WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

Masters Programmes

DISSERTATION COVER SHEET

Degree Course: **MSc Business Analytics**

Student ID Number: [redacted]

Title: **Market Basket Analysis: Discovering Customers' Purchasing Behaviour using Product Networks**

Dissertation Code: [redacted]

Submission Date: **25 August 2023**

Submission Deadline: **29 August 2023**

Word Count: **9,979 words**

Number of Pages: **49**

"I declare that this work is entirely my own in accordance with the University's Regulation 11 and the WBS guidelines on plagiarism and collusion. All external references and sources are clearly acknowledged and identified within the contents.

No substantial part(s) of the work submitted here has also been submitted by me in other assessments for accredited courses of study, and I acknowledge that if this has been done it may result in me being reported for self-plagiarism and an appropriate reduction in marks may be made when marking this piece of work."

Market Basket Analysis: Discovering Customers' Purchasing Behaviour using Product Networks

By

[redacted]

ABSTRACT

The purchasing behaviours of customers are examined from multiple perspectives, one of which involves analysing the items that customers consistently purchase together in the same transaction. The discipline under consideration is commonly referred to as market basket analysis, which involves the extraction of frequently purchased itemsets from sales transactional records. Initially, market basket analysis was conducted via association rules mining, wherein product rules deemed to be "interesting" were considered. However, the method has significant downsides. In recent years, there has been a pivot in the literature towards the utilisation of machine learning approaches for conducting market basket analysis. The primary objective of this study was to employ product network analysis to get insights into customers' purchasing behaviour and provide recommendations for operational and marketing strategies. The extraction of product sets from the pruned product network is accomplished through the utilisation of a community detection method, Louvain algorithm. Subsequently, these communities are analysed using community density, degree, betweenness, closeness, and eigenvector centralities. Based on the derived itemsets, we propose four recommendations pertaining to the store's business strategies to enhance their profitability. These recommendations encompass the shelves arrangement optimisation, a suggestion on website recommendation system, inventories monitoring, and a marketing campaign.

Keywords: market basket analysis, purchasing behaviour, product network, community detection, retail store, strategies recommendation

ACKNOWLEDGEMENT

I would like to show my best sincere to Dr Naderi Siamak, my dissertation supervisor, for the thoughtful feedbacks and impactful guidance throughout this research journey. I truly appreciate the time you invested in the project which facilitated efficient and effective communication and was critical to the successful conclusion of this dissertation.

Also, I would want to express my gratitude to my family and friends who have been so encouraging and supportive throughout the process of earning my master's degree. Their support helped me stay motivated and accomplish quality works.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AR	Association Rules
ARM	Association Rules Mining
CA	Cluster Analysis
CTM	Correlated Topic Model
DL	Deep Learning
FFNN	Feed-Forward Neural Network
LDA	Latent Dirichlet Allocation
MBA	Market Basket Analysis
NA	Network Analysis
NLP	Natural Language Processing

1. INTRODUCTION

1.1 Research Motivation

Consumer behaviour is crucial for the success of a business. Businesses that possess a comprehensive understanding of their customers' purchasing habits are well-positioned to flourish and generate profits. Businesses integrate customer trait knowledge into their strategic and day-to-day operations. Elhawary (2023) highlighted the significance of businesses comprehending the purchasing behaviour of their target customers, specifically their interests, in order to develop marketing campaigns that effectively engage them and enhance customer loyalty. Forbes' Expert Panel (2020) emphasises the importance of entrepreneurs keeping abreast of evolving customer behaviours and adjusting their operations for sustained success.

Customer purchasing behaviour can be viewed from a variety of perspectives, including the length of time customers spend inside the store, customer traffic at each section, the products customers purchase in the same basket, and so on. All of which reveal interesting characteristics for stores to manage their strategies. Knowing frequently purchased product sets could shed light on various customers' purchasing habits. The buying patterns could be conceived for a variety of reasons, such as customers' psychological associations with each product (for example, a person associates "coffee" with "pastry"). This may be due to multiple cognitive laws in our mind such as the law of similarity, contrast, contiguity, and frequency, in which our mind is reminded of things that are like specific instances, opposite to specific instances, related actions to specific instances, and events that occur repeatedly with the specific instances (PsycholoGenie, 2014). Product sets may also exist because of the complementary functionality of the product itself, such as "chair and table" or "brush and paint" that work together. Businesses can implement the extracted insights to tailor their strategies to boost the purchase of these products.

To elaborate, stores could reorganise their shelves and website layouts to influence customers' purchasing decisions by controlling what they see next when they browse for a specific product. Suppose product A is found to be associated with product B, a store could adjust the shelves or webpage item positions of product B to be next to product A such that customers see product B right after they put product A into their basket. This is because when customer sees product A, their cognitive will be reminded of product B and if the environment is set in the way that product B can be noticed instantly by customers, they are more likely to purchase product B. This may also be applied in an online store's recommendation system and in an inventory management for the store.

Furthermore, management could create a promotional campaign based on the insights of popular itemset. This would attract more attention and sales because customers would want to buy more if they perceived the campaign to be worth the value of all the products they would like to purchase together. Stores could make the product that has been reported to be more influential in the purchase of other products the star product in the campaign to increase the campaign's attractiveness.

1.2 Background of Market Basket Analysis

Market basket analysis (MBA) is a research methodology that involves the study of transaction records in order to find itemsets that are commonly purchased. The study necessitates data pertaining to the product sets engaged in the transaction, as well as a sufficient number of transactional records as desired by the organisation. The primary objective of an MBA is to analyse and interpret consumers' purchasing behaviours in order to identify patterns and associations between the goods or services that customers tend to purchase together. The objective of this initiative is to assist businesses in formulating strategies that utilise extracted information to enhance store profitability, as discussed in detail in Section 1.1: Research Motivation.

There exist several data mining strategies that can be employed to get frequently purchased itemsets. The subject matter can be broadly classified into two major groups, namely association rules mining and unsupervised machine learning. Each of these categories encompasses a range of model selections. The first group has been utilised by a significant number of earlier academics and was initially investigated by Agrawal et al. (1993), while the second category has been embraced by more modern scholars such as Reader and Chawla (2011) and Videla-Cavieres and Ríos (2014). This phenomenon can be attributed to increased utilisation of machine learning tools and their enhanced efficacy. Further elucidation will be provided in Section 2, namely in the Literature Review.

1.3 Research Aim

This study will investigate market basket analysis using an unsupervised machine learning method, specifically Network Analysis. The goal is to focus on delivering the most insightful information pertaining to frequently purchased together products in the most informative manner possible. This study will investigate the true consumer purchasing behaviour and recommend tailored potential strategies that the stores are able to employ in operation and marketing to enhance profitability.

1.4 Dissertation Outline

This dissertation consists of six main sections, excluding the Abstract, Bibliography, and Appendices. Section 1 provides an overview of the research motivation, background information on market basket analysis, and the aim and main focuses of the study. In Section 2, relevant literature is reviewed to guide the main research of this dissertation. Section 3 explains all the methodologies this dissertation will employ. Section 4 encompassing data preparation, model development, analysis, and limitations. Section 5 will analyse the findings and provide recommendations to improve store profitability through operational and marketing strategies. Section 6 will provide a summary of the findings and propose potential areas for future research.

2. LITERATURE REVIEW

This section scrutinises research on MBA that aims to extract practical insights for businesses. Scholarly works employ diverse methodologies. However, as explained in Section 1.2, there are two main categories of methodologies: association rules mining and unsupervised machine learning techniques. Due to the lack of a universally accepted framework for conducting an MBA, the following sections will be organised according to the authors' selected methodology. This will aid in determining the most appropriate approach for this dissertation.

2.1 Association Rules Mining

MBA was initially approached using the Association Rules Mining (ARM) by Agrawal et al. (1993). It is unquestionably the most well-known and widely used method for MBA. Several studies from the early to mid-2000s used this method as their primary methodology for studying frequently purchased product combinations. For example, Kaur and Kang (2016) on bakery store and Kaur and Singh (2013) on sports equipment. The most used algorithms are Apriori, FP Growth (Frequent Pattern Mining), and ECLAT (Equivalence Class Clustering and Bottom-Up Lattice Traversal) algorithms, all at which have association rules as the theoretical basis.

Association rules (AR) aid understanding of customers' purchasing behaviour by determining the probability of one product to be purchased following the purchase of another (Agrawal et al., 1996). It can be explained as follows. Let T be the customer transaction in the database D containing itemsets and I be the set of items over the binary domain $\{0,1\}$, such that T belongs to I, $T \subseteq I$. To illustrate, suppose $I = \{apple, egg, milk, corn\}$ and $D = \{T_1, T_2, T_3, T_4, T_5\}$ shown in Table 1. The transaction T_1 contains $\{apple, corn\}$, while transaction T_2 contains $\{egg, milk, corn\}$. The AR of product A and B i.e., $A \rightarrow B$, where $A \subset I, B \subset I, A \cap B = \emptyset$ implies that people who buys A also buys B.

	$apple$	egg	$milk$	$corn$
T_1	1	0	0	1
T_2	0	1	1	1
T_3	0	1	0	1
T_4	0	1	1	0
T_5	1	0	0	1

Table 1: Illustration example of transactions and itemsets

The rule must, then, be assessed by roughly two stages under the ARM. First, examine whether the rule is 'strong' by exceeding the minimum predetermined thresholds of 'support' and 'confidence'. Support assesses the probability of product A occurring in transactions. Confidence value tells probability of a customer buying product B when buying product A. Rules will then be assessed using metrices such as 'lift', 'leverage', and 'conviction' to decide whether the rule is considered as "interesting" enough to be used for implementation (Raorane et al., 2012; Ünvan, 2021). The interestingness indicator is also known as the prediction accuracy hence these metrices would indicate the reliability of the rules (Ünvan, 2021). Lift measures the probability gain of a customer likely to purchase product B when product A is purchased (Han et al., 2012). Leverage is the expected probability of product A and product B occurs in the same transaction (Hermina et al., 2022). Conviction is the expected probability of rule ($product_A \rightarrow product_B$) being incorrect assuming product A and product B has a random association. Formulas of all the metrices can be found in Appendix 1. The considerably "interesting" rules, for instance the rule having leverage value of more than 1, conviction of less than 1, lift of greater than 1, will be used for interpretation and assist strategies afterward (Ünvan, 2021).

ARM has several limitations. Because the algorithm goes through the dataset many times, it impacts the efficiency of the analysis as the algorithm requires more processing time and computing capability which is not ideal for business practices (Gupta and Mamtora, 2014). Furthermore, redundant rules may be generated, making it more difficult to appropriately choose the most impactful rules, lowering the accuracy and reliability of the analysis (Raeder and Chawla, 2011; Gupta and Mamtora, 2014; Tan and Lau, 2013). Association rules mining tends to perform poorly when given large dataset (Videla-Cavieres and Ríos, 2014). This makes it less productive for larger ventures with larger datapoints. This is especially problematic in today's world, where it is easier to capture and store data, and more data must be analysed. Moreover, the lack of specified methods to determine minimum support and confidence appropriately results in an abundance number of rules that became 'uninteresting' and are dropped (Faridizadeh et al., 2018). This incurs the opportunity cost of failing to capture useful insights that could lead to profitability.

2.2 Unsupervised Machine Learning

Consequentially, recently, many market basket researchers pivot their attentions to unsupervised machine learning models as they can analyse larger datasets and are often more accurate than ARM, according to Hruschka (2021).

2.2.1 Natural Language Processing

Natural language processing (NLP) is a broad subfield of data science that has various of applications in various areas such as topic modelling, specifically Latent Dirichlet allocation (LDA), in MBA. However, majority of literatures focusing on the MBA do not choose to use NLP as their approach. In addition, to the best of our knowledge there is no literature that used NLP as their main methodology for conducting MBA as other tools are seen to be more effective in doing so. Hruschka (2021) pointed out that topic models (LDA and the correlated topic model (CTM)) performed inferior to binary factor analysis models which performed drastically inferior to other deep learning models in their studies. This conforms with Gatzioura and Sànchez-Marrè (2014) where they stated that topic model (LDA) performing comparatively inferior to other models in producing recommendations of sets of items.

NLP is, also, commonly used alongside algorithms like ARM, rather than as a replacement, for MBA tasks. Bako et al. (2020) and Velmurugan and Hemalatha (2020) employed NLP to classify data into groups for analysis of frequent co-occurrences with ARM. It is said to enhance the comprehension of subgroup behaviours.

2.2.2 Cluster Analysis

Numerous scholars in this field commonly use cluster analysis (CA) as a supplementary tool rather than a substitute for other MBA methods. According to Raeder and Chawla (2011), the sensitivity of k-means to the initial location of cluster centres can complicate the analysis and require the determination of an additional parameter. Musalem et al. (2018) is the sole literature from my observation that utilises CA techniques, specifically k-means, to cluster product categories with high joint purchase probabilities to create product baskets. The cluster count is determined by examining the silhouette of the MDS map of product categories using the 'duplication factor' metric. This metric normalises the observed joint purchase probability by the expected probability of joint purchase under the assumption that the two products are independent. The researchers discovered that each cluster, or basket, can be distinguished based on the characteristics of its products, such as non-perishable and fresh items.

Some Literatures uses CA after ARM with the aim to explore and compare itemsets rigorously, while some adopt clustering prior to ARM with the purpose to investigate within and between groups' behaviours (Alawadh and Barnawi, 2021; Roodpishi and Nashtaei, 2015; Boztug and Reutterer, 2008; Lim 2021; Annie and Kumar, 2012). Distinctly, AL et al. (2022) employed clustering algorithm for the main purpose of dataset size reduction to better facilitate the association rules mining.

2.2.3 Deep Learning

According to Rehman and Ghous (2021) where they reviewed seventy articles related to MBA with association rules and deep learning (DL) with the purpose to extract useful insights for future research, results showed that majority research with DL gives a relatively high accuracy. Agreeably, Hruschka (2021) suggested that exploring deep belief net performs best with the lowest uncertainty in its basket recommendation, requires lesser observations, and is capable of analysing dataset with large number of product categories to other models.

Multiple studies have utilised a deep learning (DL) technique called artificial neural networks (ANN), specifically the feed-forward neural network (FFNN) and addresses challenges in association rules mining (Gangurde et al., 2018; Bhargav et al., 2014; Yang and Sudharshan, 2019; Ojugo and Eboka, 2019). This model addresses the issue of repetitive dataset scanning by modifying the neural network's weights. Additionally, it was discovered that significant time savings can be achieved when executing operations on large datasets. Other studies have also shown that neural network techniques, such as neural gas network and growing self-organising neural network, can effectively analyse large and high-dimensional datasets with unknown distributions (Decker and Monien, 2003; Decker, 2005). However, it is realised that DL models are more complex than other models (Gangurde et al., 2018).

2.2.4 Network Analysis

Network analysis (NA) is, also, widely adopted for MBA (Raeder and Chawla, 2011; Kim et al., 2012; Wu et al., 2021; Videla-Cavieres and Ríos, 2014; Valle et al., 2018; Faridizadeh et al., 2018). This pertains specifically in aspect of community detection. Reader and Chawla (2011) introduced the model of ‘product network’ that is based on the concept of NA but uses product coexistence in the transactions as links. It is agreed by various literatures including Videla-Cavieres and Ríos (2014), Faridizadeh et al. (2018), Valle et al. (2018), and Li et al. (2008) that this approach can extract more meaningful and valuable frequently purchased itemsets compared to traditional methods such as ARM. They employed the weighted product network as a methodology for capturing the relationships between individual items that are purchased. The weights assigned to the edges are derived from the frequency of transactions involving the two products. Furthermore, to initially analyse the networks, Reader and Chawla (2011) uses degree distributions to investigate networks. This represents the probability distribution of the degree of connection of nodes over the whole network. They found that product networks typically have a heavy-tailed distribution meaning there would be few clusters within the network and each node would have few neighbours.

Subsequently, majority of literatures suggested to prune the weighted product-to-product network rigorously by applying a predetermined minimum threshold to eliminate relations that exhibit significant weakness or may arise because of coincidental factors (Kim et al., 2012; Wu et al., 2021; Videla-Cavieres and Ríos, 2014; Valle et al., 2018; Faridizadeh et al., 2018; Raeder and Chawla, 2011). They also suggest adopting community detection techniques after achieving the pruned network to extract sets of highly interconnected nodes out representing associated itemsets that can interpret as frequently purchased itemsets in MBA. The process of community extraction can be accomplished using various methods, which will be elaborated upon in Section 2.2.4.2.

2.2.4.1 Network Improvement

To construct a comprehensive product network, it is important to eliminate random or low co-occurrences. Analysts should strategically select a minimum threshold to achieve this. The absence of standardised guidelines for selecting a threshold for network pruning requires researchers to exercise their judgement and consider the characteristics of the data. Kim et al. (2012) and Wu et al. (2021) used the average support value of their network as the pruning threshold. Videla-Cavieres and Ríos (2014) chose thresholds of 5% and 10% for their top three heavy edges. Valle et al. (2018) employed minimum spanning trees (MST) to connect nodes with high correlation and discard low co-occurrence or random connections. Faridizadeh et al. (2018) set the weight of edges threshold at 10, resulting in only 38% of their original network nodes being included. Raeder and Chawla (2011) proposed a parameter choice guideline. It is suggested to compare the modularity values obtained by varying the minimum thresholds. This can be done by starting with a lower threshold and gradually increasing it in an iterative manner. The minimum threshold and corresponding modularity values are plotted. The local and global maximum of modularity should be considered and chosen based on subjective judgement for the specific case. Hence, the process of network pruning is inherently subjective due to the absence of a universally accepted method.

2.2.4.2 Itemset Extraction Method

To identify highly related itemsets in co-purchasing, many literatures extract communities within the product network. Various methods are available for community detection, such as CNM, Louvain, GN, Label Propagation, etc., each employing distinct approaches to recognise communities.

Although concerns about overfitting and the extraction of non-insightful communities have been raised regarding the modularity maximisation method for community detection, it has been utilised by various researchers in their studies (Kafkas et al., 2021). Community detection using modularity maximisation measures the discrepancy between the number of edges connecting nodes within a community and the expected number of edges connecting nodes within the community in a random network with a similar distribution (Raeder and Chawla, 2011; Wu et al., 2021; Videla-Cavieres and Ríos, 2014). Modularity refers to the level of interconnectivity among nodes in a community, indicating the quality or significance of the community (Newman, 2006). The CNM and Louvain algorithms are frequently employed for maximising modularity. Woma and Ngo (2019) found that the two algorithms outperformed other alternatives in detecting communities in the YouTube network.

Though Raeder and Chawla (2011) pointed out that there are no significant differences across various community detection algorithms, researchers' algorithm choice are not aligning. There are various factors that can contribute to this, including individual preferences and the compatibility with one's dataset.

Clauset et al. (2004) introduced the CNM algorithm, a hierarchical agglomeration method, for identifying communities in networks. This algorithm effectively identified meaningful communities related to customers' purchasing patterns. The algorithm merges communities and recursively compares modularity values to find the highest modularity level and the most optimal community extraction. Many studies in the early 2000s employed this algorithm for community detection (Basuchowdhuri et al., 2014; Mitra et al., 2016).

The Louvain algorithm, similar to CNM, is a hierarchical clustering greedy algorithm that aims to maximise modularity. The algorithm assesses the impact on modularity by moving each node i from its current community to a neighbouring community, rather than solely merging communities for optimal modularity (Woma and Ngo, 2019). The Louvain algorithm treats each community as a node and calculates the sum of edges connecting the communities in each iteration to determine modularity gain. Thus, the community formed by this approach can be considered "robust," allowing us to examine the connections within each product community and gain valuable insights into consumer buying behaviours. The algorithm is considered efficient and effective for large datasets, as demonstrated in previous studies (Faridizadeh et al., 2018; Zhang et al., 2021).

There is increasing research on the Leiden algorithm, a newly developed method for detecting communities. This algorithm, similar to the Louvain algorithm, has been claimed by Traag et al. (2019) to ensure well-connected communities and be significantly faster, with a speed improvement of 10 to 100 times compared to the Louvain algorithm. This is attributed to the utilisation of the fast local move approach and the inclusion of an additional 'refinement' phrase in all the phrases employed by the Louvain algorithm for identifying the most optimal partition of communities. Roghani and Bouyer (2022) and Anuar and Abas (2021) also identified the Leiden algorithm as the most recent, stable, and efficient method for detecting node communities. Pankratz et al. (2023) suggests that Leiden exhibits lower volatility. Christensen (2022) discovered that the Leiden algorithm excels in achieving modularity objectives, while the Louvain algorithm is more accurate in accomplishing grid search objectives. This suggests that both algorithms perform equally well.

2.2.4.3 Itemset Assessment Method

After obtaining the frequently purchased itemset, a product evaluation would be conducted within the community to uncover associations between them. Evaluation measures offer insights into the practical significance of each community in the application of itemsets. Various methods exist for assessing communities in which each offers valuable insights into various aspects of customers' purchasing behaviour.

Community density is a widely used metric in research to evaluate the relationships between nodes within a community (Qi et al., 2013; Kafkas, 2021; Kim et al., 2012; Raeder and Chawla, 2011; Faridizadeh et al., 2018). The stronger the connection between nodes, the higher the likelihood that a customer who purchases one item will also purchase another. Therefore, stores can achieve better results by strategically integrating these connections into their consideration for strategies. Hence, it is essential to determine the priority of product sets for implementation.

Additionally, centrality measures the degree to which a product influences the purchase of other products. Newman (2008) compares the number of neighbours within a community for each item. Further investigation is needed to explore the correlation between products within densely connected communities (Faridizadeh et al., 2018; Kim et al., 2012). In practical terms, the method can determine the product with the highest influence on other products, enabling stores to allocate their marketing budgets more efficiently and avoid unnecessary expenses on less influential products.

Moreover, betweenness centrality is employed to measure the frequency of a specific product's inclusion in the shortest path between pairs of other products (Ding et al., 2018). It identifies a product that serves as a connection between different other products. By understanding this concept, it is possible to effectively target more products using this bridge. Ding et al. (2018) examine betweenness centrality in identifying products that are frequently co-purchased and serve as a link to the purchase of other products. Kafkas (2021) identified 17 influential bridge products using centrality measures. This shows the utility of betweenness centrality as a network analysis measure.

Furthermore, eigenvector centrality takes into account the connections of the entity in question. Nodes are assigned scores based on the weights of their edges, which include both the product itself and its neighbouring nodes. Higher weights correspond to higher scores. Newman (2008) proposed this centrality measure that regards the score as a measure of relationship quality. Kafkas (2021) used eigenvector centrality to identify the focal point in their network and prioritise the product for marketing implementation. This demonstrated the effectiveness of the measure.

2.3 Literature Reviews Summary

In summary, there is no consensus on the most effective strategies for MBA. ARM is commonly used in MBA research, but there are limitations that can affect the analysis and interpretation. Furthermore, it has been observed that NLP and CA exhibit subpar performance compared to alternative models. DL and NA models demonstrate superior performance compared to other models. DL models are considered more complex than other models. As a result, the research advanced its focus towards NA and conducted a systematic review of relevant literature. Undirected weighted product networks are commonly used in conjunction with community extraction algorithms to identify frequently purchased itemsets for MBA.

3. METHODOLOGY

Based on a comprehensive literature review, it is evident that unsupervised machine learning techniques outperform association rules mining in terms of effectiveness and efficiency for MBA. Certain machine learning models, specifically DL and NA, have demonstrated superior performance in comparison to other models (Hruschka, 2021; Faridizadeh et al., 2018). This study will utilise NA to analyse the MBA. The rationale for this research is in its objective to generate recommendations for operational and marketing strategies in a retail store. To achieve this, a transparent and comprehensible algorithm is necessary. NA employs visual depictions to portray the relationships among nodes, with each node symbolising a product in this investigation. This characteristic makes NA the most suitable approach for this research. Numerous studies endorse the use of NA as a valuable approach in MBA. Faridizadeh et al. (2018) suggest that product networks are a more practical and informative way to represent frequently purchased itemsets than ARM. Moreover, the analyses of Faridizadeh et al. (2018) and this study's objective aligns therefore the primary basis for this methodology would draw inspiration from those sources.

Reader and Chawla (2011) and other studies suggest that employing a product network is a viable method for conducting MBA. This approach represents items as nodes in a network, with edges indicating co-purchases in a single transaction. The study will use an undirected network with links between nodes. This is attributed to insufficient information regarding the sequence in which customers arrange items in their baskets. Even if this information were available, the reasons for customers' choices may not be meaningful, as the frequent purchase of two products together does not necessarily indicate a cause-and-effect relationship. The weight assigned to edges in this study represents the frequency of co-occurrence between two products in transactions, indicating their association strength (Reader & Chawla, 2011; Videla-Cavieres & Ríos, 2014; Faridizadeh et al., 2018; Valle et al., 2018; Li et al., 2008). There is a positive correlation between edge weight and relationship strength. The research suggests that product networks have higher density than social networks or phone networks. This is because product networks connect two products solely based on their presence in the same transaction (Reader & Chawla, 2011). The interpretation of this dissertation will acknowledge that the connection between two items does not always imply a definitive affiliation or tie between them (Faridizadeh et al., 2018).

It is necessary to prune the product network to eliminate potential effects of random customer purchasing behaviour and establish a meaningful network. This step is crucial before proceeding to the community extraction stage. Existing literature recommends the implementation of minimum criteria to filter out product combinations that occur fewer than a specified number of times. This filtering process has been shown to enhance the quality of the extracted product communities (Faridizadeh et al., 2018; Kim et al., 2012; Wu et al., 2021; Videla-Cavieres and Ríos, 2014; Valle et al., 2018; Raeder and Chawla, 2011). However, as outlined in Section 2.2.4.1, there is a lack of established guidelines for determining the criterion for network pruning. Instead, it is recommended that researchers exercise their own judgement in determining the suitable criteria. Consequently, this study will establish the minimum pruning threshold by carefully examining the real dataset, so ensuring the rigour and appropriateness of the threshold for this research. In their respective studies, Kim et al. (2012) and Wu et al. (2021) employed support value as the criterion for pruning and achieved favourable outcomes. Consequently, this dissertation will adopt support value as the criterion for network pruning. The pruned network is now ready to move on to the next step which is frequently purchased itemsets extraction.

This research will adopt community detection method to find the subgraphs of the pruned product network. This is to find the products that are ‘associated’ with each other regarding having high probability to be purchased by customers in the same transaction. As this study aims to extract groups of products that are strongly linked to each other, community detection algorithms that have maximum modularity as the criteria would be adopted, similar to those of Faridizadeh et al. (2018). This is because modularity is a measure of the strength of the community of nodes and is suggested to be the measure of 'quality' of the communities extracted (Newman, 2006). It is calculated as follows:

$$Q = \frac{1}{4m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})s_i s_j \quad (1)$$

, where A_{ij} is the number of edges measured from an adjacency matrix, $\frac{k_i k_j}{2m}$ is the expected number of edges randomly assigned between each node ki and kj within the community, m represents the number of edges in the graph, and $s_i s_j$ is the division of network i.e., let $s_i = 1$ if vertex i belongs to community 1 and $s_i = 1$ if vertex j belongs to community 2.

Several algorithms, including CNM, Louvain, and Leiden algorithm, have been developed with the objective of partitioning nodes into subgroups to maximise modularity values. All the algorithms share a common theoretical foundation, namely hierarchical clustering, and have been widely utilised in research for extracting frequent co-purchased product sets (Woma and Ngo, 2019; Faridizadeh et al., 2018; Raeder and Chawla, 2011; Wu et al., 2021; Videla-Cavieres and Ríos, 2014). In recent years, numerous scholarly works have highlighted the significant advancements and superior performance achieved by the Leiden algorithm. However, several literature sources continue to employ the Louvain algorithms and CNM. Hence, this research will employ all the algorithms in order to compare and determine the optimal extraction technique.

Finally, the obtained communities will be regarded as commonly purchased itemsets and afterwards subjected to analysis utilising diverse methodologies. This dissertation aims to assess the dynamics within different product groups by employing various centrality measures, namely community density, degree centrality, betweenness centrality, eigenvector centrality, and closeness centrality. The first four metrics are commonly utilised in numerous academic studies to analyse itemsets. However, the metric of "closeness centrality" has received limited attention in MBA research (Faridizadeh et al., 2018; Qi et al., 2013; Kafkas, 2021; Kim et al., 2012; Raeder and Chawla, 2011; Ding et al., 2018; Newman, 2008). The concept of 'closeness centrality' is commonly employed in network analysis research and is considered to provide valuable insights into itemsets. Therefore, this study will incorporate this matrix, considering relevant studies conducted by Das et al. (2018) and Zhang and Luo (2017). The itemsets will subsequently be placed for business implementations and the formulation of strategic recommendations.

4. MODELLING

This section implements the methods outlined on the acquired dataset to observe the outcomes for subsequent analysis. Section 4.1 explores the procedures involved in data preparation, encompassing the activities of data extraction and data engineering. Section 4.2 delves into the concept of product network modelling, providing a detailed explanation and analysis. On the other hand, Section 4.3 focuses on community detection, specifically the extraction of commonly purchased itemsets. The subsequent part, Section 4.4, investigates the retrieved itemsets. Lastly, Section 4.5 accounts for possible limitations for this research regarding methodology.

4.1 Data Preparation

This subsection covers the description of the dataset focussed by this dissertation and detailed steps of data pre-processing for product network.

4.1.1 Data Extraction

This dissertation aims to conduct an MBA which involves extracting records of co-purchased items from each transaction; hence the transaction data of a retail store are needed. The transactional dataset this dissertation concentrate on is a publicly available dataset and were obtained from the data science online community platform "Kaggle". The dataset contains 522,064 rows and 7 columns. Since the dataset contains multiple variables, some of which are not relevant to this study, only those necessary for constructing a product network and performing MBA were extracted. This includes *'BillNo'* and *'Itemname'*, that contains the transaction number and products name included in those transactions. After scrutinising the data there are missing values of 1,455 entries for item names. This is not significantly large in relative to the total number of data we have gotten, and product names are necessary for performing MBA, hence rows containing those missing values were dropped. This leaves 20,210 transactions with 4,185 unique products in the dataframe ready for further pre-processing.

4.1.2 Data Engineering for Modelling

To build the product-to-product network with number of transactions being the edges, one needs to first generate bipartite network to capture and obtain co-appearance of different products in the same transaction and measure it (Faridizadeh et al., 2018). The network was built by using Python’s package called *networkx*. Bipartite of nodes would be *‘products’* and *‘transactions’*. Product nodes and transaction nodes are then separated to form a biadjacency matrix, shown in Figure 1, that has transactions as indices and each unique product as columns. The values containing inside the matrix would be a positive real number value denoting whether and how many times the product appears in a transaction. This would then be used to product the adjacency matrix to form a full product network later.

	0	1	2	3	4	5	...	4179	4180	4181	4182	4183	4184
0	1	1	1	1	1	1	...	0	0	0	0	0	0
1	0	0	0	0	0	0	...	0	0	0	0	0	0
2	0	0	0	0	0	0	...	0	0	0	0	0	0
3	0	0	0	0	0	0	...	0	0	0	0	0	0
4	0	0	0	0	0	0	...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
20205	0	0	0	0	0	0	...	0	0	0	0	0	0
20206	0	0	0	0	0	0	...	0	0	0	0	0	0
20207	0	0	0	0	0	0	...	0	0	0	0	0	0
20208	0	0	0	0	0	0	...	0	0	0	0	0	0
20209	0	0	0	0	0	0	...	0	0	0	0	0	0

Figure 1: Head and Tail of the Biadjacency Matrix

The adjacency matrix is widely utilised as a means of data storage for converting into a weighted network, as it allows for the representation of pairwise weights (Videla-Cavieres and Ríos, 2014; Ding et al., 2018; Gao, 2023). Before proceeding to generate the adjacency matrix and generate product-to-product network, the biadjacency matrix is pruned. Steps are as follows:

I. All the transactions that only contain one product are dropped as it would not give any information about products associations.
II. Set minimum support threshold for the itemsets and filter those that do not meet this threshold out.

It is revealed that the *‘support’* value for all the itemsets is relatively low, hence the minimum criteria in this dissertation would be relatively low. Data shows that there are 17,743 itemsets that appeared only less than three times in the record and all the sets has support value of less than 0.00015. Therefore, for this study, minimum support threshold of 0.00015 is set to filter out itemsets that appears by chance or in random and may create spurious edges.

From this, the adjacency matrix can be generated by multiplying the pruned biadjacency matrix with its transpose. The resulting adjacency matrix represents the number of times each product appears in the same transaction as one another. The diagonal would represent total number of times a product appears in any transactions, hence, to allow for the study of co-purchased products and avoid self-loop of a node, the diagonal would need to be zero. After that all the rows and columns that only contain zero would be filtered out as it denotes that the product does not joint appear with any other products thus, they do not align with the aim of this research. At last, there are 163 products left in our adjacency matrix. Figure 2 displays the first and last 5 rows of this matrix. Note that the number appears in the indices and columns represents the identification number of a unique product and do not represent total counts of products in the matrix. The identification number is set in place instead of the actual item name because this dataset contains products that have a relatively longer names, hence, to facilitate a clear and informative product network numbers are used instead. Faridizadeh et al. (2018) labelled their product nodes in the network in the similar way. The data is now ready to be turned into product-to-product network.

	0	1	2	3	4	5	...	3743	3851	3854	3885	4158	4159
0	0	3	3	3	3	3	...	0	0	0	0	0	0
1	3	0	3	3	3	3	...	0	0	0	0	0	0
2	3	3	0	3	3	3	...	0	0	0	0	0	0
3	3	3	3	0	3	3	...	0	0	0	0	0	0
4	3	3	3	3	0	3	...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
3851	0	0	0	0	0	0	...	0	0	4	4	0	0
3854	0	0	0	0	0	0	...	0	4	0	4	0	0
3885	0	0	0	0	0	0	...	0	4	4	0	0	0
4158	0	0	0	0	0	0	...	0	0	0	0	0	4
4159	0	0	0	0	0	0	...	0	0	0	0	4	0

Figure 2: Adjacency Matrix

4.2 Product Network

This subsection focuses on the first half of the modelling, that is the construction of full product network. The network would be a product-to-product network. This is to establish the relationships between each product and observe the overall characteristics of products sold by this retailer.

After achieving the pruned adjacency matrix, the matrix is input into *networkx* again to generate product-to-product network. The network has 163 nodes and 1,768 undirected edges. The full product-to-product network can be seen in Figure 3. The links to nodes proportion seems to be considerably medium at 1 node per 10.85 edges. It can be observed that there are multiple isolated product groups of various sizes in which vertices are seemed to be highly connected. It is evidenced by the low density the network has at only 0.134, meaning large percentage of nodes are not linked with other nodes in the network but only selected few of its neighbours. This can denote that these cluster of nodes are strongly linked signalling that they are highly related to each other and not those of different groups. It can be speculated that these isolated groups of nodes would form a separate itemsets, while it cannot be concluded with confidence how many communities can be detected. Moreover, it can be observed that some groups hold products that are located closer together than some other groups. This may be speculated that the products are stronger associated. This, though, needs to be verified after obtaining actual communities of products.

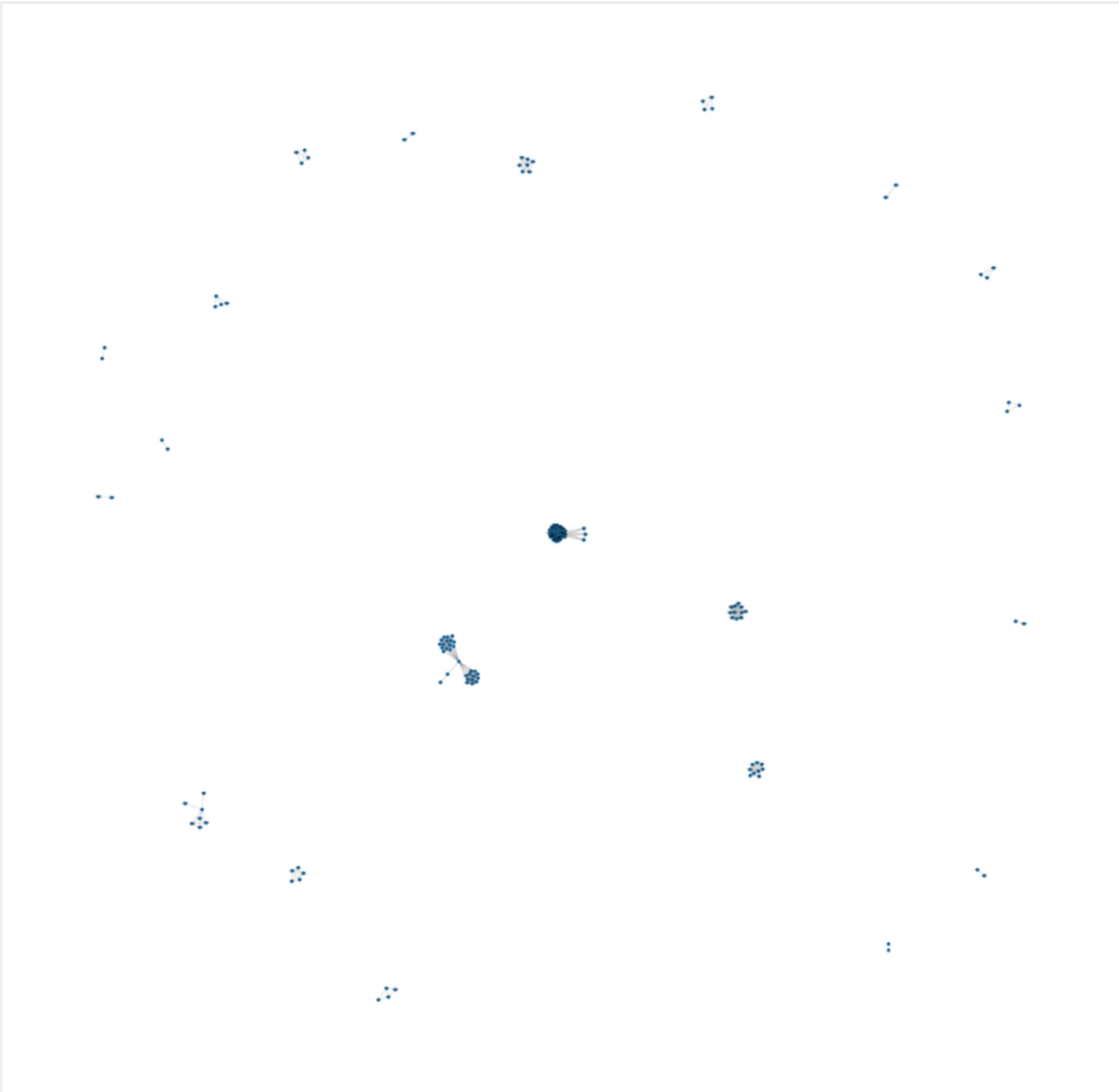

Figure 3: Product Network

Degree distribution of this product network is skewed and heavy-tailed, as illustrated in Figure 4, this is similar to Raeder and Chawla (2011). Degree of a node represents the number of neighbours the node has. The distribution signifies that there is a huge number of nodes that only have a few neighbours, though there are a few chunks of vertices that do have high number of links to its neighbours. This further insinuates that there may be several smaller itemsets detected. We could speculate to achieve this result when we adopt community detection algorithm. This conforms to the nature of MBA's product network for a retail store as customers would purchase smaller quantity of product at the time. This nature of the products and transactions may make the recommendation of customers' tailored business strategies more complex because there would be multiple groups of products that we identify as highly related, and it may not be clear which one the business should prioritise. Though the prioritisation of itemsets may be more complex if there are a greater number of sets, it is not impossible as each community could be analysed in detailed and compare. Section 4.4 will elaborate more on this point.

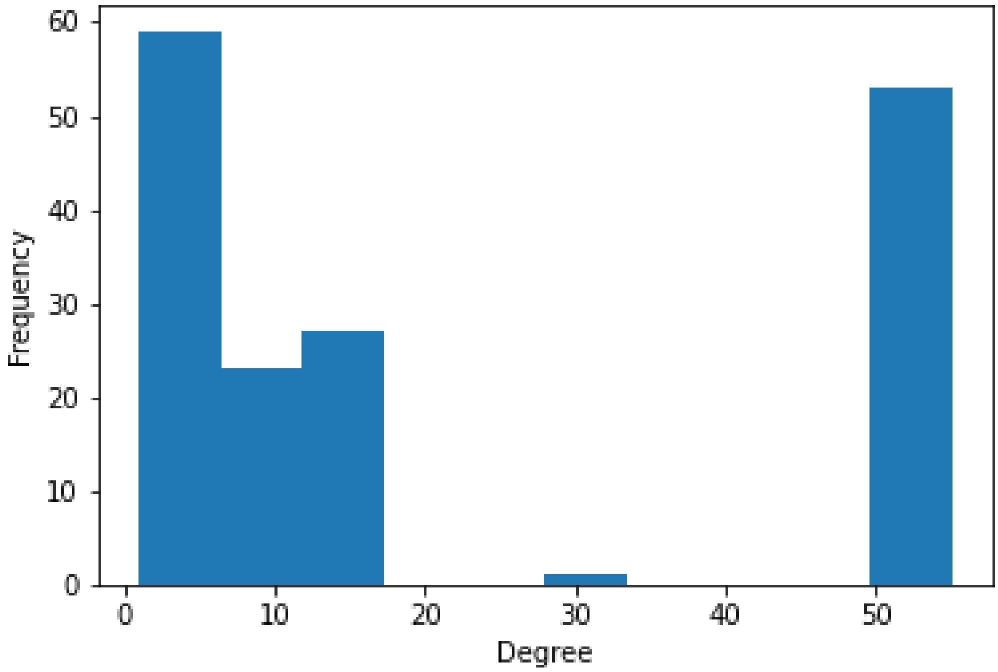

Figure 4: Degree Distribution of the Product Network

All in all, it is clear from the network that there are multiple clusters of nodes strongly linked together and appears close to each other. This can lead us to hypothesise that those clusters of nodes are product sets that have strong co-purchasing associations with each other. This may be rigorously investigated by applying community detection algorithms to differentiate between each highly related itemsets.

4.3 Community Detection

This subsection covers the other half of the modelling process, the itemsets extraction. This could be considered as the most important step of the whole project as here is where our perception of customers' purchasing traits will be based on.

Three community detection algorithms that uses modularity maximisation in separating nodes into communities are adopted in this dissertation. This includes CNM, Louvain, and Leiden algorithm. All the algorithms are applied on the same product network that we constructed in Section 4.2. The comparison is controlled by setting the same random seeds for those algorithms that requires random seed to execute, including Louvain and Leiden, setting the seed to be 4500.

Regarding the number of nodes in each community produced by each algorithm, the number of nodes in each community are moderately skewed with about 3 communities having more than 10 nodes, while the rest are seen to have up to 10 nodes. Many communities can be observed to have only 2 to 3 products in it. As discussed in Section 2.2.4.2 and Section 3, CNM and Louvain algorithm have almost identical subgroups of nodes that they detected as a strongly linked. Leiden, on the other hand, outputs some of similar communities but also detected a relatively more skewed sized communities and some in which do not appear in the other two algorithms' outputs. This is because Leiden aims to only produce communities that are fully connected. However, in some cases this is not necessarily essential to achieve especially in the application of MBA where the main purpose is to study what customers buy at the same time. If one forces certain products to be connected to the other to generate a fully connected nodes, the interpretation of consumer purchasing behaviour may not represent the true trait.

Results show in the comparison of number of communities detected that Louvain algorithm produced 22 communities, while CNM found 21 and Leiden extracted 22, which are very similar. However, considering that Leiden produced communities that are far more skewed and are a bit different from those of CNM and Louvain, this study concludes to not base the analysis on communities produced by Leiden algorithm.

Moreover, concerning the nodes within each community, Louvain and CNM are found to have a very similar nodes appearing in the same communities. For instance, community 9, community 19, and community 2. This example can be shown in Figure 5 below.

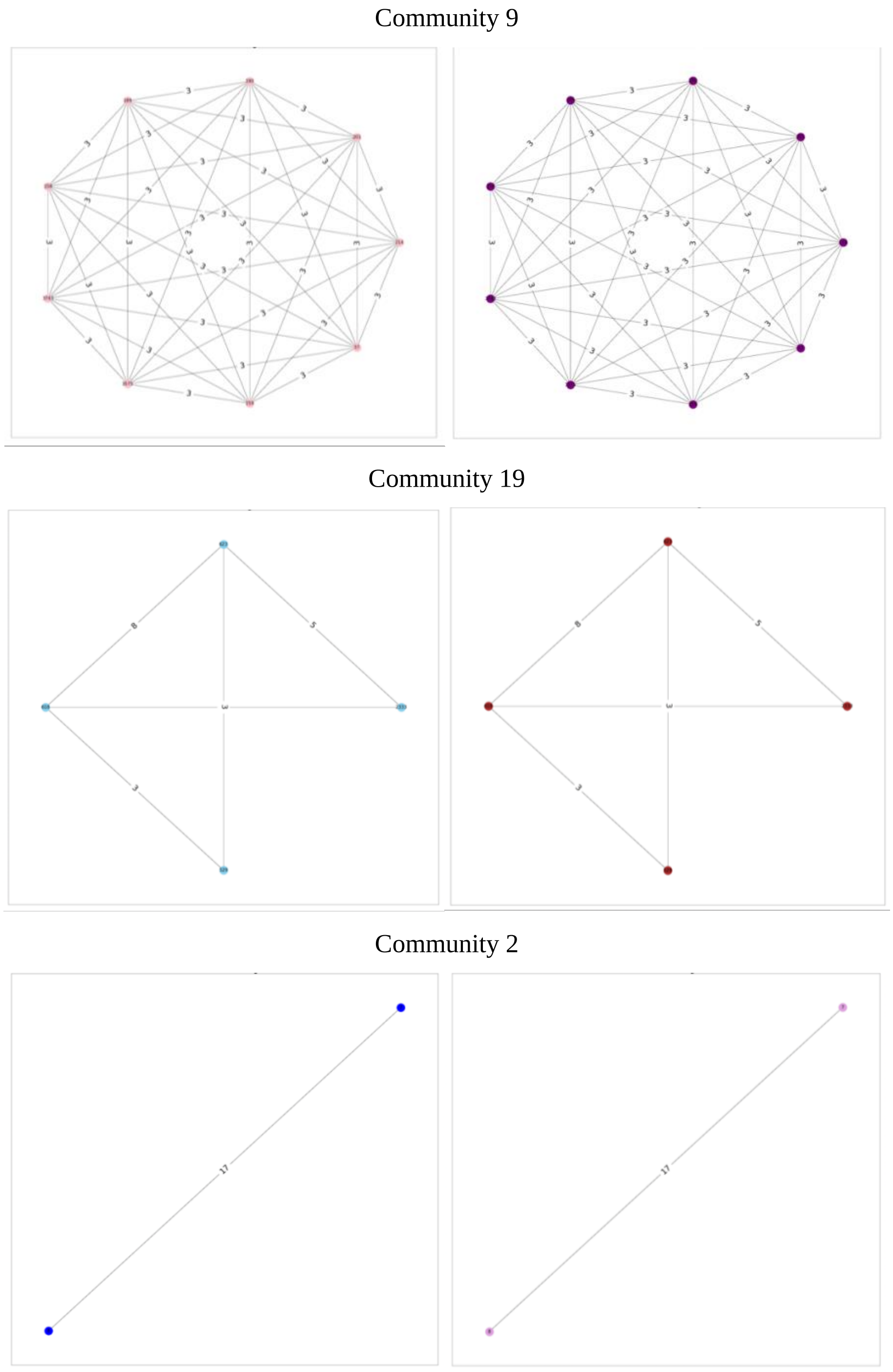

Figure 5: Example of Similar Communities Extracted by Louvain Algorithm (Left side) and CNM (Right side)

As the two algorithms are seen to have a very similar communities extracted, both in terms of number of communities in the network and nodes in the communities. Therefore, it can be deduced that performance of all the two algorithms is on par. Hence, to finally decide which of these algorithms is most appropriate for this research, we look more on the other factors. We concluded that Louvain algorithm is more appropriate to apply to this research because more of the recent research recommended them instead of CNM. Moreover, we test for stability of the model for Louvain algorithm, that requires random seeds to execute by running the algorithm to observe the output. Louvain algorithm produced 22 communities 10 times out of 10 times. The algorithm detected communities with 100% stability level which is considered very stable. Therefore, this study will base the further analysis on communities produced by Louvain algorithm.

The overall communities detected by Louvain algorithm in the full network and in separated can be found in Figure 6 and Figure 7 below. As portrayed in the network, the isolated groups of nodes are detected as a community which conforms with the speculation made in Section 4.2.

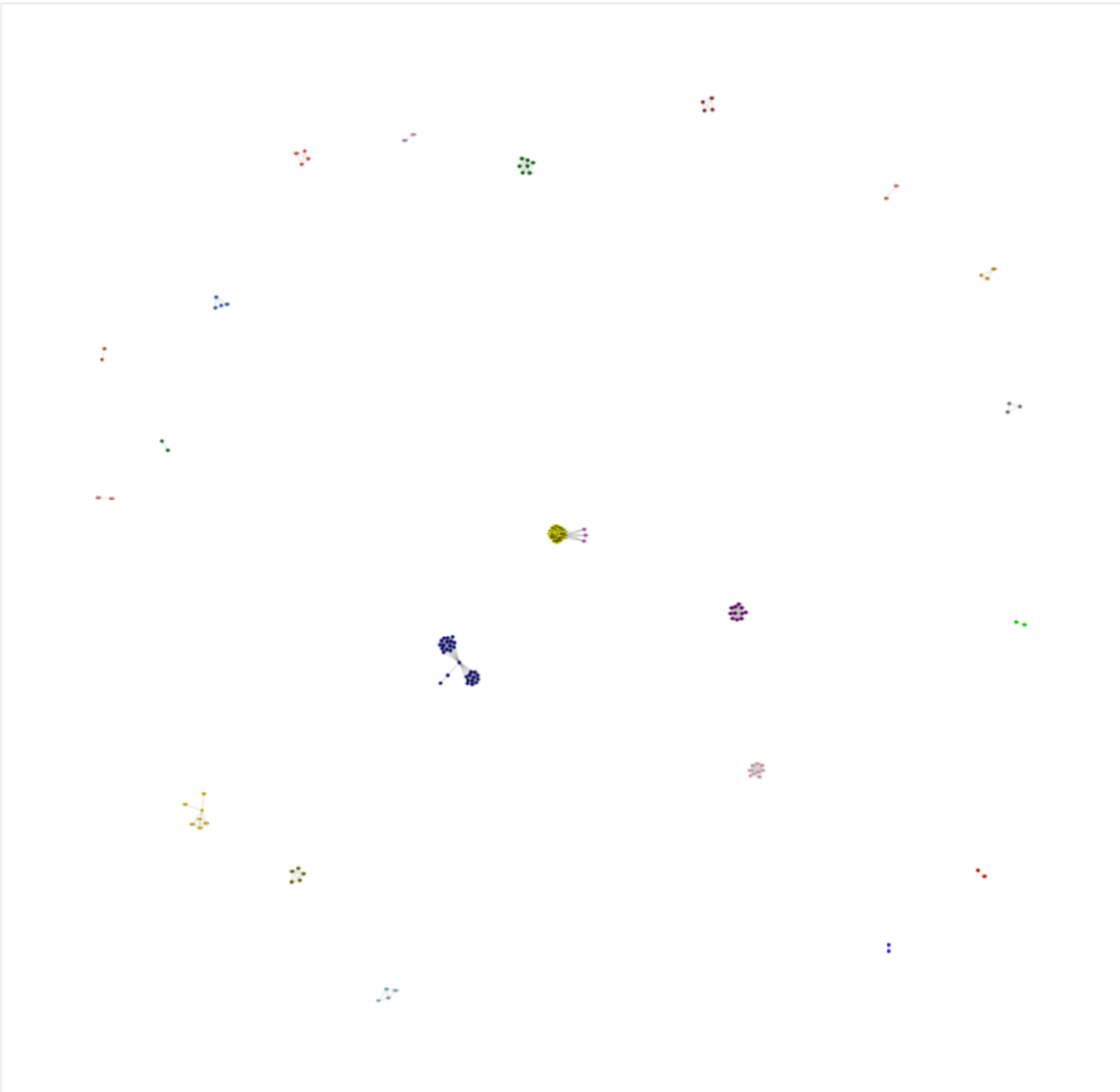

Figure 6: Full Product Network with Coloured Communities

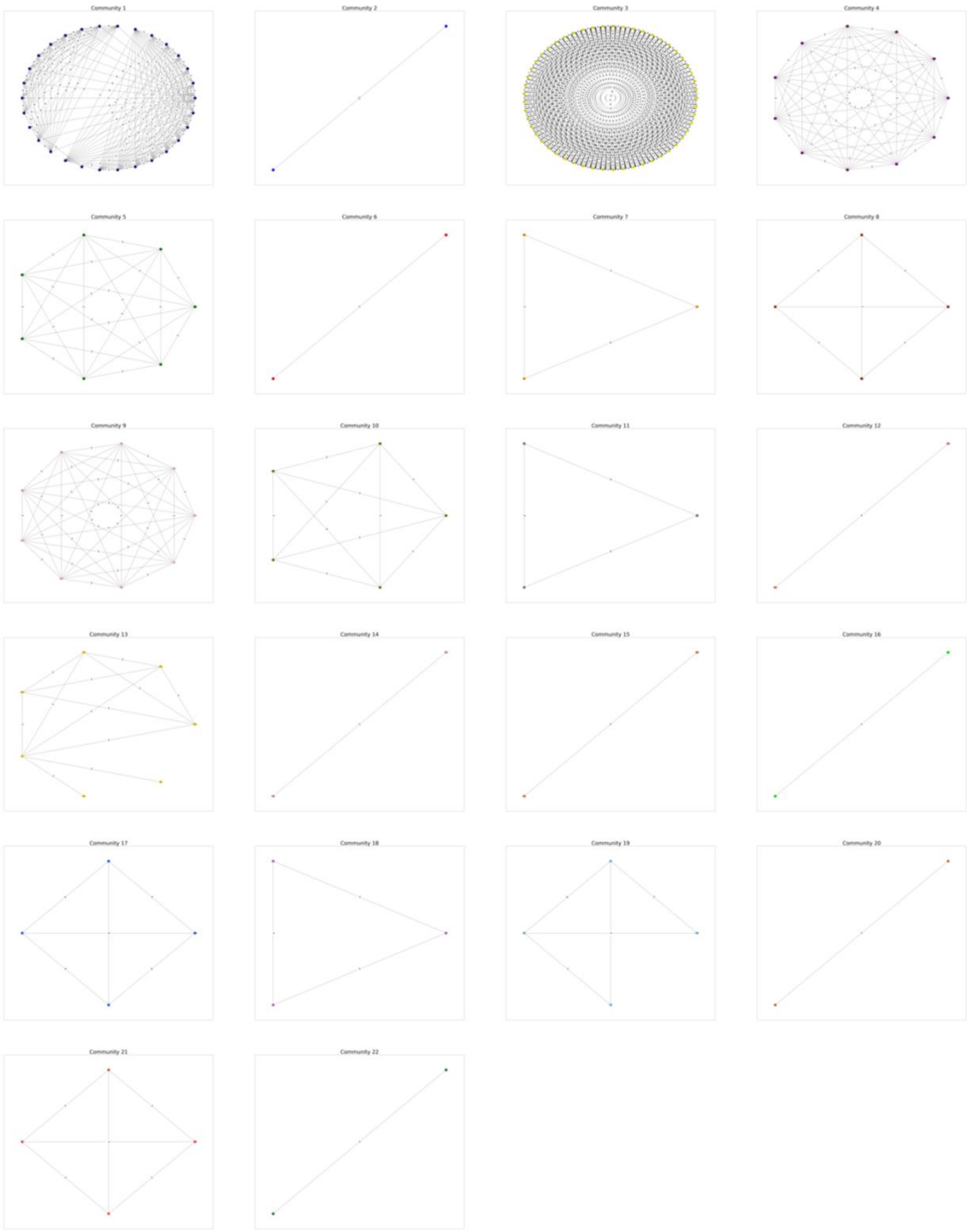

Figure 7: All Communities Detected by Louvain Algorithm

4.4 Within Community Analysis

This subsection focuses on analysing the extracted communities, which is considered in this research to be the frequently purchased itemsets for the MBA. Characteristics of each communities needs to be investigated to uncover customers' purchasing behaviour and later implement these insights to practical business strategies. The following analysis will be both in the community level and the nodes level. Also, the full product name and their corresponding analyses can be found in Appendix 2.

Firstly, density of each community is calculated and plotted in Figure 8. As seen from the diagram, most communities have density of one which means all the nodes are connected to one another. Though, there are a couple of communities that has a relatively low density. This is clearly seen in community 1 where it only has 0.462 density value, community 13 which has a density of 0.571, and community 19 that has a density of 0.833. Though several nodes are seen to not be connected to other nodes in these two communities, the connection between a product to another is relatively high in magnitude, forming a strong community.

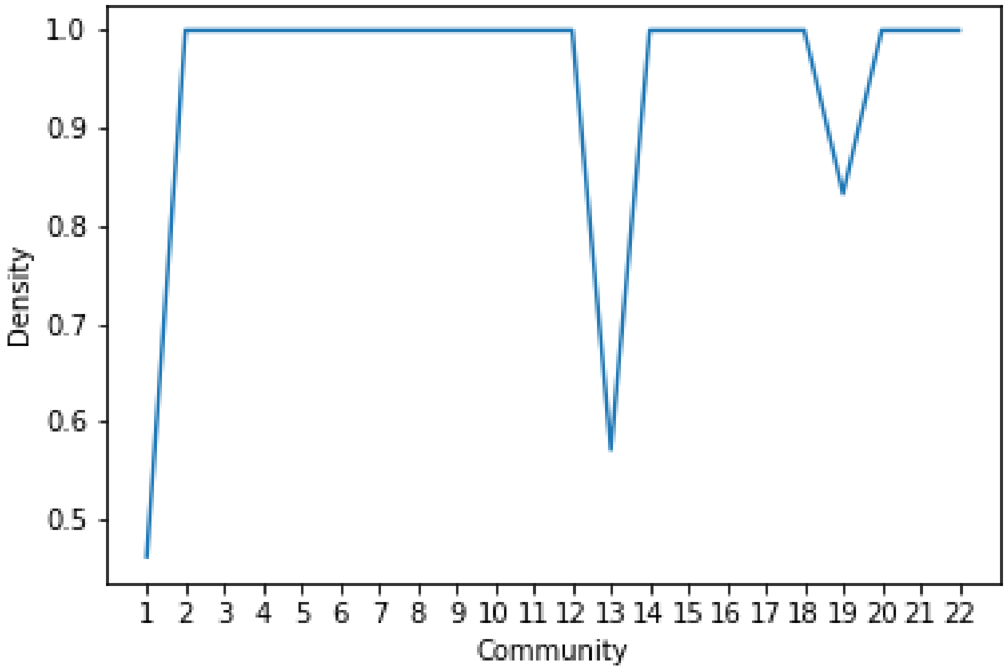

Figure 8: Density of Each Community

Regarding degree centrality, it is observed that all the communities that has density of 1 hold nodes that got degree centrality of 1. This is because all the nodes are connected to each other, hence there is one node that has a stronger connectivity degree than others. On the other hand, community 1, 13 and 19 that has density of less than 1 reveals one or more nodes that have a higher degree centrality than other nodes in that community. For instance, product number 623 and 618 of community 19 that have degree centrality of 1, while the other two products in the same community have degree centrality of 0.667. It can also be observed that communities that has higher density tends to contain nodes that has higher degree centrality for instance those of community 19 compared with community 13 which has only one product that has degree centrality of 1, four products that have degree centrality of 0.667, and the other two nodes has degree centrality of 0.167. It can be noticed that the correlation between community density and degree centrality of nodes inside a community is positive. This could be speculated that the higher the density of a community, there are more products that is considered “important” in that community. To elaborate, those “important” product is linked to multiple other products in a community hence when customer buys this product there are higher probability that they will purchase other items within that community too.

Similar to degree centrality, those community that has a density of 1 do not appears to have any nodes that have a positive betweenness centrality. This means there are no specific nodes that could act as an “important” middle person to the purchase of other nodes within that community. Whereas community 1, 13, 19 has at least one product that has a betweenness centrality value of greater than 0. Findings can be found in Table 2 below.

Table 2: Betweenness Centrality

Community Number	Product Number	Betweenness Centrality
1	0	0.576
13	45	0.6
19	623	0.167
	618	0.167

To increase rigorousness of the analysis, closeness centrality is checked. Results shows that those products that appears to have high degree centrality also have high closeness centrality values. This is because it measures the average distance of a node to other nodes so nodes that are connected directly to other nodes, the distance from one to other nodes will be shorter. This makes those communities that are fully connected and has density of 1 would have closeness centrality of 1 for all nodes.

Moreover, the test for eigenvector centrality reveals that those nodes that are fully connected in a community share the same value of eigenvector centrality. This is because the transitive influence of each node is the same as everyone is connected to each other in the same level. Moreover, in the communities that has density of less than 1, nodes that has a degree centrality of 1 and positive betweenness centrality has a stronger transitivity power from one node to another indicating that if customer buys that product first, it is more likely that they will be influenced to buy other products in that community more than when they buy products with low eigenvector centrality value.

4.5 Possible Limitations

Though, many literatures have recommended product network for performing MBA, there are a couple of points that needed to be noted. First, there is no set framework in the minimum criteria for pruning the network of products. This needs to be thought out by the analyst case by case. Though Raeder and Chawla (2011) had tried to establish a framework for choosing the minimum support parameter explained in Section 2.2.4.1, it still requires analyst's personal judgement to decide on the threshold anyway. Therefore, the itemset extracted from the network and the following analysis may be misled if the threshold of pruning is not chosen appropriately. However, this problem could be mitigated via the rigorous reviews of other previous literatures' experiments to identify the best way to assess the interested dataset.

Secondly, the interpretation of the products associations to each other may be unclear because links of the two products may be established only due to the two being in the same transaction more than x times. There is no clear explanation of why one would purchase the two together. Therefore, implementation of this information may require experts in behavioural science to put in their qualitative perceptions before the store starts building their marketing campaigns. However, this potential problem does not only appear in product network, but all the methods aim to tackle MBA would also face these difficulties as the only information input into the analysis is the transactions records of customers and nothing else.

5. RESULT INTERPRETATION

From the observation of the communities extracted, itemsets can be separated into six categories as follows, very strongly linked products, moderately linked products, minimum threshold level linked products, seasonally related products, asymmetric links, and categorically related products, in which these six types of itemsets can be referred on to aid the construction of operational and marketing strategies.

There are a couple of insights that can be of use to the operational strategies. Firstly, communities of products that contains products of same category may lead the store to understand the purchasing habits of customers within the same section of the store more. From this, the store may reorganise the shelving to attract the most sales out of customers roaming that area of the store as once the customer came in contact with one of the various associated product they would have a motivation to purchase the other related products, hence if the store makes those items available immediately when customers thought of buying it, it is more likely that they will choose to place them in their baskets. Regarding this study's dataset it is community 2 and community 22 that are of our interest, see Appendix 2. It would be recommended that the store displays *red polka dot* and *union jack style hand warmer* next to each other in the hand warmer shelves and display *white hanging heart t-light holder* and *wooden picture frame white finish* next to each other in the small desk decoration section of the store.

Secondly for strategies regarding shelves organisations, itemsets exhibits seasonal related products may be of use for the popup arrangements in the corresponding seasons, events, or festivals. For instance, community 1 that contains outdoor living facilitated products, community 3 that holds kids birthday parties related products, and community 17, 4 and 13 that shows various Christmas related products. This may be appealing for customers who have or may not have an initial intention to purchase these seasonal products.

In the case where the store has an online shopping platform, the store may make use of those itemsets that are seen to have passed the minimum support criteria but are not yet considered moderate to high associations between them to generate a recommendation system once they place one item in their online basket. For this study, community 5, 8, 9, 10, 11, 12, 14, 15, 18, 20, 21. These itemsets contain products that may or may not all be in the same category and the links may not be as high but still exhibits some motivation for the customers to purchase the other related products. Providing them with “recommended products” may increase their motivation to buy multiple products together in the same transaction, encouraging higher sales and profitability of the store.

In addition, like those of shelves optimisation, the store may establish a special popup pages or sidebands for seasonally related products during those corresponding seasons, events, or festivals. This may appeal to customer and trigger their purchasing decisions even though they may not have an initial aim to buy those products.

For itemsets that are highly associated for instance in community 2, where *hand warmer red polka dot* and *hand warmer union jack* are purchased together for as high as 17 times. This exhibits a solid customer purchasing behaviour for these two products that if they buy one of the two it is very likely that they will buy another in the same basket. Hence, if the store observes a low in stock of *hand warmer red polka dot*, it is recommended that the store check stocks of *hand warmer union jack* too. It is ideal to keep the stock of the products to be sufficient for both the purchase of each product on its own and the possibility of them purchasing together. This is because if a customer purchase *hand warmer red polka dot* and would like to purchase *hand warmer union jack* afterward but cannot be due to the latter being out of stock the store would have an opportunity cost of not selling the product.

Moreover, information regarding frequently purchased itemset may be used as a guideline for the marketing team to construct marketing campaigns. Both the seasonally related itemsets, moderately linked products, and itemsets that have few products highly related but also some lighter weighted links may be perfect for establishing the campaign. Firms may form a product bundle campaign for the former type of product sets and sell during the corresponding seasons or special events, for instance community 4 which holds various Christmas theme related products such as toys, Christmas lights, cards, and snack boxes, may allow firms to sell more of this bundle during Christmas celebration period. This is because it is seen from the result that products within these itemsets are considered associated and frequently bought together, hence there already exists the motivation to buy these products bundle and if the store further push the attractiveness to buy of these products by letting them feel of value for money, there would be higher probability that customers would purchase the bundle.

Similarly, regarding purchasing motivation of customers, itemsets that contains moderately linked products such as community 6 and 16, that have links of weight 5, may be advertised together as a product package to appeal to customers for being value for money in which would increase their motivation to buy the bundle. If the store could sell the product bundle of two, for instance, it means that stores may sell two products instead of one stand-alone product, hence increasing quantity of goods sold.

Also, the itemsets that have highly related products mixed with lower related products may aid the store to create a product bundle of few products including at least one of the highly linked product and other lower linked products. For this dataset, community 7 and 19 are of interest. To elaborate, consider community 7 in Figure 9, where product *jumbo bag red retrospot (524)* and *jumbo bag strawberry (525)* are strongly linked with the weight of 9 and are weaker linked with the *sixty cake cases dolly girl design (1666)*, the store may create a bundle of *jumbo bag* *red retrospot* and *sixty cake cases dolly girl design* to generate higher motivation for the customer to buy *jumbo bag strawberry* in addition to the bundle after they purchase the bundle. This would enhance stores' sales of products and hence revenue.

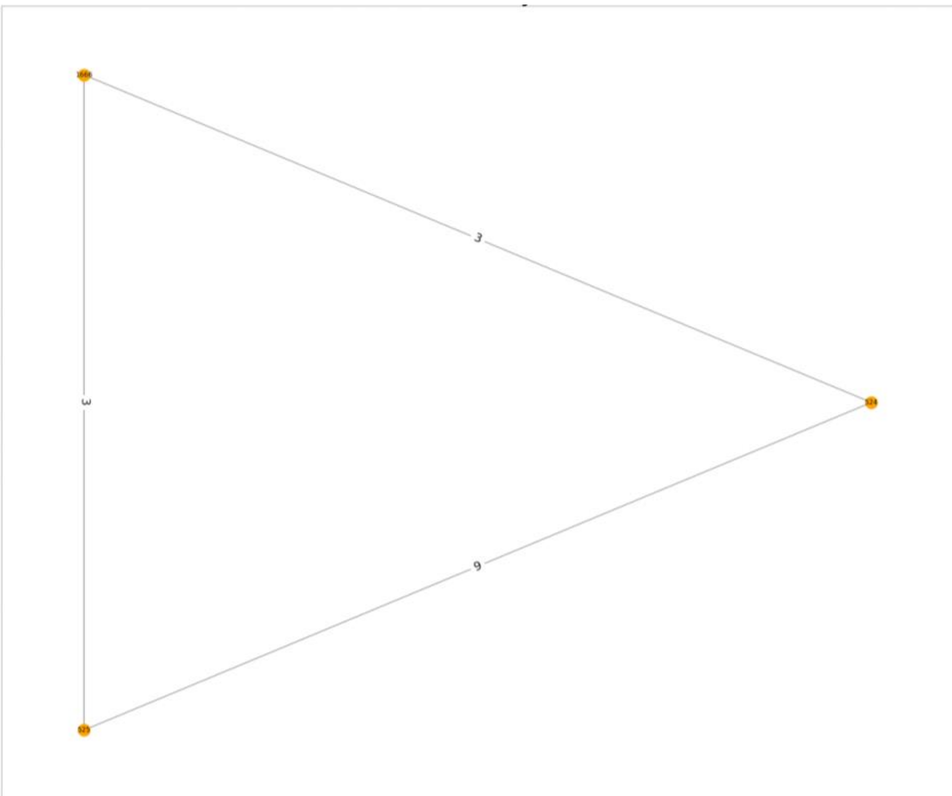

Figure 9: Community 7 Extracted by Louvain Algorithm

Note, it is noticed that a few of these itemsets contains Christmas related products, which makes us reasonably speculated that the dataset may be collected around the period of Christmas festival. This point may impact this analysis interpretation as customers' purchasing behaviour may not be considered "normal" due to the nature of festival shopping period, where some of the itemsets that normally may not have festive related products in, may appear to include some of these items such as *paper chain kit fifty's Christmas* and *heart of wicker* of different types, which appears in many itemsets and that seems unique to other items in the itemset. Moreover, it must be pointed out that customers' purchasing behaviour may be different if looked at in a different period as behaviours can vary with time and trends.

6. CONCLUSION AND FUTURE WORK

This dissertation magnifies into customer behaviour regarding their purchasing of product combinations and recommend businesses strategies for the stores to put in practice to enhance their profitability. From reviewing various related literatures, association rules mining is the widely known method to tackle market basket analysis. However, there are several limitations associated with the aforementioned method. These include its inefficiency in analysing larger datasets, the generation of a high number of rules deemed "interesting," and the failure to identify various significant rules. Hence, the utilisation of unsupervised machine learning as an alternative has gained increasing favour in recent times. This dissertation employs product network analysis as a method of analysis, with a specific emphasis on community extractions. This methodology has been demonstrated to be both effective and efficient and has been suggested by multiple scholarly sources.

Product network is pruned before further analysis is done. For this study, the minimum support threshold of 0.00015 is set to filter out those product sets that appears to be random occurrences. This is to strengthen interpretative analysis afterward. It is found that Louvain community detection algorithm is the most suitable for extracting frequently purchased itemsets for this study. 22 itemsets from 163 products are extracted as a result. Each community are then analysed from different angles both community-level and node-level to observe interesting traits. It can be observed that those communities that are fully connected have no outstanding nodes that may act as a bridge to other products or are more powerful in terms of transitive power than others within the same community. In this study, it has been observed that three communities, which are not entirely interconnected, exhibit a vice versa relationship. Irrespective of the circumstances, both forms of communities may prove beneficial for the business in utilising these itemsets to support their strategic endeavours.

For business related interpretation, communities detected can be categorised into six categories: very strongly linked, moderately linked, minimum threshold linked, asymmetric links, seasonally related, and categorically related products. In our analysis, we have identified four strategies to enhance the store's performance in terms of operations and marketing. These strategies include rearranging shelves to group associated products within the same category, implementing an online shop recommendation system to promote related products, monitoring inventory for highly related products, and conducting marketing bundling price campaigns for communities that exhibit asymmetric links between products and seasonally related items.

This study acknowledges the potential limitations associated with the chosen technique and analysis. Nevertheless, it is advisable to conduct further research in this particular field in order to address and mitigate these potential challenges. One such constraint is the absence of established guidelines for determining the minimum criterion to be employed in network pruning. In future research, it may be beneficial to employ the selection strategy proposed by Reader and Chawla (2011) in order to investigate its potential for enhancing community detection. Alternatively, other methodologies can be employed to extract itemsets, such as the utilisation of Centre-Piece Subgraphs. The aforementioned approach represents one of the methodologies employed to amplify subgraphs inside a network. In their study, Reader and Chawla (2011) discovered that this particular approach has the potential to delve into the intricate connections among nodes by generating a comprehensive list of interconnections between all nodes inside the network, which are subsequently ranked. Another important point to consider is that the existing connections between the two items may not accurately reflect their true relationship. Therefore, in order to accurately understand the retrieved itemsets, it is advisable to consult experts from various business disciplines.

7. BIBLIOGRAPHY

Agrawal, R., Imieliński, T. and Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22(2), pp.207–216. doi:https://doi.org/10.1145/170036.170072.

Agrawal, R, Mannila, H, Srikant, R, Toivonen, H & Verkamo, AI (1996), Fast discovery of association rules. in UM Fayyad, G Piatetsky-Shapiro, P Smyth & R Uthurusamy (eds), Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park, CA, pp. 307-328.

AL, M.R., Sembiring, M.T. and Maulana, R.G.R. (2022). Product Recommendations Using Market Basket Analysis with FP-Growth and Clustering Techniques . Proceedings of the First Australian International Conference on Industrial Engineering and Operations Management.

Alawadh, M.M. and Barnawi, A.M., (2021). A Survey on Methods and Applications of Intelligent Market Basket Analysis Based on Association Rule. Journal on Big Data, 4(1).

Annie, L.C.M. and Kumar, A.D., (2012). Market basket analysis for a supermarket based on frequent itemset mining. International Journal of Computer Science Issues (IJCSI), 9(5), p.257.

Anshul Bhargav, Robin Prakash Mathur and Munish Bhargav (2014). Market basket analysis using artificial neural network. International Conference for Convergence for Technology-2014. doi:https://doi.org/10.1109/i2ct.2014.7092091.

Anuar, S.H.H., Abas, Z.A., Yunos, N.M., Zaki, N.H.M., Hashim, N.A., Mokhtar, M.F., Asmai, S.A., Abidin, Z.Z. and Nizam, A.F. (2021). Comparison between Louvain and Leiden Algorithm for Network Structure: A Review. Journal of Physics: Conference Series. doi:https://doi.org/10.1088/1742-6596/2129/1/012028.

Bako, A.T., Walter-McCabe, H., Kasthurirathne, S.N., Halverson, P.K. and Vest, J.R., (2020). Reasons for social work referrals in an urban safety-net population: a natural language processing and market basket analysis approach. Journal of Social Service Research, 47(3), pp.414-425.

Basuchowdhuri, P., Shekhawat, M.K. and Saha, S.K. (2014). Analysis of Product Purchase Patterns in a Co-Purchase Network. [online] IEEE Xplore. doi:https://doi.org/10.1109/EAIT.2014.11.

Boztuğ, Y. and Reutterer, T., (2008). A combined approach for segment-specific market basket analysis. European Journal of Operational Research, 187(1), pp.294-312.

Bhargav, A., Mathur, R.P. and Bhargav, M., 2014, April. Market basket analysis using artificial neural network. In International Conference for Convergence for Technology-2014 (pp. 1-6). IEEE.

Christensen, A.P. (2022). Undimensional Community Detection: A Monte Carlo Simulation, Grid Search, and Comparison. Psyarxiv Preprint.

Clauset, A., Newman, M.E.J. and Moore, C. (2004). Finding community structure in very large networks. Physical Review E, 70(6). doi:https://doi.org/10.1103/physreve.70.066111.

Das, K., Samanta, S. and Pal, M. (2018). Study on centrality measures in social networks: a survey. Social Network Analysis and Mining, 8(1). doi:https://doi.org/10.1007/s13278-018-0493-2.

Decker, R. (2005). Market Basket Analysis by Means of a Growing Neural Network. The International Review of Retail, Distribution and Consumer Research, 15(2), pp.151–169. doi:https://doi.org/10.1080/09593960500049332.

Decker, R. and Monien, K. (2003). Market basket analysis with neural gas networks and self-organising maps. Journal of Targeting, Measurement and Analysis for Marketing, 11(4), pp.373–386. doi:https://doi.org/10.1057/palgrave.jt.5740092.

Ding, Z., Hosoya, R. and Kamioka, T. (2018). Co-Purchase Analysis by Hierarchical Network. Pacific Asia Conference on Information Systems.

Elhawary, M. (2023). Council Post: To Build A Business That Lasts, Know Your Customer Journey. [online] Forbes. Available at: https://www.forbes.com/sites/forbesagencycouncil/2023/05/08/to-build-a-business-that-lasts-know-your-customer-journey/ [Accessed 25 Jun. 2023].

Faridizadeh, S., Abdolvand, N. and Rajaee Harandi, S., 2018. Market basket analysis using community detection approach: A real case. Applications of Data Management and Analysis: Case Studies in Social Networks and Beyond, pp.177-198.

Gangurde, R., Kumar, B. and Gore, S.D., 2018. Optimized predictive model using artificial neural network for market basket analysis. International Journal of Computer Science and Communication, 9(1), pp.42-52.

Gao, H. (2023). The Impact of Topological Structure, Product Category, and Online Reviews on Co-Purchase: A Network Perspective. Journal of Theoretical and Applied Electronic Commerce Research, 18(1), pp.548–570. doi:https://doi.org/10.3390/jtaer18010028.

Gatzioura, A. and Sànchez-Marrè, M., 2014. A case-based recommendation approach for market basket data. IEEE Intelligent systems, 30(1), pp.20-27.

Gupta, S. and Mamtora, R. (2014). A survey on association rule mining in market basket analysis. International Journal of Information and Computation Technology, 4(4), pp.409–414. doi:https://doi.org/0974-2239.

Han, J., Kamber, M. and Pei, J. (2012). 6 - Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/B978012381479100006X.

Hermina, C.I., Gopalakrishnan, A.B. and Gopalakrishnan, Dr.B. (2022). Market Basket Analysis for a Supermarket. In: International Journal of Management, Technology And Engineering.

Hruschka, H., 2021. Comparing unsupervised probabilistic machine learning methods for market basket analysis. Review of Managerial Science, 15(2), pp.497-527.

Kafkas, K., Perdahçı, Z.N. and Aydın, M.N. (2021). Discovering Customer Purchase Patterns in Product Communities: An Empirical Study on Co-Purchase Behavior in an Online Marketplace. Journal of Theoretical and Applied Electronic Commerce Research, 16(7), pp.2965–2980. doi:https://doi.org/10.3390/jtaer16070162.

Kaur, H. and Singh, K., 2013. Market basket analysis of sports store using association rules. International Journal of Recent Trends in Electrical & Electronics Engg, 3(1), pp.81-85.

Kaur, M. and Kang, S., 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, pp.78-85.

Kim, H.K., Kim, J.K. and Chen, Q.Y. (2012). A product network analysis for extending the market basket analysis. Expert Systems with Applications, 39(8), pp.7403–7410. doi:https://doi.org/10.1016/j.eswa.2012.01.066.

Kumar, B., Gangurde, R. and Gore, S.D. (2017). Optimizes Predictive Model using Artificial Neural Network for Market Basket Analysis. JJCSC, 9(1), pp.42–52.

Li, X., Mazvancheryl, S., Prasad, S.Y., Rau, P. and Soyer, R., 2008. Market Basket Analysis Using Bayesian Networks.

Lim, T. (2021). K-Means Clustering-Based Market Basket Analysis: U.K. Online E-Commerce Retailer. 2021 International Conference on Information Technology (ICIT). doi:https://doi.org/10.1109/icit52682.2021.9491703.

Mitra, A., Ghosh, S., Basuchowdhuri, P., Shekhawat, M.K. and Saha, S.K. (2016). Recommendation system based on product purchase analysis. Innovations in Systems and Software Engineering, 12(3), pp.177–192. doi:https://doi.org/10.1007/s11334-016-0274-x.

Morse, G. (2006). Decisions and Desire. [online] Harvard Business Review. Available at: https://hbr.org/2006/01/decisions-and-desire.

Musalem, A., Aburto, L. and Bosch, M. (2018). Building brand authenticity in fast-moving consumer goods via consumer perceptions of brand marketing communications. European Journal of Marketing, [online] 52(7/8), pp.1387–1411. Available at: https://repositorio.uchile.cl/bitstream/handle/2250/152625/Market-basket-analysis.pdf [Accessed 8 Jun. 2023].

Newman, M.E.J. (2006a). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), pp.8577–8582. doi:https://doi.org/10.1073/pnas.0601602103.

Newman, M.E.J. (2006b). The mathematics of networks. The new palgrave encyclopedia of economics, pp.1–12.

Ojugo, A.A. and Eboka, A. (2019). Inventory prediction and management in Nigeria using market basket analysis associative rule mining: memetic algorithm based approach. International Journal of Informatics and Communication Technology (IJ-ICT), 8(3), pp.128–138. doi:https://doi.org/10.11591/ijict.v8i3.

Panel®, E. (2020). Council Post: 13 Ways For Entrepreneurs To Keep Up With Changing Customer Needs. [online] Forbes. Available at: https://www.forbes.com/sites/forbescoachescouncil/2020/09/25/13-ways-for-entrepreneurs-to-keep-up-with-changing-customer-needs/ [Accessed 25 Jun. 2023].

Pankratz, B., Bogumił Kamiński and Paweł Prałat (2023). Community Detection Supported by Node Embeddings (Searching for a Suitable Method). Studies in computational intelligence, 1078, pp.221–232. doi:https://doi.org/10.1007/978-3-031-21131-7_17.

PsycholoGenie (2014). Associationism and its Laws in Psychology Explained in Thumbnail. [online] Psychologenie. Available at: https://psychologenie.com/associationism-in-psychology-explained.

Qi, W., Ma, S. and Dai, Y., 2013, July. Shopping basket analysis based on the social network theory. In 2013 Ninth International Conference on Natural Computation (ICNC) (pp. 1093-1097). IEEE.

Raorane, A.A., Kulkarni, R.V. and Jitkar, B.D., 2012. Association rule–extracting knowledge using market basket analysis. Research Journal of Recent Sciences ISSN, 2277, p.2502.

Raeder, T. and Chawla, N.V., 2011. Market basket analysis with networks. Social network analysis and mining, 1, pp.97-113.

Rehman, I. and Ghous, H., 2021. Structured Critical Review on Market Basket Analysis using Deep Learning & Association Rules. International Journal of Scientific & Engineering Research, 12(1), pp.1-24.

Roodpishi, M. and Nashtaei, R., 2015. Market basket analysis in insurance industry. Management Science Letters, 5(4), pp.393-400.

Roghani, H. and Bouyer, A. (2022). A Fast Local Balanced Label Diffusion Algorithm for Community Detection in Social Networks. IEEE Transactions on Knowledge and Data Engineering. doi:https://doi.org/10.1109/TKDE.2022.3162161.

Tan, S.C. and Lau, J.P.S., (2013), December. Time series clustering: A superior alternative for market basket analysis. In Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013) (pp. 241-248). Singapore: Springer Singapore.

Traag, V.A., Waltman, L. and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports, 9(1). doi:https://doi.org/10.1038/s41598-019-41695-z.

Ünvan, Y.A., 2021. Market basket analysis with association rules. Communications in Statistics-Theory and Methods, 50(7), pp.1615-1628.

Vahidi Roodpishi, M. and Aghajan Nashtaei, R. (2015). Market basket analysis in insurance industry. Management Science Letters, 5(4), pp.393–400. doi:https://doi.org/10.5267/j.msl.2015.2.004.

Valle, M.A., Ruz, G.A. and Morrás, R., 2018. Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97, pp.146-162.

Velmurugan, T. and Hemalatha, B., (2020). Mining implicit and explicit rules for customer data using natural language processing and apriori algorithm. Int. J. Adv. Sci. Technol, 29(9), pp.3155-3167.

Videla-Cavieres, I.F. and Ríos, S.A., 2014. Extending market basket analysis with graph mining techniques: A real case. Expert Systems with Applications, 41(4), pp.1928-1936.

Woma, J. and Ngo, K., (2019). Comparisons of Community Detection Algorithms in the YouTube Network. CS 224W: Machine Learning with Graphs Stanford University.

Wu, J., Wang, Y., Shafiee, S. and Zhang, D. (2021). Discovery of associated consumer demands: Construction of a co-demanded product network with community detection. Expert Systems with Applications, 178, p.115038. doi:https://doi.org/10.1016/j.eswa.2021.115038.

Yang, Z. and Sudharshan, D. (2019). Examining multi-category cross purchases models with increasing dataset scale – An artificial neural network approach. Expert Systems with Applications, 120, pp.310–318. doi:https://doi.org/10.1016/j.eswa.2018.11.038.

Zhang, J. and Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. Advances in Intelligent Systems Research, 132.

Zhang, L., Priestley, J., DeMaio, J., Ni, S. and Tian, X., 2021. Measuring customer similarity and identifying cross-selling products by community detection. Big data, 9(2), pp.132-143.

8. APPENDICES

Appendix 1: Metrices for Assessing Interestingness of Association Rules

Matrix	Formula
Support	$support(product_A) = P(product_A)$
Confidence	$confidence(product_A \rightarrow product_B) = P(product_B \mid product_A)$
Lift	$lift(product_A, product_B) = \frac{confidence(product_A \rightarrow product_B)}{support(product_B)}$ $= \frac{P(product_A \cup product_B)}{P(product_A) \cdot P(product_B)}$
Leverage	$leverage(product_A \rightarrow product_B)$ $= P(product_A \cup product_B) - P(product_A) \cdot P(product_B)$
Conviction	$\frac{1 - P(product_B)}{1 - P(product_B \mid product_A)}$

Appendix 2: Details of each Community

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
0	0	1	WHITE HANGING HEART T-LIGHT HOLDER	1.0	0.576	1.0	0.293	0.462
1	1	1	WHITE METAL LANTERN	0.517	0.0	0.674	0.239	0.462
2	2	1	CREAM CUPID HEARTS COAT HANGER	0.517	0.0	0.674	0.239	0.462
3	3	1	KNITTED UNION FLAG HOT WATER BOTTLE	0.517	0.0	0.674	0.239	0.462
4	4	1	RED WOOLLY HOTTIE WHITE HEART.	0.517	0.0	0.674	0.239	0.462
5	5	1	SET 7 BABUSHKA NESTING BOXES	0.517	0.0	0.674	0.239	0.462
6	6	1	GLASS STAR FROSTED T-LIGHT HOLDER	0.517	0.0	0.674	0.239	0.462
7	257	1	PACK OF 60 PINK PAISLEY CAKE CASES	0.069	0.0	0.518	0.021	0.462
8	205	1	BLUE DINER WALL CLOCK	0.414	0.0	0.63	0.069	0.462
9	81	1	GLASS STAR FROSTED T-LIGHT HOLDER	0.517	0.0	0.674	0.239	0.462
10	722	1	CERAMIC STRAWBERRY DESIGN MUG	0.414	0.0	0.63	0.069	0.462
11	88	1	JUMBO BAG CHARLIE AND LOLA TOYS	0.414	0.0	0.63	0.069	0.462
12	1053	1	LARGE ROUND WICKER PLATTER	0.414	0.0	0.63	0.069	0.462
13	291	1	BABUSHKA LIGHTS STRING OF 10	0.414	0.0	0.63	0.069	0.462
14	1572	1	PLASTERS IN TIN SKULLS	0.414	0.0	0.63	0.069	0.462
15	101	1	CHARLIE & LOLA WASTEPAPER BIN FLORA	0.414	0.0	0.63	0.069	0.462
16	1577	1	ROCKING HORSE RED CHRISTMAS	0.414	0.0	0.63	0.069	0.462
17	1001	1	SET OF 72 GREEN PAPER DOILIES	0.414	0.0	0.63	0.069	0.462
18	48	1	HAND WARMER UNION JACK	0.517	0.0	0.674	0.239	0.462
19	49	1	WHITE HANGING HEART T-LIGHT HOLDER	0.517	0.0	0.674	0.239	0.462
20	50	1	WHITE METAL LANTERN	0.517	0.0	0.674	0.239	0.462
21	51	1	CREAM CUPID HEARTS COAT HANGER	0.517	0.0	0.674	0.239	0.462
22	52	1	EDWARDIAN PARASOL RED	0.517	0.0	0.674	0.239	0.462
23	53	1	RETRO COFFEE MUGS ASSORTED	0.517	0.0	0.674	0.239	0.462
24	54	1	SAVE THE PLANET MUG	0.517	0.0	0.674	0.239	0.462
25	55	1	VINTAGE BILLBOARD DRINK ME MUG	0.517	0.0	0.674	0.239	0.462
26	1528	1	GLASS JAR DAISY FRESH COTTON WOOL	0.414	0.0	0.63	0.069	0.462
27	58	1	WOOD S/3 CABINET ANT WHITE FINISH	0.069	0.0	0.518	0.021	0.462
28	956	1	PLASTERS IN TIN SPACEBOY	0.414	0.0	0.63	0.069	0.462
29	1399	1	RECYCLING BAG RETROSPOT	0.414	0.0	0.63	0.069	0.462

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
30	8	2	HAND WARMER RED POLKA DOT	1.0	0.0	1.0	0.707	1.0
31	7	2	HAND WARMER UNION JACK	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
32	516	3	5 STRAND GLASS NECKLACE CRYSTAL	1.0	0.0	1.0	0.137	1.0
33	134	3	CERAMIC STRAWBERRY CAKE MONEY BANK	1.0	0.0	1.0	0.137	1.0
34	775	3	DIAMANTE HAIR GRIP PACK/2 RUBY	1.0	0.0	1.0	0.137	1.0
35	392	3	PARTY INVITES JAZZ HEARTS	1.0	0.0	1.0	0.137	1.0
36	391	3	PARTY INVITES FOOTBALL	1.0	0.0	1.0	0.137	1.0
37	10	3	POPPY'S PLAYHOUSE BEDROOM	1.0	0.0	1.0	0.137	1.0
38	3851	3	SET/6 RED SPOTTY PAPER CUPS	1.0	0.0	1.0	0.137	1.0
39	12	3	FELTCRAFT PRINCESS CHARLOTTE DOLL	1.0	0.0	1.0	0.137	1.0
40	11	3	POPPY'S PLAYHOUSE KITCHEN	1.0	0.0	1.0	0.137	1.0
41	3854	3	PINK POLKADOT BOWL	1.0	0.0	1.0	0.137	1.0
42	15	3	BOX OF VINTAGE JIGSAW BLOCKS	1.0	0.0	1.0	0.137	1.0
43	649	3	PACK OF 60 MUSHROOM CAKE CASES	1.0	0.0	1.0	0.137	1.0
44	401	3	PACK OF 60 DINOSAUR CAKE CASES	1.0	0.0	1.0	0.137	1.0
45	398	3	PACK OF 60 SPACEBOY CAKE CASES	1.0	0.0	1.0	0.137	1.0
46	277	3	VINTAGE BILLBOARD LOVE/HATE MUG	1.0	0.0	1.0	0.137	1.0
47	408	3	WHITE METAL LANTERN	1.0	0.0	1.0	0.137	1.0
48	793	3	PACK 3 BOXES BIRD PANNETONE	1.0	0.0	1.0	0.137	1.0
49	1820	3	ORIGAMI LAVENDER INCENSE/CANDL SET	1.0	0.0	1.0	0.137	1.0
50	1311	3	JOY WOODEN BLOCK LETTERS	1.0	0.0	1.0	0.137	1.0
51	296	3	HAND WARMER SCOTTY DOG DESIGN	1.0	0.0	1.0	0.137	1.0
52	41	3	VINTAGE SEASIDE JIGSAW PUZZLES	1.0	0.0	1.0	0.137	1.0
53	3885	3	JAM JAR WITH GREEN LID	1.0	0.0	1.0	0.137	1.0
54	174	3	JUMBO BAG BAROQUE BLACK WHITE	1.0	0.0	1.0	0.137	1.0
55	518	3	HAND WARMER SCOTTY DOG DESIGN	1.0	0.0	1.0	0.137	1.0
56	692	3	HOMEMADE JAM SCENTED CANDLES	1.0	0.0	1.0	0.137	1.0
57	181	3	HOME BUILDING BLOCK WORD	1.0	0.0	1.0	0.137	1.0
58	567	3	5 HOOK HANGER RED MAGIC TOADSTOOL	1.0	0.0	1.0	0.137	1.0
59	1725	3	CHARLIE + LOLA RED HOT WATER BOTTLE	1.0	0.0	1.0	0.137	1.0
60	319	3	LAVENDER INCENSE IN TIN	1.0	0.0	1.0	0.137	1.0
61	3137	3	MEMO BOARD RETROSPOT DESIGN	1.0	0.0	1.0	0.137	1.0
62	3138	3	NATURAL SLATE HEART CHALKBOARD	1.0	0.0	1.0	0.137	1.0
63	3525	3	PHOTO FRAME 3 CLASSIC HANGING	1.0	0.0	1.0	0.137	1.0
64	326	3	DISCO BALL ROTATOR BATTERY OPERATED	1.0	0.0	1.0	0.137	1.0
65	455	3	WRAP COWBOYS	1.0	0.0	1.0	0.137	1.0
66	842	3	SKULLS WRITING SET	1.0	0.0	1.0	0.137	1.0
67	843	3	VINTAGE PAISLEY STATIONERY SET	1.0	0.0	1.0	0.137	1.0
68	1228	3	CHILDREN'S APRON DOLLY GIRL	1.0	0.0	1.0	0.137	1.0
69	1229	3	RETROSPOT CHILDRENS APRON	1.0	0.0	1.0	0.137	1.0
70	849	3	CINAMMON SET OF 9 T-LIGHTS	1.0	0.0	1.0	0.137	1.0
71	984	3	BLUE PAISLEY TISSUE BOX	1.0	0.0	1.0	0.137	1.0
72	601	3	SMALL POPCORN HOLDER	1.0	0.0	1.0	0.137	1.0
73	985	3	HEART OF WICKER SMALL	1.0	0.0	1.0	0.137	1.0
74	222	3	SET/20 RED RETROSPOT PAPER NAPKINS	1.0	0.0	1.0	0.137	1.0
75	489	3	5 STRAND GLASS NECKLACE CRYSTAL	1.0	0.0	1.0	0.137	1.0
76	365	3	PINK NEW BAROQUECANDLESTICK CANDLE	1.0	0.0	1.0	0.137	1.0
77	111	3	RED DRAWER KNOB ACRYLIC EDWARDIAN	1.0	0.0	1.0	0.137	1.0
78	369	3	TEA TIME TABLE CLOTH	1.0	0.0	1.0	0.137	1.0
79	1394	3	LUNCH BAG SPACEBOY DESIGN	1.0	0.0	1.0	0.137	1.0
80	497	3	PAISLEY PATTERN STICKERS	1.0	0.0	1.0	0.137	1.0
81	372	3	PICNIC BASKET WICKER SMALL	1.0	0.0	1.0	0.137	1.0
82	377	3	POSTAGE	1.0	0.0	1.0	0.137	1.0
83	379	3	CLEAR DRAWER KNOB ACRYLIC EDWARDIAN	1.0	0.0	1.0	0.137	1.0
84	511	3	6 RIBBONS SHIMMERING PINKS	1.0	0.0	1.0	0.137	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
85	2080	4	POPPY'S PLAYHOUSE KITCHEN	1.0	0.0	1.0	0.302	1.0
86	2081	4	POPPY'S PLAYHOUSE BEDROOM	1.0	0.0	1.0	0.302	1.0
87	1379	4	ROUND SNACK BOXES SET OF4 WOODLAND	1.0	0.0	1.0	0.302	1.0
88	1892	4	SWEETHEART CERAMIC TRINKET BOX	1.0	0.0	1.0	0.302	1.0
89	2119	4	CHRISTMAS LIGHTS 10 REINDEER	1.0	0.0	1.0	0.302	1.0
90	2183	4	RIBBON REEL CHRISTMAS SOCK BAUBLE	1.0	0.0	1.0	0.302	1.0
91	720	4	CREAM HEART CARD HOLDER	1.0	0.0	1.0	0.302	1.0
92	2034	4	CHRISTMAS LIGHTS 10 REINDEER	1.0	0.0	1.0	0.302	1.0
93	2035	4	WOOD BLACK BOARD ANT WHITE FINISH	1.0	0.0	1.0	0.302	1.0
94	1044	4	WHITE BELL HONEYCOMB PAPER	1.0	0.0	1.0	0.302	1.0
95	1045	4	BLACK/BLUE POLKADOT UMBRELLA	1.0	0.0	1.0	0.302	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
96	1344	5	SILVER HANGING T-LIGHT HOLDER	1.0	0.0	1.0	0.378	1.0
97	753	5	COSY HOUR GIANT TUBE MATCHES	1.0	0.0	1.0	0.378	1.0
98	129	5	CERAMIC CHERRY CAKE MONEY BANK	1.0	0.0	1.0	0.378	1.0
99	196	5	VINTAGE UNION JACK CUSHION COVER	1.0	0.0	1.0	0.378	1.0
100	3239	5	SMALL POPCORN HOLDER	1.0	0.0	1.0	0.378	1.0
101	120	5	YELLOW BREAKFAST CUP AND SAUCER	1.0	0.0	1.0	0.378	1.0
102	3244	5	RED HARMONICA IN BOX	1.0	0.0	1.0	0.378	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
103	545	6	PINK POLKADOT PLATE	1.0	0.0	1.0	0.707	1.0
104	1037	6	PLACE SETTING WHITE HEART	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
105	1666	7	60 CAKE CASES DOLLY GIRL DESIGN	1.0	0.0	1.0	0.577	1.0
106	524	7	JUMBO BAG RED RETROSPOT	1.0	0.0	1.0	0.577	1.0
107	525	7	JUMBO BAG STRAWBERRY	1.0	0.0	1.0	0.577	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
108	233	8	3 STRIPEY MICE FELTCRAFT	1.0	0.0	1.0	0.5	1.0
109	194	8	5 HOOK HANGER MAGIC TOADSTOOL	1.0	0.0	1.0	0.5	1.0
110	177	8	LIGHT GARLAND BUTTERFILES PINK	1.0	0.0	1.0	0.5	1.0
111	1070	8	LOVEBIRD HANGING DECORATION WHITE	1.0	0.0	1.0	0.5	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
112	201	9	BLACK/BLUE POLKADOT UMBRELLA	1.0	0.0	1.0	0.333	1.0
113	3743	9	12 MESSAGE CARDS WITH ENVELOPES	1.0	0.0	1.0	0.333	1.0
114	214	9	ROSE COTTAGE KEEPSAKE BOX	1.0	0.0	1.0	0.333	1.0
115	3575	9	RED RETROSPOT PUDDING BOWL	1.0	0.0	1.0	0.333	1.0
116	57	9	WOOD 2 DRAWER CABINET WHITE FINISH	1.0	0.0	1.0	0.333	1.0
117	158	9	HEART OF WICKER SMALL	1.0	0.0	1.0	0.333	1.0
118	189	9	FRIDGE MAGNETS US DINER ASSORTED	1.0	0.0	1.0	0.333	1.0
119	190	9	HOMEMADE JAM SCENTED CANDLES	1.0	0.0	1.0	0.333	1.0
120	159	9	HEART OF WICKER LARGE	1.0	0.0	1.0	0.333	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
121	320	10	TV DINNER TRAY VINTAGE PAISLEY	1.0	0.0	1.0	0.447	1.0
122	180	10	RED TOADSTOOL LED NIGHT LIGHT	1.0	0.0	1.0	0.447	1.0
123	1351	10	PAPER BUNTING COLOURED LACE	1.0	0.0	1.0	0.447	1.0
124	1355	10	METAL 4 HOOK HANGER FRENCH CHATEAU	1.0	0.0	1.0	0.447	1.0
125	295	10	HAND WARMER RED RETROSPOT	1.0	0.0	1.0	0.447	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
126	65	11	VICTORIAN SEWING BOX LARGE	1.0	0.0	1.0	0.577	1.0
127	171	11	ORGANISER WOOD ANTIQUE WHITE	1.0	0.0	1.0	0.577	1.0
128	172	11	LUNCH BAG DOLLY GIRL DESIGN	1.0	0.0	1.0	0.577	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
129	4158	12	CHILDREN'S SPACEBOY MUG	1.0	0.0	1.0	0.707	1.0
130	4159	12	HEART DECORATION PAINTED ZINC	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
131	208	13	SMALL HEART MEASURING SPOONS	0.167	0.0	0.545	0.116	0.571
132	3341	13	ALARM CLOCK BAKELIKE IVORY	0.167	0.0	0.545	0.116	0.571
133	3434	13	JAM MAKING SET WITH JARS	0.667	0.0	0.75	0.432	0.571
134	3435	13	PIN CUSHION BABUSHKA PINK	0.667	0.0	0.75	0.432	0.571
135	3436	13	PAPER CHAIN KIT 50'S CHRISTMAS	0.667	0.0	0.75	0.432	0.571
136	45	13	POSTAGE	1.0	0.6	1.0	0.477	0.571
137	3439	13	ALARM CLOCK BAKELIKE IVORY	0.667	0.0	0.75	0.432	0.571

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
138	3240	14	LARGE POPCORN HOLDER	1.0	0.0	1.0	0.707	1.0
139	3378	14	RETRO COFFEE MUGS ASSORTED	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
140	3608	15	BLUE HAPPY BIRTHDAY BUNTING	1.0	0.0	1.0	0.707	1.0
141	3275	15	ASSTD DESIGN 3D PAPER STICKERS	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
142	3115	16	TEA TIME KITCHEN APRON	1.0	0.0	1.0	0.707	1.0
143	740	16	SMALL CHINESE STYLE SCISSOR	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
144	113	17	PHOTO CLIP LINE	1.0	0.0	1.0	0.5	1.0
145	1418	17	CHRISTMAS TREE STAR DECORATION	1.0	0.0	1.0	0.5	1.0
146	76	17	WOODEN PICTURE FRAME WHITE FINISH	1.0	0.0	1.0	0.5	1.0
147	138	17	GLASS CLOCHE SMALL	1.0	0.0	1.0	0.5	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
148	402	18	LUNCH BAG SUKI DESIGN	1.0	0.0	1.0	0.577	1.0
149	380	18	PINK DRAWER KNOB ACRYLIC EDWARDIAN	1.0	0.0	1.0	0.577	1.0
150	364	18	BLUE NEW BAROQUE CANDLESTICK CANDLE	1.0	0.0	1.0	0.577	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
151	529	19	PACK OF 12 RED RETROSPOT TISSUES	0.667	0.0	0.75	0.435	0.833
152	618	19	RIBBON REEL MAKING SNOWMEN	1.0	0.167	1.0	0.557	0.833
153	2333	19	MAGIC DRAWING SLATE CIRCUS PARADE	0.667	0.0	0.75	0.435	0.833
154	623	19	HEART OF WICKER LARGE	1.0	0.167	1.0	0.557	0.833

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
155	46	20	PAPER CHAIN KIT 50'S CHRISTMAS	1.0	0.0	1.0	0.707	1.0
156	167	20	JAM MAKING SET PRINTED	1.0	0.0	1.0	0.707	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
157	104	21	JUMBO BAG PINK VINTAGE PAISLEY	1.0	0.0	1.0	0.5	1.0
158	1105	21	ASSORTED COLOUR LIZARD SUCTION HOOK	1.0	0.0	1.0	0.5	1.0
159	103	21	JUMBO STORAGE BAG SUKI	1.0	0.0	1.0	0.5	1.0
160	47	21	HAND WARMER RED POLKA DOT	1.0	0.0	1.0	0.5	1.0

	Product ID	Community Number	Product Name	Degree Centrality	Betweenness	Closeness	Eigenvector	Community Density
161	355	22	WHITE HANGING HEART T-LIGHT HOLDER	1.0	0.0	1.0	0.707	1.0
162	59	22	WOODEN PICTURE FRAME WHITE FINISH	1.0	0.0	1.0	0.707	1.0