CELL FORMATION: A REAL LIFE APPLICATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BAŞAR UYANIK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

SEPTEMBER 2005

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan ÖZGEN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Çağlar Güven
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Canan Sepil
Supervisor

Examining Committee Members

Prof. Dr. Sinan Kayalıgil	(METU, IE)	
Asst. Prof. Dr. Ayten Türkcan	(METU, IE)	
Prof. Dr. Refik Güllü	(BOUN, IE)	
Assoc. Prof. Dr. Canan Sepil	(METU, IE)	
Assoc. Prof. Dr. Erol Sayın	(METU, IE)	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Başar, Uyanık

Signature :

ABSTRACT

CELL FORMATION: A REAL LIFE APPLICATION

Uyanık, Başar
M.S., Department of Industrial Engineering
Supervisor : Assoc. Prof. Dr. Canan Sepil

September 2005, 104 pages

In this study, the plant layout problem of a worldwide Printed Circuit Board (PCB) producer company is analyzed. Machines are grouped into cells using grouping methodologies of Tabular Algorithm, K-means clustering algorithm, and Hierarchical grouping with Levenshtein distances. Production plant layouts, which are formed by using different techniques, are evaluated using technical and economical indicators.

Keywords: Tabular Algorithm, K-Means Algorithm, Hierarchical Grouping, Levenshtein Distance, Printed Circuit Board (PCB)

ÖZ

HÜCRE YERLEŞİMİ:
BİR GERÇEK HAYAT UYGULAMASI

Uyanık, Başar
Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Assoc. Prof. Dr. Canan Sepil

Eylül 2005, 104 sayfa

Bu çalışmada, Baskı Devre kartları üretiminde dünya çapında hizmet veren bir firmanın üretim sahasının yerleşimi problemi incelenmiştir. Tabular Algoritması, K-means ve Levenshtein uzaklıkları ile Hiyerarşik gruplandırma yöntemleri kullanılarak makineler hücrelere ayrılmıştır. Farklı tekniklerle oluşturulan üretim sahası yerleşim planları teknik ve ekonomik göstergeler kullanılarak değerlendirilmiştir.

Anahtar Kelimeler: Tabular Algoritması, K-Means Algoritması, Hiyerarşik Gruplandırma, Levenshtein Uzaklıkları, Baskı Devre Kartları

To My Father

ACKNOWLEDGMENTS

I would like to express gratitude and sincere thanks to my supervisor Assoc. Prof. Dr. Canan Sepil for her patience, encouragements, guidance, suggestions, and insight throughout the research.

I would like to thank to the General Manager of Solectron, Mehmet Güvey, for his guidance and support.

I would like to thank to my sister Bahar Bayrak for her motivations and suggestions that reached me to the solution. I would also like to thank Birol Bayrak for his contributions.

I would like to thank to my best friend Taylan Zengin for his invaluable friendship and motivations.

I would like to thank to my dear mother, Zümrüt Uyanık, for her understanding, motivation and most importantly for her love.

Lastly, I would thank to my dear father, Yaşar Uyanık, for his belief in me.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

INTRODUCTION

Manufacturing companies all over the world face the same global resource-finding problems in modern business marketplace. In order to survive in this environment companies are required to be committed to better quality, higher productivity, more efficient use of energy and tougher price competition. Thus, the right manufacturing strategy is important to meet the challenges of today's and future's markets.

In such a competitive marketplace, companies try to reduce production costs. Layout planning is vital for efficient utilization of available resources of a company. Therefore, in order to reduce the production costs the design of a plant layout is critical. The plant layout problem is designing a new facility or redesigning the existing facility locations. The process of developing plant layouts contains the elements of both art and science. The artist's dependence on creativity, synthesis, and style is very evident in designing plant layouts. Similarly, the scientist's use of analysis, reduction, and deduction are essential in designing plant layouts. The plant layout problem is fundamentally different from an optimization problem because it is a design problem. Furthermore, solutions to the plant layout problem depend heavily on the use of synthesis, rather than been directly driven from analysis. These distinctions are important and should not be treated lightly.

Electronics-related products have become one of the largest industries in the world. Among the most challenging operations management issues in this industry is the design of electronic assembly systems to be used for the production and assembly of building-block subsystems for electronics products. As the pressure of competition to reduce cost and lead times increases, the manufacturers adopt automated assembly systems to produce large annual volumes of high variety of Printed Circuit Board (PCB). Effective management and design of electronic assembly systems requires the development of complex integrated design and production planning support systems.

In this thesis, a real life layout problem is studied. The scope of the problem is planning the layout of a Printed Circuit Board (PCB) manufacturing plant, which is traditionally designed to produce more than 400 products having different routes.

The demand forecast of the products changes weekly in the company. The huge amount of data and the rapid changes in the parameters make the problem difficult to solve. In order to accomplish the best solution the philosophy called Group Technology is used.

Group Technology (GT) capitalizes on similar recurrent activities by bringing together and organizing common tasks to improve productivity. GT offers a system approach to the reorganization of traditional complex job shop and flow shop manufacturing systems into cellular manufacturing systems. Cellular Manufacturing system (CM) is an application of the GT in manufacturing systems.

CM is ideal for small and medium-size batch production environments. The first and the most important stage in the design of CM is the part-family machine-group formation problem that is known as Cell Formation problem. In this thesis, different approaches to Cell Formation problem are discussed. Because of the complexity and size of the data, the techniques are selected carefully. First one is the Tabular method which is a simple use of technique is applied to the problem. The K-Means algorithm is one of the other techniques used. In order to determine the similarities between the production routes Levenshtein Distance algorithm is used. Hierarchical Clustering is applied to the distances in order to classify the products with similar routes.

CHAPTER I: THE PROBLEM

1.1. A Worldwide Electronics Company

Founded in 1977, the company is a leading electronics manufacturing services (EMS) company offering a full range of integrated supply-chain solutions for the world's leading electronics original equipment manufacturers (OEMs). The company's integrated design, prototyping and test, manufacturing, packaging, systems assembly, global distribution and post-manufacturing services offer customers competitive outsourcing advantages. Some of these advantages are access to advanced manufacturing technologies, shortened product time-to-market, reduced total cost of ownership and more effective asset utilization.

The company provides integrated services to the world's leading OEMs. The company delivers a full range of services to its customers in a variety of industries, like automotive (Airbag control modules, car radio navigation systems, engine and ignition control modules, etc.), communications (cellular infrastructure equipment, core and edge routers and ethernet switches etc.), computing (mainframe computers, PCs and notebooks, server etc.), consumer electronics (cellular handsets, game consoles, personal video recorders etc.) and industrial (home appliance electronic controls, process automation equipment, security control systems etc.).

The company manufactures more than 400 different electronic circuit assemblies for telecommunication products and provides repair and return services. The physical area of the plant is 4500 square meters. 3500 square meters of this area is the manufacturing area. The remaining is the warehouse and the offices.

It is necessary to give a short description of how the assembly of PCBs is organized at the company plant. There are number of assembly lines. SMT Processes is the main assembly line of the plant. It is composed of dispenser, screen printer, chip shooter, precision placer, and oven. There are three offline processes; wave solders, axial inserter and DIP Inserter. The test operations have two stages; in circuit test systems and functional test systems. The functional tests are product specific and dedicated microprocessor controlled test systems.

1.2. The Need for Layout Planning

The company has a plant area of 3500 m^2. There are 79 machines working in the plant area. The company can produce more than 400 types of products that have different operation sequences. The material handling costs within the factory are higher than desired. Even transportation of one single part of a machine needs a carriage car with a supervisor care, which causes high labor costs. The products are fragile and care must be taken when moving from one place to another. The company aims to reduce the material handling in the plant. This makes the design of the plant very critical and important. On the other hand, the company tries to make the production as quickly as possible to cover all customer demand. The company is aiming to shorten production time for each part. The machine based production time for each part is fixed. Nevertheless, it is possible to shorten the carriage time from one machine to another.

The company's production is based on orders. According to the incoming orders, the production schedules are prepared weekly. Production based on orders that changes weekly makes the problem environment dynamic. This becomes one of the most important conflicts of the problem. On the other hand, when the operation sequences of the products are analyzed some similarities are observed.

The company, observing the symptoms of high material handling costs, evaluates redesigning the layout of the plant. However, some of the machines in the plant have very sensitive accuracies that can only be performed by a professional foreign team. Thus, the relocation decision of these machineries is reluctantly deferred. The aim of this study, considering the constraints, is to generate different layout alternatives.

The company especially needs to know the cost for designing the new layout and then will decide to change the current layout or not. During the analyses, we will try to find the trade off between the cost of designing the new layout and the reduction in the material handling.

1.3. Outline of the thesis

In Chapter 2, the Literature about the group technology and cell formation is mentioned. In addition, the literature of the techniques proposed in this thesis is discussed. These techniques are Tabular Method, K-Means Algorithm, Levenshtein Distance Algorithm, and Hierarchal Clustering Algorithm. Lastly, the literature review of performance evaluation techniques is included in this chapter.

In Chapter 3, the proposed methodologies for designing the layout are discussed. First, Tabular Method is studied in detail. Expanded Tabular method is the second technique that is mentioned in this chapter. K-Means and the Hierarchical Algorithm are also mentioned.

Chapter 4 covers the results of the techniques. The comparison and the calculations are analyzed. The conclusion is included in Chapter 5.

CHAPTER II: LITERATURE SURVEY

The long-term goals of a manufacturing enterprise are to stay in business, grow, and make profits. What is needed is the right manufacturing strategy to meet the challenges of today's and future markets. With an escalating worldwide competition and shrinking product life cycles, manufacturing managers are increasingly replacing their traditional job shops with more efficient and responsive manufacturing cells. These developments emerged in the formation of philosophy called Group Technology (GT)

Group Technology (GT) is a philosophy that capitalizes on similar recurrent activities by bringing together and organizing common concepts, principles, problems, and tasks to improve productivity. In the ensuring 20 years, this philosophy has spread throughout the manufacturing world. This philosophy largely based on the concept of Group technology (GT), a theory of management based on the principle that similar things should be done similarly. GT may be applied to all activities including administrative.

As we reached the 21-century, many of the manufacturers find themselves in a state of transition from the traditional mass-production model to a more flexible and customer-responsive operations environment. A combination of changes has had the effect of breaking down a lot of old rigidities and barriers that existed in traditional manufacturing organizations.

The major obstacle to overcome in batch manufacturing is the enormous variability of product characteristics. Such variability requires that universality be contained in the system design, which usually results in a very complex, and expensive design. The sensible approach to the problem is to reduce or restrain the variability in product characteristics, thus leading to a less complex and costly manufacturing systems design. This approach requires a method of product and process analysis that will yield a rationale for standardization. Group technology must be utilized as a philosophy that permeates the entire enterprise. For GT to grow and to grow successfully, it must become an integral part of successful systems (Burbidge, 75).

Batch Manufacturing is viewed as one of the most difficult problems to overcome for low-cost products. About 75 percent of all machined components today are produced in batches of 50 or fewer. The impact of this is the inability to take advantage of low-cost mass-production techniques to produce parts. The cost to build a product in small batches can be 10 to 100 times higher than its mass production cost (Arieh, 1998).

The majority of the factories today are arranged as functional layouts with a grouping of common process equipment, as opposed to being laid out for process flow. The result of a factory layout based on grouping common equipment is that products must flow from department to department through the manufacturing process. The result may be dramatic in production scheduling and control as well as excessively long queue times. The result is higher inventory cost, larger scrap due to material damage, more overhead personnel to handle the production control tasks, and less customer satisfaction because of excessively long delivery times (Burbidge, 1975).

The group technology is a manufacturing concept that seeks to identify similar parts to take advantage of their similarities in manufacturing parts by classifying these parts into groups and applying similar operations to the parts in each group.

Group technology can be used to aid the design process for new manufacturing technology systems. It provides to the designer the ability for target selection of opportunity. By identifying and designing to certain characteristics of the product, the manufacturing systems designer is able to apply Pareto rule to create a new design that will handle the majority of the product within a minimum system cost. In this study, Pareto rule is used in order to eliminate the data (Snead, 1989).

GT philosophy offers a systems approach to the reorganization of traditional complex job shop and flow shop-manufacturing systems into cellular manufacturing systems. Cellular Manufacturing (CM) is an application of the GT philosophy in manufacturing systems. CM is ideal for small and medium size-batch production environments. Cellular manufacturing is one of the major uses of group technology philosophy. The first and the most important stage in design of Cellular Manufacturing (CM) systems is the part-family machine-group formation problem that is known as cell formation problem.

2.1. Cellular Manufacturing System

Cellular Manufacturing (CM) is one of the major applications of group technology. Cellular Manufacturing is ideal for small and medium size batch of production environments. CM provides reducing manufacturing costs, improving quality and reducing the delivery lead-time of products in a high variety-low demand environment. Therefore, CM has recently begun to receive great attention worldwide.

A manufacturing cell is a collection of dissimilar machines or manufacturing processes dedicated to a collection of similar parts and Cellular Manufacturing is said to be in place when a manufacturing system encompasses one or more such cells. Wemmerlov and Johnson (1997) reported a study of plants involved with cellular manufacturing. A target population of high-probability users received mail questionnaires designed to collect responses related to characteristics of industry cells and the firms that have implemented them. Forty-six plants supplied detailed data on 126 of their cells, including reasons for establishing them, types of operations performed in the cells, problems faced and lessons learned during implementation, and achieved performance improvements.

There are primarily three steps in cellular manufacturing systems design:

1- Cell formation,
2- Machine layout
3- Cell layout

Among them, cell formation is the first and the most difficult step. It involves identifying part families and machine cells and then allocating each part family to corresponding machine cell.

It is clear that within the group structural decisions, the selection of part and machine types has a particular significance since most subsequent decisions depend on these choices. The term “cell formation” is used to refer to the initial activities in the cell design process dealing with the identification of parts and machines for cellular manufacturing and the evaluation of associated cell properties. However, the implementation of cells could have some disadvantages as compared to traditional functional and product layouts. The disadvantages can arise from the underlying design characteristics of cells and limitations of methods used to design and evaluate cells. These disadvantages are summarized by Irani (1999).

There have been several approaches proposed for manufacturing cell formation in GT. A comprehensive review and discussion on different approaches can be found in Offodille et al. (1994) and Singh (1993). They provide concise reviews of the usefulness and limitations of existing methods for cell formation in CM.

Some approaches for cell formation are focused on minimizing inter-cell movements (Joines et al, 1995; Cheng et al, 1995; Sofianopoulou, 1999). Joines et al (1995) also offers a comprehensive review and classification of techniques to manipulate part routing sequences for manufacturing cell formation. Chu (1990), Boctor (1991) and Chen et al. (1998) studied on maximizing parts and/or machines similarities (or minimizing dissimilarities). In 1992, Venugopal and Narendran studied minimizing cell load unbalances.

Extensive work has been performed in the area of cell formation and numerous approaches have been developed. The machine-part matrix forms the basis of many procedures for cell formation (Mansouri, Husseini & Newrnan, 2000).

.

The cell formation research in literature can be divided into three categories: Grouping part families or machines only, forming part families and then machine cells (Kamrani & Parsaei, 1993), forming part families and machine cells simultaneously (Shafer & Rogers, 1991).

At the highest level, the methods for part family/machine cell formation can be classified as design-oriented or production-oriented. Design-oriented approaches group parts into families based on similar design features. Production-oriented techniques aggregate parts requiring similar processing. (Joines, King & Culbreth, 1995).

2.1.1. Design Oriented Techniques

Design-oriented approaches group parts into families based on similar design features. The remedy for designing cells lies in sorting parts into families that have similar part design attributes and/or manufacturing attributes for specific purpose. The part design attributes include part shape (round or prismatic), size (length/diameter ratios), surface integrity (roughness, tolerance), material type, raw material state (casting, bar stock) etc. The part manufacturing attributes include operations (tooling, milling) and sequences, batch size, machine and cutting fools, processing times, production volumes etc. When these attributes are standardized, it prevents part variety proliferation, and provides accurate planning and cost estimation values. An engineering database, containing information on part design and manufacturing attributes provides a bridge between computers aided design and manufacturing (Singh & Rajamani, 1996).

Classification and coding schemes are design-oriented tools that can be used to implement CM applications. Since part codes are assigned based up on physical geometry, parts having similar design features have similar codes providing a weak connection between part features and machine groups. Classification and coding involves substantial implementation effort and cost. Much perquisite part data must be developed in order to apply design-oriented techniques.

In cases where the part variety is low, a visual /manual analysis by part and drawing can be used to determine the part families. When the part variety is large, to consider all factors it is preferable to code all the parts and classify parts by the code, similarity, or distance.

Design-oriented systems are not as popular as production-oriented systems., Burbidge states that parts may be similar in shape but have to be made in different groups (also called cells) because they may differ greatly in size, tolerance, required quantities or materials (Burbidge, 1991).

Clustering algorithms aim at assigning P parts to f part families while minimizing some measure of distance. A clustering algorithm to group the parts accesses the distance measures stored in a two- dimensional array. In such algorithms, the parts are grouped into a few broad families, each of which is then partitioned into smaller part families and so on until the final part families are generated. The parts are clustered at each step by lowering the amount of interaction between each part and a part family mean or median, to develop a tree-like structure called a dendogram.

For effective formation of part families, several attributes need to be evaluated based on certain priorities. Unlike other cell formation algorithms where only one objective is considered, a multi-objective clustering algorithm evaluates each attribute separately by considering their relative importance (Mansouri et al., 2000).

The multi-objective model is proposed for identifying flexible part families and similar digit set. However, since the method utilizes goal programming, proper selection of priorities is important to obtain meaningful results. Ham and Han as well as Jung have developed a multi-objective cluster analysis tool using design features to form machine cells (Mansouri et al., 2000).

2.1.2. Production Oriented Techniques

There has been a great amount of research in the area of production-oriented techniques, which identify and group parts sharing common processing requirements. Most production-oriented system use route sheets to record the relationship between parts and the machines that process them.

Burbidge (1991) was the first researcher working in this area. Production Flow Analysis (PFA), which was introduced by Burbidge is one of the first and most comprehensively -recognized methodologies.

2.1.3. Array-Based Methods

Array based clustering techniques are considered as the simplest classes of production-oriented cell formation methods. The array based methods group machines and parts without finding a similarity measure. It operates on a 0-1 machine-part incidences matrix performing a series of column and row manipulations trying to produce small-clustered blocks along the diagonal of the matrix. The machine-part incidence matrix, A, consists of elements a_{ij} = 1 if part j requires processing on machine i otherwise a_{ij} = 0. The rows and columns of the incidence matrix are rearranged until a diagonal pattern emerges. These methods are clustering algorithms that sort rows and columns of the machine part incidence c, f matrix according to some rules. Any tightly clustered blocks represent the candidate part families and machine cells, which are formed simultaneously.

The major drawbacks of array-based algorithms are that they do not consider any other information than the machine-part incidence matrix. Furthermore, they do not deal with the number of intercellular movements and the sequence of operations within cells.

There are many clustering algorithms in the literature. Bond Energy Algorithm (BEA) is proposed by McCormick et al (1972) . It maximizes the matrix by rearranging the rows and columns to form part families and machine groups. The algorithm permutes the rows and columns to obtain mutually exclusive cluster of 1's in the matrix, if they exist.

Rank Order Clustering (ROC) that is proposed by King and Nakornchai (1982), is not suitable for large problems. In ROC binary weights are assigned to each rows and column of the part-machine incidence matrix. The algorithm first assigns each row and column of the machine-part incidence matrix its al equivalent. This algorithm simply assigns a binary weight to each row and sorts them in decreasing order according to the corresponding decimal weights, and repeats the same steps for columns. The algorithm continues until no further changes order of rows and columns. The quality of the result is dependent upon the machine-part incidence matrix. Therefore, identification of exceptional elements and bottleneck machines is somewhat arbitrary. In addition, binary representation restricts the size of the matrix.

Modified Rank Order Clustering (MODROC) is proposed by Chandrasekharan and Rajagopalan (1987). They used ROC iteration twice to obtain an incidence matrix containing a rectangular black of 1's at its top-left comer. The rectangular black represents a candidate cell.

Direct Cluster Algorithm (DCA) was developed by Chan and Milner (1982), which was proposed to form tight groups along the incidence matrix. Rather than giving binary weights, the number of 1's in each row is counted as weights and they are sorted according to an increasing order, then the same step is followed for columns, but decreasing order is used in sorting. The algorithm stops when no further changes occur.

Cluster Identification Algorithm (CIA) is proposed by Kusiak and Chow (1987). The algorithm forms machine clusters starting with parts that have maximum subcontracting costs. The objective of the algorithm is to minimize subcontracting exceptional elements to a limited cell size.

Some other algoritihms are Occupancy Value Method by Khator and Irani (1987) ,and Hamiltonian Path Heuristic by Askin et al (1991).

2.1.4. Hierarchical Clustering Methods

Clustering is a generic name for a variety of mathematical methods, which can be used to find out which objectives in a set are similar. The main objective of cluster analysis is to group either objects, entities or their attributes into clusters such that individual elements within a cluster have a high degree of natural association between clusters. Clustering methods interchange rows and columns according to some measures until the initial matrix is transformed into a more structured form.

Hierarchical clustering refers to the formation of a recursive clustering of the data points: a partition into two clusters, each of which is itself hierarchically clustered (Johnson, 1967).

Hierarchical Clustering methods operate on an input data set described in terms of similarity or distance function and produce a hierarchy of clusters or partitions. At each similarity level in the hierarchy, there can be a different number of clusters with different numbers of members. Unlike array-based methods, hierarchical clustering methods do not form machine cells and part families simultaneously. These methods can be described as either divisive or agglomerative. Divisive algorithms start with all data (machines or parts) in a single group and create a series of partitions until each machine (part) is in a singleton cluster (D'andrade, 1978).

Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate in objects successively into finer groupings.

Simply, agglomerative clustering (bottom-up) begins with singletons (sets with 1 element), merging them until *S* is achieved as the root. It is the most common approach And Divisive Clustering (top-down): Recursively partition *S* until singleton sets are reached (Andrew Moore, 2005).

Hierarchical clustering methods involve a two-stage process that first calculates the similarity coefficients between each pair of individuals (machines or parts). This can be represented as a lower triangular matrix since the similarity relationship between individuals is independent. The second stage of the process determines how the pairs with roughly equivalent similarity levels should be merged. (Andrew Moore, 2005).

There is wide latitude in the definition of the resemblance matrix and choice of clustering method. A resemblance coefficient can be a similarity or dissimilarity coefficient. The larger the value of similarity coefficient, the more similar the two parts/machines are; the smaller the value of a dissimilarity coefficient the more similar the parts/machines. A few of the clustering methods, which will be discussed, are single linkage clustering, average linkage clustering, complete linkage clustering and linear cell clustering (Borgatti, 2005).

The values of similarity coefficient generally range from 0 to 1, except with some proposed methods in which coefficients vary from a negative number to positive value greater than one.

Some of these clustering techniques are Single Linkage Clustering (SLC), Complete Linkage Clustering (CLC), Average linkage Clustering (ALC), Linear Cell Clustering (LCC).

McAuley (1972) was the first to apply single linkage clustering to cluster machines. He developed this procedure, which makes use of a Jaccard's similarity coefficient. This similarity coefficient is first defined between two machines in terms of the number of parts that visit each machine. Once the similarity coefficients have been determined for machine pairs, SLC evaluates the similarity between two machine groups as follows: the pair of machines (or a machine and a machine group, or two-machine groups with the highest similarity is grouped together. This process continues until the desired number of machine groups has been obtained or all machines have been combined in one group. As a result of applying SLC, two groups are merged together merely because two machines have high similarity. If this process continues with alone machines that have not yet been clustered, it results in chaining. The SLC is most likely to cause chaining.

Processing times, operation sequences and production volumes are sued in calculating the similarity coefficients as in Seifoddini and Djassemi (1995). Tam (1990) defines a similarity coefficient based on Levenshtein’s distances, which are measures of distances of two sentences where operation sequences are considered as sentences.

Gupta and Seifoddini (1990) proposed CLC. The complete linkage clustering method combines two clusters at minimum similarity level, rather than at maximum similarity level as SLC. The algorithm, however, remains the same except that minimum similarity coefficient is used. Since CLC is antithesis of SLC, it is least likely to cause chaining.

SLC and CLC are clustering based on extreme values. Instead, it may be of interest to cluster by considering the average of all links within a cluster. ALC produces results between extremes SLC and CLC. ALC is proposed by Seifoddini and Wolfe (1987). In ALC, the similarity between two clusters is the average of similarity coefficients of all members of the two clusters.

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object.

The linkage method you choose determines how the distance between two clusters is defined. At each stage, there is a distance matrix. The entry, d(m,j), in row m and column j of this matrix is the distance from cluster m to cluster j. At the beginning, when each observation constitutes a cluster, the distance from cluster m to cluster j is the corresponding value in D, giving the distance from observation m to observation j. On each step of the amalgamation algorithm, the two rows (and columns) of the distance matrix corresponding to the two clusters to be joined are replaced by a new row (and column) corresponding to the new cluster created by joining the two clusters. The linkage method determines how the elements, d(m,j), of the new row, m, are calculated from the elements, d(k,j) and d(l,j), of the deleted rows, k and l (Miranda, 2005).

Distances can be found in different ways, for example, we could link two clusters together when *any* two objects in the two clusters are closer together than the respective linkage distance. Put another way, we use the "nearest neighbors" across clusters to determine the distances between clusters; this method is called *single linkage*. This rule produces "stringy" types of clusters, that is, clusters "chained together" by only single objects that happen to be close together. Alternatively, we may use the neighbors across clusters that are furthest away from each other; this method is called complete linkage.

With average linkage, the distance between two clusters is the mean distance between an observation in one cluster and an observation in the other cluster. Whereas the single or complete linkage methods group clusters based upon single pair distances, average linkage uses a more central measure of location.

Here the distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group.

With complete linkage, or "furthest neighbor," the distance between two clusters is the maximum distance between an observation in one cluster and an observation in the other cluster. This method ensures that all observations in a cluster are within a maximum distance and tends to produce clusters with similar diameters. The results can be sensitive to outliers.

In this method, the distances between clusters are determined by the greatest distance between any two objects in the different clusters. This method usually performs quite well in cases when the objects actually form naturally distinct "clumps."

2.1.5. Non-hierarchical Clustering Methods

Non-hierarchical methods use the number of clusters to be formed as an input. After the number of clusters is determined the seeds are selected and the parts or machines are assigned to theses seeds.

Chandrasekharan and Rajagopalan (1986) applied a non-hierarchical technique (ISNC) using an evaluation criterion called "grouping efficiency", which measures inter-cell movement and within cell machine utilization.

Chandrasekharan and Rajagopalan (1986) developed ZODIAC. It is a much-improved expanded version of ISNC. The evaluation criterion was expanded by the introduction of "limited efficiency", or upper bound. They suggested an upper bound on the number of possible candidate cells and applied absolute value metric for distances. After generating the required number of seeds, parts and machines are grouped independently into equal number of clusters.

K-means is one of the simplest unsupervised Non-hierarchical clustering algorithms that solve the well-known clustering problem. This popular method of classification partitions a set of cases into k clusters to minimize the “error” or sum of squared distances of the cases about the cluster means. This method is developed by MacQueen (1967) and others.

2.1.6. Heuristics

Logendran (1990) proposed an algorithm in order to minimize the total moves contributed by both intercell and intracell moves. Harhalakis, Nagi and Proth (1990) proposed a bottom-up aggregation procedure in order to minimize ‘normalized intercell traffic’. After the cells are formed, the total intracell traffic is tried to be maximized for improvement.

Del Valle, Balarezo and Tejero (1994) proposed a 4-stage workload based model that minimizes intercellular moves. Bazargan (1996) used a pairwise interchange method to form cells. Nagi, Harhalakis and Proth (1990) proposed a bottom-up aggregation procedure minimizing part traffic under the constraints of multiple routings, multiple functionally similar workcenters, operation sequences, demand and work center capacities. Ballakur and Steudel (1987) consider within-cell machine utilization, workload fractions, maximum number of machines assigned to a cell, and the percentage of operations of parts completed within a single cell in their model. The model indirectly minimizes total number of intercell moves. Gupta and Seifoddini (1990) presented a two-stage algorithm, which considers several important criteria such as within-cell machine utilization, maximum number of machines that are assigned to a cell, maximum number of cells, total material handling cost to determine best among alternatives.

Algorithm assuming uniform machine utilization in cells and no intercellular moves is proposed by Sarker and Balan (1996). The optimal number of cells is found by minimizing the average material handling, setup costs and cost of performing bottleneck operations. Askin and Subramanian (1987) proposed a cost-based heuristic to determine machine groups and part families.

2.1.7. Mathematical Methods

The mathematical programming approaches differ in the manner that the number of part families is determined. There are mathematical models that use a sequential or simultaneous approach to the cell formation problem. Mathematical methods can be classified into four major groups: Linear Programming, Linear and quadratic integer programming, Dynamic programming, Goal programming (Russell et al., 1999).

The p- median 0/1 integer programming formulation identifies part families. The p-median approach involved initially selecting p of the parts to serve as medians or seeds for clusters. Subsequently, the remaining parts were assigned to the seed parts such that the sum of part similarity in each part family was maximized. A significant contribution of this method was that it was one of the first procedures developed to process a similarity matrix using mathematical programming as opposed to hierarchical clustering. A major limitation of this method was that only part families were identified and that a second procedure was needed to identify machine cells (Fan et al., 2004).

The cell formation problem can be formulated as a general assignment problem. Specifically, the formulation was for minimizing the cost of assigning parts to cells such that minimum and maximum usage levels for each cell were achieved. Wang (1998) proposed a linear assignment algorithm for formation of machine cells and part families in cellular manufacturing (Wang, 1998).

The cluster algorithms and p-median model minimize the distance or maximize the similarity between parts by considering the group mean or median. However, the parts within a group interact with each other. Therefore, it becomes important to account for the total group interaction. A quadratic programming model is proposed for this purpose (Hiller et al, 1966).

Thomas (2004) proposed an algorithm combining dynamic programming and genetic search for solving a dynamic facility layout problem. A model coped with equal sizes, which may change from one period to in time to the next.

Wei and Gaither (1990) presented the first 0/1 integer programming model for minimizing the cost associated with intercellular transfers. An empirical analysis was conducted by Fan et al. (1999) to assess the relative effectiveness of four integer-programming models for the regular permutation flowshop problem.

Graph partitioning methods treat the machines/parts as vertices and the processing, of parts as arcs connecting the nodes. These models aim at obtaining disconnected sub graphs from a machine-machine or machine part graph to identify manufacturing cells. The algorithms select a key machine or part according to a criterion. The cell formation problem is defined by Kandiller (1998) using the hypergraph representation of the manufacturing systems. The proposed method approximates the hyphergraph model by graphs so that the cuts are less affected by approximation. The algorithm is subjected to an experimentation of randomly generated manufacturing situations.

Choobineh (1988) proposed a two-stage algorithm that determines the part families and machine groups sequentially. Zou and Askin (1995) also proposed a sequential procedure that forms the part families by using a similarity coefficient based method and the machine groups are determined for each part family by using a composite operation set.

Lee and Chen's (1997) model minimizes normalized intercell movement under cell size, capacity and workload balance among the duplicated machines constraints. A three-stage procedure is proposed in order to minimize intercell movements and to balance the workload among duplicated machines. Lin et al. (1996) proposed a model, which minimizes intercell material handling, intracell processing and cell imbalance costs. In Vakharia and Chang's (1997) model, additional machine investment cost and intercell material handling is minimized under the cell size and machine capacity constraints. Adil, Rajamani and Strong (1996) proposed a nonlinear IP to identify part families and machine groups simultaneously, considering alternative routings.The objective is minimizing the total number of voids and exceptional elements.

Joines, Culbreth and King (1996) proposed an IP in order to minimize intercell movements. The model uses binary part-machine incidence matrix. Each machine and part can be assigned to only one cell or family. Rajamani, Singh and Aneja (1992) proposed a solution procedure to cell formation problem in a manufacturing environment where there are significant sequence dependent setup times and costs. Beaulieu, Gharbi and Ait-Kadi (1997) proposed a MIP model in order to minimize annual machine cost under the machine capacity and cell size constraints. Alternative routings are considered and no intercell movement is allowed. A two-stage heuristic is proposed to solve the model. Heragu and Gupta (1994) used a mathematical programming formulation only to determine the required number of each machine types. A search heuristic is used to solve the cell formation problem. An integrated approach that solves the part-family and machine-cell formation problem simultaneously is proposed by Akturk and Turkcan (2000). The algorithm considers the efficiency of both individual cells and the overall system. The algorithm that determines the within-cell layout with equal weighted backward and skipping costs provides two alternative solutions: independent cells, inter-cell movement.

2.2. Performance Evaluation Techniques

2.2.1. Technical Criteria

One of the important issues involved in the design of CM systems is the evaluation of cell formation solutions. Although, there have been quite a number of techniques developed, the evaluation of cell formation solutions has remained somewhat qualitative (Sarker, B. R, 1998).

As stated before, each cell formation algorithm within the Cell Formation module reorders the rows and columns in part/machine matrix to obtain a nearly block diagonal form. Based on this final matrix, the user forms the alternative cells in an interactive mode. The last step in cell formation process is to evaluate the performance of each alternative cell design using the following criteria (Sarker, B. R, 2001):

- Grouping efficiency,
- Grouping efficacy,
- Grouping measure,
- Machine Utilization Index
- Grouping efficiency for jobs with alternative routings

In this thesis we use five technical measures to quantify the within cell utilization, inter and intra cell movements, the ability to convert a random matrix into block diagonal form, and the ability to cluster is together. Thus, these performance measures are comprehensive in their evaluation of cell formation.

Notation:

M: number of machines

P: number of parts

c: number of cells (diagonal blocks)

d: number of 1 s in diagonal blocks

e: number of exceptional elements in the solution

o: number of ls in the part machine matrix

z: number of zeros in the part machine matrix

v: number of voids in the solution

w: weighting factor

2.2.1.1. Group Efficiency (η)

Group efficiency, which was proposed by Chandrasekaran and Rajagopalan (1986), was one of the first measures to evaluate the result obtained by different algorithms. The 'goodness' of solution depends on the utilization of machines within cell and inter-cell movement. Grouping efficiency was therefore proposed as a weighted average of the two efficiencies η_1 and η_2.

$$\eta_1 = (o-e)/(o-e+v)$$

η_1: The ratio of number of ls in the diagonal blocks to the total number of elements in the diagonal blocks.

$$\eta_2 = (M \times P - o - v)/(M \times P - o - v + e)$$

η_2: The ratio of number of 0s in the off-diagonal blocks to the total number of elements in the off-diagonal blocks.

$$\eta = w \times \eta_1 + (1-w) \times \eta_2$$

Where

$$0 \le w \le 1 \text{ and } 0 \le \eta \le 1$$

If $\eta = 1$ this implies that, there are no voids and no exceptional elements in the block diagonal (perfect clustering)

It is difficult to assign the value of w as the range of grouping efficiency mostly varies from about 75 % to 100 % (Kumar and Chandrasekharan, 1990). In this thesis, a value of 0.75 is used for w. The weighting factor allows the designer to alter on the emphasis between utilization and inter-cell movement. In this evaluation if $M \times P$ is large, the presence of exceptional elements is not reflected.

2.2.1.2. Grouping Efficacy (τ)

Kumar and Chandrasekaran (1990) proposed the grouping efficacy between well-structured and ill-structured matrices to overcome the low discriminating power of grouping efficiency. It has a more meaningful 0-1 range. Unlike grouping efficiency, the grouping efficacy is not affected by the size of the matrix.

$$\tau = (o - e)/(o + v)$$

If $\tau = 0$ implies all the 1s are outside the diagonal blocks.
If $\tau = 1$ implies a perfect grouping with no exceptional elements and voids.

However, the influence of exceptions and voids is not symmetric. Thus, the change in exceptional elements has a greater influence than the change in the number of voids in the diagonal blocks. Finally, the voids in the diagonal blocks become less and less significant at lower efficacies.

2.2.1.3. Grouping Measure (η_g)

It is also a direct measure of effectiveness of an algorithm to obtain a final grouped matrix, which is proposed by Miltenburg and Zhang (1991). The value of η_g is high if the utilization of machines is high and few parts require processing on machines in more than one cell.

$$\eta_u = d/(d+v) \qquad 0 \le \eta_u \le 1$$

η_u: is a measure of usage of parts in the part-machine cell.

$$\eta_m = 1 - d/o \qquad 0 \le \eta_m \le 1$$

η_m: is a measure of part movement between two cells.

$$\eta_g = \eta_u - \eta_m \qquad -1 \le \eta_g \le 1$$

η_g: is a measure of usage of parts in the part-machine cell.

Thus to maximize η_g values, large values of η_u and small values of η_m are preferred.

2.2.1.4. Machine Utilization Index

The machine utilization is the percentage of the time the machines within the clusters are being utilized most effectively (Kumar, 1990).

$$\text{Machine Utilization Index} = o / \sum (m_i \times p_i)$$

Where $m_{i:}$ number of machines in cell i

$p_{i:}$ number of parts in cell i

2.2.1.5. Grouping Efficiency For Jobs With Alternative Routings (η_{ARG})

To evaluate the grouping effect in the presence of alternative routings, Sarker and Li (1998) developed a generalized group efficiency measure for cell formation with alternative routings and they coined to it a name alternative routing efficiency (ARG efficiency) which is defined as

$$\eta_{Arg} = ((o-e)/(o+e))/((z-v)/(z+v))$$

2.2.2. Economic Analysis

The cost expression used in such analysis consists of four components. They are within-cell material-handling cost, between-cell material-handling cost, machine replacement cost, and total material handling cost. An explanation of each cost component follows (Sule, 1994).

2.2.2.1. Cost Of Between-Cell Material Handling

It is assumed that the transportation cost per unit includes the cost of handling the unit in the parent cell, between the parent cell and the host cell, and within the host cell. The job-related expense is due to additional clerical accounting when the job is transferred between cells.

Cost of between-cell material handling= $\sum_j \sum_k u_{jk}$

u_{jk} : Number of units transported between cells j and k; $j < k$

2.2.2.2. Cost Of Designing The Layout

Carrying cost of machine is defined as

Carrying Cost = Carrying Distance × Unit Carrying Cost

2.2.2.3. Total Material Handling Cost

After determining the layout of the problem, the coordinates of all machines can be defined. By using the coordinates, distances between machines can be determined.

Rectilinear distance between two machines I and j is defined as

$$D(i,j) = |x_i - x_j| + |y_i - y_j|$$

Where x_i : is the X coordinates of machine i

y_i : is the Y coordinate of machine i

Total Material Handling Cost= $\sum_i \sum_j D(i,j) \times T(i,j) \times U$

Where $T(i,j)$: is the total number of parts transport between machines i and j.

U: is the cost of transporting one part for one unit of distance

CHAPTER III: THE PROBLEM ENVIRONMENT

3.1. The Products

The company can produce over 400 types of product. Arriving demands are weekly evaluated on an MRP system and production schedules are generated. In this thesis, data produced from the MRP system on the 42^{nd} week of 2004, is used for analysis purposes. This data contains information of products to be produced in the next 52 weeks. Data taken from MRP system consists of

1. Id of the product demanded
2. Amount
3. Delivery date of the order

The table including this data is given in the Appendix C (Table C1). From the table, it can be easily seen that there has been a demand for the same product on different dates. Annual demand for the products is calculated by rearranging the table. The annual demands of products are given in the Table C2 in Appendix C.

It is assumed and confirmed by the company that the annual demand figures for the given year will be representative of the annual demand figures for the next couple of years.

Referring to disposed annual demand table, the company will produce 368 different types of product in 52 weeks. However, 27 of them are subcontracting products manufactured in subcontracting partners. Therefore, the remaining 341 products are taken into account. In order to reduce the problem into a manageable size, a Pareto analysis is conducted.

Because the purpose is to decrease the level of material handling, the products that are demanded more, should play an important role for the analysis. Besides, the rate of utilization in the facility has a great importance. Therefore, the products using the resources of the facility at most would be significant on deciding the facility layout. If it is assumed that, the company has bared one unit of fixed cost for every one unit of production time, then the product that uses the capacity the most becomes the most costly product. In order to control the production cost, products should have low transportation costs. Table C3, in which the total capacity utilization is calculated, by multiplying total demands with operation time for products, is given in Appendix C. In "Cumulative % of capacity usage" area of the table, capacity usage percentages are sorted from high to low. Based on the data in Table C3, we generate the P-Q chart shown in Figure 1.

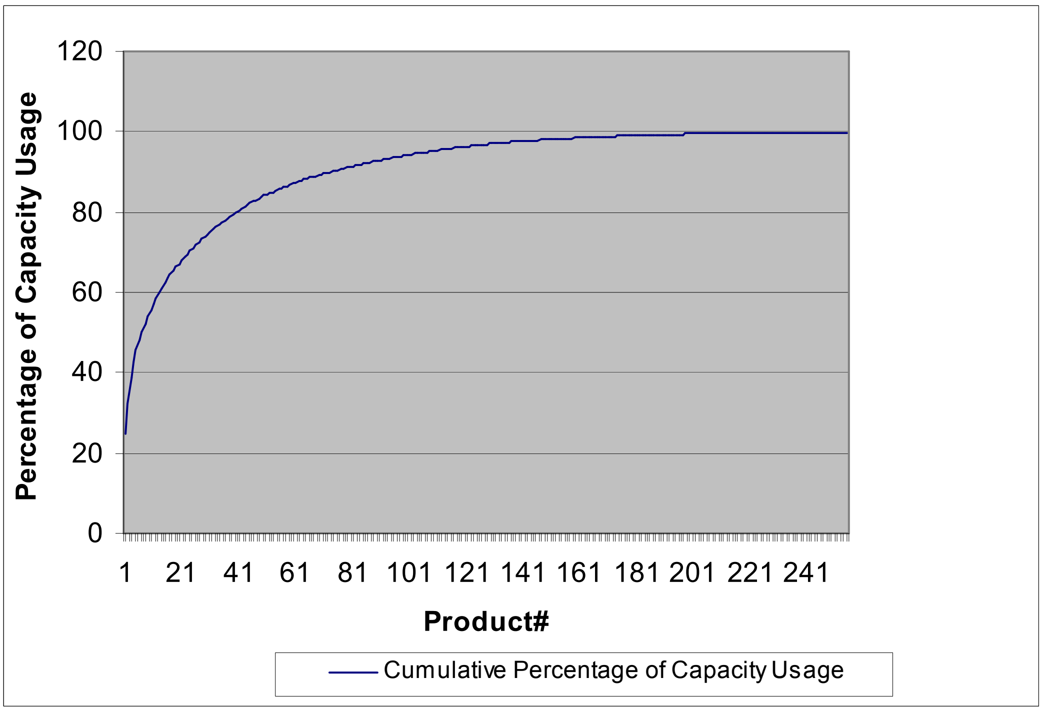


Figure 1. Pareto Chart of Products Sorted by Percentage of Capacity Usage

Figure 1 shows trend of the cumulative percentage of capacity. In the scope of this thesis, A Class products that consist of 80 % of the capacity usage are used. Other products are ignored. There are 36 products having usage of first 80 % of the capacity, (A). 55 products use 15 % of the capacity. The rest of the products, 250 products, consist of 5 % of the capacity usage. The list of A Class products, their rates, and amounts are given in Table A1 in Appendix A. The From-To Chart showing transportation amounts between machines is prepared by using the route information. This From-To chart is given in the Table C4 in Appendix C.

The machine/product incident matrix is used to determine the product's travel through machines during its manufacturing. This incident matrix is also given in Table C5 in Appendix C.

In the matrix, it can be easily noticed that some machines are on the route of all products. These machines can be assumed to be shared by all products and they will not be included in the cells. This property of the data set prevents us from forming the diagonal block structure. So, some machines are decided to be eliminated from the incidence matrix. The information of percentages and numbers of items passing through machine (i) is given in Table A2 in Appendix A. For instance, nine items pass through the machine 63. This means 25 % of all items visit this machine. It is decided that the machines whose passing through percentage is bigger than 25 % can be eliminated from the incident matrix. By using the number of items passing through machine (i), the following chart is formed. The chart shows that if the product amount passing through the same machine is more than 10, then this machine is eliminated from the incident matrix and the eliminated machines are evaluated separately (Figure 2).

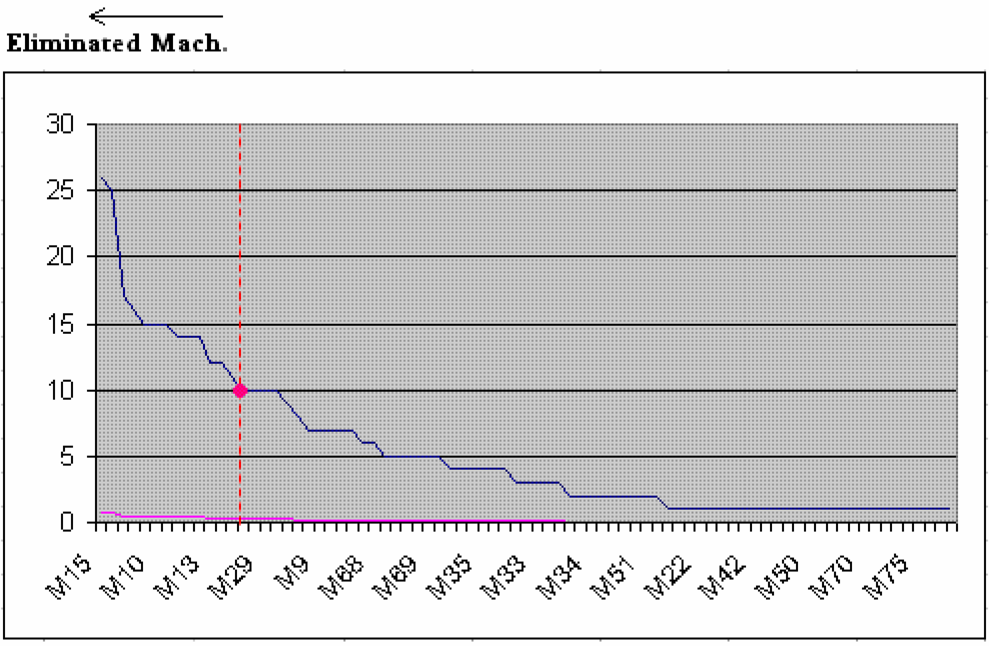


Figure 2. Eliminated Machines

This "Eliminated Incident Matrix" is given in Table C6 in Appendix C. After 17 machines are eliminated, 62 machines are left to group. Only two of the products are eliminated because they are processed on the eliminated machines. Thus, the number of products that would be processed is decreased to 34.

The machine-to-machine relationship table is formed by using the Eliminated Incident Matrix. Each entry of the machine component data is obtained by comparing columns m and n (for $\forall n, m$) computing the number of components requiring both machines. The machine-to-machine matrix is given in the Table C7 in Appendix C. The visual basic code used for forming the table is given in Appendix B1.

3.2. The Current Layout

The current layout of the company is shown in Figure 3. The plant is composed of four main divisions: Raw Material Storage, Production Area, Final Product Storage, and Offices. The differences between the operation sequences of the products make the raw material handling complicated. Therefore, in the company the required raw material is transferred to the manufacturing area before the production starts. As the production continues, the raw material for the later process is prepared. By the way of parallel processing, production does not wait for the raw material. The places that are separated with yellow-black tapes are arranged in order to keep the raw material together. The main purpose of this arrangement is to reduce the raw material handling from the raw material storage because nearly every item starts to be processed from a different machine.

Similarly, the final products are stored in the places with green colored tapes in the production area. The reason of making such an arrangement is the same. Many items complete their operation sequence in a different machine. Therefore, final products are transferred to the finished good area after they are collected in the production area. This arrangement provides us easiness and we are able to ignore the relationships between the storages and the machines. Therefore, in the following sections, only the production area will be taken into account.

The “ARC” is a special production line that produces only one item to only one customer. The line consists of only one unit and it works independent from the other machines. There is no constraint for its location. Therefore, machine has not been included in the calculations. Nevertheless, after designing the plant layout it was located in appropriate place.

In order to make a simpler visual expression of analysis results, some simplifications are made on the layout. All the machines are converted to smallest rectangular areas that they can fit into. These rectangular areas are prepared with the company staff by considering the walking and material handling path between machines. Therefore, when those rectangular areas (representing machines) are placed touching each other on the layout, walking and material handling paths are automatically formed. The simplified layout is given in Figure 4.

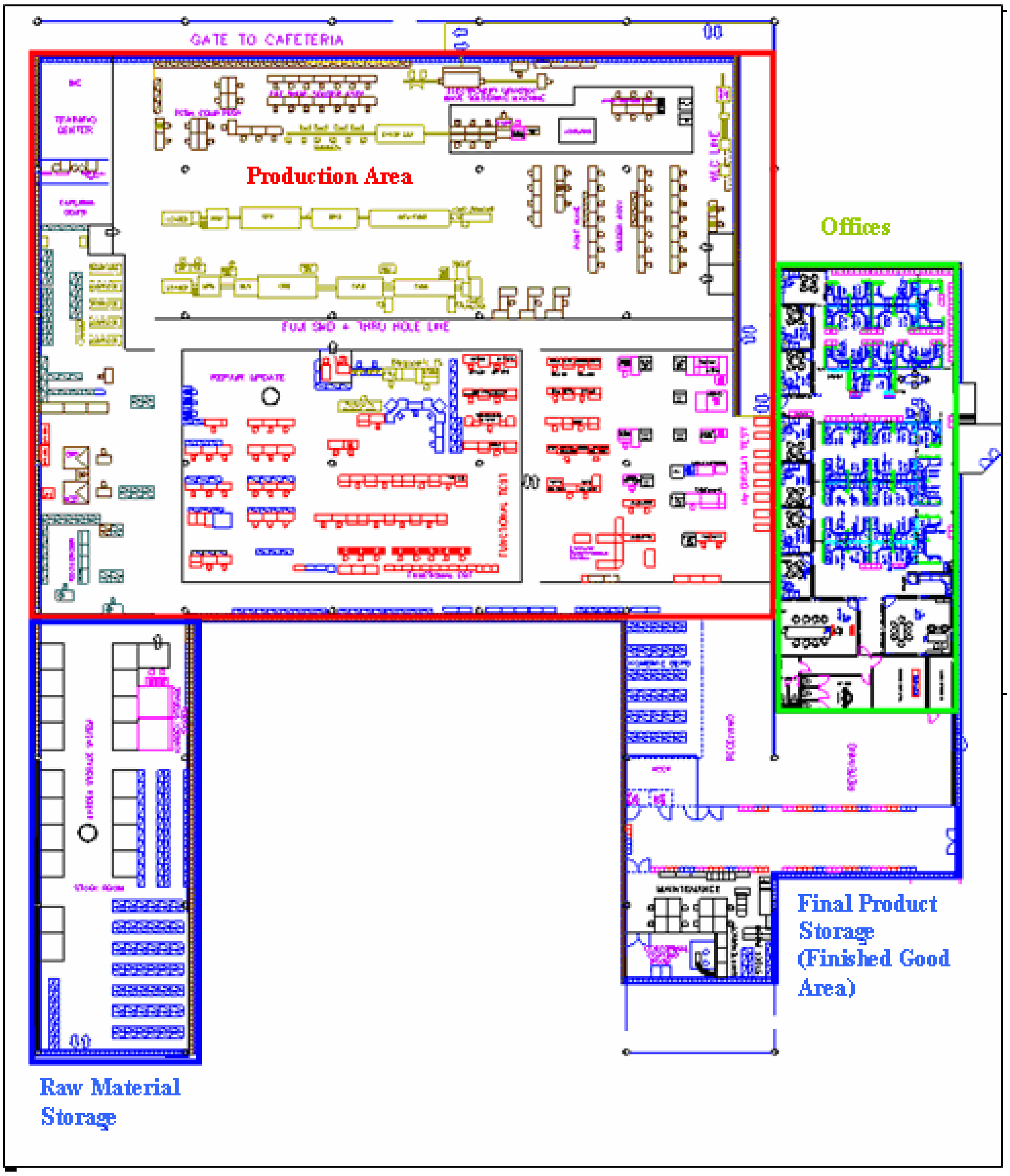


Figure 3. Current Layout (Original)

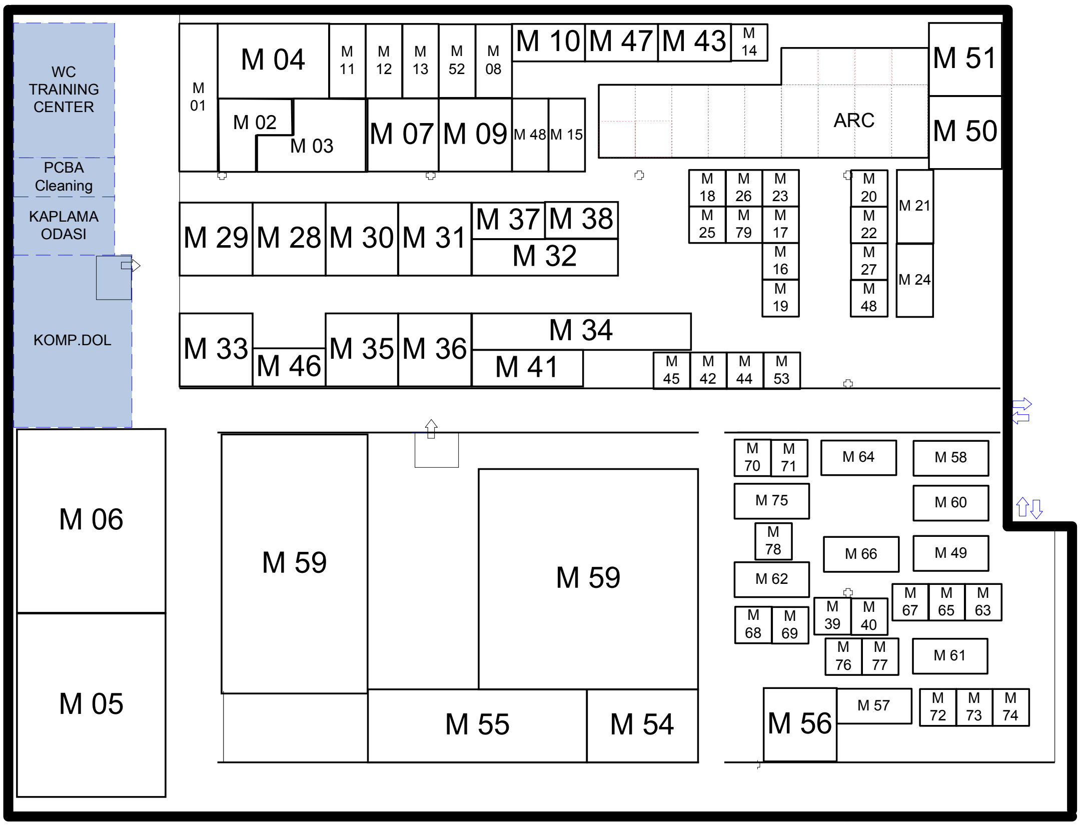


Figure 4. Current Layout (Simplified)

It is assumed and accepted that the left most and below point is (0,0) The co-ordinates are defined by this assumption. The representation of the X-Y Coordinates is shown in Figure 5. Table A3 showing the co-ordinates of present layout of the machines is given in Appendix A.

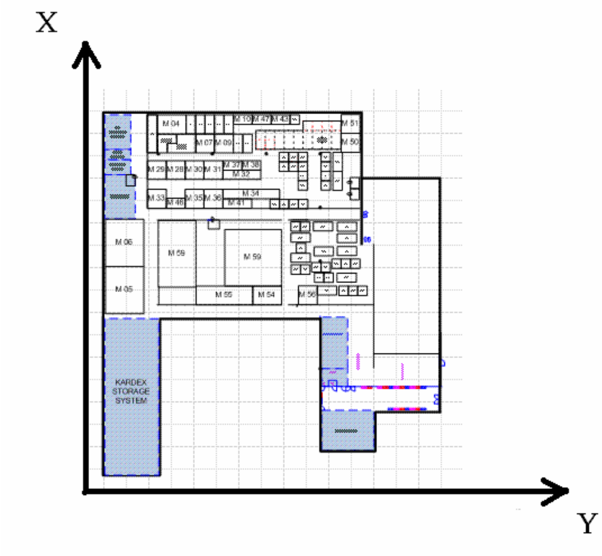
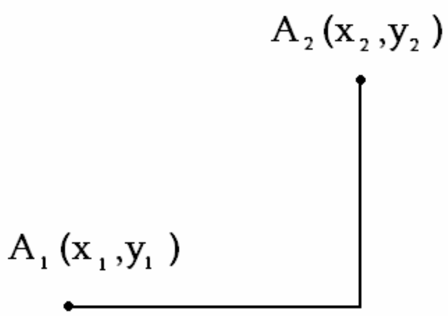


Figure 5. **X-Y Coordinates Representation**

In the plant, the walking and material handling paths have a rectilinear structure. The paths are defined with colored tapes on the floor. Therefore, passing through the machines with Euclidean paths are restricted. Therefore, in this thesis it is decided that distance between two machines can be approximated by the rectilinear distances. In this case, the distance matrix, whose elements are the distances between machines, is computed by the formula below. The Distance Matrix (D_r) between machines is given in Table C8 in Appendix C.



$$D_r = |x_1 - x_2| + |y_1 - y_2|$$

CHAPTER IV: PROPOSED METHODOLOGY

After the data was analyzed, results below can be reached.

1. Data:
    - The number of machines is 62 and the number of products is 34.
    - Volatile changes on demand occur throughout the year.

2. Operation sequences:
    - Routes of the products are varying. When the routes are analyzed in detail, it is observed that there were as many routes as the number of products. This assortment of the routes makes the problem more complex to solve.

The complexity of the operation sequences, the production dynamics and the immenseness of data; directed the researches towards heuristic techniques for the cell formation.

Considering the constraints, the layout is redesigned by using four techniques:

1. Tabular Method
2. Expanded Tabular Method
3. K-Means Clustering Method
4. Hierarchical Clustering with Levenshtein Distances

The problem solving procedure has three steps:

Step 1: Forming the Clusters

Step 2: Distributing the machines to clusters and forming suitable layout of inside the clusters.

Step3: Locating the clusters by using the From-to matrix between clusters.

All calculations of these techniques are mentioned in the further sections.

In the second step, the machines are located in appropriate places according to the relationship between each other. It is obvious that the machines that have strong relationship must be close to each other. Therefore, first the machines with the highest value of the material handling are located into the cluster. The machine(s) with the next highest value is located second. This continues while all machines are located into clusters. Such an arrangement provides us to reduce the material handling between machines within the clusters.

Although, the machines with close relationships are located into the same clusters, there would be material handling between machines that are in different clusters. The aim of the third step is to reduce the material handling between clusters. Therefore, the locations of cells would become critical. In order to arrange the cell locations, the material handlings between the clusters are calculated. By using this information, the same approach is applied to the cells.

4.1. Tabular Method

The most important reason to use this method is its simplicity. The technique is called the tabular method because the method involves successive calculations that can be tabulated easily (Sule, 94).

The method mainly involves two phases. In the first phase of the method, a machine is assigned to a group based on its affinity to all the machines that are presently in the group. It automatically identifies the bottleneck machines and distributes them to appropriate cells. The second phase distributes the jobs in the cells generated in the first phase (Sule, 94).

The data set used in algorithm is the “Incident Matrix”. Using this data machine-to-machine matrix is created.

A measure of effectiveness (P) of a machine joining to a group must be defined by the analyzer at the beginning of the problem. It states the closeness of an entering machine with all existing machines within a group in order for the entering machine to join that group. (Sule, 94) In this thesis, the P is defined as 0,5.

The algorithm gives the chance of duplicating the machines when needed. However, at the meetings with the company authorities, it is mentioned that none of the machines can be duplicated. The reason is there is no budget for buying a new machine in the short and long-term investment plans of the company. Moreover, another reason is that the production plant was working under capacity. According to this information taken from the company, “duplication” part of the algorithm is not used. The algorithm by nature starts assigning machines into cells with two machines having the largest relationship. Therefore, the benefit created by the assignment of the machine into the first cell is the highest, and the benefit that can be created by assigning into the second cell with duplicating the machine is ignored.

The algorithm is coded by using visual basic. Macros are applied on the data set held in Excel. This use is important for the company, because the company will not bear to any additional software cost for the applied technique.

The visual basic code of the algorithm is given in Appendix B2.

4.1.1. Forming the Clusters

Unlike some clustering algorithms, the tabular method determines the number of clusters. The algorithm suggests the best number of cluster.

Applied tabular method, resulted with a solution of eight clusters (K=8). The machines that are assigned to each cluster are shown in Table 1.

Table 1. Clusters With Tabular Method

Clusters	Machines				
1	Machine 31	Machine 67	Machine 14	Machine 22	Machine 74
	Machine 5	Machine 6	Machine 59	Machine 34	Machine 75
	Machine 9	Machine 61	Machine 25	Machine 41	Machine 76
	Machine 20	Machine 78	Machine 38	Machine 42	Machine 77
	Machine 24	Machine 16	Machine 4	Machine 45	Machine 79
	Machine 66	Machine 18	Machine 12	Machine 65	Machine 72
	Machine 11	Machine 58	Machine 39	Machine 70	Machine 73
	Machine 71	Machine 40			
2	Machine 68	Machine 69	Machine 3	Machine 17	Machine 57
3	Machine 63	Machine 64			
4	Machine 49	Machine 52	Machine 50		
	Machine 51	Machine 46	Machine 53		
5	Machine 48	Machine 54	Machine 55		
6	Machine 33	Machine 35	Machine 43	Machine 44	Machine 47
7	Machine 1	Machine 37			
8	Machine 60	Machine 62			

In the algorithm, the P value is set as 0.5 at the beginning of the solution procedure. In order to find out the sensitivity of P value, the algorithm is run with $P = \{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$. It is observed that no changes occur in the results.

There are thirty-seven machines in the first cluster. The other clusters have more homogenous number of machines. This result shows that most of the machines have close relationship with each other. This forces us to design the layout of the cells. All machines within the clusters are arranged by using the from-to chart. The machines that have strong relationship are judged to be close.

Table 2 shows the total relationships between the eliminated machines and the clusters. The table is formed by summing the material handling between the machines within a cluster and each of eliminated machines. For instance, the total material handling between M2 and the machines in cluster 1 is 59581 items. It is represented in the last column of the table that the total material handling between M2 and the other eliminated machines is 109020 items. As we analyze the values in the columns, we observe an apparent difference in the last column. The last column's values are larger than the other columns', except for M36. This means that the relationships between the eliminated machines are stronger. The eliminated machines have a cluster behavior. In the calculations, eliminated machines are taken into account as the ninth cluster (The exception of M36 is ignored.). For the other techniques, it is observed that the similar results occur. Therefore, in addition to the clusters the eliminated machines are assumed to form a cluster.

Table 2. Relationships Between Eliminated Machines And The Clusters (Number of items)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Group of eliminated machines
M 2	59.581	27.336	7.114	803	0	803	0	5.582	109.020
M 7	145.216	29.075	23.529	0	0	2.704	680	5.582	334.657
M 8	143.224	30.864	1.666	803	0	3.507	0	0	264.042
M 10	125.961	26.376	1.666	0	0	0	0	0	279.666
M 13	58.164	4.673	22.839	0	0	0	680	5.582	123.923
M 15	168.420	29.187	23.189	115.288	45.794	3.507	0	5.582	445.425
M 19	115.999	13.194	0	803	0	803	0	0	167.512
M 21	123.373	18.781	690	803	0	3.507	0	0	257.218
M 23	137.102	30.482	1.666	114.485	45.794	0	0	0	334.039
M 26	123.060	16.806	690	803	0	3.507	0	0	243.690
M 27	109.502	20.296	690	0	0	0	0	0	274.271
M 28	104.174	12.083	976	803	0	803	0	0	237.827
M 29	99.356	12.083	976	114.485	45.794	0	0	0	322.991
M 30	123.163	12.083	976	114.485	45.794	0	0	0	322.991
M 32	81.016	12.083	976	115.288	45.794	97.822	0	0	329.415
M 36	117.572	4.785	22.326	803	81.978	84.682	680	5.582	77.622
M 56	29.925	4.600	13.622	0	0	0	680	5.582	167.377

4.1.2. Designing the Cell Layouts

By rearranging the rows and the columns, the original machine-to-machine material-handling matrix is divided into eight sectors one for every cluster. The group of eliminated machines is added to the bottom of the matrix. By this way, the material handling matrices for cells are formed. The arranged material handling table is shown in Table C9 in Appendix C. The cell layouts are generated by locating the most related machines close.

The methodology is summarized below:

Start with the first cluster

  Find the largest value in the material handling matrix (within cluster)

    Locate the related machines as close as possible

    Erase the data

  Stop when all machines are located into the cell.

Pass to the next cluster

Stop when all clusters are arranged.

In the first cluster, the most related machines are M66 and M67. Firstly, these machines are located in the cell. It is followed with the nearest machine(s). When all machines are placed in the cell, we started to the second cell.

The most related machines in the clusters are;

M69 and M68 in the second cluster,

M64 and M63 in the third cluster,

M49 and M51-M52 in the fourth cluster,

M54 and M48 in the fifth cluster,

M33 and M35 in the sixth cluster,

M01 and M37 in the seventh cluster,

M60 and M62 in the eight cluster,

M29 and M30-M32 in the group of eliminated machines.

4.1.3. Locating the Clusters

In order to reduce the material handling between clusters, we arrange the cell locations. Firstly, the total material handlings between the clusters are calculated. In order to find out the material handling between the clusters, the From-to matrix between clusters is formed and shown in Table 3. The total sums of the total amount transferred between machines, which are in different cells, are calculated. After designing the cluster layouts, the locations of the clusters are determined with similar approach using the table below.

The methodology is summarized below:

Find the largest value in the material handling between clusters

Locate the related clusters as close as possible

Erase the data

Stop when all clusters are located.

Table 3. The From-To Matrix Between Clusters (Number of items)

Clusters	1	2	3	4	5	6	7	8	Eliminated Machines
1	-								
2	123.328	-							
3	7.462	3.113	-						
4	6.424	0	0	-					
5	0	0	0	3.980.594	-				
6	225.060	0	0	803	409.890	-			
7	1.360	0	680	0	0	0	-		
8	0	0	0	0	0	0	0	-	
Eliminated Machines	1.864.808	304.787	123.591	579.652	310.948	201.645	2.720	33.492	-

The machines which are eliminated from the incident matrix are assumed here as a group. The material handling between the eliminated machines and the machines within clusters are also added to the bottom of the table above. It is obvious that the group of eliminated machines have a strong relationship between all clusters. It is not a surprising solution because 75% of items are visiting these machines while being produced.

On the other hand, the most significant relationship is between the clusters four and five. The fourth cluster is composed of machines 46, 49, 50, 51, 52 and 53. Cluster 5 consists of machines 48, 54 and 55. The relationships between these machines are shown in Table 4. All these machines may be in the same cluster. This means that in the plant layout the fourth and the fifth clusters must be near to each other.

Table 4. From-to Matrix between Cluster 4 and 5 (Number of items)

Machine	46	49	50	51	52	53	48	54	55
46									
49	22897								
50	0	468953							
51	22897	491850	468953						
52	22897	491850	468953	491850					
53	22897	22897	0	22897	22897				
48	0	468953	468953	468953	468953	0			
54	22897	491850	468953	491850	491850	22897	496279		
55	22897	22897	0	22897	22897	22897	27326	50223	

After these three steps, the layout of the plant is shown in Figure 6.

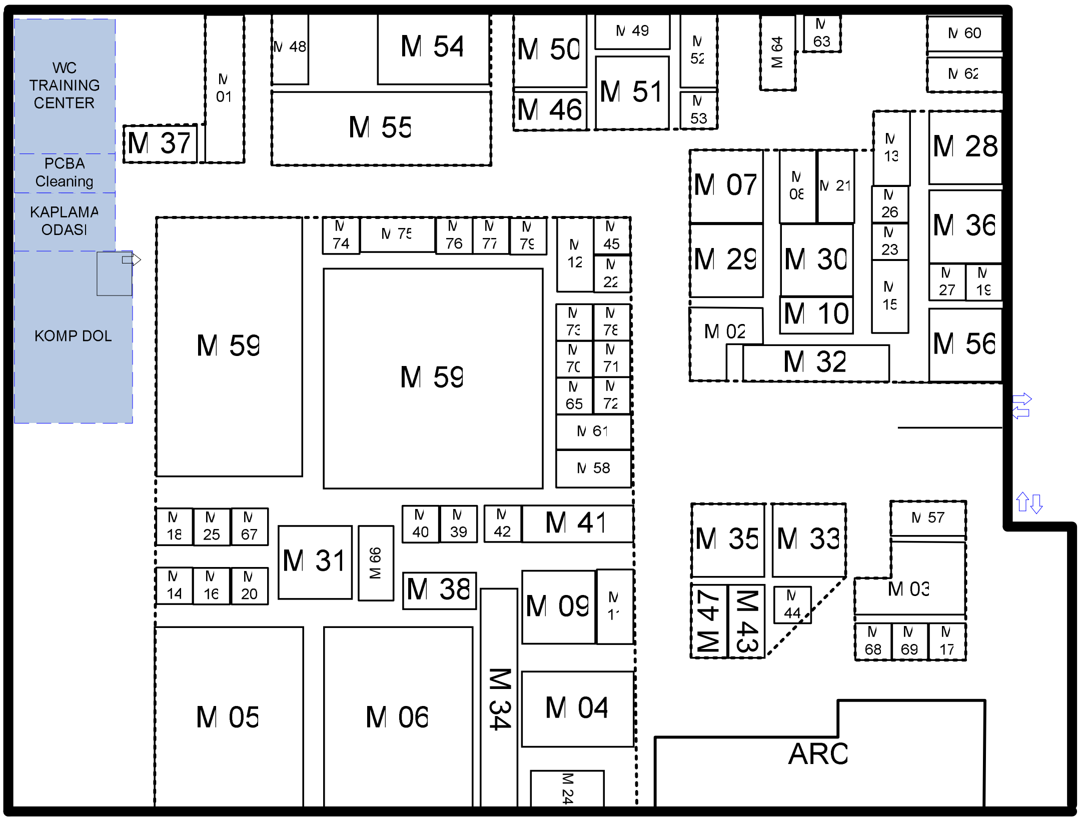


Figure 6. The Plant Layout Designed with Tabular Method

4.1.4. Block Diagonal Structure

Block diagonal structure is partitioning the matrix such that “ boxes” on the main diagonal contain 0’s and 1’s but off-diagonal boxes contain only 0’s. Block diagonalization is considered as the best approach to form part-families and machine cells. In an ideal solution, all the 1s will remain in the diagonal blocks of the incidence matrix and all 0s in the off diagonal blocks, but an ideal case is rarely obtainable in practice (Sarker, 2001).

Many algorithms have been developed to form block diagonalization of machine-part incidence matrices. Most of the algorithms are suitable for the formulation of block diagonals of well structured matrices, but in case of poor structured incidence matrices no researchers were sure of their capability whether they give a better solution or not. Again a wide range of heuristics have been developed to solve a problem that may or may not give the optimum result; but whatever the method used, one should choose the method that is based on some criteria to indicate the reliability of the solution (B.R.Sarker, 2001).

The aim is to collect the 1s in rectangles that are vertically bounded with the clusters. In addition, the rectangles would form a diagonal. In order to form this diagonal structure, the 1s in columns are summed and the matrix is vertically sorted according to the sums within the clusters. By changing the order of the rows, the diagonals are formed. By this way, the Block Diagonal Structure of the solved machine-part incidence matrix is formed. The matrix has thirty-four rows and sixty-two columns. The table is given in the Table C10 in Appendix C.

The block diagonal results are listed below:

The number of 1 s in diagonal blocks *(d)* is 135,
The number of exceptional elements in the solution *(e)* is 43,
The number of l s in the incidence matrix *(o)* is 178,
Number of voids in the solution *(v)* is 551,
Number of cells (diagonal blocks) *(c)* is 8.

According to this information, the Performance Measures that are discussed in Chapter 2 are calculated. In order to provide the ease of comparing, the solutions for all techniques used are summarized in a table (Table 22 in Section 4.5).

4.2. Expanded Tabular Method

Machine to machine transportation data (I_{ij}) is used in the traditional tabular method procedure. Here, I_{ij} refers how many times a transportation occurred between machine i and j. However, this approach ignores the demands of the product. If there are differences between the demands of the products, the impact of the demand to the results cannot be ignored.

Table 5 shows the unit loads (u_i) of material movement for each product. The material handling for parts is obtained by applying $\sum_j I_{ij} \times u_i$. By using this data Machine-to-machine material handling table is formed. The Visual Basic code for these calculations is given in Appendix B1. Machine-to-machine material handling table is also shown in Table C11 Appendix C

Table 5. Unit Loads (U_i) Of Material Movement (Number of items)

Product No	u_i Unit Load Required	Product No	u_i Unit Load Required	Product No	u_i Unit Load Required
1	954	13	878	25	3769
2	1806	14	2048	26	2791
3	3755	15	2375	27	382
4	1784	16	22897	28	690
5	9155	17	2184	29	976
6	7953	18	1352	30	340
7	6717	19	803	31	185
8	1467	20	1046	32	468953
9	3016	21	513	33	27326
10	2380	22	3112	34	19430
11	741	23	506	35	7880
12	1254	24	531	36	5000

To begin the grouping process the value of P (the ratio) is set to 0.5 (step 5). The traditional Tabular Method steps are followed in order to solve the problem.

4.2.1. Forming the Clusters

The steps that are explained in the tabular algorithm are terminated after all machines are assigned. The result consists of six clusters, with the following machine arrangements are shown in Table 6.

Table 6. Clusters With Expanded Tabular Algorithm

Clusters	Machines				
1	Machine 64	Machine 40	Machine 4	Machine 74	Machine 39
	Machine 35	Machine 42	Machine 12	Machine 22	Machine 73
	Machine 33	Machine 38	Machine 65	Machine 76	Machine 37
	Machine 34	Machine 25	Machine 75	Machine 77	Machine 62
	Machine 45	Machine 72	Machine 1	Machine 60	
2	Machine 54	Machine 55	Machine 43	Machine 47	Machine 53
	Machine 48	Machine 50	Machine 44	Machine 46	
3	Machine 52	Machine 49	Machine 51	Machine 41	
4	Machine 66	Machine 5	Machine 9	Machine 78	Machine 3
	Machine 61	Machine 69	Machine 68	Machine 6	Machine 57
	Machine 17				
5	Machine 31	Machine 24	Machine 58	Machine 71	
	Machine 20	Machine 67	Machine 70		
6	Machine 63	Machine 14	Machine 16	Machine 79	
	Machine 11	Machine 59	Machine 18		

There are twenty-four machines in the first cluster. The diversity of the clusters with this technique is better than the results of the traditional tabular methods. Still the number of the machines in the first cluster is higher than the others.

4.2.2. Designing the Cell Layout

The material handling matrices within the cells are formed by using the procedure that explained in Section 4.1.2. The arranged material handling table is shown in Table C12 in Appendix C. By using this table, the best cluster inside layout is formed. The cell layouts are obtained by locating the most related machines close as mentioned in Section 4.1.2.

The most related machines in the clusters are;

M33 and M35 in the first cluster,
M54 and M48 in the second cluster,
M48, M51, and M52 in the third cluster,
M56 and M66 in the fourth cluster,
M31 and M67 in the fifth cluster,
M14 and M11 in the sixth cluster,
M29 and M30-M32 in the group of eliminated machines.

4.2.3. Locating the Clusters

The from-to matrix between clusters is formed and shown in Table 7. The forming procedure is mentioned in Section 4.1.3. Locations of the clusters are determined with the approach that was explained in Section 4.1.3.

Table 7. The From-To Matrix Between Clusters (Number of items)

Clusters	1	2	3	4	5	6	Eliminated Machines
1	471.686						
2	327.912	1.954.161					
3	6.424	4.495.341	1.476.353				
4	65.105	0	1.606	142.321			
5	35.663	0	3.212	119.703	104.535		
6	24.028	0	1.606	71.137	30.660	37.976	
Eliminated Machines	406.940	621.896	357.909	869.666	851.779	383.699	1.992.421

The material handling between the eliminated machines and the clusters are decreased. On the other hand, the most significant relationship is between the clusters two and three. The relationships between these machines are shown in Table 8. All these machines may be in the same cluster. This means that in the plant layout the third and the fifth clusters must be near to each other.

Table 8. Relationship Matrix between Cluster Two and Three (Number of items)

MCH	43	44	46	47	48	50	53	54	55	41	49	51	52
43													
44	27326												
46	0	0											
47	27326	27326	0										
48	27326	27326	0	27326									
50	0	0	0	0	468953								
53	0	0	22897	0	0	0							
54	27326	27326	22897	27326	496279	468953	22897						
55	27326	27326	22897	27326	27326	0	22897	50223					
41	0	0	0	0	0	0	0	0	0				
49	0	0	22897	0	468953	468953	22897	491850	22897	0			
51	0	0	22897	0	468953	468953	22897	491850	22897	0	491850		
52	0	0	22897	0	468953	468953	22897	491850	22897	803	491850	491850	

The layout of the plant is shown in Figure 7.

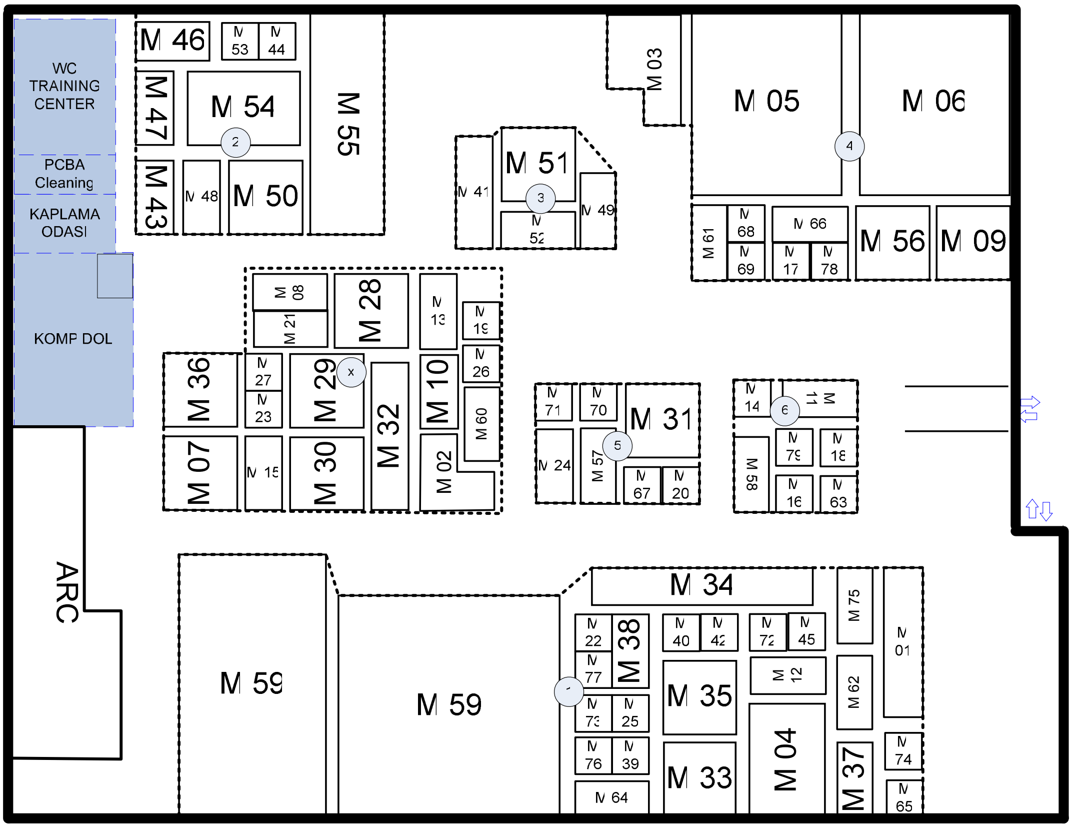


Figure 7. The Plant Layout Designed with Expanded Tabular Method

4.2.4. Block Diagonal Structure

The formed block diagonal structure of the incidence matrix is given in Table C13 in Appendix C.

The results of the variables:

The number of 1s in diagonal blocks (*d)* is 106,

The number of exceptional elements in the solution *(e)* is 72,

The number of ls in the part machine matrix *(o)*, is 178,

Number of voids in the solution *(v)* is 210,

Number of cells (diagonal blocks) *(c)* is 6.

With this technique, the number of 1s in diagonal blocks is decreased from 135 to 106. Therefore, the number of exceptional elements in the solution is automatically increased from 43 to 72. Because the total number of 1s in diagonal blocks and exceptional elements in the solution is constant. The material handling between clusters would be higher with this cluster structure. It is obvious that the material handling within the clusters would be less than the Traditional Tabular Method's.

The Performance Evaluation Results that are calculated according to this information are summarized in Table 22.

4.3. K- Means Clustering Algorithm

In this technique, the grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroids.

We use the K-Means Algorithm in order to group our machines into K number of clusters. In this technique, the user is able to define the number of clusters. The decision maker would want to decide the number of clusters by comparing results for different number of clusters. A set of numbers can be tried as the number of clusters.

The data of Eliminated Incident Matrix is clustered by using K-Means Clustering Algorithm with MINITAB 14.

In order to determine the number of clusters, different number of classes is used to solve the problem. The solutions of different number of classes were compared with each other. The comparison criteria are

- Maximum Sum Of Squares Within Clusters
- Average Distance From Centroid
- Maximum Distance From Centroid
- Maximum Distance Between Cluster Centroids

The results for every K is given in the Table C14 in Appendix C

The results of comparison criteria for K= {6, 7, 10, 15, 20} are summarized in Table 9.

Table 9. The Results Of Comparison Criteria

K	Within Cluster Sum of Squares (i=1)	Average Distance From Centroid (i=2)	Maximum Distance From Centroid (i=3)	Distances Between Cluster Centroids (i=4)
6	74,6670	1,8000	2,2640	3,2146
7	46,4000	1,3310	2,1350	3,6742
10	48,5630	1,2150	1,7140	3,3442
15	34,0000	1,1010	1,1010	2,8710
20	26,9570	1,1180	2,2720	3,4641

In order to compare the results normalization is needed. Normalized matrix is calculated by dividing all values in the columns by the sum of the each column. The normalized matrix is given in Table 10. By using the normalized matrix, the chart shown in Figure 8, is formed.

Table 10. Normalized Matrix Comparison Criteria

K	Within Cluster Sum of Squares (i=1)	Average Distance From Centroid (i=2)	Maximum Distance From Centroid (i=3)	Distances Between Cluster Centroids (i=4)
6	0,3238	0,2742	0,2387	0,1940
7	0,2012	0,2027	0,2251	0,2218
10	0,2106	0,1851	0,1807	0,2018
15	0,1474	0,1677	0,1161	0,1733
20	0,1169	0,1703	0,2395	0,2091

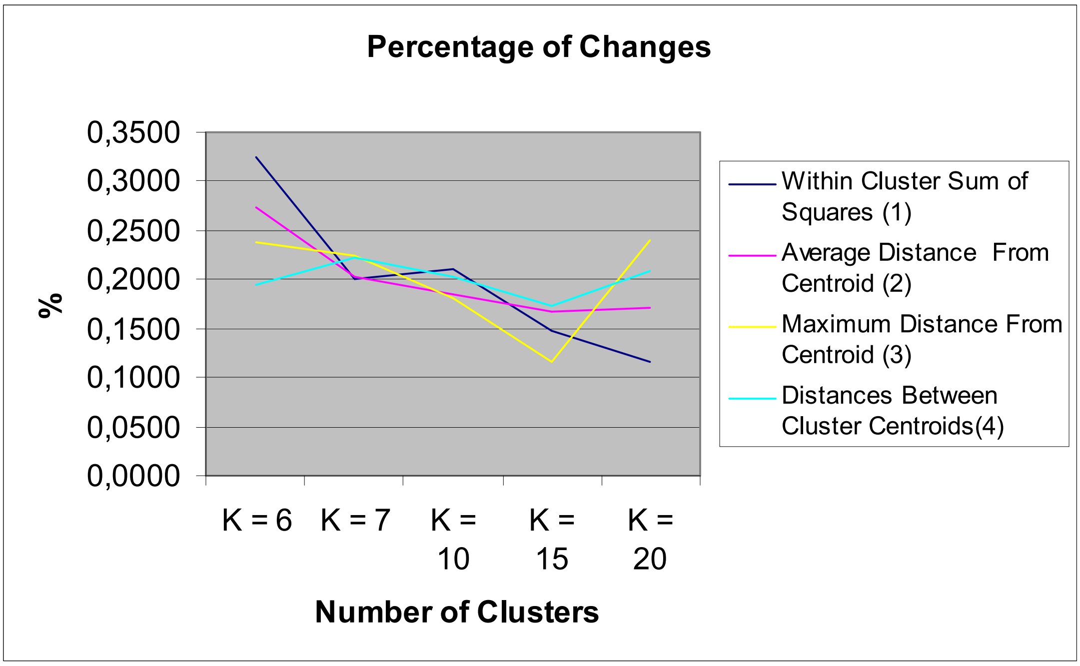


Figure 8. The Percentage Of Change

As shown in the figure, a remarkable decrease is occurred between K=6 and K=7. By analyzing the differences between all k values the number of clusters is defined as seven (K=7).

4.3.1. Forming the Clusters

The results of K-Means Algorithm for K = 7 is given in Table 11.

Table 11. Clusters With K-Means Clustering Algorithm

Clusters	Machines				
1	Machine 31				
2	Machine 3	Machine 17	Machine 6	Machine 68	Machine 16
	Machine 5	Machine 61	Machine 9	Machine 69	
3	Machine 63	Machine 64			
4	Machine 20	Machine 58	Machine 34	Machine 72	Machine 41
	Machine 33	Machine 59	Machine 35	Machine 73	Machine 45
	Machine 78				
5	Machine 24	Machine 57			
6	Machine 66	Machine 67			
7	Machine 1	Machine 25	Machine 44	Machine 52	Machine 70
	Machine 4	Machine 37	Machine 46	Machine 53	Machine 71
	Machine 11	Machine 38	Machine 47	Machine 54	Machine 74
	Machine 12	Machine 39	Machine 48	Machine 55	Machine 75
	Machine 14	Machine 40	Machine 49	Machine 60	Machine 76
	Machine 18	Machine 42	Machine 50	Machine 62	Machine 77
	Machine 22	Machine 43	Machine 51	Machine 65	Machine 79

Only one machine is located to first cluster. On the other hand, thirty-five machines are assigned to seventh cluster. Each of the fifth and the sixth clusters consist of 2 machines. The separation of the clusters is not homogenous.

4.3.2. Designing the Cell Layout

The material handling matrices within the cells are formed by using the procedure that explained in Section 4.1.2. The arranged material handling is shown in Table C15 in Appendix C. By using this table, the best cluster inside layout is formed. The cell layouts are obtained by locating the most related machines closer, as mentioned in Section 4.1.2.

The most related machines in the clusters are;

M69 and M68 in the second cluster,
M63, and M64 in the third cluster,
M33 and M35 in the fourth cluster,
M24 and M57 in the fifth cluster,
M66 and M67 in the sixth cluster,
M54 and M48 in the seventh cluster,
M29 and M30-M32 in the group of eliminated machines.

4.3.3. Locating the Clusters

The From-to matrix between clusters is formed and shown in Table 12. The forming procedure is mentioned in Section 4.1.3. Locations of the clusters are determined with the approach that was explained in Section 4.1.3.

Table 12. The From-To Matrix between Clusters (Number of items)

Clusters	1	2	3	4	5	6	7	Eliminated Machines
1	-							
2	2.208	-						
3	0	5.183	-					
4	26.116	15.327	0	-				
5	12.081	13.242	1.952	33.125	-			
6	37.068	13.770	0	55.958	21.014	-		
7	6.885	40.947	4.120	556.251	10.599	1.806	-	
Eliminated Machines	246.772	442.173	123.591	570.258	172.009	453.006	1.413.834	-

The material handling between the clusters is significantly decreased. This means that the material handlings within the clusters are increased.

The layout of the plant is shown in Figure 9.

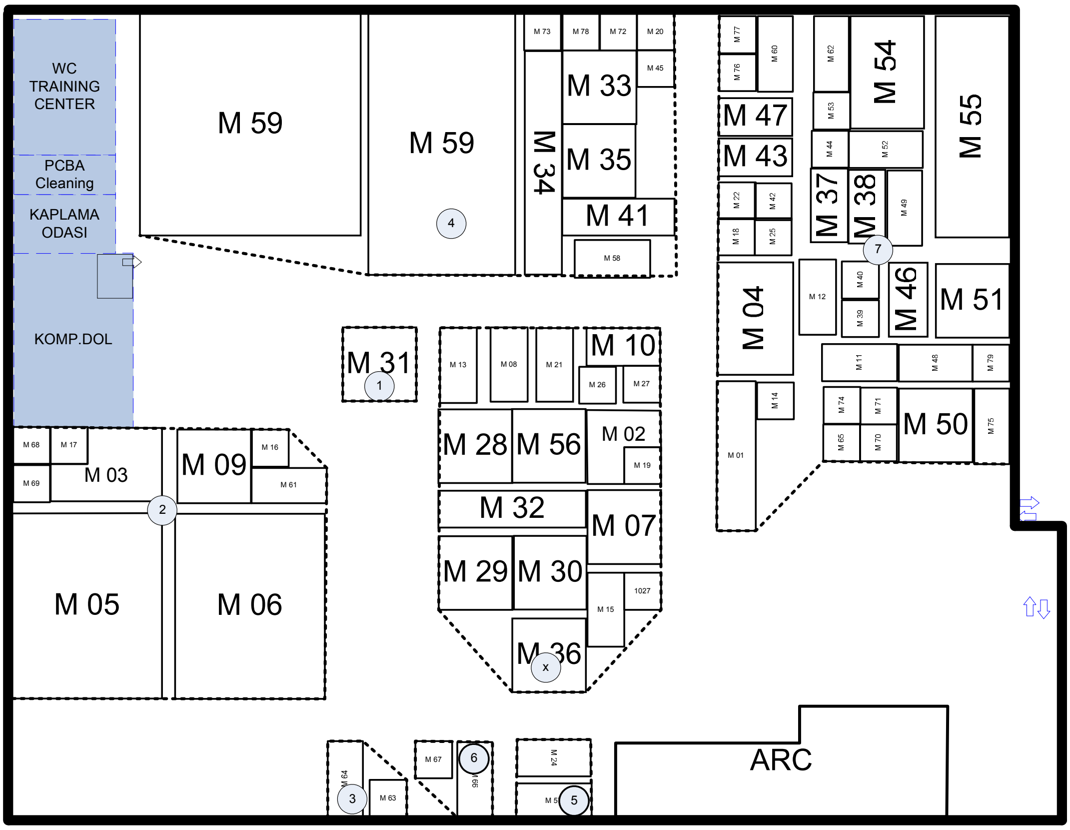


Figure 9. The Plant Layout Designed with K-Means Algorithm

4.3.4. Block Diagonal Structure

The Block diagonal structure of the incidence matrix is given in Table C16 in Appendix C.

The results of the variables are:

The number of 1 s in diagonal blocks (*d)* is 105,

The number of exceptional elements in the solution *(e)* is 73,

The number of ls in the part machine matrix *(o)* is 178,

Number of voids in the solution *(v)* is 100,

Number of cells (diagonal blocks) *(c)* is 7.

The Performance Evaluation Results are summarized in Table 22.

4.4. Hierarchical Clustering with Levenshtein Distance

As we mentioned before, in our problem, there are too many products with different operation sequences. Although the routes of the products are different from each other, they have some similar cycles in their routes.

In the other cell formation techniques used in this thesis, the operation sequences for the products are avoided. Other techniques mainly group the machines, which have strong relationship. Nevertheless, in this methodology, we classify the parts, which have similar operation sequences. This different point of view provides us to take into account of the routes. The usage of this methodology would probably reduce the complexity of the routes inside the cells.

The goal of this methodology is to avoid confusing flow pattern when a large number of flows are shown together (Irani, 2005).

The methodology is listed below (Irani, 2005):

1. Generate the Levenshtein distance matrix for all pairs of parts produced in the facility.
2. Based on the distance matrix generated in the previous step, perform a cluster analysis. A suitable threshold is chosen in the clustering dendogram to group the parts into clusters.

3. Adjacency graphs are developed for each of the part clusters.
4. Each of these adjacency graphs is embedded on the block layout.

4.4.1. What is the Levenshtein Distance?

In information theory, the Levenshtein distance or edit distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution. It is named after the Russian scientist Vladimir Levenshtein, who considered this distance in 1965. It is useful in applications that need to determine how similar two strings are.

In more detail, Levenshtein distance (LD) is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t.

This means, the greater the Levenshtein distance, the more different the strings are. It can be considered as a generalization of the Hamming distance, which is used for strings of the same length, only considers substitution edits.

The Levenshtein procedure is given in the Appendix B3. In addition, an example is provided in Appendix B4.

4.4.2. Hierarchical Clustering With Levenshtein Distance Results

The first step of the methodology is generating the Levenshtein Distance matrix for all pairs of products produced in the facility.

In order to determine the Levenshtein distance the operation sequences must be defined. The total volume of products whose operation sequences involve material movement from operation *i* to operation *j* for the operation sequences of the products shown in Table 13.

By using the algorithm that is explained in Appendix B3, the Levenshtein Distance matrix is calculated. This distance calculates the similarity between operation sequences. The Levenshtein Distance matrix is shown in Table 14.

The Visual basic code of this algorithm used is given in Appendix B5.

Table 13. Operation Sequences Of The Products

Products	Operation Sequence																						Batch Quantity
Product 1	M 02	M 07	M 07	M 15	M 24	M 27	M 61	M 57	M 27	M 57													954
Product 2	M 02	M 11	M 03	M 05	M 06	M 07	M 08	M 15	M 10	M 23	M 17	M 21	M 26	M 21	M 27	M 66	M 68	M 69						1806
Product 3	M 36	M 07	M 15	M 13	M 56	M 63	M 64																	3755
Product 4	M 19	M 28	M 30	M 31	M 32	M 29	M 32	M 07	M 08	M 15	M 10	M 23	M 21	M 20	M 26	M 27	M 66	M 67						1784
Product 5	M 28	M 30	M 31	M 32	M 29	M 32	M 07	M 08	M 15	M 10	M 23	M 20	M 27	M 66	M 67	M 57								9155
Product 6	M 11	M 13	M 14	M 19	M 21	M 26	M 11	M 13	M 19	M 21	M 26	M 14	M 79											7953
Product 7	M 11	M 13	M 19	M 21	M 26	M 14	M 79																	6717
Product 8	M 19	M 09	M 28	M 30	M 31	M 32	M 29	M 32	M 07	M 08	M 15	M 10	M 23	M 26	M 21	M 27	M 66	M 67	M 56					1467
Product 9	M 28	M 30	M 31	M 32	M 29	M 32	M 15	M 10	M 19	M 26	M 21	M 27	M 66	M 67	M 56									3016
Product 10	M 36	M 07	M 15	M 13	M 56	M 64	M 63																	2380
Product 11	M 02	M 28	M 32	M 30	M 31	M 32	M 29	M 32	M 29	M 07	M 08	M 15	M 10	M 14	M 19	M 09	M 09	M 26	M 14	M 58	M 59	M 78	M 56	741
Product 12	M 02	M 03	M 19	M 09	M 05	M 07	M 08	M 15	M 10	M 23	M 16	M 26	M 17	M 21	M 27	M 61	M 68	M 69	M 56					1254
Product 13	M 02	M 32	M 28	M 28	M 30	M 31	M 32	M 29	M 32	M 07	M 08	M 15	M 10	M 23	M 24	M 20	M 21	M 26	M 58	M 59				878
Product 14	M 28	M 30	M 31	M 32	M 29	M 32	M 11	M 07	M 24	M 23	M 20	M 21	M 26	M 27	M 58	M 70	M 71							2048
Product 15	M 36	M 02	M 07	M 15	M 13	M 64	M 63																	2375
Product 16	M 46	M 30	M 32	M 29	M 52	M 15	M 23	M 51	M 49	M 54	M 54	M 55	M 53	M 53	M 53									22897
Product 17	M 19	M 05	M 06	M 07	M 08	M 15	M 10	M 14	M 13	M 13	M 18	M 21	M 58	M 59	M 27	M 78								2184
Product 18	M 33	M 35	M 36	M 34	M 45	M 07	M 08	M 15	M 08	M 20	M 66	M 67	M 21	M 26	M 24	M 72	M 73	M 78	M 56					1352
Product 19	M 02	M 32	M 28	M 35	M 36	M 34	M 45	M 52	M 08	M 15	M 45	M 20	M 58	M 59	M 21	M 19	M 24	M 41	M 26	M 78				803
Product 20	M 02	M 03	M 19	M 09	M 05	M 07	M 08	M 15	M 10	M 23	M 16	M 26	M 17	M 21	M 27	M 61	M 68	M 69	M 56					1046
Product 21	M 02	M 07	M 15	M 13	M 63																			513
Product 22	M 28	M 30	M 31	M 32	M 29	M 32	M 07	M 08	M 15	M 10	M 23	M 21	M 20	M 26	M 27	M 66	M 67	M 56	M 78					3112
Product 23	M 36	M 07	M 15	M 13	M 56	M 64	M 63																	506
Product 24	M 02	M 11	M 05	M 07	M 08	M 15	M 10	M 23	M 24	M 18	M 21	M 27	M 68	M 68	M 69	M 57								531
Product 25	M 36	M 07	M 15	M 13	M 63																			3769
Product 26	M 36	M 02	M 07	M 15	M 13	M 56	M 60	M 62																2791
Product 27	M 02	M 09	M 03	M 05	M 06	M 07	M 08	M 15	M 10	M 23	M 22	M 25	M 16	M 21	M 27	M 61	M 76	M 77	M 57					382
Product 28	M 02	M 04	M 12	M 09	M 05	M 06	M 07	M 08	M 15	M 10	M 23	M 21	M 27	M 75	M 64	M 65	M 26							690
Product 29	M 02	M 28	M 30	M 32	M 29	M 23	M 08	M 10	M 23	M 09	M 24	M 24	M 63	M 68	M 69	M 57								976
Product 30	M 01	M 36	M 37	M 07	M 38	M 13	M 63	M 74	M 56															340
Product 31	M 02	M 36	M 07	M 15	M 13	M 63	M 68																	185
Product 32	M 52	M 48	M 51	M 49	M 50	M 54																		468953
Product 33	M 33	M 35	M 36	M 47	M 48	M 43	M 44	M 54	M 55															27326
Product 34	M 33	M 35	M 36	M 25	M 42	M 38	M 39	M 40																19430

Table 14. Levenshtein Distance Matrix

Products	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0	14	8	15	13	13	13	16	13	8	20	14	16	15	8	14	13	17	18	14	7	16	8	10	8	8	13	13	13	10	7	10	10	10	10	10
2	14	0	16	12	12	15	15	12	12	16	17	9	15	14	16	16	11	14	17	9	15	12	16	7	16	16	8	10	13	17	14	18	18	18	18	18
3	8	16	0	16	14	12	12	17	14	2	21	17	18	16	3	14	13	16	18	17	3	17	2	14	2	3	17	14	15	6	3	7	8	8	6	6
4	15	12	16	0	4	16	16	5	8	16	12	15	8	10	16	15	14	17	17	15	16	3	16	14	16	16	15	12	12	17	16	18	18	18	18	18
5	13	12	14	4	0	16	16	5	6	14	13	14	8	9	14	13	15	16	18	14	14	4	14	12	14	14	13	11	11	15	14	16	16	16	16	16
6	13	15	12	16	16	0	3	18	13	12	20	18	18	14	12	15	14	17	19	18	12	17	12	14	12	12	18	16	16	12	12	13	13	13	13	13
7	13	15	12	16	16	3	0	18	13	12	20	17	18	15	12	15	13	17	19	17	12	17	12	14	12	12	18	16	16	12	12	13	13	13	13	13
8	16	12	17	5	5	18	18	0	5	17	12	13	9	11	17	16	15	17	17	13	17	6	17	14	17	17	14	13	13	17	17	19	19	19	19	19
9	13	12	14	8	6	13	13	5	0	14	12	13	12	9	14	12	14	16	17	13	14	7	14	13	14	14	15	14	12	14	14	15	15	15	15	15
10	8	16	2	16	14	12	12	17	14	0	21	17	18	16	2	14	13	16	18	17	3	17	0	14	2	3	17	14	15	6	4	7	8	8	6	6
11	20	17	21	12	13	20	20	12	12	21	0	17	10	14	21	19	16	17	18	17	20	12	21	18	21	21	18	18	15	21	20	23	23	23	23	23
12	14	9	17	15	14	18	17	13	13	17	17	0	16	17	17	17	13	15	17	0	16	14	17	8	17	17	10	10	13	17	15	19	19	19	19	19
13	16	15	18	8	8	18	18	9	12	18	10	16	0	11	18	17	16	18	16	16	17	10	18	14	18	18	16	14	12	19	17	20	20	20	20	20
14	15	14	16	10	9	14	15	11	9	16	14	17	11	0	16	14	16	17	18	17	16	10	16	15	16	16	17	16	13	16	16	17	17	17	17	17
15	8	16	3	16	14	12	12	17	14	2	21	17	18	16	0	14	13	16	18	17	2	17	2	14	2	3	17	14	15	6	4	7	8	8	6	6
16	14	16	14	15	13	15	15	16	12	14	19	17	17	14	14	0	15	18	18	17	14	14	14	14	14	14	17	17	13	15	14	11	13	15	15	15
17	13	11	13	14	15	14	13	15	14	13	16	13	16	16	13	15	0	15	17	13	13	14	13	11	13	13	12	13	16	14	13	16	16	16	16	16
18	17	14	16	17	16	17	17	17	16	16	17	15	18	17	16	18	15	0	11	15	17	16	16	15	16	16	16	16	17	16	16	19	16	16	18	18
19	18	17	18	17	18	19	19	17	17	18	18	17	16	18	18	18	17	11	0	17	18	17	18	16	18	18	17	16	17	19	17	19	18	18	19	19
20	14	9	17	15	14	18	17	13	13	17	17	0	16	17	17	17	13	15	17	0	16	14	17	8	17	17	10	10	13	17	15	19	19	19	19	19
21	7	15	3	16	14	12	12	17	14	3	20	16	17	16	2	14	13	17	18	16	0	17	3	13	1	4	16	14	14	6	2	6	9	8	5	5
22	16	12	17	3	4	17	17	6	7	17	12	14	10	10	17	14	14	16	17	14	17	0	17	13	17	17	14	12	13	17	17	19	19	19	19	19
23	8	16	2	16	14	12	12	17	14	0	21	17	18	16	2	14	13	16	18	17	3	17	0	14	2	3	17	14	15	6	4	7	8	8	6	6
24	10	7	14	14	12	14	14	14	13	14	18	8	14	15	14	14	11	15	16	8	13	13	14	0	14	14	9	10	11	15	12	16	16	16	16	16
25	8	16	2	16	14	12	12	17	14	2	21	17	18	16	2	14	13	16	18	17	1	17	2	14	0	4	17	15	15	5	2	6	8	7	4	4
26	8	16	3	16	14	12	12	17	14	3	21	17	18	16	3	14	13	16	18	17	4	17	3	14	4	0	17	15	16	6	5	8	9	8	7	7
27	13	8	17	15	13	18	18	14	15	17	18	10	16	17	17	17	12	16	17	10	16	14	17	9	17	17	0	10	14	18	16	19	19	18	19	19
28	13	10	14	12	11	16	16	13	14	14	18	10	14	16	14	17	13	16	16	10	14	12	14	10	15	15	10	0	14	16	14	17	17	17	17	17
29	13	13	15	12	11	16	16	13	12	15	15	13	12	13	15	13	16	17	17	13	14	13	15	11	15	16	14	14	0	15	13	16	16	16	16	16
30	10	17	6	17	15	12	12	17	14	6	21	17	19	16	6	15	14	16	19	17	6	17	6	15	5	6	18	16	15	0	5	9	9	8	8	8
31	7	14	3	16	14	12	12	17	14	4	20	15	17	16	4	14	13	16	17	15	2	17	4	12	2	5	16	14	13	5	0	7	8	7	6	6
32	10	18	7	18	16	13	13	19	15	7	23	19	20	17	7	11	16	19	19	19	6	19	7	16	6	8	19	17	16	9	7	0	8	8	6	6
33	10	18	8	18	16	13	13	19	15	8	23	19	20	17	8	13	16	16	18	19	9	19	8	16	8	9	19	17	16	9	8	8	0	6	8	8
34	10	18	8	18	16	13	13	19	15	8	23	19	20	17	8	15	16	16	18	19	8	19	8	16	7	8	18	17	16	8	7	8	6	0	7	7

The second step of the methodology is performing a cluster analysis based on the Levenshtein distance. The chosen threshold value grouped the parts into seven clusters.

The third step of the methodology is developing the adjacency graphs for each of the product clusters. The Tables 15-23 and Figures 10-18 show the index sets and the directed graphs representing the From-To charts developed for each of the six clusters (Irani, 2005)

Table 15. Index Of Graph G1

Graph Index	Samples of products Used	Samples of products Used
G1	Product 1	Product 26
	Product 3	Product 30
	Product 10	Product 31
	Product 15	Product 32
	Product 21	Product 33
	Product 23	Product 34
	Product 25	Product 35
		Product 36

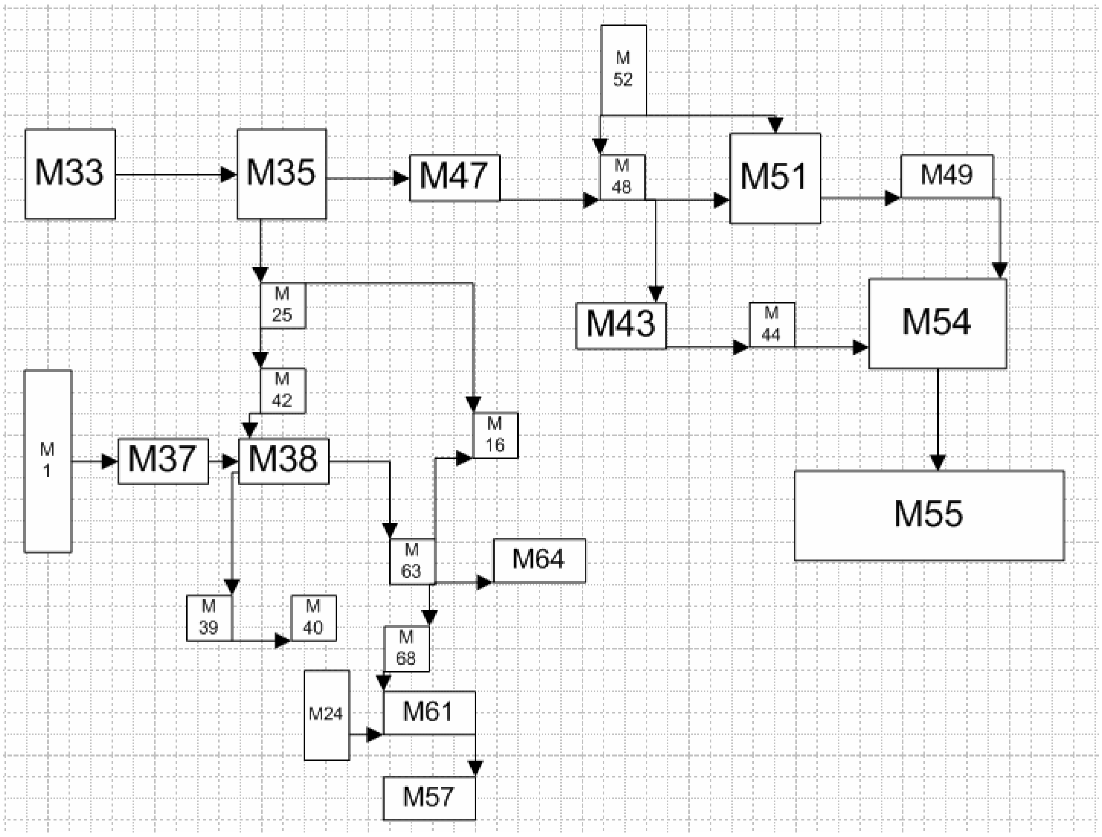


Figure 10. Adjacency Graph for G1

Table 16. Index Of Graph G2

Graph Index	Samples of Products Used
G2	Product 2
	Product 12
	Product 17
	Product 20
	Product 24
	Product 27
	Product 28
	Product 29

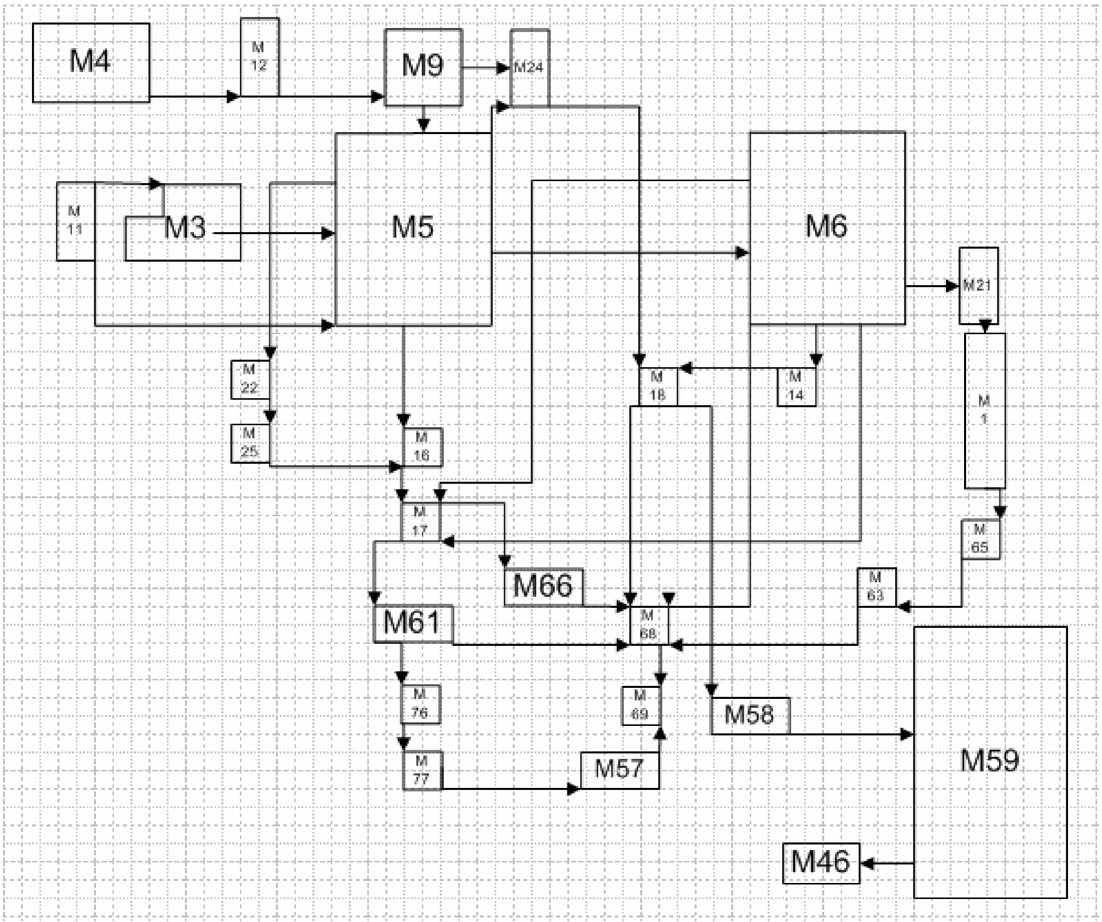


Figure 11. Adjacency Graph for G2

Table 17. Index of Graph G3

Graph Index	Samples of Products Used
G3	Product 4
	Product 5
	Product 8
	Product 9
	Product 13
	Product 14
	Product 22

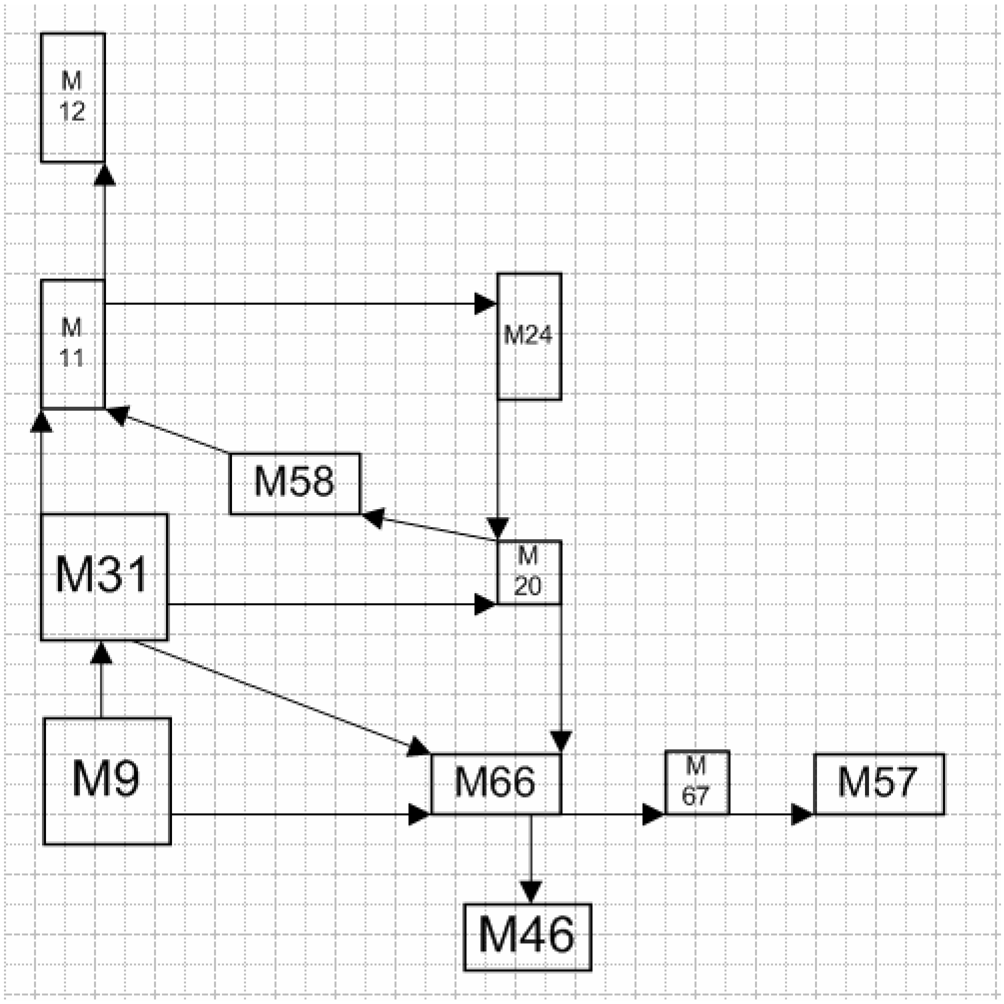


Figure 12. Adjacency Graph for G3

Table 18. Index of Graph G4

Graph Index	Samples of Products Used
G4	Product 6
	Product 7

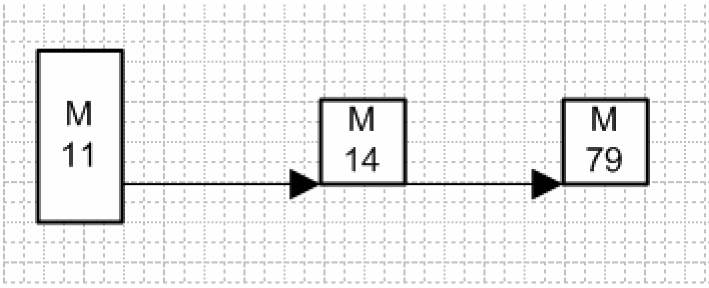


Figure 13. Adjacency Graph for G4

Table 19. Index of Graph G5

Graph Index	Samples of Products Used
G5	Product 11

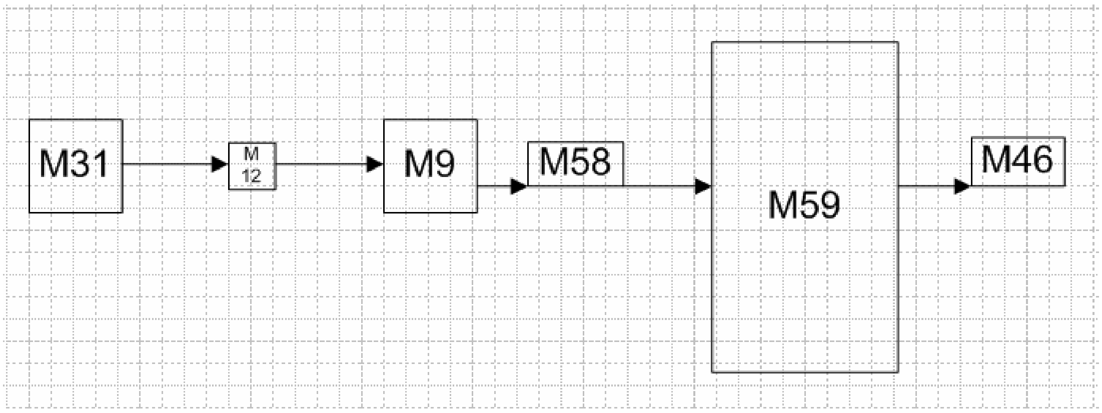


Figure 14. Adjacency Graph for G5

Table 20. Index of Graph G6

Graph Index	Samples of Products Used
G6	Product 16

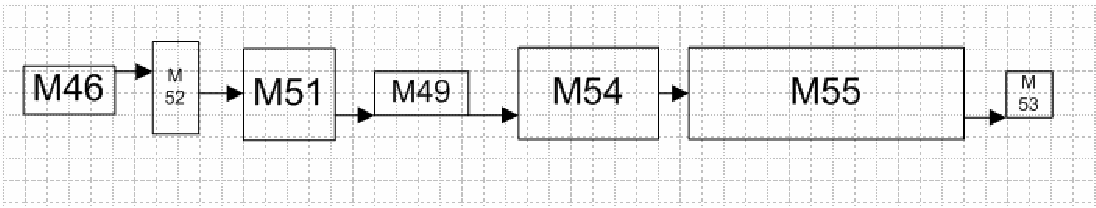


Figure 15. Adjacency Graph for G6

Table 21. Index of Graph G7

Graph Index	Samples of Products Used
G7	Product 18
	Product 19

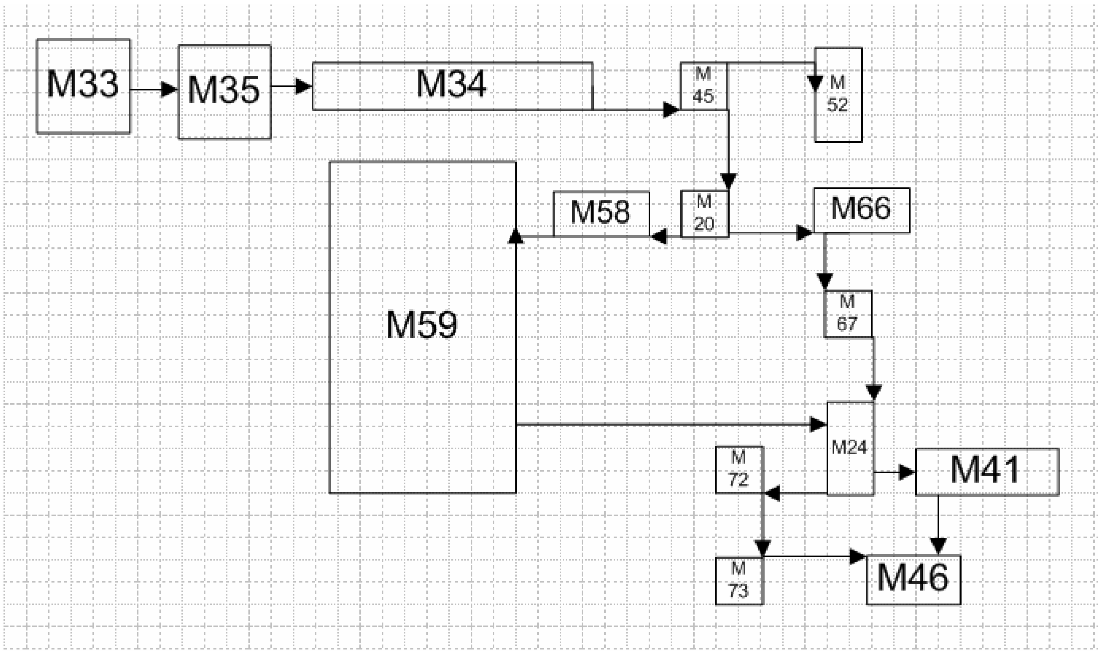


Figure 16. Adjacency Graph for G7

The fourth step of the methodology is embedding each of all adjacency graphs on the layout. The layout is given in Figure 17.

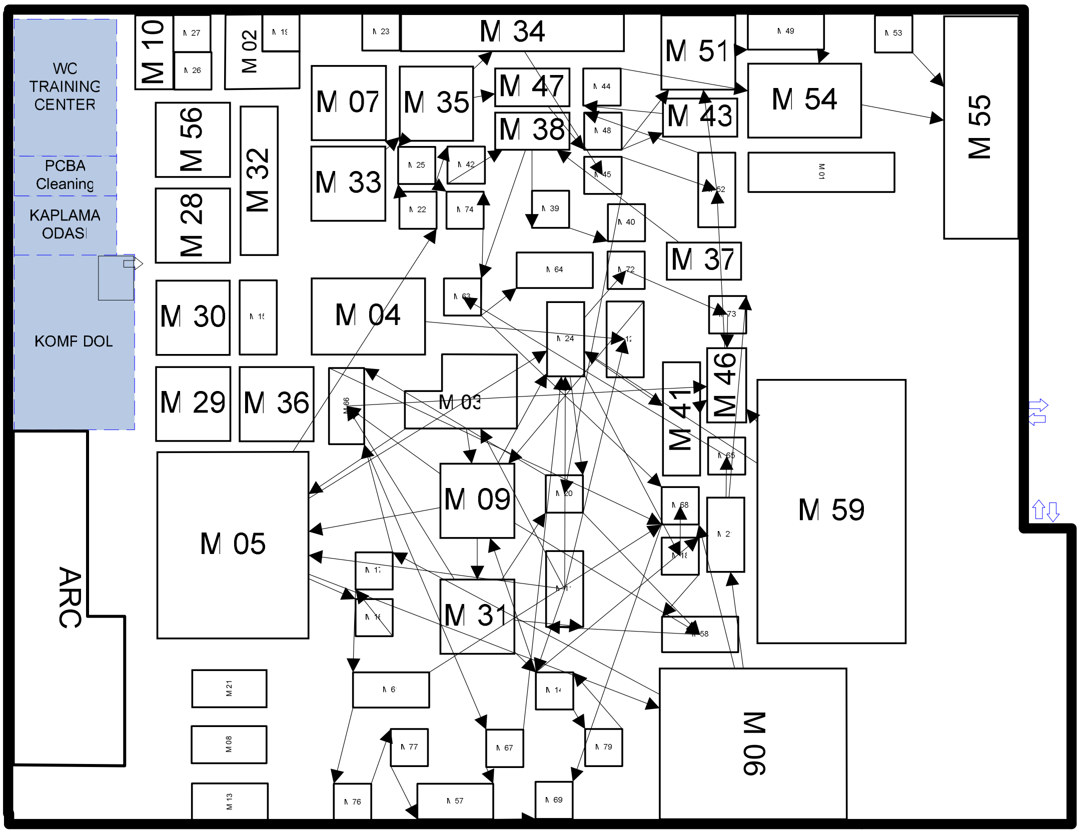


Figure 17. The Plant Layout Designed with Hierarchical Clustering with Levenshtein Distance

4.5. Performance Evaluation

In the scope of this thesis, four techniques are used. A comparison is made for these techniques by using technical and economical measurements. The definition of all criteria is mentioned in the literature survey section. The comparison of techniques is summarized in Table 22 to be overviewed.

After forming the block diagonalization matrix values of the number of 1s in diagonal blocks (*d)*, the number of exceptional elements in the solution *(e),* the number of ls in the part machine matrix *(o),* the number of voids in the solution *(v),* number of cells are obtained. By using the formulations of technical criteria mentioned in Section 2.1.3.2.1, the first five rows of the Table 22 are formed.

After designing the layouts for the applied techniques, the co-ordinates of the machines are defined. Table A3 in Appendix A0 shows the coordinates. The rectilinear distances between machines are calculated. In the solution procedure, we have already formed the machine-to-machine material handling matrix (Table C11). The distances between machines are multiplied with the material handling quantities in order to find the Material handling cost. In this calculation, it is assumed that transporting one unit of product for one unit of distance is equal to \$1 and it is same for all products. This calculation forms the sixth row of Table 22.

The unit carrying costs for machines are defined according to the area requirements of the machines. The unit cost of carrying a machine for one unit of distance is assumed directly proportional to the area requirement of the machine. The distances, which the machines carried in order to form the new layouts, are calculated by using the coordinates of current layout and new layouts. The rectilinear distances are used in these calculations. The distances that the machines carried are multiplied with the unit carrying costs of machines. It is assumed that carrying one machine for one unit of distance is equal to \$1000 and it is same for all machines. These results formed the seventh row of the Table 22.

Cost of between-cell material handling is explained in Section2.1.3.2.2. This expression is calculated by multiplying the total of material handling between clusters and the unit cost of transportation between cells. Total of material handling between clusters is calculated by summing the values of material handling between clusters matrices (Table 6-10-15). The cost of transportation between cells refers the managerial cost of transferring an item between cells. It is different from material handling cost. Solutions formed the last row of Table 22.

All the data in this thesis is changed in order not to share the critical information for the company. Especially, the cost information is kept as basic as possible. While changing the data, we have paid attention for not to lose the comparable logic of the solutions.

The solutions are summarized in Table 22

Table 22. Performance Evaluation Table

	Indicators	Current Layout	Tabular Method	Exp.Tabular Method	K-Means	Hier. Clus. With Levenshtein Dis.
Technical Criteria	**Group Efficiency**	-	0,625	0,492	0,625	-
	Group Efficacy	-	0,185	0,273	0,378	-
	Grouping measure	-	-0,045	-0,069	0,102	-
	Machine Utilization	-	0,259	0,569	0,868	-
	Eff. For alt routing	-	0,340	0,341	0,377	-
Economic Analysis	**Total Material Handling Cost ($)** *(Number of products × unit material handling cost × distance)*	464.628.281	236.147.010	260.979.403	190.040.066	212.422.588
	Cost Of Designing The Layout ($) *(Carrying distance × unit carrying cost)*	-	6.969.000	8.289.000	7.065.000	6.832.000
	Cost of Between-Cell Material Handling ($) *(Number of products × Unit cost)*	-	8.180.357	8.674.286	4.279.295	-
	Total Cost ($)	464.628.281	251.296.367	277.942.689	201.384.361	219.254.588

Because the current layout is not distinguished into cells, the technical criteria of the current layout are empty. Similarly, there will be no machine carrying activity in the plant when we continue using the current layout, so the machine carrying cost is zero for the current layout. Nevertheless, when we look at the total material handling cost the calculated value is 464.628.281 that is higher than the values of all other techniques used. By comparing, the total material handling costs, K-Means algorithm gave the best result. On the other hand, Hierarchical Clustering with Levenshtein Distance algorithm constructed a minimum value of Machine Carrying Cost.

The results of K–Means Algorithm dominate the results of Expanded Tabular Method, because all the values of Technical Criteria for K-Means are bigger than the results for Expanded Tabular Method. Similarly, all the economical measures for K-Means is less than the results for Expanded Tabular Method. Therefore, there is no need to consider the Expanded Tabular Method's results.

By comparing the results of Tabular Method and the K-Means Algorithm, we realize that a significant improvement is obtained with K-Means algorithm for the criteria of machine utilization and cost of between-cell material handling. The machine utilization is increased from 0,259 to 0,868. Cost of between-cell material handling is decreased from 8.180.357 to 4.279.295. It is obvious that while designing layout, the machine carrying activity with Tabular Method will be less than K-Means. Therefore, the machine carrying cost of Tabular Method is less than the K-Means'. However, this cost saving is not as significant as the machine utilization's and cost of between-cell material handling provided by K-Means.

By comparing the result of K-Means and Hierarchical Clustering with Levenshtein Distance these solutions can be obtained:

- Machine Replacement Cost is reduced from 7.065.000 to 6.832.000 by using Hierarchical Clustering with Levenshtein Distance. (%3,3 improvement)
- Total Material Handling Cost is reduced from 212.422.588 to 190.040.066 by using K-Means (%10,5 improvement).

Along with the discussed findings, it is concluded that the layout designed with K-Means clustering algorithm is the best solution.

CHAPTER V: CONCLUSION AND RECOMENDATIONS

In this thesis, production plant layout of a worldwide electronics company is designed. First stage of the study was to determine the principal problems and to collect data. The current production plant layout of the company and production process is analyzed thoroughly. By studying with the company authorities, problems and constraints are defined and the data collected is examined.

The complexity of operation sequences, the production dynamics, and the data immenseness, directed researches towards heuristic techniques. Considering the constraints, the layout is redesigned by using four techniques: Tabular Method, Expanded Tabular Method, K-Means Algorithm, and Hierarchical clustering method with Levenshtein distances. The results are compared by technical and economical evaluations. By dissecting the results, best layout is opted.

A solution procedure with three steps is examined. First, clusters are formed by using the proposed methodologies. Second, layouts of the cells are designed. Determining the best location for cells was the last step.

The solution of Tabular Method provided an initial improvement in the current layout. The method resulted in eight clusters. The nature of the algorithm does not allow considering the annual demand for products. In order to amend the layout the Expanded Tabular Method is applied. The main extension of this technique is using the material handling machine-to-machine matrix as an input. By using this technique, group efficacy, grouping measure and machine utilization index is positively affected. On the other hand, the group efficiency is significantly decreased. The cost of material handling between cells and total material handling cost are also kincreased.

The third technique used is K-means Clustering Algorithm. Unlike the others, this technique does not determine the number of clusters. Defining the number of clusters before running the algorithm is essential. Hence, the decision of the designer affects the layout's performance. In order to decrease the subjectivity, we run the algorithm for different number of clusters. The number of cluster is selected by comparing the clustering measures. Results that were more satisfying are obtained by using the methodology. Therefore, the designed layout by this methodology is preferred.

Finally, The Hierarchical Clustering with Levensthein distances is applied to the problem. The technique provides taking into consideration of the operation sequences. Unlike the others, this methodology clusters the parts having similar routes.

Results obtained are evaluated by using technical and economical criteria. There are five technical and three economical measures are examined in this study.

The solutions for five technical measurements depend mostly on the perfection of block-diagonalization matrix. The creativity of designer mostly influences the performance of block-diagonalized matrix. This prevents objectivity of the measurements. In order to decrease this subjectivity, some heuristics are suggested in the literature. An important factor, cost, is introduced in the economic analysis section. The unit costs are set as one therefore; the relative performance of the costs is not comparable. In addition, the other performance measures such as solving CPU time are not included in this study.

The grouping heuristic procedure descriptions, along with the discussed findings, support the following conclusive statements:

- The best grouping results are achieved by means of proposed heuristics with the selection strategy, K-Means Algorithm, which best fits, the nature of our case constraints.
- All of the considered performance indicators are positively affected by the proposed methodology, K-Means Algorithm.

In this study, by evaluating different techniques, an alternative layout is designed and suggested. A 59 % improvement is obtained in material handling.

In this study, it is not possible to find out the amount of work in process inventory. By using arrival and production rates, a simulation analysis can be done as a future work. In addition, the queuing theory can be applied in order to find out the expected waiting times. On the other hand, different techniques can be used such as an optimization technique. It is mentionable that the studies can be collected to form a Decision Support System.

REFERENCES

Adıl, G.K., Rajamanı, D. And Strong, D., 1996, "Cell Formation Considering Alternate Routings", International Journal Of Production Research, 34(5), 1361-1380.

Akturk, S.M. And Turkcan, A., 2000, "Cellular Manufacturing System Design Using A Holonistic Approach", International Journal Of Production Research, Vol.38, No.10, 2327-2347.

Andrew Moore, "K-Means And Hierarchical Clustering - Tutorial Slides" Http://Www-2.Cs.Cmu.Edu/~Awm/Tutorials/Kmeans.Html, 2005.

Arieh, D.B., 1998, "Analysis Of A Distributed Group Technology Methodology," Computers And Industrial Engineering, Vol. 35, No. 1-2, 69-72.

Askin, R G., Creswell, J. B., Goldberg, J. B., Vakharia, A. J., 1991, "A Hamiltonian Path Approach To Reordering The Part-Machine Matrix For Cellular Manufacturing", International Journal Of Production Research, Vol. 29, Pp. 1081-1100.

Askın, R.G. And Subramanıan, S.P., 1987, "A Cost-Based Heuristic For Group Technology Configuration", International Journal Of Production Research, 25(1), 101-113.

Ballakur, A. And Steudel, H.J., 1987, "A Within-Cell Utilization Based Heuristic For Designing Cellular Manufacturing System". International Journal Of Production Research, 25(5), 639-665.

Bazargan-Ları, M. And Kaebernıck, H., 1996, “Intracell And Intercell Layout Designs For Cellular Manufacturing”, International Journal Of Industrial Engineering, 3(3), 139-150.

Beaulıeu, A., Gharbı, A. And Aıt-Kadı, 1997, “An Algorithm For Cell Formation And Machine Selection Problems In The Design Of A Cellular Manufacturing System”, International Journal Of Production Research, 35(7), 1857-1874.

Boctor, F.F., 1991, “ A Linear Formulation Of The Machine Part Cell Formation”, International Journal Of Production Research, Vol. 29, No. 2, 343-356.

Brian T. Luke, 2005 “K-Means Clustering”, Http://Fconyx.Ncifcrf.Gov~ Lukeb/Kmeans.Html

Burbidge, J.L., 1991, “Production Flow Analysis For Planning Group Technology”, Journal Of Operations Management, Vol. 10, No.1, 5-28.

Burbidge,J.L. 1975. Introduction Of Group Technology. John Wiley And Sons., New York.

Carrie A.S., 1975, “Layout Of Multi-Product Lines”, International Journal Of Production Research, Vol. 13, No. 6, 451-467.

Chan, H. M, Milner, D. A., 1982, “Direct Clustering Algorithm For Group Formation In Cellular Manufacturing”, Journal Of Manufacturing Systems, Vol. 1, Pp. 65-74.

Chandrasekharan, M. P., Rajagopalan, R., 1989, “Group ability: An Analysis of the Properties of Binary Data for Group Technology”, International Journal of Production Research, Vol. 27, Pp. 1035-1052.

Chandrasekharan, M.P. And Rajagopalan, R., 1987, “Zodıac - An Algorithm For Concurrent Formation Of Part Families And Machine Cells”, International Journal Of Production Research, 25(6), 835-850.

Chandrasekharan M.P. And Rajagopalan, R., 1986, “ An Ideal Seed Non-Hierarchical Clustering Algorithm For Cellular Manufacturing”, International Journal Of Production Research, Vol. 24, No. 2, 451-464.

Chandrasekharan M.P. And Rajagopalan, R., 1986, “Modroc: An Extention Of Rank Order Clustering For Group Technology”, International Journal Of Production Research, Vol. 24, No. 5, 1221-1233.

Chen, J.S., And Heragu S.S., 1999, “Stepwise Decomposition Approaches For Large-Scale Cell Formation Problems, European Journal Of Operations Research”, Vol. 113, 64-79.

Cheng, C.H., Kumar, A., And Motwani , J, 1995, “ A Comparative Examination Of Selected Cellular Manufacturing Clustering Algorithms”, International Journal Of Operations And Productions Research, Vol. 15, No. 1-2, 86-98.

Choobıneh, F., 1988, “A Framework For The Design Of Cellular Manufacturing Systems”, International Journal Of Production Research, 26(7), 1161-1172.

Chu, C., H., Kumar, A., And Motwani, J., 1995, “A Comparison Of Three Array-Based Clustering Techniques For Manufacturing Cell Formation”, International Journal Of Production Research, Vol. 28, No. 8, 1417-1433.

D'andrade R., 1978, "U-Statistic Hierarchical Clustering", Psychometrika, Vol. 4, 58-67.

Del Valle, A.G., Balarezo, S. And Tejero, J., 1994, “A Heuristic Workload Based Model To Form Cells By Minimizing Intercellular Movements”, International Journal Of Production Research, 32(10), 2275-2285.

Fan, T. T., Edward F. S. J. And Jatinder N.D.G, 2004, “Empirical Analysis Of Integer Programming Formulations For The Permutation Flowshop”, The International Journal Of Management Science, Vol. 32, 285-293.

Gupta, T. And Seıfoddını, H., 1990, “Production Data Based Similarity Coefficient For Machine Component Grouping Decisions In The Design Of A Cellular Manufacturing System”, International Journal Of Production Research, 28(7), 1247-1269.

Harhalakıs, G., Nagı, R. And Proth, J.M., 1990, “An Efficient Heuristic İn Manufacturing Cell Formation For Group Technology Applications”, International Journal Of Production Research, 28(1), 185-198.

Heragu, S.S. And Gupta, Y.P., 1984, “A Heuristic For Designing Cellular Manufacturing Facilities”, International Journal Of Production Research, 32(1), 125-140.

Hiller, F.S, Connors, M.M., 1966, “Quadratic Assignment Problem Algorithms And The Location Of Indivisible Locations”, Management Science, Vol. 13, No. 1, 42-57.

Irani, S. A., Khator, S. K., 1987, “Cell Formation In Group Technology: A New Approach”, Computers And Industrial Engineering, Vol. 12, Pp. 131-142.

Irani, S.A., 1999. Handbook Of Cellular Manufacturing Systems. John Wiley And Sons., New York.

Irani, S.A., 2005. “Enhancements in Facility Layout Tools Using Cell Formation Techniques”, Department of Industrial, Welding and Systems Engineering, The Ohio State University Columbus, OH 43210

Jain, A.K., Murty, M.N., Flynn, P.J., 1999, “ Data Clustering A Review”, Acm Computing Surveys, Vol. 31, No. 3, 265-323.

Johnson, S. C., 1967, “Hierarchical Clustering Schemes", Psychometrika, Vol. 2, 241-254.

Joines, A.J., King R.E., And Culbreth, C.T., 1995, “A Comprehensive Review Of Production-Oriented Manufacturing Cell Formation Techniques”, International Journal Of Flexible Automation And Integrated Manufacturing”, Vol. 3, No. 3-4, 225-264.

Joınes, J.A., Culbreth, C.T. And Kıng, R.E., 1996, “Manufacturing Cell Design: An Integer Programming Model Employing Genetic Algorithms”, IIE Transactions, 28(1), 69-85.

Kamrani A.K., And Parsaei H.R., 1993, “A Group Technology Based Methodology For Machine Cell Formation In A Computer Integrated Manufacturing Environment”, Computers And Industrial Engineering, Vol. 24, 431-447.

Kandiller, L., 1998, “A Cell Formation Algorithm: Hypergraph Approximation- Cut Tree”, European Journal Of Operational Research, Vol. 109, 686-702.

Kıng, J.R. And Nakornchai, V., 1982, “Machine Component Group Formation In Group Technology: Review And Extension”, International Journal Of Production Research, 20(2), 117-133.

Kumar, K., R., Chandrasekharan, M.P., 1990, “Grouping Efficacy: A Quantitative Criterion For Block Diagonal Forms Of Binary Matrices In Group Technology”, Inernational Journal Of Production Research, Vol. 28, No. 2, 233-243.

Kusiak, A., Chow, W. S., 1987, “Efficient Solving Of The Group Technology Problem”, Journal Of Manufacturing Systems, Vol. 6, Pp. 117-124.

Lee, H. And Garcıa-Dıaz, A., 1996, “Network Flow Procedures For The Analysis Of Cellular Manufacturing Systems”, IIE Transactions, 28, 333-345.

Lın, T.L., Dessouky, M.M., Kumar, K.R. And Ng, S.M., 1996, “A Heuristic Based Procedure For The Weighted Production Cell Formation Problem”, IIE Transactions, 28, 579-589.

Logendran, R., 1990, “Aworkload Based Model For Minimizing Total Intercell And Intracell Moves”, International Journal Of Production Research, 28(5), 913-925.

Mcauley, J., 1972, “Machine Grouping For Efficient Production”, The Production Engineer, 51(2), 53-57.

Mccormıck, W.T., Schweıtzer, P.J. And Whıte, T.W., 1972, “Problem Decomposition And Data Reorganization By A Clustering Technique”, Operations Research, 20(5), 993-1009.

Macqueen, J. B., 1967, "Some Methods For Classification And Analysis Of Multivariate Observations, Proceedings Of 5-Th Berkeley Symposium On Mathematical Statistics And Probability", Berkeley, University Of California Press, 281-297.

Mansouri, S.A., Husseini, S.M.M, And Newman, S.T., 2000, “Review Of Modern Approaches To To Multi-Criteria Cell Design”, International Journal Of Production Research. Vol. 38, No. 5, 1201-1218.

Maria Irene Miranda, "Clustering Methods And Algorithms", Http://Www.Cse. Iitb.Ac.In/Dbms/Data/Courses/Cs632/1999/Clustering/Dbms.Htm, 2005.

Miltenburg, J., And Zhang, W., 1991, "A Comparative Evaluation Of Nine Well-Known Algorithms For Solving The Cell Formation Problem In Group Technology", Journal Of Operations Management, Vol. 10, 44-72.

Nagı, R., Harhalakıs, G. And Proth, J.M., 1990, "Multiple Routings And Capacity Considerations İn Gt Applications", International Journal Of Production Research, 28(12), 2243- 2257.

Offodile, O.F., Mehrez. A., And Grznar. J., 1994, "Cellular Manufacturing: A Taxomomic Review Framework", Journal Of Manufacturing Systems, Vol. 13, No. 3, 196-220.

Pardalos P.M. And H.Wolkowicz, 1994, "Quadratic Assignment And Related Problems", American Mathematical Society, Vol. 16.

Rajamanı, D., Sıngh, N. And Aneja, Y.P., 1992, "A Model For Cell Formation İn Manufacturing Systems With Sequence Dependence", International Journal Of Production Research, 30(6), 1227-1235.

Reza, H.A, Panagiotis, K., 1999, "Design Of Electronics Assembly Lines: An Analytical Framework And Its Application", European Journal Of Operational Research, Vol. 115, 113-137.

Russel D. Meller And Kia-Yin Gau, 1996, "An Investigation Of Facility Layout Objective Functions And Robust Layouts", International Journal Of Production Research, Vol. 34, No. 10, 2727–2742.

Russel D. Meller And Kia-Yin Gau, 1996. “The Facility Layout Problem: Recent And Emerging Trends And Perspectives”, Journal Of Manufacturing Systems, Vol.15, No. 5, 351- 366.

Russell D.M., Venkat, N., Pamela, H. V., 1999, “Optimal Facility Layout Design”, Operations Research Letters, Vol.23, 117-127.

Sarker, B.R. And Balan, C.V., 1996, “Cell Formation with Operation Times Of Jobs For Even Distribution Of Workloads”, International Journal Of Production Research, 34(5), 1447-1468.

Sarker, B. R, 2001, “Measures Of Grouping Efficiency In Cellular Manufacturing Systems”, European Journal Of Operational Research, Vol. 130, 588-611.

Sarker, B. R, Li, Z., 1998, “Measuring Matrix Based Cell Formation With Alternative Routings”, Journal Of The Operational Research Society, Vol. 49, No. 9, 953-965.

Sarker, B. R., Mondal, S., 1999, “Grouping Efficiency Measures In Cellular Manufacturing: A Survey And Critical Review”, International Journal Of Production Research, Vol. 37, No. 2, 285- 314.

Seıfoddını, H. And Djassemı, M., 1995, “Merits Of The Production Volume Based Similarity Coefficient In Machine Cell Formation”, Journal Of Manufacturing Systems, 14(1), 35-44.

Seıfoddını, H. And Wolfe, P.M., 1987, “Selection Of A Threshold Value Based On Material Handling Cost In Machine Component Grouping”, IIE Transactions, 19(3), 266-270.

Singh, N., 1993, “Design Of Cellular Manufacturing: An Invited Review”, European Journal Of Operational Research, Vol. 69, 284-291.

Singh, N., Rajamani, D., 1996. Cellular Manufacturing Systems / Design Planning And Control. Chapman And Hall, London.

Snead, C. S., 1989, Group Technology: Foundation For Competitive Manufacturing. Van Nostrand Rein Hold., New York.

Sofianopoulou, S., 1999, “Manufacturing Cells Design With Alternative Process Plans And Replicate Machines”, International Journal Of Production Research, Vol.37, No.3, 707-720.

Stephen P. Borgatti, “How To Explain Hierarchical Clustering”, Http://Www.Analytictech.Com/ Networks/Hiclus.Htm

Sule D.R., 1994. Manufacturing Facilities, Location, Planning, And Design. Pws Publishing Company, Boston.

Tam, K.Y., 1990, “An Operation Sequence Based Similarity Coefficient For Part Families Formation”, Journal Of Manufacturing Systems, 9(1), 55-68.

Thomas, D., Günter, R., 2004, Engelbert W., “Combining Evolutionary Computation And Dynamic Programming For Solving A Dynamic Facility Layout Problem”, Vol. 60, 311-324.

Vakharıa, A.J. And Chang, Y.-L., 1997, “Cell Formation In Group Technology: A Combinatorial Search Approach”, International Journal Of Production Research, 35(7), 2025- 2043.

Venugopal , V. And Narendran, T.T., 1992, “Cell Formation In Manufacturing Systems Through Simulated Annealing: An Experimental Evaluation”, European Journal Of Operations Research”, Vol. 63, 409-480.

Wang, J.,1998, “A Linear Assignment Algorithm For Formation Of Machine Cell And Part Families In Cellular Manufacturing” , Computers And Industrial Engineering, Vol. 35, No. 1-2, 81-84.

Wei, J.C., And Gaither, N., 1990, “A Capacity Constrained Multi Objective Cell Formation Method”, Journal Of Manufacturing Systems, Vol. 9, 222-232.

Wemmerlov, U., And Johnson, D.J., 1997, “Cellular Manufacturing At User Plants: Implementation Experiences And Performance Improvements”, International Journal Of Production Research, Vol. 35, No. 1, 29-49.

Zhou, M. And Askın, R.G., 1995, “Formation Of General Group Technology Cells: An Operation Based Approach”, Technical Report, University Of Arizona.

APPENDIX A

Table A.1. A Class Products, Their Rates And Cumulative Percentages

Item	Total	Rate	Total * Rate	Cum %	Class
5	106608	12,29	1310532,14	25	A
7	63416	6,42	407130,72	32	A
351	468953	0,74	348901,03	39	A
193	2791	68,64	191571,45	42	A
6	31150	5,39	168023,10	46	A
89	3755	34,47	129416,08	48	A
352	27326	4,37	119523,92	50	A
88	1806	53,93	97393,97	52	A
158	22897	4,17	95480,49	54	A
160	2184	40,44	88329,70	56	A
119	2380	35,41	84268,66	57	A
176	3112	23,49	73100,88	59	A
138	1254	57,42	72004,68	60	A
98	9155	7,71	70585,05	61	A
362	7880	8,88	70005,92	63	A
155	2375	26,17	62149,00	64	A
167	1046	47,22	49392,12	65	A
151	2048	22,93	46964,74	66	A
111	3016	14,73	44425,68	66	A
363	5000	8,88	44420,00	67	A
173	513	83,60	42886,80	68	A
120	741	57,79	42823,87	69	A
163	1352	31,47	42547,44	70	A
188	3769	11,13	41945,20	70	A
236	690	55,87	38547,54	71	A
106	7953	4,80	38174,40	72	A
107	6717	5,34	35868,78	72	A
87	954	37,05	35349,52	73	A
288	976	35,46	34605,06	74	A
148	878	39,38	34579,15	74	A
4	3959	8,71	34471,01	75	A
216	382	89,20	34075,16	76	A
110	1467	20,94	30718,98	76	A
355	19430	1,57	30543,96	77	A
166	803	37,51	30117,32	77	A
180	506	56,63	28655,79	78	A
28	276	103,32	28516,32	79	A
313	340	83,43	28365,52	79	A
3	2995	8,59	25712,08	80	A
183	531	47,89	25427,47	80	A
94	1784	13,86	24726,24	81	A
320	185	129,76	24005,60	81	A

Table A.2. The information of percentages and numbers of items passing through machine (i) is given in

	(X)	(X/36)		(X)	(X/36)
Machines (i)	Number Of Items Pass Through Machine(i)	Percentage Of Items Pass Through Machine(i)	Machines (i)	Number Of Items Pass Through Machine(i)	Percentage Of Items Pass Through Machine(i)
M15	26	72%	M17	3	8%
M7	25	69%	M33	3	8%
M21	17	47%	M52	3	8%
M8	16	44%	M54	3	8%
M2	15	42%	M18	2	6%
M10	15	42%	M25	2	6%
M26	15	42%	M34	2	6%
M23	14	39%	M38	2	6%
M27	14	39%	M45	2	6%
M36	14	39%	M48	2	6%
M13	12	33%	M49	2	6%
M56	12	33%	M51	2	6%
M32	11	31%	M55	2	6%
M19	10	28%	M1	1	3%
M28	10	28%	M4	1	3%
M29	10	28%	M12	1	3%
M30	10	28%	M22	1	3%
M63	9	25%	M37	1	3%
M31	8	22%	M39	1	3%
M5	7	19%	M40	1	3%
M9	7	19%	M41	1	3%
M20	7	19%	M42	1	3%
M24	7	19%	M43	1	3%
M66	7	19%	M44	1	3%
M67	6	17%	M46	1	3%
M68	6	17%	M47	1	3%
M11	5	14%	M50	1	3%
M57	5	14%	M53	1	3%
M58	5	14%	M60	1	3%
M64	5	14%	M62	1	3%
M69	5	14%	M65	1	3%
M78	5	14%	M70	1	3%
M3	4	11%	M71	1	3%
M6	4	11%	M72	1	3%
M14	4	11%	M73	1	3%
M35	4	11%	M74	1	3%
M59	4	11%	M75	1	3%
M61	4	11%	M76	1	3%
M16	3	8%	M77	1	3%
			M79	1	3%

Table A.3. Coordinates of Machines

	Current Layout		Tabular Method		Exp Tabular Method		K-Means		Clus.With Leven. Dis.	
	X	Y	X	Y	X	Y	X	Y	X	Y
M1	10,8	76,9	14,6	71,6	49,6	45,4	40,2	38,5	40,9	64,9
M2	14,1	74,7	36,4	56,2	24,4	52,6	32,6	53	15,8	69,5
M3	17,3	74,8	44,5	48,9	29	77,7	8,7	55,1	26,9	52,9
M4	15,1	79	30,8	43,4	44,3	40,2	38,1	42,8	27,1	61,5
M5	4,6	42	12,5	42,2	40,7	76,6	7,6	47,7	19,6	53,9
M6	4,6	52,6	21,1	42,2	40,7	68	16,2	47,7	39,9	44,5
M7	22,5	74,8	36,4	64,6	24,3	65,3	32,6	48,8	19,9	69,6
M8	27,8	79	39,6	64,6	33,7	61	27,1	57,3	13,2	55,6
M9	26,7	74,8	29,7	47,6	33,3	65,8	14	55	28,2	57,1
M10	30,9	80,1	40,7	57,2	28,6	53,6	32,6	58,4	10,4	69,5
M11	19,3	79	32,9	47,6	39,8	62	42,3	40,7	31,3	50,8
M12	21,5	79	30,7	64,6	44,2	44,3	42,3	42,9	31,3	61,5
M13	23,6	79	43,9	66,8	32,7	53,6	25	57,3	13,2	53,4
M14	42,5	80,1	9,4	48,6	36,6	62	41,2	44,9	31,3	47,6
M15	32	74,8	43,8	58,3	24,3	62,1	31,6	44,6	20,9	65,4
M16	44,3	67,5	11,5	48,6	38,7	57,7	17,6	56,1	25	47,7
M17	44,3	69,6	46,7	45,8	32,2	75,6	65,8	56,2	24,9	49,8
M18	40	71,7	9,4	50,7	40,9	59,8	37	47,1	33,5	51,9
M19	44,3	65,4	48	59,3	31,7	51,5	33,7	51,9	16,8	70,6
M20	49,3	71,7	13,7	48,6	42,8	50,9	41,5	73,9	31,3	54,1
M21	52	70,6	41,8	64,6	31,7	61	29,3	57,3	35,6	53
M22	49,3	69,6	32,8	63,5	34,6	46,3	37,1	49,3	23	64,5
M23	44,3	71,7	43,8	61,5	27,5	62,1	33,7	45,6	20,9	72,8
M24	52	66,4	29,7	40,1	36,5	52	33	40,9	31,4	57,2
M25	40	69,6	11,6	50,7	36,7	42,1	39,1	47,1	22,9	66,6
M26	42,1	71,7	43,8	63,6	30	51,5	31,5	56,2	12,6	68,5
M27	49,3	67,5	46	59,4	30	62,1	33,6	56,2	12,5	70,6
M28	16	68,8	47	66,9	32,7	56,8	24,1	53	11,1	61
M29	11,8	68,8	36,4	60,4	28,5	58,9	24,1	46,6	17,6	61,2
M30	20,2	68,8	40,7	60,4	24,3	58,9	28,4	46,6	17,6	65,3
M31	24,4	68,8	17	49,7	41,8	54,8	21	57,8	28,2	46,6
M32	30,7	67,8	40,6	55,1	26,4	55,7	26,2	49,8	14,4	63,2
M33	11,8	62,4	40,8	51,5	39,9	39,1	38,3	70,8	24,2	69,9
M34	32,8	63,5	26,4	43,3	40	48,6	35,1	66,5	28,4	73,2
M35	32,8	63,5	36,6	51,4	40	43,2	38,3	66,6	28,3	69,9
M36	24,4	62,4	47	62,6	28,6	65,3	28,3	42,4	21,8	61,2
M37	28,6	69,8	11,4	68,4	47,4	39,1	41,2	48,2	34,5	64,7
M38	32,8	69,9	23,2	48,6	36,7	45,3	43,4	50,4	28,3	66,8
M39	47,2	47,1	24,3	50,7	36,7	40	43,4	44,9	31,5	66,8
M40	49,3	47,1	22,1	50,7	38,8	46,3	43,4	47,2	33,7	66,8
M41	29,7	61,4	30,8	50,8	23,4	71,2	39,4	63,4	33,5	58,4
M42	40,1	61,2	26,5	50,8	41	46,3	39,2	49,3	25,1	66,8

Table A.3. Coordinates of Machines (cont'd)

M43	39,3	80,1	37,6	47,3	9,9	54,8	38,1	51,5	39	71,3
M44	42,2	61,2	39,8	48,3	16,2	62,2	41,3	51,4	35,8	73,1
M45	38	61,2	32,8	65,7	45,4	46,5	41,5	71,9	31,5	68,9
M46	16	61,4	31,5	68,4	10,9	42,2	45,6	39,5	35,7	59,5
M47	35,1	80,1	35,6	47,3	9,9	59	38,1	53,7	32,6	71
M48	29,9	74,8	17,5	73,7	12,1	54,8	46,7	47,1	35,8	70,9
M49	54,1	50,7	35,7	72,6	29,7	70,1	45,5	50,4	47,3	75,6
M50	54,9	74,9	31,5	71,5	15,2	54,8	46,6	43,9	47,4	72,5
M51	54,9	79,1	35,7	69,4	26,6	72,3	48,7	50,3	38,9	74,6
M52	25,7	79	38,9	71,5	26,5	69	44,5	53,6	35,8	67,8
M53	44,3	61,2	38,9	68,3	14,1	42,2	41,4	53,6	46,1	64,9
M54	36,3	40,8	21,8	73,7	14,2	59,9	44,5	57,9	43,1	73,5
M55	26,8	40,8	22,7	69,5	19,5	58,9	48,7	58,8	43,2	68,1
M56	45,4	40,9	47	56,2	33,3	70,1	28,3	53,1	11,2	65,3
M57	49,6	41,9	45,5	52	38,6	52	33	38,9	26	43,5
M58	54,1	56,2	31,7	53	36,5	58,8	38,3	61,2	34,6	49,8
M59	24,75	49,65	17,9	58,75	22,05	43,85	24,6	67,2	40,9	56,2
M60	54,1	53,6	47,2	72,5	26,5	51,5	39,3	56,9	10,4	66,3
M61	54	44,8	31,7	55,1	33,2	79,8	18,3	53,9	25,2	47,7
M62	43,7	49,2	47,2	70,5	47,4	43,3	41,4	56,9	10,4	64,3
M63	55,9	47,9	43,5	70,6	40,9	57,6	24,1	37,9	30,7	64,6
M64	48,7	56,2	41,4	69,5	35,7	38	22	38,9	33,9	64,6
M65	53,8	47,9	30,6	57,2	49,6	38	47,7	38,5	39,1	56,3
M66	48,9	50,7	20,1	49,7	34,3	74,5	24,5	41,7	28,4	57,2
M67	51,7	47,9	13,7	50,7	40,7	50,9	22,5	42,9	28,4	45,6
M68	42,7	46,6	42,4	45,8	34,4	77,7	4,4	56,2	36,9	54,1
M69	44,8	46,6	44,5	45,8	32,2	77,7	4,4	53,9	32,6	43,5
M70	42,6	56,2	30,6	59,3	38,6	55,2	49,8	38,5	37	45,5
M71	44,7	56,2	32,8	59,3	36,5	55,2	49,8	40,6	34,9	45,5
M72	53,3	41,9	32,8	57,2	43,2	46,4	39,4	73,9	36,9	62,7
M73	55,4	41,9	30,6	61,4	34,6	42,1	35,1	73,9	39,1	62,7
M74	57,5	41,9	17,9	65,7	49,6	40,1	47,7	40,6	28,6	64,6
M75	43,7	53,7	21,1	65,7	47,4	47,5	49,8	43,8	15	72,7
M76	47,9	44,8	24,4	65,6	34,6	40	37,1	55,8	24	45,6
M77	50	44,8	26,5	65,7	34,6	44,2	37,1	57,9	26,2	45,6
M78	43,8	51,4	32,8	61,4	32,2	73,4	37,3	73,9	18,2	72,7
M79	42,1	69,6	28,6	65,6	38,7	59,8	49,8	47,1	36,9	47,6

APPENDIX B

B.1. The Visual Basic Code Of Forming Machine To Machine Matrix

```
*************************************************************
* Machine to machine Algorithm
*
* Basar Uyanık
* 05 / 2005
*************************************************************
Sub Macro1()
Dim a As Integer
Dim i As Integer
Dim j As Integer
Dim toplam As Integer
For a = 1 To 79
        For j = 1 To 79
            toplam = 0
            For i = 0 To 36
               Range("a1").Select
               ActiveCell.Offset(i, a - 1).Select
               If ActiveCell.Value = 1 Then
                       ActiveCell.Offset(0, j).Select
                       If ActiveCell.Value = 1 Then
                               toplam = toplam + 1
                       End If
                       ActiveCell.Offset(0, -j).Select
               End If
            Next i
        Range("CE1").Select
        ActiveCell.Offset(j + a, a - 1).Select
        ActiveCell.Value = toplam
        Next j
Next a
End Sub
```

B.2. The Visual Basic Code for Tabular Method

```
*****************************************************************
* Tabular/ Expanded Tabular Algorithm
*
* Basar Uyanık
* 05 / 2005
*****************************************************************
* Public variable declarations
*****************************************************************
Public RC As Integer
Public P As Double
Public MTV As Double
Public MCR As Double
Public grp_counter As Integer
Public temp_RC
*
******************************************************************
* This fuction calculates maximum RC value
******************************************************************
*
Sub find_RC()
        * P value is entered from the cell B1 by the user
        P = Worksheets("RC").Range("B1").Value
        max = 0
        * Activate data matrix, first cell is A3
        Worksheets("veri").Activate
        Range("a3").Select
        counter = 1
        * Search through all the matrix
        For j = 1 To 79 Step 1
                For i = 1 To counter Step 1
                        ActiveCell.Offset(0, 1).Select  ' row,col
                        temp = ActiveCell.Value
                        If temp > max Then
                                max = temp
                        End If
                Next i
                n_counter = counter * -1
                counter = counter + 1
                ActiveCell.Offset(1, n_counter).Select
        Next j
        * Write the RC value into RC sheet B2 cell
        Worksheets("RC").Range("B2").Value = max
        RC = max
        MTV = RC * P
End Sub
*
*****************************************************************
```

```
* This function creates a relational matrix(RM) in a new sheet
* Here all matrix elements are grouped under the same RC value
**************************************************************
*
Sub create_matris()
n = 0
* Find all cells with this RC value
For m = RC To 1 Step -1
        k = 1
        Worksheets("RM").Activate
        Worksheets("RM").Range(Chr(65 + n) + Trim(Str(k))).Value = m
        counter = 1
        hit = 0
        * Search all matrix
        For j = 2 To 80 Step 1
                Worksheets("veri").Activate
                Range("a" & Trim(Str(j + 1))).Select
                For i = 1 To counter Step 1
                        ActiveCell.Offset(0, 1).Select  ' row,col
                        temp = ActiveCell.Value
                        If temp = m Then  ' RC hit
                                If i > 9 Then
                                        istr = Trim(Str(i))
                                Else
                                        istr = "0" & Trim(Str(i))
                                End If
                                If j > 9 Then
                                        jstr = Trim(Str(j))
                                Else
                                        jstr = "0" & Trim(Str(j))
                                End If
                                sstr = "C" + istr + "C" + jstr
                                k = k + 1
                                Worksheets("RM").Activate
                                Worksheets("RM").Range(Chr(65+n)+Trim(Str(k))).V
                        alue=sstr
                                Worksheets("veri").Activate
                                Range("a" & Trim(Str(j + 1))).Select
                                ActiveCell.Offset(0, i).Select
                                If hit = 0 Then
                                        hit = 1
                                End If
                        End If
                Next i
                counter = counter + 1
        Next j
        If hit Then
                n = n + 1
        End If
```

```
Next m
End Sub
******************************************************************
******
* This button click calls two functions above
******************************************************************
******
Private Sub Find_RC_Create_RM _Click()
   Call find_RC
   Call create_matris
End Sub
******************************************************************
*******
* This function creates two dimentional matrix(2DM) from m*n matrix
* 2DM will be used in grouping function
******************************************************************
*******
Private Sub Create_2DM_Click()
*
k = 0
For m = RC To 1 Step -1
        counter = 1
        For j = 2 To 80 Step 1
                Worksheets("veri").Activate
                Range("a" & Trim(Str(j + 1))).Select
                For i = 1 To counter Step 1
                        ActiveCell.Offset(0, 1).Select  ' row,col
                        temp = ActiveCell.Value
                        If temp = m Then
                                If i > 9 Then
                                        istr = Trim(Str(i))
                                Else
                                        istr = "0" & Trim(Str(i))
                                End If
                                If j > 9 Then
                                        jstr = Trim(Str(j))
                                Else
                                        jstr = "0" & Trim(Str(j))
                                End If
                                sstr1 = "C" + istr + "C" + jstr
                                sstr2 = "C" + jstr + "C" + istr
                                k = k + 1
                                Worksheets("2DM").Activate
                                Worksheets("2DM").Range("a" & Trim(Str(k))).Value
                        = sstr1
                                Worksheets("2DM").Range("b" & Trim(Str(k))).Value
                        = m
                                k = k + 1
```

```
                        Worksheets("2DM").Range("a" & Trim(Str(k))).Value
                = sstr2
                        Worksheets("2DM").Range("b" & Trim(Str(k))).Value
                = m
                        Worksheets("veri").Activate
                        Range("a" & Trim(Str(j + 1))).Select
                        ActiveCell.Offset(0, i).Select
                End If
            Next i
            counter = counter + 1
        Next j
Next m
End Sub
*
***********************************************************************
******
* This button processes the Relational Matrix(RM) and Two Dimention Matrix
(2DM)
* and create groups
***********************************************************************
******
*
Private Sub Process_Click()
*
Worksheets("RM").Activate
Range("a1").Select
temp_RC = ActiveCell.Value
grp_counter = 1
* Work on the relational matrix
For k = 0 To RC Step 1
        For i = 1 To 7000 Step 1
            Worksheets("RM").Activate
            ActiveCell.Offset(1, 0).Select  ' row,col
            Value = ActiveCell.Value
            If Value = "" Then
                Exit For
            End If
            val1 = Left(Value, 3)
            val2 = Right(Value, 3)
            v1 = 0
            v2 = 0
            For x = 1 To 999 Step 1
                Worksheets("grps").Activate
                grp_value = Worksheets("grps").Range("a" &
            Trim(Str(x))).Value
                If grp_value = "" Then
                        Exit For
                Else
                        If v1 = 0 Then
```

```
                    v1 = InStr(grp_value, val1)
                End If
                If v2 = 0 Then
                    v2 = InStr(grp_value, val2)
                End If
                If v1 > 0 And v2 > 0 Then
                    Exit For
                End If
            End If
        Next x
        If v1 > 0 And v2 > 0 Then
            * This machines are already in same group
        End If
        If v1 > 0 And v2 = 0 Then
            * v2 belongs to this group or not
            Call Entering_Machine(val2, val1)
        End If
        If v1 = 0 And v2 > 0 Then
            * v1 belongs to this group or not
            Call Entering_Machine(val1, val2)
        End If
        If v1 = 0 And v2 = 0 Then
            * New group for v1 and v2
            Worksheets("grps").Activate
            Worksheets("grps").Range("a"&
            Trim(Str(grp_counter))).Value = val1 + val2
            grp_counter = grp_counter + 1
        End If
    Next I
ActiveCell.Offset(-i, 0).Select  ' row,col
ActiveCell.Offset(0, 1).Select  ' row,col
temp_RC = ActiveCell.Value
Next k
End Sub
*******************************************************************
*******
* This sub session determines which group the new entering machine belong to
*******************************************************************
*******
*
Sub Entering_Machine(entering, other)
entering_val = 0
entering_count = 0
maxCR = 0
maxt = 0
For t = 1 To grp_counter Step 1
    Worksheets("grps").Activate
    temp_grp = Worksheets("grps").Range("a" & Trim(Str(t))).Value
    For i = 1 To Len(temp_grp) Step 1
```

```
            temp_str = Mid(temp_grp, i, 3)
            i = i + 2
            entering_str = entering + temp_str
            k = 1
            * Search from 2DM
            For k = 1 To 99999 Step 1
                Worksheets("2DM").Activate
                sstr = Worksheets("2DM").Range("a" & Trim(Str(k))).Value
                If sstr = "" Then
                    Exit For
                End If
                If sstr = entering_str Then
            entering_val=entering_val+Worksheets("2DM").Range("b"&Trim(Str
        (k))).Value
                    entering_count = entering_count + 1
                    Exit For
                End If
            Next k
        Next i
        If entering_count = 0 Then
            CR = 0
        Else
            CR = entering_val / entering_count
            entering_val = 0
            entering_count = 0
        End If
        If maxCR < CR Then
            maxCR = CR
            maxt = t
        End If
    Next t
    MTV = temp_RC * P
    MCR = maxCR
    i = 20
    If MCR >= MTV Then
        * dd entering to the group
        Worksheets("grps").Activate
        temp_grp = Worksheets("grps").Range("a" & Trim(Str(maxt))).Value
        temp_grp = temp_grp + entering
        Worksheets("grps").Range("a" & Trim(Str(maxt))).Value = temp_grp
    Else
        * Create new group for the entering machine
        Worksheets("grps").Activate
        Worksheets("grps").Range("a" & Trim(Str(grp_counter))).Value = entering
        grp_counter = grp_counter + 1
    End If
    End Sub
```

B.3. The Levenshtein Procedure

A commonly-used bottom-up dynamic programming algorithm for computing the Levenshtein distance involves the use of an $(n + 1) \times (m + 1)$ matrix, where n and m are the lengths of the two strings.

1) Set n to be the length of s and set m to be the length of t.
    If n = 0, return m and exit.
    If m = 0, return n and exit.
    Construct a matrix containing m rows and n columns.
2) Initialize the first row to 0..n.
    Initialize the first column to 0..m.
3) Examine each character of s (i from 1 to n).
4) Examine each character of t (j from 1 to m).
5) If s[i] equals t[j], the cost is 0.
    If s[i] doesn't equal t[j], the cost is 1.
6) Set cell d[i,j] of the matrix equal to the minimum of:
    The cell immediately above plus 1: d[i-1,j] + 1.
    The cell immediately to the left plus 1: d[i,j-1] + 1.
    The cell diagonally above and to the left plus the cost: d[i-1,j-1] + cost.

The invariant maintained throughout the algorithm is that we can transform the initial segment s[1..i] into t[1..j] using a minimum of d[i,j] operations. At the end, the bottom-right element of the array contains the answer.

B.4. Example of Levenshtein Distance

This section shows how the Levenshtein distance is computed between the route of product 3 and route of product 21.

Steps 1 and 2

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1							
M 07	2							
M 15	3							
M 13	4							
M 63	5							

Steps 3 to 6 When i = 1

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1	1						
M 07	2	2						
M 15	3	3						
M 13	4	4						
M 63	5	5						

Steps 3 to 6 When i = 2

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1	1	2					
M 07	2	2	1					
M 15	3	3	2					
M 13	4	4	3					
M 63	5	5	4					

Steps 3 to 6 When i = 3

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1	1	2	3				
M 07	2	2	1	2				
M 15	3	3	2	1				
M 13	4	4	3	2				
M 63	5	5	4	3				

Steps 3 to 6 When i = 4 and 5

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1	1	2	3	4	5		
M 07	2	2	1	2	3	4		
M 15	3	3	2	1	2	3		
M 13	4	4	3	2	1	2		
M 63	5	5	4	3	2	3		

Steps 3 to 6 When i = 6 and 7

	i	M 36	M 07	M 15	M 13	M 56	M 63	M 64
j	0	1	2	3	4	5	6	7
M 02	1	1	2	3	4	5	6	7
M 07	2	2	1	2	3	4	5	6
M 15	3	3	2	1	2	3	4	5
M 13	4	4	3	2	1	2	3	4
M 63	5	5	4	3	2	3	2	3

Step 7

The distance is in the lower right hand corner of the matrix, i.e. 3. This corresponds to our intuitive realization that The route of product 3 can be transformed into the route of product 21 by substituting "M02" for "M36" and adding "M56" and “M64” (one substitution and 2 insertion = 3 changes).

B.5. The Visual Basic code for Levensthein Distances

```
*****************************************************************
* Levenshtein Algorithm
*
* Basar Uyanık
* 05 / 2005
*****************************************************************
Private Sub Process_Click()
Worksheets("matrix").Activate
Range("b2").Select
For b = 2 To 99
        mach1 = Worksheets("Route").Range("A" & Trim(Str(b))).Value
        str1 = Worksheets("Route").Range("B" & Trim(Str(b))).Value
        If str1 = "" Then
                Exit For
        End If
        For bb = 2 To 99
                mach2 = Worksheets("Route").Range("A" & Trim(Str(bb))).Value
                str2 = Worksheets("Route").Range("B" & Trim(Str(bb))).Value
                If str2 = "" Then
                        Exit For
                End If
                ActiveCell.Value = LD(str1, str2)
                ActiveCell.Offset(1, 0).Select
        Next bb
ActiveCell.Offset(-(bb - 2), 0).Select
ActiveCell.Offset(0, 1).Select
Next b
End Sub
```