# Declaration

I declare that this research project is my original work and has not been submitted for the award of a degree in this or any other University. To the best of my knowledge and belief, this thesis proposal contains no material previously published or written by another person except where due reference is made on the proposal itself.

SIGNED ………………………………………………………………………………… DATE ……………………………………………

Tarus Kiprop Martin

Registration Number: 056935

**Approval**

This research project has been submitted for examination with my approval as university supervisor.

Dr. Marwanga ---------------- ---------------

SIGNATURE DATE

Lecturer Faculty of Information technology, Strathmore University

# Dedication

To my late father Joseph, my mother Army, my brothers Allan, Collins and Dennis and my fiancée Brenda. I dedicate this work to because you have been with me all this while through the strain of the study. It is my sincere hope that I will, in one way or the other be able to return a similar favor.

# Abstract

This research investigates the benefits of studying social networks in a bar setting to enable them be able have more customized promotions from data collected and analysis these data by identifying relationships of customers drinking preferences with the aim of identifying usable value in these relationships. It is aimed at developing and using an algorithm that will categorise an individual as per what he or she drinks thereby creating an avenue of prediction on what would be the next item he or she would want by studying the probabilities of similar preferences between other individuals.

The study adopts a social network analysis tool that brings together data about a person’s dinking habit relating it to the number of times the individual visits a particular pub or night club then aggregates this data to different other people with the same habit providing critical information.

The study group for this study are patrons of drinking establishments, while the audience of the resulting information are the owners and managers of the same establishments.

# Acknowledgement

I would like to express my sincere gratitude to persons whose assistance during this period has been of great inspiration.

To Dr. Marwanga, my supervisor, I am particularly grateful for his patience and guidance towards enabling me to complete this proposal on time. I hope to prove his trust to not be in vain.

To Anthony Mwangi, my ground supervisor, who accepted to help me convince various bars and night clubs to use business intelligence as a source of information to better their decision and also in believing what I hoped to be of use came true.

Finally, to my friends who were by me all the nights testing the logics and everything on the system, know I shall not forget the assistant and encouragement that you have provided me.

May God bless you all.

# Definition of abbreviations

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Terms** | **Abbr.** | **Description** |
| Business Intelligence | BI | Business intelligence (BI) is a discipline made up of several inter-related fields including data mining, online analytical processing, social network analysis querying and reporting. (Mulcahy,2014) |
| Social Network Analysis | SNA |  |
| Social Networks | SN |  |
| Radio Network Controller | RNC |  |

Table of Contents

[Declaration i](#_Toc413774273)

[Dedication ii](#_Toc413774274)

[Abstract iii](#_Toc413774275)

[Acknowledgement iv](#_Toc413774276)

[Definition of abbreviations v](#_Toc413774277)

[List of Figures viii](#_Toc413774278)

[List of tables ix](#_Toc413774279)

[CHAPTER 1: INTRODUCTION 1](#_Toc413774280)

[1.1 Background to the Study 1](#_Toc413774281)

[1.1.1 Business intelligence 1](#_Toc413774282)

[1.1.2 Social network analysis 1](#_Toc413774283)

[1.2 Problem Statement 3](#_Toc413774284)

[1.3 Research objectives 4](#_Toc413774285)

[1.4 Research Questions 4](#_Toc413774286)

[1.5 Justification 4](#_Toc413774287)

[1.6 Scope of the study 5](#_Toc413774288)

[CHAPTER 2: LITERATURE REVIEW 6](#_Toc413774289)

[2.1 Introduction 6](#_Toc413774290)

[2.2 Business intelligence 6](#_Toc413774291)

[2.2.1 Role of business intelligence in businesses 7](#_Toc413774292)

[2.2.2 Impact of business intelligence in business 7](#_Toc413774293)

[2.2.3 Business intelligence implementation 8](#_Toc413774294)

[2.3 Social network analysis 9](#_Toc413774295)

[2.3.1 Introduction 9](#_Toc413774296)

[2.3.2 Social network analysis illustrations 10](#_Toc413774297)

[2.3.3 Theories of SNAs’ 16](#_Toc413774298)

[CHAPTER 3: RESEARCH METHODOLOGY / DESIGN 20](#_Toc413774299)

[3.1 Introduction 20](#_Toc413774300)

[3.2 Research design 20](#_Toc413774301)

[3.3 Definition of variables 21](#_Toc413774302)

[3.3.1 Alcohol products / categories 22](#_Toc413774303)

[3.3.2 Frequency 22](#_Toc413774304)

[3.4 Unit of analysis 22](#_Toc413774305)

[3.5 Data collection methods 22](#_Toc413774306)

[3.6 Data analysis 22](#_Toc413774307)

[CHAPTER 4: Presentation and Analysis of Findings 23](#_Toc413774308)

[4.1 Comparison of Sharing between categories 23](#_Toc413774309)

[4.1.1 Determining which categories are most sharable 23](#_Toc413774310)

[4.1.2 Determining which percentile to use 24](#_Toc413774311)

[4.2 Determining sharing between categories 25](#_Toc413774312)

[4.2.1 Sharing categories at 90th Similarity 25](#_Toc413774313)

[4.2.2 Sharing categories at 80th Similarity 26](#_Toc413774314)

[4.2.3 Sharing categories with respect to the ratio of the number of shared to the total number of customers per category 27](#_Toc413774315)

[4.3 Determining Relationships between individual Categories 28](#_Toc413774316)

[4.3.1 Comparison at the 90th Similarity 28](#_Toc413774317)

[4.3.2 Comparison at 80th to 100th Percentile 29](#_Toc413774318)

[CHAPTER 5: Conclusions and Recommendations 33](#_Toc413774319)

[5.1 Conclusions 33](#_Toc413774320)

[5.2 Recommendations 34](#_Toc413774321)

[CHAPTER 6: REFERENCES 35](#_Toc413774322)

# List of Figures

[Figure 2‑1: Roles of business intelligence in businesses (Olszak and Ziemba, 2007) 7](#_Toc413773160)

[Figure 2‑2 Knowledge creation process through BI (Sacu & Spruit 2010) 8](#_Toc413773161)

[Figure 2‑3 Challenges online game companies face (Idiro Analytics, n.d) 14](#_Toc413773162)

[Figure 4‑1 Ratio of people consuming one product to other products 23](#_Toc413773163)

[Figure 4‑2: 70th percentile is the ideal percentile to use 24](#_Toc413773164)

[Figure 4‑3: Relationships of categories with respect to the percentage number of shared customers to total number of customers per category 27](#_Toc413773165)

[Figure 4‑4: Comparison at 90th Similarity 28](#_Toc413773166)

[Figure 4‑5: Comparison at 80th Similarity 29](#_Toc413773167)

[Figure 4‑6: Cigarettes and Brandy relationship 30](#_Toc413773168)

# List of tables

[Table 4‑1: Similarity table at 90th similarity 25](#_Toc413773169)

[Table 4‑2: Similarity of categories at 80th similarity level 26](#_Toc413773170)

# INTRODUCTION

## Background to the Study

The background to the study will provide an overview of business intelligence and social network analysis which are the two key concepts used in this research.

### Business intelligence

Mulcahy (2014) defines Business Intelligence (BI) as a discipline made up of several inter-related activities that includes data mining, online analytical processing and creating graphical representations of the information analysed. Sacu & Spruit (2010) looks at BI as a tool that helps business organizations make sound decisions by transforming data into knowledge through the use of various analytical tools.

BI contains sets of actions points that help businesses use the concept in achieving their goals. These actions points are a) data collection which basically involves the collection of data using various technologies i.e. mobile applications, desktop applications social network sites etc. this is supported by a data warehousing component. b) Gathering action which is supported by an extract-transform-load component. c) Analysis action that is supported by on-line analytical products i.e. algorithms to identify the not obvious information from the gathered data. Examples of analysis processes are like identifying relationships between data, checking for data similarity etc. Lastly we have reporting which involves the representation of the information from the analysis in a manner that can be easily understood by the organization members (Olszak & Ziemba, 2007).

### Social network analysis

Butts (2008) describes Social Network Analysis (SNA) as a process whose objective is to find relationships between people in a community or a study group with the aim of identifying usable value in these relationships. Some of the industries benefiting from the use of SNAs are introduced below, the study will attempt relate to four similar characteristics drawn from the illustration. These characteristics are a) the community studied, b) the characteristics used to identify the similarity between individuals in the community, b) the limited resource each company wants to maximize and d) the business case for undertaking the analysis. These aspects will thereafter be used to show how the successful implementation of SNA in these seemingly unrelated areas can apply in the bar business scenario. The following are the illustrations of SNAs’.

#### Loyalty cards

A loyalty card is a card given to a customer for the purpose of enforcing the role of loyalty programs in a business. The customer uses this card to earn points, sometimes redeem the earned points and to also use the same as a credit card in paying of services or products. These cards have been used in businesses such as supermarkets. Example is the Nakumat Global card.

Data collected by this card can be used by organisations to calculate factors including the frequency of customer purchases and the combination of products purchased which can be information enough for marketing strategies that will thereby be profitable to the organizations (Worthington, 2009). The author states that through studying behavioural patterns of multiple customers over time, the organization is able to better optimize its operations and better design promotions.

#### Mobile network masts positioning

In the telecommunication industry one challenge that mobile telecommunication companies face has been brought about by the emergence of the mobile phone technology *Jonge, Pelt and Ross (2012)*. According to the authors, the emergence of this technology has brought about a challenge in allocating data storage which ought to always handle the varying number of mobile network user in a place. Ayyub and Jain (2009) from the school of engineering noted that mobile network users’ movements greatly affect the data storage distribution in a country, this means that in places with large number of people there should be a relative amount of data allocated in the region to handle the large number of network users.

Mobile network companies benefit from studying the movements of people all through the year through using this information as a guide on determining the required data storage capacity for a particular region keeping in mind the number of users in that region at a particular time. This means that data allocation can be done periodically though with some background of what has changed in a region therefore reducing wastage of data and also reduce the chances of having limited data in a region where people are many (Ayyub & Jain, 2009).

#### Player analytics on online games

Online games are the games where gamers can access them over the internet. These games allow gamers to compete with each other, create profiles and also purchase items to make the game more interesting. Gamers profile information is key since it can be used in SNAs in this industry to enable determine different structures to modify and what not to modify through an analysis of preferences gamers have. Some games have been equipped with profile monitoring schemes that can be used in tracking the different aspects the developers of the games have in order to help them improve on the delivery of their product.

An organization such as Electronic Arts (EA sports) benefits by being able to design better games, targeted promotions which thereby increases the organizations’ revenue and also reducing the time and cost to design and develop a product. This is attained by employing the end user feedback through their preferences i.e. choice of cars in a game of racing and also studying the levels that different people have attained over a period of time.

## Problem Statement

Tamasha Bar, Lounge and Restaurant (Tamasha) approached the researcher with the understanding that studying customer behaviour and habits could help them improve their marketing strategy i.e. promotions, customer loyalty and also increasing their profits and customer share in the market. Bearing in mind that for Social Network Analysis (SNA) to work, there is need to have a source of data, Tamasha has a system that captures customer details and orders through which the data can be identified and used for analysis. This information include: customer name, order details, payment information and also preference on the orders they want.

According to Worthington (2009), Hannover Research (2010) and Bellizzi & Bistol (2004) articles describing loyalty cards, Ayyub and Jain (2009) and Jonge, Pelt and Ross (2012) describing mobile mast positioning and Kirman and Lawson (2009) and Applying social network analysis (SNA) to gaming (n.d) articles describing player analytics on on-line games which shows how the retail , telecommunication and gaming industries respectively benefited from studying people’s habits with an aim of identifying relations through extracting important information and using the information to make decisions for the benefit of their organizations.

The study will sample three Social Network Analysis (SNA) illustrations describing how each of them determined the non-obvious relations developing to important information that the business can use. The study will take into account the similarities of the illustration mapping them to the area of study with an aim of using relevant architectures implemented in the study to attain the objectives.

## Research objectives

1. Identify applications using social network analysis for business intelligence
2. Identifying algorithms used in studying study human behaviours
3. Apply the algorithms to determine valuable information from applying the algorithms to the case of bars
4. Translate information for the business use

## Research Questions

1. What applications use social network analysis for business intelligence?
2. Which algorithms can be used in studying human behaviours?
3. What information can be derived from applying Social Network Analysis algorithms to data in the bar?
4. How will the business benefit from combining and similarities of products for the purpose of increasing revenue?

## Justification

According to Olszak and Ziemba (2006) description of Business Intelligence (BI), knowledge can be created through analysing data following certain steps which Sacu and Spruit (2010) describe as an ideal model how this can be achieved. Social Network Analysis (SNA) on the other hand defined by Butts (2008) can be related as a subset of BI meaning that through applying SNAs’, BI advantages can be realised in a business.

The problem statement describes the angle of the study which is using SNA in Tamasha’s case in order to realise the goal of predicting and customizing promotions through the analysis of customer information. It has been previously noted that the study will use three illustrations or case studies that will recommend how the study will be conducted, some of the reasons of using these case studies are; the nature of the community of study applies for both cases. Secondly, in both cases relationship between the community members can be determined through the different items of considerations, thirdly it will be realised that there is a limited resource that cannot be changed and therefore must be utilised in order to realise business benefit. Lastly we have the business advantage which both cases look up for in order to attain the business goal.

The study will use the customer as the community, the relationship between customers will be determined through how often, what the customer prefers, what quantity the persons takes and also the sequence the customers takes the products in the bar. The limited resource in this case is the promotions that the clubs do often seeking to increase their profits and also creates loyalty to their customers, lastly the business advantage is creating room for customized promotions to realise customer loyalty and also improved sales.

In summary the three illustrations on how SNA can be an important tool for a business to make use of information that previously was thought to be just data and not so important. The study can achieve its objective by applying the same concepts applied in the illustrations thereby a bar can be able to increase its revenue and also have the promotions with guided knowledge or information.

## Scope of the study

The study concentrates on analysing Social Networks (SN) in a bar setting with an aim of finding relationships between individual drinking habits and also fitting these individuals into categories defined by their drinking preference in relation to other drinkers. The outcome of this analysis should help Tamasha capitalize on promotion through prediction the analysis will give. The information will enable the business maintain high revenues and also customer loyalty

This therefore, maintains that the organization should have in place a system collecting the information about its customers and their orders over a period of time. The study will use this collected data to analyse and cluster individuals in categories which fits their drinking preference through the use of a categorising algorithm. The results of this will help in predicting certain preferences which can be found from a customer with a same drinking trait therefore objective promotions of products can be attained easily.

# LITERATURE REVIEW

## Introduction

Literature review section of the document will discuss related studies and research done by various authors in relation to the study. The research will borrow the components highlighted from the illustrations in the previous chapter of this document and thereby applying the same components to the study in order to meet the research objectives.

This chapter starts with an introduction to general concepts about business intelligence , the effect or importance of using it in a business environment and how well can this concept be used in different type of businesses around (Section 2.1). The introduction on business intelligence will be followed by definition of concepts about social networks and how it influences business intelligence, how to what extent has business intelligence been used as a resource centre of information on business intelligence systems and lastly works done in relation to the study of social networks in improving business intelligence (Section 2.2). This will be followed by a review of how social networks can be analysed in bistro cases in Nairobi to improve on sales and customer service (Section 2.3).

## Business intelligence

Olszak and Ziemba (2006) describe Business Intelligence (BI) systems, to be knowledge creating tools that helps businesses in decision-making. This tool comprises several related action points that include data mining, online analytics, querying and reporting (Mulcahy,2014).

The concept of using intelligent systems in a business setting has been described by (Solomon, 2004) to complement the decision making process. Intelligent systems combine operational data with various analytical tools to present complex, reliable and competitive information to decision makers for fact based decision-making.

In order for any business to meet its objectives, the business operations should be completed through effective and efficient ways to realise success. One of the ways to attain success is through intelligence systems which adapts to any environment change without losing an opportunity (Olszak and Ziemba, 2007).

### Role of business intelligence in businesses

Business intelligence has a role in the strategy / decision arm of an organization; this means that it is a technology used by them to effectively and efficiently make abstract decision rather that decisions made out of wit or feelings (Solomon, 2004).

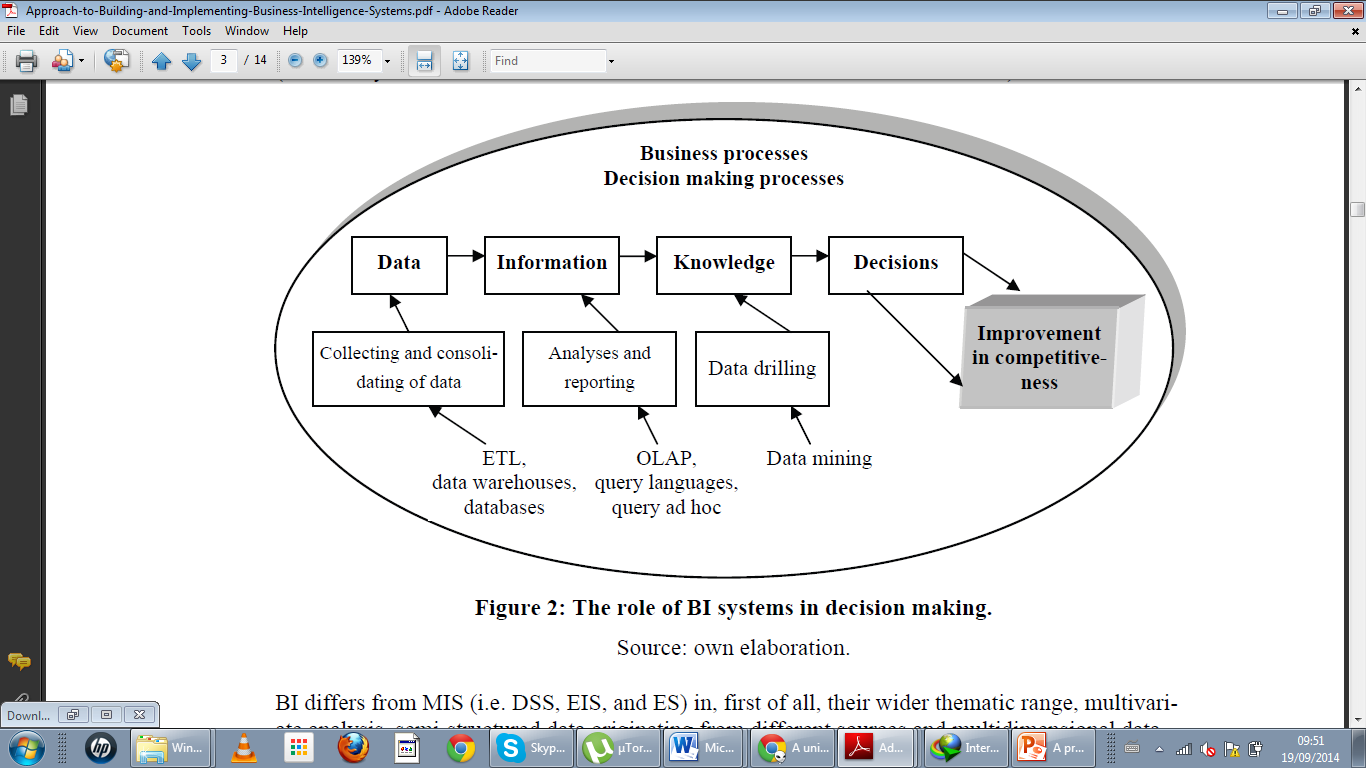


Figure 2‑1: Roles of business intelligence in businesses (Olszak and Ziemba, 2007)

Every organization aims to better their structures in order to improve the quality they offer to the community. Through employing intelligence systems in business processes the realization of improved competitiveness shown in **Figure 1** in the business will be achieved. Intelligent system is one major stronghold that many organizations use as strategy in making decisions (Olszak & Ziemba, 2007).

### Impact of business intelligence in business

There are various reasons why many businesses have embarked on going forward in implementing business intelligence as a strategy of decision making. Sacu and Spruit (2010) outline the several things that came about when business intelligent systems are used by organizations. a) possibility of accessing historical, summarized and consolidated information, b) single source of true information, because of the consistent nature of the data in the data warehouse as they have been cleaned, transformed and integrated, c) OLAP technology plus various reporting tools offer a summarized/detailed access to the data stored and d) There is a separation of the operational and decisional or analytical processing as they have a different architecture.

An illustration of how business intelligence implements the described action has been explained by Olexova (2014) in a journal where business intelligence was adopted in one of the retail chain stores. According to the author several benefits were realised over time, some of the benefits are a). There was an improved ability to anticipate earlier changes in the markets, this is so because the BI system looked into the trends of how and what customers shop, by doing so the retail store was able to better display their products in a manner the customers could easily access the different varieties of the same product. b) Stock management optimization was realised. The BI system used by the retails store was able to monitor the stock turn over from the purchase made by the customers thereby predicting through the number of purchases when to have more stock than usual because of high number of people coming to purchase the products.

### Business intelligence implementation

According to Olszak and Ziemba (2006) description of Business Intelligence (BI) systems as knowledge creating tools, Sacu and Spruit (2010) describes the steps of knowledge creation in BI system. a) Data selection and cleaning b) Data enrichment c) Data transformation d) Data mining e) Interpretation Knowledge.

The figure below shows the steps of knowledge creation



Figure 2‑2 Knowledge creation process through BI (Sacu & Spruit 2010)

From this illustration on how knowledge is attained when using BI systems, the same steps when applied in my study area can assist in identifying the following towards attaining the objectives of the study:

1. Step1 – Data selection and cleaning: This will determine and sort the relevant variables required to categorise individuals i.e. customer information, customer order details
2. Step 2 – Data enrichment: this step will help identify relations that can be described out of the many other relations on the variables identified i.e. summarising a customer’s drinking history
3. Step 3 – Data transformation: This will help in creating several outcomes of a relation based on the variable attached to the relations, i.e. relating a customer’s history to another with respect to their preferences
4. Step 4 – Data mining: the step represents the querying activity of the system to pull out relationships after the data has been analysed. i.e. assigning similarity quotient of related individuals according to some algorithm
5. Step 5 – Interpretation knowledge: knowledge can be used in prediction of the next item of preference therefore business can customize promotions to individuals according to their preferences and prediction.

## Social network analysis

In this section, the study will cover definition of social network analysis, impact of social network analysis and lastly illustrations social networks analysis implementation.

### Introduction

As Solomon (2004) state, BI is a composition of tools that analyse data to extract meaningful information, therefore it is not only important to know the relevant data required to be transformed to gain knowledge but also through what channels can this data be transformed to generate the knowledge required and what sources can this data be collected from.

In an article written by Chaudhuri, Dayak & Narasayy (2011) the various sources of data in many organizations come from Customer Relationship Management systems (CRM’s), RFID tags for inventory tracking, social media i.e. email, blogs, websites and social network platforms, social networks etc. Chaudhuri and colleagues continue saying that all these sources contain data of varying quality, use inconsistent representations, code and even formats. It is at this point that business intelligence architecture has to contain a component that will manipulate these different formats of data and relate all of them to one particular format that is understood by the business decision makers.

The study will source its data from the system installed in the bar which contains data about the Social Networks (SN), i.e. revellers and social drinkers in the pub and use Social Network Analysis as the model of analysing the habits of these individual with an aim of creating other opportunities for the business to make more money.

### Social network analysis illustrations

Oinas-Kukkonen, Lyytinen & Yoo (2010) describes Social Network Analysis (SNA) as channels through which various businesses study habits of people living in a community. This are habits can be structured in terms of relationships based on similarities identified from the habits each individual has thereby creating a study area of identifying usable value in the relationships (Butts, 2008). The previous chapter outlined the aspect which this study will later use to find the similarity of these cases therefore explain how the same aspects can be used in the study in attaining the objectives.

The outlined aspects were a) the community studied, b) the characteristics used to identify the similarity between individuals in the community, c) the limited resource each company wants to maximize and d) the business case for undertaking the analysis.

#### Loyalty cards

According to an article done by Bellizzi and Bistol (2004)the history of loyalty cards came during the time retailers and wholesalers used merchandised set collection promotion back in the 50’s and early 60’s. The two authors continue to say that with time various technologies such as point-of-sale scanners and cheaper means of collecting and storing customer purchase histories in a databank has greatly seen the growth of loyalty programs improve during the 1990s.

A Loyalty card is a way of enforcing loyalty programs by companies through giving rewards “points” to customers with an aim of strengthening the bond between them and the business.

Worthington (2009) in his study described a company which benefited from using the loyalty card. From the reading Tesco a United Kindom (UK) retailer launched its Tesco Clubcard which it has been noted that it has a participation rate of 80 per cent.

Tesco notes that the customer details are important to them as this information enables them to do promotions on their products directly to their customers. Example; they use the customer data to promote products that are likely to appeal to parents of students going to the university at the start of an academic year.

The company realised increased sales through in-store advertising, differential pricing and also better merchandising. Apart from this the company used the customer habits details to identify opportunities for them to increase its share of customer spending outside its store, i.e. Tesco company started offering insurance policies and loans to its customers.

Another company according to the Hannover Research (2009) that has noticed positive feedback from the use of the loyalty cards is Starbucks coffee franchise, which have in a place a loyalty program with a four-tier (Gold, Green, Welcome and non-member) loyalty program. According to the research the Gold Card carries in itself immense status as its noted that status and luxury are the focal messages in the company’s program description.

The loyalty program in Starbucks coffee shop runs by awarding a customer 1 star every time the customer pays for products with his card. Some of the enticers for customers to enrol in this program are like; free birthday beverage for all registered members of the My Starbucks card, the green level is attained when you have 5 stars which immediately will earn you rewards like free select syrups and milk etc.

Some of the benefits that Starbucks got from implementing this loyalty programs was the “loyalty to brand”, many people knew about Starbucks and in turn so many of them got themselves the loyalty card for them to enjoy rewards plus get the status and luxury by being members of Starbucks.

The two illustrations can be summarised accordingly with respect to the structure of the study. The various aspects being taken into consideration are as follows:

1. Community of study: The customers whose details have been captured in a data repository through a loyalty card in order to help in tracking the purchase histories, i.e. In the case Tesco retail store the community will be the shoppers e.g. parents.
2. Characteristics of relationship identification: These are defined through information about frequency of visits, the mix of products purchased and also the quantity of each product purchased in a visit. i.e for Tesco retail store the number of times a parent shops there and also the products he or she purchases determines is he or she is a parent.
3. Limited resource: How to display the new products in the market being a new season or year i.e. the Tesco retail store parents should be able to notice the new items for them to purchase for their children as they go back to school every new year.
4. Business advantage: Tesco retail store is able to advertise the school products i.e. bags, mattresses, shoes directly to the parents as they have their details and therefore also in the store the arrangement should have changed to have these products at the ground floor rather than the other floors.

The aspects to the study derived from the illustrations can be seen to match the business of running a bar. First it is seen that the number of customers a retail store serves daily is huge the same applies to the number of customers visiting the bar. Second is the relationship the retail store come up with, for them they use the information on customer purchases, i.e. if a parent buys form 1’s text books in a year it obvious that the next year he or she is going to get a form 2 text book. The same scenario can be applied in a bar setting where the relationship can be determined through the data on orders a customer makes, i.e. a customer purchasing beer on the first round then order a round of whisky, chances of ordering a cigarette could be high since previous analysis show that with the sequence of how the orders are placed there is a high probability of ordering a cigarette. This means that the business can be able to make strategic marketing to the customer about promotions on cigarettes available in the bar.

#### Mobile network mast positioning

In the previous chapter as explained Ayyub and Jain (2009) stated a challenge faced by telecommunication companies brought by the emergence of the mobile phone technology. Since this emergence, the companies have been left with no choice but to allocate usable data appropriately to different places in the country, keeping in mind the number of people in the area because of the resource being limited.

Jonge, Pelt and Ross (2012) approve the fact that the movements of people affects the network distribution therefore monitoring should be paramount. The authors outlined their hypothesis in order for them to study the movement of network users, some of them are; the economic activities which is related to regional call activity, the frequency and time frame an individual makes call can determine the calling behaviour and lastly the approximation of the number of people per time and location is possible with the number of calls per time.

Ayyub and Jain (2009) went further describing how mobile network is dispersed all over a region using the boosters. The two authors described the function of a Radio Network Controller (RNC) as a functional unit that is used to hold the data allocated and required for a particular place usage. This means that if an area’s RNC has less data allocated compared to the number of people it can handle, then chances of users experiencing problems with the network is high.

The RNCs’ can be configured to raise the average response time a sender sends requests and receive back a response with no faults in the network (Ayyub and Jain, 2009). They used an algorithm called the Dijikstra’s algorithm to determine the shortest path between boosters containing the multiple RNCs’. The aspects of the study can be drawn from the illustration and concept explained above and explained as follows:

1. Community of study: Number of network users in a geographical area.
2. Characteristics of relationship identification: The relationship can be identified through studying the nature of movements the members of the community have across the year, i.e. over the holidays it is known that people move from cities over to rural areas.
3. Limited resource: Data allocation required to serve the number of people in a geographical region on a specific period of time.
4. Business advantage: Data allocation in different areas of a country can be managed over by allocating data depending on the number of people in that area therefore reducing wastage of data where people are less and also there being a situation where the data is less to serve a large number of people.

The significance of this illustration to this study is that it complements it. The illustration marks similar traits to the study at hand i.e. the community to be studied is large, seen that, the number of mobile network users are many, same to the number of people visiting a bar. Secondly, how relationship of members in the community is determined can also be done for this case, given that, the illustration used an algorithm to determine how close the booster can be arranged to ensure seamless network flow. Lastly, the process on how telecommunication companies managed to get their business advantage in data allocation can also guide this study in mining information about the customers in a pub.

#### Player analysis in the gaming industry

Applying social network analysis (SNA) to gaming (n.d) article states that the gaming industry is now shifting its method of distribution from retail to online, the article continues to mention in order to improve the profitability of the games industry there was need to have in place a company that oversees the analysis of the online community, in this case online gamers. The company GamesAnalytics Ltd aims to improve the game industry by optimizing games plays and encouraging players to play longer, invite other people to join and also make purchases within the game through information collected at individual player levels (Applying social network analysis (SNA) to gaming, n.d).

According to a study by Kirman & Lawson (2009) on social network analysis for web-based online games, a key driver on how the social network analysis will positively impact the game industry was identified to be the “hardcore” players who are recognized to be the ones defining the social environment for the game. Applying social network analysis (SNA) to gaming (n.d) article describes the advantages using the “hardcore” players as a standard way of analysis in this industry. The insights brought about are used to convey personalized experiences, targeted contact strategies and improve company revenues.

IDIRO technologies established in 2003 a company with an aim to offer client the ability to gain insights and business intelligence into their industries, designed a business intelligent system that would help the online gamers solve their mystery on social network analysis in their industry. The figure below shows some challenges online game industries face in the market currently according to Idiro Analytics (n.d).

Figure 2‑3 Challenges online game companies face (Idiro Analytics, n.d)

According to Idiro Analytics (n.d) the solution created has been structured to perform the following; a) predict defection – through which the company can target the player prone to defect with retention offers, b) segmentation – through which the company can use community segmentation of the customer base to create more targeted marketing campaigns. c) Cross-sell & Up-sell – the company can identify gamers that are most susceptible to purchase value added services, e.g. power parks etc. d) Acquisition - The company can stimulate new gamers to be rooted through social activity of the game thus increase the game uptake.

1. Community of study: The online gamers.
2. Characteristics of relationship: The relationship can be identified through studying the nature of purchases a gamer makes, the levels a gamer has attained.
3. Limited resource: The levels of the game.
4. Business advantage: Through targeted marketing and acquisition strategies the business can ensure the customer base is maintained and also have more gamers attracted to the game.

The significance of this illustration to the study can be drawn from the structure of the solution describe by Idiro Analytics (n.d). Through advanced segmentation of online gamers, the gaming company can benefit from targeted marketing campaigns; the same can be applied to the area of study where the bar customers can be categorised according to their drinking habits with an aim of trying to find relationships between individuals who have similar drinking traits.

#### Summary

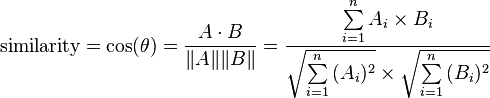
This research as mentioned before, is going to use these illustrations as the foundation of the study through employing some of the different tools in the illustrations to come up with a solution that will fits a bar setting. The illustrations explained above have outlined several characteristics with explanations on how each of the illustration resembles the bar business setting and therefore considering the level at which the illustration match with the area of study it can be concluded that, with a proper application of the tools or algorithms used to achieve SNAs in the respective industries the illustration lie, then chances of having the same working for the bar business is high.

### Theories of SNAs’

This section describes the various theories behind Social Network Analysis (SNAs), their advantages and the limitation that the theory has in the area of study.

#### Cosine similarity text rule

Similarity between two vectors of an inner product space can be identified from the cosine of the angle between the two vectors using the cosine rule. The cosine rule states that the cosine of 00 is 1, and it is less than 1 for any other angle (Sree & Murthy, 2012) the formula of finding text similarity is as follows:



The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity meaning that the closer the relationship the higher the cosine value

An example on how the cosine similarity rule works is displayed below:

Text 1: Julie loves me more than Linda loves me

Text 2: Jane likes me more than Julie loves me

1. We want to know how similar these texts are, purely in terms of word counts (and ignoring word order). We begin by making a list of the words from both texts:

me Julie loves Linda than more likes Jane

1. Now we count the number of times each of these words appears in each text:

me 2 2

Julie 1 1

likes 0 1

loves 2 1

Jane 0 1

Linda 1 0

than 1 1

more 1 1

1. We are not interested in the words themselves though. We are interested only in those two vertical vectors of counts. For instance, there are two instances of 'me' in each text. We are going to decide how close these two texts are to each other by calculating one function of those two vectors, namely the cosine of the angle between them.

The two vectors are, again:

a: [2, 1, 0, 2, 0, 1, 1, 1]

b: [2, 1, 1, 1, 1, 0, 1, 1]

The cosine of the angle between them is about 0.822.

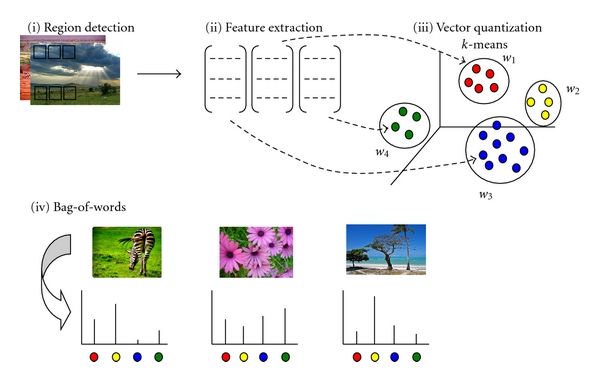
According to Sree and Murthy (2012) the cosine rule plays a role in a) Modelling of psychological tasks i.e. recognition, identification, and categorization of objects, where a common assumption is that the greater the similarity between a pair of objects, the more likely one will be confused with the other. b) Similarity also plays a key role in the modelling of preference and liking for products or brands, as well as motivations for product consumption etc.

The implication of this theory to the study is on how the cosine rule can be used to identify relationships between two vectors. It is noted that the above rule can be used to model preferences of liking. Therefore, in the case of a bar, the cosine rule will result to similarity quotient that can determine how close the relationship with regards to drinking habits of two individuals can be determined. The business will benefit from the similarity quotient by using the value determined to predict what could be the next preferred brand or product since the relationship is closer together.

The cosine similarity rule according to Sree and Murthy (2012) has some limitations when used in finding similarities between two vectors. These limitations are, a) time consuming, i.e. the rule has to repeat several times in order to check similarities for two relations one at a time, b) more space complexity, as seen the cosine similarity rule is complex due to the power factor it bring in the equation and c) not accurate clustering: i.e. since the similarity rule compares only two vectors at a time some of the stronger relationships can be harder to be identified in large masses of data. In conclusion the cosine similarity rule could be used in this research to determine the relationship between customers’ drinking habits, though the efficiency of the algorithm will be compromised since the environment of study covers a number of different parameters.

#### Bag of words theory

The bag-of-words (BoW) model can be used to classify images by using the features contained in the image. According to Tsai (2012) BoW model is used in document classification as a sparse vector of occurrence counts of words, same as the cosine similarity model describe above. The BoW model contains a number of steps used for coming up with classifications of the images, these steps are: a) identify points of interest on the image, b) calculate local descriptors over the points, c) quantize the descriptors into words forming a visual vocabulary and lastly find the occurrences in the image of each specific word in the vocabulary. The figure below shows the steps used in the BoW model for classifying images.



Based on the architecture of this model, classification can be achieved through use of different features of an image then using this information to modify the meaning of the image at the point of quantizing. This means that the more the number of features, the more complex the process. This models implication to the area of study is that, it can be used as a model to cluster customers in the environment of a bar, as the process entails identifying unique features in the drinking habits of a customer, then using this information to classify the customer as a person closely related to the other person with a similar drinking habit.

The BoW similarity rule according to Tsai (2012) has some limitations. These limitations are, a) time consuming, i.e. clustering huge, diverse and rapidly changing data can be very complex b) Narrow result, the results for this cluster will mainly depend on the set of event upon which the clustering criteria was applied c) “any-of-problem” challenge: in cases where an item belongs to more than one cluster it could be difficult to identify the actual cluster it falls under, therefore the algorithm should allow overlapping clusters. In conclusion the BoW model could be used in this research to determine the relationship between customers’ drinking habits, though the any-of-problem can crop up thereby compromising the result of relationships and the ones mentioned above.

# RESEARCH METHODOLOGY / DESIGN

## Introduction

This section of the document will largely discuss how the research’s design and data collection and analysis was structured with an aim of answering the research question and to meet the research objectives.

The discussion will proceed as follows – it will start with a definition of the various variables that will basically form the basis of data collection, they will be identified from the proposed application on how social networks analysis will be conducted in bistros influenced by the customers. This will be followed by an outline of how the required data was collected for analysis how the data was analysed and lastly the consideration that were made in proposing the new way that enhances business decision making processes in bistros using social network analysis.

The research first question requires a comparison of different technologies of social networks based on the components required for social network analysis with an aim of identifying whether the technologies have exhausted the capabilities of analysing a social network. The second research question necessitates in identification of data that is required when doing social network analysis in a pub setting with an aim of critically determining which data is critical and which one is not. The third research question seeks to get a review of how the data collected can be analysed using a model in social network analysis. The fourth question basically seek to identify a practically way in which the data can be collected analysed and later presented to the business owners.

## Research design

The research design used was the quasi-experimental research design. Shuttleworth (n.d) describes quasi-experimental research design as a form of experimental research that basically involves selection of groups, which are not randomly selected then the groups identified can be assigned a variable which will be tested on them.

Example, in order to perform drinking analysis, a group of party people can be arbitrarily defined and divided by what they love drinking after which a variable is put to test the different groups over a period of time.

According to cf. Grant & Wall (2009), the strong points of this design are when randomization to treatment conditions is not possible and the other point is that it takes advantage of un-controllable environmental events.

The research will use this kind of research design to identify influences of testing variables between drinking groups which will have the intention of determining or predicting under what circumstances people change their norm to another. The quasi-experimental design best suited this research and therefore the research was obliged to establish a procedure for obtaining, analysing and eventually presenting the relevant information.

## Definition of variables

As stated in the previous chapter the nature of study is going to be based on the foundation described by the illustrations on Social Network Analysis (SNA) applied in the various business industries, this means that the study will adopt the characteristics from the illustrations then use it to meet the outlined objectives of the study. The characteristics defines some sort of structure if used can be of importance in identifying the variables of the study.

The structure is as follows; a) The Community - customers going to the bar, b) relationships – can be identified from the frequency of the visits, the alcohol products customers get and also the quantity of the alcohol products the people take, c) limited resource – conducting promotions in a bar is a limited resource because they can only be done one at a time, d) lastly business advantage – the advantage towards the business is that they need to make the customers loyal to them and by this they need to ensure that the products they sell make them more money through customized promotions.

In order for the business advantage to be realised the limited resource should be considered in the marketing strategy as it will enable the business be able to make decisive decisions towards achieving the advantage. The basic features describing the limited resource can be determined from how the community relationship is identified, which from the point of view of the study they form part of the variables that will be used in the research. These variables are: a) different categories of products in the pub, b) quantity consumed by an individual lastly c) frequency of visits in the club

### Alcohol products / categories

These are products that contain alcohol content in them, for this case it is the products sold at a bar or night club. These products vary in quantity and how they are brewed; examples of categories are beers, liquor, wine, and many others.

### Frequency

The frequency refers to information about how often a customer consumes a particular category of products or how often does the customer combine one category to another, i.e. does the customer always take a particular category e.g. beer or does he or she combines at times e.g. beer, whiskey or liqueur.

## Unit of analysis

The unit of analysis of in this research come from the similarities identified by the Cosine similarity rule that will try and run through the available dataset to come up with similarity indexes which could be grouped and analysed to create useable information

## Data collection methods

The variables of the study as described in definition of variables section in this chapter are identified through studying the profiles of customers who visit the bar. The method for data collection is through the use of the platforms already existing within the club. The data collected will be used to identify relationships between customers who purchase alcohol products from the club.

## Data analysis

Data collected will be analysed as and when an algorithm of categorization is ran through the data available in the database of Tamasha. This analysis will be looking for matching order records between one customer and another order histories therefore showing the chances the two individuals match and the probabilities the customer is going to order the predicted product through knowing information about the orders of the customers.

# Presentation and Analysis of Findings

## Comparison of Sharing between categories

### Determining which categories are most sharable

Figure 4‑1 Ratio of people consuming one product to other products

The figure above considers a number of aspects that creates important information to be used in this study. The first item is the total number of customers who consumes a particular product, this is represented by the blue marker named total. The second item is the marker in red representing a subset of customers who consumes a particular product and also consume another product, lastly the percentage marker in green representing the average number of customers consuming a product and another product.

Considering the items mentioned above, it is noticeable that the other drinks bar in red do not go above the total bar in blue, this remains the same in all cases because as mentioned the total bar in blue contains the whole number customers consuming a product, the other drinks bar in red shows a subset of the whole number who consumes other products offered by the bar. The other noticeable information is based on the ratio of how many customers consume one product and another product represented by the percentage bar in green. Taking an example of beer and whiskey, it is noticeable that beer consumers are reasonably many compared to whiskey consumers denoted by the blue bar. It is also noticeable that the percentage bar in green marks very low in beer compared to the one in whiskey, meaning that the number of customers who take beer and another product are fewer compared to the ones who take whiskey and another product. Therefore it can be concluded that it is easier to combine whiskey and another product compared to combining beer and another product during promotions as the information tells us that not many beer lovers mix products often.

### Determining which percentile to use

Figure 4‑2: 70th percentile is the ideal percentile to use

The figure above shows a representation of percentile range taking into account the percentage similarities between consumers and the product they consume. The percentiles is derived from the degree of similarity between two consumers through the use of the Cosine similarity rule described in Chapter two of the document in identifying the relations of what products Consumer A happens to consume that Consumer B consumes. The information about the relations has been represented by the bars showing the number of consumers who share a product at different levels of similarity and per product i.e. the navy blue bar in whiskey shows the number of consumers who consumes whiskey and another product at the range of 70 – 79 percentiles.

The averages of the percentiles described by Average and Average 2 lines have been derived from calculating the average of the data in ranges 90 – 100, 80 – 89, 70 – 79 and 90 – 100, 80 – 89, 70 – 79, 60 – 69 for Average and Average 2 respectively. This results to determining the percentile to which significant information can be retrieved from therefore meeting the purpose of the graph. According to the graph it is noted that when the percentile is reduced away from 100% there is little information change, meaning that at the point where information changes slightly then it could explain why information could be less significant. Looking at the results on the averages i.e. Average and Average 2 it is noted that at 60th percentile represented by Average 2 does not change much when compared to the 70th percentile average represented by Average. Therefore it can be concluded that the study will best use the 70th percentile average and over to meet the objective of the study.

## Determining sharing between categories

### Sharing categories at 90th Similarity

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # | Category Name | Clients | Whiskey | Wine | Vodka | Soda | Beer | Rum | Brandy | Gin | Liqueur | Tequilla | Energy Drinks | Cognac | Juices | Cigarattes | Water | Tea |
| 1 | Whiskey | 15 | 8 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 2 | Wine | 6 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Vodka | 6 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Soda | 22 | 2 | 0 | 0 | 14 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | Beer | 56 | 1 | 0 | 0 | 2 | 48 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| 6 | Rum | 3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | Brandy | 13 | 0 | 0 | 0 | 0 | 2 | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Gin | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | Liqueur | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | Tequilla | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | Energy Drinks | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | Cognac | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | Juices | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | Cigarettes | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | Water | 9 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 |
| 16 | Tea | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4‑1: Similarity table at 90th similarity

The table above shows a relation at the 90th - 100th Similarity level showing the number of customers shared between two categories of products, i.e. the number of whiskey consumers who also take soda are 2. The reading can be confirmed if you take the same example and compare the value of at the center of the whiskey and soda to the value of soda to whiskey both values should read 2.

### Sharing categories at 80th Similarity

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # | Category Name | Clients | Whiskey | Wine | Vodka | Soda | Beer | Rum | Brandy | Gin | Liqueur | Tequilla | Energy Drinks | Cognac | Juices | Cigarattes | Water | Tea |
| 1 | Whiskey | 15 | 13 | 1 | 2 | 5 | 5 | 1 | 4 | 0 | 1 | 1 | 2 | 0 | 0 | 2 | 2 | 0 |
| 2 | Wine | 6 | 1 | 5 | 0 | 2 | 2 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 |
| 3 | Vodka | 6 | 2 | 0 | 5 | 2 | 2 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | 0 |
| 4 | Soda | 22 | 5 | 2 | 2 | 19 | 5 | 0 | 5 | 0 | 1 | 1 | 2 | 0 | 0 | 2 | 1 | 0 |
| 5 | Beer | 56 | 5 | 2 | 2 | 5 | 54 | 1 | 5 | 0 | 1 | 2 | 2 | 0 | 0 | 2 | 4 | 0 |
| 6 | Rum | 3 | 1 | 0 | 0 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | Brandy | 13 | 4 | 2 | 2 | 5 | 5 | 2 | 11 | 0 | 1 | 1 | 2 | 0 | 0 | 2 | 1 | 0 |
| 8 | Gin | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | Liqueur | 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 10 | Tequilla | 3 | 1 | 0 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 0 | 0 |
| 11 | Energy Drinks | 3 | 2 | 1 | 1 | 2 | 2 | 0 | 2 | 0 | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 0 |
| 12 | Cognac | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | Juices | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | Cigarettes | 3 | 2 | 0 | 2 | 2 | 2 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | 0 |
| 15 | Water | 9 | 2 | 2 | 0 | 1 | 4 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 9 | 0 |
| 16 | Tea | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4‑2: Similarity of categories at 80th similarity level

The table above shows a relation at the 80th - 100th Similarity level which contains the number of customers shared between two categories of products. When comparing the Table 4-1 and Table 4-2 it can be noticed that the number of customers sharing the different categories of products increases significantly meaning that the data grows down the similarity levels though at some point the data can be less significant as explained in Figure 4-2.

### Sharing categories with respect to the ratio of the number of shared to the total number of customers per category

|  |  |
| --- | --- |
|  |  |

Figure 4‑3: Relationships of categories with respect to the percentage number of shared customers to total number of customers per category

The table above describes the how sparse a category can be with regards to the ratio of the shared subset of customers who consume a category to another and the total number of customers each category has. The purpose of this chat is to explain how significant the ratio of shared customers to the total number is in regards to the weight it applies when determining which category could be targeted for promotion.

From the scatter graph, the point furthest from the Y axis on the X axis, marked as Beer shows that there are many customers who love consuming from the category though with very little of the total take another product. The implication of this relation is that when considering a category which can be combined during promotions the ideal category would be the one that has a ratio higher since there is a high probability that customers at that category combine with other categories i.e. the category that would suit combinations of products would be the ones far away from the X axis and closer to the Y axis e.g. Tequila.

## Determining Relationships between individual Categories

### Comparison at the 90th Similarity

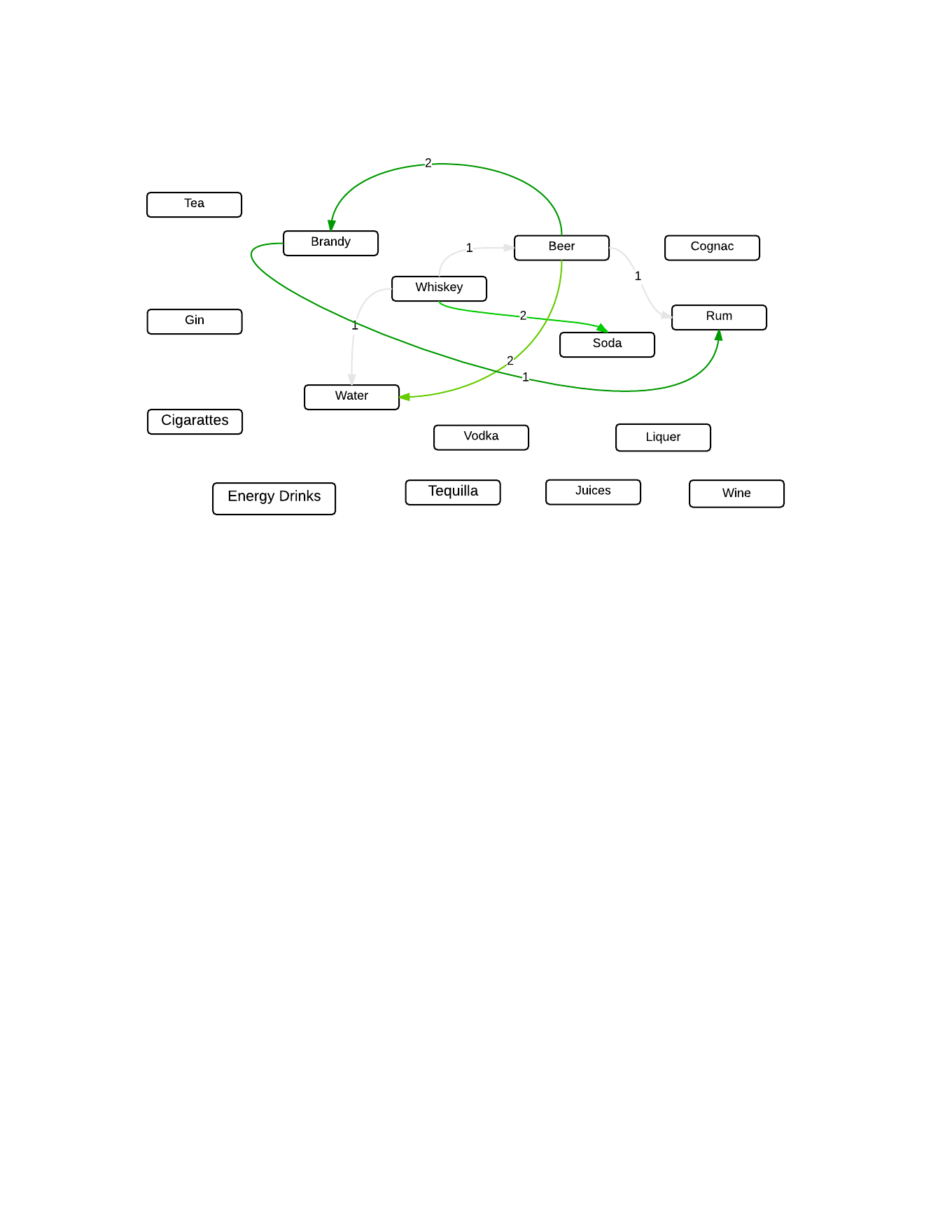


Figure 4‑4: Comparison at 90th Similarity

The figure above shows how categories relate to each other with respect to the number of consumers who devours these categories. The representation of this relation is characterized by a line weighted to the number of consumers who consume either of two categories joined by the line, for instance the line from Beer to Brandy weighing as 2 mean that, 2 of Beer consumers also take Brandy. Figure 4-3 is a graphical representation of the data in the relation of Table 4-1 at the 90th similarity level which also shows that only 2 customers consume Beer and Brandy.

Since the purpose of the figure is to determine whether the relations displaying in the figure can amount to some descriptive value which can be used to notice a pattern in the data represented, it can be concluded that at this point the representation does not show any pattern with regards to how consumers devour more than one category. Some of the pointers showing the information is not enough are a) the relations between categories only cover 1/3rd of the total categories and b) looking at the number of consumers who relate at the 90th Similarity level, only a few of them are in this relations, therefore more it is necessary to move a step to the 80th Similarity level to see the structure of the relations.

### Comparison at 80th to 100th Percentile

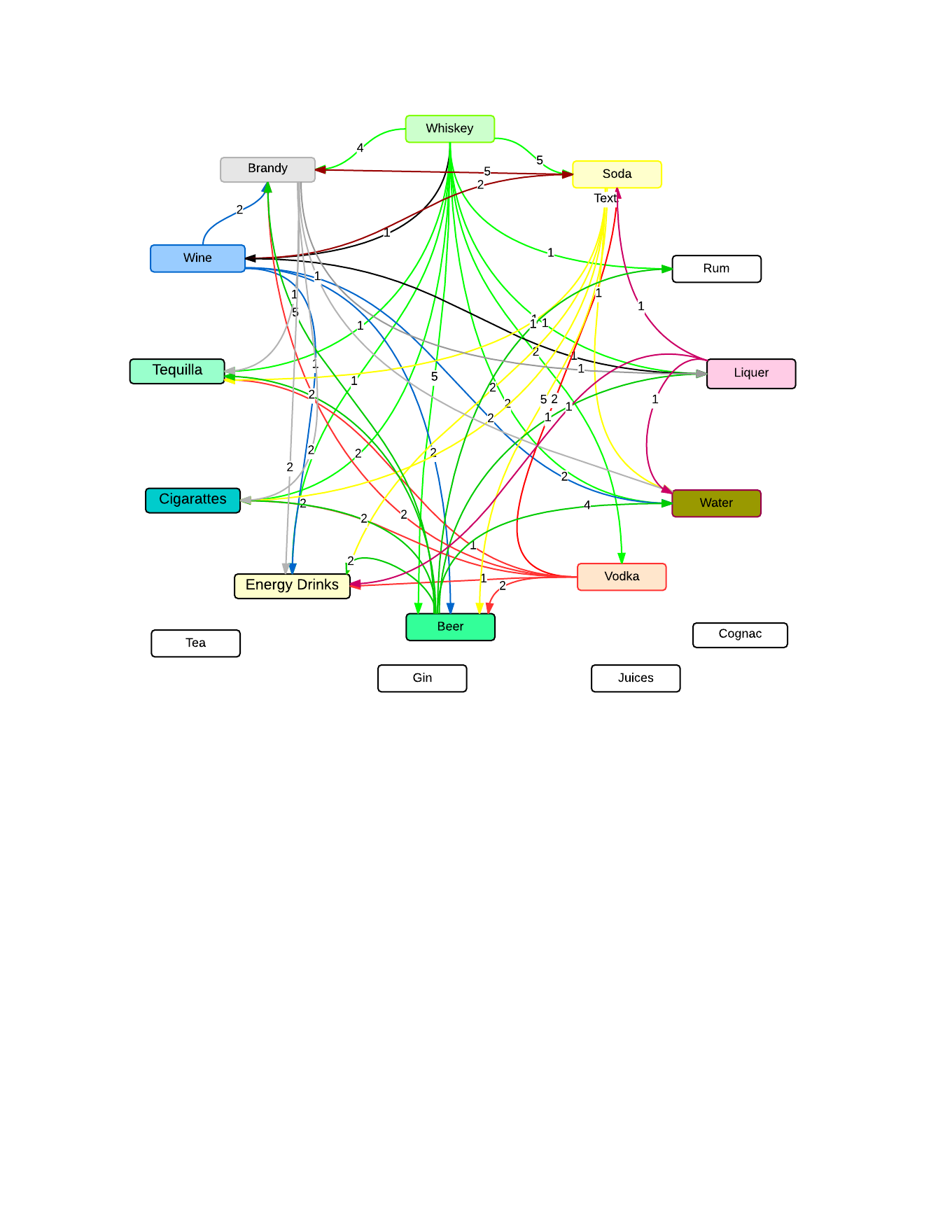


Figure 4‑5: Comparison at 80th Similarity

The figure above shows how categories relate to each other with respect to the number of consumers who devours these categories at the 80th similarity level and above. This means that the information captured on the graphical representation refers to all the similarities between categories with respect to consumers who consumes a category to another at the level of 80 – 100th similarity level. The relationships captured on Figure 4-4 has been generated from the dataset on Table 4-2 whereby the values on the arrows show the relation a category has to the other.

The figure displays color coded categories and also the representation of the relations i.e. line have been color coded displaying where the relation start from to where it relates to. When reading this figure, for instance you want to find the all the relations Beer has, check the green colored lines out of the category marking all the relations that beer has to the other categories and all the other colored categories showing the relations another category has to Beer i.e. for Soda and Beer its marked by a yellow line.

From this table it is noticeable that the relations the dataset generates is so enormous i.e. the lines in and out of a category. This representation also contrasts with the first representation i.e. Figure 4-3 whereby only few categories were connected but in this one there are quite a number of relations being captured at the level of similarity being selected. This can deduce that the range of similarity affects the structure of the network diagram generated from the data, i.e. the greater the range more relations are identified. Just from the Figure 4-3 it is seen that the information is too crowded therefore it can be related that when the similarity level is reduced from 70 -100, the relationships will increase therefore making it harder to create meaningful information from the data.

To make the concept of similarity clear, the next diagram will show an extracted section of Figure 4-3 to show relations of the categories at the level of 80Similarity. The first example we will use is the relation of Cigarettes and Brandy:

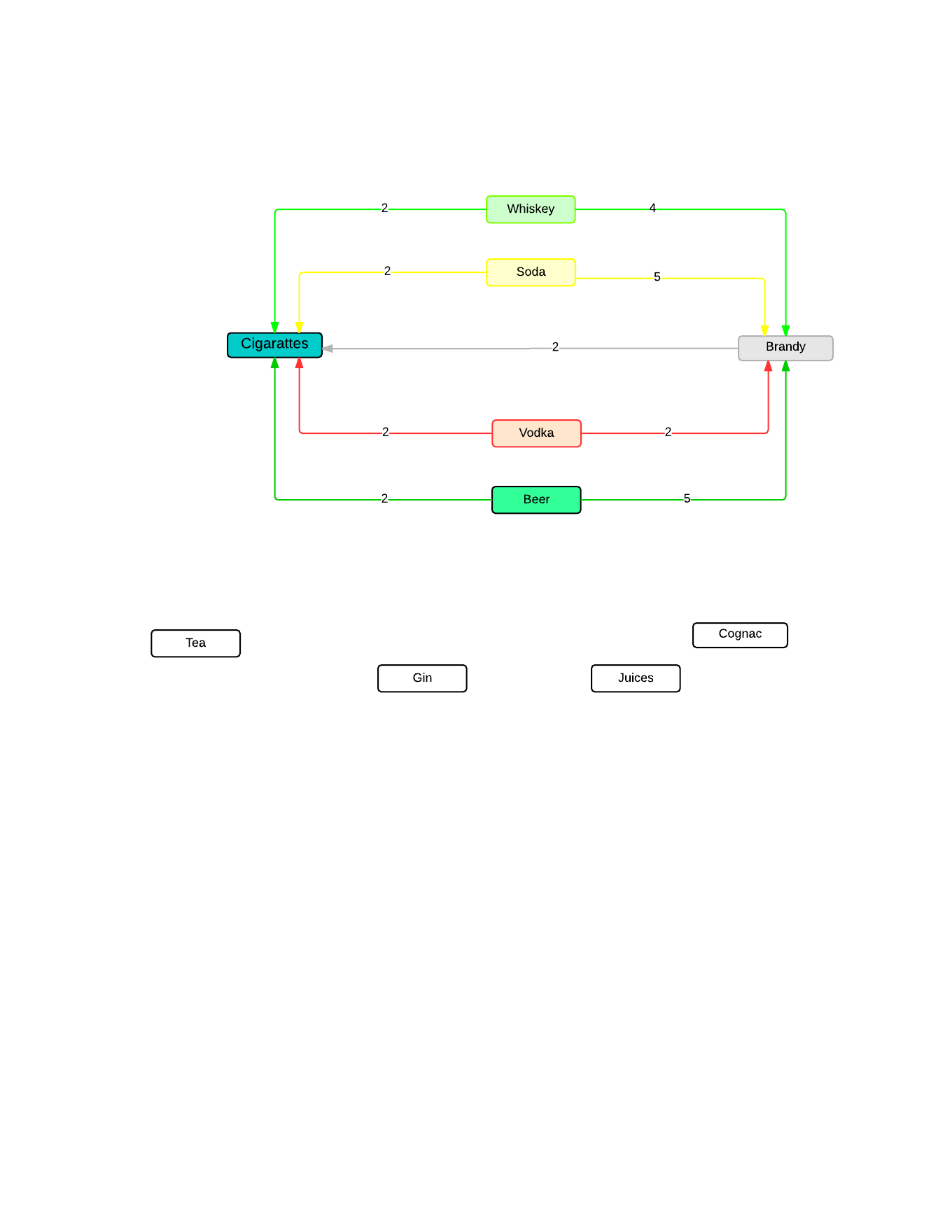


Figure 4‑6: Cigarettes and Brandy relationship

The figure above is an extracted section from Figure 4-5 showing all the connected products to the relation between Cigarettes and Brandy. The numbered label on the arrows shows the number of customers who consume one product in relation to the other product. i.e. The arrow from Brandy to Cigarettes has a value of 2, meaning that 2 people who consumes Brandy also consumes Cigarettes, same to the relation between Brandy and Whiskey where the relations shows that 4 of Whiskey takers also take Brandy.

The purpose of this table is to determine, how relationships can be used to identify the most valuable combinations of categories when conducting a promotion of Brandy and Cigarettes. From the Figure 4-5 an ideal combination should have the most number of customers who will purchase the products therefore increasing the revenues they collect in that day. To identify the greatest number of people between the categories of Brandy and Cigarettes we would perform the following steps:

1. List down the combinations of categories connected to Brandy and Cigarettes.
2. Include the weights each of them have as represented in the Figure 4-5
3. Calculate totals of all the combination and indicate them beside each of the combination
4. Arrange the categories in ascending order of the total customers the combination brings
5. The combination with the highest number of sharing customers would be the ideal combination for the business to target. The illustration is as follows:
6. List down the combination of categories connected to Brandy and Cigarettes
   1. Cigarettes – Whiskey – Brandy
   2. Cigarettes – Soda – Brandy
   3. Cigarettes – Vodka – Brandy
   4. Cigarettes – Beer – Brandy
7. Including the weight of each combination
   1. Cigarettes -- 2 -- Whiskey -- 4 -- Brandy
   2. Cigarettes -- 2 -- Soda -- 5 -- Brandy
   3. Cigarettes – 2 -- Vodka -- 4 --Brandy
   4. Cigarettes -- 2 -- Beer -- 5 – Brandy
8. Calculating the total of each combination
   1. Cigarettes -- 2 -- Whiskey -- 4 -- Brandy : 2 + 4 = 6
   2. Cigarettes -- 2 -- Soda -- 5 – Brandy : 2 + 5 = 7
   3. Cigarettes – 2 -- Vodka -- 4 –Brandy: 2 + 4 = 6
   4. Cigarettes -- 2 -- Beer -- 5 – Brandy: 2 + 5 = 7
9. Arranging the relations with regards to the totals
   1. **Cigarettes – Beer – Brandy OR Cigarettes – Soda – Brandy :** Total 7
   2. **Cigarettes –Whiskey – Brandy OR Cigarettes – Vodka – Brandy :** Total 6
10. In conclusion the best alternative or a complementary category to add when doing a promotion of Cigarettes and Brandy would be the first combinations of option i, though when we factor the explanation given on Figure 4-3, the ideal combination would be adding Soda as complement to the combination of Cigarettes and Brandy since the relation shows that there are high chances of Soda consumers combining with another product compared to consumers who take Beer.

# Conclusions and Recommendations

## Conclusions

In conclusion, the study has been able to attain all the objectives set at the beginning of the chapter since it has been able to illustrate how various Social Network Analysis (SNAs) have been applied in different sectors of the economy, i.e. in the case of Supermarkets, the use of loyalty card representing one way the business applies the concept of SNA in order to meet the objectives of the organization. The illustration of Mobile network positioning the use of information about people’s movement when considering data allocation to different place at different times of the year has created positive impact toward resource utilization and maximization of available data without wastage.

Secondly the study has been able to outline some of the algorithms used when studying Social Networks (SN) and habit or people, which include the Cosine Similarity rule which attempts to find similarities in a data set, therefore deductions of how similar an occurrence of a particular set of data can be used in prediction of future occurrence of the same set of data, also the Bag of Words (BoW) model which tries to identify relations in different arrangements of datasets and stating whether they words are similar or not. For instance the study used the Cosine Similarity rule in identifying similarities between different categories of products which helped in determining different relations of different sets of data with at different ranges.

The study has been able to apply some of the algorithms example the Cosine Similarity rule to identify recommended scope of datasets required to person the study i.e. it identified the 70th similarity level to be significant break off point when studying data in a bar as the information below this value did not show significant difference. The other one is the use of the vice versa of the rule of finding the shortest route between the categories which identified information on how the business could combine products in order to make promotions more effective and durable in the long run.

Last but not least the study was able to show how identifying similarities in categorize could be used in creating a promotion strategy targeting to identify different combinations of products that could reap reaping more benefits in terms of profits through just studying the relations these categories have, relating relevant weights to each category and lastly obtaining the best combination that will reap the most benefits.

## Recommendations

This research proposed a way in which a bar business could use the neglected data about customer orders to be a factor in the study of human behaviours and preferences with an aim of identifying relations and later relating the relations to different set of implications that could lead to important information being created. It is the hope of the author that the two algorithms i.e. Cosine similarity and also the shortest path rule can be tested in a different bars and restaurants to see its applicability and faults when applied in different contexts.

# REFERENCES

Ayyub S. and Jain S. (2009). Data allocation in a mobile computing network with minimized response time and communication cost. Journal of Computer Science & Technology, 2(2), 141-144

Bellezzi J. & Bristol T (2004). An assessment of supermarket loyalty cards in one major US market. Journal of Consumer Marketing, Vol 21(2), pp.144 - 155

Butts T. C (2008). Social network analysis: A methodological introduction. Asian Journal of Social Psychology, 11(1), doi: 10.1111/j.1467-839X.2007.00241.x

Elaina R. (2013). *Getting started with Business Intelligence.*Retrieved from [http://blog.glider.com/getting- started-business-intelligence/](http://blog.glider.com/getting-%20started-business-intelligence/)

Hannover research (2011). Consumer loyalty programs

Idiro Analytics (n.d). Idiro Analytics - Social Network Analysis for Online Gaming. Retrieved February 26, 2015, from http://www.slideshare.net/Idiro/sna-for-gaming-jan-2015

Jain. Y., and Nandanwar. K., A (2014), A Theoretical Study of Text Document Clustering. International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2246-2251

Joseph A. Bellizzi and Terry B., (2004),"An assessment of supermarket loyalty cards in one major US market", Journal of Consumer Marketing, Vol. 21 Iss 2 pp. 144 – 154

Applying social network analysis (SNA) to gaming. (n.d.). Retrieved February 26, 2015, from <http://www.iidi.napier.ac.uk/c/casestudies/casestudyid/13368224>

Mulcahy R.(2014). Business Intelligence definition and Solution. Retrieved From

[http://www.cio.com/article/40296/Business\_Intelligence Definition\_and\_Solutions](http://www.cio.com/article/40296/Business_Intelligence_Definition_and_Solutions)

Oinas-Kukkonen, H., Lyytinen, K. and Yoo, Y. (2010), "Social networks and Information Systems: Ongoing and Future Research Streams", Journal of the association for information systems, Vol. 11 Iss 2.

Olexova, C (2014). Business intelligence adoption: a case study in the retail chain.

Olszak, C.M., & Ziemba, E. (2006). Business intelligence systems in the holistic infrastructure development supporting decision-making in organizations. Interdisciplinary Journal of Information, Knowledge and Management, 1, 47- 58. doi:10.1.1.99.8329

Sree. S, and Murthy. R, (2012), Clustering Based On Cosine Similarity Measure, International Journal of Engineering Science & Advanced Technology, Vol . 2 Iss 2 pp.508 -512

Solomon. N, (2004) Communications of the Association for Information Systems: Vol. 13, Article 15.Available at: <http://aisel.aisnet.org/cais/vol13/iss1/15>

Tsai (2012). Bag-of-Words Representation in Image Annotation, International Scholarly Research Network , Vol. 2 , doi:10.5402/2012/376804

Worthington. S., & Fear. J (2009) The hidden side of loyalty car programs

Zhong, Y (2010). Advanced Intelligence: Definition, Approach and Progresses, International Journal of Advanced Intelligence, 2 (1), 15-23