Thesis Skeleton

Network Analysis on Post Trade data: variation in topology and stability of the Scale-free Networks over time (Possible title)

Andrea Ierardi

September 2021

Abstract

The aim of the study is the construction of Social Networks from Settlement Instructions of T2S. The topological structure of the Network may change over time due to disruptive events. Two Case-studies are conducted on disruptive events: COVID19 and BTP Italia, BTP Futura emissions. The identification over time of a Scale-free behavior and a ranking for the most central nodes is conducted. Moreover, a networks resiliency analysis is performed using random and targeted attacks.

1 Introduction

Network Analysis is a set of integrated techniques to depict relations among actors and to analyze the social structures that emerge from the recurrence of these relations. It can be applied in a lot of domains: Biology, Demography, Economics, Geography, History, Health, Social psychology and Computer science. It is vastly used for the potentiality of simplification of complex system.

1.1 Network definition

A network can be represented by a *Graph*. In mathematics, and more specifically in graph theory, a graph is a structure composed by a set of objects in which some pairs of the objects are in some sense "related". A graph consists of two types of elements, *nodes* and *links* between them. Links can have weights that represent the importance of the relationship between nodes. If a link from one node to another one, has a direction the graph is called *Directed*, otherwise if there is not a direction the graph is called *Undirected*. Social network analysis allows to identify local and global patterns, locate influential entities, and examine network dynamics.

1.2 Network Metrics and Topology

Network Analysis allows to obtain useful measures and insights from a graph representation. [RPA21]

1.2.1 Size and Order

The size of a Network is the total number of nodes while the order is the total number of links.

1.2.2 Degree distribution

The degree distribution take into account the number of link in a graph. In particular, the average mean degree is the average number of links per node in the network. In case of a directed graph, the degree can be divided divided in in-degree and out-degree. The in-degree measures the number of in-going links of a node, the out-degree measures the number of out-going links of a node.

1.2.3 Assortativity, Density and Average Clustering

Assortativity is the measure in which nodes with a certain degree are more likely to have links with nodes of similar degree and it is called disassortative when the contrary occurs.
Density is the ratio of the number of existing connections versus the maximum number of connections.
Average Clustering Coefficient is a measure of the degree to which nodes in a graph tend to cluster together.

1.2.4 Average Path Length

Average Path Length is the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network. Most real networks have a very short average path length leading to the concept of a small world (Section 1.3.2 where everyone is connected to everyone else through a very short path.

1.2.5 Connected Components

In a directed network the largest component is known as the Giant Weakly Connected Component (GWCC) in which all nodes connect to each other via undirected paths. The core is the Giant Strongly Connected Component in which the nodes can reach each other through a directed path. The Disconnected Components (DC) are smaller components for which there exist two vertices in the Component graph such that no path has these vertices as endpoints.

1.3 Scale-free Networks

1.3.1 Scale-free Definition

A network is called *scale-free* if the characteristics of the network are independent of the size of the network (i.e. the number of nodes). This means that when the network grows, the underlying structure remains the same. A scale-free network is defined by the distribution of the number of edges of the nodes following a so called *power law* distribution. An example of central concern for macroeconomics are production networks whose scale-free nature has recently been put forward [ACOTS12] as a potentially major driver of macroeconomic fluctuations.

1.3.2 Small World Theory

Small world theory is based on the idea that two individuals will be connected through a series of intermediaries. In the 1960s, Stanley Milgram tested this theory [TM69] in which all nodes are distance from each other for a short path. The "Six degrees of separation" is the idea that all people on average are six, or fewer, social connections away from each other. As a result, a chain of "friend of a friend" statements can be made to connect any two people in a maximum of six steps.

1.3.3 Scale-free Properties

- **Growth**: a growth process where, over an extended period of time, new nodes join an already existing system.
- **Preferential Attachment**: In real networks new nodes prefer to link to the more connected nodes.
- **Scale-free Resiliency**: Scale-free networks are more resistant to random disconnection of nodes. It can be eliminated a considerable number of nodes randomly and the network's structure is preserved and will not break into disconnected clusters. When the most connected nodes are targeted, the diameter of a scale-free network increases and the network breaks into isolated clusters. This occurs because when removing these nodes, the damage disturbs the heart of the system, whereas a random attack is most likely not.

1.4 Centrality Definition

In graph theory and network analysis, indicators of centrality assign numbers or rankings to nodes within a graph corresponding to their network position. Applications include identifying the most influential persons in a social network, key infrastructure nodes in the Internet or urban networks, super-spreaders of disease. The word "importance" has a wide number of meanings, leading to many different definitions of centrality. Two categorization schemes have been proposed. "Importance" can be conceived in relation to a type of flow or transfer across the network. This allows centralities to be classified by the type of flow they consider important. "Importance" can alternatively be conceived as involvement in the cohesiveness of the network.

2 Methodology

2.1 Data description

Settlement Instructions of T2S are retrieved using Spark queries and stored in a specific bucket in Amazon AWS. The data comprehends instructions where dates range from May 2018 to July 2021. The used features are Instruction Deliver and Receiver, Amount Counter-value, Cross Border Instruction Indicator, ISIN, Type of instruments, Status of the instruction.

2.2 Network construction

Two years instruction data cannot be easily managed in tabular data. Network structures help in the simplification process and in both visualisation and inspection of a large amount of data. In particular, they can be useful in providing a representation of the relationship between different companies. The strategy is the simplification of all the instructions exchanged between companies building Networks where nodes represent companies and links represent instruction exchanges. Links are weighted using the amount of instruction counter value. Since there are different instructions during a business day, the networks are constructed in this way:

- **Daily aggregation**: for each business day the data are aggregated for *Deliver* and *Receiver* name, *Cross Border Instruction Indicator*, *ISIN*, *Type of financial instrument*, *Status of the instruction.* For each of this feature different networks are obtained.
- **Monthly aggregation**: for each month data are aggregated and information such as *Deliver* and *Receiver* name, *Type of financial instrument* and *Status of the instruction* are used for the networks construction.
- **Monthly cumulative aggregation**: same as Monthly Aggregation type, but in this case nodes and links are added cumulative each month to the previous month network.

For each aggregation feature combination a different network is obtained. Each Network has different nodes, links and characteristics.

3 Experiments

Different sets of experiments are conducted based on the different type of data aggregation. The analysis is focused on computing different metrics and topological structure for both the two techniques of data aggregation. The metrics calculated are those considered in the Section 1.2:

- Size and order is useful for understanding the number of companies and instruction exchanges in a certain period
- Degree distribution helps understanding how much instruction are exchanged between companies over time
- From Assortativity is possible to understand if companies tend to exchange instructions with those similar (or dissimilar in case of disassortative behavior)
- Density can be useful for understanding if companies exchange with only few participant
- Clustering coefficient indicates if there is a link between two companies who have a common trading partner.
- Average path length is useful for understanding a Small word theory may be applied or not (See section 1.3.2). It means that any node can be reached from any other node in only a few steps.
- Connected Components may help in identify the structure of the network and the different paths between the companies.

The estimate of all these metrics over time is a key part of the analysis. They help to identify local and global patterns, locate influential entities, and examine how network dynamics are changing in a daily and monthly bases

3.1 Experiments on Daily Aggregated data

3.1.1 Disruptive Events

Disruptive Events could reversely change the Network topology and structure. Studies on disruptive events such as the September 11^{th} 2001 terrorist attack have shown how this event can alter economics systems. [SBA^{+}06] Financial system of the USA have been disrupted, this affected the structure of interbank payments. This is due to the fact that the attacks damaged property and communications systems making impossible for many bank to execute payments. A similar study is performed taking into consideration two different case-studies of disruptive events:

- The impact of Covid19 on the network metrics and topology during the first lockdown (January 2020 - June 2020)
- The impact of large emission of Government Bonds such as BTP Italia and BTP Futura on the network metric and topology. The analysis takes into consideration the next ten days after the BTP announcement. This is considered a disruptive event since large amount of instructions are exchanged during the emission dates.

3.2 Experiments on Monthly Aggregated data

3.2.1 Scale-free Behavior

A Scale-free analysis is conducted for the monthly aggregated networks. In particular, the idea is the detection of the Scale-free behavior for each month based on the type of financial instrument and settlement status of the instructions. Scale-free properties (Section 1.3.3) can be useful for understanding how companies prefers to connect with each other. Preferential Attachment implies that a new company will prefer to connect with a more central and connected company. This means that when the network grows, the underlying structure remains the same. It can help in understanding how future connection will be in the networks. Moreover, it is important to define at which degree the scale-free behavior is present: Super-Weak, Weakest, Weak, Strong, Strongest. [BC19]

3.2.2 Centrality Analysis

The idea is to detect the most central nodes in the network each month. Different techniques have been applied for the computation of the most Central nodes: Degree Centrality, Betweenness Centrality, Closeness Centrality, Eigenvector Centrality, PageRank Centrality. Using these techniques, for each month a nodes ranking is obtained. In particular, the objective is the identification of the most 10^{th} central nodes and also their position changes over time. If a node does not change position or is not excluded from the ranking, it means that Preferential Attachment property is confirmed.

3.2.3 Network Resilience Analysis

A Network Resilience Analysis is performed for the monthly aggregated networks. In a network a node deletion means that a Company may goes to bankrupt or most commonly the company decides to exit the system. This deletion may alter the structure of the networks. To verify the structure alterations, it is possible to detect any changes in the Connected Component (CC) of a Network. If after a deletion, a change in the CC is present, this means that the node is a vulnerable one and the Network has been damaged and compromised. If after a deletion, there is no changes in the Connected Component, this implies that the networks has not been compromised. Moreover, the scale-free property strongly correlates with the network's robustness to failure. The major hubs are closely followed by smaller ones. These smaller hubs, in turn, are followed by other nodes with an even smaller degree and so on.
Different node deletion approaches are applied:

- Random Node deletion: a node is deleted randomly form the Network
- Localized Node deletion: a deletion of a precisely selected node

If failures occur at random and the vast majority of nodes are those with small degree, the likelihood that a hub would be affected is almost negligible. Localized attacks makes the scale free network more vulnerable compared to random attacks. In this case the targets are the most central nodes with high number of connection and amount counter-value delivered. Network of payments have shown scale-free properties in literature [dlTKKE16] and they are resilient to random damage. This means that it is barely possible to destroy the network of payments by random removal, but if an exact portion of particularly selected nodes is removed, it breaks completely.

References

[ACOTS12] Daron Acemoglu, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80:1977–2016, 2012.

[BC19] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nat Commun*, 2019.

[dlTKKE16] Stephanie Rendòn de la Torr, Jaan Kald, Robert Kit, and Juri Engelbrecht. On the topologic structure of economic complex networks: empirical evidence from large scale payment network of estonia. *Chaos Solitons*, 2016.

[RPA21] Jeniffer Rubio, Bryan Pérez, and John Arroyo. Risk monitoring in ecuador's payment system: Implementation of a network topology study. *Latin American Journal of Central Banking*, 2:100039, 2021.

[SBA+06] Kimmo Soramak, Morten L. Bech, Jeffrey Arnold, Robert J. Glass, and Walter E. Beyeler. The topology of interbank payment flows. *Federal Reserve Bank of New York Staff Report*, 2006.

[TM69] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.