IMPACT OF BEST MANUFACTURING ENGINEERING PRACTICES ON SOFTWARE ENGINEERING PRACTICES

By

Shahina Akhter

A Thesis submitted to the Faculty of

The College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Computer Engineering

Florida Atlantic University

Boca Raton, Florida

August 1997

UMI Number: 1385324

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Impact of Best Manufacturing Engineering Practices on Software Engineering Practices

by

Shahina Akhter

This thesis was prepared under the direction of the candidate's advisor Dr. Neal Coulter, Department of Computer Science and Engineering and has been approved by the members of her supervisory committee. It was submitted to the faculty of the College of Engineering and was accepted in partial fulfillment of the requirements for the degree of Master of Science in Computer Engineering.

SUPERVISORY COMMITTEE:

Thesis Advisor, Dr. Neal S. Coulter

Dr. Thomas B. Horton

Dr. Mohammad Ilyas

Chairperson, Department of Computer Science and Engineering

Dean, College of Engineering

Dean of Graduate Studies and Research

7-14-97

Date

Acknowledgments

The work required to finish this thesis could not have been carried out without the support of many people. Dr. Neal S. Coulter provided the guidance and encouragement for this research. I would like to thank him very much for his magnificent support and tremendous patience. I want to thank Dr. Thomas B. Horton and Dr. Mohammad Ilyas for being on my thesis committee and for giving me their vigorous support. I am also thankful to Dr. Imad Mahgoub for his superb encouragement Mr. Ira Monarch and Dr. Suresh Konda of Carnegie Mellon University greatly assisted me in this thesis.

I also wish to recognize the contribution of my family members for their support, forbearance, and understanding during my studies. Especially, I am exceedingly grateful to my beloved father Mohammad Jamshed Ali Sheik and my precious mother Mrs. Mamtaz Begum. I want to acknowledge them for their continued devotion to my success and for continuously believing in me. I cannot imagine having a better parent. I thank Aminul Islam, my treasured husband, for sharing his life with me and for his endless love and support. My gratitude goes also to my lovely younger sister Farhana Akhter, and my adorable younger brothers Sheik Ariful Islam, and Sheik Tariqul Islam. Heartfelt thanks go also to Jahanara Begum, Moniruzzaman, Zual Islam, Nasreen Momen, Fookloy Ford, Nurul Momen, Dr. Thomas Dorworth, and Miriam Carpenter.

ABSTRACT

Author: Shahina Akhter

Title: Impact of Best Manufacturing Engineering Practices to Software Engineering Practices

Institution: Florida Atlantic University

Thesis Advisor: Dr. Neal S. Coulter

Degree: Master of Science in Computer Engineering

Year: 1997

This thesis involves original research in the area of semantic analysis of textual databases (content analysis). The main intention of this study is to examine how software engineering practices can benefit from the best manufacturing practices. There is a deliberate focus and emphasis on competitive effectiveness worldwide. The ultimate goal of the U.S. NAVY's Best Manufacturing Practices Program is to strengthen the U.S. industrial base and reduce the cost of defense systems by solving manufacturing problems and improving quality and reliability. Best manufacturing practices can assist software engineering practices in a way that when software companies use these practices they can:

- Improve both software quality and staff productivity
- Determine the current status of the organization's software process
- Set goals for process improvement
- Create effective plans for reaching those goals
- Implement the major elements of the plans

Table of Contents

List of Figures

List of Tables

TO MY BELOVED PARENTS

Impact of Best Manufacturing Engineering Practices on Software Engineering Practices

1. Introduction

Software Engineering has expeditiously appeared as a discipline in its own right. The chronicle of software development is one of heightening scale. According to I. Sommerville,

> "There have been a number of proposed definitions of software engineering. Their common factors are that software engineering is concerned with building software systems which are larger than would normally tackled by a single individual, uses engineering principles in the development of these systems and is made up of both technical and non-technical aspects. As well as a thorough knowledge of computing techniques, the software engineer, like any other engineer, must be able to communicate, both orally and in writing" [SOMM85].

Manufacturing engineering (*also called industrial and process engineering*) made enormous contributions to improving productivity in the early decades of the industrial revolution. "When operations were disorganized, unstructured, and subject to changing whims of individuals, developing and applying "standard" methods and times repaid the costs incurred many times over" [PLOS91].

Presently, software systems are typically not developed in a systematic way. Software engineering's goal should be to develop and to introduce software technologies which will enable companies to develop software systematically. Manufacturing engineering is concerned with the economical use of technology and the management of resources for producing and delivering quality goods and services. After examining the best manufacturing practices it is encouraging to perceive that software engineering practices can benefit enormously from it. Throughout this entire thesis, there is an increased emphasis on software engineering practices. Most of the analysis attempts to delineate the co-relation between software engineering practices and best manufacturing engineering practices. This thesis also has a great deal of information about lexical maps, and how to interpret them in a usable manner.

1.1 Ingredients For Best Practices

Good software is good business. Some of many factors which are important ingredients, or best practices, for software development include [Will97]:

- Definition, adoption, and maintenance of a rigorous software development process from product concept to end of life (phase containment).
- Definition, collection, analysis, and action associated with process, code, and test metrics [GRAD92].
- Successful communication between cross-functional team members (i.e., employees working well with other individuals from different disciplines).
- Timely response to problems (e.g., discovery, follow-up, and closure).
- Quality engineering (testing and analysis using metrics, standards, and automated tools)
- Fault recovery (trap handlers and test assertions), defensive design, and requirements, design, and source code modularity [WILL95].
- Code maintainability using standards and guidelines.
- Usability engineering (task analysis and usability testing) [PRES92]
- Reliability engineering (testing and analysis based on use-modules and operational profiles).
- Functional testing, definition, design, and development (e.g., mapping requirements to designs and source code).
- Structural testing (e.g., static and dynamic source code analysis).

1.2 Software Engineering Practices

Software Engineering has many facets. Software engineering certainly is not the same as programming, although programming is an important ingredient of software engineering. The practice of software engineering however also has to deal with such matters as the management of huge development projects, human factors (regarding both the development team and the perspective users of the system) and cost estimation and control.

Competitive software development and delivery methods, tools, and processes are critical ingredients for releasing successful software products. The requirement for software engineering best practices applies to both application and system software companies [WILL97].

Just understanding the development tools (e.g. C, C++, **dbx, lint**) is not enough to ensure project and product success. Understanding software engineering principles such as task analysis, assertions, configuration management, paper prototypes, design and code reviews, and other best practices is equally important to tool knowledge [WILL97].

2 Best Manufacturing Practices

2.1 Introduction[1]

Best Manufacturing Practices means practices that are geared to produce a functional, mostly reliable product for the purpose intended. Manufacturing covers both the economical use of technology and the management of resources. "Best Manufacturing Practices" help to convert raw materials into usable products at reasonable cost.

> "Benchmarking is the continuous process of measuring, analyzing, and comparing one organization's products, services, and practices against those of its competitors. This concept is embodied in the Office of Naval Research's Best Manufacturing Practices (BMP) Program, which was established in 1985. The intention of the BMP program to identify the best practices in the areas of design, test, production, facilities, logistics, and management, and to encourage industry and government to share information about these practices. To accomplish this goal, independent teams of experts have been established to survey organizations that are ready to share information about their own best processes. Participation in the survey is voluntary and at no cost to the requesting organization. Once the BMP team completes its work, the surveyed activity or organization reviews a draft of the report. Copies of the final report are distributed to representative of government, industry, and academia throughout the country. The information in the reports is designed to help organizations evaluate their own processes by identifying, analyzing, and emulating the processes of organizations that excel in those areas. Since the inception of the BMP Program, more than 75 organizations including McDonnell Douglas, General Dynamics, Hughes Aircraft, Department of Energy, and NASA- have participated. Navy participants include Naval Surface Warfare Center Division, Crane, IN; Charleston Naval Shipyard; Naval Air Warfare Center, Aircraft Division, Indianapolis; Philadelphia Naval Shipyard; and Naval Undersea Warfare Center Division, Keyport, WA. The results of these surveys have been disseminated throughout government and industry, leading to many instances of exchanged information and cooperative effort. The BMP Program recently moved to its new Center of Excellence (COE) location in College Park, MD. The BMPCOE is a joint effort of the BMP Program, the National Institute of Standards and Technology, and the University of Maryland at College Park Engineering Research Center. The BMPCOE's mission is to identify and promote exemplary manufacturing practices and disseminate this information to U.S. manufacturers of all sizes. The BMP Program is a sharing program that benefits everyone who uses it. For more

1. This information has been taken from the intenet site: http://www.bmpcoe.org

information, or for a list of organization surveyed. Contact the BMP Center of Excellence, 4321 Hartwick Road, Suite 308, College Park, MD, 20740; phone (800)-789-4BMP or (301) 403-8100" [INTE1].

2.2 Form and Content of BMP Data

BMP's are connected of Best Practices (BP) and Information (INFO) sections for each organization. BP's are above and beyond what the relevant industry does; BP's must have a track record demonstrated by an improvement metric contrasting the BP and the previous practice. INFO's are practices on par with similar firms or just implemented with no track record. BP's are the main basis for this study.

2.3 Using Inquery to Sort BMP Data

Inquery is a probabilistic information retrieval (IR) system. Given a database of textual documents Inquery can retrieve documents relevant to a user's query. The documents do not need to be pre-classified by hand, though they do need to be indexed by Inquery's parsing system. The queries can be formulated either in natural language (i.e. a regular English sentence) or in a more exact structured query language. In this study, I used Inquery to sort BMP data.[2]

2.4 Organization and Distribution of BMP Data

Each BP statement is treated as a document. These are usually a few sentences to a few paragraphs each. The BMP database has 1,230 BP documents (and 680 INFO documents, for a total of 1,910 documents). In this study 238 documents among those 1230 BP documents had the word ***software*** in them. Inquery was used to sort out these documents.

2. INQUERY has been discussed in detail in Appendix B

2.5 Determination of Best Practices

A typical network consists of descriptors and links. Network structures suggest common interpretations of subsets of documents. An information retrieval system is used to search these documents and evaluate interpretations. This process identifies common best manufacturing practice themes.

2.6 Limitations of Analysis the BMP Data

First, due to voluntary participation in the BMP, several limitations exist concerning the data itself [STEV94]:

- Companies that believe they have implemented distinctive practices may be more likely to have participated in the study,
- Some sophisticated practices may not be disclosed because of propriety reasons,
- Participants have final editorial control over the final report, and
- Defense contractors are likely to have somewhat different product and organizational requirements than commercial firms.

3 Co-Occurrence Analysis

3.1 Introduction

Co-Occurrence analysis is a technique for extracting the strength and breadth of term associates in large corpora. It is used on large bibliographic databases since 1986. Co-Occurrence analysis incorporates graphic representations of networks of term associations [CALL86].

Lexical analysis and mapping methods are constituents of a data analysis methodology which may be addressed to both qualitative and quantitative data collected and corroborated by various teams or organizations. A lexical map is a graphical presentation of a network composed of nodes holding words and phrases (terms) conjoined by links.

3.2 The Data

Co-word methodology operates on textual data. In this study, software created at Carnegie Mellon University's Software Engineering Institute (SEI) is used to generate index terms. SEI's software finds only noun phrases.

3.3 The Metric and the Algorithm

Co-word analysis enables the structuring of data at various levels of analysis [CALL86]:

(1) as networks of links and nodes;

(2) as distributions of networks called super networks;

Co-word analysis reduces a large space of related terms to multiple related smaller spaces that are easier to comprehend but are also indicative of actual partitions of interrelated concepts in the literature being analyzed. This analysis requires an association measure and algorithm for searching through the space.

The analysis is designed to identify areas of strong focus that interrelate. This scheme allows us to construct a mosaic of software engineering topics.

3.3.1 The Metric

Metrics for co-word analysis have been studied extensively [CALL86], [CALL91], [COUR89], [LAW92], [WHIT89]. The basic metric most suitable for this study is strength S (called Equivalence Index by Callon). It is described as follows:

Two descriptors, i and j, co-occur if they are used together in the classification of a single document. Take a corpus consisting of N documents. Each document is indexed by a set of unique descriptors that can occur in multiple documents. Let C_k be the number of occurrences of descriptor k; i.e., the number of times k is used for indexing documents in the corpus. Let c_{ij} be the number of co-occurrences of descriptors i and i (the number of documents indexed by both descriptors). Then Strength S of association between descriptors i and i is given by the expression:

$$S(C_i, C_j, C_{ij}) = C^2_{ij} / C_iC_j \text{ , } 0 <= S <= 1$$

Two descriptors that appear many times in isolation but only a few times together will yield a lower S value than two descriptors that appear relatively less often alone but have a higher ratio of co-occurrences.

3.3.2. The Algorithm

The algorithm makes two passes through the data to produce pair- wise connections of descriptors in networks. A network consists of nodes (descriptors) connected by links. Each node must be linked to at least one other node in a network. The first pass (Pass-1) generates the primary associations among descriptors; these descriptors are called internal nodes and the corresponding links are called internal links. A second pass (Pass-2) generates links between Pass-1 nodes across networks, thereby forming associations among completed networks. Pass-2 nodes and links are called external ones. Pass-1 builds networks that can identify areas of strong focus; Pass-2 can identify descriptors that associate in more than one network and thereby indicate pervasive issues. This pattern of networks yields a mosaic of the data being analyzed [CALL86].

3.3.2.1 Pass-1

During Pass-1, the link that has the highest strength is selected first. These linked nodes become the starting points for the first network. Other links and their corresponding nodes are then determined breadth-first.

This Pass-1 strategy does not necessarily (or usually) yield S strengths in strict descending order, either within individual networks or among sequentially generated networks with respect to the sum or average of S strengths. The first network becomes the first network only because it starts with the highest link; The second network then starts with the highest link among remaining links, and so forth. This order of generation is not especially significant because it is possible that the links included in a network after the initial link do not have co-occurrence strengths in the same high range as this initial link.

Figure 1 shows Pass-1 links and nodes (the dark lined links are pass-1 links, and the dark lined nodes are the Pass-1 nodes) [CALL86].

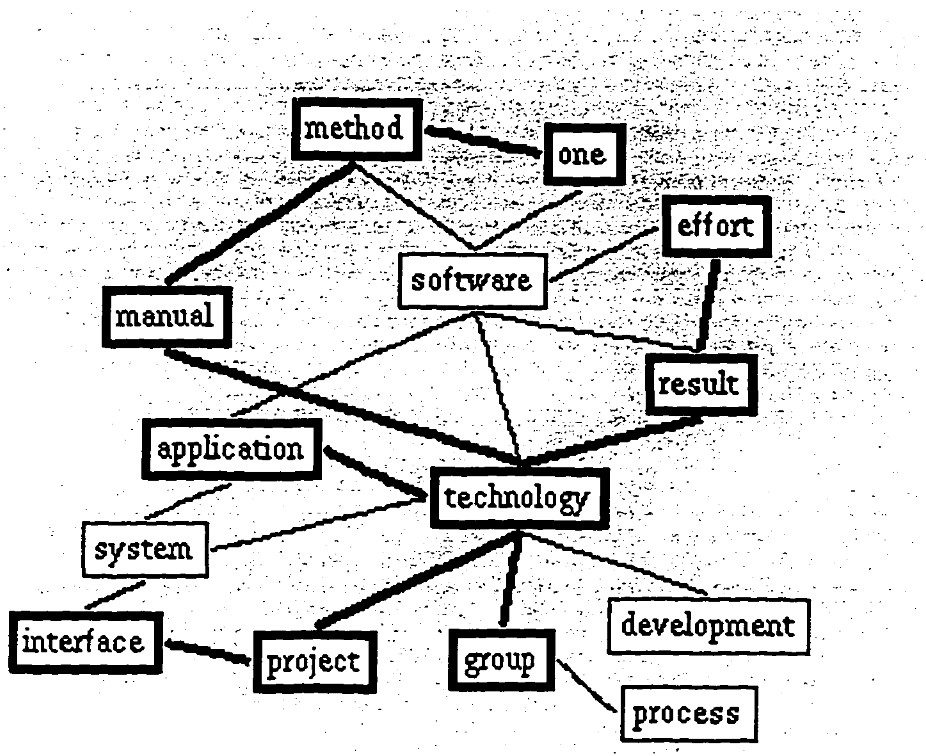

Figure 1: Sample Network Number 1 from BMP data

3.3.2.2 Pass-2

The second pass (Pass-2) is designed to seek further associations among descriptors found in Pass-1 to determine how networks fit together in larger super networks. To be a candidate for inclusion in Pass-2, both nodes (descriptors) of a link must be in some network in Pass-1. Pass-2 nodes and Pass-2 links are represented by thin boxes and by thin lines connecting them with Pass-1 nodes, respectively. As in Pass-1, candidate links are included in Pass-2 based on their strengths and co-occurrence counts. The order of Pass-2 links by descending S values for qualifying links. A node can appear in only one Pass-1 network, but can appear in more than one Pass-2 link and network [CALL86]. A hashed line is displayed in the event when two Pass-1 nodes from the same Pass-1 networks are linked during Pass-2.

Figure 2 illustrates Pass-2 Links and nodes (the thin lined links are pass-2 links, and the thick lined nodes are the Pass-1 nodes).

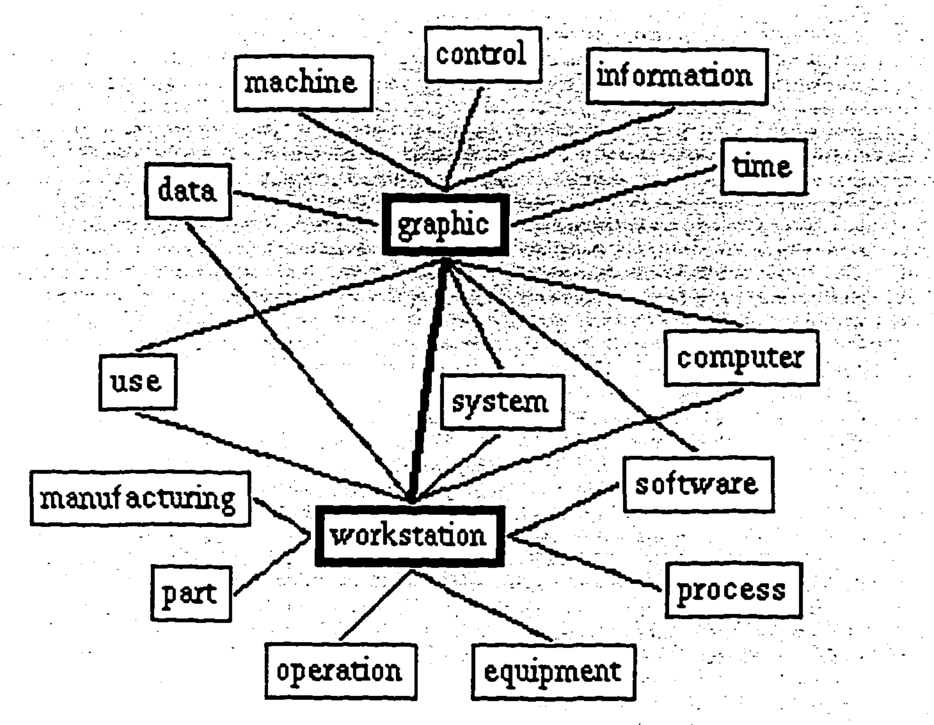

Figure 2: Sample Network Number 2 From BMP Data

3.3.2.3 Algorithm Constraints

Without some minimum constraints, descriptors appearing infrequently but almost always together could dominate networks [CALL86]; hence a minimum co-occurrence C_{ij} value is required to generate a link. At the same time, some maps can become cluttered due to an excessive number of legitimate links (but of generally decreasing S values); hence, restrictions on numbers of nodes and links are sometimes required to facilitate the discovery of major partitions of concepts.

In this study, I analyzed the four sets of data:

i) Entire BMP data

ii) Only the documents with the word "Software" in it

iii) Excluding only the word "Software"

iv) Excluding the word "Software" and also all the "Software" related phrases entirely

Software related phrases were phrases like ***software development, software design, software testing,*** etc., not only the word ***software*** alone.

Table 1: Parameters and Resulting Networks

Different Networks	***Minimum Co-Occurrence***	***Maximum Links***	***Networks Generated***
Entire BMP data	50	10	12
Only the documents with the word "Software" in it	15	10	12
Excluding Only the word "Software"	15	10	11

Table 1: Parameters and Resulting Networks

Different Networks	*Minimum Co-Occurrence*	*Maximum Links*	*Networks Generated*
Excluding the word "Software" and all the "Software" related phrases entirely	15	10	11

For the networks generated by the entire BMP data the co-occurrence cutoff was set at fifty to accommodate the lesser volume of data. For all networks the number of links and nodes in each network, both Pass-1 and Pass-2, was set at 10 links and maximum 100 maps.

Table 1 depicts that the entire BMP data generated 12 networks when the minimum co-occurrence was set to 50, and maximum links were set to 10. Similarly, the rest of the three sets of data can be observed in **Table 1.**

These four sets of data have plenty of differences. Some of the differences are depicted in the following **Table 2.**

Table 2: Comparisons of the Number of Generated Networks According to Different Co-occurrence Levels

Different Co-occurrence Level (Keeping the number of links l=10 as a fixed value here)	*Entire BMP Data*	*Only the documents with the word "Software" in it*	*Excluding Only the word "Software"*	*Excluding the word "Software" and all the "Software" related phrases entirely*
10	97	20	19	20
15	62	12	11	11
25	26	5	4	4
40	16	3	2	2
50	12	3	2	2

Table 2 illustrates that various number of networks are generated by our four sets of data depending upon the various co-occurrence level. When the co-occurrence is 10 then the entire BMP data generates 97 networks. Only the documents with the word "Software" in it generates 20 networks. Excluding only the word "Software" generates 19 networks. Excluding the word "Software" and also all the "Software" related phrase entirely generates 20 networks.

Similarly, when the co-occurrence is 15 then the entire BMP data generates 62 networks. Only the documents with the word "Software" in it generates 12 networks. Excluding only the word "Software" generates 11 networks. Excluding the word "Software" and also all the "Software" related phrase entirely generates 11 networks.

Throughout the entire study, I used the 12 networks generated by the entire BMP data, 12 networks generated by only the documents with the word "Software" in it, 11 networks generated excluding only the word "Software", and 11 networks generated by excluding the word "Software" and also all the "Software" related phrases entirely (Please refer to appendix for all of the networks).

3.3.2.4 Algorithm Summary

A summary of the algorithm is [COUL96]:

1 Select a minimum for the number of co-occurrences, C_{ij} for descriptors i and j.

2 Select maxima for the number of Pass-1 links and nodes;

3 Select maxima for the total (Pass-1 and Pass-2) links and nodes;

4 Start Pass-1;

5 Generate the highest S value from all possible descriptors to begin a Pass-1 network;

6 From that link, from other links in a breadth-first manner until no more links are possible due to the co-occurrence minima or to Pass-1 link or node maxima. Remove all incorporated descriptors from the list of subsequent available Pass-1 descriptors;

7 Repeat Steps 5 and 6 until all Pass-1 networks are formed; i.e., until no two remaining descriptors co-occur frequently enough to begin a network;

8 Begin Pass-2;

9 Restore all Pass-1 descriptors to the list of available descriptors;

10 Starting with the first Pass-1 network, generate all links to Pass-1 nodes in that network with any Pass-1 nodes having at least the minimal co-occurrences in descending order of S value; stop when no remaining descriptors meet co-occurrence minima or when total node or link maxima are met. Do not remove any descriptors from the available list;

11 Repeat Step 10 for each succeeding Pass-1 network.

A maximum number of Pass-1 networks can be specified in cases where an excessive number of networks will be generated otherwise; this restriction was not necessary here.

Numerous variations of this algorithm is possible.

3.4 Comments on Selection of Network Parameters

Link and node limitations mostly determine how networks will be generated in concert with the corresponding co-occurrence minimum. If the co-occurrence minimum is too high, few links may be formed; If it is too low, an excessive number of links may result. In the former case, subspecialities in a field may not emerge; in the latter case, a field may look disproportionately cluttered [CALL86].

So, while the link, node, and co-occurrence parameters effectively control the generation of networks, small changes in their values appear to affect only marginal links, at least in this study. Of course, additional and subsequent data can affect the generation of core themes without changes in parameters, which is the intent of co-word analysis.

4 Network Analysis

Network analysis reveals an abundance of information. Representing the interactions of nodes and links within larger contexts enables the identification of research and commentary trends in software engineering publications.

Maps of all networks appear in the Appendix C. A wealth of information emerges from these maps, but much of it is difficult to distill. A combination of qualitative and quantitative approaches are used in analyzing teh networks.

4.1 Network Names

4.1.1 Methodology

In this study, the naming of the maps have been an attempt to summarize their main thrusts. This is not a precise activity. In some cases the maps had highly connected descriptors; in others, two (and sometimes three) descriptors exhibited high connectivity. In the latter cases, hyphenated names have been used. Descriptors contained in nodes have always been used. Generally, descriptor(s) from the node(s) with the most connections have been used here, giving greater weight to Pass-1 nodes in close calls.

4.1.2 Findings

The names chosen are as follows:

4.1.2.1. Networks on Entire BMP Data

Table 3: Network Names for the Entire BMP Data

Number of the Network	*Network Names*
1	Manufacturing-Process
2	Quality Assurance
3	Process Improvement
4	Cycle Time
5	Test Equipment
6	Personnel-Training
7	Data base
8	System-Analysis
9	Total-Performance
10	Inspection
11	Cost-Reduction
12	Manufacturing-Standard

The naming of the networks in **Table 3** are taken from the **Figures A. 1-1** to **Figure A. 1-12** in the **Appendix C**. In **Figure A. 1-1** there was high strength between descriptors "Manufacturing" and "Process". Therefore, I preferred to name this network as "Manufacturing-Process." Some of the networks have few connections that are insignificant. **Figure A. 1-29 of Appendix C** has the connections among the descriptors "Shop Floor", "Shop", and "Floor" which are very redundant, although their strength may be numerically high. **In future studies,** these kind of connections can be taken out for better understanding.

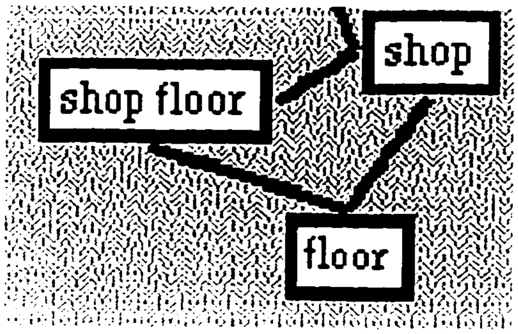

Figure 3: Example of Redundant Connections

5 Types of Networks and Their Interactions

5.1 Methodology

There are essentially three types of networks: principal, secondary, and isolated. Principal networks are connected to one or more (secondary) networks. Secondary networks generally are linked to principal networks through a number of external links. Isolated networks have an absence (or low intensity) of links with other networks. Isolated networks often have links with high S values, usually accompanied by low co-occurrence C_{ij} values. While isolated networks are easy to recognize, principal and secondary networks may not be. Therefore, we will define and operationalize terms that characterize these functionalities.

Density is defined as the mean of the Pass-1 S values of a network; centrality is defined as the square root of the sum of the squares of the Pass-2 S values of a network in order to distinguish among relatively close values. Density represents the internal strength of a network, while centrality represents a network's position in strength of interaction with other networks.[3]

3. These terms are accepted ones in co-word analysis literature. We recognize that density and centrality have others domain-specific connotations--say, in statistics. Alternative choices include ***coupling*** and ***cohesion***, but these already have meanings in software engineering literature. ***Adhesion*** and ***density*** could be used, but that is not conventional in co-word literature, and it confounds the use of the term ***density*** even more. We trust the intent of the terminology is clear.

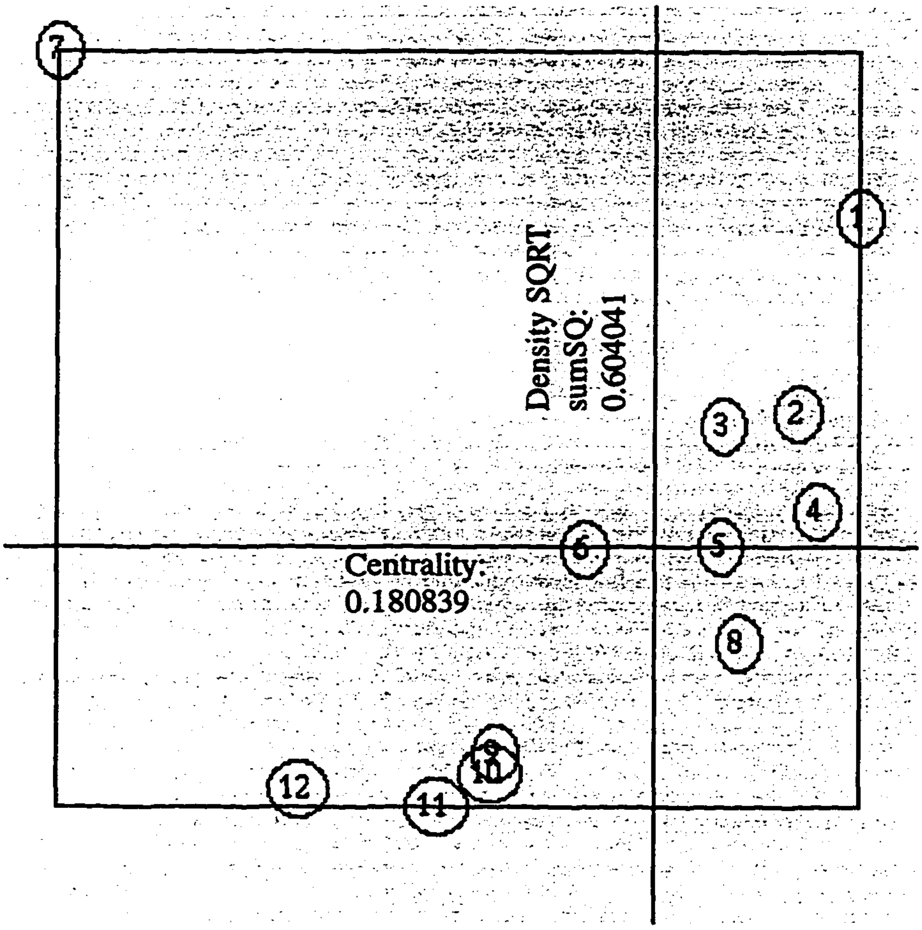

Figure 4: Density and Centrality Map for the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

From **Figure 4**, it can be observed that network number 7 is an isolated network. The graph's origin is the median of the Axes Values. We can observe the **Figure A. 1-7** from the **Appendix C**, and we can see the nodes and links in that particular network does not relate to the other eleven networks. Furthermore, from **Figure 4**, it can be observed that networks numbered 1, 2, 3, and 4 are high in density and high in centrality. Network number 8 is high in density but low in centrality. Network number 9, 10, 11, and 12 are low in density and low in centrality. Network number 5 and 6 are approximately neutral in density and centrality. We can look at the corresponding maps from **Figure A. 1-1 to Figure A. 1-13 of Appendix C** for better understanding.

6 Super Network Analysis

6.1 Methodology

It is possible to be more specific in describing how networks interact with other specific networks; this addresses centrality in a more focused fashion, but does not substitute for the general centrality measure. Centrality is a composite measure of a network's intersection with all other networks generated from the same descriptor data; super networks describe pair-wise links of networks from those data.

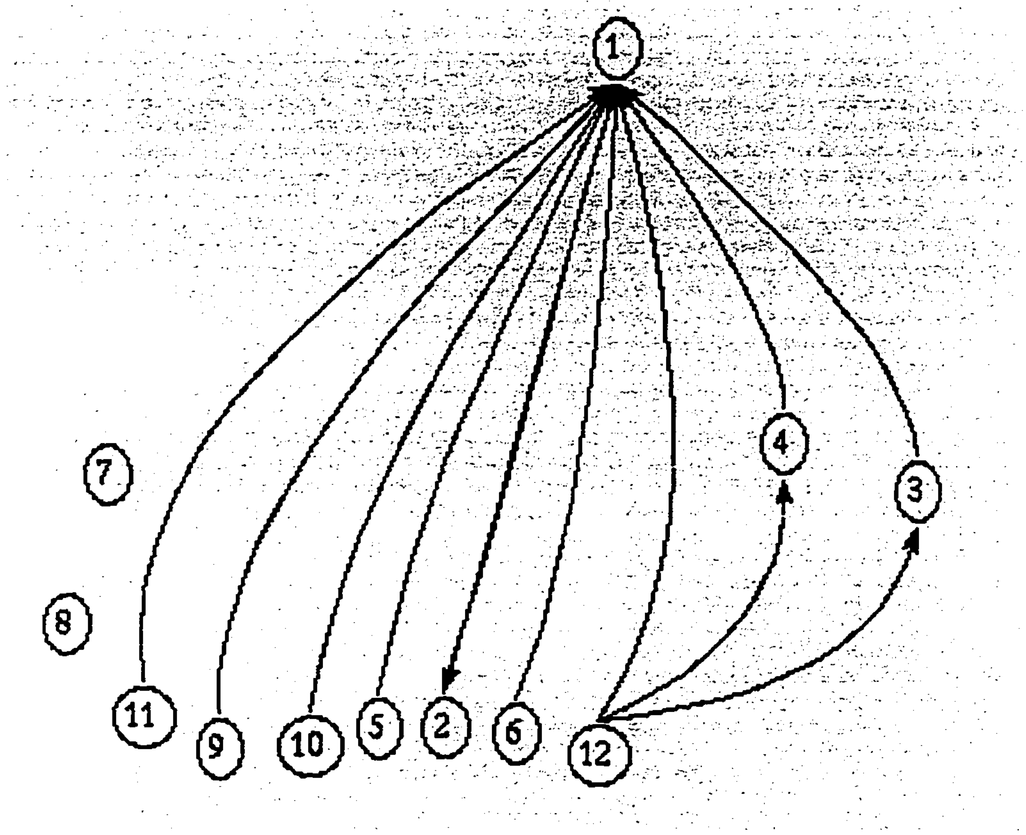

Figure 5: Sample Super Network for only the documents With the Word "Soft ware" when the Cutoff for Co-Occurrence Level is 15 and Links are10

From **Figure 5**, it can be observed that network 7 and network 8 are isolated networks.

More formally, principal and secondary networks have been chosen to be defined as follows: If Network-A has internal nodes that are Pass-2 nodes in x links of Network-B and each of these links has a Pass-2 S value that exceeds the minimum Pass-1 S value of Network-B, then Network-A is a secondary network of Network-B. Super network (directed) links to capture this situation are of the form

x: Network-A ------> Network-B.

Notice that Network-A and Network-B can share more than x nodes.

Using this way of determining principal and secondary networks, we can describe super networks of networks. The relationships in these super networks are not inherently bi-directional, as are network links (at least as defined using S).

It should further be noted that, super networks show a lot about centrality, but they do not display much about density.

7 Sample Network Analysis of BMP Data

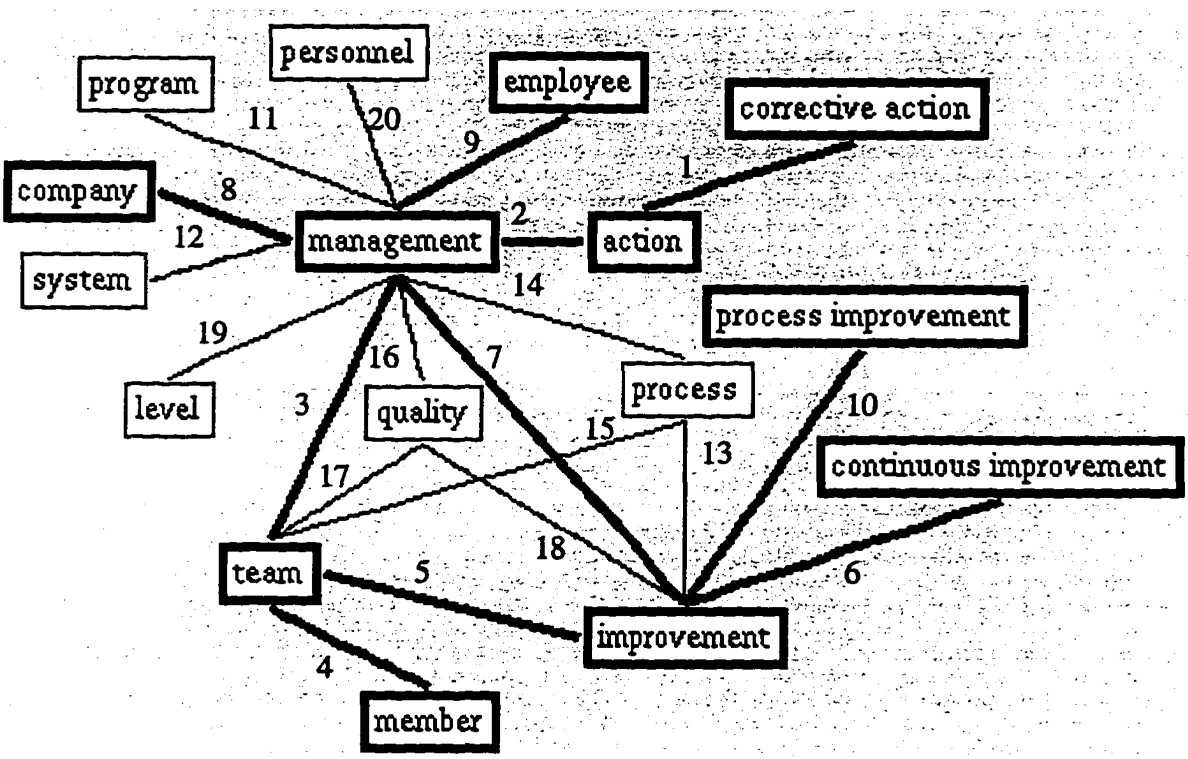

Figure 6: Network Number 3 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

Figure 6 is a depiction of different Pass-1 and Pass-2 descriptors and links. The numbering of the links are given according to the seriality of the breadth-first creation of the links. The link between descriptors "corrective action" and "action" is link number 1 because it has the highest strength. The second link (link between descriptors "management" and "action") then has the highest strength among remaining links, and so forth. As I mentioned earlier, in this study I have four sets of data. Therefore, I have four results file (Please refer to **Appendix A** to see how ***results file*** are generated).

Figure 7: Sample Results File Number 1 for Network Number 3 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

```
Run Parameters: Eliminate by Nodes
Pass Two Node Filter: Both nodes. Link Selection: Strength and Max. Nodes
Min. Strength: 0.000000. Min. Co-Occurrence: 50. Max links: 10.
Max maps 100. Max nodes 10. Maps Produced: 12

3 16 20
^management^ 412 5 11 1 3
^improvement^ 278 1 6 1 3
^team^ 270 2 5 1 3
^process^ 699 12 3 2 1
^quality^ 418 6 3 2 1
^action^ 179 1 2 1 3
^program^ 558 9 1 2 1
^system^ 817 12 1 2 1
^level^ 270 2 1 2 9
^personnel^ 308 2 1 2 6
^corrective action^ 86 1 1 1 3
^member^ 97 1 1 1 3
^continuous improvement^ 90 1 1 1 3
^company^ 251 2 1 1 3
^employee^ 192 2 1 1 3
^process improvement^ 50 1 1 1 3
^action^ ^corrective action^ 86 0.480447 1 0
^continuous improvement^ ^improvement^ 90 0.323741 1 0
^management^ ^team^ 160 0.230133 1 0
^member^ ^team^ 74 0.209087 1 0
^improvement^ ^team^ 122 0.198295 1 0
^improvement^ ^management^ 148 0.191241 1 0
^company^ ^management^ 138 0.184157 1 0
^employee^ ^management^ 120 0.182039 1 0
^improvement^ ^process improvement^ 50 0.179856 1 0
^action^ ^management^ 111 0.167069 1 0
^management^ ^program^ 252 0.276229 2 1
^management^ ^system^ 291 0.251575 2 1
^improvement^ ^process^ 206 0.218380 2 1
^management^ ^process^ 248 0.213564 2 1
^process^ ^team^ 189 0.189270 2 1
^management^ ^quality^ 179 0.186051 2 1
^quality^ ^team^ 142 0.178664 2 1
^improvement^ ^quality^ 143 0.175975 2 1
^level^ ^management^ 138 0.171197 2 9
^management^ ^personnel^ 147 0.170289 2 6
0.234606 2.031195 0.424525 1007 596
```

Figure 7 has been taken from the results file number 1 for network number 3 generated by the entire BMP data when the cutoff for co-occurrence level is 50 and links are 10. This figure depicts the various descriptors and links created for the network number 3 generated by the Entire BMP data when the cut-off for co-occurrence level is 50 and links are 10. Figure 10 also shows that Pass-2 links are in order by decreasing strengths, but that is not necessarily true for Pass-1 links.

Table 4: Analysis of the Links of Network Number 3 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

Number of the Links from Figure 3	***Co-Occurrence***	***Strength***	***Number of Passes***
1	86	.480447	Pass-1
2	111	.167069	Pass-1
3	160	.230133	Pass-1
4	74	.209087	Pass-1
5	122	.198295	Pass-1
6	90	.323741	Pass-1
7	148	.191241	Pass-1
8	138	.184157	Pass-1
9	120	.182039	Pass-1
10	50	.179856	Pass-1
11	252	.276229	Pass-2
12	291	.251575	Pass-2
13	206	.218380	Pass-2
14	248	.213564	Pass-2
15	189	.189270	Pass-2
16	179	.186051	Pass-2
17	142	.178664	Pass-2
18	143	.175975	Pass-2
19	138	.171197	Pass-2
20	147	.170289	Pass-2

Table 4 illustrates that link number 1 (Link between "action" and "corrective action") which is a Pass-1 link has a co-occurrence of 86 and strength of 0.480447. Similarly, link number 2 ("management" and "action") which is also a Pass-1 link has a co-occurrence of 111 and strength of 0.167069 and so on.

Table 5: Analysis of the Nodes of Network Number 3 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

Name of the Nodes from Figure 3	*Number of Documents it Appears*	*Number of Maps it Appears*	*Number of Passes*
Management	412	1	Pass-1
Improvement	278	1	Pass-1
Team	270	2	Pass-1
Process	699	12	Pass-2
Quality	418	6	Pass-2
Action	179	1	Pass-1
Program	558	9	Pass-2
System	817	12	Pass-2
Level	270	2	Pass-2
Personnel	308	2	Pass-2
Corrective Action	86	1	Pass-1
Member	97	1	Pass-1
Continuous Improvement	90	1	Pass-1
Company	251	1	Pass-1
Employee	192	2	Pass-1
Process Improvement	50	1	Pass-1

Table 5 illustrates that the descriptor "management" which is a Pass-1 descriptor appeared in 412 documents (Our entire BMP Data has 1230 documents). It also appeared in only one map among the twelve maps (There are twelve maps created by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10).

8. Sample Network Analysis of Only the Documents With the Word "Software"

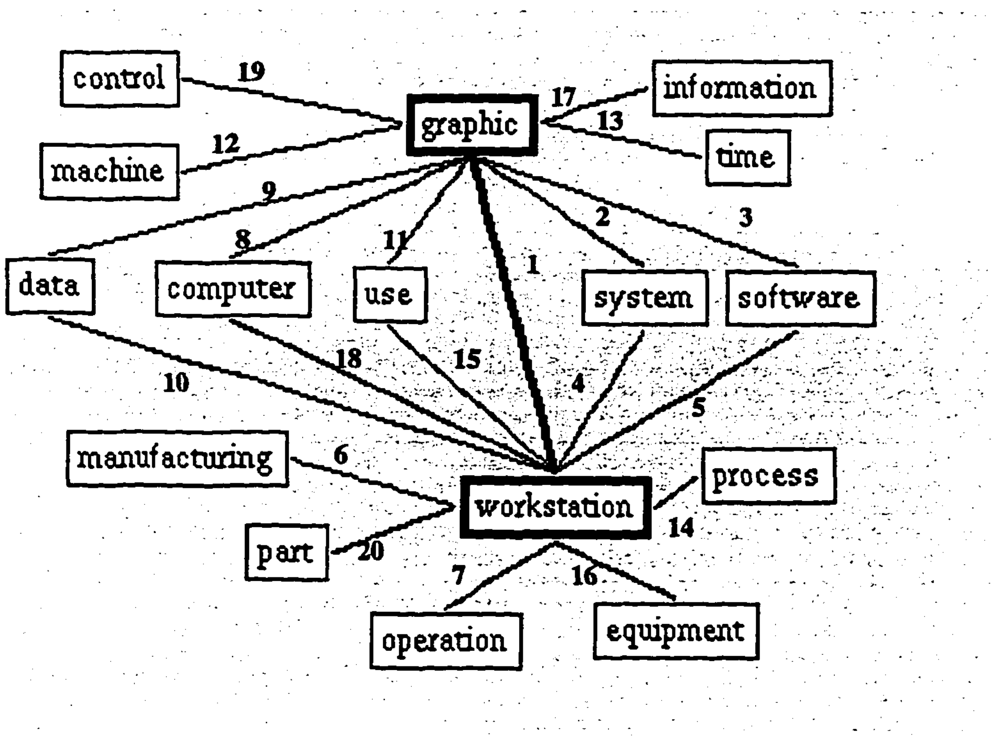

Figure 8: Network Number 11 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

Figure 8 is a depiction of different Pass-1 and Pass-2 descriptors and links. The numbering of the Pass-2 links are given according to the seriality of the creation of the links. The order can be derived from the ***results file*** for network number eleven. The link between descriptors "graphic" and "workstation" is link number 1 because it starts with highest strength.

Figure 9: Sample Results File Number 2 for Network Number 11 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

```
Run Parameters: Eliminate by Nodes
Pass Two Node Filter: Both nodes. Link Selection: Strength and Max. Nodes
Min. Strength: 0.000000. Min. Co-Occurrence: 15. Max links: 10.
Max maps 100. Max nodes 10. Maps Produced: 12

11 16 20
^workstation^ 36 1 11 1 11
^graphic^ 45 1 10 1 11
^software^ 235 12 2 2 1
^use^ 76 3 2 2 3
^system^ 195 12 2 2 1
^computer^ 109 4 2 2 1
^data^ 138 7 2 2 1
^part^ 96 6 1 2 2
^manufacturing^ 75 2 1 2 3
^machine^ 57 3 1 2 2
^time^ 120 7 1 2 1
^process^ 138 8 1 2 1
^equipment^ 106 6 1 2 2
^information^ 87 4 1 2 4
^operation^ 70 4 1 2 5
^control^ 89 4 1 2 2
^graphic^ ^workstation^ 15 0.138889 1 0
^graphic^ ^system^ 43 0.210712 2 1
^graphic^ ^software^ 44 0.183073 2 1
^system^ ^workstation^ 34 0.164672 2 1
^software^ ^workstation^ 36 0.153191 2 1
^manufacturing^ ^workstation^ 20 0.148148 2 3
^operation^ ^workstation^ 19 0.143254 2 5
^computer^ ^graphic^ 26 0.137819 2 1
^data^ ^graphic^ 28 0.126248 2 1
^data^ ^workstation^ 25 0.125805 2 1
^graphic^ ^use^ 20 0.116959 2 3
^graphic^ ^machine^ 17 0.112671 2 2
^graphic^ ^time^ 24 0.106667 2 1
^process^ ^workstation^ 23 0.106481 2 1
^use^ ^workstation^ 17 0.105629 2 3
^equipment^ ^workstation^ 20 0.104822 2 2
^graphic^ ^information^ 20 0.102171 2 4
^computer^ ^workstation^ 20 0.101937 2 1
^control^ ^graphic^ 20 0.099875 2 2
^part^ ^workstation^ 18 0.093750 2 2
0.138889 2.443884 0.332558 224 66
```

Figure 9 is a depiction of different Pass-1 and Pass-2 descriptors and links.

Table 6: Analysis of the Links of Network Number 11 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

Number of the Links from Figure 13	*Co-Occurrence*	*Strength*	*Number of Passes*
1	15	.138889	Pass-1
2	43	.210712	Pass-2
3	44	.183073	Pass-2
4	34	.164672	Pass-2
5	36	.153191	Pass-2
6	20	.148148	Pass-2
7	19	.143254	Pass-2
8	26	.137819	Pass-2
9	28	.126248	Pass-2
10	25	.125805	Pass-2
11	20	.116959	Pass-2
12	17	.112671	Pass-2
13	24	.106667	Pass-2
14	23	.106481	Pass-2
15	17	.105629	Pass-2
16	20	.104822	Pass-2
17	20	.102171	Pass-2
18	20	.101937	Pass-2
19	20	.099875	Pass-2
20	18	.93750	Pass-2

Table 6 illustrates that link number 1 (Link between "graphic" and "workstation") which is a Pass-1 link has a co-occurrence of 15 and strength of 0.138889. Link number 2 (Link between "graphic" and "system") which is a Pass-2 link has a co-occurrence of 43 and strength of 0.210712 and so on.

Table 7: Analysis of the Nodes of Network Number 11 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

Name of the Nodes from Figure 3	*Number of Documents it Appears*	*Number of Maps it Appears*	*Number of Passes*
Workstation	36	1	Pass-1
Graphic	45	1	Pass-1
Software	235	12	Pass-2
Use	76	3	Pass-2
System	195	12	Pass-2
Computer	109	4	Pass-2
Data	138	7	Pass-2
Part	96	6	Pass-2
Manufacturing	75	2	Pass-2
Machine	57	3	Pass-2
Time	120	7	Pass-2
Process	138	8	Pass-2
Equipment	106	6	Pass-2
Information	87	4	Pass-2
Operation	70	4	Pass-2
Control	89	4	Pass-2

Table 7 illustrates that the descriptor "workstation" which is a Pass-1 descriptor appeared in 36 documents (Our entire BMP Data has 1230 documents). It also appeared in only one map among the twelve maps (There are twelve maps created by only the documents with the word "Software" when the cutoff for co-occurrence level is 15 and links are 10).

9. Best Practices Analysis with a Sample Phrase

Maps do not always give a clear depiction of relationships between various descriptors. Therefore, sometimes it is essential to read the documents for clarification. But documents can be very large, and it is very time consuming to read the entire documents. Inquery, which is a probabilistic information retrieval (IR) system, can be very helpful for solving this crucial problem. I used Inquery to analyze the relationships among numerous descriptors. The following examples were taken from the BMP document with the help of Inquery. Desired words may be given at the Inquery prompt to get the corresponding documents.

9.1 Entering the Phrase "Software Development"

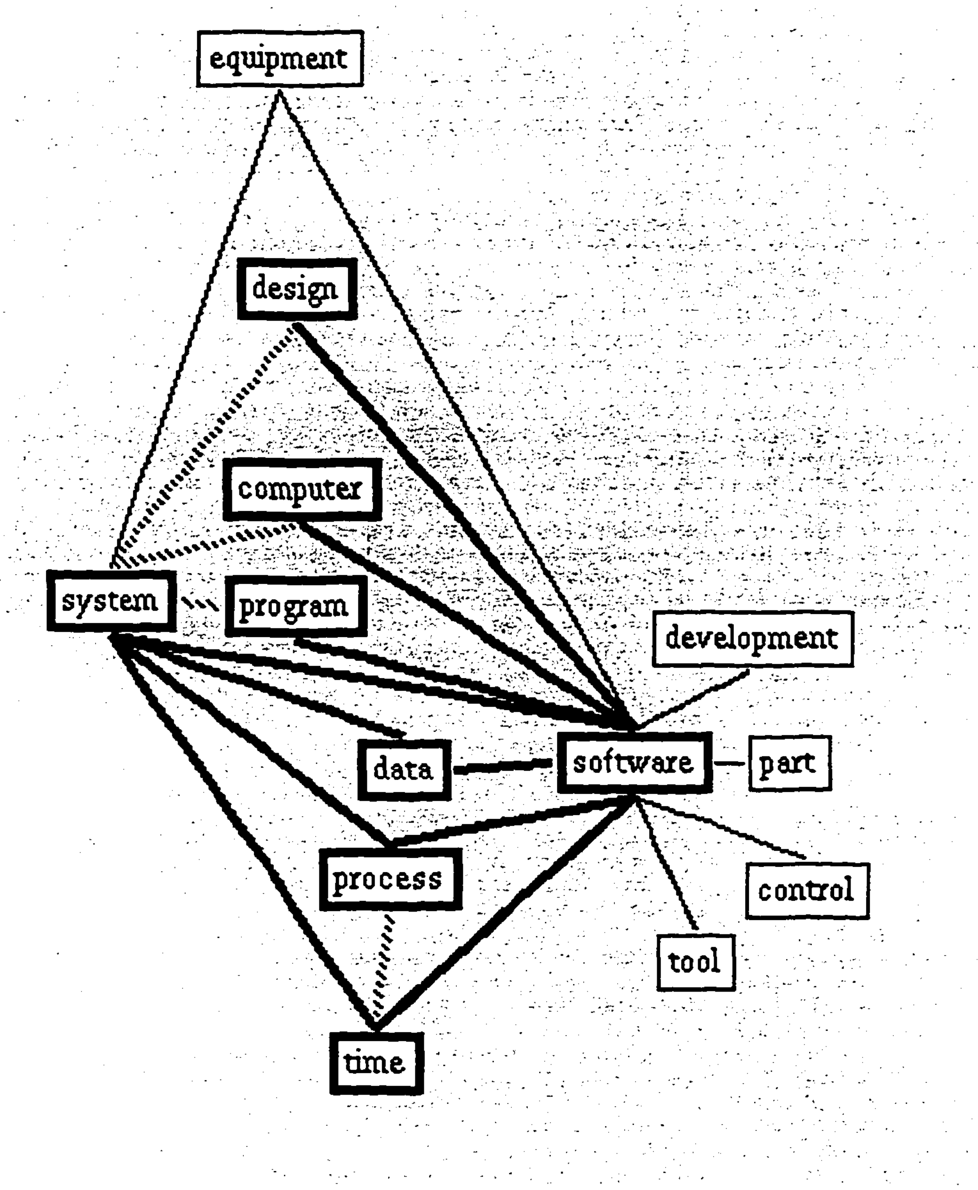

Figure 10: Network Number 1 Generated by Only the Documents With the Word "Soft ware" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

<u>Figure 10 can be interpreted as follows:</u>

Software design, software tool, and software development are all related to one another. Software design is a process through which requirements are translated into a representation of software.

Design is the phase where quality is promoted in software development. Software engineering tools provide automated or semiautomated support for methods.

The following is an excerpt taken from the BMP document. This paragraph was extracted with the help of INQUERY (here the phrase "software development" was entered at the Inquery prompt):

> "Software developers at Sandia represent most stakeholders for the preferred pro cesses for software development. Team representatives include a Process Analyst, Process Quality Management and Improvement Consultant, Management Advo cate, Education and Training Representatives, and Software Quality Managers. The teams meet on a regular schedule and review processes that address DOE orders, DOD Standards, and Sandia's Strategic Plan. The teams also conduct research into standards for software development, and identify and evaluate alter native core processes such as software requirements, design, and implementation. Software life cycle models, design processes, test processes, and review techniques are also considered preferred processes for software development. Process defi nitions include a step-by-step description of what should be performed for a spe cific phase of software development. Flexibility is built into the independent meth odology to avoid dictating how each task is accomplished. The process is meas urable -- for improvement and repeatability -- for control, and scalable for applic ability to various size projects. The preferred process for software design include definitions for overviews, entrance criteria and inputs, process summary tables, exit criteria and outputs, process changes, expected improvements, metrics, tools, training, documents and detailed task descriptions (develop and review products). Detailed task descriptions include objectives, dependencies, responsibilities, inputs, entrance criteria, task descriptions, verifications, exit criteria, outputs, an standards.
>
> The software development process begins with a set of requirements, and the design process tasks consist of eleven steps. Templates have been developed for document control for software requirements specification, design description, sy stem test plans, and DOD Standards. The Sandia preferred process for software development team has established a considerable list for private industry external distribution. By actively soliciting suggestions for improvements, Sandia has decreased the required time and effort to produce documentation, while seeing a significant savings due to early detection of defects. Formal development metho dology has fostered increased productivity results. Changes and improvements will continue with the development and implementation of feedback surveys" [INTE1].

However, there are ample documents which has the phrase "software development" in them. Of course, Figure 10 alone by itself gives us some idea about software development. However, if we read the actual document along with the maps the concept gets much more unambiguous.

10. Elaborated Interpretation of Some Sample Figures:

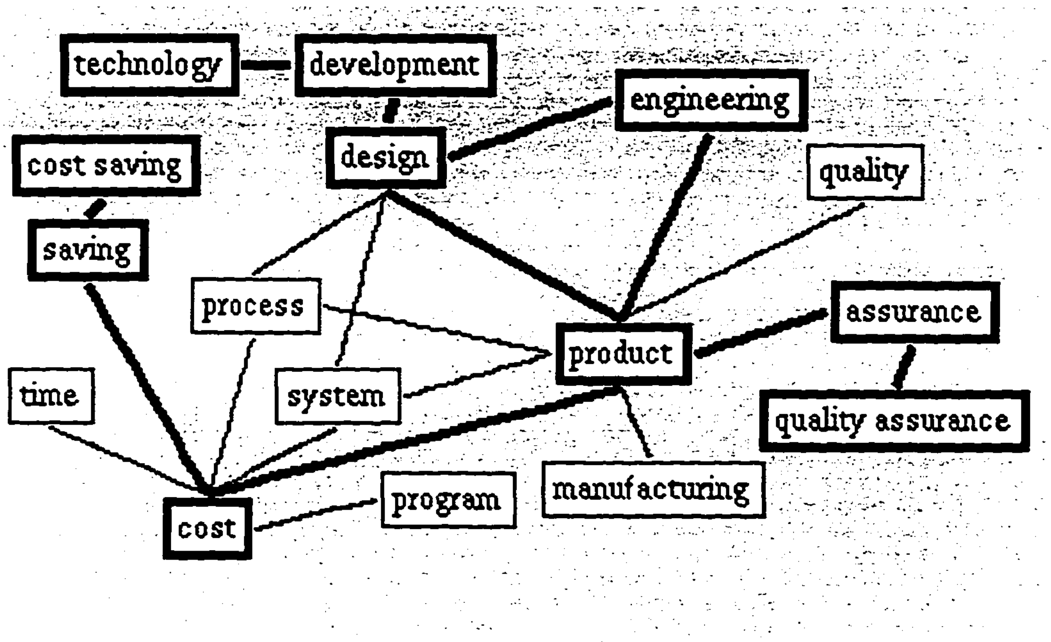

Figure 11: Network Number 2 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

<u>Thorough Interpretation of Figure 11:</u>

Figure 11 can be interpreted in the following manner:

- It is the responsibility of design to define the materials and the format into which they must be processed, and to make a product that will meet the design objectives of quality, reliability, and cost.
- It is the responsibility of manufacturing engineering to develop and define the manufacturing process, including the tools and methods to be used, that will make the product to the specified quality and reliability, on schedule at the agreed cost. In some companies manufacturing engineering also has the responsibility of taking the basic design and reconfiguring it for best manufacturability [INTE1].
- Manufacturing process must use the defined tools, methods, materials, and people to make the

product according to the developed process, on schedule at the agreed cost.

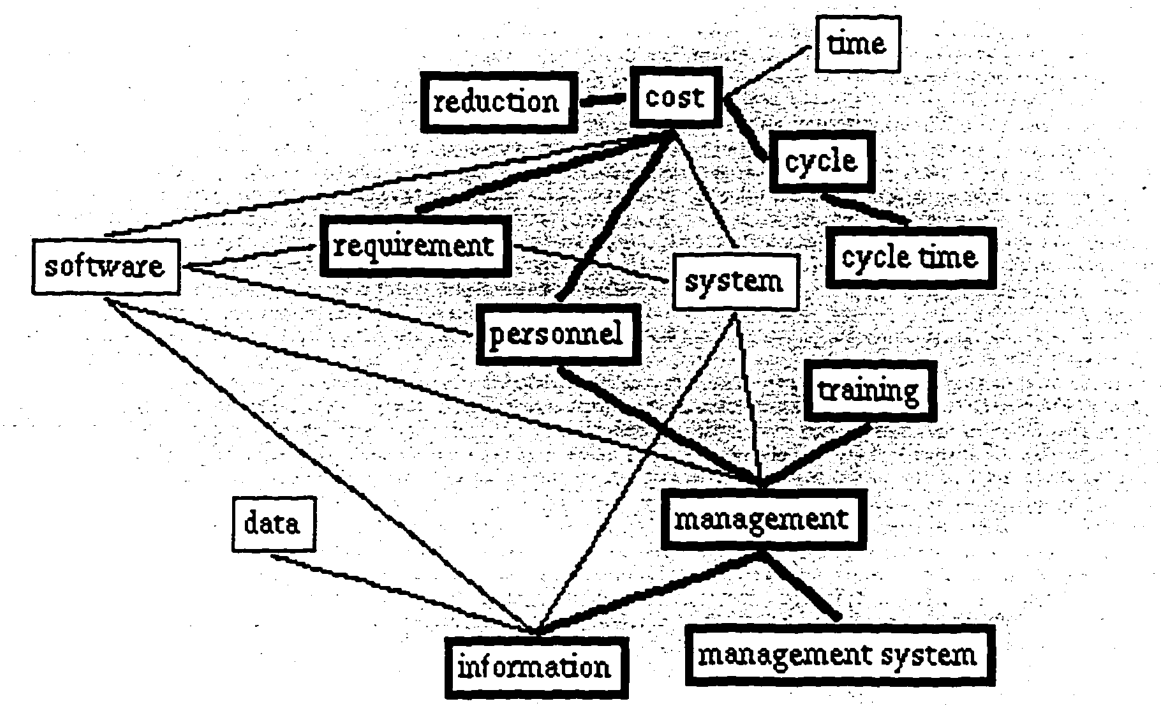

Figure 12: Network Number 4 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

Thorough Interpretation of Figure 12:

The main theme of Figure 12 is ***cycle time***, and the Figure can be interpreted in the following manner:

- Critical factor in any company's success, time-to-market refers to the cycle time necessary from product initiation through introduction. The faster the time-to-market, the greater the opportunity to successfully sell the product.
- An effective management system indefectibly marks revisions and a contention system to resolve product and period conflicts and determine the balance between line and staff. Proper managerial training can help to meet the goals of total customer satisfaction and

total cycle time reduction.

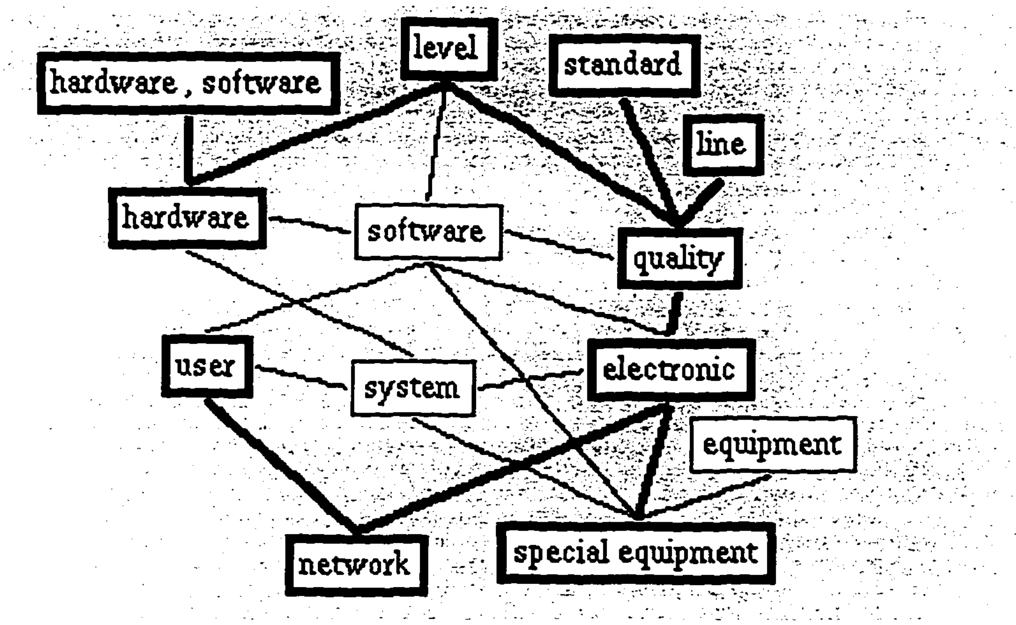

Figure 13: Network Number 6 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

Thorough Interpretation of Figure 13:

Figure 13 can be interpreted in the following manner:

- Quality must be built into a software product during its development to satisfy quality requirements established for it.
- The quality plan should clearly state which quality attributes are most significant for the product being developed. It should also state which standards are appropriate to the product and to the process, and define the plan for developing these standards.
- Quality should be engineered into the software product; it is not an attribute which can be added to a product. Quality should be the driver of the software process; at all stages of development, achieving high quality results should be the goal.

Software quality is often defined as the degree to which software meets requirements for reliability, maintainability, transportability, etc., as contrasted with functional, performance, and interface requirements that are satisfied as a result of software engineering. Primary responsibility of managers, the fundamental one around which all others will revolve, is to break projects into manageable steps, then deliver the new software produced by each step into the client's hands, so they can give you objective feedback for the next step.

11 Extended Analysis

Foremost, lexical maps were created from the entire BMP data. The intention behind this is to get a general idea of the best manufacturing practices. Now what about software? To understand the software process, I selected the database related to only the word "Software." The objectives of this were to analyze software engineering, to find the distinguishing characteristics of well-engineered software and to corroborate the notion of software process, that is, the exertions involved in software production. Also, another purpose was to correlate software engineering with manufacturing engineering. This was the reason behind creating the second set of data. My intention of creating the third set of data by taking off the word "Software" from the BMP document was to investigate more on the BMP data without any software ascendancy on it. At this point, I realized that there were many software related phrases (e.g., software development, software design, etc.) that were still not isolated from the original BMP document. Therefore, by taking off all of the word "software" and all of the "software" related phrases I created the fourth set of data. The main intention of doing this was to see how the variations of including or excluding the word "software" can enlighten us to correlate software engineering practices to manufacturing engineering practices.

The study of the maps from **Appendix C (Figure A. 1-1 to Figure A. 1-14)** for the entire BMP data shows us the various important aspect of manufacturing practices. Especially, "Quality" and "Design" are the two dominant themes depicted in the **Figure A. 1-2 of Appendix C**. If the manufacturing process is truly in control, if all of the variables are eliminated, then each unit of the product will be exactly the same and the product quality and reliability will be determined solely by the design. If the designer has taken into account any possible variations in the materials and parts and in the tools and equipment, controlling the process will not be difficult and minor variations will not affect the product. Design cannot be separated from production. They are all part of the manufacturing process and must recognize their reliance on each other. Manufacturing must understand the problems of the design team, and design must recognize the limitations of the manufacturing process. The only quality objective is perfection. The product specification must define the limits below which the product will never fall [INTE1]. The quality control department must provide the overall monitoring to assure that materials, machines, equipment, and processes, exactly adhere to the developed process.

Figure A. 1-3 of Appendix C depicts how management plays an important role in the realm of process improvement. Consistency in management's attitude toward change and evolution of the organization to deal with the requirements is paramount. If the key success factor is new product introduction there is a requirements for management and the work for to be adept at coping effectively with continual change and to ensure that the planning and control system is able to deal with an increasing number of products and associated activities. Another subtle requirement occurs when technical leadership is identified as a key success factor. Effective management in today's environment presents challenges. Managers who survive and thrive will be those who develop the ability to manage their organizations as a combination feedback and feed-forward control system. In a similar way, **Figure A. 1-45 of Appendix C** portrays the same main emphasis on management. This Figure was created after I took off all of the word "Software" and "Software" related phrases.

Figure A. 1-17 from Appendix C shows a correlation between Manufacturing Engineering and Software Product. Like all engineering, software engineering is not just about producing products but encompasses producing products in a cost-effective way. Given unlimited resources, the majority of software problems can be presumably be resolved but the challenge for software engineers is to produce high-quality software with a finite amount of resources and to a predicted schedule. Sound engineering practices are needed to get useful products.

The central theme of **Figure A. 1-19 of Appendix C** is Testing. The descriptor "Testing" is a Pass-1 descriptor. According to our algorithm, we know from our earlier **section 3.3.2,** that Pass-1 build networks that can identify areas of strong focus; Pass-2 can identify descriptors that associate in more than one network and thereby indicate pervasive issues. The descriptor "Testing" in this Figure is a Pass-1 descriptor, and central idea flows to it. The testing process focuses on the logical internals of the software, ensuring that all statements have been tested. Quality is not tangible. The purpose of testing is to make quality visible. There are obvious practical difficulties associated with testing software system, such as having poorly expressed requirements, informal design techniques and nothing executable available until the coding stage. I can also comprehend that on top of these are the various psychological problems. In the typical business system usability and maintainability are key factors, while for a one-time scientific program neither may be significant.

According to the definition of density and centrality, we can analyze super networks should be low in the internal strength but high in the position in strength of interaction with other networks. If we look at the **Figure A. 1-14** from Appendix C, we shall get a better understanding of this concept. Notwithstanding, this super network does not show much about density. On the other hand, **Figure A. 1-14** shows a lot about centrality.

12. Conclusions

12.1 Future Research

Here are some topics for future research:

- Find more efficient ways of eliminating redundant terms from maps.
- Investigate other criteria for forming super networks (i.e.e, besides just the 2 Pass-2 links higher than the lowest Pass-1 link)
- Investigate other measures for centrality and density (i.e., varying the statistical test to see how sensitive the results are)
- Investigate use of parsing based on verb phrases, and not just on noun phrases.

12.2 Conclusions

On the whole, Software Engineering Practices can learn immensely from best manufacturing practices. The ultimate goal of the Best Manufacturing Practices program is to strengthen the U.S. industrial base and reduce the cost of defense systems by solving manufacturing problems and improving quality and reliability. Best Manufacturing Practices also maintains a substantial aggregate capability in many diverse areas of manufacturing, manufacturing support, research and development. Knowledge of basic processes, tool design, and computer-aided manufacturing can be applied directly to various fields of software engineering. Best Manufacturing Practices can assist Software Engineering Practices in a way that when Software Companies use these practices they can:

- Improve both software quality and staff productivity
- Determine the current status of the organization's software process

- Set goals for process improvement
- Create effective plans for reaching those goals
- Implement the major elements of the plans

Implementing software inspections must be an important step along the path to a more mature software development process. Product quality is one of the best measures of a software development project.

Companies either manufacturing or software should concentrate on customer satisfaction, should have an open mind to new application and change, and should be able to adapt to new technologies. The exchange of knowledge between the best practices of manufacturing engineering and software engineering will enhance quality and competitiveness and will optimize the use of technology and resources for the twenty first century.

13 Definitions, Acronyms, and Abbreviations

BMP	Best Manufacturing Practices
Co-Occurrence	Co-Occurrence analysis is a technique for extracting the strength and breadth of term associates in large corpora.
Descriptor	A word from the index terms
External Nodes	Pass-2 nodes
External Links	Links between external nodes
GUI	Graphical User Interface
Index Term	Same as descriptor
Internal Nodes	Pass-1 nodes
Internal Links	Links between internal nodes.
Network	A set of nodes (descriptors) connected by links
Node	An occurrence of a descriptor which is linked to at least one other node within a network.
Pass-1	The first pass of the co-word algorithm which produces pair-wise connections of descriptors in networks. These networks can identify areas of strong focus.
Pass-2	The second pass of the co-word algorithm which generates links between Pass-1 nodes across networks, thereby forming associations among completed networks.
S value	The strength association between two descriptors (index terms).

Appendix A

Listing of the Mark-up Tags and Commands for Processing Text and Creating Leximaps

Markup Tags for Prep Files:

* beginning of an item (document)

\# beginning of the unique identifier

<identifying alphanumeric>

\# close of unique identifier

\! beginning of the contents of an item (document)

<text>

\! close of an item (document)

* close of the item

Table A. 1-1: Phrase Based

Commands	*User Defined Input* (* stands for prefix appended to various suffixes like .prep .stop .tag etc.)	*User Defined Output* (* stands for prefix appended to various suffixes like .prep .stop .tag etc.)	*Implied Output*
before_tagger	*.prep	*.pre-tag	
tagger	*pre-tag	*tag	
reg-exp-parser	*tag	*parse	

Table A. 1-2: Table to Create Lexical Map Analysis

Commands	*User Defined Input*	*User Defined Output*	*Implied Output*
clust1			
sort_files (must be done in the same directory)			
clust2			
check -t 0 -L -S -l 10	*		*.index *.cl *.cl-names (*.index needs to be edited to remove freq count to left of the index term. Each field is separated by a blank. If a field is blank then it does not apply).

Table A. 1-3: Table to Create Lexical Maps

Commands	***User Defined Input***	***User Defined Output***	***Implied Output***
lm1	*.index		*.index.LMDB.not-strict
lm2 -c <#> -m <#> -l <#>	*.index.LMDB.not-strict	*.c#.m#.l#.results	

Table A. 1-4: Table to Show Lexical Maps

(The DISPLAY needs to be setup to show lexical maps).

Commands	***User Defined Input***	***User Defined Output***	***Implied Output***
gui&			

Go to file menu and open *results file. There the first map is shown which has its own menu from the menu pull down layout menu and also set default edge (default is 100 - start with that). Neato layout can be chosen after that.

Choose help on original menu for further features.

APPENDIX B

Inquery

Overview

Inquery is a Probabilistic Information retrieval (IR) system developed at the University of Massachusetts, Amherst. Given a database of textual documents Inquery can retrieve documents relevant to a user's query. The documents do not need to be pre-classified by hand, though they do need to be indexed by Inquery's parsing system. The queries can be formulated either in natural language (i.e. a regular English sentence) or in a more exact structured query language.

Creating Inquery Database:

The Inquery system requires that a collection of documents be indexed to form an INQUERY database before the collection can be queried and specific documents retrieved. Two forms of the query processor are available. A batch version for running INQUERY from files containing queries and relevance judgements, is invoked from the "inquery" program. This version prints results in a form that allows statistical tests and performance changes to be evaluated. An interactive "inquery" program is provided to allow users to type queries from a command line and browse through the retrieved documents.

There are several steps to build and run inquery:

After getting the bp.dat file, the following command needs to be executed to build inquery:

inbuild bp bp.dat -fields -stoplist stopwrd

It is important to note that the file stop.wrd must be in the directory.

Note: Here stop.wrd is the stopfile.

Once inbuild completes (will take several minutes and will build several files) run Inquery with command inquery bp

For In build the command line:

inbuild database_name collection -fields -stoplist stop.wrd

Note: stop.wrd is the stopfile in the directory with info.dat the command in that directory would be:

inbuild info info.dat -field -stoplist stop.wrd

Typing inbuild -h at the Unix prompt provides info on this plus other flag and defaults. The suffix to the dat file should not be a long name. For example use 'r.dat' instead of 'repository.dat' for the collection name but 'repository for the databae_name.'

To Run Inquery on the Database Type:

path_name_to_compiled_inquery_files../inquery database_name

The number of documents that need to be brought back requires to be set and is needed to issue queries.

Type ? <return> to get help and x to get back to the query prompt. In order to leave inquery type exit and the y.

Some Helpful Inquery display Commands:

The inquery program will ask the user for the number of documents to display and then prompt for a query. Once a set of documents have been received, the bottom of the screen will display the following options:

page ***FORWARD/BACK***: Move forward/back one display page

TOP/END: Move to top /end of display

LINE UP/DOWN: Move up/down one line of display

WRITE: write the retrieved summary list to the specified on the commandline or docs.out by default.

#: Enter a document rank number to view the contents of a document. To view the first ranked document, type 'I'

Query: Submit a new query

?: Display a help screen. some additional options not shown on bottom of display are available

REFRESH: Refresh the display in event of outside screen interruption

EXIT: Exit the inquery program.

Note:_Except for the '#' option, the user should type the character displayed in upper case to select teh option. The typed letter itself should not be in upper case. <Return> key after every command.

APPENDIX C

Maps of All Networks

Following are maps of all networks generated by the co-word analysis used in this study. These images were captured directly from the output of a graphical user interface and are presented in that form.

Pass-1 descriptors are enclosed by thick boxes; while Pass-2 descriptors are enclosed by thin boxes. Pass-1 links are shown by thick lines, Pass-2 links are shown by thin lines. Hashed lines indicate two Pass-1 nodes linked during Pass-2; such links are treated as Pass-1 links because they join two Pass-1 nodes.

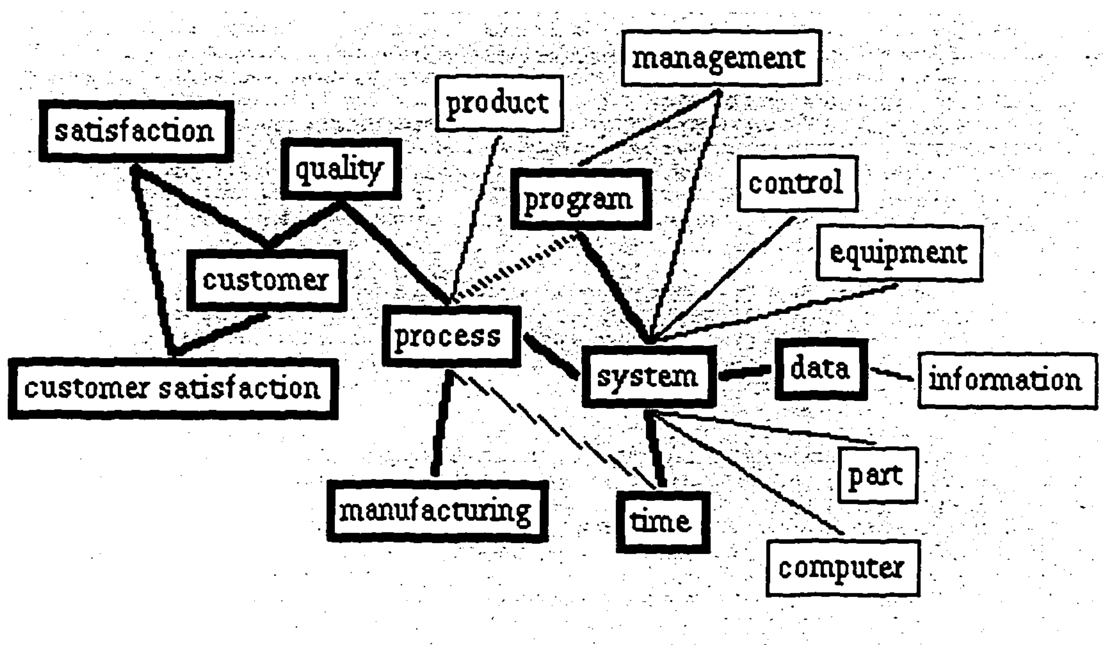

Figure A. 1-1: Network Number 1 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

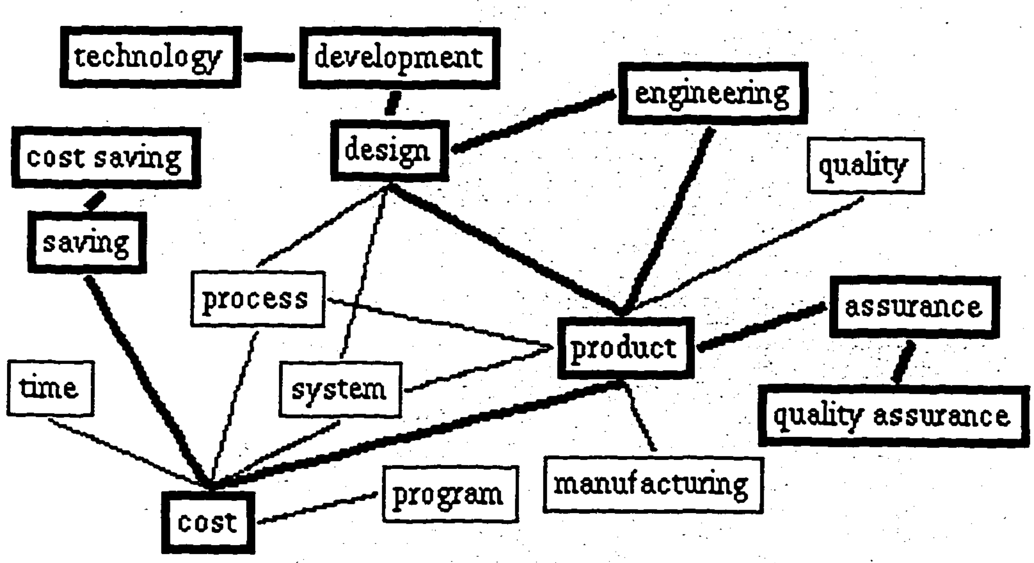

Figure A. 1-2: Network Number 2 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

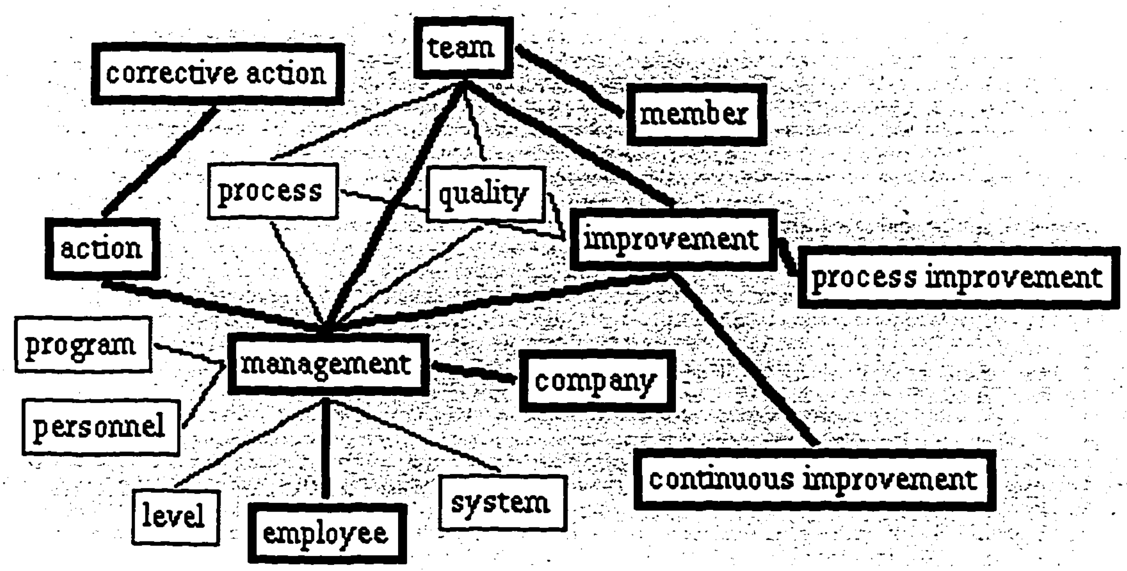

Figure A. 1-3: Network Number 3 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

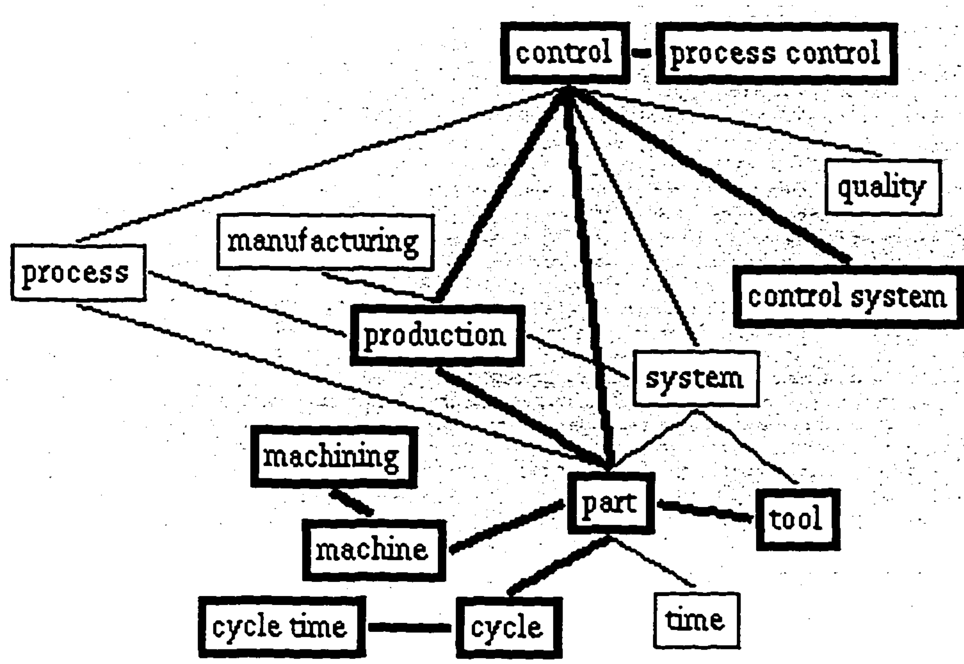

Figure A. 1-4: Network Number 4 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

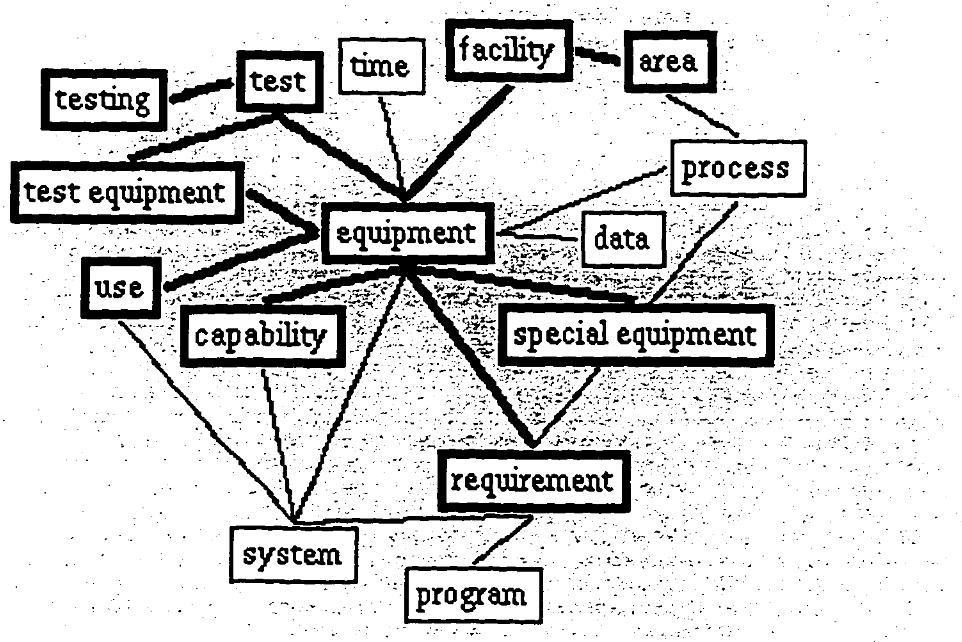

Figure A. 1-5: Network Number 5 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

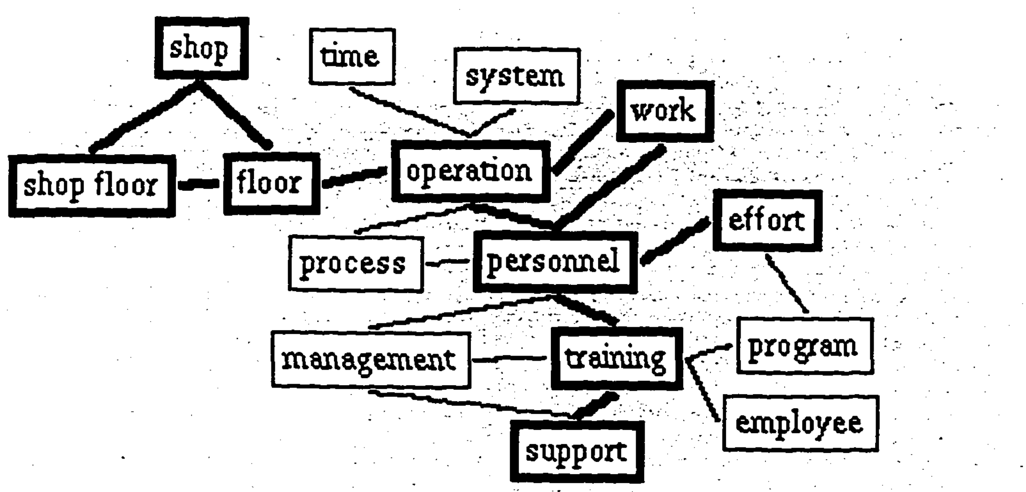

Figure A. 1-6: Network Number 6 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

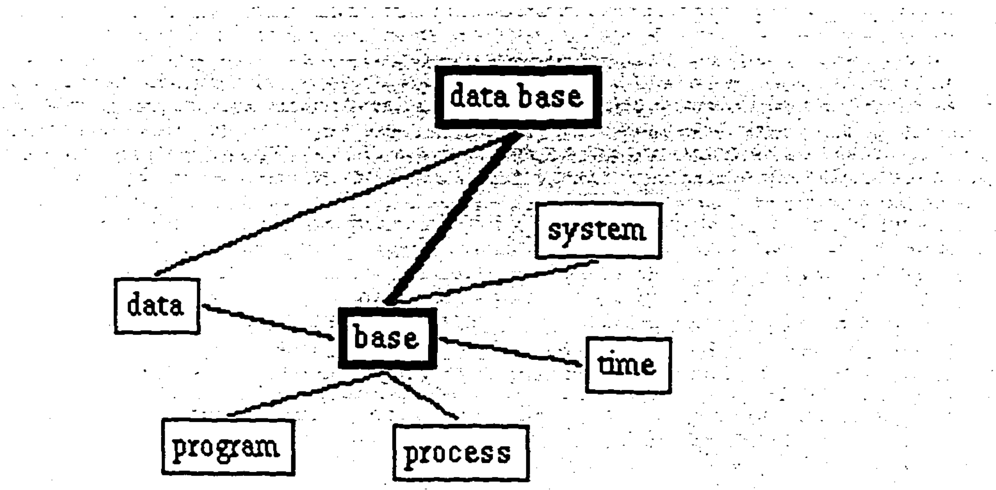

Figure A. 1-7: Network Number 7 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

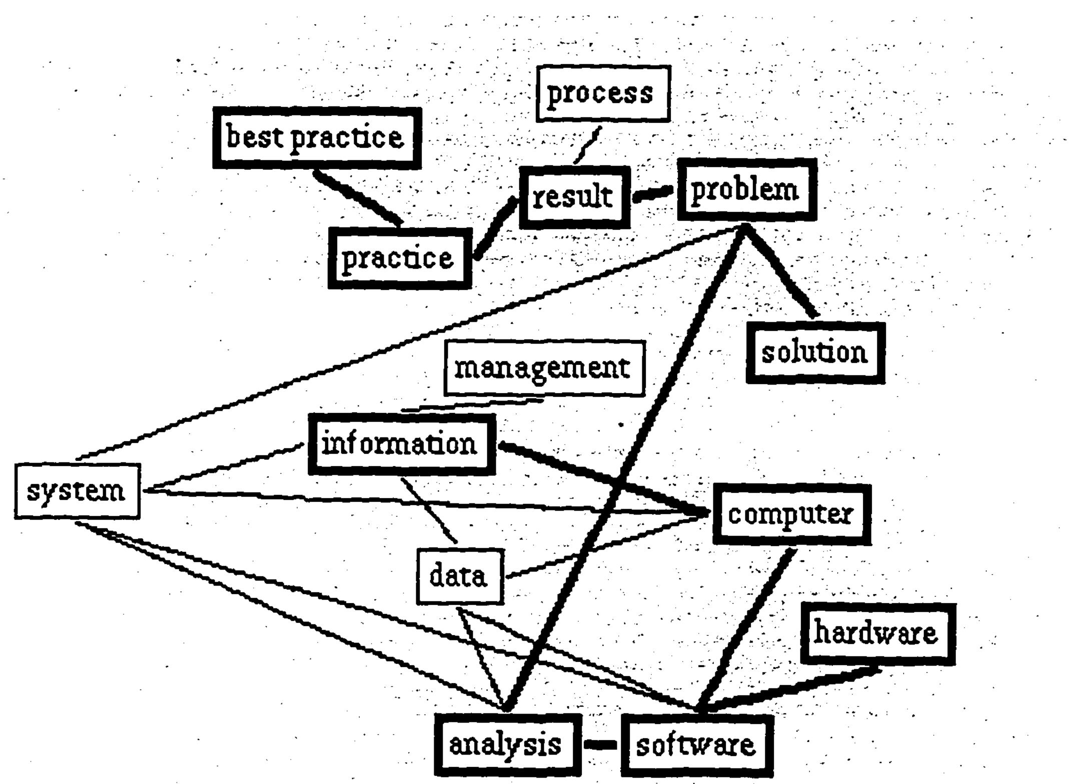

Figure A. 1-8: Network Number 8 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

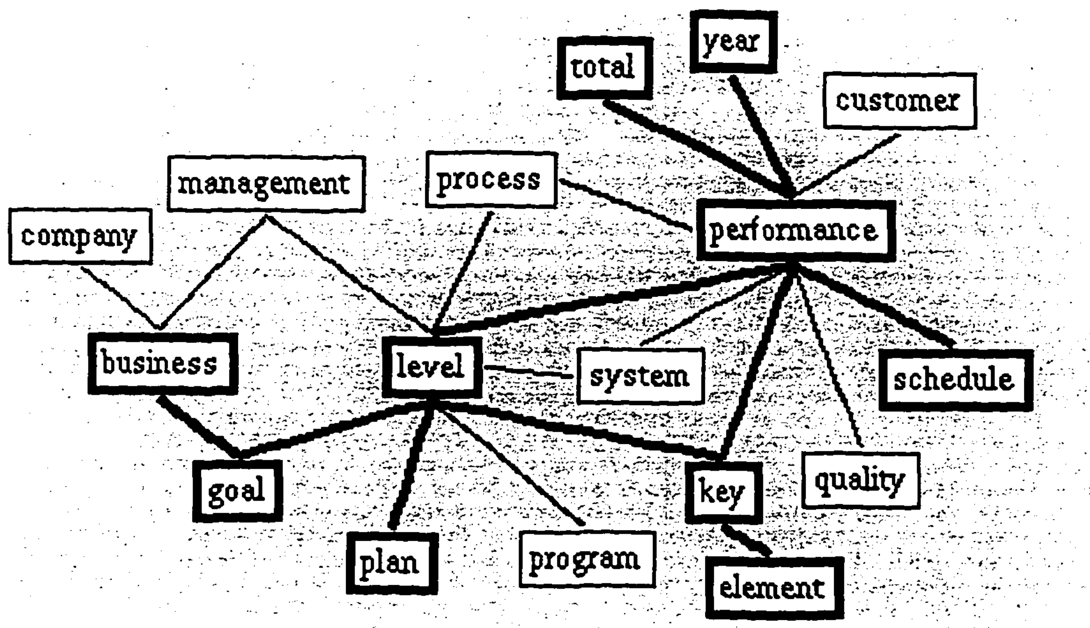

Figure A. 1-9: Network Number 9 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

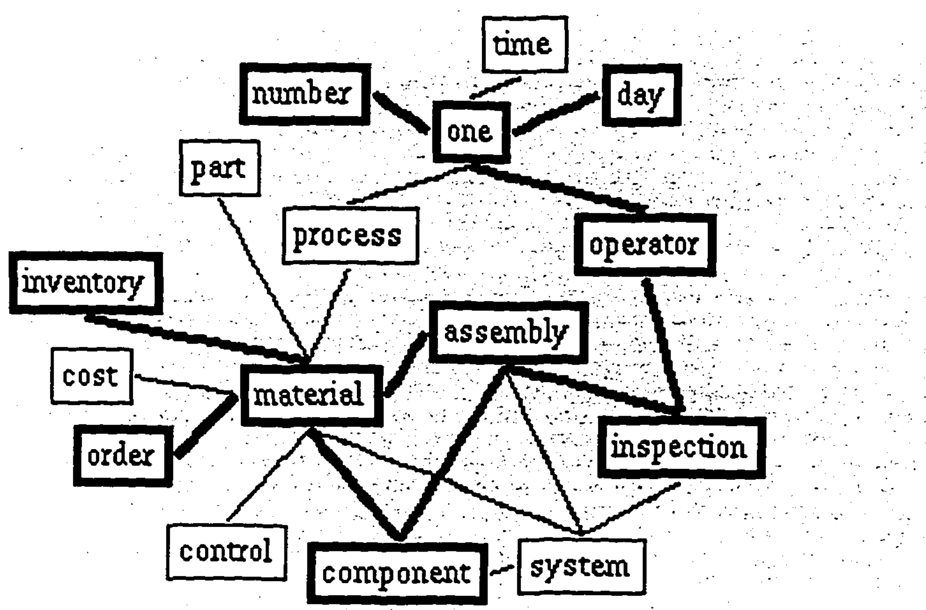

Figure A. 1-10: Network Number 10 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

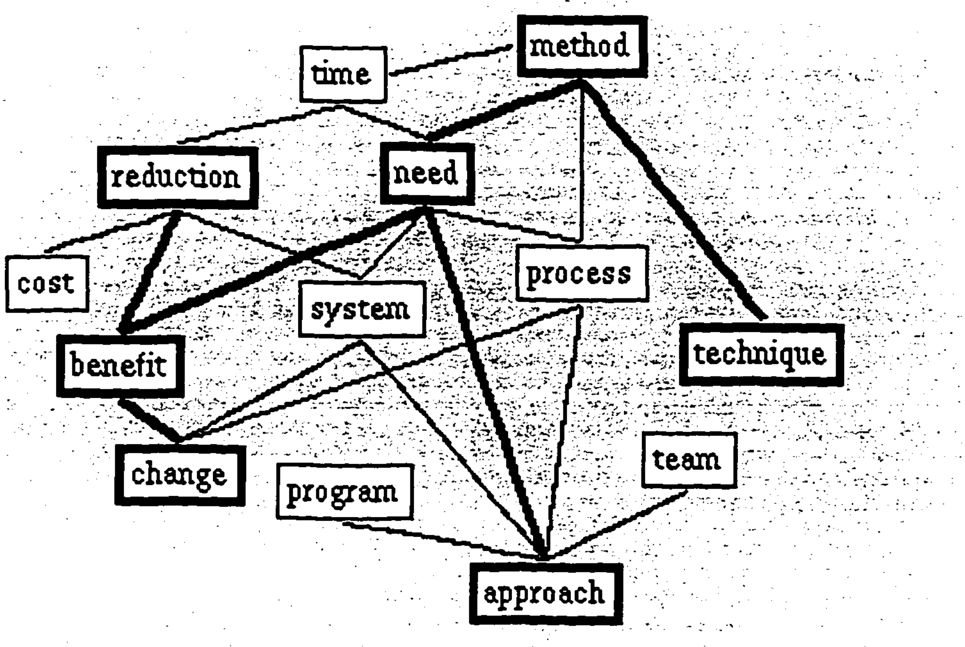

Figure A. 1-11: Network Number 11 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

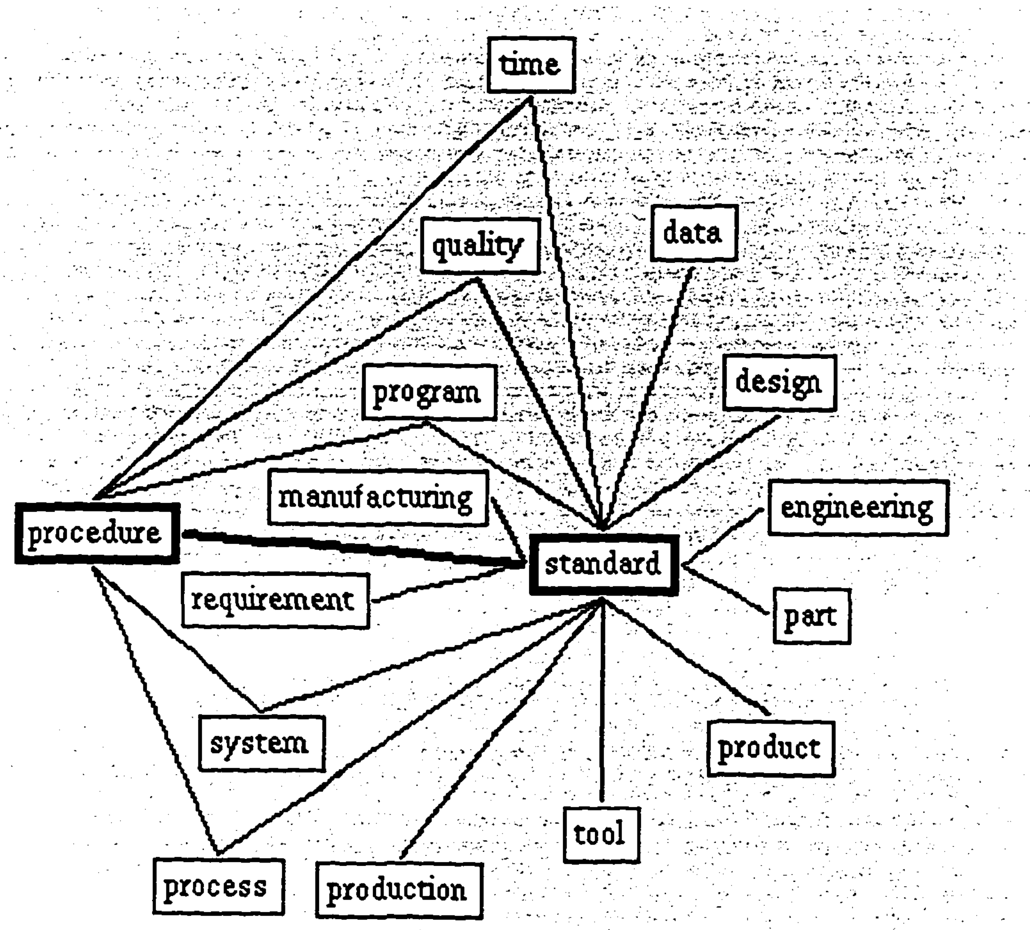

Figure A. 1-12: Network Number 12 Generated by the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

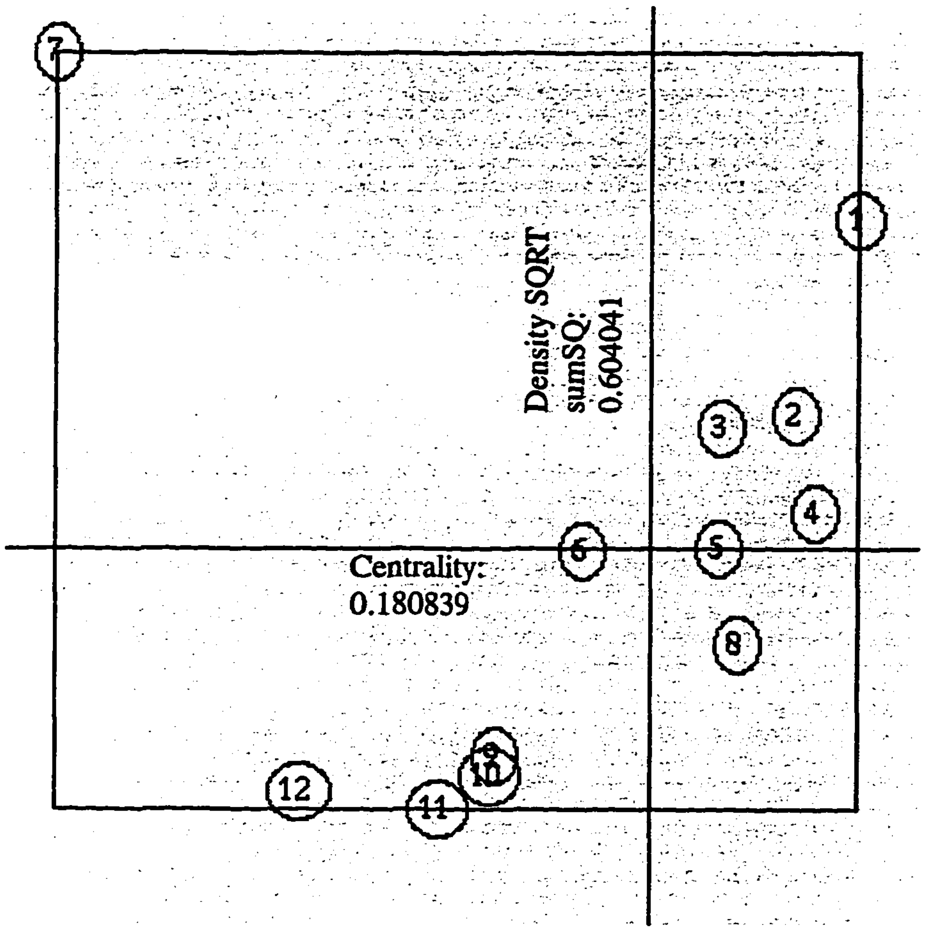

Figure A. 1-13: Density and Centrality Map for the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

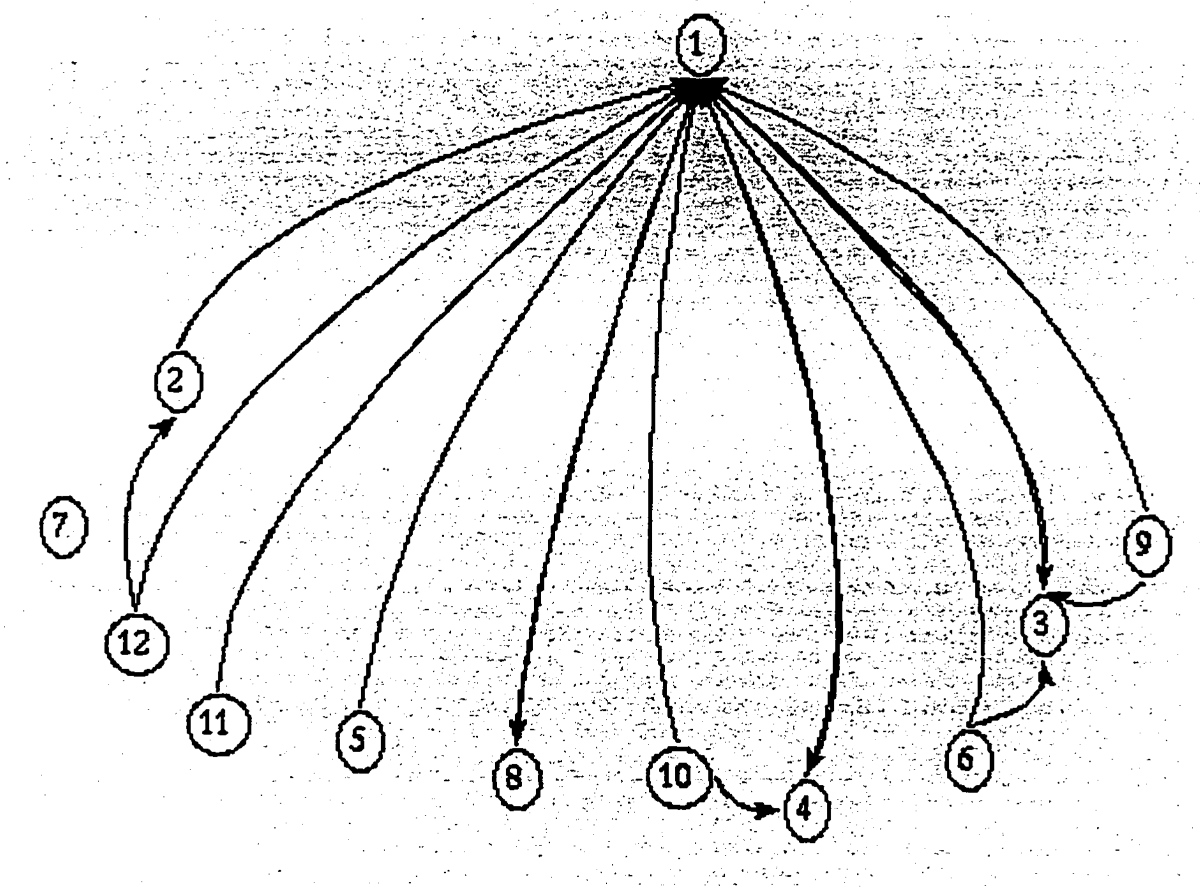

Figure A. 1-14: Super Network for the Entire BMP Data when the Cutoff for Co-Occurrence Level is 50 and Links are 10

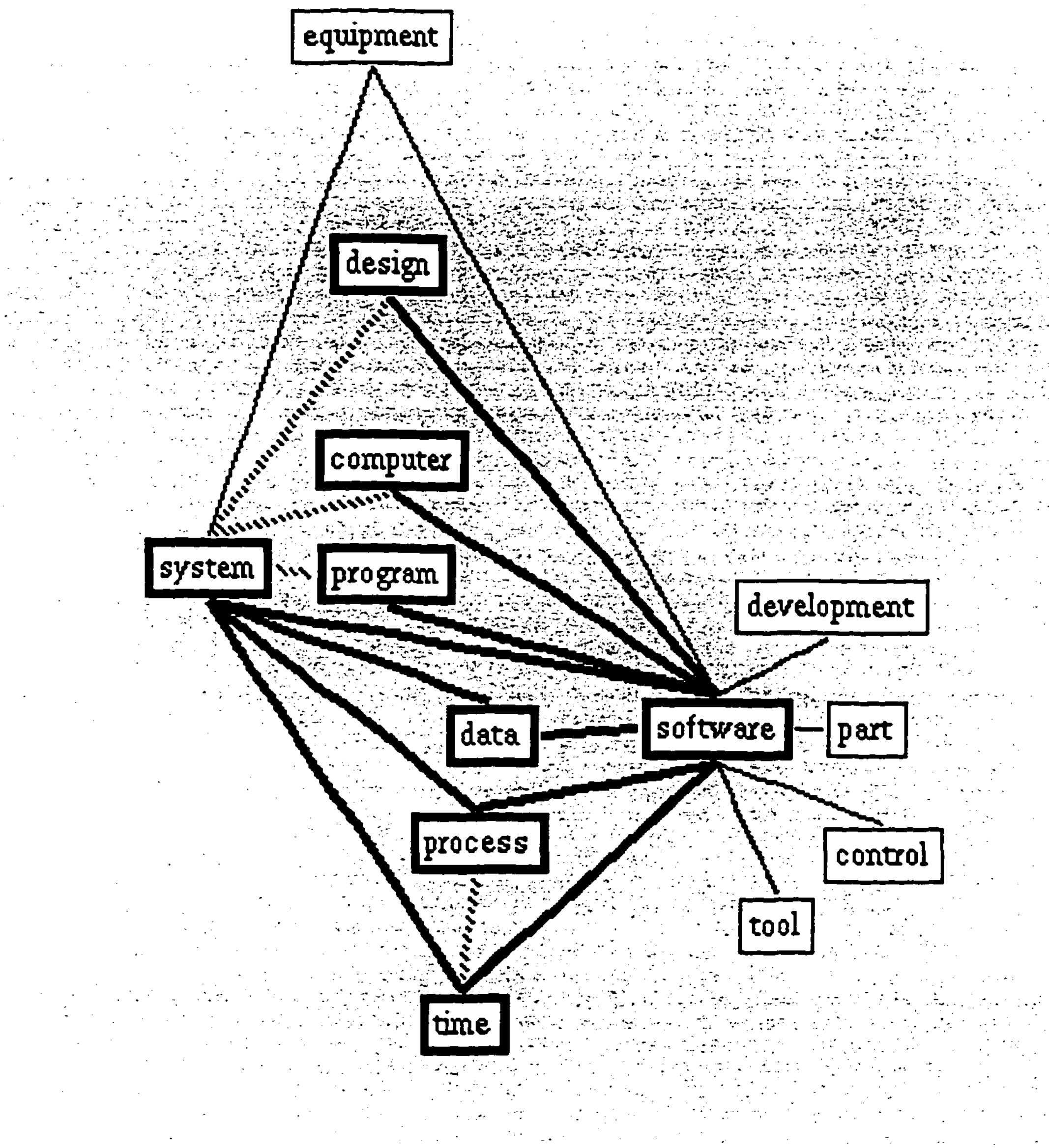

Figure A. 1-15: Network Number 1 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

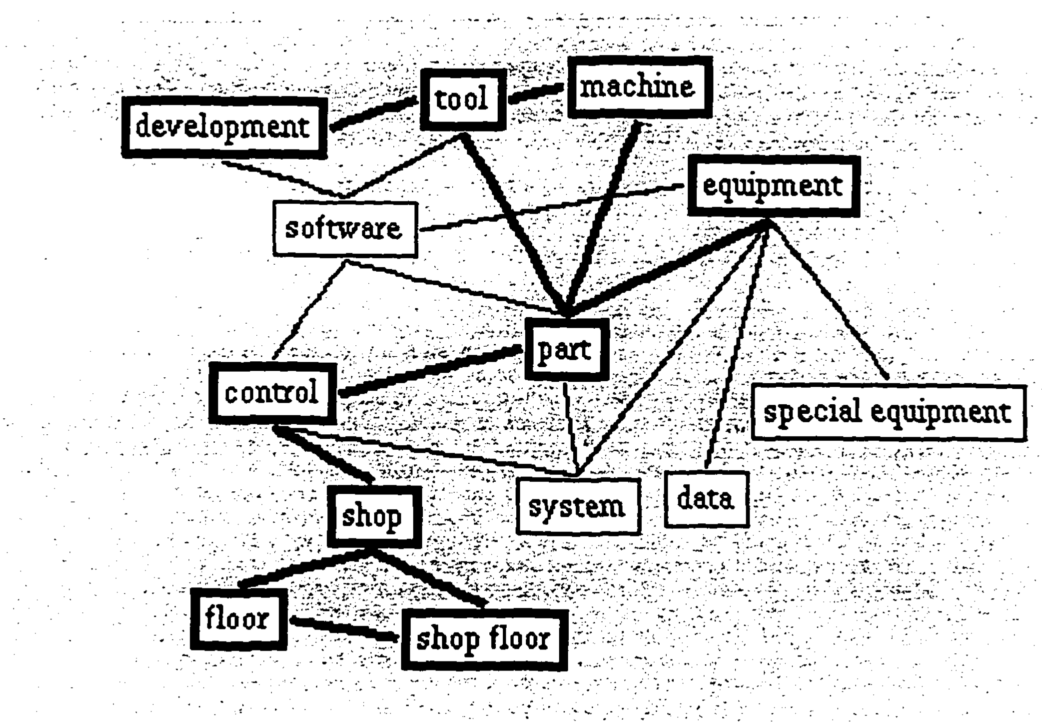

Figure A. 1-16: Network Number 2 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

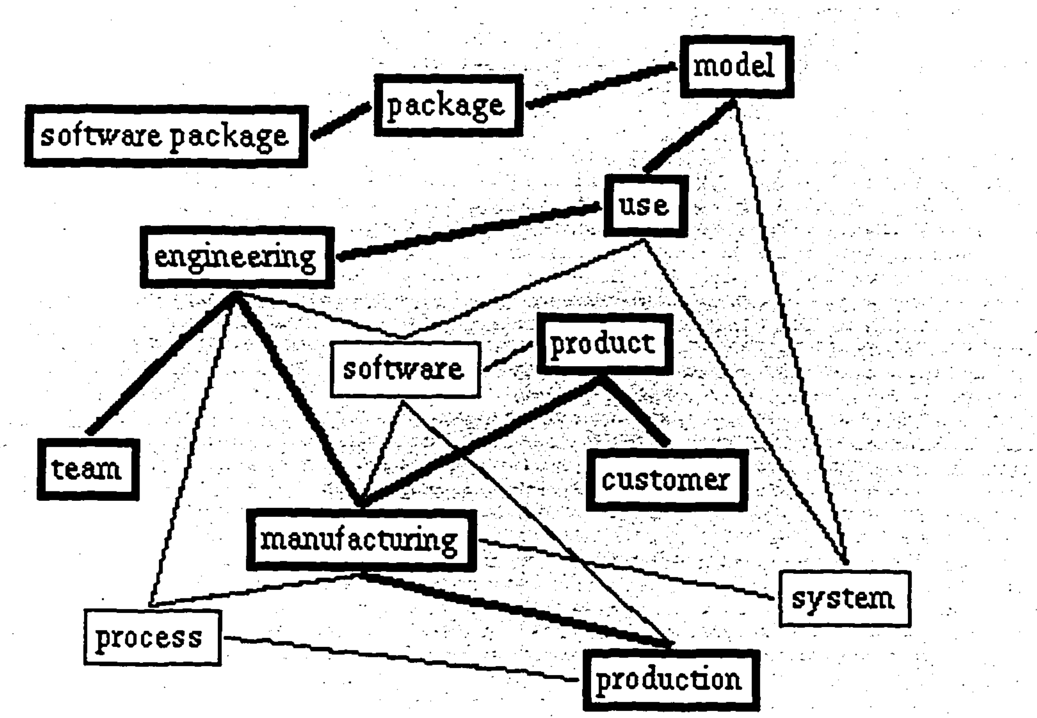

Figure A. 1-17: Network Number 3 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

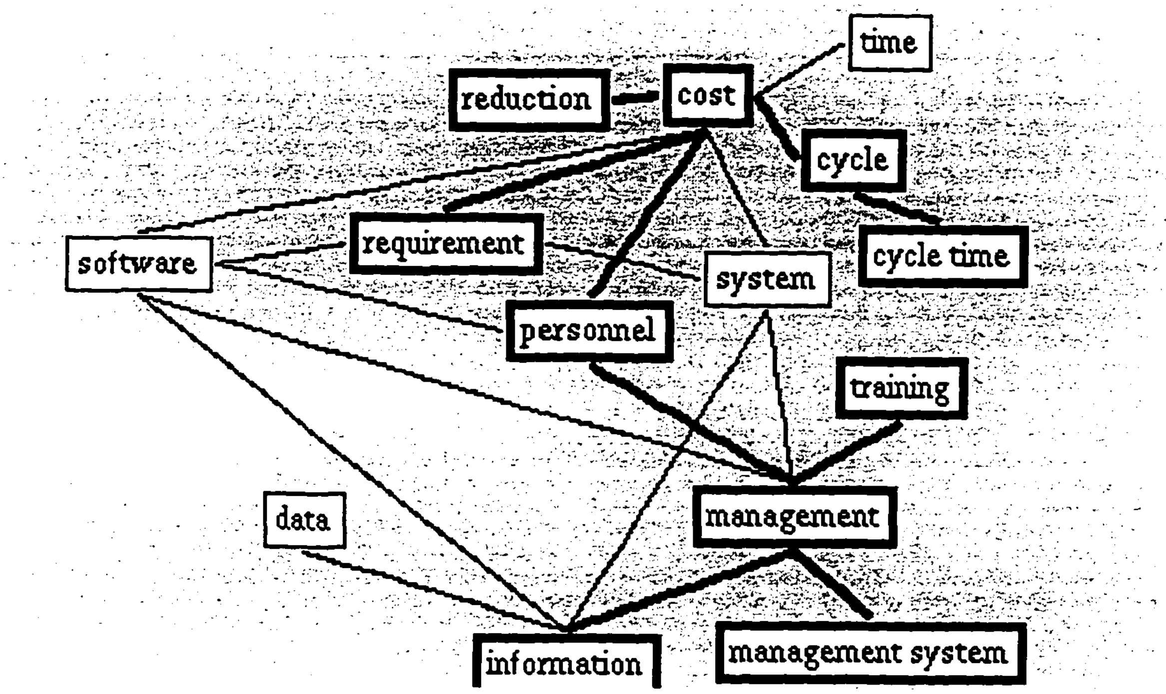

Figure A. 1-18: Network Number 4 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

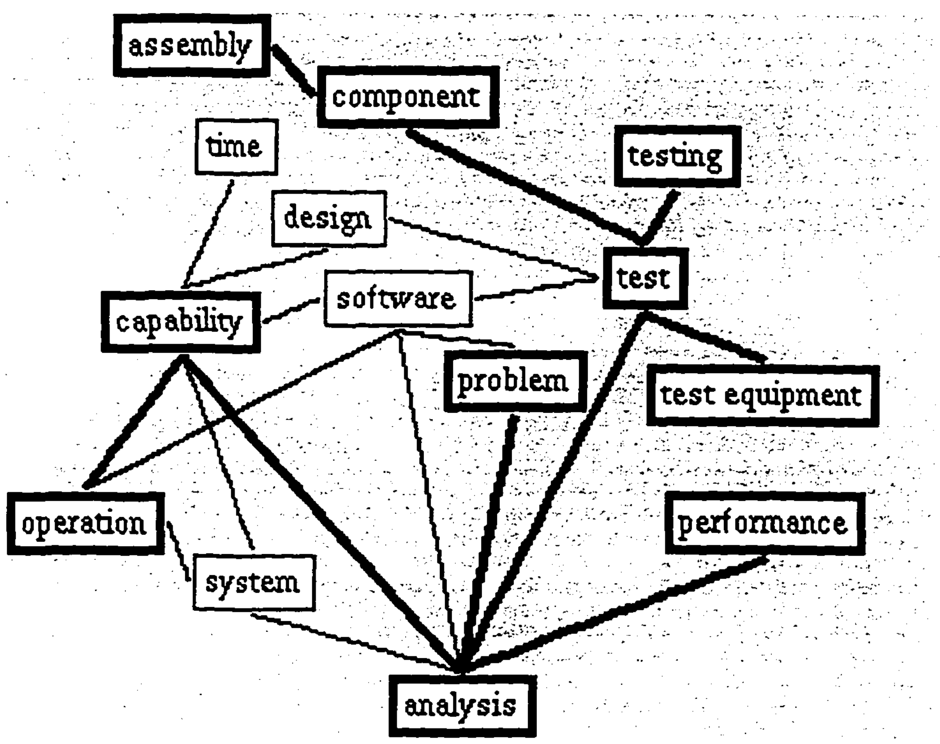

Figure A. 1-19: Network Number 5 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

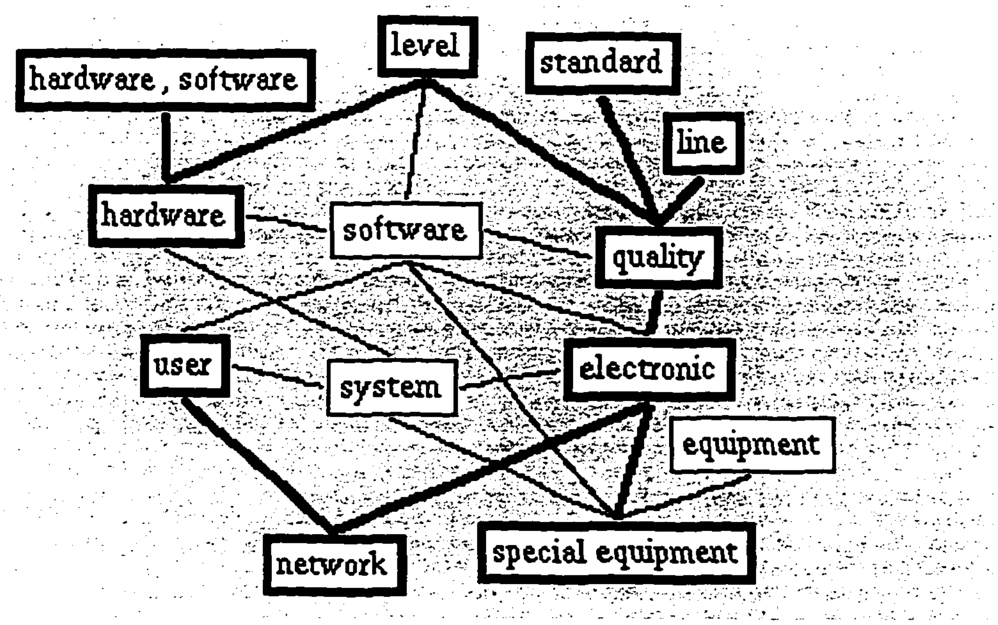

Figure A. 1-20: Network Number 6 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

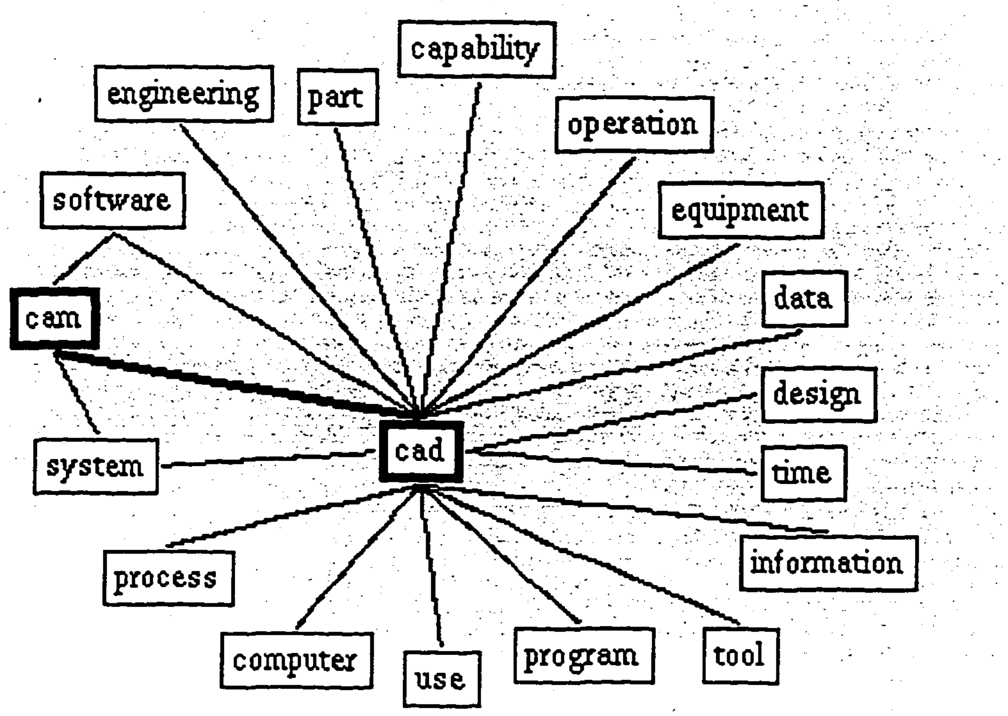

Figure A. 1-21: Network Number 7 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

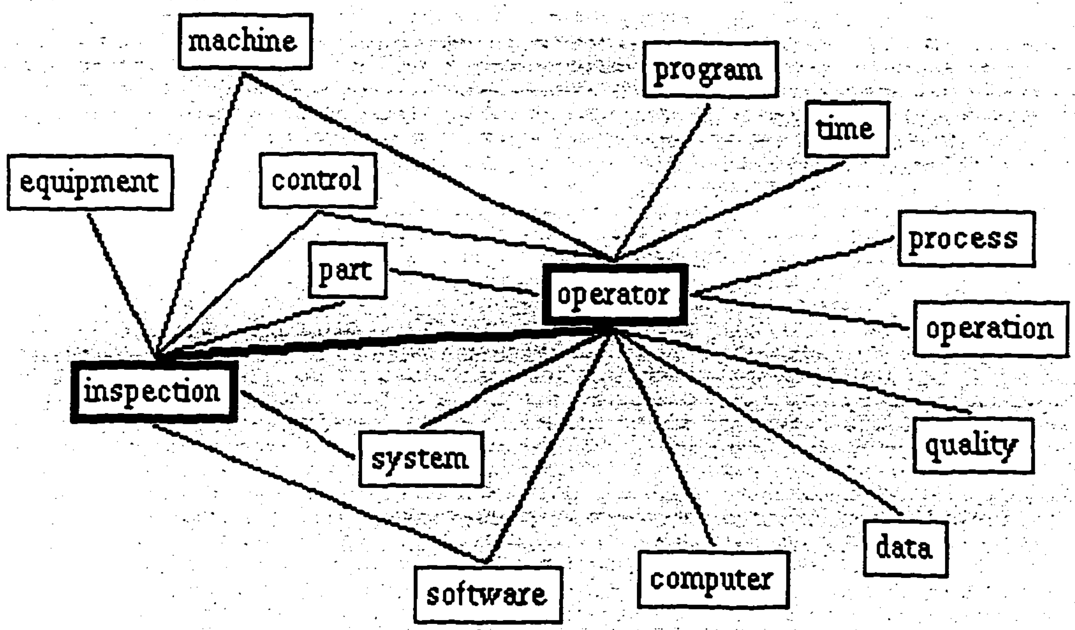

Figure A. 1-22: Network Number 8 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

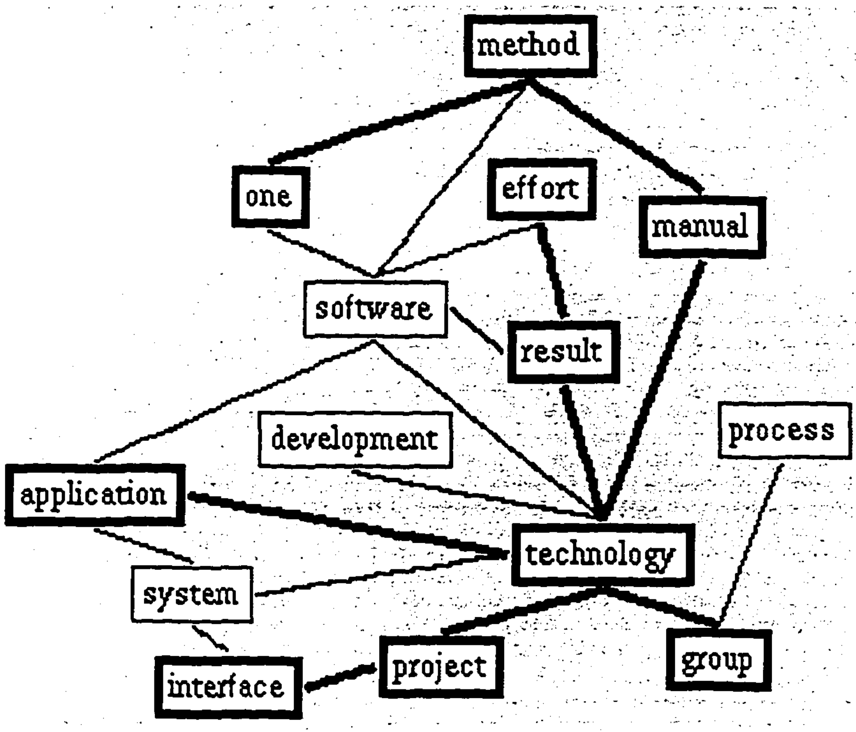

Figure A. 1-23: Network Number 9 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

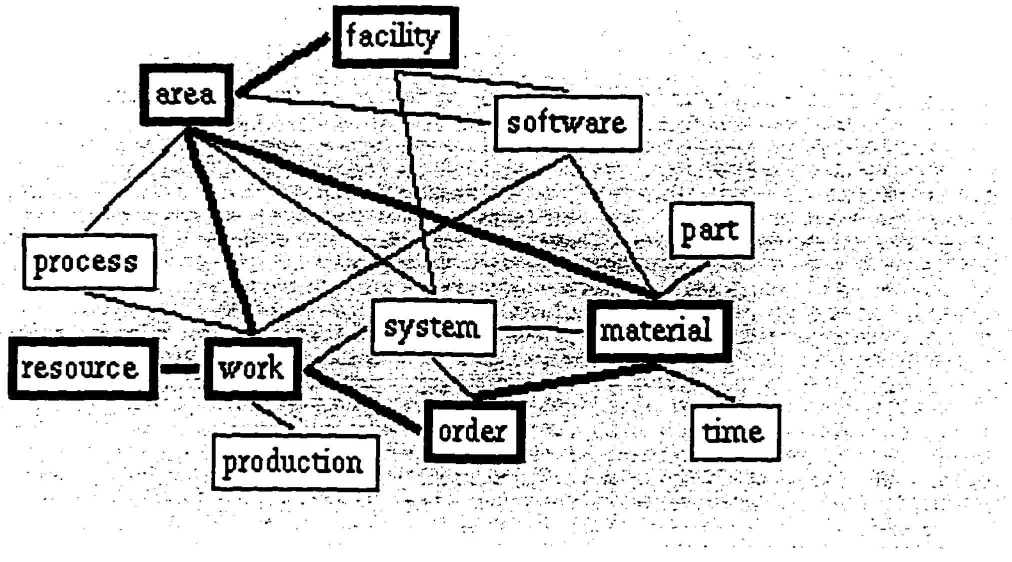

Figure A. 1-24: Network Number 10 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

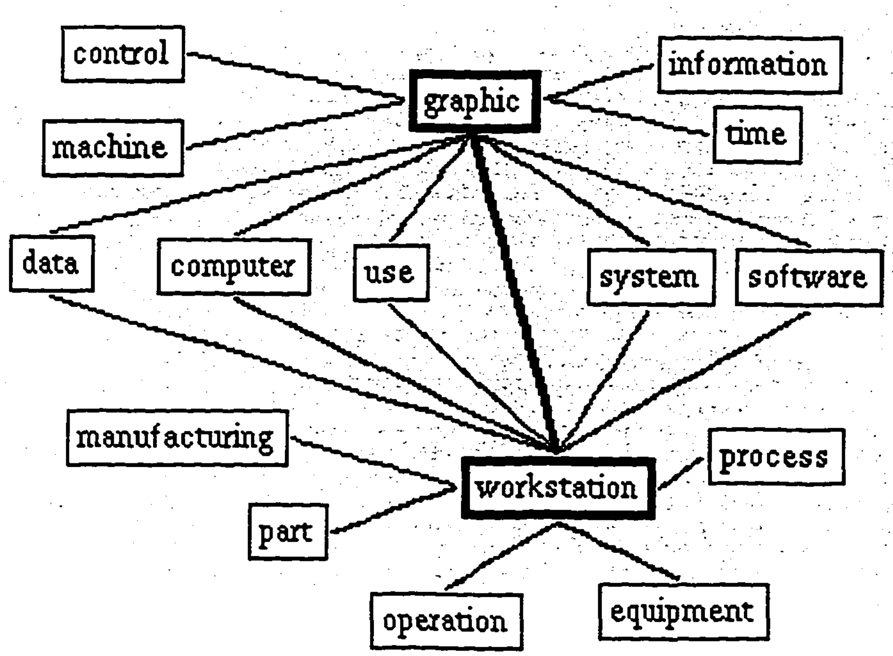

Figure A. 1-25: Network Number 11 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

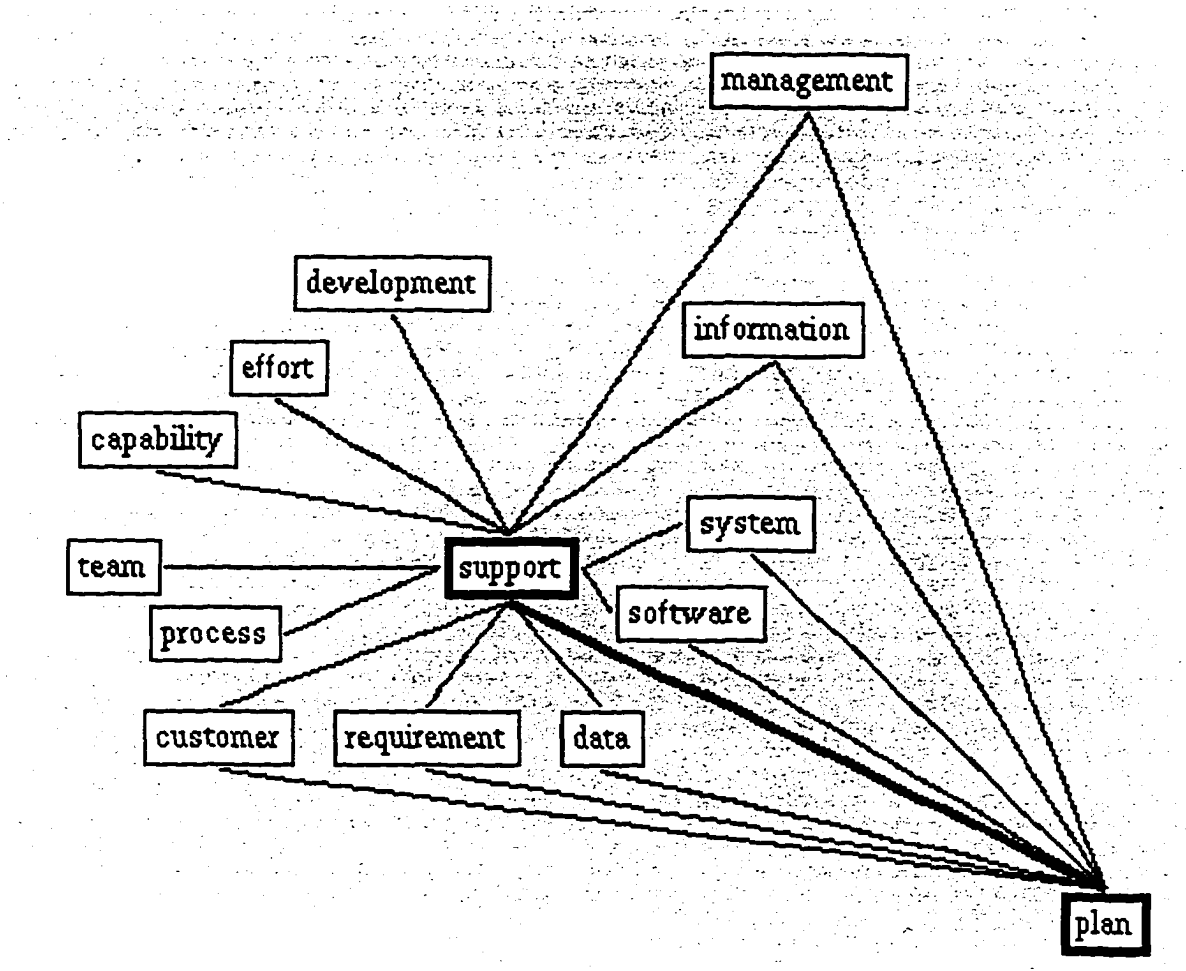

Figure A. 1-26: Network Number 12 Generated by Only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

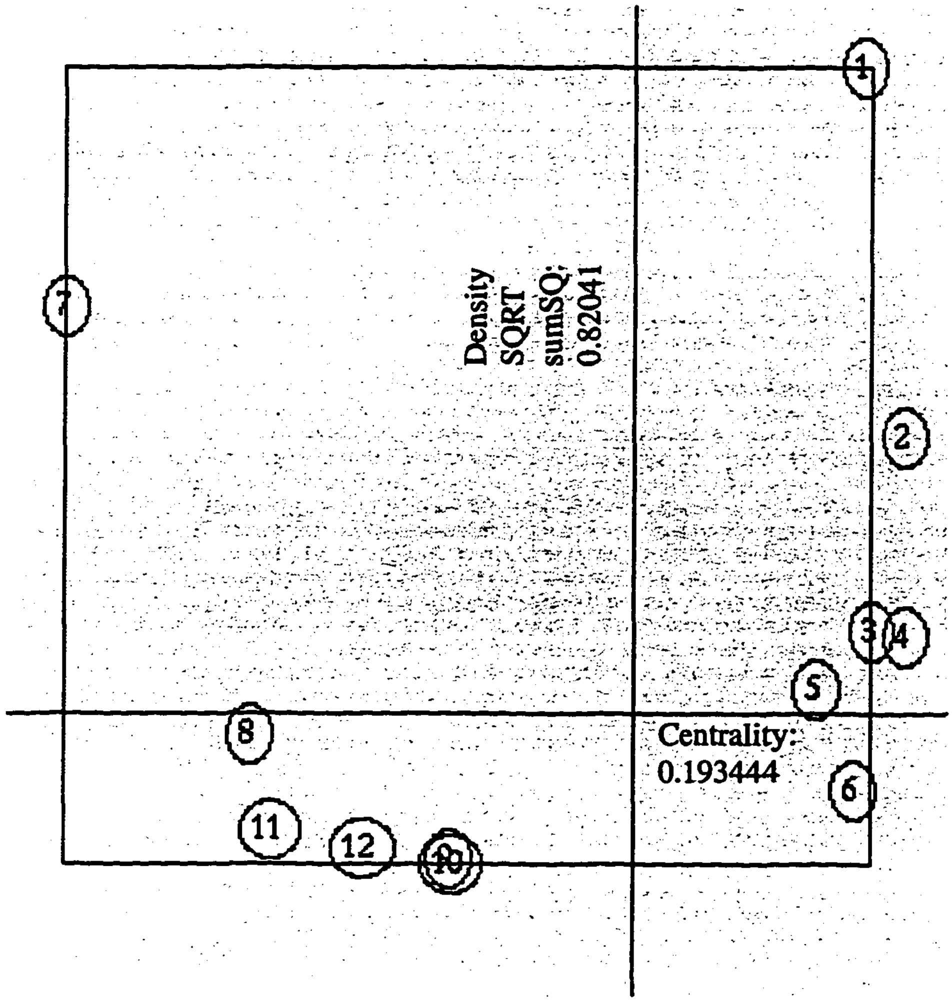

Figure A. 1-27: Density and Centrality Map for the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

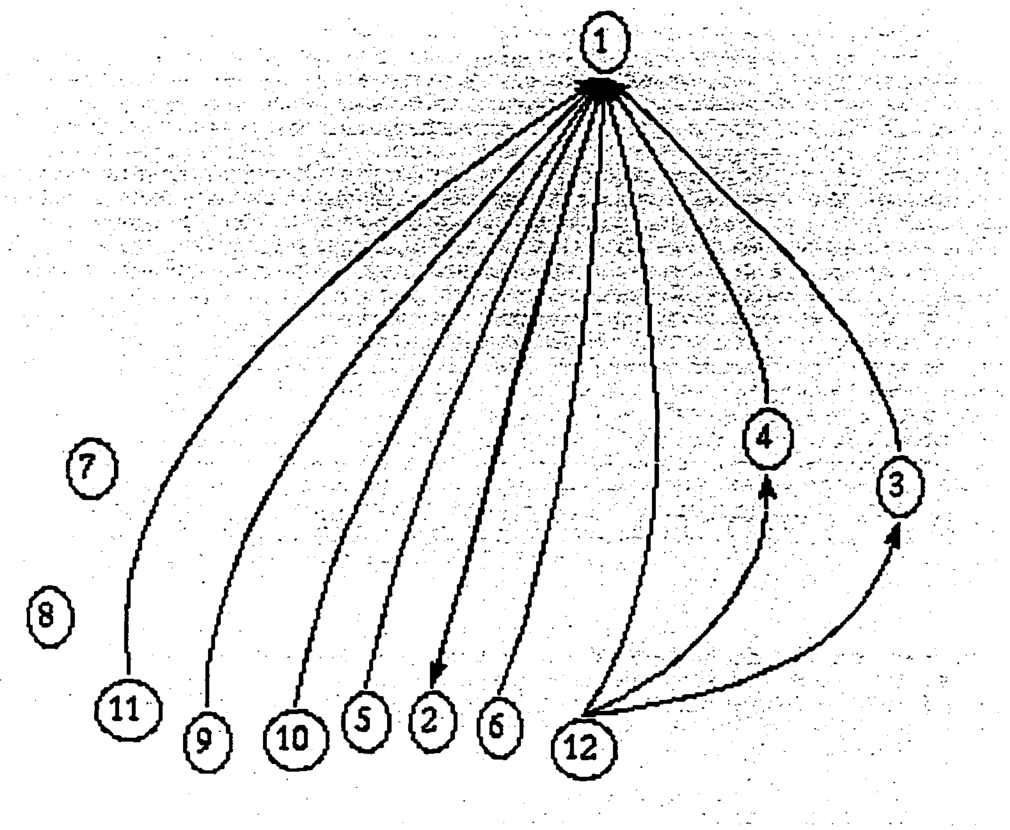

Figure A. 1-28: Super Network for only the Documents With the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

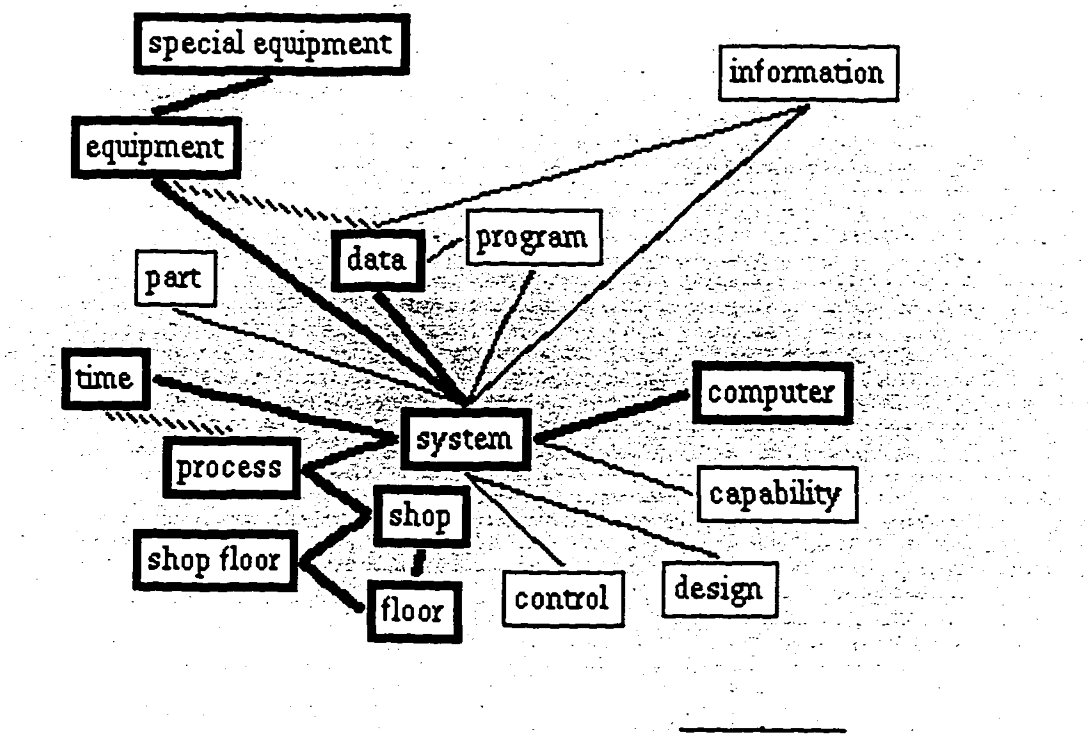

Figure A. 1-29: Network Number 1 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

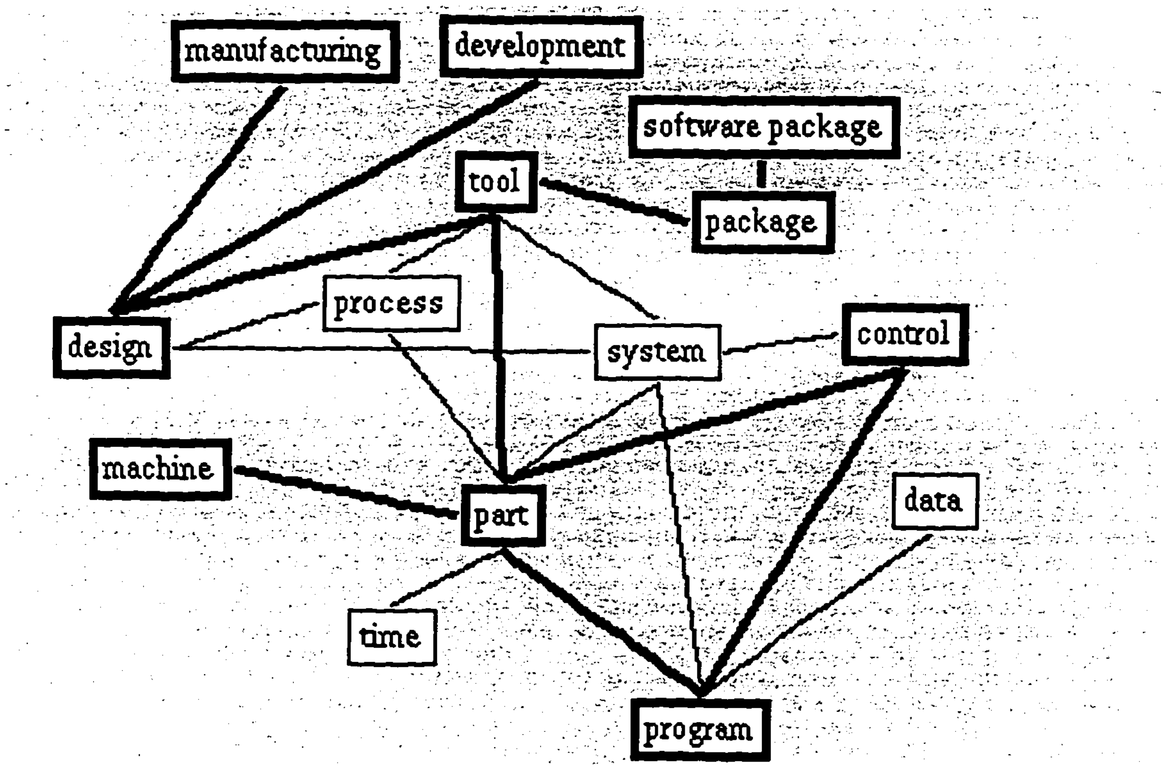

Figure A. 1-30: Network Number 2 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

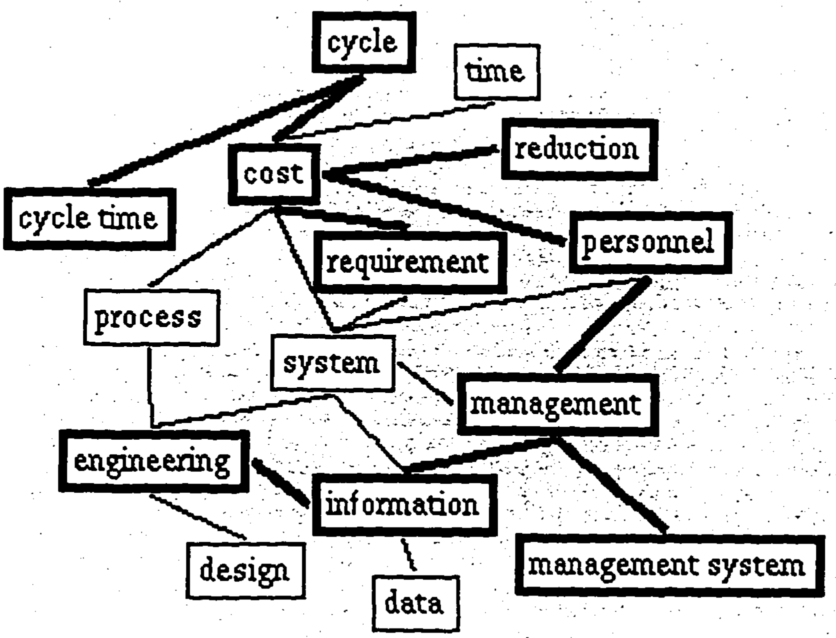

Figure A. 1-31: Network Number 3 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

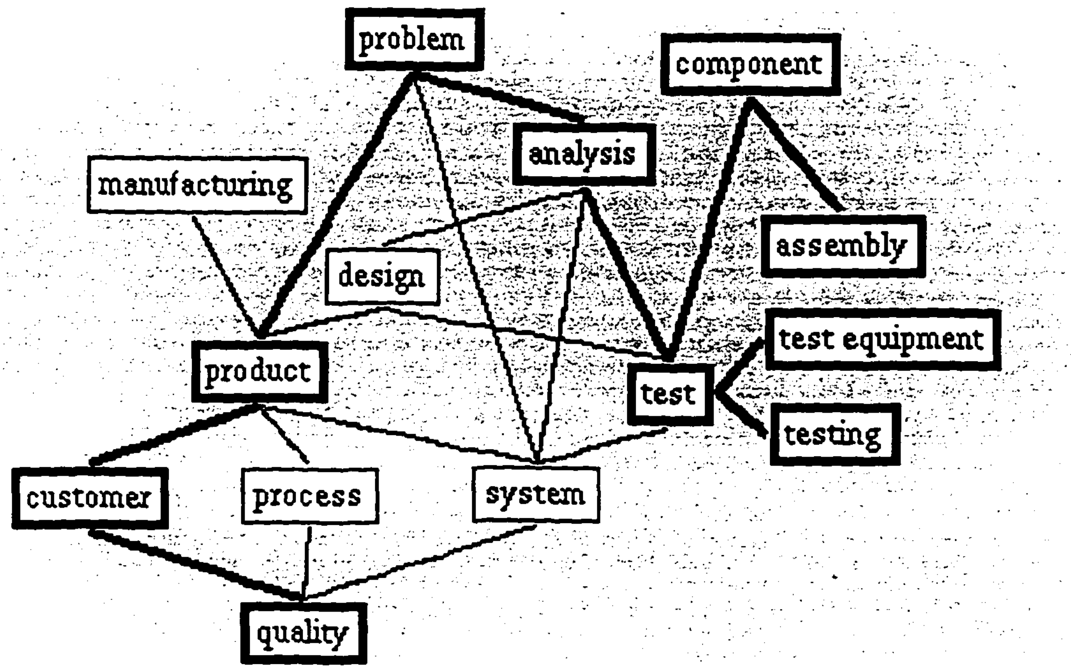

Figure A. 1-32: Network Number 4 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

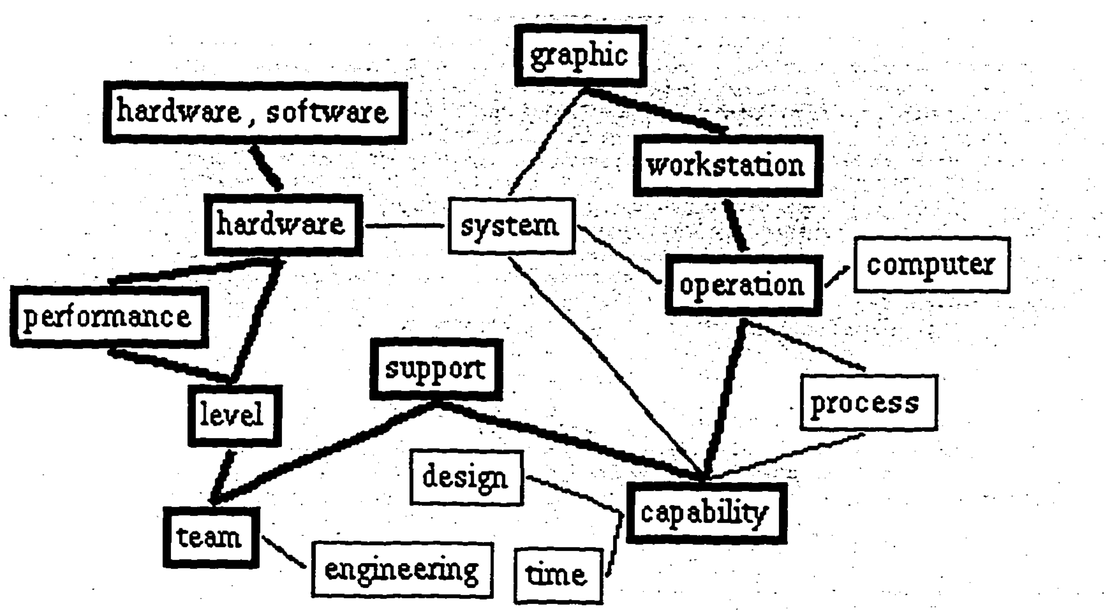

Figure A. 1-33: Network Number 5 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

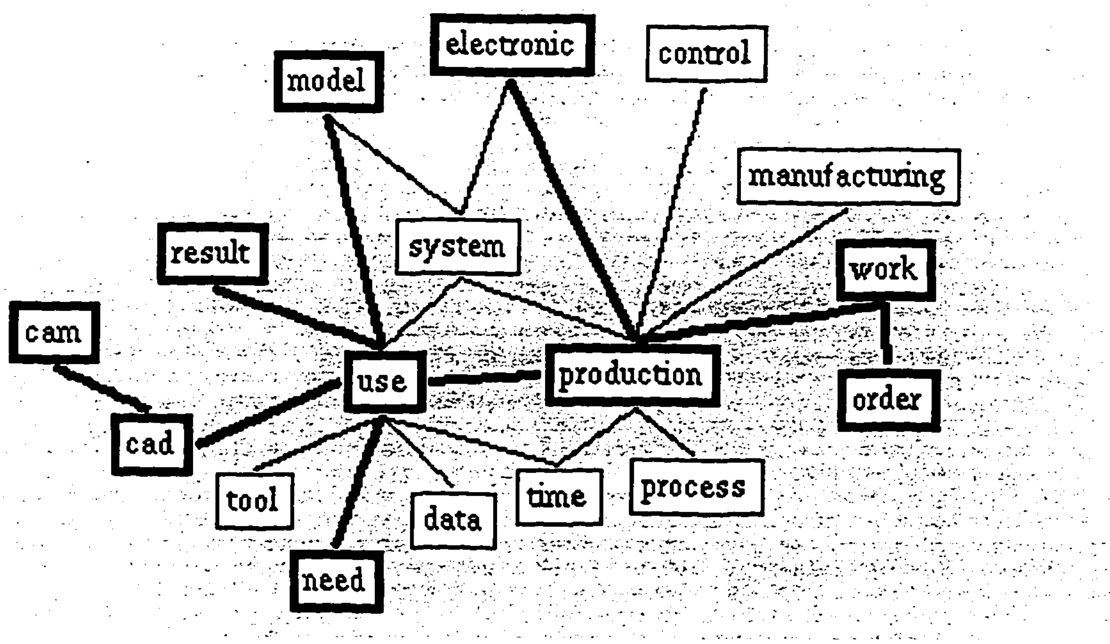

Figure A. 1-34: Network Number 6 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

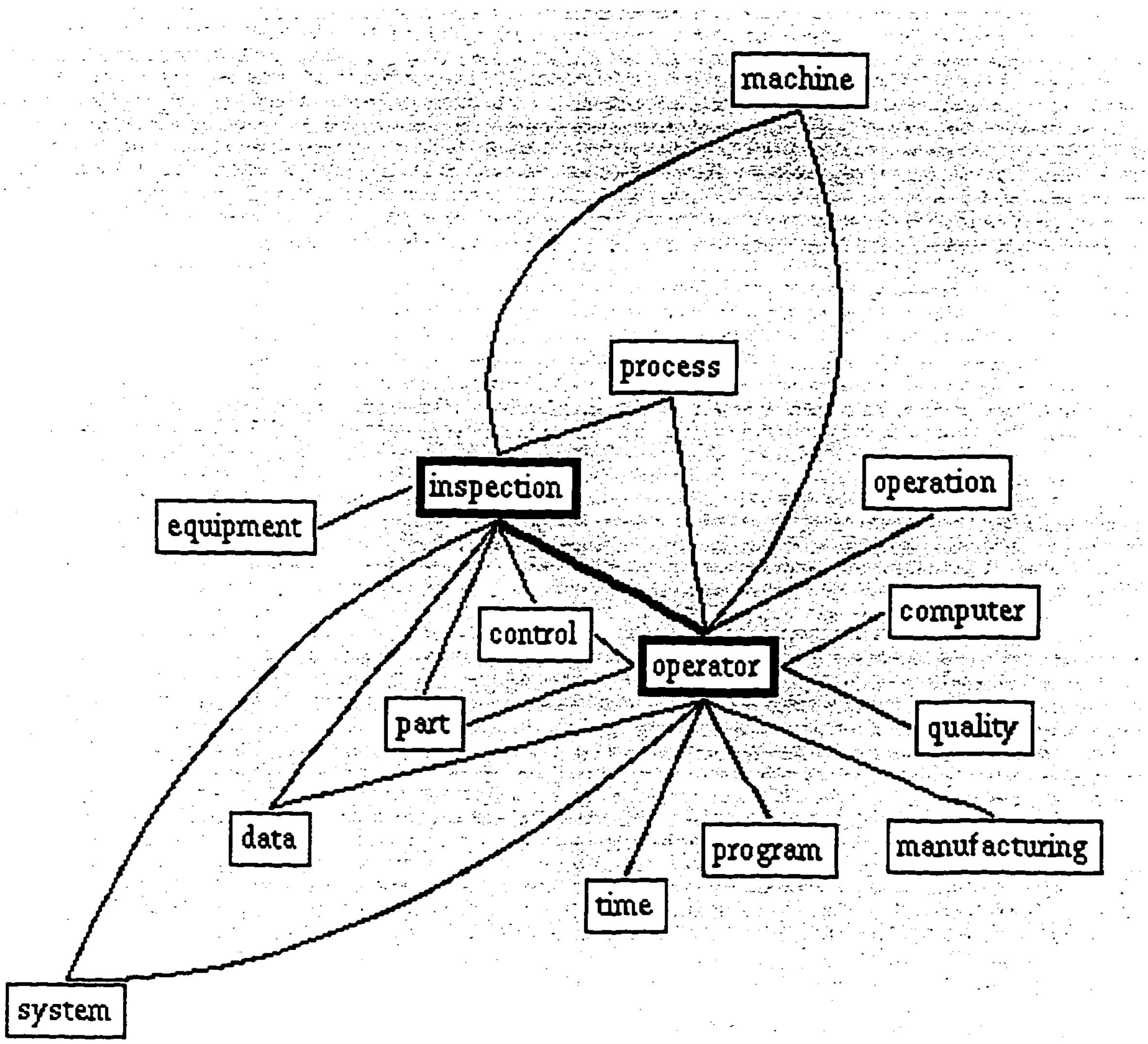

Figure A. 1-35: Network Number 7 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

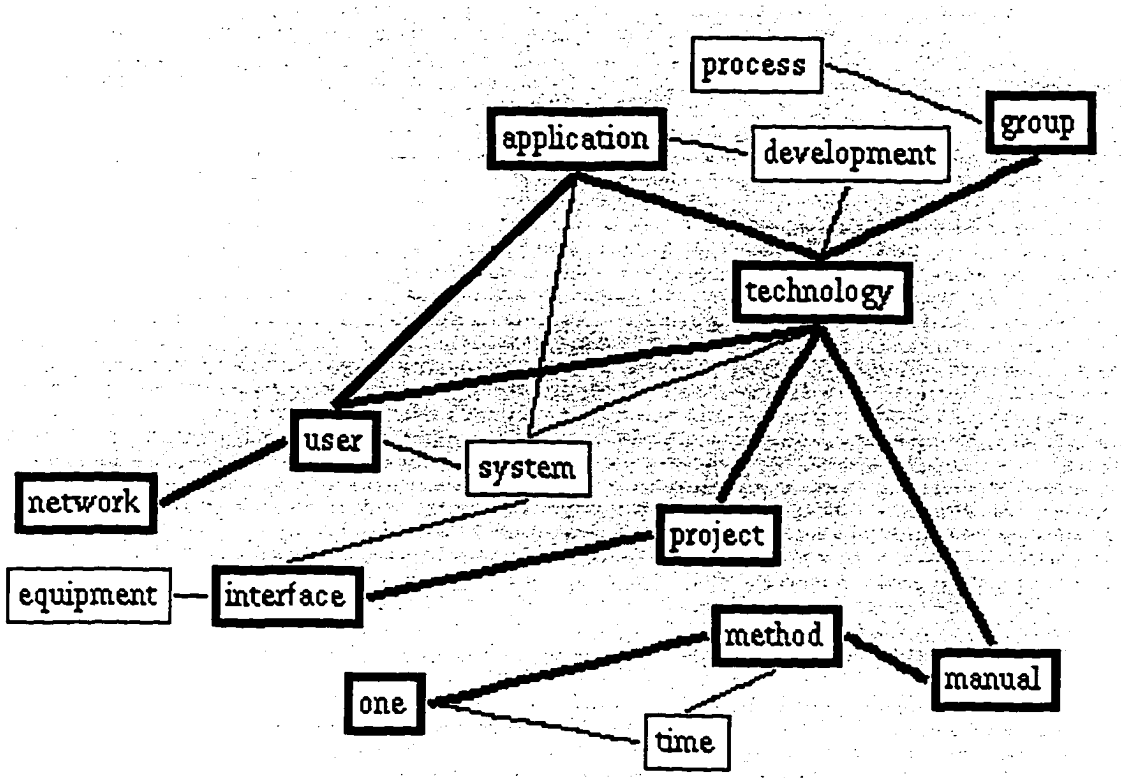

Figure A. 1-36: Network Number 8 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

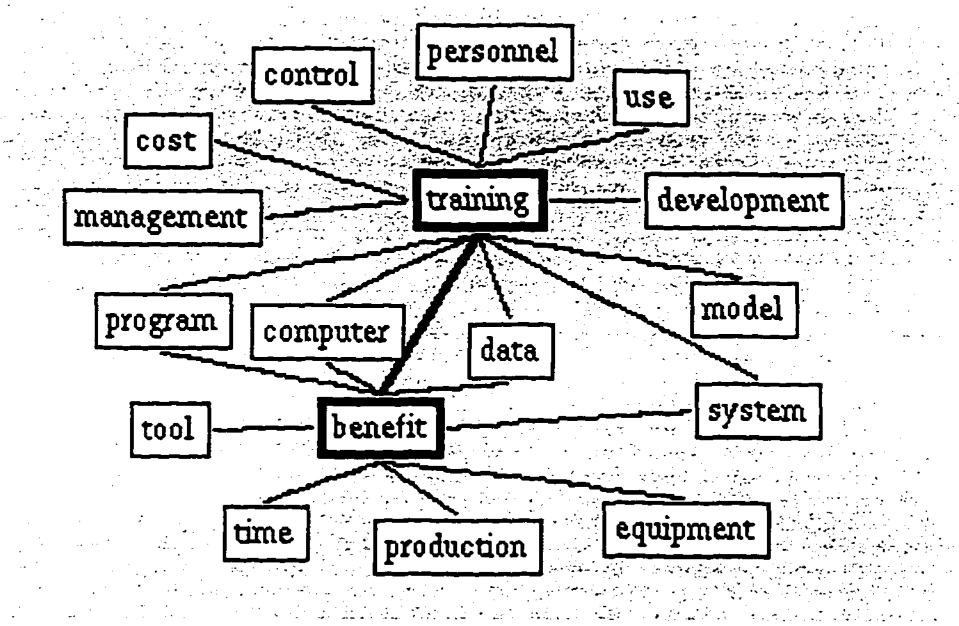

Figure A. 1-37: Network Number 9 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

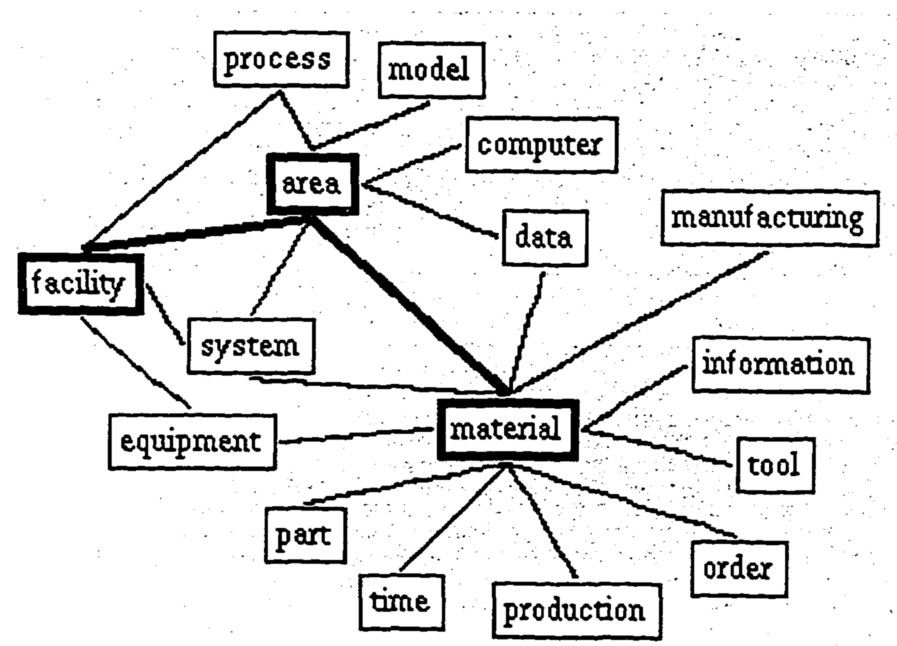

Figure A. 1-38: Network Number 10 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

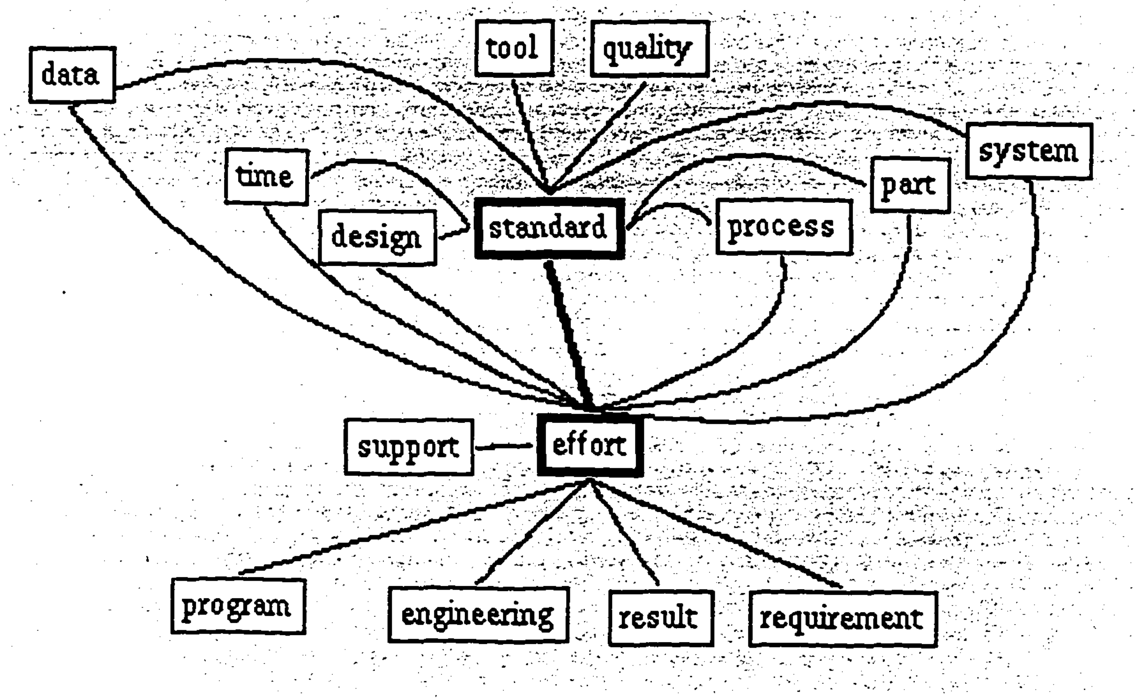

Figure A. 1-39: Network Number 11 Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

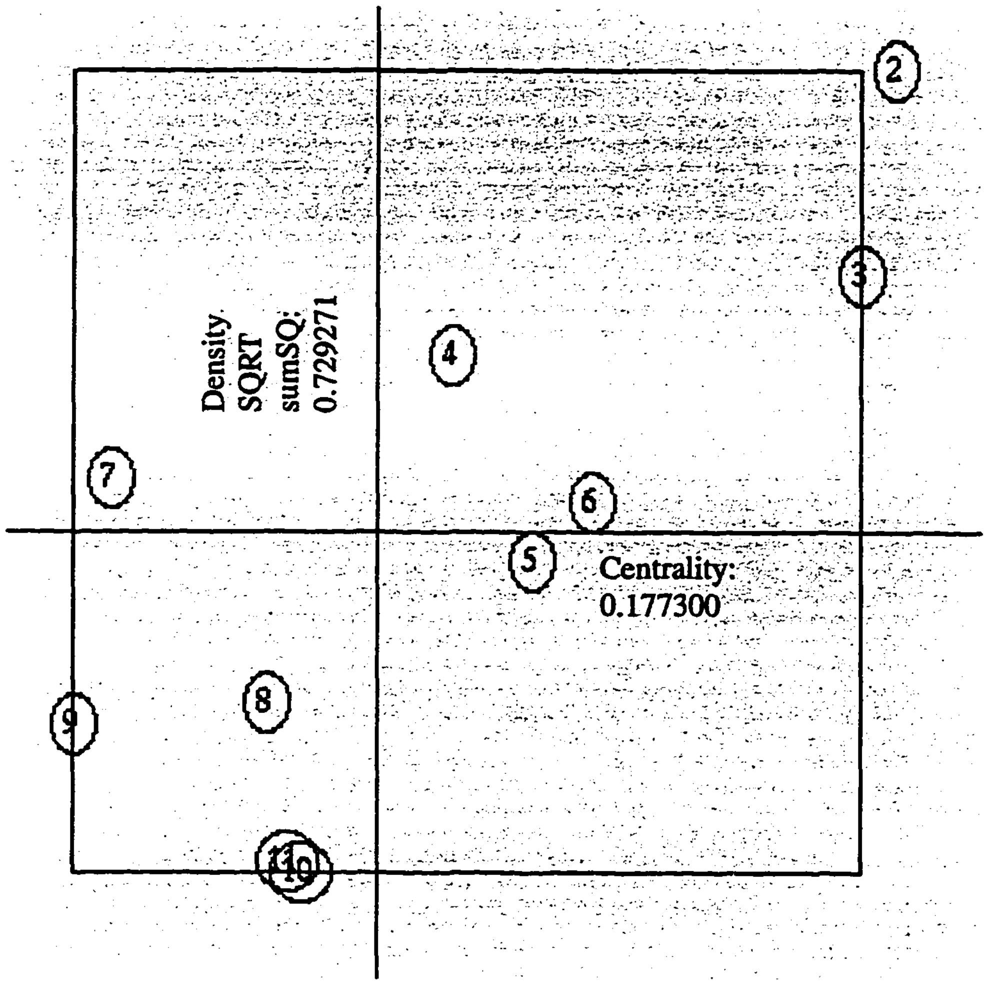

Figure A. 1-40: Density and Centrality Map Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

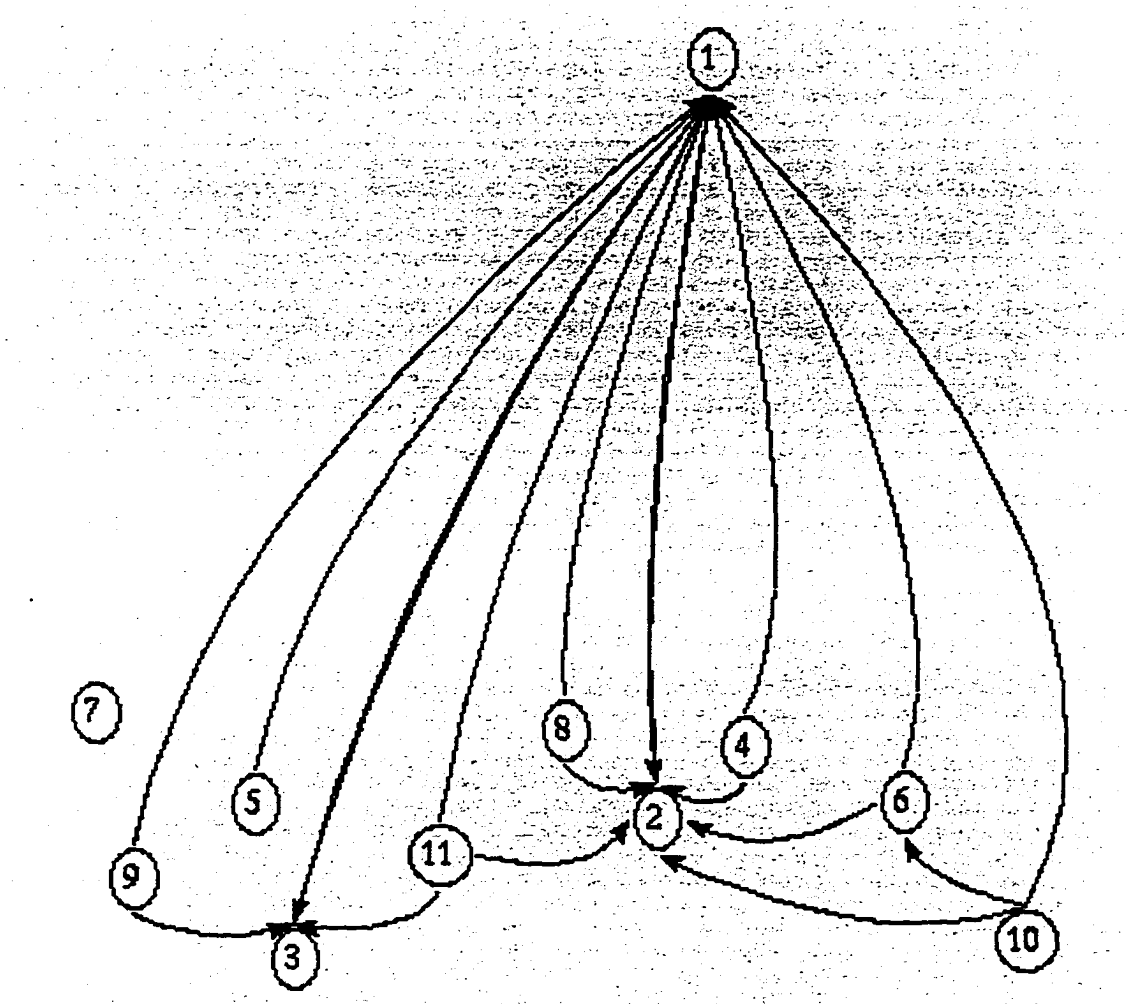

Figure A. 1-41: Super Network Generated by Excluding Only the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

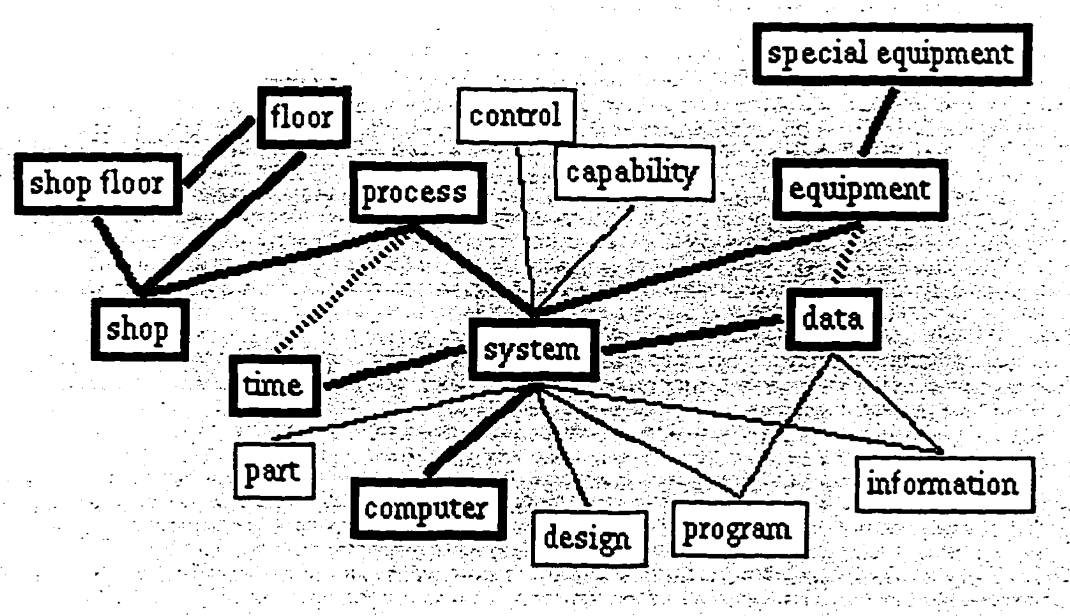

Figure A. 1-42: Network Number 1 Generated by Excluding All of the "Software" Related Phrases when the Cutoff for Co-Occurrence Level is 15 and Links are 10

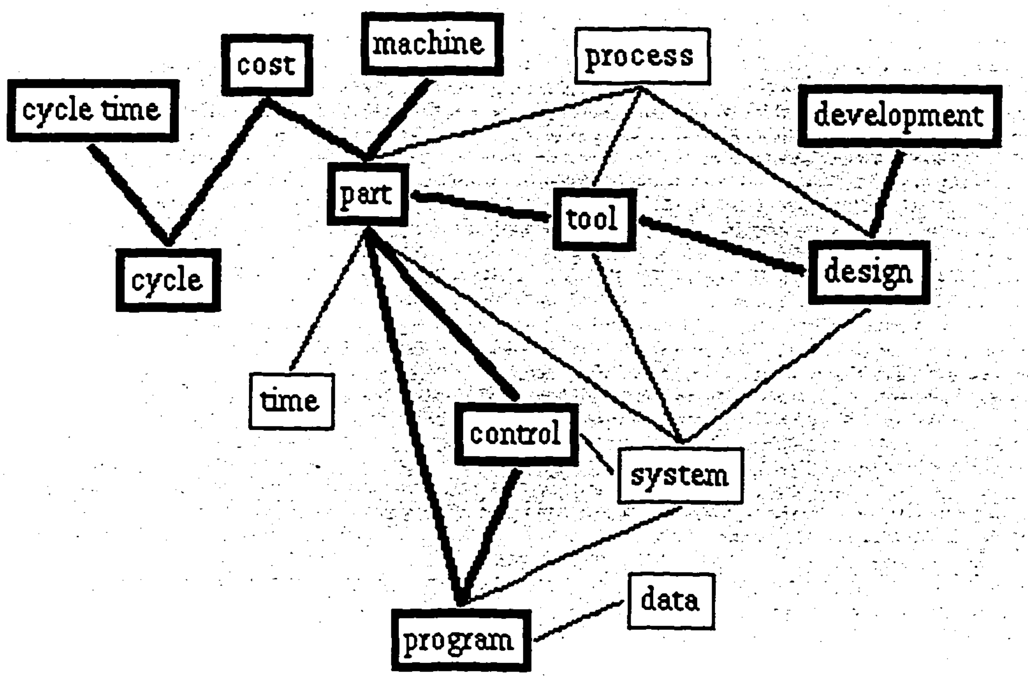

Figure A. 1-43: Network Number 2 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

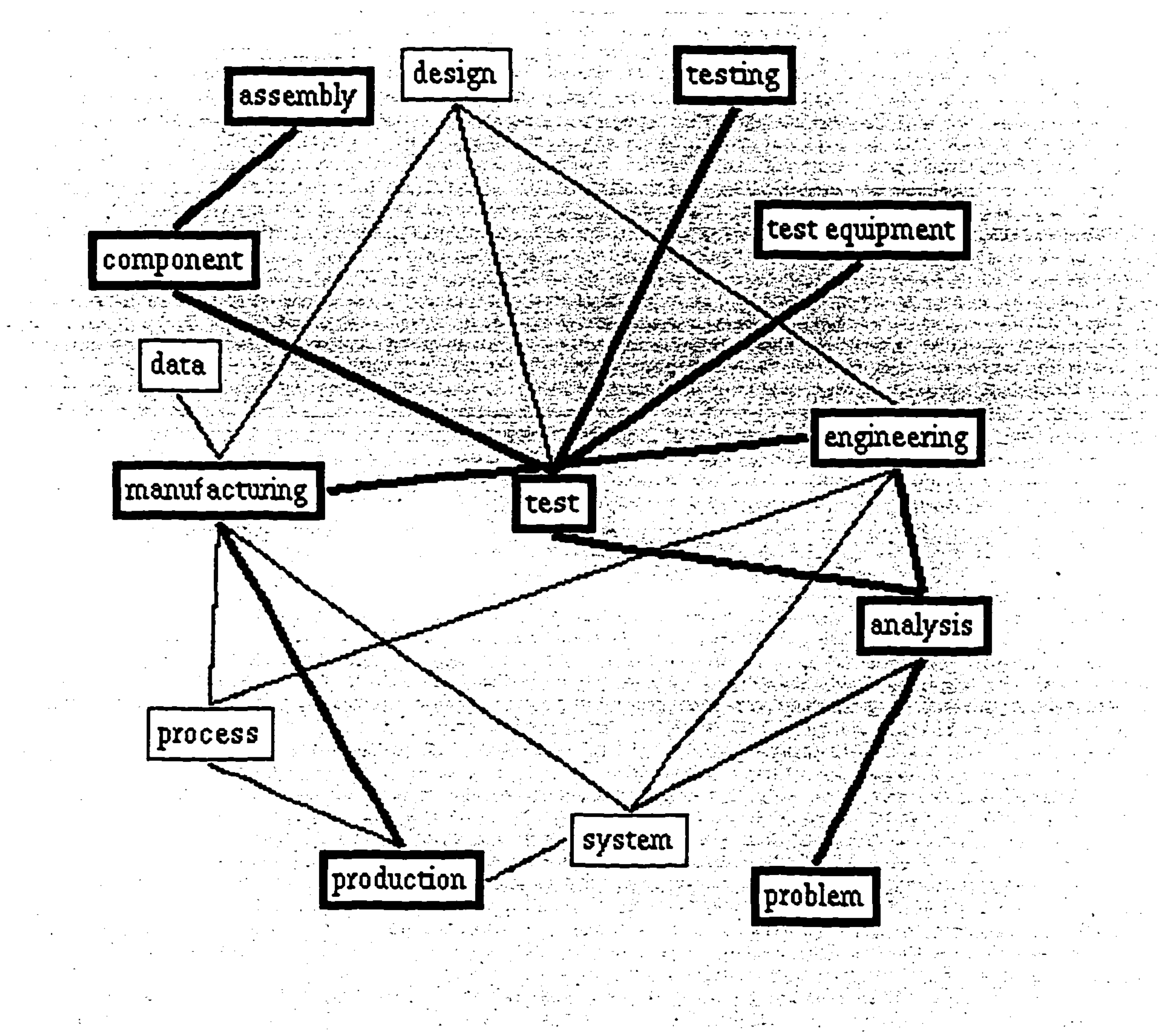

Figure A. 1-44: Network Number 3 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

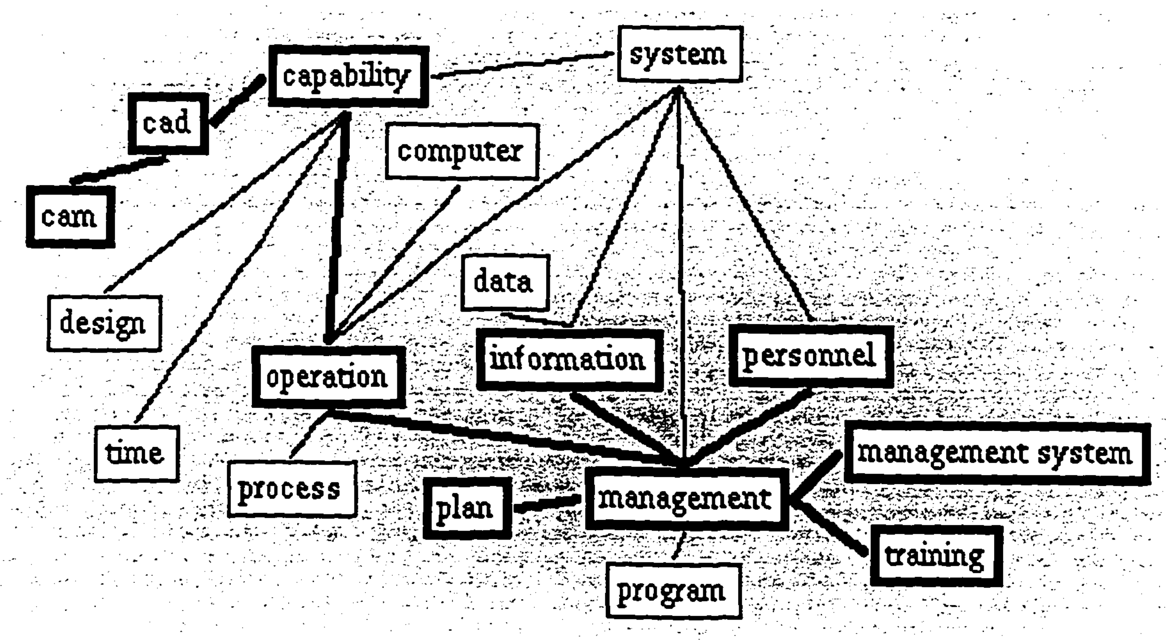

Figure A. 1-45: Network Number 4 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

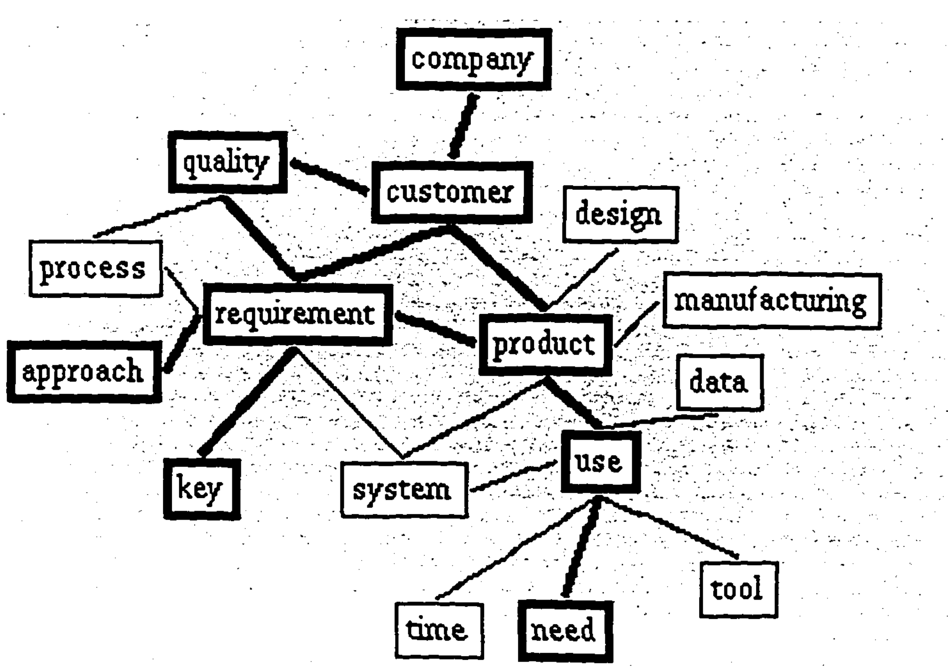

Figure A. 1-46: Network Number 5 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

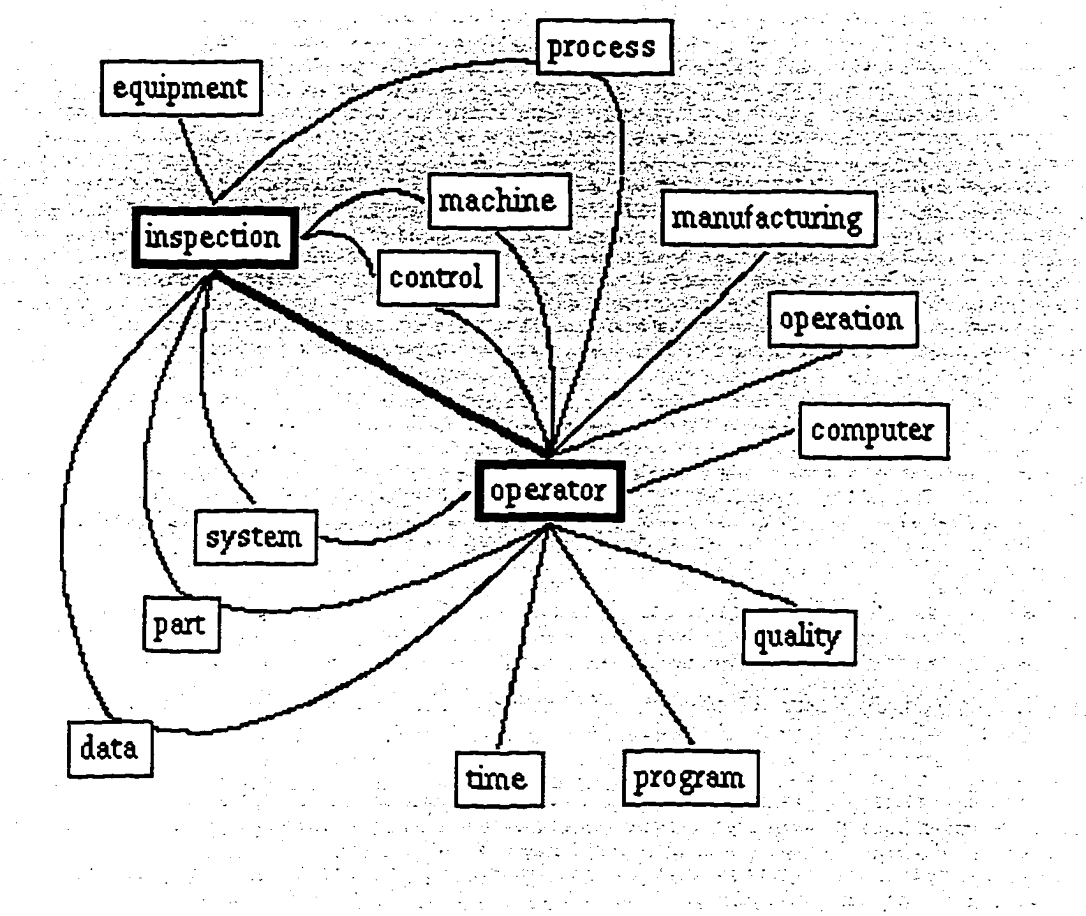

Figure A. 1-47: Network Number 6 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

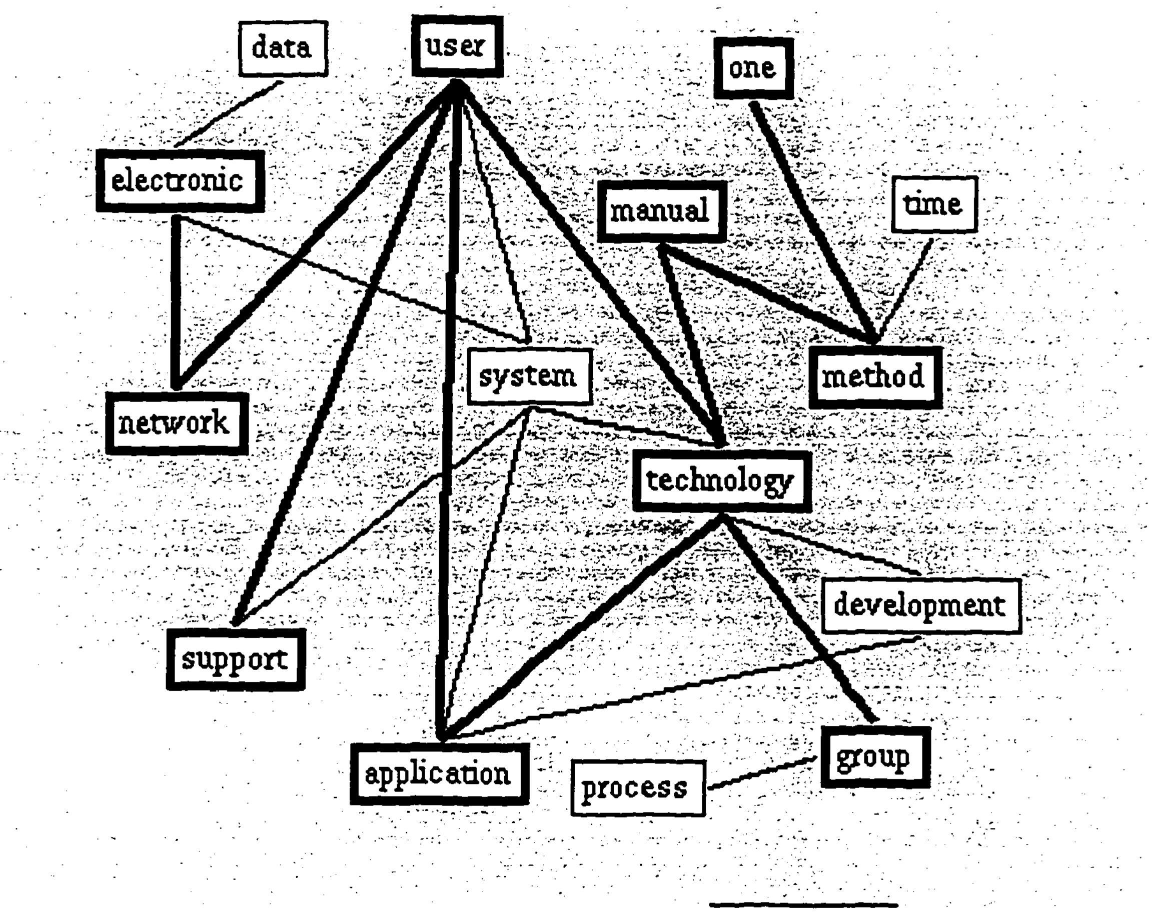

Figure A. 1-48: Network Number 7 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

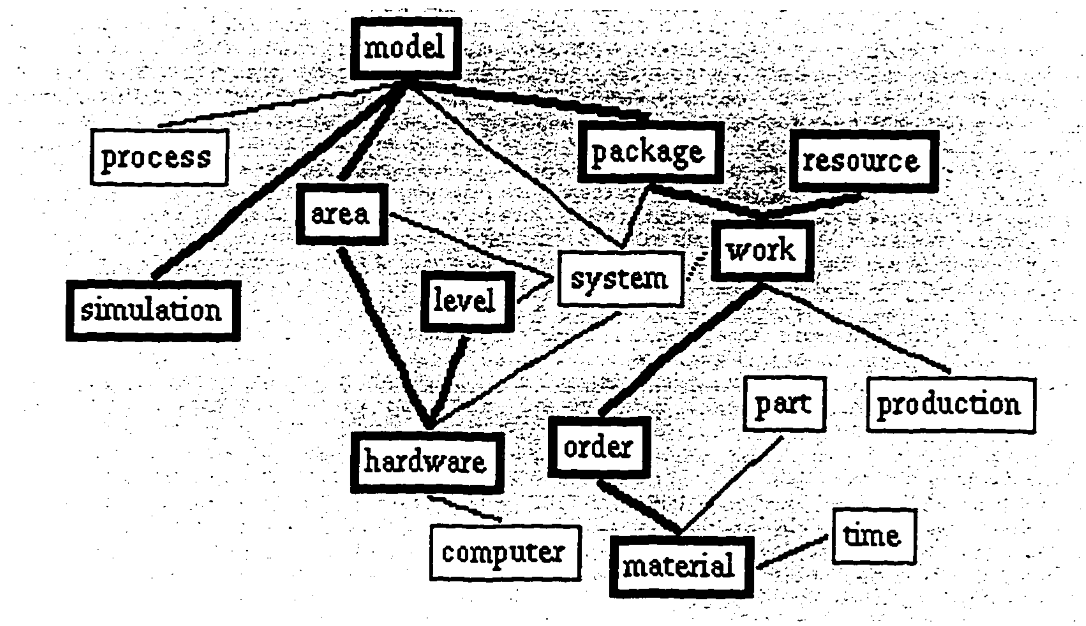

Figure A. 1-49: Network Number 8 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

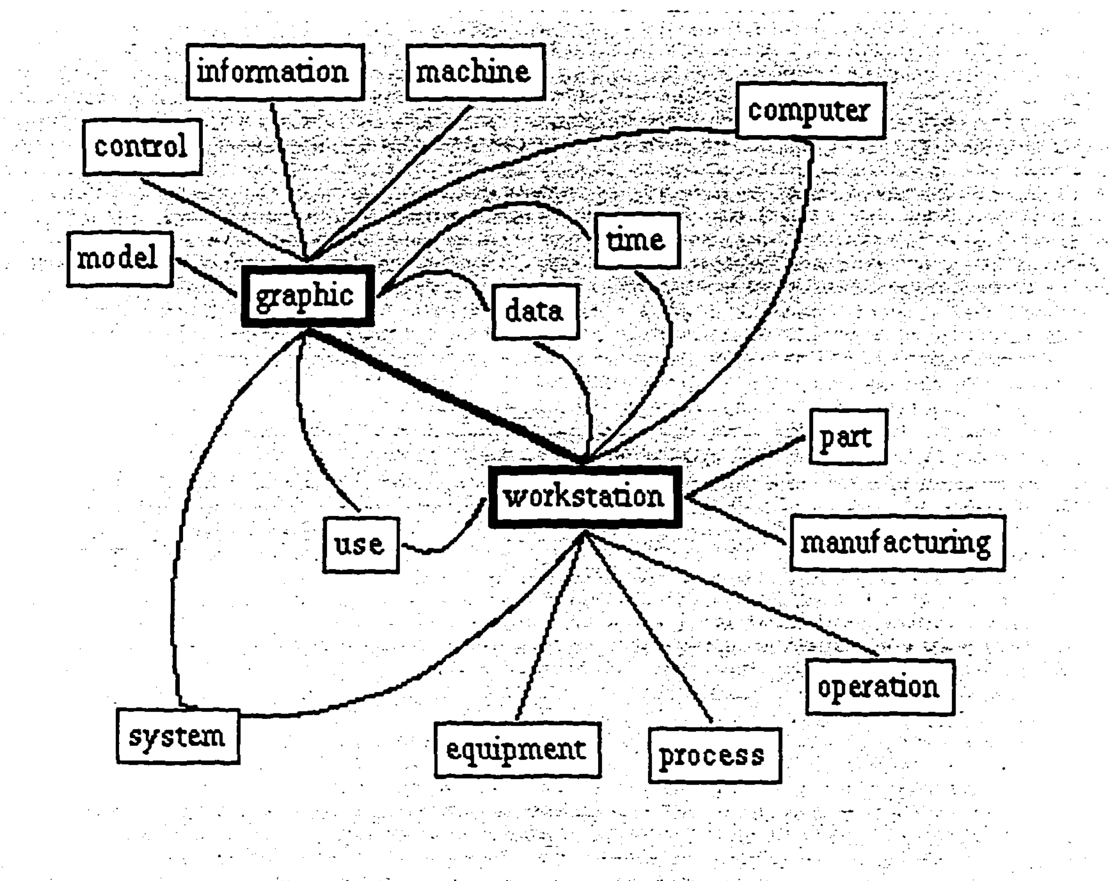

Figure A. 1-50: Network Number 9 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

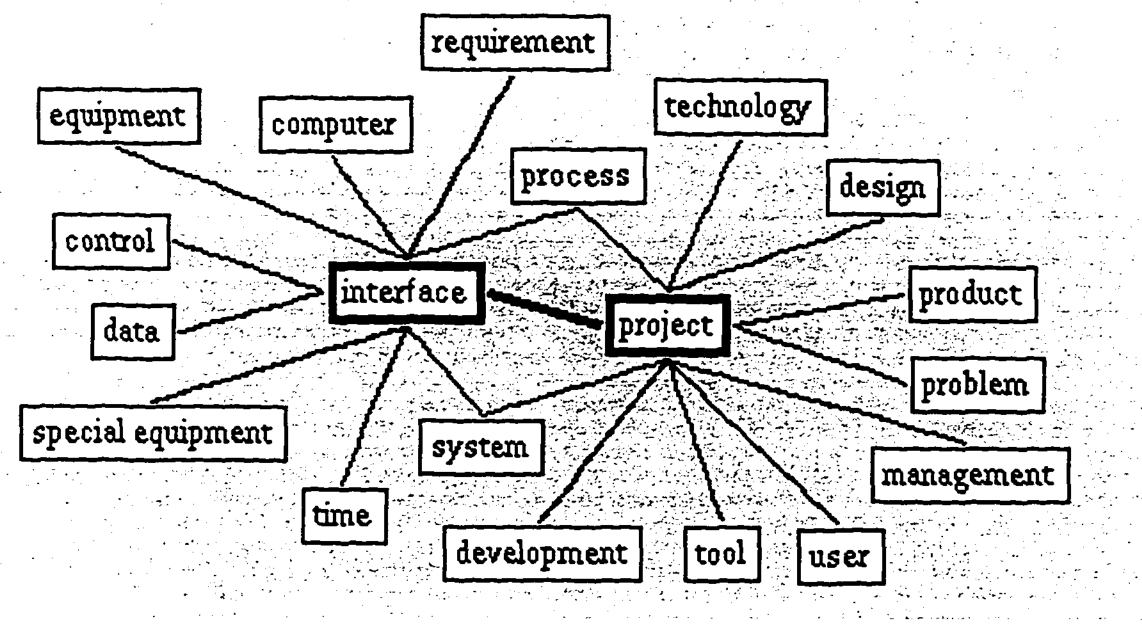

Figure A. 1-51: Network Number 10 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

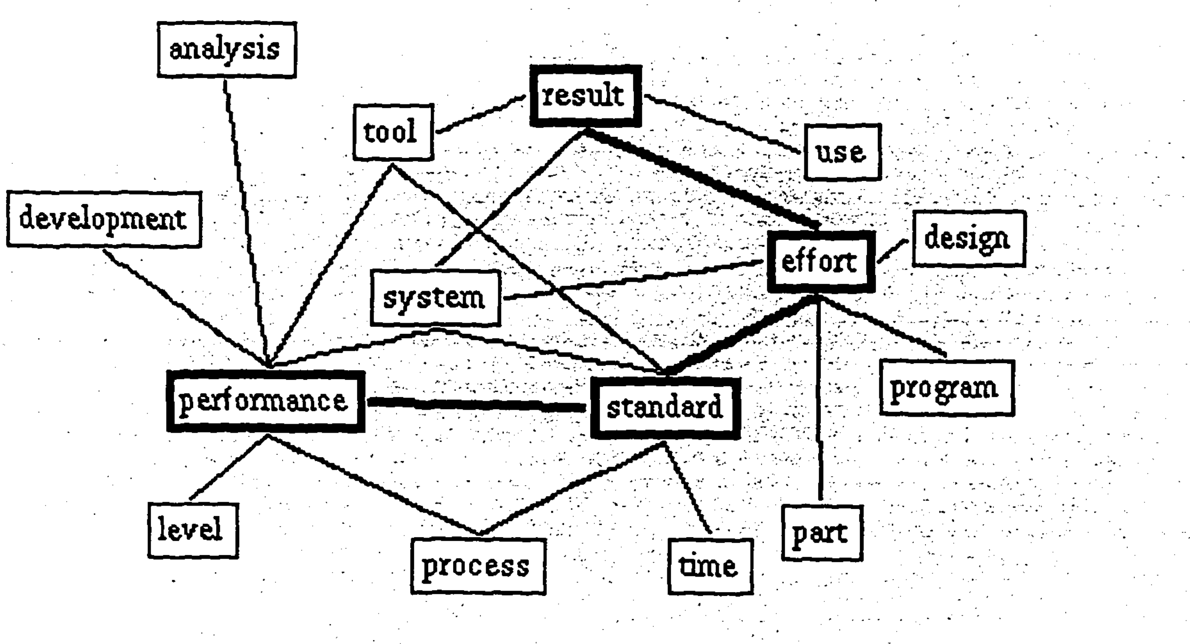

Figure A. 1-52: Network Number 11 Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

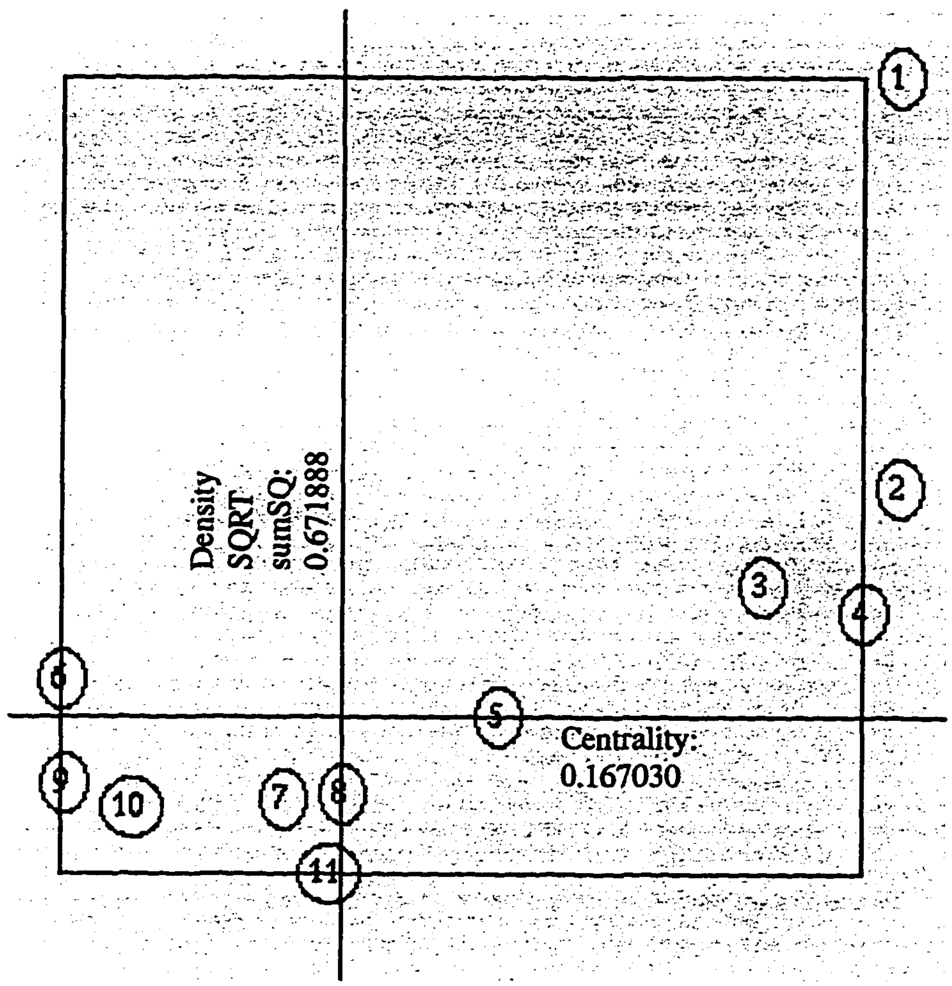

Figure A. 1-53: Density and Centrality Map Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

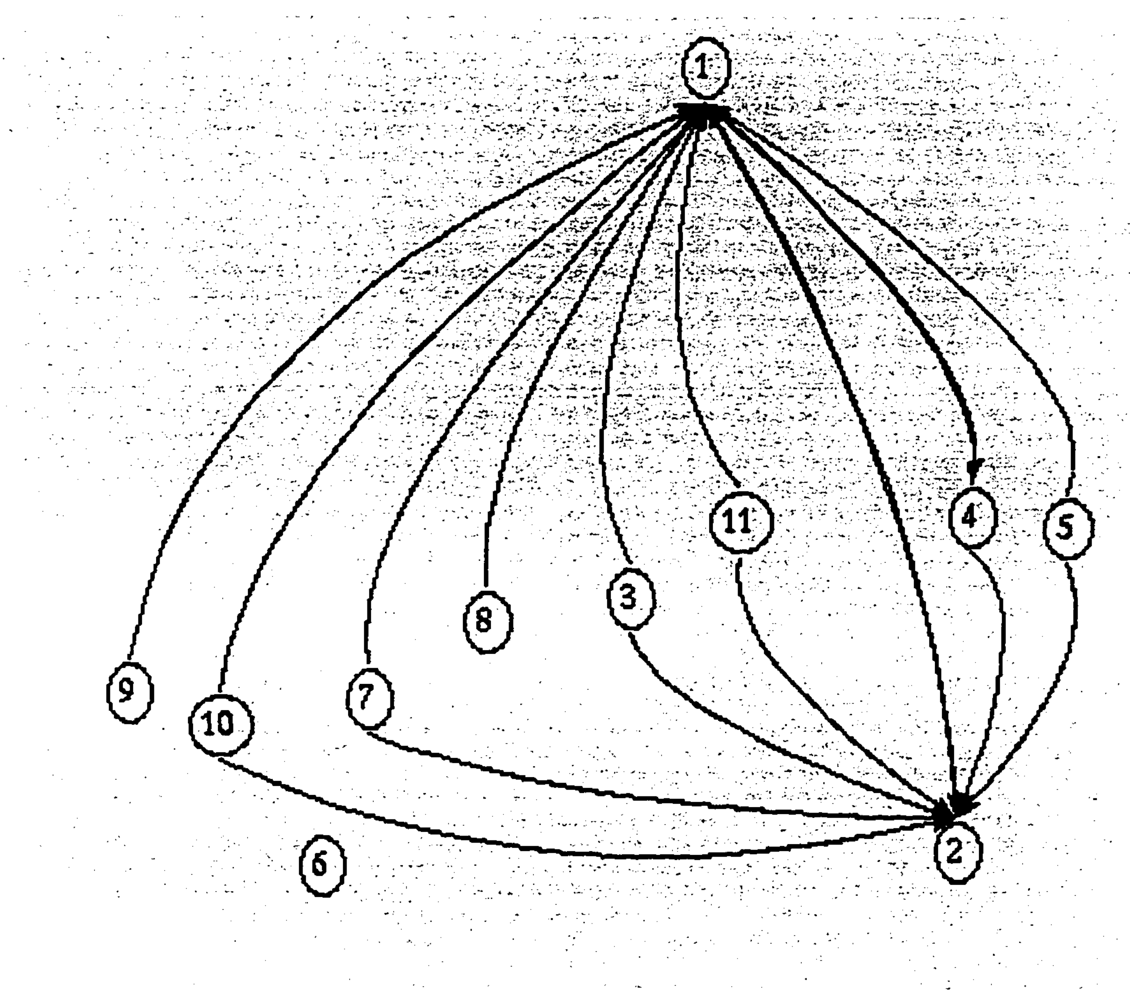

Figure A. 1-54: Super Network Generated by Excluding all of the Word "Software" when the Cutoff for Co-Occurrence Level is 15 and Links are 10

References

[CALL86] Callon M., Law J., and Rip A. "Qualitative scientometrics." *Mapping of the Dynamics of Science and Technology*, London: McMillian, 1986.

[CALL91] Callon M., Courtial, et al. "Co-word Analysis as a Tool for describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemistry." *Scientometrics 22*, 1 (January 1991): 153-203.

[COUL96] Coulter N., Suresh, et al. "An Evolutionary Perspective of Software Engineering Research Through Co-Word Analysis." *Technical Report CMU/SEI-95-TR-019 ESC-TR-95-019*, Software Engineering institute, Carnegie Mellon University, Pennsylvania.

[COUR89] Courtial, Law J., et al. "A Co-Word Study of Artificial Intelligence." 301-311. *Social Studies in Science*, London: SAGE, 1989.

[GRAD92] Grady, Robert B. *Practical Software Metrics for Project Management and Process Improvement*, Prentice-Hall, 1992.

[INTE1] Best Manufacturing Practices Home Page. (http://www.bmpcoe.org)

[LAW92] Law, Whittaker J. "Mapping Acidification Research: A test of the Co-Word Method." *Scientometrics 23*, 3 (1992): 417-461.

[OLSE95] Olsen N. "Survival of the Fastest: Improving Service Velocity," *IEEE Software*, September 1995, pp 28-38.

[PLOS91] Plossl G. *Managing in the New World of Manufacturing*, Prentice Hall, 1991.

[PRES92] Pressman R. *Software Engineering A Practitioner's Approach.* McGraw-Hill: New York, 1992.

SOMM85] Sommerville I. Software Engineering. Addison-Wesley: Massachusetts, 1985.

[STEV94] Stevenson S. et al. "The Use of Best Design Practices: An Analysis of US Navy Contractors," *Research in Engineering Design* (1994) 6: 14-24.

[TAYL59] Taylor E. *An Interim Report on Engineering Design*, Massachusetts Institute of Technology, 1959.

[WHIT89] Whittaker J. "Creativity and Conformity in Science: Titles, Keywords, and Co-Word Analysis." *Social Science in Science*, London: SAGE, 1989.

[WILL95] Wilson R. *UNIX Test Tools and Benchmarks*, Prentice Hall, 1995.

[WILL97] Wilson R. *Software Secrets of Engineering Quality Software*, Prentice Hall, 1997.

[YOUN89] Youngblut, Christine. "SDS Software Testing and Evaluation: A Review of the State of the Art in Software Testing and Evaluation with Recommended R&D Tasks." *Institute for Defense Analysis*, IDA Paper P. 2132, February 1989.