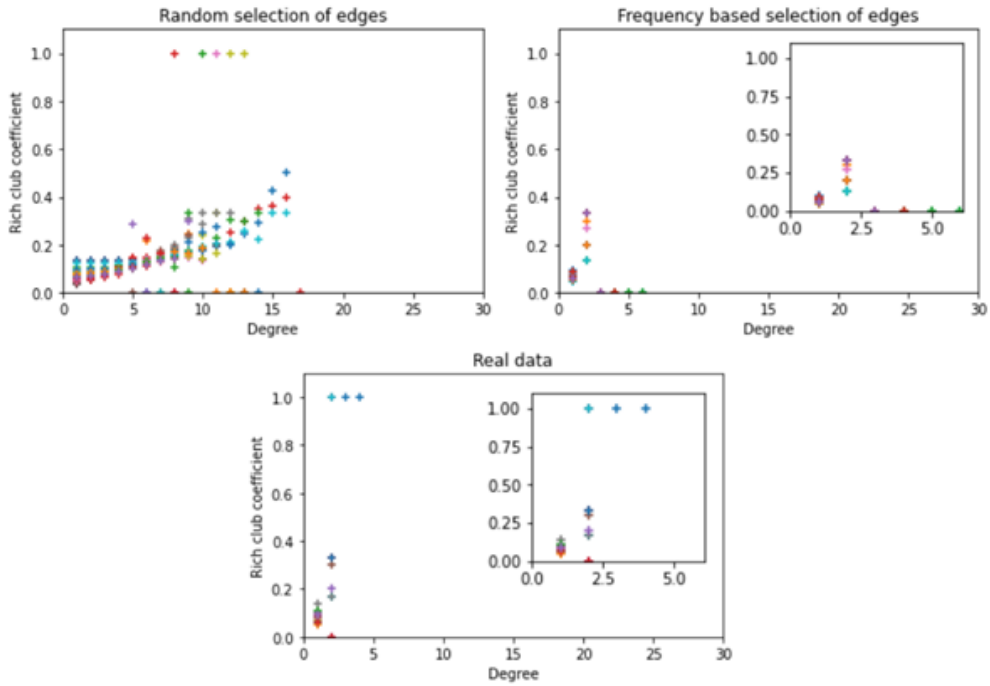


Figure 37: Rich club coefficients for the degrees of nodes in the network for the FTSE-A instrument. The colours represent snapshot networks aggregated every 10 minutes. Inset plots are presented for the frequency and real data with a reduced x axis scale.

largest degree have very low rich club indices, which is driven by the presence of disconnected star subnetworks containing a single hub node (in all cases a CCP) connected to a number of other nodes

that are only connected to the hub. This property is also reproduced by frequency based edge selection but not for random selection.

The second property I consider is the aggregated degree distributions. Figure 38 shows an example for instrument A of the degree distributions for the different edge selection methods along with the real data. Here I observe that the real data and frequency based edge selection data show similar

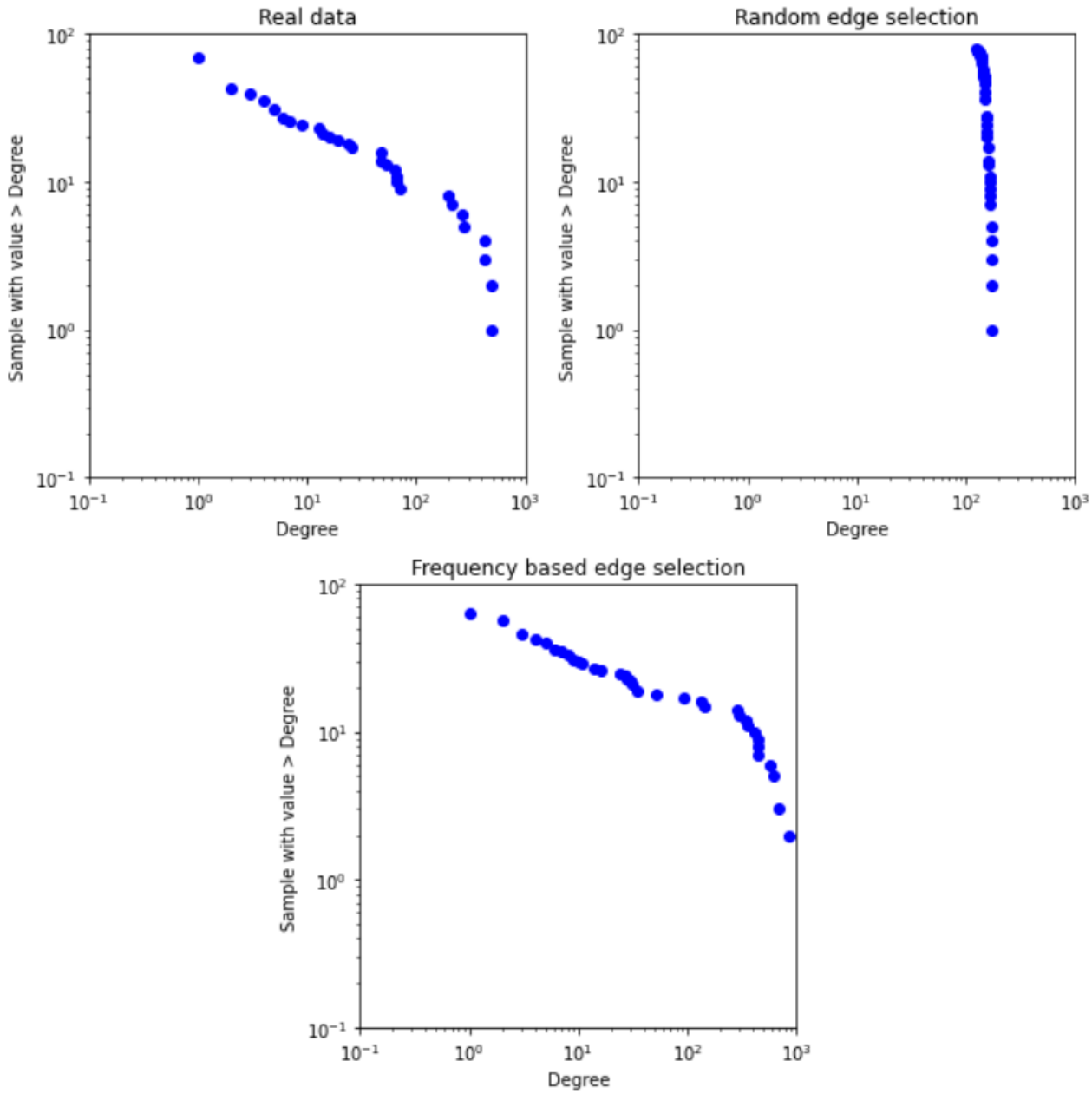


Figure 38: Cumulative degree distributions for the full networks computed over the entire time period, plotted on a log-log scale

shaped degree distributions, with a small number of nodes with very high degree, but the majority of nodes with very low degree. The random based edge selection methods instead produce networks which have a large number of nodes with a reasonably high degree. Table 3 shows the p-values of a two dimensional, two sided Kolmogorov-Smirnov test applied to each of the degree distributions of the simulated data, in comparison to the real data. I can see that the null hypothesis of the samples having the same distribution is not rejected for all of the datasets for the frequency based edge selection, but

Data	Random	Frequency
A (full)	2.78×10^{-48}	0.10
A (single venue)	7.12×10^{-61}	0.15
A (off exchange)	3.73×10^{-27}	0.01
B (full)	4.24×10^{-71}	0.99
B (single venue)	1.79×10^{-63}	0.15
B (off exchange)	4.53×10^{-27}	4.03×10^{-9}
C (full)	5.43×10^{-40}	0.06
C (single venue)	1.86×10^{-34}	0.12
C (off exchange)	3.40×10^{-18}	1.75×10^{-5}
D (full)	2.20×10^{-65}	0.99
D (single venue)	4.79×10^{-72}	0.45
D (off exchange)	6.25×10^{-34}	0.02
E (full)	2.41×10^{-91}	0.99
E (single venue)	4.71×10^{-98}	0.24
E (off exchange)	3.73×10^{-30}	0.99

Table 3: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the degree distributions for the two different edge selection methods in comparison to the real data, for the 5 datasets considered.

observe very small p-values for the other random method of selection confirming that this method is not able to reproduce the degree distribution of the real data.

Finally, I present the results of reciprocity and assortativity observed across the entire time period and in rolling windows across time. Table 4 and figure 39 shows the reciprocity computed across the entire

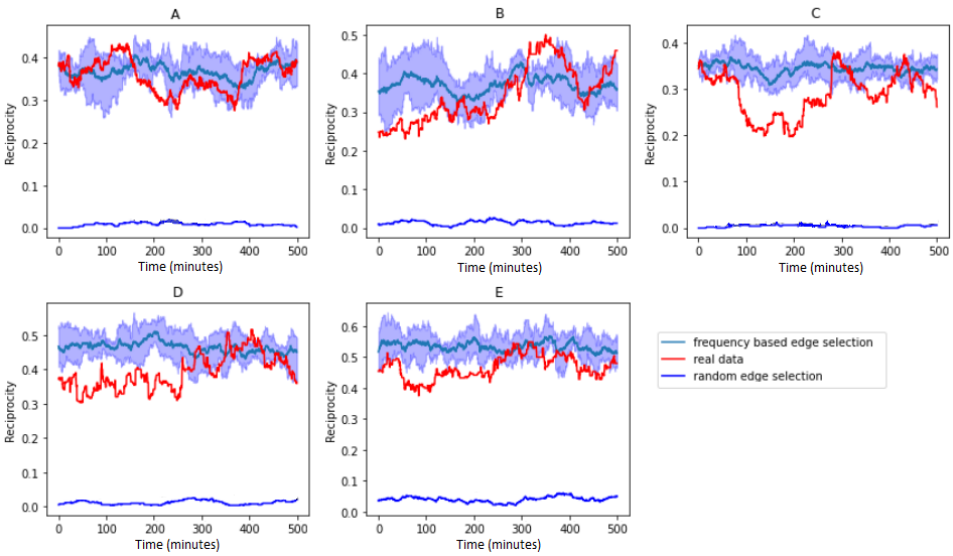


Figure 39: Reciprocity computed in rolling windows of 200 transactions for the 5 different datasets, presented for the two methods of edge selection along with the real data.

time period and over a rolling window for the 5 different datasets for the 3 different methods of edge selection, demonstrating consistency with figure 5.4 in chapter 5.

Table 5 and figure 40 shows the assortativity computed across the entire period and over the same rolling window. I can clearly see that the frequency based edge selection method produces an assortativity that is similar to that of the real data, whereas the random selection is unable to reproduce the

Data	Real	Random	Frequency
A (full)	0.53	0.07	0.52
A (single venue)	0.52	0.10	0.36
A (off exchange)	0.29	0.08	0.28
B (full)	0.45	0.16	0.47
B (single venue)	0.56	0.16	0.52
B (off exchange)	0.28	0.21	0.33
C (full)	0.46	0.04	0.51
C (single venue)	0.46	0.03	0.46
C (off exchange)	0.50	0.10	0.34
D (full)	0.38	0.12	0.41
D (single venue)	0.56	0.17	0.57
D (off exchange)	0.68	0.23	0.50
E (full)	0.64	0.28	0.64
E (single venue)	0.66	0.47	0.69
E (off exchange)	0.42	0.25	0.40

Table 4: Reciprocity across all transactions for the 5 different datasets, presented for the two methods of edge selection along with the real data.

Data	Real	Random	Frequency
A (full)	-0.24	0.01	-0.15
A (single venue)	-0.61	-0.02	-0.61
A (off exchange)	0.79	0	0.75
B (full)	-0.44	-0.03	-0.42
B (single venue)	-0.63	-0.01	-0.59
B (off exchange)	0.98	-0.03	0.98
C (full)	-0.41	-0.01	-0.42
C (single venue)	-0.46	0	-0.44
C (off exchange)	0.89	-0.01	0.87
D (full)	-0.39	0	-0.41
D (single venue)	-0.62	0	-0.62
D (off exchange)	0.84	0	0.86
E (full)	-0.37	0	-0.39
E (single venue)	-0.37	0	-0.39
E (off exchange)	0.93	0	0.91

Table 5: Assortativity across all transactions for the 5 different datasets, presented for the two methods of edge selection along with the real data.

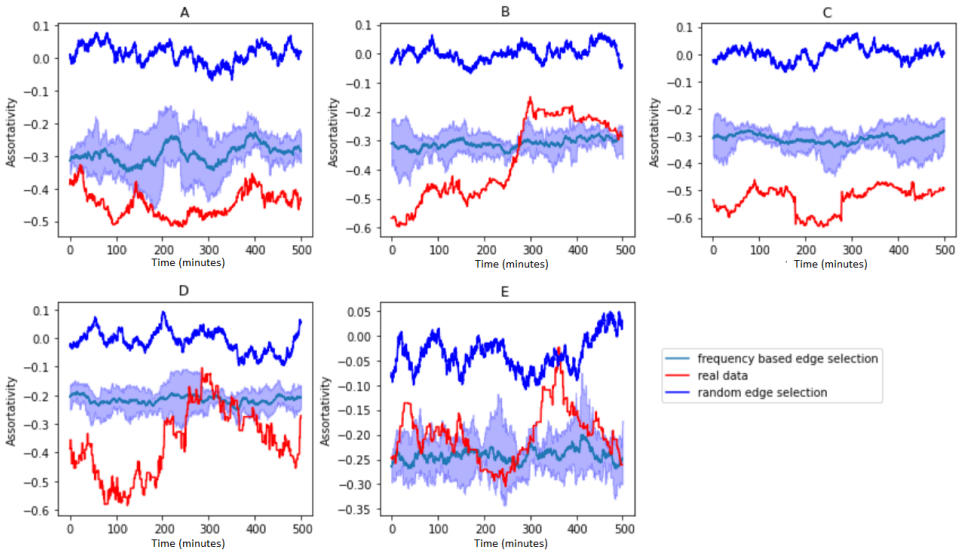


Figure 40: Assortativity computed across time in rolling windows of 200 transactions for the 5 different datasets, presented for the two methods of edge selection along with the real data. In the case of the frequency based method, the shaded blue area corresponds to 95% confidence intervals.

assortativity of the real data. The blue bands represent 95% confidence intervals for the assortativity for the frequency based method and although the real data assortativity falls within the confidence intervals some of the time, this is generally less than 95% of the time so I cannot conclude that the frequency based edge selection method is producing sequences with the same assortativity as the real data.

Additional model results - multivariate Hawkes processes

In addition to the results in chapter 5, I now present further observations for the multivariate Hawkes process. Figure 41 visualises the results presented in table 5.5. Here I observe that the univariate process better reproduces the real burstiness than the multivariate, however often has larger confidence intervals suggesting a lower stability of this model.

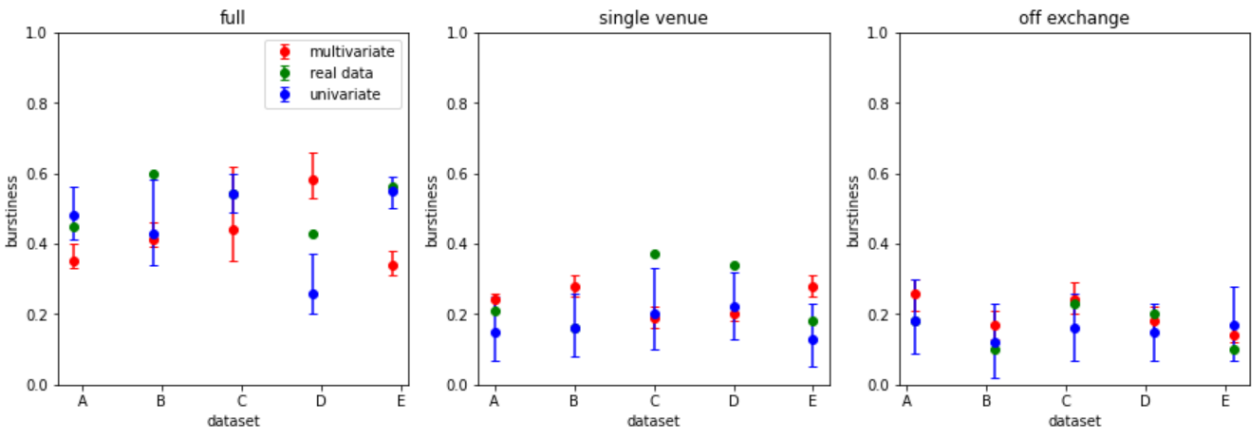


Figure 41: Burstiness of the simulated timestamps for both univariate and multivariate edge level Hawkes, for the 5 different datasets across the different subsets, in comparison to the burstiness of the real data.

For the cross sectional properties, starting with the rich club coefficient distributions and degree distributions, figures 42 and 43 demonstrates the similarities in both simulation methods to the real data that are confirmed in all cases using the same Kolmogorov-Smirnov test as before, for which the null hypothesis of differences between the distributions is rejected for both models, for all datasets.

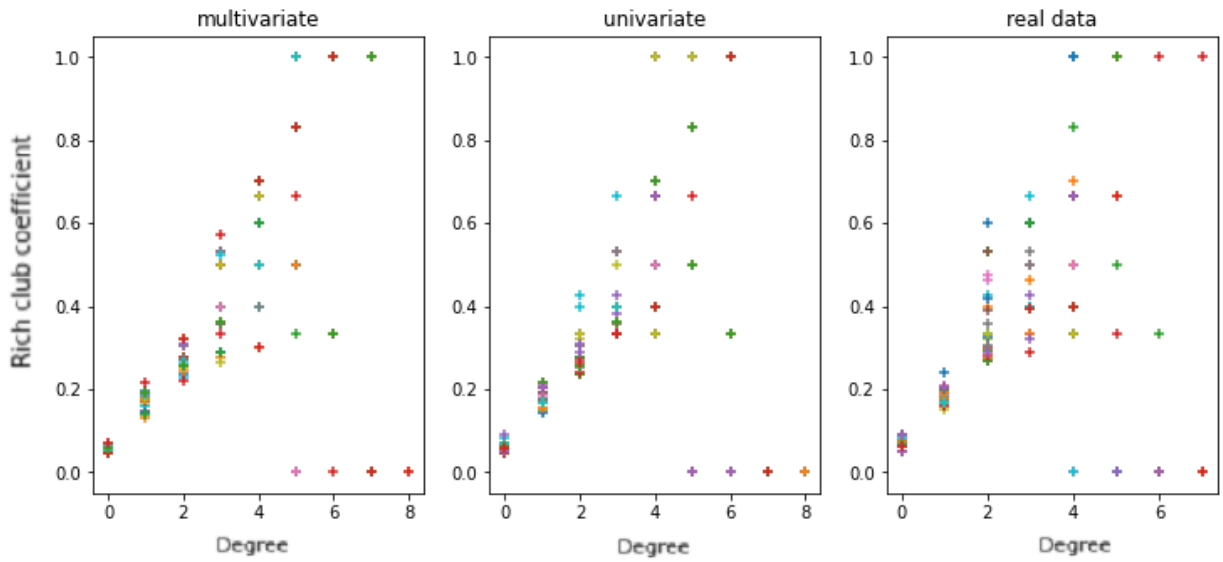


Figure 42: Distributions of rich club coefficients for the degrees of nodes in networks from snapshots aggregated hourly.

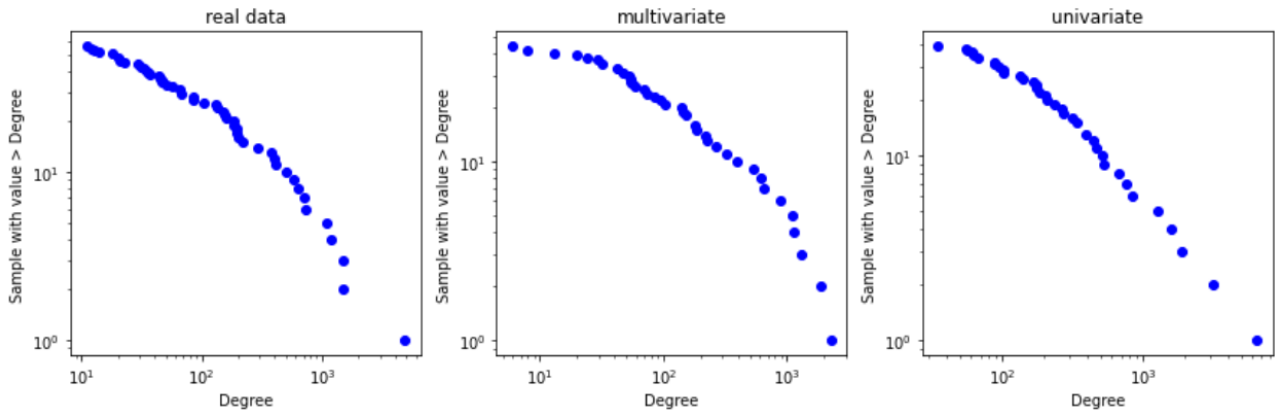


Figure 43: Degree distribution of networks from snapshots aggregated hourly.

Figures 44 and 45, and tables 6 and 7 show the assortativity and reciprocity computed over the entire time period and rolling windows for both methods of applying Hawkes processes to generate edge level transaction sequences.

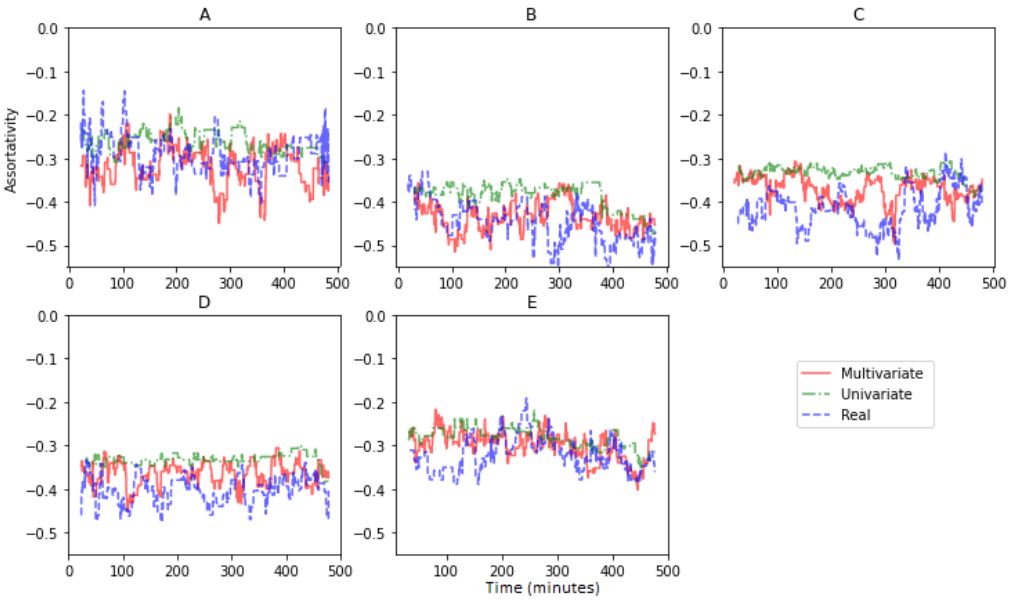


Figure 44: Assortativity computed across time in rolling windows of 200 transactions for the 5 different datasets, presented for the two different methods of generating edge level transaction sequences, in comparison to the real data.

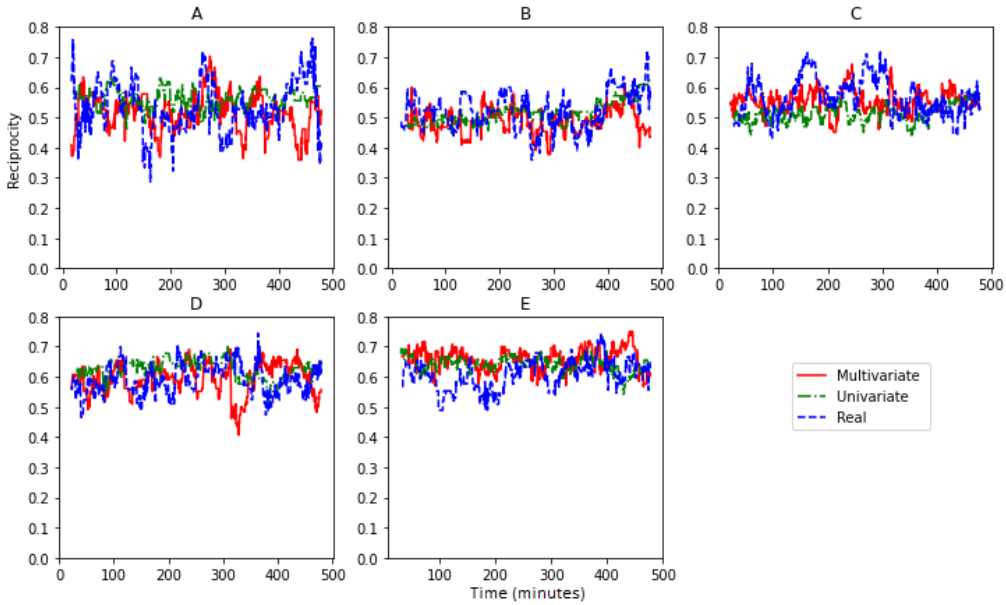


Figure 45: Reciprocity computed across time in rolling windows of 200 transactions for the 5 different datasets, presented for the two different methods of generating edge level transaction sequences, in comparison to the real data.

Data	Univariate	Multivariate
A (full)	0.38	0.22
A (single venue)	-0.63	-0.68
A (off exchange)	0.14	0.57
B (full)	-0.40	-0.53
B (single venue)	-0.58	-0.68
B (off exchange)	0.57	0.98
C (full)	-0.58	0.50
C (single venue)	-0.60	-0.68
C (off exchange)	0.85	0.92
D (full)	-0.35	-0.45
D (single venue)	-0.62	-0.69
D (off exchange)	0.72	0.90
E (full)	-0.29	-0.42
E (single venue)	-0.62	-0.68
E (off exchange)	0.95	0.97

Table 6: Assortativity across all transactions generated by the two multidimensional Hawkes models

Data	Univariate	Multivariate
A (full)	0.56	0.57
A (single venue)	0.64	0.54
A (off exchange)	0.21	0.37
B (full)	0.30	0.44
B (single venue)	0.44	0.54
B (off exchange)	0.46	0.70
C (full)	0.19	0.48
C (single venue)	0.25	0.56
C (off exchange)	0.36	0.76
D (full)	0.37	0.35
D (single venue)	0.43	0.55
D (off exchange)	0.46	0.75
E (full)	0.60	0.64
E (single venue)	0.73	0.63
E (off exchange)	0.29	0.46

Table 7: Reciprocity across all transactions generated by the two multidimensional Hawkes models

Data	Baseline int.	Kernel int.	Kernel dec.
A	$[0.24, 0.11]$	$\begin{bmatrix} 0.57 & 0.42 \\ 0.15 & 0.10 \end{bmatrix}$	$[1.23, 0.24]$
A (SE)	$[0.05, 0.07]$	$\begin{bmatrix} 0.04 & 0.03 \\ 0.07 & 0.03 \end{bmatrix}$	$[0.03, 0.08]$
B	$[0.03, 0.01]$	$\begin{bmatrix} 0.15 & 0.0 \\ 0.21 & 0.58 \end{bmatrix}$	$[0.15, 0.76]$
B (SE)	$[0.03, 0.01]$	$\begin{bmatrix} 0.03 & 0.01 \\ 0.04 & 0.05 \end{bmatrix}$	$[0.03, 0.06]$
C	$[0.29, 0.02]$	$\begin{bmatrix} 0.59 & 0.01 \\ 0.05 & 0.20 \end{bmatrix}$	$[0.67, 0.25]$
C (SE)	$[0.07, 0.02]$	$\begin{bmatrix} NA & 0.03 \\ 0.09 & 0.21 \end{bmatrix}$	$[NA, 0.30]$
D	$[0.20, 0.53]$	$\begin{bmatrix} 0.21 & 0.07 \\ 0.05 & 0.27 \end{bmatrix}$	$[0.33, 0.37]$
D (SE)	$[0.13, 0.07]$	$\begin{bmatrix} 0.04 & 0.03 \\ 0.02 & NA \end{bmatrix}$	$[0.08, NA]$
E	$[0.00, 0.04]$	$\begin{bmatrix} 0.42 & 0.01 \\ 0.27 & 1.07 \end{bmatrix}$	$[0.43, 1.67]$
E (SE)	$[0.00, 0.02]$	$\begin{bmatrix} 0.04 & 0.02 \\ 0.05 & 0.10 \end{bmatrix}$	$[0.04, 0.22]$

Table 8: Hawkes parameters estimated using a Maximum Likelihood approach for the chosen datasets, for a bivariate Hawkes process applied to buys and sells through a central clearing house

Additional model results - bivariate Hawkes process

The estimated parameters for the bivariate Hawkes model are shown in table 8. I see here that all datasets show similar value ranges for the parameters and that higher or lower parameter values for buys in comparison to sells are not observed consistently across the different datasets. The kernel intensity is interesting to compare values across datasets since it captures the level of cross - and self-excitation present. I don't observe consistently across the different dataset a dominance of either of these from the parameter values. For example, instruments C, D and E show larger parameter values for self excitation than cross excitation, however this is not the case for instruments A and B. Figure 46 shows the burstiness values presented in table 5.7 for the bivariate Hawkes process applied to buys and sells.

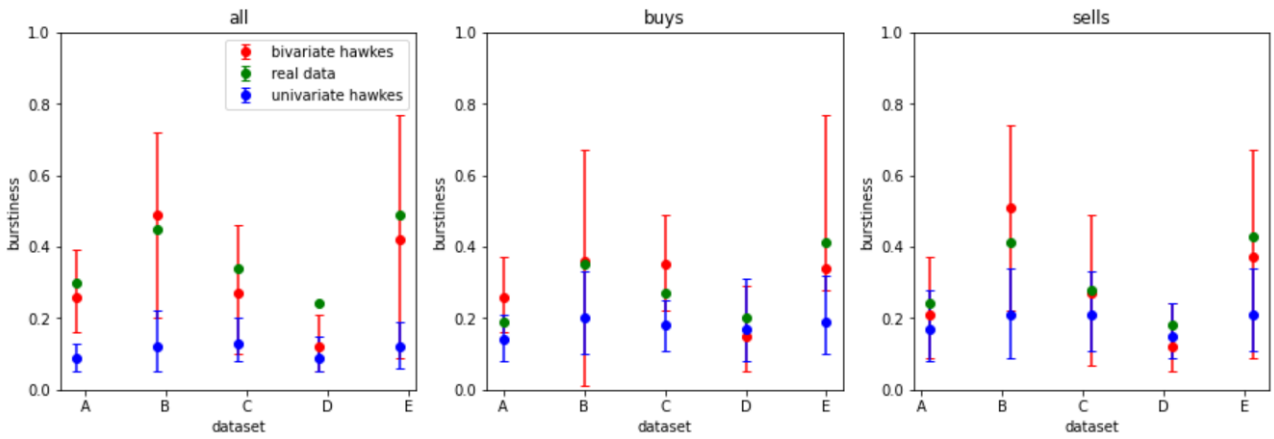


Figure 46: Burstiness of the simulated timestamps using a bivariate Hawkes, for the 5 different datasets considering trades through a single central clearing house, in comparison to the burstiness of the real data.

Application of models for anomaly detection

Here, I present a brief example of how the Hawkes models I present in chapter 5 can be used in practice. I demonstrate this using the multidimensional, univariate Hawkes process as this model showed the strongest performance of the models in reproducing both the temporal and cross-sectional properties of the real data.

I start by fitting the Hawkes process to a single day for a randomly selected instrument, and use the fitted process to generate an ensemble of 200 transaction sequences. I then measure the reciprocity across time for each of these sequences, and calculate the time series of zscores for the reciprocity values for the ensemble. I then consider an additional 10 days of data for the same financial instrument, and calculate the reciprocity across time. I then use the zscores for the simulated ensemble and flag instances when the reciprocity of the new days deviates outside of the zscore range.

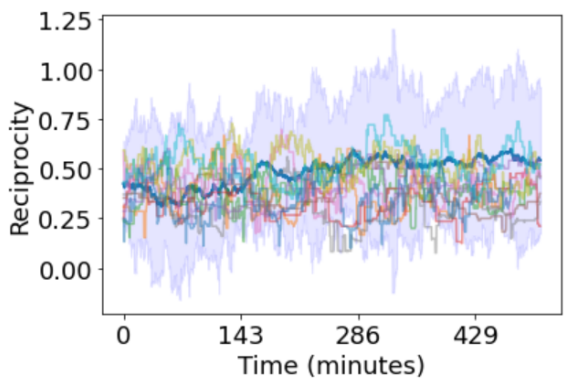


Figure 47: Mean of the reciprocities of 200 simulations using a univariate Hawkes process (thick line), with shaded area represents 3 standard deviation from the mean of this average. Other lines represent the reciprocity for 10 additional days.

Figure 47 shows the mean (thick line) of the reciprocities across time for the ensemble of transaction sequences. The shaded area represents 3 standard deviations above and below the mean, and the additional lines are the reciprocities for the additional days. Visually, I observe that the additional days rarely deviate outside of the shaded area. 3 of the 10 days are found to deviate outside at least once, so would be flagged as anomalous.

.1.4 Robustness assessment of chapter 5 results

Here I present results for additional instruments to complement those presented in chapter 5.

Univariate Hawkes processes

Table 9 shows the burstiness of transaction timestamps for an additional 10 FTSE 100 instruments, across the different subsets. As in the main results in chapter 5, I observe here that although in all cases the burstiness is positive, there are only a few cases where the real burstiness falls within the 95% confidence intervals of the burstiness of the generated data.

Dataset	simulated burstiness	real burstiness
D1 Full	0.07 (0.06, 0.08)	0.65
D1 Single venue	0.54 (0.49, 0.59)	**0.58**
D1 Off exchange	0.26 (0.24, 0.28)	0.35
D2 Full	0.31 (0.29, 0.34)	0.59
D2 Single venue	0.11 (0.10, 0.12)	0.55
D2 Off exchange	0.15 (0.10, 0.21)	0.39
D3 Full	0.28 (0.26, 0.31)	0.56
D3 Single venue	0.36 (0.28, 0.43)	**0.43**
D3 Off exchange	0.16 (0.13, 0.17)	0.28
D4 Full	0.51 (0.47, 0.54)	**0.49**
D4 Single venue	0.78 (0.66, 0.86)	0.32
D4 Off exchange	0.19 (-0.06, 0.39)	0.42
D5 Full	0.28 (0.26, 0.31)	0.57
D5 Single venue	0.18 (0.16, 0.20)	0.42
D5 Off exchange	0.15 (0.12, 0.19)	0.47
D6 Full	0.48 (0.32, 0.60)	**0.52**
D6 Single venue	0.26 (0.25, 0.27)	0.39
D6 Off exchange	0.13 (0.10, 0.15)	0.45
D7 Full	0.31 (0.29, 0.34)	0.57
D7 Single venue	0.45 (0.41, 0.47)	0.48
D7 Off exchange	0.15 (0.12, 0.18)	0.19
D8 Full	0.86 (0.77, 0.92)	0.57
D8 Single venue	0.44 (0.38, 0.51)	**0.48**
D8 Off exchange	0.38 (0.20, 0.59)	**0.42**
D9 Full	0.07 (0.06, 0.08)	0.58
D9 Single venue	0.24 (0.19, 0.27)	0.51
D9 Off exchange	0.08 (0.05, 0.10)	0.32
D10 Full	0.50 (0.18, 0.67)	0.47
D10 Single venue	0.15 (0.13, 0.17)	0.34
D10 Off exchange	0.18 (0.08, 0.41)	0.42

Table 9: Burstiness of the timestamps generated by a univariate Hawkes process, in comparison to the real data

Table 10 shows the parameters estimated for the univariate Hawkes process. As for the original 5 datasets, I observe a reasonable amount of variation across the parameter values for the different datasets for all three subsets considered, with the baseline intensities ranging between 0.0005 and 1.89, the kernel intensities ranging between 0.2 and 11.14 and the kernel decays ranging between 0.3 and

24.15. Whereas before I noted that the kernel decay appeared to show larger values for the off exchange subset and smaller values for the full subset, this is not consistently observed for these datasets.

Dataset	baseline intensity	kernel intensity	kernel decay
D1 Full	0.44 ± 0.14 **	0.37 ± 0.004 ***	2.28 ± 0.08 ***
D1 Single venue	0.10 ± 0.24	0.52 ± 0.02 ***	0.57 ± 0.59
D1 Off exchange	1.98 ± 0.14 ***	1.16 ± 0.04 ***	1.10± NA
D2 Full	0.48 ± 0.08 ***	0.61 ± 0.006 ***	0.68 ± 0.03 ***
D2 Single venue	1.23± NA	0.67 ± 0.01 ***	1.88± NA
D2 Off exchange	1.31 ± 0.27 ***	0.71 ± 0.02 ***	0.74± NA
D3 Full	0.1± NA	0.2 ± 0.002 ***	0.3± NA
D3 Single venue	0.05 ± 0.05	0.44 ± 0.008 ***	0.66± NA
D3 Off exchange	0.005 ± 0.003	0.26 ± 0.006 ***	0.66± 0.06 **
D4 Full	0.65 ± 0.04 ***	0.50 ± 0.005***	0.54± NA
D4 Single venue	1.33 ± 0.14 ***	1.87 ± 0.04 ***	1.87 ± 0.10 ***
D4 Off exchange	0.78 ± 0.22 *	0.56 ± 0.02 ***	0.60 ± 0.21 **
D5 Full	0.12 ± 0.12	0.36 ± 0.003 ***	0.38± NA
D5 Single venue	0.06± NA	10.38 ± 0.06 ***	10.63± NA
D5 Off exchange	0.005 ± 0.005	0.27 ± 0.005 ***	0.63± NA
D6 Full	0.1 ± 0.04	0.2 ± 0.002 ***	0.3 ± 0.06 ***
D6 Single venue	1.59± NA	1.08 ± 0.02 ***	1.56± NA
D6 Off exchange	1.89± NA	1.06 ± 0.02 ***	1.11± NA
D7 Full	0.03 ± 0.02	1.29± NA	1.30± NA
D7 Single venue	0.33± NA	0.62 ± 0.01 ***	0.65± NA
D7 Off exchange	0.80 ± 0.22 ***	0.66 ± 0.02 ***	0.66 ± 0.11 ***
D8 Full	1.27± NA	0.54 ± 0.006 ***	0.76 ± 0.06 ***
D8 Single venue	0.33 ± 0.07 ***	0.62 ± 0.01 ***	0.65± NA
D8 Off exchange	0.16 ± 0.19	3.53 ± 0.09 ***	3.53 ± 0.21 ***
D9 Full	0.90 ± 0.01 ***	0.36 ± 0.003 ***	1.59± NA
D9 Single venue	0.09 ± 0.08	11.14 ± 0.12 ***	24.15 ± 0.13 ***
D9 Off exchange	2.84± NA	$0.75 \pm NA$	0.80± NA
D10 Full	0.0005 ± 0.0004	2.16 ± 0.02 ***	2.31± NA (NA)
D10 Single venue	1.66 ± 0.64 *	0.88 ± 0.03 ***	1.09 ± 0.39 **
D10 Off exchange	0.005 ± 0.005	0.27 ± 0.005 ***	0.63± NA

Table 10: Parameters for the univariate Hawkes model estimated using a Maximum Likelihood approach for the chosen datasets, across the different subsystems considered as dynamic networks. P-values indicated as follows: *** <0.001, ** <0.01, * <0.05. P-values and standard error calculated by taking the inverse of the Hessian, so are not calculated when the Hessian is non-invertible.

Cross sectional properties of univariate models

Table 11 shows the p-values of a two dimensional, two sided Kolmogorov-Smirnov test applied to the rich-club distributions of the simulated data, in comparison to the real data. As in the main research in chapter 5 for the original 5 datasets, here I can see that the null hypothesis of the samples having the same distribution is not rejected for the majority of the datasets for the frequency based edge selection, but I observe very small p-values for the random method of selection which confirms that this method is not able to reproduce the degree distribution of the real data.

Dataset	Random	Frequency
D1 Full	1.82e-27	0.11
D1 Single venue	1.14e-22	0.26
D1 Off exchange	1.77e-07	0.01
D2 Full	6.40e-26	0.42
D2 Single venue	9.26e-13	0.51
D2 Off exchange	6.02e-09	0.01
D3 Full	1.17e-12	3.43e-5
D3 Single venue	1.20e-12	0.08
D3 Off exchange	3.89e-07	0.003
D4 Full	1.84e-30	0.01
D4 Single venue	1.45e-11	0.11
D4 Off exchange	8.95e-06	0.02
D5 Full	6.77e-17	0.06
D5 Single venue	1.65e-11	0.11
D5 Off exchange	3.86e-06	0.11
D6 Full	4.33e-36	0.07
D6 Single venue	1.07e-29	2.75e-5
D6 Off exchange	4.79e-11	0.0002
D7 Full	7.40e-26	0.03
D7 Single venue	2.08e-16	0.0002
D7 Off exchange	3.67e-11	0.01
D8 Full	1.55e-16	0.01
D8 Single venue	2.27e-17	0.008
D8 Off exchange	1.96e-08	0.007
D9 Full	4.83e-43	1.65e-12
D9 Single venue	3.30e-12	0.003
D9 Off exchange	2.41e-06	0.0008
D10 Full	5.24e-31	0.01
D10 Single venue	6.65e-13	6.96e-6
D10 Off exchange	1.31e-06	0.009

Table 11: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the distributions of the Rich Club index over different degrees for the different edge selection methods in comparison to the real data, for the 10 additional instruments.

Table 12 shows the p-values again for a two dimensional, two sided Kolmogorov-Smirnov test applied to each of the degree distributions of the simulated data, in comparison to the real data. As for the rich club distributions, I observe consistency with my results presented in the chapter 5, with the null hypothesis of the samples having the same distribution not rejected for many, but not all, of the datasets for the frequency based edge selection. I observe very small p-values for the random method of selection confirming that this method is not able to reproduce the rich club distributions of the real data.

Dataset	Random	Frequency
D1 Full	5.22e-91	0.05
D1 Single venue	9.61e-66	0.95
D1 Off exchange	7.18e-27	0.72
D2 Full	7.05e-78	0.99
D2 Single venue	1.42e-30	0.99
D2 Off exchange	2.04e-26	0.003
D3 Full	1.45e-68	0.99
D3 Single venue	3.04e-45	0.16
D3 Off exchange	6.25e-47	1.43e-8
D4 Full	1.53e-52	0.99
D4 Single venue	2.07e-28	0.002
D4 Off exchange	2.81e-29	0.04
D5 Full	6.72e-72	0.13
D5 Single venue	5.94e-40	0.002
D5 Off exchange	2.84e-23	0.68
D6 Full	3.34e-82	0.93
D6 Single venue	2.63e-37	0.19
D6 Off exchange	7.18e-28	0.05
D7 Full	1.39e-82	0.43
D7 Single venue	7.39e-83	0.76
D7 Off exchange	5.21e-27	0.006
D8 Full	5.07e-28	0.99
D8 Single venue	2.32e-119	0.15
D8 Off exchange	8.26e-18	5.43e-9
D9 Full	9.61e-74	0.11
D9 Single venue	5.01e-32	0.99
D9 Off exchange	1.64e-34	1.27e-9
D10 Full	3.74e-24	0.02
D10 Single venue	6.90e-71	0.21
D10 Off exchange	4.54e-17	0.005

Table 12: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the degree distributions for the different edge selection methods in comparison to the real data, for the 10 additional instruments.

Multidimensional Univariate Hawkes processes

I now present additional results for multidimensional, univariate Hawkes processes. Table 13 shows the burstiness of the simulated transactions for the different datasets across the different subsets, in comparison to the real burstinesses for the multidimensional univariate 'edge level' Hawkes processes. The results here are in agreement with those seen in chapter 5, showing the true burstiness falling within the 95% confidence intervals in all cases.

Dataset	simulated burstiness	real burstiness
D1 Full	0.30 (0.20, 0.41)	**0.34**
D1 Single venue	0.12 (0.01, 0.24)	**0.14**
D1 Off exchange	0.08 (-0.04, 0.25)	**0.07**
D2 Full	0.41 (0.33, 0.50)	**0.41**
D2 Single venue	0.26 (0.16, 0.37)	**0.24**
D2 Off exchange	0.19 (0.09, 0.30)	**0.28**
D3 Full	0.46 (0.35, 0.56)	**0.47**
D3 Single venue	0.18 (0.08, 0.30)	**0.12**
D3 Off exchange	0.48 (0.25, 0.67)	**0.67**
D4 Full	0.31 (0.22, 0.40)	**0.34**
D4 Single venue	0.21 (0.09, 0.33)	**0.14**
D4 Off exchange	0.05 (-0.05, 0.16)	**0.15**
D5 Full	0.26 (0.18, 0.35)	**0.29**
D5 Single venue	0.28 (0.18, 0.37)	**0.35**
D5 Off exchange	0.13 (0.03, 0.22)	**0.16**
D6 Full	0.33 (0.23, 0.42)	**0.32**
D6 Single venue	0.29 (0.20, 0.37)	**0.25**
D6 Off exchange	0.32 (0.12, 0.53)	**0.38**
D7 Full	0.21 (0.10, 0.36)	**0.24**
D7 Single venue	0.21 (0.01, 0.37)	**0.19**
D7 Off exchange	0.17 (-0.05, 0.42)	**0.13**
D8 Full	0.48 (0.41, 0.53)	**0.49**
D8 Single venue	0.27 (0.18, 0.35)	**0.27**
D8 Off exchange	0.16 (0.10, 0.23)	**0.20**
D9 Full	0.29 (0.18, 0.38)	**0.29**
D9 Single venue	0.13 (0.03, 0.27)	**0.16**
D9 Off exchange	0.19 (0.11, 0.29)	**0.16**
D10 Full	0.20 (0.12, 0.30)	**0.14**
D10 Single venue	0.15 (0.07, 0.24)	**0.21**
D10 Off exchange	0.17 (0.09, 0.27)	**0.21**

Table 13: Burstiness of the timestamps generated by univariate edge level Hawkes processes, in comparison to the real timestamps. Results shown in bold are the cases where the true burstiness falls within the 95% confidence intervals of the simulation results.

Multivariate Hawkes processes

Here I present additional results for multidimensional, multivariate Hawkes processes in table 14. Since the runtime of the multivariate method is high, I have only considered five additional datasets here. The results support those observed in chapter 5, showing the true burstiness falling outside the confidence intervals for the majority the trials.

Dataset	simulated burstiness	real burstiness
D1 Full	0.32 (0.30, 0.34)	**0.34**
D1 Single venue	0.37 (0.36, 0.40)	0.14
D1 Off exchange	0.25 (0.23, 0.26)	0.07
D2 Full	0.39 (0.37, 0.41)	**0.41**
D2 Single venue	0.42 (0.37, 0.45)	0.24
D2 Off exchange	0.15 (0.14, 0.16)	0.28
D3 Full	0.37 (0.35, 0.40)	0.47
D3 Single venue	0.38 (0.36, 0.44)	0.12
D3 Off exchange	0.24 (0.22, 0.27)	0.67
D4 Full	0.40 (0.39, 0.41)	0.34
D4 Single venue	0.39 (0.43, 0.43)	0.14
D4 Off exchange	0.22 (0.20, 0.25)	0.15
D5 Full	0.30 (0.29, 0.31)	**0.29**
D5 Single venue	0.35 (0.34, 0.36)	**0.35**
D5 Off exchange	0.18 (0.16, 0.22)	**0.16**

Table 14: Burstiness of the timestamps generated by multivariate Hawkes processes, in comparison to the real timestamps. Results shown in bold are the cases where the true burstiness falls within the 95% confidence intervals of the simulation results.

Cross sectional properties of multidimensional models

Here I consider the cross sectional properties for the two multidimensional models applied to five additional instruments, for the full dataset only. I present additional results of Kolmogorov-Smirnov tests to compare the rich club and degree distributions of for the two multidimensional models I consider in chapter 5. In both tables 15 for the rich club distributions and 16 for the degree distributions, I observe in all but one test p-values exceeding 0.05, meaning that the null hypothesis of the samples having the same distribution is not rejected, supporting my results in chapter 5.

Dataset	Univariate	Multivariate
D1 Full	0.75	0.63
D2 Full	0.07	0.40
D3 Full	0.08	0.13
D4 Full	0.86	0.73
D5 Full	0.01	0.64

Table 15: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the Rich Club distributions for the multidimensional Hawkes models to the real data, for the 5 additional instruments.

Dataset	Univariate	Multivariate
D1 Full	0.26	0.20
D2 Full	0.04	0.35
D3 Full	0.10	0.14
D4 Full	0.97	0.87
D5 Full	0.35	0.47

Table 16: p-values for a two dimensional two sided Kolmogorov-Smirnov test comparing the degree distribution for the multidimensional Hawkes models to the real data, for the 5 additional instruments.

Bivariate Hawkes processes

Finally, here I present additional results for 10 instruments for the bivariate Hawkes process to model buy and sell trade executions as separate sequences and allow for mutual excitation between the buys and sells. As in the main results presented in chapter 5, I see in table 17 that in the majority of cases, the true burstiness falls within the confidence intervals of the simulations for both the full transaction sequences as well as the buys and sells themselves.

Data	Real data	Univariate	Bivariate
D1 (all)	0.48	0.12 (0.06, 0.21)	**0.30** (0.13, 0.54)
D1 (buys)	0.40	0.21 (0.10 , 0.37)	**0.29** (0.12,0.48)
D1 (sells)	0.38	0.20 (0.10, 0.33)	**0.29** (0.11,0.56)
D2	0.46	0.18 (0.13, 0.23)	**0.42** (0.27,0.55)
D2 (buys)	0.41	0.22 (0.16, 0.28)	**0.36** (0.24,0.49)
D2 (sells)	0.38	0.24 (0.17, 0.31)	**0.38** (0.23,0.55)
D3	0.45	0.13 (0.06, 0.20)	**0.32** (0.16, 0.53)
D3 (buys)	0.35	0.21 (0.11, 0.34)	**0.24** (0.01,0.57)
D3 (sells)	0.36	0.21 (0.10, 0.34)	**0.29** (0.13,0.47)
D4	0.37	0.12 (0.06, 0.21)	**0.26** (0.12, 0.41)
D4 (buys)	0.30	**0.21** (0.10, 0.34)	**0.21** (0.10,0.34)
D4 (sells)	0.36	0.20 (0.10, 0.35)	**0.40** (0.21, 0.63)
D5	0.47	0.12 (0.06, 0.21)	**0.51** (0.22, 0.71)
D5 (buys)	0.39	0.21 (0.10, 0.36)	**0.44** (0.18, 0.70)
D5 (sells)	0.38	0.20 (0.10, 0.33)	**0.45** (0.21,0.71)
D6	0.49	0.13 (0.07, 0.19)	0.15 (0.09, 0.21)
D6 (buys)	0.40	0.21 (0.11, 0.33)	0.29 (0.20,0.38)
D6 (sells)	0.42	0.20 (0.10, 0.31)	0.10 (0.04,0.16)
D7	0.47	0.12 (0.05, 0.21)	0.31 (0.17, 0.47)
D7 (buys)	0.39	0.20 (0.10, 0.37)	**0.29** (0.15, 0.48)
D7 (sells)	0.46	0.21 (0.10, 0.36)	**0.35** (0.21, 0.49)
D8	0.49	0.32 (0.24, 0.38)	**0.51** (0.28, 0.74)
D8 (buys)	0.38	**0.36** (0.26, 0.45)	**0.44** (0.17,0.71)
D8 (sells)	0.43	**0.35** (0.26, 0.44)	**0.46** (0.22, 0.72)
D9	0.49	0.09 (0.04, 0.14)	0.31 (0.27, 0.35)
D9 (buys)	0.43	0.15 (0.08, 0.24)	0.26 (0.21, 0.30)
D9 (sells)	0.41	0.17 (0.07, 0.27)	0.30 (0.26, 0.34)
D10	0.35	0.10 (0.03, 0.19)	**0.24** (0.09, 0.41)
D10 (buys)	0.33	0.17 (0.07, 0.31)	**0.16** (-0.008, 0.40)
D10 (sells)	0.28	**0.18** (0.08, 0.29)	**0.29** (0.17, 0.45)

Table 17: Burstiness of timestamps generated with a bivariate Hawkes process. Results highlighted in bold indicate where the real burstiness falls within the 95% confidence intervals of the model burstinesses.