VITA

Nicole M. Radziwill

EDUCATION

Ph.D. Indiana State University, Terre Haute, IN. Technology Management, 2009.
MBA Regis University, Denver, CO, 2004.
B.S. Pennsylvania State University, University Park, PA. Meteorology, 1995.

PROFESSIONAL EXPERIENCE

Assistant Director, National Radio Astronomy Observatory (2006 - 2009).
Division Head, National Radio Astronomy Observatory (2002 – 2006).
Engagement Manager, Nortel Networks/Senior Solutions Architect, Clarify (2000-2001).
Software Development Manager, Rhythms Netconnections (1999).
Systems Analyst, National Oceanic and Atmospheric Administration (1997-1998).
Technical Project Manager, Quintet International (1995-1998).

SELECTED PUBLICATIONS

Radziwill, N. M. and R. F. DuPlain, 2009: Quality and Continuous Improvement in Knowledge Management. Chapter in *Knowledge Management Strategies for Business Development*, IGI Global.

Radziwill**,** N. M., Olson, D., Vollmar, A., Lippert, T., Mattis, T., Van Dewark, K. and J.W. Sinn, 2008: Starting from Scratch. *Quality Progress,* September, p. 40-47.

Radziwill, N. M., and R.F. DuPlain, 2008: A Model for Business Innovation in the Web 2.0 World. In *Web 2.0 Business Models*, IGI Global.

Radziwill, N. M., 2008: Implementing Communities of Practice to Manage Knowledge and Drive Innovation. In *Knowledge management strategies: a handbook of applied technologies.* IGI Global.

Radziwill, N. M., 2006: Foundations for Quality Management of Scientific Data Products. *Quality Management Journal, v13 Issue 2* (April), p. 7-21.

Radziwill, N. M., 2005: An Ethical Theory for the Advancement of Professionalism in Software Engineering. *Software Quality Professional, v7 Issue 3 (June),* p. 14-23.

SERVICE

Examiner, Malcolm Baldrige National Quality Award (2009)
Chair-Elect, American Society for Quality (ASQ) Software Division
Editorial Reviewer, *Journal of Industrial Technology* and *Software Quality Professional*

BLANK

TOPOLOGY, EVOLUTION, AND NETWORK-BASED CONTINUOUS IMPROVEMENT OF THE QUALITY MANAGEMENT JOURNAL

A Dissertation

Presented to

The School of Graduate Studies

College of Technology

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Nicole M. Radziwill

May 2009

School of Graduate Studies
Indiana State University
Terre Haute, Indiana

CERTIFICATE OF APPROVAL

DOCTORAL DISSERTATION

This is to certify that the Doctoral Dissertation of

Nicole M. Radziwill

entitled

Topology, Evolution, and Network-Based Continuous
Improvement of the Quality Management Journal

has been approved by the Examining Committee for the dissertation requirement for the

Doctor of Philosophy degree in

Technology Management (Quality Systems)

May 2009

M. Affan Badar, Ph.D. Date
Member
Indiana State University

Ji Yao Shen, Ph.D. Date
Member
NC A&T State University

Suhansa Rodchua, Ph.D. Date
Member
Central Missouri State University

Todd A. Waggoner, Ph.D. Date
Member
Bowling Green State University

John W. Sinn, Ph.D. Date
Chair
Bowling Green State University

Jay D. Gatrell, Ph.D. Date
Dean, School of Graduate Studies
Indiana State University

ABSTRACT

Knowledge flows are ubiquitous in research and industrial practice, and sharing knowledge is essential for both innovation and solving practical problems. It is through knowledge flows that research results are translated into practice to benefit industry. In quality management, one of the primary mechanisms to facilitate these knowledge flows between executives, researchers, professionals and managers is the *Quality Management Journal (QMJ),* published quarterly by the American Society for Quality (ASQ) since October 1993. The purpose of this study was to model all of the knowledge flows captured to date by the *QMJ* as a citation network, use the model to survey the research in quality management, and determine an appropriate measure (*attachment-weighted quality*) for continually improving the collection of research. Citation network analysis was chosen as the predominant methodology to avoid the biases that are inherent in citation counting methods. The citation network model was validated by assessing the goodness of fit of the empirical data with characteristic network topologies and idealized models of growth patterns in networks. The network was then used to objectively determine the most central and influential articles, as well as the relationship between the quality of the citation network and its topological characteristics.

The results from this research demonstrate how to use the citation network as a measurement system to continually improve the quality of the *QMJ*. Continuous improvement is necessary to ensure that important research results are readily identified, and that the research within a discipline is effectively communicated and translated into practice. The research makes novel contributions to the fields of quality management and

knowledge management, and in addition to the study of complex systems using networks. First, the results provide academics and educators in quality management with a record of the most influential research and review articles that have shaped the discipline. The study provides knowledge managers, information technology managers and journal editors with a measurement process that can be used to effect continuous improvement of an organizational knowledge base. The capability to assess the quality of individual articles within the context of a citation network also has potential practical implications in ranking search results, which can allow researchers easier access to promising new journal articles that have not yet become popular.

Finally, this is the first empirical study to examine the genesis of a real citation network, to predict the quality of a citation network from its topological characteristics, and to explore the relationship between fluctuations in the perceived quality of the articles over time through measures of dynamic centrality. To accomplish these goals, the research introduces a windowing algorithm to measure preferential attachment in terms of quality and dynamic centrality on the systemic level, called *attachment-weighted quality*. It is demonstrated that for the *QMJ*, this measure can be effectively predicted by the change in nodes and edges over the time window and the average path length over the network. Using the mean quality of the network and attachment-weighted quality, this study illustrates that although specificity of action of preferential attachment decreases over time in the *QMJ* citation network, the attachment-weighted quality of the network increases.

ACKNOWLEDGMENTS

Writing this dissertation has been a long and exhausting process – complicated, no doubt, by the fact that I started the PhD program when my son, Alex, was three months old. He is now turning four. First and foremost, I'd like to thank him for all the hours that he spent entertaining himself while I was on my computer, and for the times he said "mommy, go do your research." That's about the best motivation you can get to stay focused and finish a project like this.

I would like to thank the Chair of my Dissertation Committee, Dr. John W. Sinn, for leading me through this process in such a supportive way. I thank Dr. M. Affan Badar, Dr. Ji Yao Shen, Dr. Suhansa Rodchua and Dr. Todd Waggoner for their contributions, their insightful questions, and their excellent teamwork which made this committee a pleasure to work with. I would also like to thank Dr. Michael Hayden, my Program Planning Chair, who helped me through the initial stages of the program and the process of preliminary exams, Dr. David Beach whose early comments on this manuscript helped shape it as well, and the Consortium staff for all they do.

I could not have made it this far without the help and moral support of the other students in this PhD program, especially Ted Mattis, Greg Anderson, Cori Zuppo, Jeff Cunion, Kevin Van Dewark, Diane Olson, Mark Chandler, and Randy DeMik. I am also thankful to Noel Wilson at ASQ Headquarters who provided me with data, thoughts, and insights that helped bring this to fruition, and Jill Pearson-Wood at the Juran Center who provided information about their resources.

My family has also been extremely supportive throughout this whole process. My husband, Todd DeBarba, has been a continuous source of positive thinking, enthusiasm, love, and confidence. He has also done more than his share of dishes and housework. I need to thank my mother, Mary Eisele Radziwill, who says the "P" in "PhD" actually belongs to her for all the diapers she's changed on my behalf while I've been off studying and dissertating these past few years. Her belief in my ability to get through this process was unwavering, and she has been my strongest advocate for visualizing success. I would also like to thank my in-laws, Joan and Al DeBarba – Al for putting way too much pressure on me to finish this, and Joan for telling him to back off every now and then (all with love, of course!) I also have to recognize Amtrak, because most of the final text was written while riding the California Zephyr, the Southwest Chief, or the Cardinal in early 2009.

Finally, I'd like to thank my father, Diethard Radziwill, for being so enthusiastic about my work in this program and on this dissertation. And thanks for letting me quit German School and go to the Math Club in 7th grade, because without that, I wouldn't have attended NCSSM where I learned about fractals and gained an appreciation for the concept of self-similarity that I revisited frequently in the course of completing this research. So, finally, to answer the question you started asking me in 1990 – "how are fractals?" – I guess I now finally have an answer almost 20 years later, in the form of a few pounds of PhD dissertation. They're doing pretty well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Chapter Overview

The purpose of this chapter is to provide a conceptual foundation for the proposed research. The study will model knowledge flows in quality management as a citation network, derived from the references of *Quality Management Journal* (*QMJ)* articles, and examine its topological characteristics and its evolution to determine how the discipline has evolved since the debut of the *QMJ* in October 1993. First, a description of the *QMJ* is presented in terms of its history and editorial goals. Next, the ontological relationship between knowledge flows, citation networks, and the Body of Knowledge (BOK) concept is described to explain the significance of the method of citation network analysis. Finally, the blueprint for the research itself is summarized: the problem statement, the research questions and hypotheses to be investigated, a justification of why the research is necessary, and the assumptions and limitations underlying the proposed methodology.

Background

The *QMJ* publishes approximately four research articles in each of its quarterly issues. The average acceptance rate for the journal has consistently hovered between 15-20% of submissions (ASQ, 2009) which is typical for a peer-reviewed academic journal n the social sciences. (Whitworth, 2008) The journal was initiated with a threefold purpose: to educate executives about the foundations of quality management, to provide researchers with an outlet for their work in quality management, and to provide professionals and managers a venue to keep track of trends in quality management. Despite this vision to serve as a journal for both researchers and practitioners, and a review process that includes academics and business managers, the original editor clearly established that the *QMJ* would be first and foremost a research journal. (Golomski, 1993) Early supporters for the journal agreed that “managerially oriented quality research” was the goal, and that the *QMJ* would “encourage a more substantive dialogue between academicians and business practitioners.” (Benson, 1993) This would rectify the imbalance in quality management literature, which until that point, had been heavily weighted towards practice with little emphasis on theory. Continuing its early policies, the *QMJ* accepts articles from each of the categories in ASQ’s quality management Body of Knowledge (BOK), which is included as Appendix B for reference.

For the *QMJ*, the Body of Knowledge plays a very important role, because it helps journal reviewers and editors identify which articles will be relevant for their readership. As a complement, the evolution of a BOK is observable by studying changes in citation patterns over time. (Zhuge, 2006) Important articles often have high incoming citation counts, but this measure is incomplete because not all citations can be considered equal. Citation patterns can thus be more comprehensively examined using graph theory, a branch of discrete mathematics, which makes it possible to characterize and examine large-scale relational structures using statistical methods. Using this formalism, *networks* consisting of *nodes* are connected to one another by *edges*, and the structure and dynamics of the network are studied in terms of metrics evaluated at both the local (node-based) and global (network-based) levels. (Newman, 2003) A citation network consists of nodes (e.g. journal articles, books) that cite references which were used to develop the content of those nodes; that is, if article A cites article B, then there is a directed citation relationship from A to B (as shown in Figure 1).

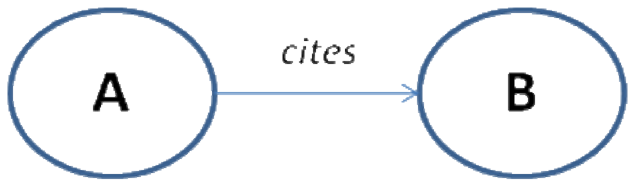


Figure 1. The Citation Relationship between Articles A and B in a Citation Network.

Modeling knowledge flows as a citation network provides the basis for objectively understanding the structure of a research discipline using the network topology. By taking snapshots of the network as new articles and new citation relationships are added over time, the dynamics of how a discipline forms, changes, and eventually defines a body of knowledge can be directly observed. This data presents many practical opportunities for learning how new ideas emerge within a discipline, for developing knowledge maps or personal research roadmaps, for determining the references that are most closely related to a particular article, for identifying peers with closely related interests, for objectively and dynamically evaluating the quality of research, for estimating the development stage or maturity of a discipline or a topic within a discipline, and for identifying emergent research areas. (Zhuge, 2006) The proposed study will explore what the topology of the citation network in quality management reveals about knowledge flows, how the evolution of topological characteristics can be used to assess the maturity of quality management as a discipline, and how these measures can be used to continually improve the *QMJ*. Understanding how to improve the quality of a citation network using its topological characteristics will provide insight into using quantitative measures to improve any knowledge base.

Statement of the Problem

The problem for this study is to model the evolution of knowledge flows in quality management by constructing a citation network from articles in the *Quality Management Journal* from October 1993 to October 2008. This model will be used to: 1) characterize the knowledge flows and the maturity of quality management as a research discipline, 2) identify which articles are hubs, authorities or emerging leaders, 3) determine whether the mean fitness (or quality) of the network at any time can be predicted by the network's topological characteristics, and 4) determine whether there is a significant difference in the mean fitness between the earlier period from 1993-2000, and the later period from 2001-2008. Results will indicate how the citation network and the quality measures derived from its topology might be used to enable continuous improvement of the *QMJ*.

Statement of the Research Questions and Hypotheses

This research will construct a network of journal articles in quality management from the *QMJ*, where nodes are individual articles and connections reflect citation relationships between the articles. Then, applying graph theoretical methods to analyze the network, this project will answer the following research questions:

RQ1. What does the citation network of articles in the *QMJ* reveal about knowledge flows in the study of quality management and the maturity of the discipline?

RQ2. What are the most influential resources in the *QMJ* citation network?

RQ3. Can the mean quality of the network $<\eta>$ or attachment-weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics at that time? Can $<\eta>$ or $<\eta>\zeta$ be predicted using only the exponent of the degree distribution at that time, γ_t?

RQ4. Are there significant differences in the mean quality of articles in *QMJ* between 1993-2000 and 2000-2008 in terms of either $<\eta>$ or $<\eta>\zeta$? Which measure is more appropriate as a quality indicator?

The first research question will be exploratory, and will seek to discover the characteristics of the quality management citation network from the empirical data using descriptive statistics and network analysis. Network analysis illuminates the topology, or structure, of the relationships between citations. This approach requires constructing a graph theoretical model of the available citations in quality management from the *QMJ* and testing the following hypotheses, where each reveals different characteristics of the knowledge flows in the development of the discipline:

Hypothesis 1a: The citation network of articles in *QMJ* is random.

Hypothesis 1b: The citation network of articles in *QMJ* is scale-free.

Hypothesis 1c: The citation network of articles in *QMJ* exhibits the small world property of short average path length and high clustering.

The second research question will be answered by calculating four centrality measures for each of the nodes within the citation network (degree centrality, Google PageRank, Kleinberg's hub and authority scores) and rank ordering them to identify the most influential references. The third question will be addressed using multiple regression to determine whether the quality of the network at time t in terms of $<\eta>$ or $<\eta>\zeta$ can be predicted using the network's topological characteristics (particularly the exponent of the degree distribution at that time, γ_t). The fourth question will examine the changes in quality of the network in terms of both $<\eta>$ and $<\eta>\zeta$ using a t-test.

Statement of the Purpose

The purpose of this study is to survey the research in quality management using the *QMJ* citation network to identify critical articles and the growth mechanisms that generated the network. Using these patterns, the research explores a quantitative approach for evaluating the quality of knowledge flows within the research discipline. This outcome is possible because topological measures on a citation network reflect how easy it is to find resources, how thematically connected those resources are, and ultimately whether the knowledge base is nascent or mature. Using these measures to express the quality of a citation network, the results from this research can improve search capabilities within the collection of articles in quality management, make the connections more visible to researchers to improve research practices and editorial policies, provide an objective quality assessment and metric for continuous improvement, and facilitate bringing academic theory into practice more readily.

This study contributes to the academic literature in quality management by providing the first analytical survey of the articles within the *QMJ* citation network to identify critical articles and topic clusters. This research also contributes to industrial practice by developing the concept of *attachment-weighted fitness* on a network as a proxy of its quality, a measure that could potentially be useful to assess and continually improve the quality of any knowledge base. Finally, the work adds to the body of knowledge in complex systems research by examining the genesis and evolution of a citation network from the unique perspective of quality estimation and continuous improvement.

Statement of the Need

The results from this study will be useful for educators, researchers, journal editors and publishers, professional societies, and business or information technology leaders who are responsible for managing and continually improving a knowledge base. Professors and other educators in quality management need to know which resources are most critical or influential within the field to guide their curricula. Researchers need to find the most critical, significant and relevant studies to inform their studies. Journal editors and publishers need to continually improve the quality of their publications to retain and increase readership. Professional societies such as the ASQ, whose mission includes stimulating the development and flow of new knowledge to catalyze innovation with the field, are motivated to ensure that their publications effectively support this outcome. Business executives and information technology leaders need to continually improve their corporate knowledge bases. Thus, a measurement system to assess the quality of a knowledge base and provide metrics for continuous improvement has practical implications for knowledge managers, information technology managers, and executives who are responsible for increasing the return on investment for an organization’s knowledge repositories. Finally, the work will result in a searchable database of articles and their citations derived from multiple sources, as well as a graphical representation of the quality management body of knowledge which heretofore has not existed. Both of these artifacts will help future researchers in quality management more easily conduct comprehensive literature reviews.

Statement of the Assumptions

The following assumptions were made in pursuit of this study. The data to be used for this research are the complete set of full-length, peer-reviewed journal articles and their references, published in the *Quality Management Journal* from its inception in 1993 through the end of 2008. This only represents a subset of the full network of journal articles that have been published on the topic of quality management, and an even smaller subset of all intellectual contributions to the field (e.g. books, theses, dissertations and magazine articles). It will be assumed that the test data provide a relevant and appropriate subnet of the full network due to a) the currency of the data, and b) the mission of *QMJ*. Regarding currency, a recent sample of data is more likely to cover a broader swath of the quality management body of knowledge, since older articles will be limited to drawing from the references that came before them in time. Regarding the mission of *QMJ*, the assumption is made that the journal has met its goal of becoming the journal of choice for publishing research in quality management in the U.S., and so examining articles in *QMJ* (while not comprehensive) is useful for understanding the foundations and evolution of the field of study relative to this country.

It will also be assumed that the information is complete, that is, that authors have appropriately cited all of the references that informed their research papers. Furthermore, this research assumes that there has been a sufficient quantity and diversity of reviewers, and that a double-blind peer review process was consistently applied. (Hamacher, 2005)

This means that the citation network can be considered an objective representation of the knowledge flows and growth of adequately demonstrated ideas within the field.

The citation network is assumed to be an appropriate model for knowledge flows. (Zhuge, 2006) There are, however, potential biases introduced due to subnet selection, because invariably a network does not capture all of the knowledge flows within a field. However, sampling properties of networks are not yet well understood (Stumpf, 2005) and so the proposed study is subject to the same challenges as all other recent research that uses graphs to model complex dynamical systems.

Statement of the Limitations

The primary limitation for this study is that it will be based only on a subset of all available research in quality management, namely, the research published in the U.S.-based *Quality Management Journal* from its inception in 1993 to 2008. Even though many foreign researchers publish in the *QMJ*, characterization of knowledge flows in quality management in terms of the *QMJ* may not adequately capture a global perspective on the topics of interest to both researchers and business.

A more comprehensive view of the quality management citation network and its coverage of the Body of Knowledge could be determined by expanding the sample to include other research outlets. These might include the *International Journal of Quality and Reliability Management*, *Total Quality Management & Business Excellence*, and quality management articles from the *Academy of Management Journal* and *Academy of Management Executive*, for example. Also, because comprehensive tracking of conference proceedings related to topics in quality management will not be included in the research data, the ability of the proposed study to detect innovations and emerging topics may be limited. New results are often presented at conferences in advance of being published in journals. Furthermore, the study of emerging trends in citation networks using network-based methods has traditionally utilized cocitation data (e.g. Wallace et al., 2008) or alternatively, with text mining using Markov models. (Le et al., 2005) Using these approaches, emerging trends are not diagnosed by the citation network, and as a result are detected at the concept level rather than the article level.

Statement of the Methodology

Subjects

The subjects for this study will be the complete collection of 261 full-length articles that appeared in the *Quality Management Journal* from the first time the journal was published in October 1993 through its October 2008 issue. This sample does not include the extended discussions, redactions, dissertation summaries, and book reviews that were also published in *QMJ* during this time. The citation network that was constructed from these articles and the references that they cited included a total of 7091 nodes and 9414 references.

Research Design

The proposed research will leverage three distinct approaches: analysis of the topology, analysis of node centrality, and statistical analysis of metrics derived from the network topology. Analyzing the topology involves building the network and extracting topological measures from the network, which is the functional equivalent of examining descriptive statistics for a non-network dataset. Hypotheses on the degree distribution of the network will be tested to establish whether the observed citation network is random, scale-free, or exhibits the small world property of high clustering with low average path length through the network. Increased clustering and the formation of a small world would reflect that a maturation of the research practices within the field has occurred. Changes in the topological characteristics at 60 equally spaced time intervals, as new nodes and edges were added to the network from October 1993 through October 2008, will be examined qualitatively to infer information about growth patterns within the network. Centrality analysis involves extracting measures for each of the nodes in the network, and ranking the nodes based on those measures. Finally, this research will explore mechanisms for predicting the quality of the network over time using the topological characteristics of the network and the centrality measures. These measures will be investigated to determine whether there was an appreciable change in quality within the *QMJ* citation network between the early years of the citation network (1993-2000) and its later years (2001-2008).

Procedures

The first research question will be answered by examining the degree distribution of the full network using the Shapiro-Wilk test of normality, a maximum likelihood based method for power law fitting, and a heuristic test to uncover small world characteristics within the network. The second question will be answered by computing four centrality measures for each node in the network (degree centrality, Google PageRank, and Kleinberg’s authority and hub scores) and presenting rank-ordered lists of nodes. The third research question will apply the technique of multiple linear regression to determine whether quality, in terms of the mean fitness of nodes within the network, can be predicted from the network’s topology and recent growth. The fourth question will use a two-sample t test to determine whether there was an appreciable change in quality of the network between the periods 1993-2000 and 2001-2008.

Definitions of the Key Terms

The following terms have been defined for the purposes of this study.

- Authority: In citation networks, an article that is cited by many others articles, but in particular by other hubs. (Fowler & Jeon, 2008)
- Centrality: A measure of the relative importance or significance of a node within a network. There are several centrality measures, each of which applies a different algorithm for quantifying importance. These include degree centrality, Kleinberg's authority and hub scores, and Google PageRank. (Newman, 2003b)
- Citation: An acknowledgement by one published article that its content has drawn upon the knowledge within another article. (Zhuge, 2006)
- Citation Network: A graph in which the nodes are published articles and the edges occur when one article references, or cites, another. (Batagelj, 2003)
- Clustering: The tendency for two nodes connected to a shared node to also be connected to one another (i.e. "the friend of my friend is also my friend"). (Newman, 2003b)
- Cohesiveness: A qualitative description of the network density and clustering within a network; a highly clustered network is often cohesive, whereas a random network is cohesive only if the connection probability between nodes is high. (Albert & Barabasi, 2002)

- Degree: The number of connections coming into or emerging from a node. (Newman, 2003b)
- Degree Distribution: The histogram of the frequency of occurrence of a node associated with a given number of degrees, that is, $P(k)$ as a function of k. (Albert & Barabasi, 2002)
- Density: The ratio of the number of edges observed within a network to the number of edges that are theoretically possible in a random network with connection probability of 1. (Albert & Barabasi, 2002)
- Directed Graph/Directed Network: A graph in which the relationships between the nodes are not equivalent based on which node is used as the point of reference. A citation network is directed because an article that cites an article that has been previously published is rarely cited by that previous article. (Zhuge, 2006)
- Dynamic Centrality: The tendency of articles to become more or less pivotal over time (Hill & Braha, 2009)
- Edge: A connection between two nodes in a graph, typically reflecting a relationship or an exchange of materials or information. (Newman, 2003b)
- Erdos-Renyi Graph: A random graph uniquely specified by a number of nodes n and a connection probability p, which represents the probability of an edge connecting any two nodes. (Albert & Barabasi, 2002)
- Google PageRank: The algorithm, developed by Google co-founders Larry Page and Sergey Brin, that describes the centrality of a node within a network recursively based on the centrality of the nodes with which it

connects during a random walk with a predetermined probability of randomly jumping to a completely new node. (Jezek et al., 2008)

- Graph: A mathematical representation of a network consisting of two sets: a set of nodes, and a set of edges that reflect relationships between those nodes. (Newman, 2003b)
- Hop: A connection between two adjacent nodes. (Jezek et al., 2008)
- Hub: In citation networks, an article that cites many other articles, e.g. a review article comprehensively describing research to date on a particular concept. (Fowler & Jeon, 2008)
- In-Degree: The number of connections pointing into a node. In a citation network, the number of citations an article receives. (Vazquez, 2001; Bilke & Petersen, 2001)
- Isolate: A node in a network that is not connected to any other nodes. (Bilke & Petersen, 2001)
- Network: A representation of relational data in the form of a mathematical graph. (Newman, 2003b)
- Node: The most basic element of a graph, connected to other nodes by edges. (Newman, 2003b)
- Out-Degree: The number of connections emanating from a node. In citation networks, the number of citations an article makes. (Vazquez, 2001; Bilke & Petersen, 2001)
- PageRank: see Google PageRank. (Jezek et al., 2008)

- Path: A connection between any two nodes in a graph. (Newman, 2003b)
- Preferential Attachment: The dynamical process whereby a node forms edges with other nodes based on some measure of the quality of those nodes. In citation networks, this refers to the tendency of an article to cite other frequently cited articles.(Albert & Barabasi, 2002)
- Random Graph: see Erdos-Renyi Graph. (Albert & Barabasi, 2002; Denning, 2004)
- Reference: see citation. (Zhuge, 2006)
- Scale-Free: A property of a network in which some nodes, called hubs, are connected to many other nodes, whereas other nodes are not connected to many other nodes at all. Hub-and-spoke transportation systems (such as airline route maps) are scale-free networks. (Albert & Barabasi, 2002; Denning, 2004)
- Small World: A property of a network in which the average distance between two nodes is small as compared to the size of the network. (Albert & Barabasi, 2002; Denning, 2004)
- Topology: The configuration and structure of a network. (Newman, 2003b)
- Transitivity: see clustering. (Newman, 2003b)

Summary

The purpose of this chapter was to provide a conceptual foundation for the proposed research. An introduction to the history and purpose of the *QMJ,* the concept of knowledge flows and their relation to citation networks and the Body of Knowledge concept, and the technique of modeling knowledge flows using citation networks was presented. Next, the outline for the proposed research was described in terms of the problem statement, the research questions and hypotheses, the purpose and the need for the study, and the assumptions and limitations that are related to the proposed methodology. The chapter concluded with a list of key terms that are essential in this research.

CHAPTER 2

REVIEW OF LITERATURE

Chapter Overview

The purpose of this chapter is to review four conceptual areas within the literature that are critical for understanding the citation network in quality management and the implications of its topology and evolution. The areas are: 1) characterizations of the quality management body of knowledge (BOK), that describe the conceptual clusters that are expected in the *QMJ* citation network, 2) the technique of network analysis of complex dynamical systems (from which the methods of this experiment will be drawn), 3) the theory underlying what citation networks represent, and a survey of prior work on citation networks in other fields, and 4) the roles of quality and continuous improvement in the development and maintenance of a body of knowledge will be explored, including a summary of the ways in which research quality and productivity are currently assessed.

Bodies of Knowledge in Quality Management

“Body of Knowledge” (BOK) is a term used to characterize a taxonomy of interrelated concepts, the literature that supports them, or both. (Fleisher, 2007) The two established BOKs in quality management, one from the American Society for Quality (ASQ) and another from the Juran Center, will now be described to provide a framework for understanding how the quality management discipline is organized.

The Juran Center BOK

The Joseph M. Juran Center of the Carlson School of Management at the University of Minnesota maintains "The Body of Knowledge in Quality" on the World Wide Web at http://www.csom.umn.edu/Page5508.aspx. This BOK is a database only, described by the Juran Center as "containing citations and abstracts of literature (journal articles and books) on the subject of quality and performance improvement. While we do not claim to have all of the relevant literature in the quality and performance improvement field, we do think we have the vast majority of it." (Juran Center, 2008)

The key strength of this BOK is its purported comprehensiveness; the Juran Center claims that it covers the full discipline of quality, of which quality management is a component. However, although it contains "classic earlier citations" its core references are from the 1990's to present. The Juran Center BOK classifies its citations into ten categories, based on the categories, criteria and core values of the Malcolm Baldrige National Quality Award. Each citation in the Juran Center BOK is associated with one or more of these categories, outlined in *Table 1*. However, beyond this classification scheme, no taxonomy for the contents of the BOK is provided and so it is not possible to examine how well the Juran Center BOK covers specific topics or subtopics.

Table 1

Categories in the Juran Center Body of Knowledge in Quality

Leadership for Quality	Process Management
Strategic Planning for Quality	Business and Quality Results
Customer and Market Focus	Tools for Quality Improvement
Information and Analysis	General Approaches and Philosophies
Human Resource Focus	Definitions of Quality

The ASQ Certified Manager of Quality/Organizational Excellence BOK

The ASQ maintains 14 separate BOKs that correspond to each of the certifications the organization offers, unified under the banner of the "Quality BOK" or "Q-BOK" for short. The body of knowledge in quality management is described by the ASQ Manager of Quality/Organizational Excellence Certification requirements. This BOK consists of seven major sections: Leadership, Strategic Plan Development and Deployment, Management Elements and Methods, Quality Management Tools, Customer-Focused Organizations, Supply Chain Management, and Training and Development. (A complete listing of the major elements and all of the subtopics under each element can be found in Appendix B.)

In addition to providing this taxonomy, the ASQ also maintains its "Quality Information Center" (QIC) online at http://www.asq.org/qic. This database is intended to be a supplement to the taxonomies that ASQ publishes as its formal BOKs, and contains over 24,000 citations and abstracts from the publications and conferences that ASQ has been affiliated with since 1944. Non-ASQ resources and all books (including those published by ASQ's Quality Press) are not indexed. As a result, despite its exemplary coverage over time, the ASQ QIC will only provide coverage of the topics covered by ASQ sponsored publication channels.

ASQ is highly committed to keeping its BOKs accurate, relevant and up-to-date. In 2005, the organization launched a multi-year effort to understand the processes that contribute to development of the BOK, the products that emerge as a result of using the BOK, and the generation of new research ideas that are later integrated into the BOK.

Their internal conceptualization of the process model, from ASQ internal documents, is illustrated in *Figure 2*.

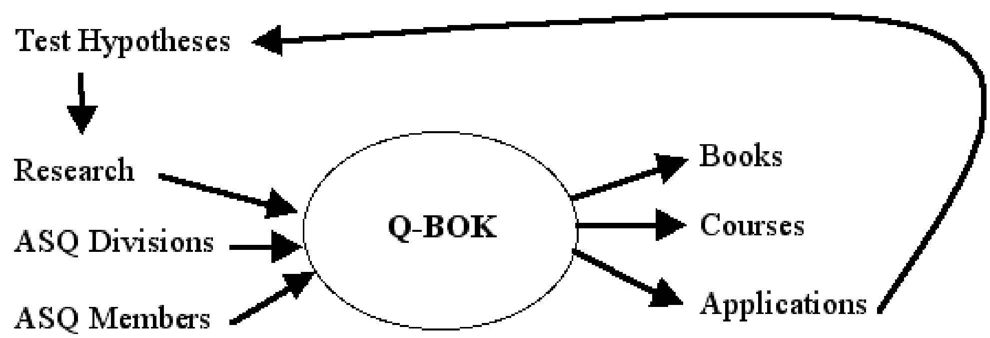


Figure 2. Knowledge flows with respect to the ASQ quality BOK. (ASQ, 2008a)

A Comparison of the Juran Center BOK and ASQ Quality Management BOK

Although the Juran Center BOK uses a classification scheme derived from the Malcolm Baldrige National Quality Award (MBNQA) program, and the ASQ BOK categories were the result of expert opinion, the schemes are very similar. With the exception of “Definitions of Quality” each of the Juran Center BOK categories maps onto one or more ASQ categories, as shown in *Table 2*. This suggests that if an alternative method is used to derive the structure of the BOK, such as citation network analysis, its structure should reflect at least the minimum categories as defined by the ASQ BOK. For the purposes of this study, the quality management Body of Knowledge is assumed to contain seven categories, derived from the union of the ASQ and Juran Center taxonomies. These seven categories, which will be used for the concept analysis of the citation network, are: 1) Organizational Leadership, 2) Strategy Development, 3) Customer-Focused Organizations, 4) Training and Development, 5) Quality Management Tools, 6) General Management, and 7) Definitions of Quality.

Table 2

Comparison of Categories in the Juran and ASQ Bodies of Knowledge

Juran Center BOK:	ASQ Quality Management BOK:
1. Leadership for Quality	1. Organizational Leadership
2. Strategic Planning for Quality	2. Strategy Development
3. Customer and Market Focus	3. Customer-Focused Organizations
4. Information and Analysis	
5. Human Resource Focus	4. Training & Development
6. Process Management	
7. Business and Quality Results	
8. Tools for Quality Improvement	5. Quality Management Tools
9. General Approaches/Philosophies	6. General Management
10. Definitions of Quality	(no match)

Network Analysis of Complex Dynamical Systems

Many relational structures have been studied using graph theory and network methods, including letter-passing between acquaintances (Milgram, 1967), servers and routers connected by the internet (Yook et al., 2001), documents and hyperlinks on the World Wide Web (Huberman & Adamic, 1999), and social, metabolic and communications networks. (e.g. Onnela et al., 2007; Lu, 2000; Jeong et al., 2000; Lu, 2000) Research results indicate that most networks observed in real life are not random, but reflect varying degrees of structure which suggest that self-organizing principles are at work. For example, new web pages have been shown to link to existing pages with greater perceived utility, a linear growth process called “preferential attachment”. (Barabasi & Albert, 1999) Dynamical principles also influence the breakdown of a network. The “scale-free” topological structure of the social network of terrorists suggests that the network is highly resilient to attacks unless the “hubs” are strategically targeted. (Krebs, 2000)

Statistical and analytical methods on networks can also be employed to investigate how the body of knowledge in quality has evolved into what it is today, which would provide the technology manager with insights into how quality methods have emerged as a consequence of scholarship. Such conclusions are possible because the growth and evolution of an academic body of knowledge is a complex dynamical system which can be modeled using a citation network. In a citation network, individual publications are modeled as nodes in a graph, and edges reflect a relationship in which one paper references another. By statistically examining the topology of the citation network and how it changes over time, many global elements of the network can be studied including its most influential contributors, the most significant publications, the cohesiveness of information and efficiency of knowledge transfer within the field, and periods in time where key ideas gained momentum and flourished.

Topological Properties of Networks

When measures on a network are considered collectively, they describe the topology (or interconnection structure) of the system as a whole. They suggest (but do not determine) which growth mechanisms influence the dynamical system, and can be either local or global. Local properties reflect the state of the system in the neighborhood of a particular node. Global measures suggest which organizing principles and mechanisms for growth may be operable within a network. One commonly used local measure is *centrality,* and three commonly used robust measures at the global level are *average path length*, *degree distribution,* and *clustering coefficient*. Each of these diagnostics is now briefly described, with centrality discussed in the context of the degree distribution.

Average Path Length

Average path length of a network (or “degrees of separation”) is one intuitive and appealing measure that illustrates the utility of modeling a dynamical system as a network. This metric has been used to identify Kevin Bacon as one of the more central figures among movie actors (Watts & Strogatz, 1998), and Paul Erdös as one of the most influential researchers in the historical development of mathematics. (Barabasi, 2001) Networks with small average path lengths were originally referred to by Milgram as “small worlds” which, when mathematically formalized by Watts and Strogatz, were shown to be sparse, yet well-connected.

Degree Distribution

The degree distribution reveals some topological features of a network. The nodes within a network are described by their *degree*, which is the number of connections to that node. In directed networks such as a citation network, the connections are distinguished by the number of incoming connections (in-degree) as well as the number of outgoing connections (out-degree). The degree distribution plots number of degrees attached to a node on the horizontal axis, and the frequency at which nodes with that number of degrees are observed on the vertical axis. The degree correlation is evaluated by looking at the relationship between the degree of a node and the average degree of its neighbors. This reflects whether the network is assortative, meaning that high-degree hubs tend to be linked to high-degree hubs, or disassortative indicating that nodes tend not to attach to hubs. Centrality reflects the relative position or popularity of a node in the network, providing an additional measure to confirm whether nodes with the highest degrees are indeed the most central. The hub score, authority score, eigenvector centrality, and the PageRank score which is related to eigenvector centrality (popularized by Brin and Page as the algorithmic engine running Google) all point to the most significant nodes in the network. (Newman, 2003)

Clustering Coefficient

In many real networks, there is an element of clustering. For example, social networks are characterized by smaller groups and communities where members are more familiar with one another. The clustering coefficient**,** which describes the probability that two neighbors of one node in a graph will also be connected to each other, is defined as the ratio of twice the number of links incident to nearest neighbors to the total number of possible nearest neighbor links. The clustering coefficient is a monadic measure, so each node in a network can be described by its clustering coefficient. The relative amount of clustering in an entire network can be approximated by considering the mean of all the monadic clustering coefficients. For the purposes of this study, clustering coefficient will refer to the mean clustering coefficient of the network.

Characteristic Network Topologies

According to the Random House Unabridged Dictionary, *topology* is “the study of the properties of geometric forms that remain invariant under certain transformations, such as bending or stretching.” The topology of a network reflects the interconnection patterns between its nodes, and for large networks, the topology can be inferred from its statistical properties. Four characteristic network topologies, which are commonly encountered in modeling studies, are reviewed here: lattice, random, scale-free and small world. The random, lattice and small world network types are illustrated in Figure 3. The scale-free network, illustrated for three different degree distribution exponents and one exponential, are shown in Figure 4. Note that the scale-free networks become more tree-like as the exponent increases from 2.5 to 4. The random and scale-free network types are

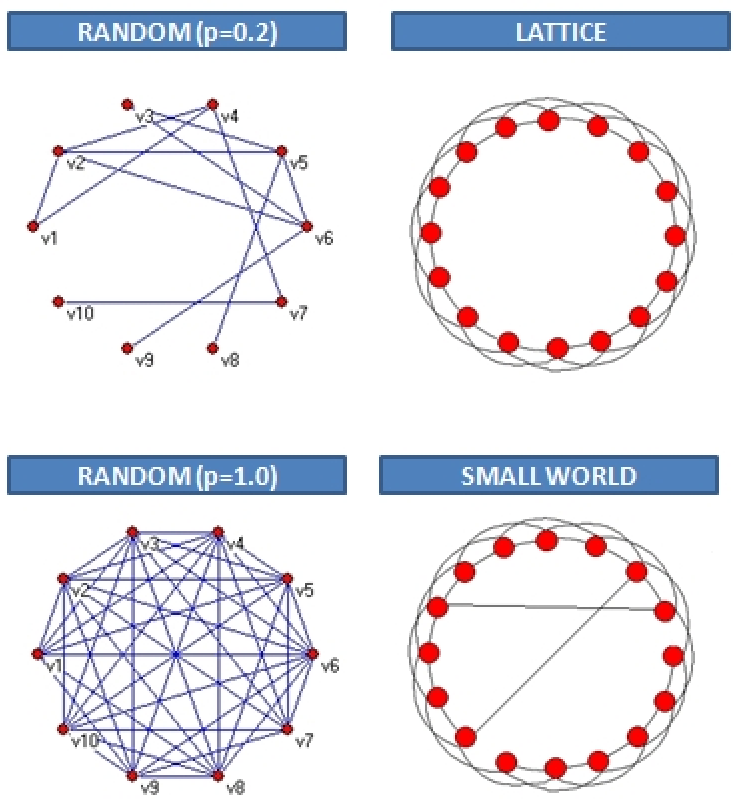


Figure 3. Random, lattice and small world network types.

contrasted with the hierarchical network type in Figure 5, in terms of the defining characteristics of their degree distributions and clustering coefficient distributions. Each of the network types is observed in biological, physical and socio-technical systems.

Lattice Networks

Lattice networks, also called regular or local networks, represent classes of systems with structural uniformity. Crystalline structures, the orientation of data processing machines in grid computing, distributed instrument control systems, sensor networks, and constellations of satellites can all be modeled using this representation. Mature citation networks, however, are never uniform. For this reason, lattice networks are presented for contrast only, and will not be examined relative to the citation network in this study.

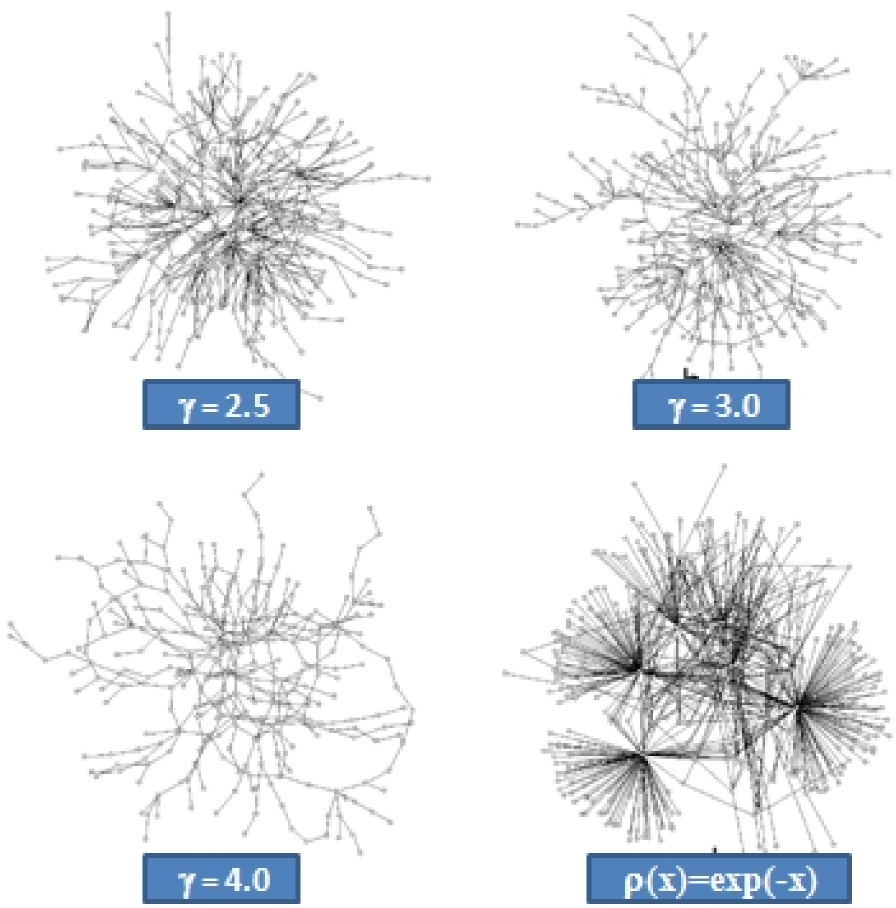


Figure 4. Scale-free networks with degree distributions fit by power laws of varying exponent values, contrasted with a network with degree distribution fit by an exponential. Images from Servedio et al. (2004)

Random (Erdos-Renyi) Networks

Random networks (also called Erdos-Renyi models after the mathematicians who initially explored the concept) consist of nodes which are randomly connected to one another with a uniform connection probability p. For example, road systems can be considered random networks because they do not preferentially attach to intersections (even though strictly speaking, the road network is not random because it has a spatial

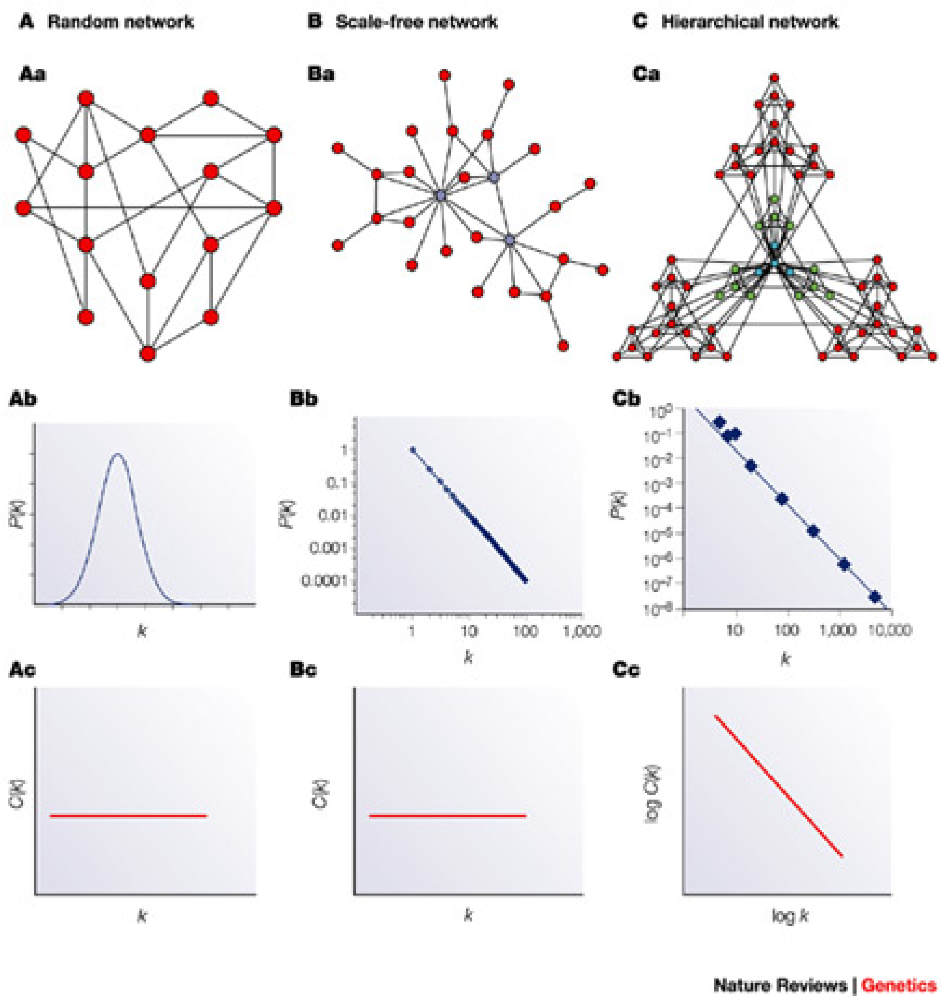


Figure 5. Characteristic network topologies. (From Barabasi & Oltvai, 2004) P(k) represents the probability that a node has exactly k edges. C(k) is the expected monadic clustering coefficient for a particular network type for a node with exactly k degrees.

dependency. Because the degree distribution of a random network is Poisson, both concentrated clustering and the formation of communities are usually not observed

Scale-Free Networks

The distinguishing mark of a scale-free network is that its degree distribution follows a power law, lognormal, or exponential. This indicates that most nodes have few connections to other nodes, whereas some nodes have many connections. If the degree distribution follows a power law, then the probability that a new connection attaches to an existing node with k degrees is $P(k) \propto k^{-\gamma}$. Airline routes can be modeled as a scale-free network, because small airports typically fly to hubs that connect the traveler with a wide variety of potential destinations, and national hubs connect to international hubs. The growth process of *preferential attachment* typically results in a scale-free network. This means that new nodes joining the network will tend to attach to existing nodes with higher *degree* relative to the other pre-existing nodes in the network. Garfield (1964) and de Solla Price (1965) noticed this tendency in citation systems and referred to it as "cumulative advantage". Through preferential attachment, the articles that are already well cited are more likely to gain additional references.

Small Worlds

Scale-free networks are often also small worlds. Small world networks are characterized by sparse density and a small average path length across the network as compared to the size of the network. (Watts & Strogatz, 1998) For example, a person in a sports stadium filled with fans might be well acquainted with only a handful of companions. However, observed networks like this often display a "small world" effect where through chains of friends and acquaintances, that person has the potential to connect with a large fraction of the entire audience. The small world effect has been observed more recently on the global level by researchers at Microsoft. They studied one month of instant messaging (IM) communications across three continents, and found an average path length through the network of 6.6 degrees (Leskovec & Horvitz, 2008) With respect to a citation network, the presence of a small world indicates that researchers have made connections between less closely related topics or articles. This is an indicator of navigability or searchability within the network (Kleinberg, 1999) which reflects that the network has achieved a level of maturity.

Summary of Topological Features

In a lattice network, P(k) is a constant and clustering is uniform throughout the network. In a random network, the degree distribution is Poisson with a peak at <k> and amount of clustering is uniform throughout the network. In a scale free network, the degree distribution is typically a power law (but can also be exponential, truncated exponential, or lognormal) and clustering is uniform. In a hierarchical network, the degree distribution is typically a power law, but very few nodes are connected to the whole network (e.g. the CEO of a company); most people are minimally connected. Any network, however, can be a small world if "short cuts" are added to generate a scenario where clustering is high and average path length is low compared to a random network.

Evolution of Networks

A commonly observed characteristic of a dynamical system, which is not captured by the mathematical model for constructing a random network, is that it grows over time. Social networks expand and contract, diseases spread and infect more people, and technological systems like the World Wide Web (WWW) support the ever-increasing publication of documents and the constant adjustments of hyperlinks. The topological characteristics of a network also change over time as the number and connectivity of nodes and edges within the network fluctuate.

Preferential Attachment ∏(k)

As a random network grows, the probability of a new link attaching to a node of degree k, ∏(k), is constant and equal to the connection probability p. In contrast, the process of preferential attachment reflects a state of anisotropy in the network. That is, a newly introduced node will connect to an existing node with a probability that is proportional to the number of degrees already incident on the existing node, or $\prod(k) \propto k$. If an existing node has a high number of degrees, then it is more likely to attract new connections than a node with few links (a "rich get richer" model). When this process is continued over time, the result is a distribution in the clustering within the network, where some groups of nodes are tightly clustered and others are more loosely connected.

Evolving networks characterized by growth and preferential attachment to nodes with higher degrees, referred to as the Barabasi-Albert (BA) model by the researchers who derived it, has been studied in depth. (Albert & Barabasi, 1999) One of the results is that this model analytically predicts that the resultant network will be scale-free with a degree distribution exponent of $\gamma = 3$. The BA model reflects linear growth of nodes and edges, and preferential attachment that depends solely on the degree of the pre-existing nodes.

BA-style preferential attachment can be measured directly for networks where the node and connection data are available at different timesteps. The desired result from this measurement is a representation of the functional form of $\prod(k)$, derived by calculating the change in incoming degrees for each node compared to the total number of degrees (dk_i/dk)which were assigned to nodes over a time interval ($t - \Delta t$), where Δt is much shorter than the total age of the network. Plotting this rate of change as a function of k yields the probability distribution for $\prod(k)$. (Albert & Barabasi, 2002)

Generalized Preferential Attachment

Since the BA model was first studied, many other generation processes have been posited to explore the mechanisms by which networks grow and evolve. Most of these approaches are based on preferential link creation, in terms of various aspects of the network such as node attributes (Söderberg, 2003), Euclidean distance (Fabrikant et al., 2002), subtopic or subcategory (Menczer, 2004), econometric variables (Powell et al., 2005), or competitive trade-offs and network optimization. (Colizza et al., 2004) These are all examples of generalized preferential attachment, which refers to “any kind of preference to interact with other agents (nodes) with respect to any node property.” (Roth, 2008) Using generalized preferential attachment, the connection probability $\prod$ is a function of other variables, which may or may not include a dependence on the degree k. Preferential attachment based on measures of quality, or fitness, is the growth process in focus for the present research.

Fitness and Attachment-Weighted Quality

According to Bianconi and Barabasi (2001), a real network will exhibit a competitive dynamic where some nodes will have an innately better ability to compete for new connections. The fitness of a node η thus represents how effective that node is in competing for new links. As a result, the fitness is a proxy for the quality of a node, and the mean fitness of the growing nodes in a network will the degree of preferential attachment that is active in the network over a time interval that is small compared to the age of the network. Even if this value is not normally distributed, it will be higher whenever a smaller portion of nodes are capturing a greater proportion of the available links that are assigned each time new nodes and edges are added to the network.

To continually improve the network, however, it would be more effective to use a metric that is systemic rather than monadic. Extending the node-based definition of fitness, the quality of a citation network can be diagnosed by four distinguishing characteristics: 1) it will be a small world, 2) the pre-existing nodes will be highly effective in attracting new links, 3) there will be a constant influx of new articles into the network to stimulate innovation, and 4) the mechanism of preferential attachment will be active in the network. First, define a time interval $(t - \Delta t, t)$ that is much smaller than the age of the network. Because there are no metrics to accurately assess the "strength" or "weakness" of a small world, a systemic quality metric could consider only the latter three elements at the present time. If the capability of the pre-existing nodes within the network to capture additional links is calculated in terms of the probability that a pre-existing node will capture a link, this can be expressed as $\zeta = 1 - \Delta N/\Delta E$ where ΔN is the number of new nodes added over the time interval $(t - \Delta t, t)$, and ΔE is the number of new edges. As the pre-existing nodes capture all the new incoming links, $\Delta N/\Delta E \rightarrow 0$ and $\zeta \rightarrow 1$. There will be some optimal value of ζ which is less than 1 that corresponds to the probability $1 - \zeta$ that an entirely new article enters the network (the innovative potential).

For the mechanism of preferential attachment to be active, the fitness of each node η is considered to be its rate of acquisition of new nodes dn/dt over (t - Δt, t). This can be directly measured from the data. The mean fitness over all the nodes $<\eta>$ represents the average rate of acquisition of new links among growing nodes, and will be lower if the links are distributed widely, and higher if the links are concentrated among a smaller subset of nodes. Thus, a higher $<\eta>$ will reflect a greater concentration of the acquisition of links on a smaller subset of nodes. Note that the distribution of $<\eta>$ over all the nodes may or may not be Poisson, and as a result, the value may not be physically meaningful. In any case, it provides a rough approximation of how active preferential attachment is in a network over a given time interval. This study also assumes that the sampling is structured such that the rate of increase of nodes and edges with time is linear.

Using these values, the *attachment-weighted quality* of the network at time t is defined to be the product of the mean fitness and the probability of attachment to the pre-existing network:

$$A_{\eta} = <\eta>_t \zeta_t = \frac{\sum_{\Delta t} (t_i - \bar{t})(dk_i/dt - \overline{dk/dt})}{\sum_{\Delta t} (t_i - \bar{t})^2} \left[1 - \frac{\Delta N}{\Delta E} \right]$$

The first term, the expansion for $<\eta>_t$, represents the mean slope of the ordinary least squares regression line through the rate of change of acquisition of links, averaged over all the nodes. The second term represents the probability that a pre-existing node receives a connection each time new nodes are added to the network. As a result, the second term is an attachment-based weighting of the mean fitness of the network $<\eta>_t$.

Preferential Attachment with Fitness

Under preferential attachment with fitness, new nodes joining the network will tend to attach to existing nodes with higher *utility* relative to the other pre-existing nodes in the network. This means that the probability of connection is not entirely dependent upon the number of degrees incident on the older node, but on an intrinsic fitness (or quality) that is an intrinsic characteristic of each node.

The preferential attachment with fitness model of Bianconi and Barabasi (2001) assumes that at every time step, a new node *j* is added to the network. The new node is described by a fitness η_j which is drawn from a fitness distribution $\rho(\eta)$. The probability of connecting to node *i* is proportional to both the degree and the fitness of that node:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

Thus the preferential attachment is an indicator of how effective the node is in achieving a proportion of the new connections within a competitive environment. In the calculation for attachment-weighted quality, note that the assessment of fitness does not depend upon knowledge of the prior number of links incident on a node, k_i, as in the traditional preferential attachment model above.

Citation Networks

Citation networks have been studied extensively for their mathematical interest (Bilke & Peterson, 2001; Vazquez, 2001) as well as their applicability in information science (Garfield, 1955; Garfield, 1974; Jezek et al., 2008; Sanyal, 2006; Scharnhorst & Thelwall, 2005). They have been used to illuminate growth patterns of the bodies of knowledge in business (Biehl et al., 2006), educational research (Carolan & Natriello, 2005), computing and information science (Polites & Watson, 2008), management of technology (Pilkington & Teichert, 2006), high energy physics (Lehmann et al., 2008), and all specializations in physics since 1893. (Redner, 1998) Insights into the cohesiveness of a dynamical system such as a citation network have been explored, as an offshoot of percolation theory in chemical networks, by identifying the emergence of large clusters in citation networks. (Cuesta, 2002; Bianconi & Barabasi, 2001) Findings such as these help researchers identify when a discipline has matured.

Because citation networks have been extensively explored as graph theoretical constructs, there is a wealth of information to support an investigation into the quality citation network. Preferential attachment as a growth mechanism in citation networks (Lehmann, 2008), dynamics of a time-evolving citation network (Leicht, 2007; Li-Chun, 2006), how citation networks compare to the networks of hyperlinked pages on the World Wide Web (Menczer, 2004), accelerated growth patterns (Sen, 2004), the effect of aging on the utility of past publications (Hajra, 2004), the effects of weak and strong coupling between disciplines (Zheng, 2003), and dealing with incomplete information due to sampling problems (Hamacher, 2008) have all been explored. Shibata et al. (2007) have used citation networks to identify which publications in a field are most likely to become core articles in the future.

Mature citation networks are not random. In general, they are sparsely connected, have scale-free degree distributions with a characteristic exponent between 2 and 3, and exhibit the small-world feature of high clustering and a small average path length compared to the size of the network across a single primary component. (Vasquez, 2001)

The mature citation network exhibits high overall knowledge transfer efficiency. (Li, 2007) A citation network that is not mature, consequently, may have limited clustering or long average path lengths, a degree distribution exponent that is below 2 or exceeds 3, or potentially even no statistically significant model fit to the degree distribution.

Citation Networks as a Model for Knowledge Creation

A prerequisite for the academic practice of research is that a scholar becomes familiar with the research and literature in which he or she will specialize. The citation network and the Body of Knowledge can both be considered constructs for understanding this basis. Although knowledge is dynamic, published articles are typically the primary medium for communication between academics, and citations are an objective acknowledgement of that communication. (Zhuge, 2006)

What is the relationship between a citation network, a Body of Knowledge, and the process of generating new knowledge (both by an individual and by a community)? First, it is important to recall that a Body of Knowledge for a field of study can be represented as a taxonomy of related topics, or as a collection of publications and articles that contain the information described by that taxonomy. According to Moore et al. (2005), these elements are inseparable: a field is “not simply an amorphous collection of articles… [but] a collection of articles with a particular structure… that emerges from the citation practices of the scholars who engage in that field.”

Van der Veer Martens (2001) establishes the relationship between a citation system and a Body of Knowledge by examining the correspondence, coherence, and consensus theories of truth. The correspondence theory promotes a singular connection between physical reality and the notion of “belief”, and is exemplified by the progress of science through the empirically verifiable scientific method. The coherence theory of truth “relates to the strength with which the system’s theoretical propositions are mutually supportive.” The principle of *stare decisis* in law, which promotes upholding past legal decisions as precedent, is an example. The consensus theory establishes truth on the basis of “converging assessments of probability assigned by observers,” but this approach is easily criticized because a widely held belief may in fact be false, while a truth may exist even if no one believes in it. A Body of Knowledge can be constructed according to any one of these theories, and often, multiple means for ascertaining truth are employed.

A citation network provides a mechanism to directly observe how these three modes of deriving truth combine to bring forth new ideas from an established base of information and knowledge. Although the process of citation may appear mundane on the surface, each time a prior article is cited, the contemporary author acknowledges both the existence of the source and the impact of the source on the new knowledge. “While classification systems explicitly facilitate the origination and organization of ideas, citation systems implicitly facilitate their association and recombination.” (Van der Veer Martens, 2001) As a result, the citation network complements a Body of Knowledge by illuminating the interconnections between ideas, and providing a basis for understanding how new ideas have emerged within the domain. The citation network is natively dynamic, whereas the Body of Knowledge is static and represents the cumulative knowledge to date.

The citation network thus describes a pattern of knowledge flows over time, whereas a Body of Knowledge captures only the cumulative state of a discipline. This state is represented by the comprehensive collection of information resources, accompanied by the taxonomy which summarizes the content. Because the citation network contains a representation of each of those information resources, and explicitly defines the ideological relationships between them, the citation network can be considered functionally equivalent to the Body of Knowledge if all available citations to date are included.

However, a citation network can provide researchers with additional utility, including outlining personal research roadmaps, modeling the evolution of knowledge and researchers' interests, recommending a network of appropriate references, identifying closely related peers, automatically discovering communities with shared interests, objectively and dynamically evaluating research, detecting emergent research areas, and estimating the maturity stage of a discipline. (Zhuge, 2006) Study of citation practices can reveal "the cognitive structure of research fields, the prominence of certain articles and scholars, and the developmental history of disciplines and areas of specialization" as well. (Moore et al., 2005) Because the citation network contains more embedded metadata than a Body of Knowledge, it can be a useful tool for objectively constructing and continually improving a Body of Knowledge.

Artificial Intelligence on Citation Networks

Artificial intelligence (AI) is the branch of computer science that seeks to automate the process of human reasoning. AI employs techniques such as semantic structures for knowledge representation, complex logic, rule-based or expert systems, neural networks and game playing to draw conclusions that otherwise would require human consideration. Algorithms for "evolutionary computation" based on genetic algorithms and swarm intelligence are also included among the modern topics of investigation in AI. (Luger & Stubblefield, 2004)

With the exception of neural networks, AI methods have not traditionally been applied to citation networks. Whereas a neural network is an active dynamical system where nodes can communicate with each other to change the state of neighboring nodes, once published, a node in the citation network typically does not change its state. In contrast with a neural network, a citation network evolves through growth only and does not decay. Additionally, all references from one article to another in a citation network are of equal strength, whereas in a neural network, it is through the differences in strengths between node connections that changes in state across the network as a whole are triggered (a process called “learning”).

However, this does not imply that it is impossible to use neural networks to model a dynamical system of citations. In one citation study that employed the technique of Kohonen maps from neural networks (Campanario, 1995), edge weights were artificially established by determining the frequency of citations between pairs of journals across the entire network before training the system. Although the results are intriguing, there are no studies that directly compare empirical data modeled as a neural network using Kohonen maps to models of comparable unweighted networks, and so the relative merits of the two approaches remain unclear. In addition, prior studies such as Campanario’s Kohonen maps study have not been for citation networks that treat articles as the quantum of observation, but instead have considered larger scales (e.g. technology field, institutions, countries).

Quality and Continuous Improvement

Previous sections have established that a citation network provides a plausible model for the knowledge flows that result in the formation of a body of knowledge. To determine whether measures on citation networks can be used to improve the quality of a body of knowledge, the notion of quality as it applies to a body of knowledge will first be articulated. Next, the meaning of continuous improvement as it relates to a body of knowledge will be explained, and traditional methods for continuous improvement in the domain of knowledge management will be outlined. Finally, because it represents the state of the art in monitoring and managing bodies of knowledge, the practice of automated citation analysis will be explored.

The Quality of a Citation Network

The International Organization for Standardization (ISO) defines quality as the "totality of characteristics of an entity that bear upon its ability to satisfy stated and implied needs." To determine how the quality of a citation network can be assessed, using this definition, two elements must be described: 1) the needs that a body of knowledge satisfies, and 2) the characteristics of the body of knowledge that would satisfy those needs. The collection of information conceptually linked together as a citation network can fulfill many needs. These include serving as a benchmark for standards of knowledge, expertise and professionalism, establishing certification requirements, and guiding curriculum development. Additionally, a body of knowledge can act as a point of reference to more readily recognize the limits of understanding within a field, and be better able to identify areas for further research. (Fleisher, 2007;

Abbott, 1988) To serve this purpose, however, it must be easy to browse, explore and understand the citation network.

Two other studies have focused on the problem of strategically evolving a body of knowledge. Recall that the citation network can be used as one tool to evaluate the accuracy of a body of knowledge, by providing visibility into themes within the research and how those themes have evolved. Pfeffer (2007) examined the 50-year history of research in management on behalf of the prestigious *Academy of Management Journal* to determine how that publication could increase the impact of its published research. Hagan (2000) addressed the general problem of improving journal quality from a publisher's perspective, referenced Day & Peters (1994) to explain that "quality in journal publishing is well defined: The utility, innovation and clarity of a journal's articles backed by the prestige of its authors and peers and presentation features are the key elements of quality. What has not been so well defined to date is a systematic means of achieving this quality." Their findings, which are summarized in Table 3, provide the foundation for finding a systematic means of not only achieving but improving journal quality.

These studies indicate that five general attributes of a body of knowledge should be assessed to determine its quality, and to uncover ways to continually improve a body of knowledge. These are: *utility* (the results are widely applicable to researchers and practitioners), *novelty* (there is evidence that new ideas are forming and spreading through the community), *accessibility* (the "barrier to entry" problem must be addressed so that useful research results are not blocked from being shared), *permeability* (a resistance to the common biases of traditional journal review), and *visibility* (members of

Table 3

Quality Attributes for a Citation Network

Source	Description
Pfeffer (2007)	Few articles have low utility and end up uncited
	Research results are practical and relevant to practitioners
	Null or controversial results are included
	Effective knowledge flows between academic and practitioner journals (ie. theory and practice)
	An unbiased review process
	Rejection rates lower than 90% average in social sciences
Hagan (2000)	Articles are well written and substantiated
	Peer review process is expedient
	Articles exhibit utility, innovation and clarity

the community must be able to view, browse and explore the citation network). It is important to establish at this point that the only attribute in this class that is under investigation is the utility of the network as a whole, and that none of the other attributes will be addressed by this study. However, identification of the measures that can be used to diagnose utility is a prerequisite for making those measures visible to the community of researchers and enabling them to make decisions based on the quality of articles.

The Philosophy of Continuous Improvement

According to Boer et al. (2000), continuous improvement is the “planned, organized and systematic process of ongoing, incremental, and company-wide change of existing work practices aimed at improving customer performance.” This aligns with the definition by Jha et al. (1996), which describes continuous improvement as the “collection of activities that constitute a process intended to achieve improvement,” such as simplifying processes, reducing waste, enhancing individual and team empowerment, and improving customer service. Lillrank et al. (2001) provides a slightly more transcendent view, describing it as “a purposeful and explicit set of principles, mechanisms and activities within an organization adopted to generate ongoing, systematic and cumulative improvement in deliverables, operating procedures, and systems.” Continuous improvement is both a philosophy and a process.

Boer’s definition of continuous improvement can be extended to the process of generating and sharing the lessons and research results that yield a body of knowledge. Using this analogy, continuous improvement would be defined as the planned, organized and systematic process of ongoing, incremental, and cross-disciplinary change of research and publication practices aimed at increasing utility to academics and practitioners. Integrating Jha’s definition, continuous improvement on a body of knowledge would include simplifying the processes required to publish peer-reviewed research, limiting the conduct of studies that are not useful, promoting and catalyzing innovative research to expand its utility, and enhancing accessibility to engender empowerment. These actions realize the quality attributes of *utility, novelty, accessibility, permeability* and *visibility* that were previously identified. To devise a method for continually improving a body of knowledge based on measures from the citation network, the typical stages of a continuous improvement effort should be addressed. These elements have been drawn from Jha et al. (1996), and are listed in Table 4.

Continuous Improvement in Knowledge Management

According to Davenport et al. (1998), knowledge management is “concerned with the exploitation and development of the knowledge assets of an organization with a view to furthering the organization’s objectives.” This includes all processes associated with

Table 4

Activities in a Continuous Improvement Effort (from Jha et al. 1996)

Stage	Activity
Understand and document the process	Identify value-added versus non-value-added activities Analyze cost, quality, and other relevant measures for equipment, labor, and material inputs
Simplify and improve	Reduce, combine or eliminate activities Improve the performance of equipment, labor and material inputs with respect to cost, quality and other relevant criteria Implement low-grade (incremental) automation Revise business rules as needed
Standardize and integrate	Reintegrate remaining activities Stabilize the process at its new level
Monitor performance	Measure and monitor Set new targets

identifying, sharing and creating knowledge, the systems used to implement knowledge repositories, and the operating procedures to facilitate knowledge sharing and organizational learning. Knowledge management deals with both explicit, documented knowledge, and tacit (internalized or subjective) knowledge.

Through an extensive survey of the knowledge management literature, Ford (2001) further clarified this definition by identifying four aspects of all knowledge management practices: generation, codification, transfer, and application. These correspond to the business practices of information processing, business intelligence, organizational learning, and organizational development. Continuous improvement is practiced on each of these practices in business, providing insight into how continuous improvement has been done in the domain of knowledge management already. Continuous improvement in knowledge management focuses on the data warehouse, which may or may not include a knowledge base. Each of four areas relies the data warehouse or knowledge base to store and index the information that represents the knowledge to be managed. In information processing, the quality of the data warehouse is improved either by addressing data integrity issues or by improving the software or systems that access that data. (Loshin, 2005)

With respect to business intelligence, quality is enhanced when more effective or more reliable reports are produced, or when better ways are identified to make decisions based on those reports. Quality improvement in organizational learning occurs when people can more readily internalize and act on new knowledge, gradually developing new capabilities that benefit themselves and their organizations. In organizational development, quality improvement refers to the collective development of new capabilities, enabled by organizational learning. (Christensen, 1997)

This information provides direction regarding what activities should occur to continually improve a body of knowledge modeled by a citation network in an automated fashion. First, the data integrity within the citation network and the systems supporting the citation network should be continually improved. Search and reporting functionality on the citation network should be continually improved. Individual interaction with the system, and feedback from the system, should also be continually improved so that researchers can more readily develop new capabilities. Finally, the monitoring system should detect the emergence of new topics and areas in which quality is increasing.

Automated Citation Analysis

Investigators typically perform their literature searches by using references from available articles, or by manually searching the Web. This process can be tedious, repetitive, and end in so many matches that the researcher cannot practically review the findings. To remedy this problem, one research team engineered an "assistant agent" system to facilitate this process, called CiteSeer. This system is implemented as a Web service and deployed at http://citeseer.psu.edu. It is a data-driven Automated Citation Indexing (ACI) service that builds its database of citations as each new search is conducted by a researcher. (Bollacker et al., 1998)

ACI "[completely automates] the citation indexing process without requiring any extra effort from authors or institutions… [and] improves on other technologies by extracting and making the contents of citations easy to access." (Lawrence et al., 1999) Additionally, the system is supported by a mature data model, and has been made freely available for non-commercial use by NEC Research. Giles (n.d.) has identified that one area for future development of CiteSeer is adding a citation graph browser to visualize the citation network, making it easier for researchers to navigate citations. However, the question of exactly how to support this navigation has not been answered.

The results of the proposed dissertation research could provide guidance for the requirements of an effective citation navigation system. These recommendations would enable a project like CiteSeer to employ a continuous improvement approach to using their citation graphs to improve the quality of the body of knowledge that they represent. As a result, the results from the proposed research have potentially wide impact on the academic community in any given discipline, helping it improve the quality of its collective research and better bridge the chasm between academics and practitioners.

Summary

The purpose of this chapter was to provide a review of the literature detailing four conceptual areas that will be critical for understanding and interpreting the citation network in quality management. First, two existing characterizations of the quality management body of knowledge (BOK) were presented to explain the concepts that quality management encompasses. Second, an outline of the mathematical foundations for network analysis of complex dynamical systems was then presented as a basis for citation network analysis. Third, the theory behind how citation networks represent knowledge flows was presented to explain the link between bodies of knowledge, citation networks, and communication patterns in the academic literature. This provided the required background for understanding research on citation networks in other fields, which has established that citation networks are typically sparse, scale-free (with characteristic exponents between 2 and 3), and exhibit “small-world” characteristics. Fourth, the theoretical foundations for continuous improvement in the domain of knowledge management were described. This was done to ascertain a basis for applying the metrics from a citation network analysis to continually improving a body of knowledge.

CHAPTER 3

METHODOLOGY

This chapter details the methods that will be used to model knowledge flows in quality management, using citation networks, to determine how the discipline has grown and evolved since 1993 and how it can be continually improved in the future. The methods to be applied to satisfy these research questions include topology analysis, centrality analysis, and regression analysis. The research questions that will be investigated are:

RQ1. What does the citation network of articles in the *QMJ* reveal about knowledge flows in the study of quality management and the maturity of the discipline?

RQ2. What are the most influential resources in the *QMJ* citation network?

RQ3. Can the mean quality of the network $<\eta>$ or attachment-weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics at that time? Can $<\eta>$ or $<\eta>\zeta$ be predicted using only the exponent of the degree distribution at that time, γ_t?

RQ4. Are there significant differences in the mean quality of articles in *QMJ* between 1993-2000 and 2000-2008 in terms of either $<\eta>$ or $<\eta>\zeta$? Which measure is more appropriate as a quality indicator?

Citation Network Analysis

Mathematical Foundations

Citation network analysis is based on graph theory, a discipline that emerged in 1736 when the mathematician Euler invented the method to solve the puzzle of the seven bridges in the East Prussian town of Königsberg. The puzzle asked whether it was possible to take a walk through the city, crossing each of its seven bridges only once before returning to your starting point (cf. Gross & Yellen, 1999). Although it was a favorite pastime of the town's residents to explore this problem on their feet, Euler's proof indicated that mathematically that their efforts were in vain: only a graph where all vertices have an even-numbered degree will support an *Eulerian walk*, as this became known. Since its development by Euler, graph theory has grown to solve problems and model systems important to each of the physical and biological sciences as well as the study of technological systems.

The first published work that leveraged citation networks to understand the evolution of a discipline was by Garfield et al. (1964), who studied the history of research on the topic of DNA. Through his analysis, he uncovered a strong relationship between historical accounts of scientific progress in understanding DNA, and the results from the citation network analysis. This study provided the basis for others to explore further applications of the methodology, and to apply citation network analysis to understand the structure and growth of knowledge. (e.g. Garner, 1965; de Solla Price, 1965; Redner, 1998; Vazquez, 2001; Bilke & Peterson, 2001; Zhuge, 2006) These include analyses of centrality (Polites & Watson, 2008) and exploring trends in network data over time (e.g. Fowler & Jeon, 2008), which extend the method of citation network analysis.

Characteristics of a Citation Network

In a citation network, the nodes (or vertices) of the network are published references (typically journal articles, books, or news/magazine articles) and the edges (or connections) between the references are citations from one paper to another. The network is *directed*, meaning that the more recent publication references the prior publication. The network grows by the addition of new nodes at each time $t_0 + \Delta t$. These new nodes preferentially attach to the pre-existing nodes that have been used to inform the new study; the preference is based on a measure of *fitness* or *utility*, which is presumed to be proportional to the number of citations the prior reference has already received.

At the same time, the new node draws additional new nodes into the network. These new nodes are references that have not previously been cited by other nodes already contained within the network. For any node within the network, the in-degree k_{in} is the number of citations that the node has received. The out-degree k_{out} is the number of citations the node makes to other references. The age of a pre-existing node is considered to be an influence in the probability that a new node connects to that pre-existing node; it is less likely that a researcher is aware of the newest research results in contrast with the most established papers in the field. (Hajra, 2004)

The network grows, but does not tend to decay; once a paper is published and its citations are fixed, it is unlikely that the paper is removed from the network and it is also unlikely that citations or added or deleted. This is in contrast to the World Wide Web, where the addition and deletion of links is commonplace. New edges do not form spontaneously between old papers, and new connections emerge only between newly added nodes, or between newly added nodes and old nodes. The direction of an edge is rigorously determined by the ages of the connecting nodes, so "one may [typically] forget about directedness" in topological analysis. (Dorogovtsev, 2002) The number of edges typically increases nonlinearly with the number of nodes over time, a symptom of network densification.

The network is not infinite dimensional, like a biological network. It is confined to the finite body of published and cited works, none of which are unknown to the analysis. In a citation network, there are no loops (self-references), multiple edges or weights (although a publication can cite the same reference repeatedly within the same article, this is not considered to give the reference additional weight). The network is immune to failure, attack or breakdown since node removal is atypical.

It is also important to remember that a citation network is a *tracer* for a dynamical system of knowledge flows, and does not represent the dynamical system itself which is more complicated. Actual knowledge flows can be achieved through personal communication, e-mail, or other methods that are less formal than publications. In addition, knowledge can be transferred not only explicitly through words and pictures, but also tacitly through mentorship and experience. (Nonaka & Takeuchi, 1995)

Despite these limitations, the citation network can provide a useful mathematical model for diagnosing the knowledge flows within and between disciplines. It is a representation that objectively characterizes the dynamics of relationships between people and their information so that these relationships can be explored quantitatively. Because measurement is critical for the continuous improvement of processes, understanding how citation network analysis can be used to measure knowledge flows is essential for understanding how those systems can be collectively improved.

Research Design

Description of the Data

All 261 full-length articles that appeared in the *Quality Management Journal (QMJ)* between October 1993 and October 2008 will be collected and used as source data. *QMJ* was selected because it is the only U.S.-based publication that is a) solely focused on quality management, and b) employs a double-blind peer review process to verify the integrity of submitted articles. If the study is to accurately characterize the discipline of quality management, it is critical to ensure that high-quality fact based articles form the core of the network rather than opinion pieces. A total of 133 discussions, redactions to discussions, and dissertation abstracts which were also published in *QMJ* during this time period will not be included, due to differences in the vetting criteria for these articles as compared to the full-length articles.

Using these 261 articles and the 9,433 citations made by these articles as source data, a citation network will be constructed to reflect the relationships between the articles published and the works they reference. In a citation network, any artifacts supporting the flow of knowledge can be nodes. This includes journal articles, books, magazine articles, web pages, television programs, personal communication, and so forth. Edges in a citation network represent a flow of knowledge from an earlier source to the more recent publication. Each of the 261 source articles will themselves become nodes in the network that will be constructed for this research project. A reference that is cited by an article in this sample will also become a node, and an edge will be constructed from the original article to its reference to indicate a citation relationship. The result is a *directed network* in which a time dependence from A to B, where A has been published prior to B, is expected. Each node in the network is treated as a single, indivisible object.

Data Collection and Data Reduction

The data was gathered from the ASQ Quality Information Center (QIC), an online repository where the full texts of all back issues of the Quality Management Journal are archived. The citations from each article were stored in a single text file, labeled with the ASQ-assigned QIC identifier (QICID) as the filename. The format was standardized to remove incomplete references (where not enough information is specified by the author to find the source that is being cited), carriage returns, and line feeds which may interfere with database creation. A Perl program was written to extract the information in each of the 261 files generated into a relational database with one table to store the nodes and attributes, and a second table to store the relationships between nodes. A second Perl program was written to create the attributes files and relations files that will be import into the R statistical software package for analysis. A fuzzy search was used to avoid duplication of nodes, for example, when different editions of books are referenced. A diagram illustrating the flow of data collection and reduction is shown in Figure 6.

Restatement of the Research Questions

This dissertation examines the topology of the citation network in quality management, and using this information, formulates measures that can be used to continually improve the network based on its topology. The first research question defines the topology of the citation network and relates it to an appropriate growth process. The second uses node centrality information (both degree centrality and eigenvector-based centrality measures) to determine which resources are the most influential. The final two research questions use the topology and centrality information to define quality measures on the network, explore the evolution of those quality measures, and determine whether the quality of the citation network of *QMJ* articles changed over time. From this, the study will determine whether the quality measures could be effective as measures to continually improve the citation network over time. The research questions and the sections of the study that they address (topology, evolution/centrality, and network-based continuous improvement) are fully outlined in Table 5. Finally, the research methodologies selected to answer each question are outlined in Table 6. The steps in network analysis are shown in Figure 7.

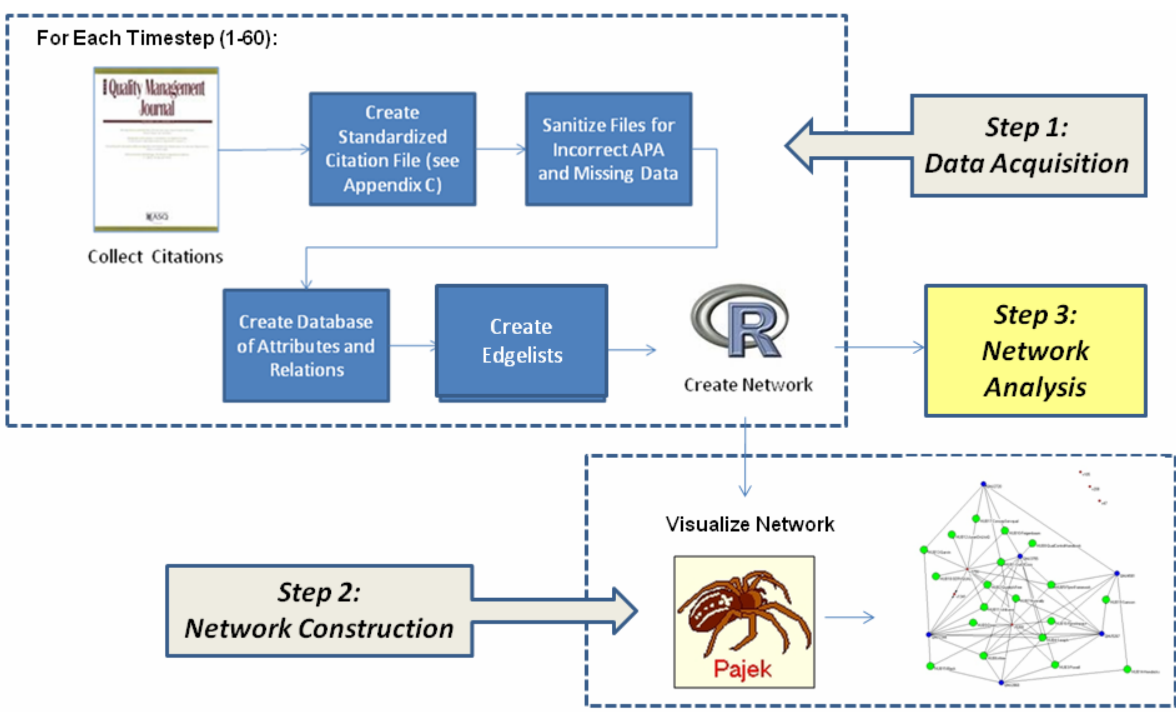


Figure 6. Data acquisition, network construction & network analysis methodology.

Table 5

Summary of Research Questions and Hypotheses

Section of Study	Description
Topology	RQ1: What does the citation network of articles in the *QMJ* reveal about knowledge flows in the study of quality management and the maturity of the discipline?
	Hypothesis 1a: The citation network is random.
	Hypothesis 1b: The citation network is scale-free.
	Hypothesis 1c: The citation network is a small world.
Evolution/ Centrality	RQ2: What are the most influential resources in the *QMJ* citation network?
Network-Based Continuous Improvement	RQ3: Can the mean quality of the network $<\eta>$ or attachment weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics at that time? Can $<\eta>$ or $<\eta>\zeta$ be predicted using only the exponent of the degree distribution at that time, γ_t?
	RQ4: Are there significant differences in the mean quality of articles in *QMJ* between 1993-2000 and 2000-2008 in terms of either $<\eta>$ or $<\eta>\zeta$? Which measure is more appropriate as a quality indicator?

Table 6

Summary of Research Methodologies

RQ	Methodology
RQ1	Network Analysis
Hypothesis 1a	*Shapiro-Wilk Test for Normality*
Hypothesis 1b	*Kolmogorov-Smirnov Goodness of Fit*
Hypothesis 1c	*Small-World Heuristic Test*
RQ2	Centrality Analysis & Concept Analysis
RQ3	Multiple Regression/Linear Regression
RQ4	t-test

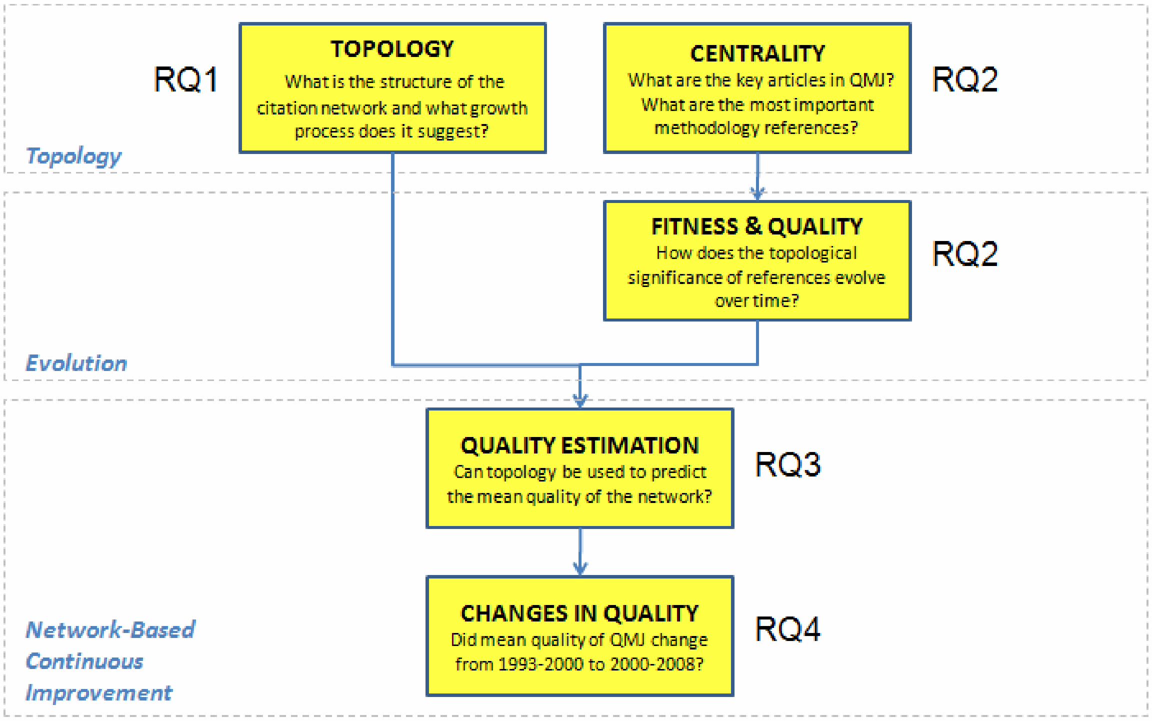


Figure 7. Stages in network analysis and relationship to research questions.

Procedures

The procedures for conducting this study can be broadly classified into three phases: data acquisition, network construction, and network analysis. A diagram summarizing the methods used to acquire data and prepare the networks for analysis is presented in Figure 6. Figure 7 includes additional details regarding how the network analysis relates to the research questions.

After the network is constructed, the topological characteristics of the quality management citation network are determined from the citation data. This step is similar to examining descriptive statistics on a dataset prior to a traditional statistical analysis. A total of 11 measures will be calculated, which are outlined and described in Table 7. These values all reflect global properties of the network, rather than local properties that reflect characteristics of the network relative to the nodes, and are typically used for intercomparisons of citation networks derived from different datasets. (Batagelj, 2003; Watts & Strogatz, 1998) The notable deviation from this characteristic is the clustering coefficient. Although measured at the node level, only the mean clustering coefficient across nodes is considered as an estimate of clustering throughout the network.

Table 7

Descriptions of Topological Measures

Measure	Description
Number of Nodes	The total number of articles in the network.
Number of Edges	The total number of references between articles in the network.
Average/Min/Max In-Degree	The average, minimum, and maximum number of citations to nodes within the network.
Average/Min/Max Out-Degree	The average, minimum, and maximum number of citations from nodes to other nodes within the network.
Number of Clusters	The total number of clusters (inclusive of the largest cluster).
Size of Largest Component	The number of nodes in the largest cluster in the network.
% of Isolates	The proportion of articles within the network that are not cited by any other articles within the network.
Clustering Coefficient	The probability that two nodes connected to a third node are also connected to each other.
Network Density	The ratio of the number of edges observed in a network, and the number of edges that would be observed in a random network with the same number of nodes and a connection probability of 1
Network Diameter	The largest number of hops required to connect any two nodes.
Average Path Length	The average number of hops between any two nodes in the network.

Once the topological characteristics of the network are extracted, they can be inspected by comparing the network to an equivalent random network. The equivalent random network is constructed by forming a network with the same number of nodes as the original network, and a connection probability equal to the ratio of the average degree and the total number of nodes.

The topological analysis will include examining the structure of the full network of nodes and edges, constructed from all source data from 1993 to 2008, in terms of measures that are commonly used in citation network analysis. Then, the analysis will involve 1) determining whether the network is random, 2) determining whether the network is scale-free, that is, whether it is reasonable to fit the degree distribution using a power law (Clauset et al. 2007), and 3) determining whether the network exhibits the "small world property" of average path length comparable to an equivalent random network, but clustering coefficient much greater than a comparable random network. (Watts & Strogatz, 1998) Each of these steps is necessary to completely characterize the topology of the citation network in quality management.

The procedures associated with testing hypotheses for the research questions will now be described. To satisfy RQ1, three tests will be applied: the Shapiro-Wilk test of normality, a Kolmogorov-Smirnov goodness of fit test, and a heuristic test. For Hypothesis 1a:

H_{01A}: The citation network is random. The degree distribution for the network is normal around the mean in-degree $<k_{in}>$.

H_{A1A}: The citation network is not random. The degree distribution for the network is normal around the mean in-degree $<k_{in}>$.

T.S.$_{1A}$: The Shapiro-Wilk test of normality will be used. If the Shapiro-Wilk D statistic exceeds the critical value, the null hypothesis will be rejected.

For Hypothesis 1b:

H_{01B}: The citation network is scale-free. The degree distribution for the network is a power law with exponent γ.

H_{A1B}: The citation network is not scale-free. The degree distribution for the network is not a power law with exponent γ.

T.S.$_{1B}$ The Kolmogorov-Smirnov goodness of fit test will be used. If the D statistic exceeds the critical value, the null hypothesis will be rejected.

For Hypothesis 1c:

H_{01C}: The citation network is a small world.

H_{A1C}: The citation network is not a small world.

T.S.$_{1C}$ The small world heuristic test will be used. If $L/L_r > 2$ or $C/C_r < 5$ then the null hypothesis will be rejected.

To satisfy RQ2, determining the most influential resources in the *QMJ* citation network, three measures of centrality will be extracted for each node: degree centrality (the number of in-degrees), Google PageRank, and Kleinberg's hub and authority scores. The following questions will be answered using the metric in brackets:

2a. What are the Top 20 most frequently cited references in the *QMJ* citation network? *[Degree Centrality; Number of In-Degrees]*

2b. What are the Top 20 most central references in the *QMJ* citation network? *[Google PageRank]*

2c. What are the Top 20 most authoritative references in the *QMJ* citation network?

[Kleinberg's Authority Score]

2d. What are the Top 20 hubs/review articles published in *QMJ*?

[Kleinberg's Hub Score]

2e. What academic journals provided the most references for *QMJ* articles?

[Degree Centrality; Number of In-Degrees when Nodes Classified by Journal]

2f. What methodology resources were most frequently referenced by *QMJ* articles?

[Degree Centrality; Number of In-Degrees]

To satisfy RQ3, determining whether the mean quality of the network can be determined in terms of $<\eta>$ or $<\eta>\zeta$, multiple regression will be applied. The statistical significance of each of the predictor variables, and the ability of the predictive model to explain the variation in the data, will be used as indicators of the quality of the model. For Hypothesis 3:

H_{03}: The regression model is appropriate.

H_{A3}: The regression model is not appropriate.

$T.S._3$: If the F statistic exceeds the critical value, the null hypothesis will be rejected.

To satisfy RQ4, determining the most influential resources in the *QMJ* citation network, a t test will be applied to identify whether there are statistically significant differences between the mean quality scores between 1993-2000 and 2001-2008. For Hypothesis 4:

H_{04}: $\mu_1 - \mu_2 = 0$

H_{A4}: $\mu_1 - \mu_2 \neq 0$

T.S.$_4$: If the t statistic exceeds the critical value, the null hypothesis will be rejected.

For most of the analysis, Type I error will be set at $\alpha = 0.05$. Albert & Barabasi (1999) noted that the Type I error of 0.05 is typically used for this type of research. While this alpha gives a higher probability of Type I error than incorrectly accepting a false alternative hypothesis with an alpha of 0.01, the results of this study are informational only and the greatest expected impact is to determine future directions of research. The exception to this choice is that for evaluating the goodness of fit for power law models, the Type I error will be set to $\alpha = 0.10$ and examined in terms of the level of significance that corresponds to the Kolmogorov-Smirnov D statistic from the fit. The rationale for this choice is that the citation network is being constructed from its earliest origins in the first issue of the *QMJ*, and it is expected that the observed degree distribution will not immediately conform to a model. For broadly assessing what constitutes a good fit at each timestep, it is reasonable for the purposes of this study to accept the potential to be incorrect one out of ten times. There is not a high cost associated with data collection or the decisions that could be made as a result of this study, and there is no risk in implementing faulty recommendations. Because the risk involved in a Type I error is small, reducing the Type I error to 0.01 for any of the evaluations is not necessary.

Data Analysis Approach

The R statistical software was used to perform the network analyses and hypothesis tests. (R Development Core Team, 2008) The `netmodels` package in R was employed to determine the basic topological characteristics of the networks, calculate measures for equivalent random networks, and do small world tests. The `igraph` package was used to compute and compare centrality scores. The base R package was used for regressions and to plot and analyze time series of network measures to explore the evolution of the citation network. Pajek (Batagelj, 2003) was used for network visualization.

Summary

This chapter explained the basis for the method of citation network analysis, described the research dataset, and presented the approach that will be used to collect, parse, store, reduce and examine the data. Next, each of the research questions were examined sequentially, and the qualitative and quantitative methods that will be used to answer those questions were described in depth. The software to be used and the Type I errors that will be used for hypothesis testing were then noted.

CHAPTER 4

RESULTS

The emphasis of this study was to model the evolution of knowledge flows from a citation network in quality management, derived from the references made by articles in the *Quality Management Journal (QMJ)* from its first issue in 1993 through 2008. As a consequence of using this model, this study aimed to identify the most influential resources, and determine whether the quality of the network can be predicted from its topology. Finally, this research determined whether there was a statistically significant change in quality in the *QMJ* citation network using these measures between 1993-2000 and 2000-2008.

This chapter presents the results of the data analysis procedures in the following five sections: 1) discussion of the data, 2) topology and evolution, 3) centrality and fitness, 4) quality estimation, and 5) changes in quality. The first section provides a discussion of the data, and reviews the data collection procedures, assumptions and variables used. The second section presents the *QMJ* citation network and the evolution of its topological parameters throughout the period of interest. The third section establishes the centrality of articles within the network in terms of multiple centrality measures, and also includes a summary of the most frequently cited journals, the proportion of academic references to non-academic references, and the mean fitness of the network over time. The fourth section examines whether the quality of the network can be estimated from the network's topology. The fifth section determines whether there is an observable quality improvement in the *QMJ* citation network between 1993-2000 and 2000-2008. Finally, the last section provides a summary of the key findings presented in this chapter.

Discussion of the Data

The following section describes the process for data collection and quality control, describes variable names, and outlines assumptions associated with the analysis.

Data Collection and Quality Control

The main factors of interest in this study were the structure and evolution of the citation network constructed from all 261 source articles in the *QMJ*. The citation data was collected from the American Society for Quality (ASQ) Quality Information Center (QIC) and stored in citation files as ASCII text. The format of these files was standardized to aid in analysis (an example is shown in Appendix C). An interactive Perl program, `process.pl` (see Appendix D) was written to automate the process of assimilating data into the network, adding new edges, and ensuring that duplicate nodes were not introduced.

Each of the new nodes added to the network was inspected manually to avoid unintended duplication of nodes, as well as to provide overall quality control for data collection. This was necessary because there were several errors routinely observed in the *QMJ* citations, including: incorrect author order, author last names spelled incorrectly, missing coauthors, wrong year recorded for citation, title paraphrased or only partially recorded (e.g. "TQM" and "Quality Management" often juxtaposed incorrectly), confusion between US and UK English in titles, volume of journal incorrect, and lack of uniformity referring to award criteria and standards (ie. Baldrige, EFQM, ISO/ANSI).

When the construction of the master node and edge lists (one for each of the 60 timesteps) was completed, the data was then transferred to the statistical data analysis package R for further work. Several R scripts were written to compute the topological measures and time series that were examined to satisfy the research questions for this study. The R scripts used for data analysis are available (see Appendix E). The complete data sets and instructions for retrieval are also available online.

Variable Names and Codes

To investigate the topology, evolution and centrality of nodes within a network, several variables must be calculated and evaluated simultaneously. The variables and symbols used in this study, and their values in the QMJ citation network, are shown in Table 8. Note that the same variables are used for the static topological analysis of the quality management citation network and the investigation of the evolution of the system. The primary difference is that for the dynamical analysis, the topological characteristics are computed at several times throughout the growth of the network and the resultant time series are then evaluated.

Assumptions and Limitations of Analysis

Citation network analysis is affected by the same challenges that impact traditional citation analysis. These include conflicting results due to slight differences in citation practices between authors and journals, inappropriate citations (e.g. authors attempting to cite likely reviewers to increase the chance that their article is accepted), inappropriately citing foundational references simply to acknowledge their existence (and

Table 8

Variable and Values for Topological Analysis of the QMJ Citation Network

Name	Value	Description
n	7091	Number of nodes in the network
e	9433	Number of edges in the network
k	N/A	Number of degrees (sum of edges) connecting to a node; k_{in} denotes that only incoming connections are counted, k_{out} indicates that only outgoing connections are counted
$\langle k_{in} \rangle$	1.369	Average number of incoming degrees incident on a node
$\langle k_{out} \rangle$	N/A	Average number of connections a node makes with other nodes
$P(k)$		Probability of a node having exactly k degrees
γ	2.96	Exponent in power-law fit to the degree distribution: $P(k) \sim k^{-\gamma}$
L	4.987	Average path length across the network
L_{er}	8.949	Average path length across an equivalent random network
C	0.00396	Clustering coefficient of the network
C_{er}	0.00036	Clustering coefficient of an equivalent random network
c_i	38	Number of isolated nodes
c_n	7029	Number of nodes in the primary connected cluster
$c_{\%}$	99.1%	Percentage of nodes in the primary connected cluster
d	6	Diameter (mean geodesic distance) across the network
ρ	0.000188	Density of a network (ratio of total edges to total possible edges)
t_0	Jan 1993	The first time network data was collected
Δt	18 mos.	The time increment between observations

not to establish a relationship with the new research), and time lags created by the disparity in review processes between popular media outlets, practitioner journals, and academic journals. (Biehl et al., 2006)

The primary assumption made by the current research is that the methods applied by previous researchers, and the conclusions derived from those methods, are legitimate. This implies that the results from the present study can be compared to preexisting studies of other complex dynamical systems. Because the field of dynamical systems analysis using networks is less than a decade old, it is possible that future insights from mathematics and statistics might invalidate some of the techniques used herein.

There are three primary limitations for the topological analysis: falsification, methodological biases, and sampling biases. First, it is not possible to definitively establish which dynamical process has produced a network, only to eliminate which processes did *not* produce the network. That is, potential models can be excluded but this study cannot definitively establish which dynamical process actually generated a network. For example, even though a power-law fit may have a high statistical significance, this does not mean that an exponential or log-normal fit are precluded.

Second, there are established methodological biases that may impact the results, most importantly regarding the determination of the power law exponents from degree distributions. Clauset et al. (2007) established that the traditional practice of estimating the power law exponent by fitting a linear model to a log-log plot of $P(k)$ versus k was so flawed that a statistically significant power law relationship could be identified even if there was little or no support for goodness of fit. These investigators established that using maximum likelihood estimation (and employing a cutoff value on the x-axis of

x_{min}) was not only more effective, but it also removed the systematic underestimation of the power law exponents by regression for datasets where a power law was indeed observed. The limitation of the maximum likelihood method is that unlike a regression-based or Bayesian approach, there is not enough information to construct confidence intervals around the power law exponent that is determined, and so confidence must be determined using the Kolmogorov-Smirnov goodness of fit test.

Finally, there are also sampling biases that can impact the effectiveness of a network analysis. The “missing past” problem, where data about the genesis of the network is not available and must be inferred (Leskovec et al., 2003), is not an issue for this study since all the data from the inception of the *QMJ* to present is available. Some topological measures may actually be artifacts of the measurement and network growth process that is used to construct the model of the system. For example, when topological properties of simulated networks are evaluated, low variability profiles can lead to high sensitivity in the resultant topological measures. As a result, the statistical significance of tests involving topological measures should be cautiously interpreted. (Airoldi & Carley, 2005)

Topology and Evolution

The purpose of this section is to present results of the topological analysis of the *QMJ* citation network. This involves two phases: *network characterization*, which establishes the structural features of citation networks in general and the specific features of the observed citation network in quality management, and *time evolution of topological characteristics*, which involves interpreting the shifts in and relationships between topological characteristics as the network developed.

Network Characterization

To characterize the topology of the citation network in quality management, the full set of nodes and edges representing all articles in the *Quality Management Journal* from 1993 through 2008, and the references used by those articles, were examined. The next three subsections present 1) the general characteristics of the citation network, followed by 2) the specific results from the observed citation relationships within quality management, and finally 3) a comparative analysis exploring how the observed network compares to other networks that have been observed in sociology, biology, and technology.

Quality Management Journal Citation Network

The *QMJ* citation network is the complete set of nodes and edges, expressed as a network, for all of the *QMJ* articles and the references they cited from October 1993 to October 2008. This network has 7091 nodes and 9433 edges. The topological characteristics of the citation network were summarized in Table 8 above. A complete visualization of the network, with all connected nodes and edges displayed, is shown in Figure 8. A total of 62 nodes are isolated and not connected to the primary cluster.

Degree Distribution

Each node in the *QMJ* citation network has been cited by other articles within the network a certain integer number of times. This is called k_{in}, or the number of in-degrees. Examining the degree distribution of a network can provide insight into its topological structure, because a random network is typically characterized by a Poisson degree distribution about a peak at the mean in-degree, $<k_{in}>$, whereas a network that is not random typically displays a power law or exponentially decaying degree distribution. The

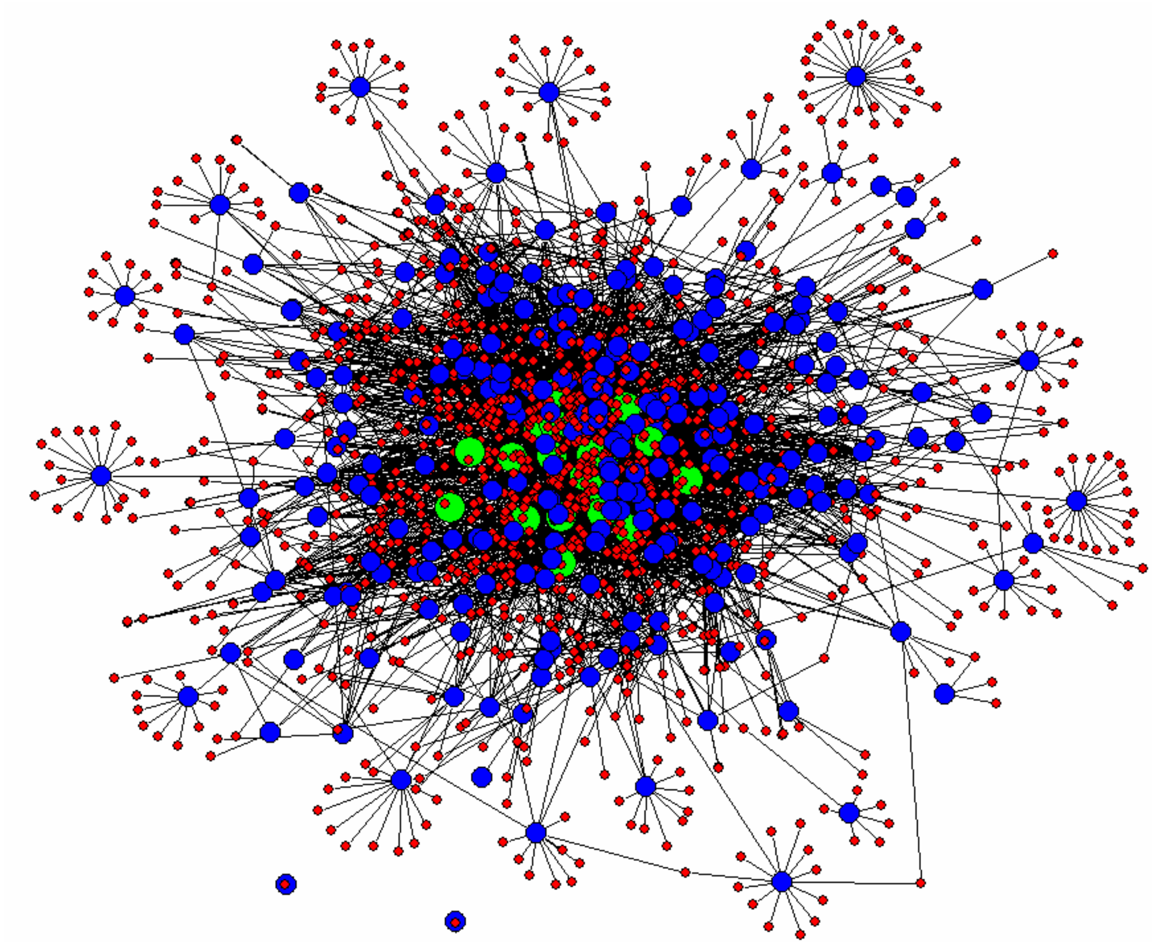


Figure 8. The Quality Management Journal citation network .

degree distribution for the 7091 node *QMJ* citation network is displayed in terms of rank is shown in Figure 5, and in terms of the number of degrees k and the observed frequency of that degree P(k) in Figure 6. Both appear linear on log-log plots, suggesting that a power law fit may be appropriate. To determine a candidate model for the degree distribution, a Poisson distribution about the mean, power laws for the full data set, and power laws for data sets truncated on either the high or low end of k are applied to investigate these distributions. The results from these fits can be used to determine whether a model is feasible, but not whether it definitively represents the data.

Topological Analysis

The first two research questions, which explore the observed degree distributions, were constructed to examine the global structural characteristics of the citation network (at the level of the whole graph). Three sub-hypotheses supporting Research Question 1 (RQ1) were also constructed to examine specific structural characteristics. RQ1 asked: What does the citation network reveal about knowledge flows in the study of quality management? The hypotheses that were examined to explore this research question were:

Hypothesis 1a: The citation network is random.

Hypothesis 1b: The citation network is scale-free.

Hypothesis 1c: The citation network exhibits the small world property.

The first two hypotheses were investigated using the degree distribution for the network, which is a plot of the probability $P(k)$ that a node has exactly k incoming citations (or in-degrees). This was illustrated in Figure 9. To test Hypothesis 1a, a Shapiro-Wilk test of normality was conducted to determine whether the citation network was random, because in random networks, the degree distribution is normally distributed about a peak at $\langle k \rangle$. The Shapiro-Wilk test was selected instead of the Kolmogorov-Smirnov goodness of fit test because it is designed to test for normality, and thus has a higher power as a statistical test for this particular distribution. The results from the Shapiro-Wilk test indicated that the null hypothesis should be rejected at the $\alpha = 0.05$ level in favor of the alternative ($W = 0.1135$; $p < 0.001$). Therefore it can be concluded, based on the sample results, that the degree distribution cannot be fit by a Poisson model and thus the observed citation network is not consistent with a random network.

To test Hypothesis 1b, a power-law was fit to the data using two techniques. These fits are displayed in Figure 10. First, a linear fit to the log-log representation of the

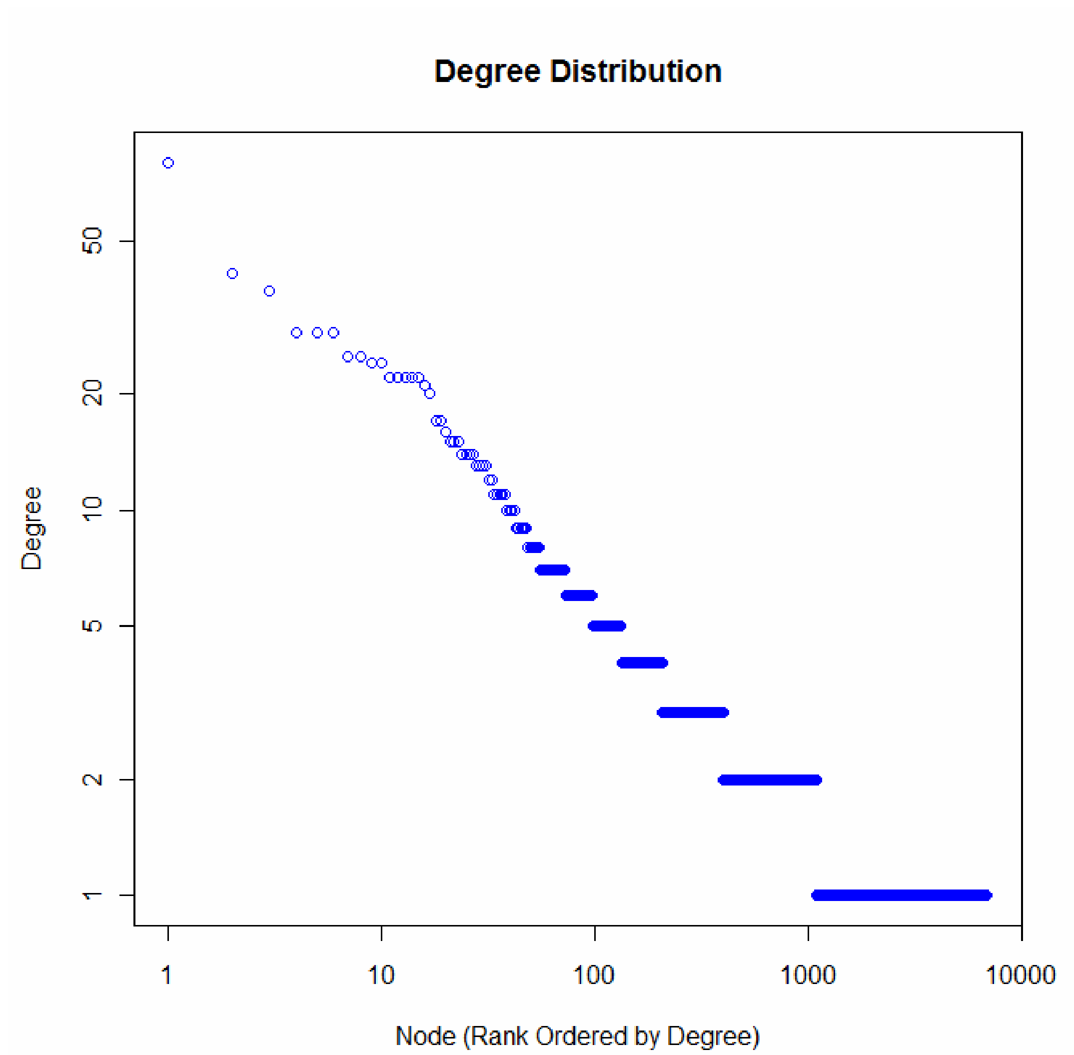


Figure 9. The rank degree distribution of the *QMJ* citation network on a log-log scale.

degree distribution, a technique consistent with most other networks examined over the past decade, was applied. Using all of the data points, an exponent of $\gamma = 2.07$ was determined ($D = 0.132$; $p = 0.9725$; $R^2 = 0.8519$). The best result from the linear fitting approach (maximizing the correlation coefficient, R^2) was achieved by fitting to the first

20 data points on the degree distribution. This yielded an exponent of $\gamma = 2.84$ ($D = 0.193$; $p = 0.8005$; $R^2 = 0.9784$).

Second, a maximum likelihood estimation of the power law exponent was

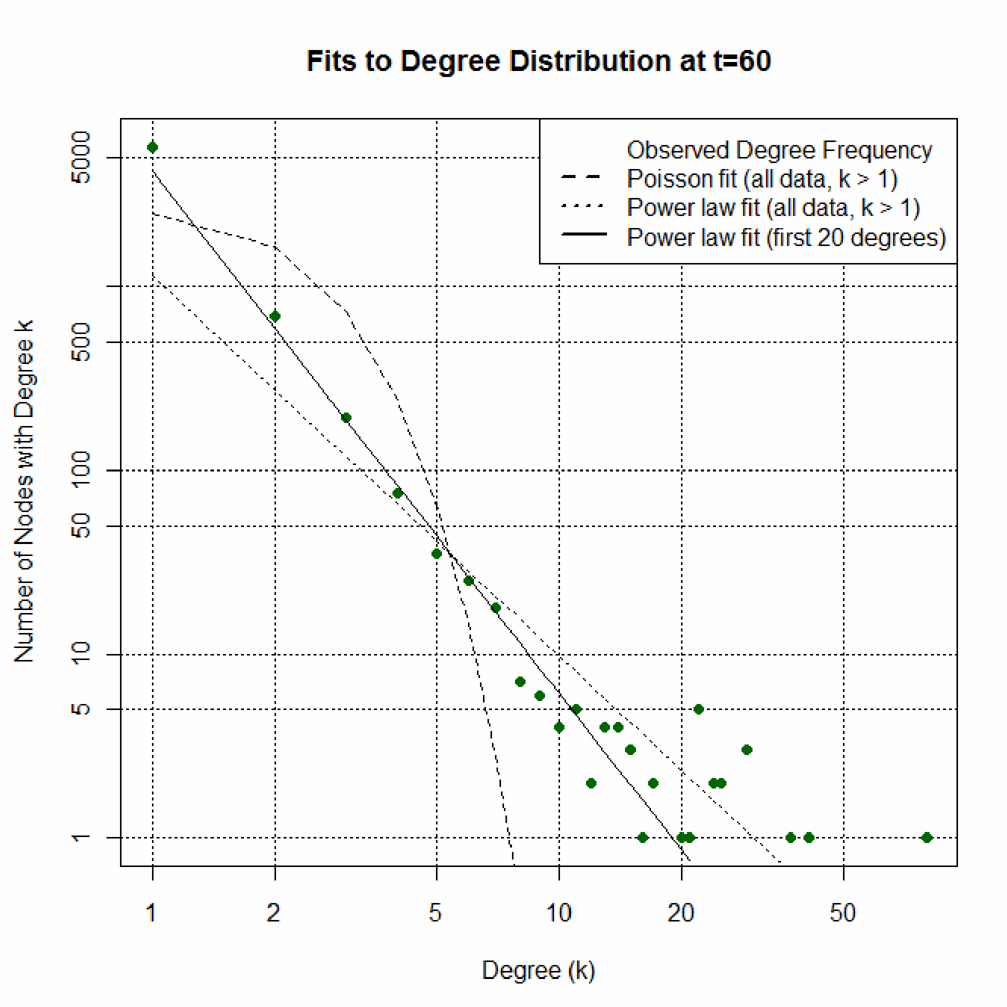


Figure 10. Power law and Poisson fits to the *QMJ* citation network degree distribution.

performed using the guidelines and R source code developed by Clauset et al. (2007). A Kolmogorov-Smirnov goodness of fit test was used in addition to examine the significance of this power law fit. Using maximum likelihood estimation, the value of the power law exponent was determined to be $\gamma = 2.96$ with a cutoff of $k_{min} = 2$ ($D = 0.194$; p

= 0.978). The results from the maximum likelihood estimation of the power law exponent indicate that the fit cannot be rejected at the $\alpha = 0.05$ level, and thus the alternative is not acceptable. Therefore, it can be concluded based on the sample results that the observed citation network is consistent with a scale-free network.

The result from the maximum likelihood estimation of the power law exponent is also consistent with Clauset's expectation that the linear fit method underestimates the exponent of the power law fit. A high p-value suggests that a power law with these characteristics may provide a reasonable fit to the data, and that the null hypothesis that the citation network is scale-free cannot be rejected. However, this test does not guarantee that the power law is the *only* candidate model for the empirical data; rather, that it is one potential model that could explain the organization of the network.

The final test performed was a heuristic test to determine whether the quality management citation network exhibits the small world property, which is the simultaneous presence of a short average path length and strong clustering. The definitions of "short" and "strong" are determined by comparing the observed network with an equivalent random network, that is, a network with the same number of nodes as the observed network, and a connection probability that assumes all observed edges are equally distributed. In a small world, the growth of the network is not random; instead, organizing principles have influenced its evolution. For this reason, the small world test provides some indication of whether there is underlying order in how the network has evolved. The small world test thus provides additional information that is not available just by examining the degree distribution.

The null hypothesis for the small world heuristic test is that the network exhibits the small world property. This test is performed by evaluating the ratio of the average path length in the network L to the path length of a comparable random network L_r. Similarly, the ratio of the clustering coefficient C in the network to the clustering coefficient of an equivalent random network C_r is examined. If $L/L_r > 2$ or $C/C_r < 5$ then the null hypothesis is rejected, and the network may be considered a small world. For the citation network from the *QMJ*, $L/L_r = 0.557$ and $C/C_r = 11$, so the null hypothesis cannot be rejected and the network can be considered a small world. Goodness of fit is not assessed using the heuristic test due to insufficient data to support its calculation.

Time Evolution of Topological Measures

To explore the growth patterns within the *QMJ* citation network, the time series constructed from the following variables or measures recorded each time a new *QMJ* was issued is examined: 1) growth of nodes and edges, 2) densification (number of edges at time t as a function of number of nodes at time t), 3) changes in the mean in-degree $<k_{in}>$, 4) changes in clustering coefficient C, 5) changes in the power law exponent γ, and 6) changes in the goodness of fit of the power law for which the exponent was derived (in terms of the Kolmogorov-Smirnov D statistic). Together, these topological measures provide insight into the time evolution of the citation network.

Growth of Nodes and Edges

The growth of nodes and edges at each timestep is illustrated in Figure 11. The number of new nodes is increasing linearly with time, as is the number of edges, indicating a process of accelerated (or nonlinear) growth. (Albert & Barabasi, 2002; Dorogovtsev & Mendes, 2001) The rate of increase is slightly higher for the edges than

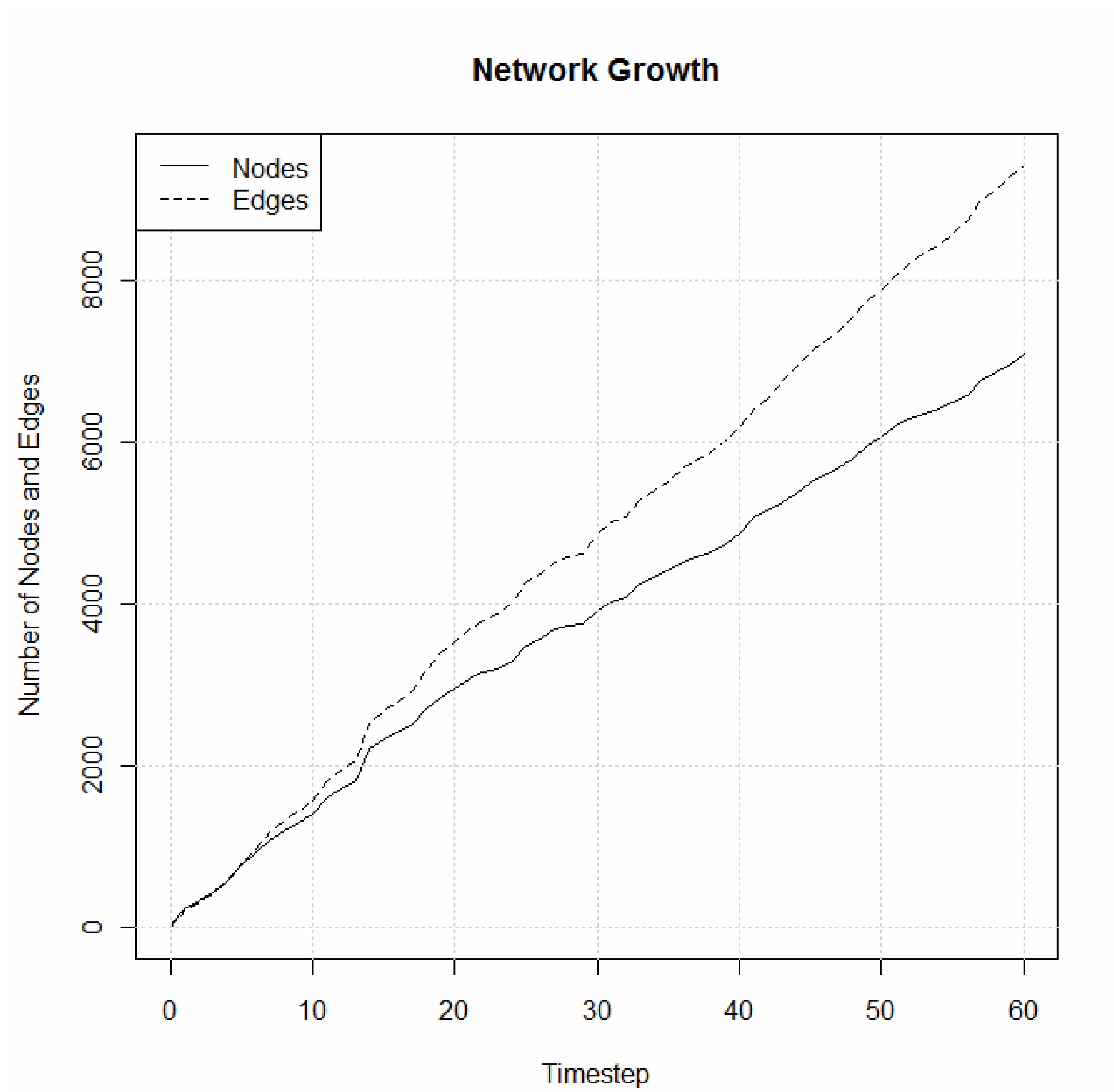


Figure 11. The number of nodes and edges in the network at the end of each timestep.

for the nodes. The same data is shown again in Figure 12, only this time on a log-log chart. The time axis is removed and the number of nodes is plotted with respect to the number of edges for each timestep in the evolution of the network. A fit to the data (Figure 12) indicates that $e(t) \propto n(t)^{\alpha}$ where the exponent $\alpha = 1.155$ ($R^2 = 0.9999$; $p < 0.001$) is remarkably consistent. This densification is also consistent with the *forest fire model* for network growth, one mechanism for the development of a citation network.

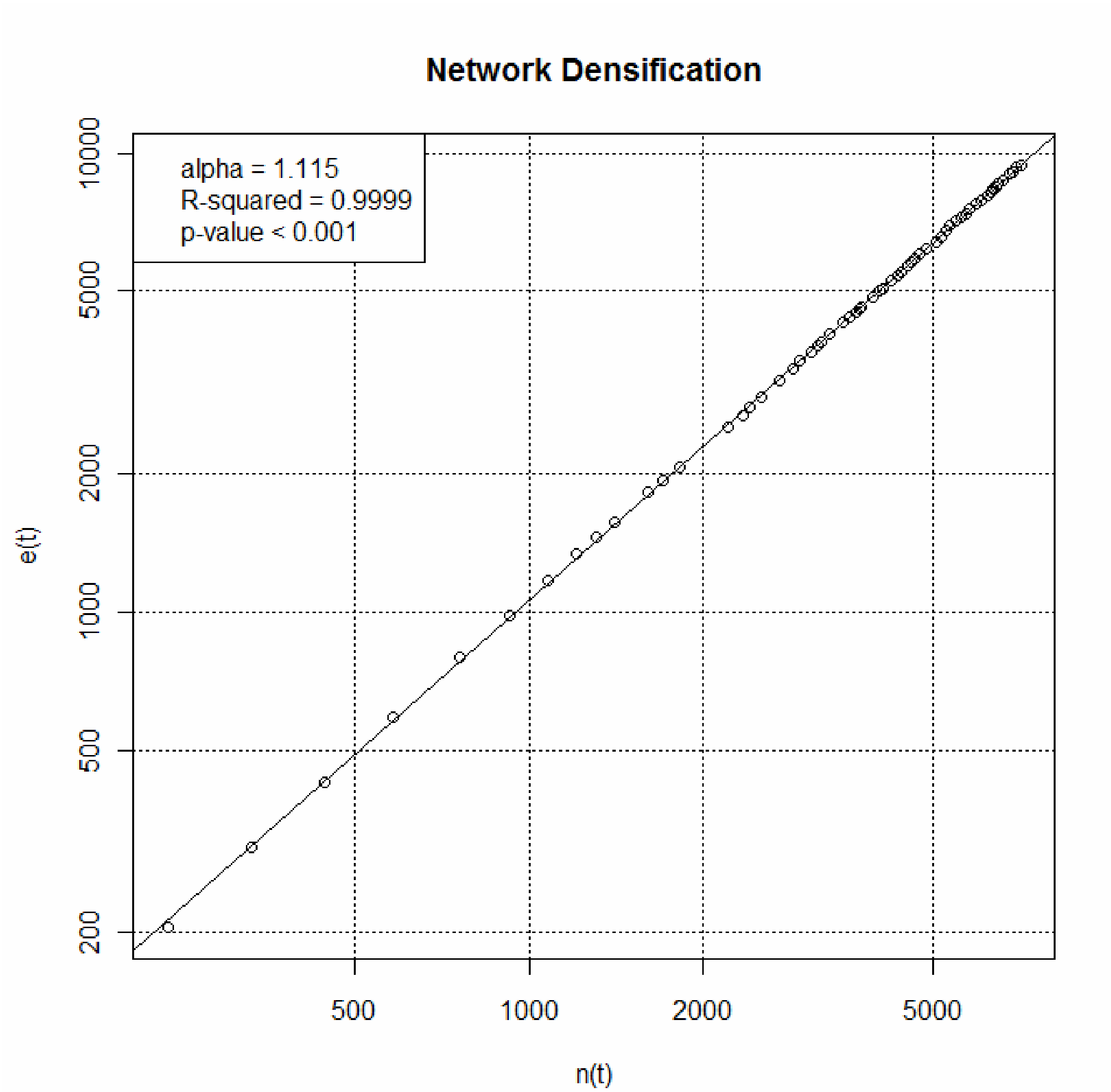


Figure 12. Log-log plot of number of edges vs. number of nodes for t=1 through t=60.

According to the forest fire growth model, the author of a new paper initially consults one of the articles within the existing citation network, finds its references, and then preferentially attaches to the references from the initial paper the author discovered. The process, described by Leskovec et al. (2006), continues until the author ends the search process and finds a new source. These authors also indicate that the mean in-degree of the network should also be increasing linearly at a constant rate.

Mean In-Degree

Figure 13 corroborates the linear increase in the mean in-degree <k_{in}> of the network. This value increases at a rate of 0.00718 degrees per timestep prior to t=24 (October 1999), and 0.00311 degrees per timestep between this point and t=60 (October 2008). The abrupt shift in the slope of the mean in-degree over time is not captured by Figures 11 or 12, and suggests a phase transition around t=24 in October 1999.

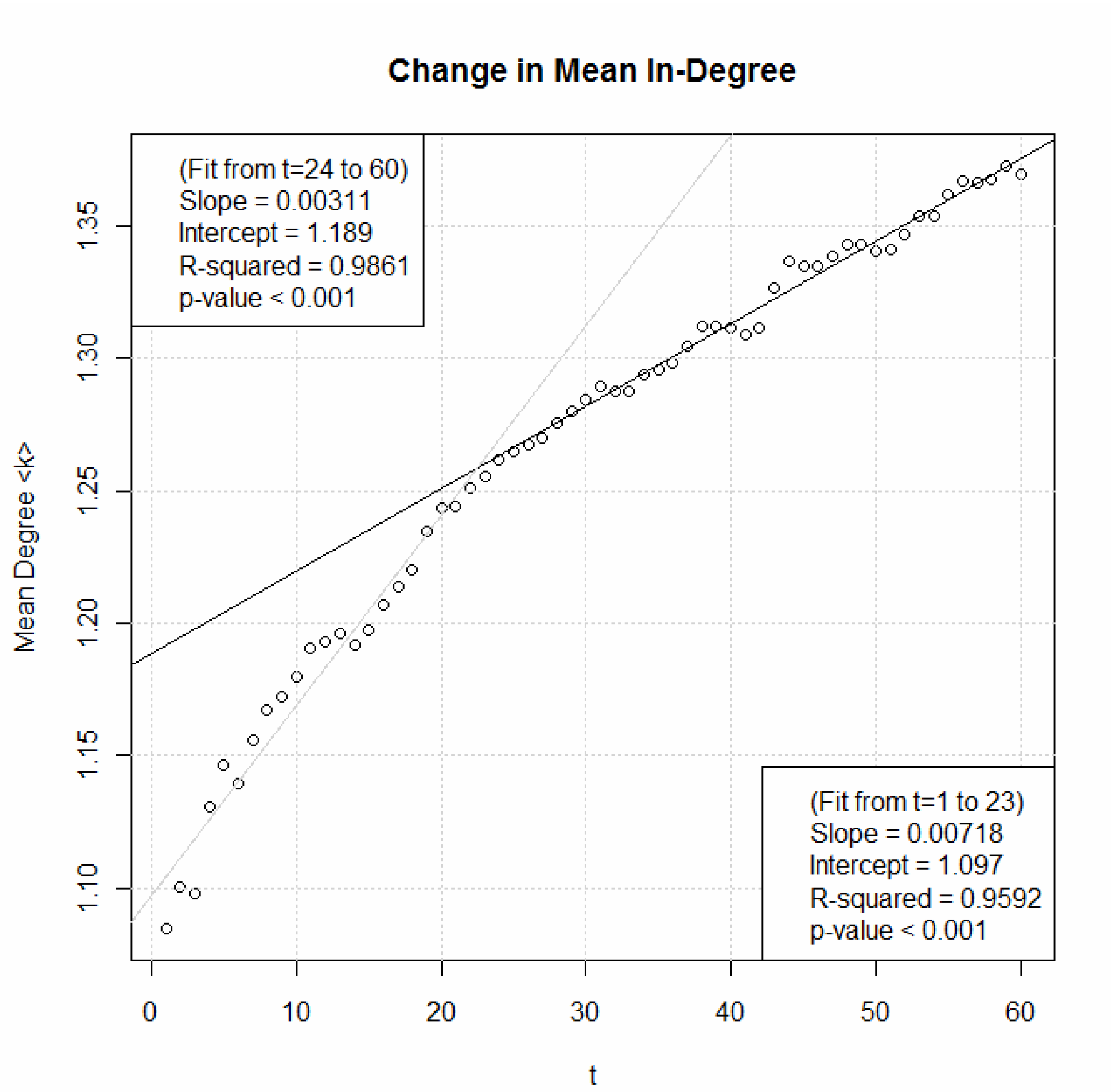


Figure 13. Evolution of the mean in-degree. Two regimes appear (t=1 to 23; t=24 to 60).

Clustering Coefficient

Although there is little clustering evident within the network at its inception, clustering rapidly increases between t=1 and t=20 and continues increasing throughout the period of observation as shown in Figure 14. The clustering plateaus around t=20 and again by t=45. Throughout its evolution, the *QMJ* citation network continues to self-organize and form communities.

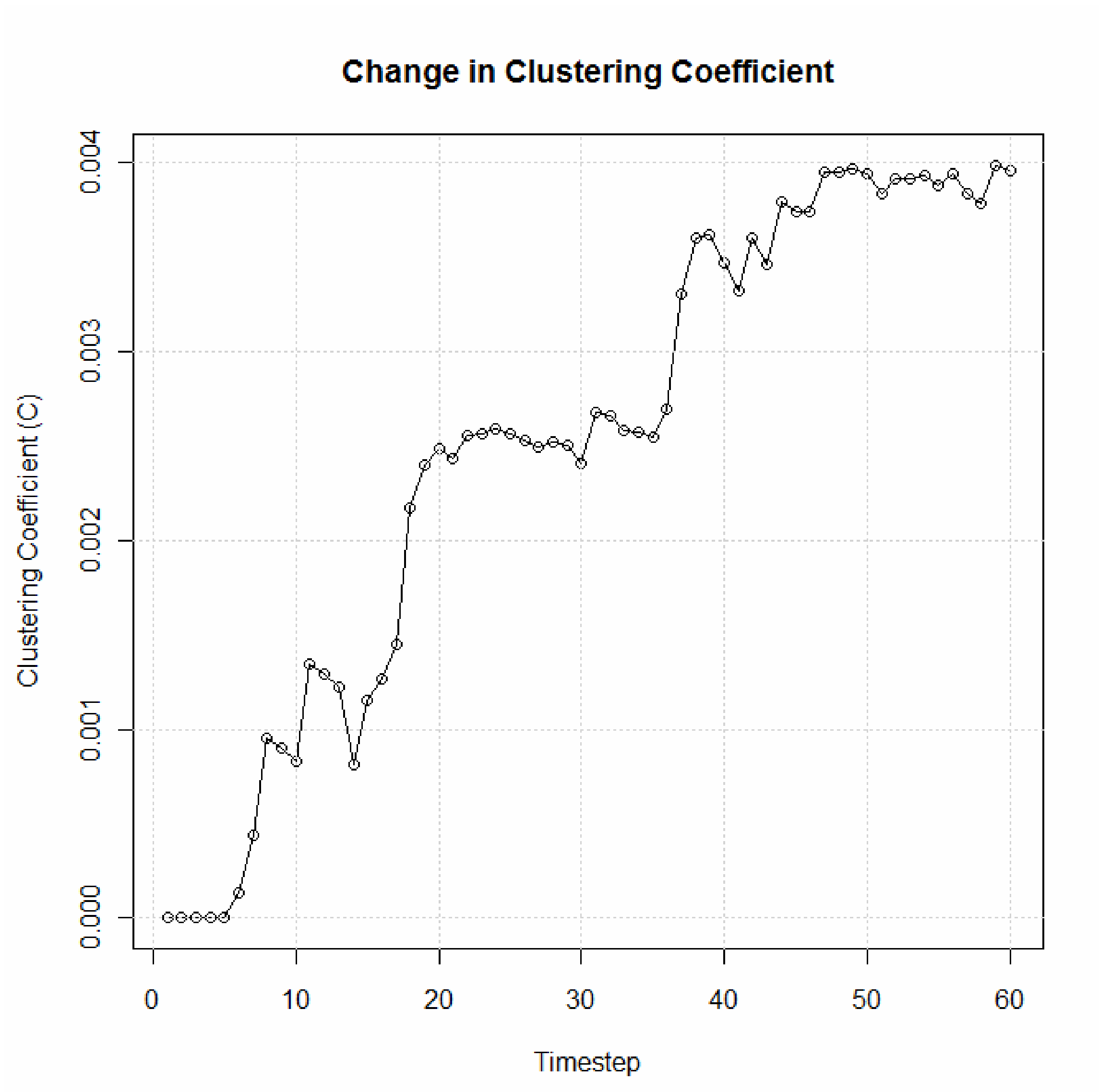


Figure 14. Growth of the average clustering coefficient in the network over time.

Average Path Length

The average path length varies significantly until t=13, then stabilizes and begins decreasing. Around t=25, the value of the average path length begins to increase linearly. Figure 15 shows the changes in the average path length. The dashed line indicates a linear fit to the data from t=25 to t=60 ($R^2 = 0.9394$, $p < 0.001$). At the end of the observed evolution, there are on average five degrees of separation between each resource.

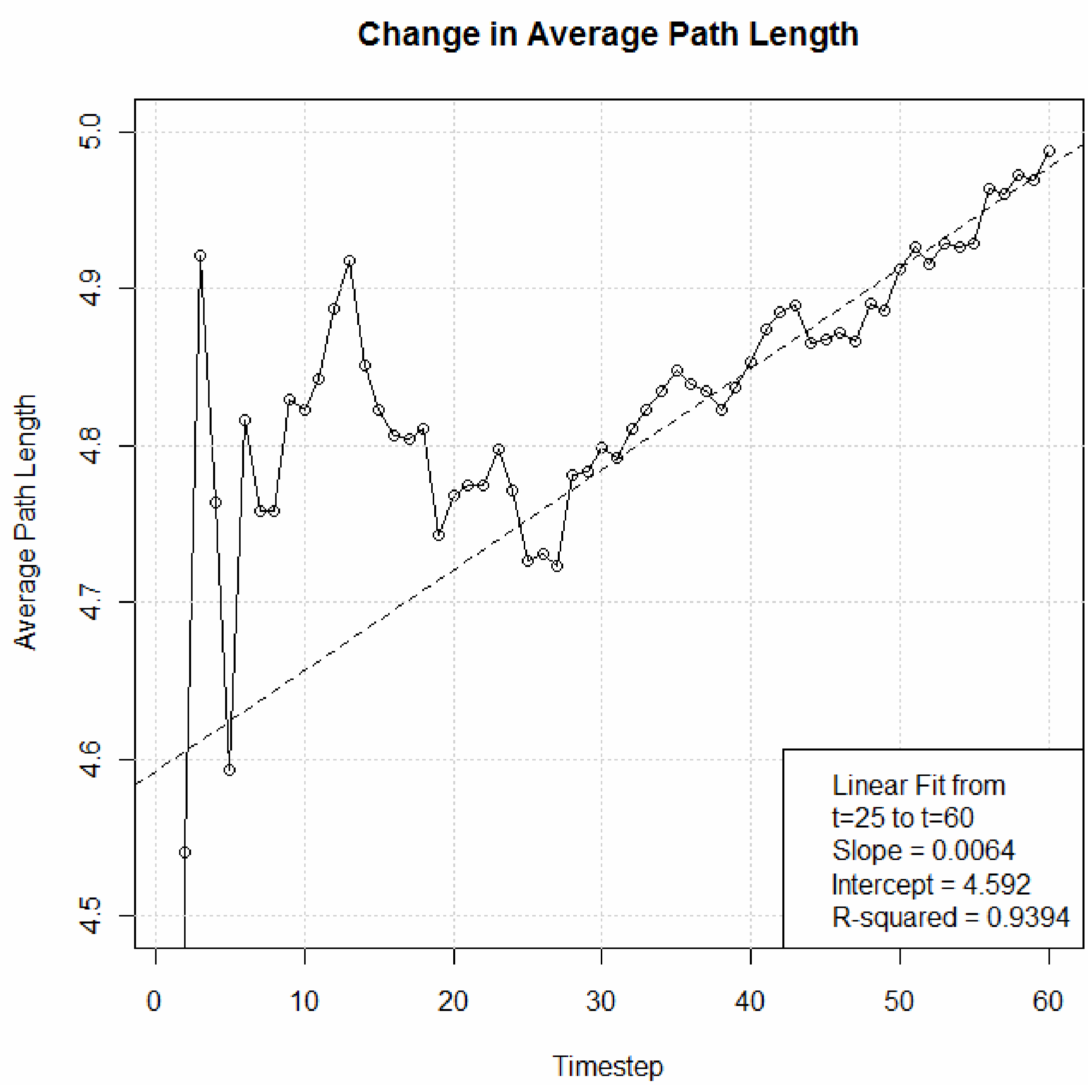


Figure 15. Changes in the average path length across the network over time.

Degree Distribution Exponent

The power law exponent vacillates significantly over the first 10 timesteps, then stabilizes and begins decreasing linearly. Between t=23 and t=24, the exponent crosses the $\gamma = 3$ mark, which corresponds to a transition between a random growth process and preferential attachment. The exponent stabilizes around t=40 and becomes unstable between t=50 and t=56 before leveling out near the $2.9 \leq \gamma \leq 3.0$ level.

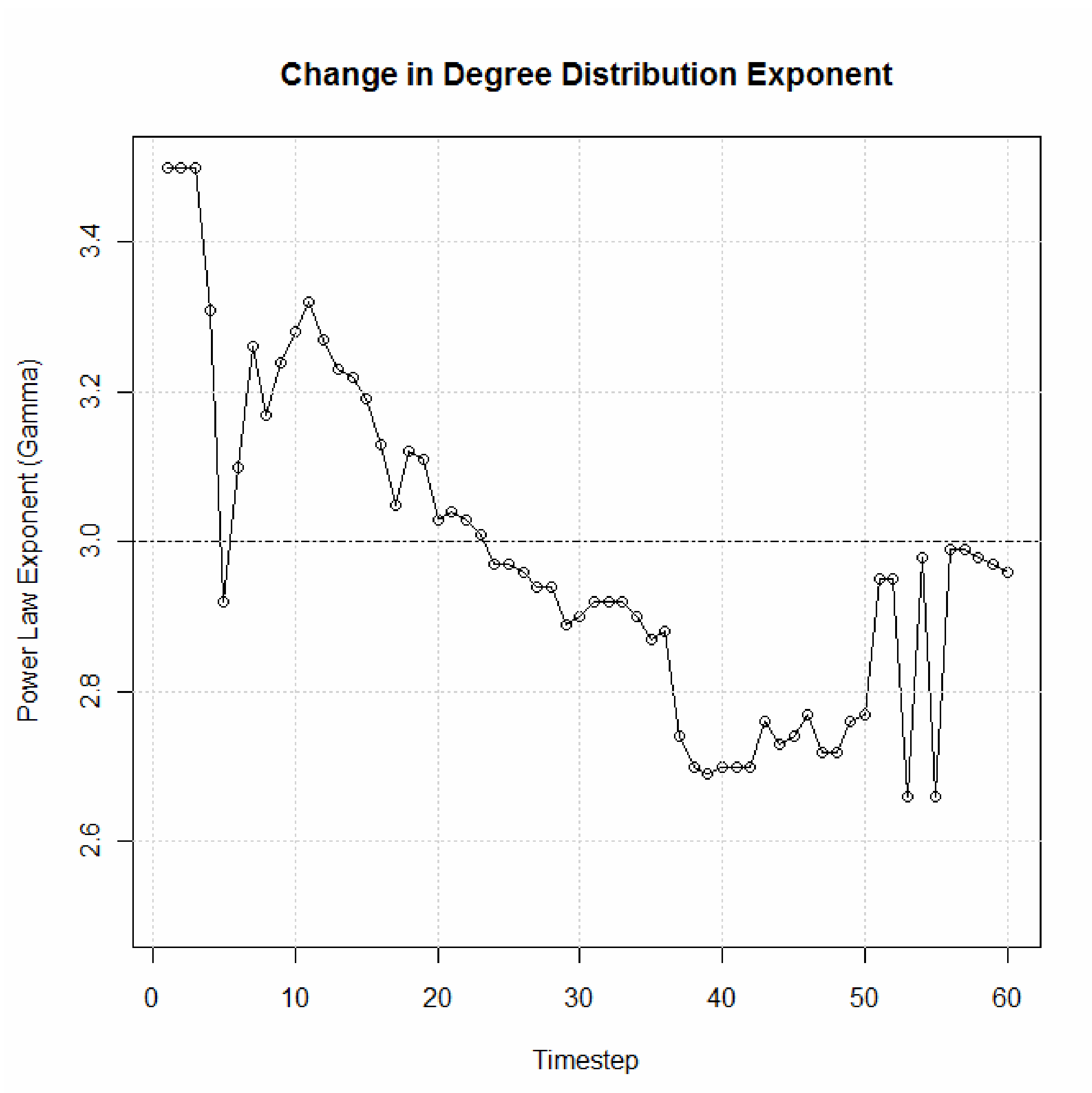


Figure 16. Variation in the maximum likelihood derived power law exponent.

Degree Distribution Goodness of Fit

The validity of the exponent of the fit to the degree distribution of the network is critical. If goodness of fit is not indicated, then inferences about the topology of the network using the exponent are not valid. Figure 17 shows the Kolmogorov-Smirnov D statistic for each of the fits in the evolution. Beyond t=12 (July 1996), the null hypothesis that the power law fit is reasonable cannot be rejected at the $\alpha = 0.10$ level.

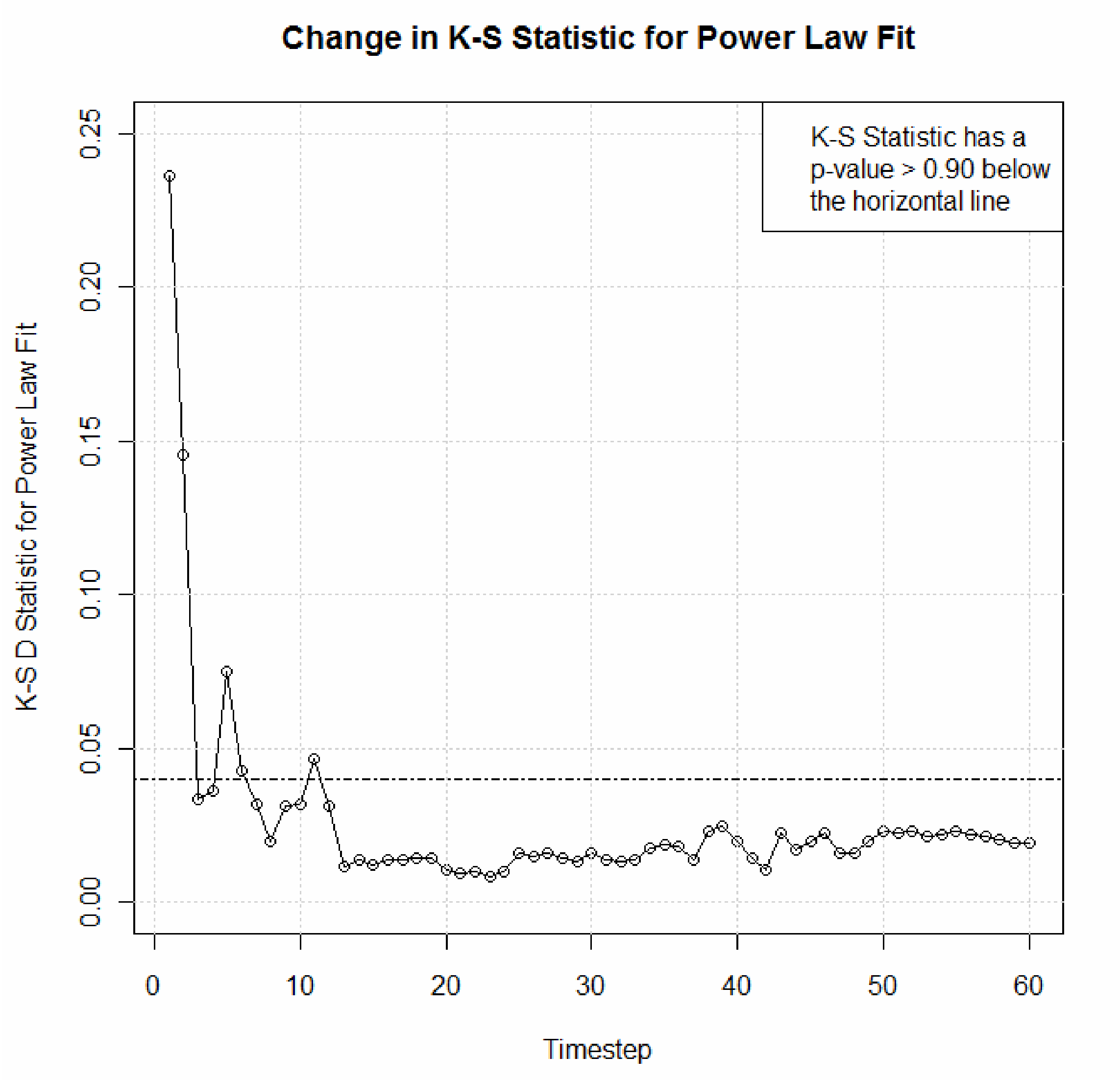


Figure 17. Goodness of fit from maximum likelihood estimation of exponents.

Centrality Analysis

Centrality analysis involves identifying which nodes within a network have the greatest influence, that is, which nodes are the most topologically central. There are several different measures that are used to assess node centrality, including number of in-degrees (k_{in}), Google PageRank, and Kleinberg's authority and hub scores. Each of these scores is monadic and thus measured relative to node *n*. The next five subsections present results from examining the *QMJ* citation network in terms of each of these measures.

Most Frequently Cited Resources

The most frequently cited articles are the nodes with the most in-degrees, k_{in}. In the *QMJ* citation network, the top 20 references each have 16 or more incoming citations. These are presented in Table 9 below. These results include 12 books (ranks 1-3, 6, 8, 12-17 and 19) and 8 peer-reviewed journal articles (ranks 4, 5, 7-11, 18 and 20). Only one book in the top 20, Nunnally's 1978 textbook on psychometric theory, is a methodological reference. The historically recognized quality gurus appear in this list as expected: Deming with the highest incoming citation count for the 1986 book *Out of the Crisis* plus two other references, Crosby with *Quality is Free*, and Juran with the *Quality Control Handbook* and one other resource.

Table 9

Top 20 Most Frequently Cited Resources in the QMJ Citation Network

Rank	k_{in}	Article
1	80	Deming, W. E. (1986). Out of the Crisis. Cambridge, MA: MIT Center for Advanced Engineering Study.
2	41	Crosby, P. B. (1979). Quality is Free. New York: McGraw-Hill.

3	37	Juran, J. (1951). Juran's Quality Control Handbook. New York: McGraw-Hill.
4	29	Powell, T. C. (1995). Total Quality Management as competitive advantage: A review and empirical study. *Strategic Management Journal,* 16(1), 15-37.
5	29	Saraph, J., Benson, P. & Schroeder, R. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences,* 20(4), 810-829.
6	29	Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.
7	25	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, Spring 1988, 12-37.
8	25	Garvin, D. A. (1988). Managing quality. New York: Free Press.
9	24	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, Fall 1985, 41-50.
10	24	Flynn, B. B., Schroeder, R. G. & Sakakibara, S. (1994). A framework for Quality management research and an associated measurement instrument. *Journal of Operations Management*, 11, 339-366.
11	22	Ahire, S. L., Golhar, D. Y. & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision Sciences*, 27(1), 23-56.
12	22	Feigenbaum, A. V. (1961). Total quality control: Engineering and

		management. New York: McGraw-Hill.
13	22	Senge, P. (1990). The fifth discipline. New York: Doubleday.
14	22	Ishikawa, K. (1991). What is total quality control? The Japanese way. Upper Saddle River, NJ: Financial Times/Prentice Hall.
15	22	Deming, W. E. (1995). The new economics for industry, government and education. MIT Center for Advanced Educational Services.
16	21	Deming, W. E. (1982). Quality, productivity and competitive position. Cambridge, MA: MIT Press.
17	20	Juran, J. M. (1989). Juran on leadership for quality. New York: Free Press.
18	17	Dean, Jr., J. W. & Bowen, D. E. (1994). Management theory and total quality: Improving research and practice through theory development. *Academy of Management Review*, 19(3), 392-418.
19	17	Evans, J. R. & Lindsey, W. M. (1989). The management and control of quality. St. Paul, MN: West Publishing.
20	16	Flynn, B. B., Schroeder, R. G. & Sakakibara, S. (1995). The impact of quality management practices on performance and competitive advantage. *Decision Sciences*, 26(5), 659-692.

Most Frequently Cited Academic Journals

In addition to examining which references are most frequently cited, the references can also be grouped by the publication in which they appear. This approach was used to identify which academic journals are most frequently referenced by *QMJ* authors. The background material for the *QMJ* is primarily drawn from the marketing literature, which captures the top two positions. The results, which indicate strong ties between quality management, marketing, and general management, are shown in Table 10. Operations management resources are moderately referenced, followed by academic articles in psychology and strategic management.

Table 10

Top 20 Most Frequently Cited Academic Journals in the QMJ Citation Network

Times Cited	Journal Name
209	Journal of Marketing
170	Journal of Marketing Research
154	Harvard Business Review
136	Technovation
127	Academy of Management Review
109	Quality Management Journal
102	Management Science
91	Journal of Operations Management
65	Academy of Management Journal
58	Journal of Retailing
54	Sloan Management Review
50	Journal of Applied Psychology
50	Strategic Management Journal
49	Administrative Science Quarterly
32	Psychological Bulletin
32	Production and Operations Management
32	Psychological Bulletin

26	Academy of Management Executive
26	International Journal of Quality and Reliability Management
24	Journal of Business Research

Most Frequently Cited Methodology References

The incoming citation information was also scanned to determine which methodological resources are most frequently employed by *QMJ* authors. Although most articles include a reference to at least one methodology reference, only 5 were cited 10 or more times (the remainder were cited only 1 to 3 times each). This indicates that quality management employs a wide variety of methodologies to satisfy the research objectives of the various studies. The top 5 methodology results, clearly indicating Nunnally's *Psychometric Theory* as a key reference, are displayed in Table 11.

Table 11

Top 5 Most Frequently Cited Methodology References in the QMJ Citation Network

Times Cited	Reference
29	Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.
22	Feigenbaum, A. V. (1961). Total quality control: Engineering and management. New York: McGraw-Hill.
14	Hair et al. (1995). Multivariate data analysis. New York: Macmillan.
11	Cronbach, L. J. (1951). Coefficient alpha. *Psychometrika*, 16, 297-334.
10	Eisenhardt, K. (1989). Building theories from case study research. *Academy of Management Review,* 14(4), 532-550.

Most Central Resources Using Google PageRank

The previous sections have examined the centrality of articles based on the number of incoming citations, and the centrality of journals based on the number of citations to papers within those journals. However, it is widely recognized that the number of incoming citations (or in-degrees) may not accurately reflect the significance of a reference. For example, a reference is certainly more influential if it is cited by other highly regarded papers, rather than by a collection of small and obscure works. Additionally, the oldest high-impact articles are often not referenced once their results are integrated into textbooks. (Maslov & Redner, 2008)

These shortcomings are somewhat alleviated by algorithms like Google PageRank that calculate the significance of an article based on how influential its citing papers are. The algorithm starts at one node, and randomly traverses that node's citations. At each step, the algorithm assumes a 0.15 probability of balking, indicating that the reader has stopped following citations and is instead searching for a new article. At this point, the algorithm again randomly selects a node and traverses citations. The procedure continues iteratively until the time spent at each node during the random walk converges. This value is the Google PageRank, presented in Table 12 for articles within the *QMJ* citation network. The PageRank result is scaled by 10^4 for easier interpretation.

Table 12

Top 20 Resources in the QMJ Citation Network Using Google PageRank

Rank	PR_n	Article
1	20.2	Deming, W. E. (1986). Out of the Crisis. Cambridge, MA: MIT Center for Advanced Engineering Study.

2	13.7	Heineke, J. & Meile, L., Eds. (1995). Games and exercises for operations management. Englewood Cliffs, NJ: Prentice Hall.
3	10.8	Deming, W. E. (1982). Quality, productivity and competitive position. Cambridge, MA: MIT Press.
4	10.7	Juran, J. (1951). Juran's Quality Control Handbook. New York: McGraw-Hill.
5	9.75	Crosby, P. B. (1979). Quality is Free. New York: McGraw-Hill.
6	7.40	Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.
7	6.85	Finch, B. J. (1999). Internet discussions as a source for consumer product customer involvement and quality information: an exploratory study. *Journal of Operations Management*, 17(5), 535-556.
8	6.75	Garvin, D. A. (1988). Managing quality. New York: Free Press.
9	6.26	Deming, W. E. (1995). The new economics for industry, government and education. MIT Center for Advanced Educational Services.
10	5.77	Saraph, J., Benson, P. & Schroeder, R. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences,* 20(4), 810-829.
11	5.42	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, Fall 1985, 41-50.
12	5.38	Juran, J. M. (1989). Juran on leadership for quality. New York: Free Press.
13	5.36	Buzzell, R. D. & Gale, B. T. (1987). The PIMS principles. New York: Free Press.

14	5.35	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, Spring 1988, 12-37.
15	5.31	Flynn, B. B., Schroeder, R. G. & Sakakibara, S. (1994). A framework for quality management research and an associated measurement instrument. *Journal of Operations Management*, 11, 339-366.
16	5.23	Hauser, J. R. & Clausing, D. (1988). The house of quality. *Harvard Business Review*, 66(3), 63-73.
17	5.19	Senge, P. (1990). The fifth discipline. New York: Doubleday.
18	5.16	Groocock, J. M. (1986). The chain of quality. New York: Wiley & Sons.
19	5.08	Baldrige Criteria (all years)
20	5.04	Montgomery, D. C. (1987). Introduction to statistical quality control, 2^{nd} Ed. New York: Wiley & Sons.

Authorities

Rather than using a random walk based approach, Kleinberg's authority score uses a method based on the eigenvectors of the graph's adjacency matrix to determine node centrality. The strength of this method is that added weight is given to citations from more highly cited articles than from less highly cited articles, as well as references received from review articles (called hubs). The results in Table 13 use Kleinberg's authority score to rank the most central references. Whereas only 6 articles in the Top 20 using Google PageRank are from academic journals, a total of 12 are present in the Top 20 using the authority score.

Table 13

Top 20 Authorities in the QMJ Citation Network

Rank	AU_n	Article
1	100	Deming, W. E. (1986). Out of the Crisis. Cambridge, MA: MIT Center for Advanced Engineering Study.
2	56.4	Crosby, P. B. (1979). Quality is Free. New York: McGraw-Hill.
3	52.2	Powell, T. C. (1995). Total quality management as competitive advantage: A review and empirical study. *Strategic Management Journal*, 16(1), 15-37.
4	49.7	Saraph, J., Benson, P. & Schroeder, R. (1989). An instrument for measuring the critical factors of quality management. *Decision Sciences,* 20(4), 810-829.
5	38.1	Flynn, B. B., Schroeder, R. G. & Sakakibara, S. (1994). A framework for quality management research and an associated measurement instrument. *Journal of Operations Management*, 11, 339-366.
6	37.7	Ahire, S. L., Golhar, D. Y. & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision Sciences*, 27(1), 23-56.
7	33.2	Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.
8	32.2	Juran, J. (1951). Juran's Quality Control Handbook. New York: McGraw-Hill.
9	31.5	Dean, Jr., J. W. & Bowen, D. E. (1994). Management theory and total quality: Improving research and practice through theory development.

		Academy of Management Review, 19(3), 392-418.
10	31.4	Feigenbaum, A. V. (1961). Total quality control: Engineering and management. New York: McGraw-Hill.
11	28.8	Ishikawa, K. (1991). What is total quality control? The Japanese way. Upper Saddle River, NJ: Financial Times/Prentice Hall.
12	28.2	Juran, J. M. (1989). Juran on leadership for quality. New York: Free Press.
13	28.0	Garvin, D. A. (1988). Managing quality. New York: Free Press.
14	27.5	Hendricks, K, and V. Singhal. 1997. Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. *Management Science* 43, no. 9:1258 –1274.
15	26.9	Black, S. A., and L. J. Porter. 1995. Identification of the critical factors of TQM. *Decision Sciences* 27, no. 1: 1–21.
16	26.4	Flynn, B. B., Schroeder, R. G. & Sakakibara, S. (1995). The impact of quality management practices on performance and competitive advantage. *Decision Sciences*, 26(5), 659-692.
17	22.6	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, Fall 1985, 41-50.
18	22.1	Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, Spring 1988, 12-37.
19	21.9	Samson, D., & Terziovski, M. (1999). The relationship between total

		Quality management practices and operational performance. *Journal of Operations Management*, 17, 393-409.
20	21.9	Phillips, L. W., Chang, D. R. & Buzzell, R. D. (1983). Product quality, Cost position, and business performance: A test of some key hypotheses. *Journal of Marketing*, Spring 1983, 26-43.

Hubs

To determine which articles are hubs (major review articles that draw together concepts in different themes or topic areas), Kleinberg's Hub Score was used. Kleinberg's hub scores are calculated iteratively, in conjunction with the authority scores. As expected, all of the hubs are *QMJ* articles because these are more heavily weighted with out-degrees in the citation network, as shown in Table 14. The top three hubs discuss frameworks for quality-driven development of strategy.

Table 14

Top 20 Hubs in the QMJ

Rank	HU_n	Article
1	100	Ford, M. W. & Evans, J. R. 2000. Conceptual Foundations of Strategic Planning in the Malcolm Baldrige Criteria for Performance Excellence. *QMJ*, 7(1), 8-26.
2	63.1	Issac, G., Rajendran, C. & Anantharaman, R. N. 2004. A Holistic Framework for TQM in the Software Industry: A Confirmatory Factor Analysis Approach. *QMJ*, 11(3), 35-60.
3	58.5	Jabnoun, N., Khalifah, A., & Yusuf, A. 2003. Environmental Uncertainty, Strategic Orientation, and Quality Management: A Contingency Model.

QMJ, 10(4), 17-31.

4	54.6	Issac, G., Rajendran, C, & Anantharaman, R. N. 2004. Significance of Quality Certification: The Case of the Software Industry in India. *QMJ*, 11(1),. 3-32.
5	50.5	Grandzol, J. R. & Gershon, M. 1997. Which TQM Practices Really Matter: An Empirical Investigation. *QMJ*, 4(4), 43-59.
6	48.3	Balbastre Benavent, F. 2006. TQM Application Through Self-Assessment and Learning: Some Experiences from Two EQA Applicants. *QMJ*, 13(1), 7-25.
7	48.2	Kujala, J. & Lillrank, P. 2004. Total Quality Management as a Cultural Phenomenon. *QMJ*, 11(4), 43-55.
8	46.5	Ryan, C., Deane, R. H. & Ellington, N. P. 2001. Quality Management Training in Small to Midsized Manufacturing Firms. *QMJ*, 8(2), 44-52.
9	46.2	Sun, H. 2001. Comparing Quality Management Practices in the Manufacturing and Service Industries: Learning Opportunities. *QMJ*, 8(2), 53-71.
10	43.8	Handfield, R., Ghosh, S. & Fawcett, S. 1998. Quality-Driven Change and Its Effects on Financial Performance. *QMJ*, 5(3), 13-30.
11	41.7	Jain, B. & Tabak, F. 2002. Organizational Quality Management in Emerging Economies. *QMJ*, 9(2), 10-24.
12	40.5	Cameron, K. & Sine, W., 1999. A Framework for Organizational Quality Culture. *QMJ*, 6(4), 7-25.
13	40.4	Prajogo, D. I. & Brown, A., 2004. The Relationship Between TQM

		Practices and Quality Performance and the Role of Formal TQM Programs: An Australian Empirical Study. *QMJ*, 11(4), 31-42.
14	37.6	Martinez-Lorente, A. R., Gallego-Rodriguez, A., & Dale, B. G. 1998. Total Quality Management and Company Characteristics: An Examination. *QMJ*, 5(4), 59-71.
15	37.3	Witcher, B. 1995. The Changing Scale of Total Quality Management. *QMJ*, 2(4), 9-29.
16	36.8	Pinar, M. & Ozgur, C. 2007. The Long-Term Impact of ISO 9000 Certification on Business Performance: A Longitudinal Study Using Turkish Stock Market Returns. *QMJ*, 14(4), 21-40.
17	35.8	Schniederjans, M. J., Parast, M. M., Nabavi, M., Rao, S. S. & Raghu -Nathan, T. S. 2006. Comparative Analysis of Malcolm Baldrige National Quality Award Criteria: An Empirical Study of India, Mexico, and the United States. *QMJ*, 13(4), 7-21.
18	35.7	Masters, B. & Frazier, G. V. 2007. Project Quality Activities and Goal Setting in Project Performance Assessment. *QMJ*, 14(3), 25-35.
19	35.4	Huff, L., Fornell, C., & Anderson, E. 1996. Quality and Productivity: Contradictory and Complementary. *QMJ*, 4(1), 22-39.
20	34.0	Barringer, B., Foster Jr., S. T., & Macy, G. 1999. The Role of Quality in Determining Export Success. *QMJ*, 6(4), 55-70.

Concept Analysis

To understand the interrelatedness of the top ranked articles from the *Quality Management Journal* at a more conceptual level, this study examines the most connected articles visually. The procedure of k-core decomposition can be used to extract the most central, connected components of a graph. (Alvarez-Hamelin et al., 2006) In a k-core decomposition, only the nodes with at least k incoming degrees are retained. The 3-core from the *QMJ* citation network, which is displayed in Figure 18, scales each node in terms of its number of incoming citations where larger nodes reflect a greater degree centrality. Green denotes nodes that appear in the top 20 using Kleinberg's authority score, blue indicates a *QMJ* article that is not an authority, and yellow represents all non-*QMJ* articles that are not authorities.

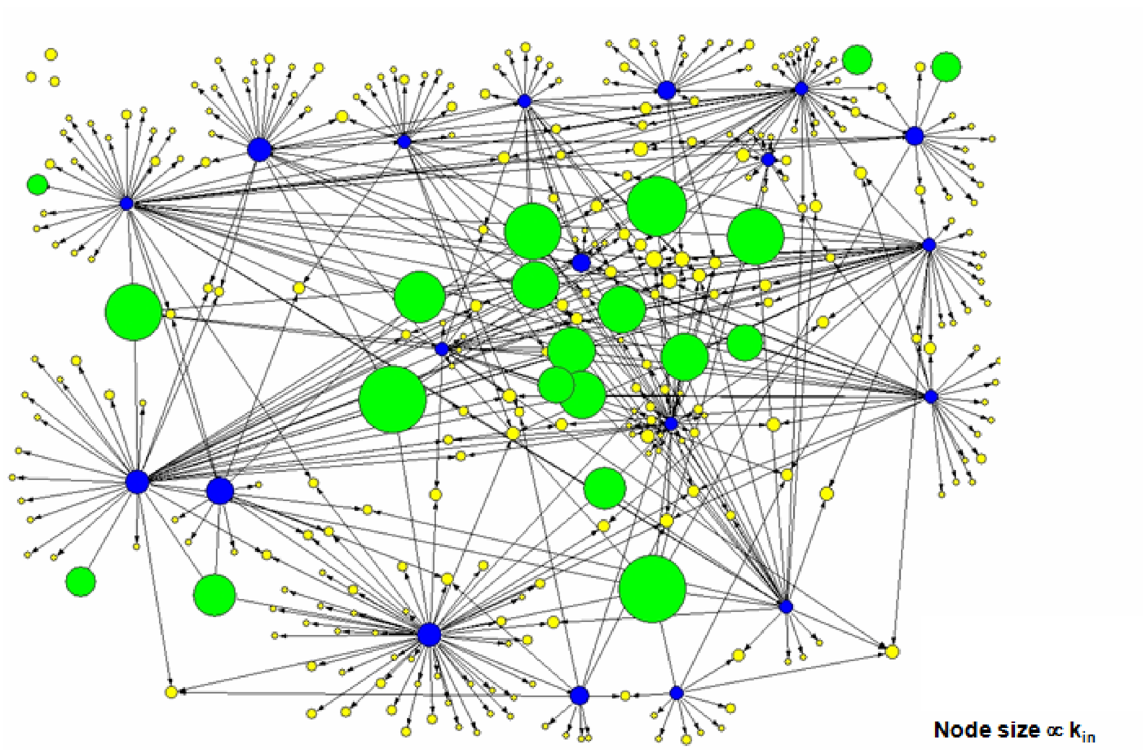


Figure 18. The 3-core decomposition of the *QMJ* citation network (n=1089).

Recall from the literature review in chapter 2 that both the Juran Center and the ASQ have established categories that define bounds of the quality management body of knowledge. These categories are listed in Table 15; an asterisk indicates that the category is shared by both the Juran and ASQ BOKs, and two asterisks indicate that the category is part of the ASQ BOK but is not covered by the Juran Center. Categories with no asterisk appear in the Juran body of knowledge but not in the ASQ taxonomy.

Table 15

Categories in the Quality Management Body of Knowledge

* Leadership for Quality	Process Management
*Strategic Planning for Quality	Business and Quality Results
*Customer and Market Focus	*Tools for Quality Improvement
Information and Analysis	*General Approaches and Philosophies
Human Resource Focus	Definitions of Quality
** Training & Development	

The 3-core decomposition of the *QMJ* citation network was examined to uncover the themes that are present in the research. The first pattern that is evident is the tight cluster in the middle, anchored by many of the authorities in the network, including Deming's *Out of the Crisis*, Feigenbaum's *Total Quality* Control methodology book, and Juran's *Quality* Handbook. The second pattern is a distinct outline of conceptual themes, which are pictured in Figure 19. Note that seven themes are apparent: TQM/validation, quality tools, international aspects of quality improvement, service quality, strategy development and the Baldrige criteria, quality impacts on business results, and quality culture. Crosby's *Quality is Free* is clustered with the section on business results, suggesting that this reference is most useful when the business impacts of quality are studied. Three of the seven categories are covered by the body of knowledge categories in Table 16: quality tools, strategy development, and business results. Eight areas are absent entirely: leadership, customer/market focus, information and analysis, human resource focus, training and development, process management, and definitions of quality. This does not suggest that these topical areas are insignificant, only that they are not major themes in the research in quality management in the *QMJ*.

The four categories that are evident as a result of studying the *QMJ* citation network, but are not part of the established BOKs described in Table 16, are: validation of quality methods, international aspects of quality, service quality, and quality culture. The prominence of these categories in the citation network analysis indicates that they may be more important than initially realized with respect to the body of contemporary research in quality management.

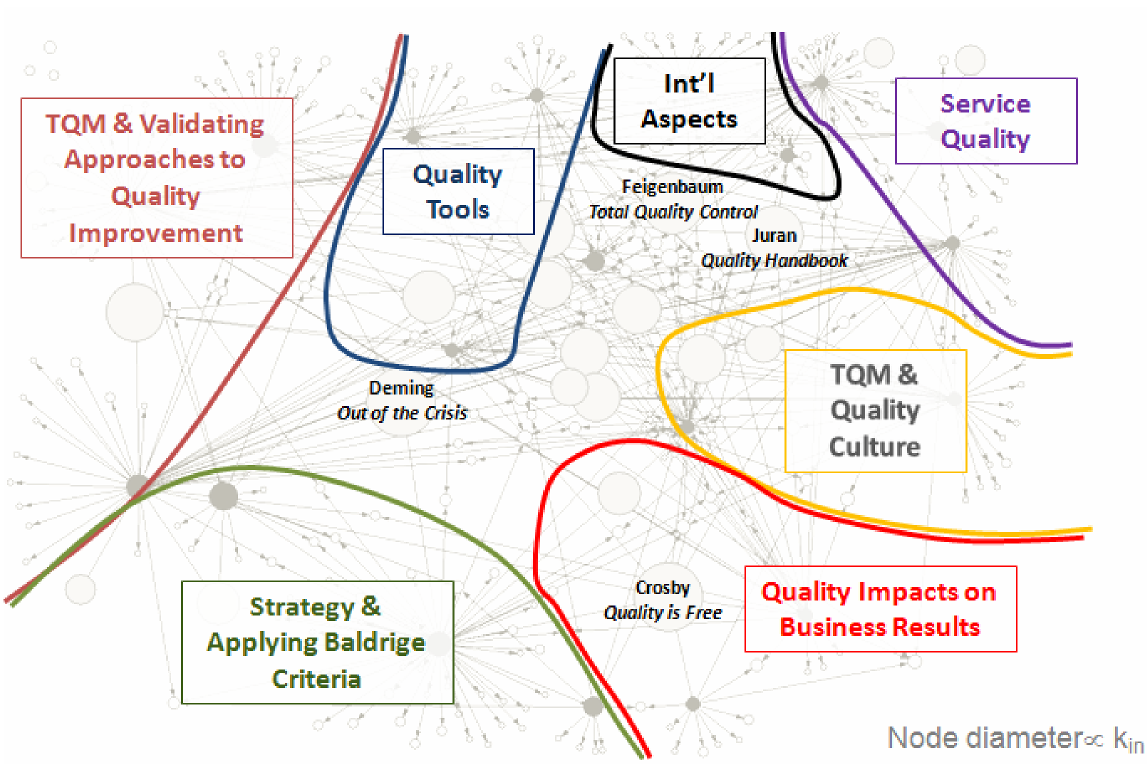


Figure 19. Conceptual themes in the 3-core decomposition of the *QMJ* citation network.

Quality Estimation

The purpose of this section is to present results related to the third research question, which explores whether the mean quality of the network $<\eta>$ or the attachment-weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics. Furthermore, this study aims to examine whether $<\eta>$ or $<\eta>\zeta$ can be predicted using only the exponent of the degree distribution, γ_t, at time t. Note that in this section the terms "fitness" and "quality" are used interchangeably, on the basis that quality is defined as fitness for use by a new connection.

Predicting the Mean Quality of the Network

Multiple linear regression was used to test the null hypothesis that a linear model would be appropriate to predict the mean quality of the network at a given time. For the first test case, the dependent variable was the mean fitness of the network at time t, $<\eta>_t$. The following independent variables were utilized: number of nodes in the network at time I, number of edges in the network at time t, number of new nodes over the window Δt, number of new edges over the window Δt, the ratio of new edges to new nodes over Δt, average path length, clustering coefficient, power law exponent γ, network density, diameter, number of growing nodes, number of stagnant nodes that acquired no links over Δt, change in the power law exponent over Δt, and probability of attachment to a pre-existing node. Pairwise correlations were examined to ensure that the assumptions of the regression technique could be met. Linear models were attempted iteratively, selecting independent variables to avoid multicollinearity, until only the most statistically significant coefficients remained and the multiple correlation coefficient, R^2, was maximized.

The observed mean fitness is plotted in Figure 20. Note that mean fitness starts very high, indicating that there is strong preferential attachment in the early formation of the citation network. This measure peaks between t=20 and t=30 and then falls. Clusters of low mean fitness are noted from t=5 to t=8 (corresponding to calendar year 1994), t=41 to t=44 (calendar year 2004) and t=53 to t=56 (calendar year 2007).

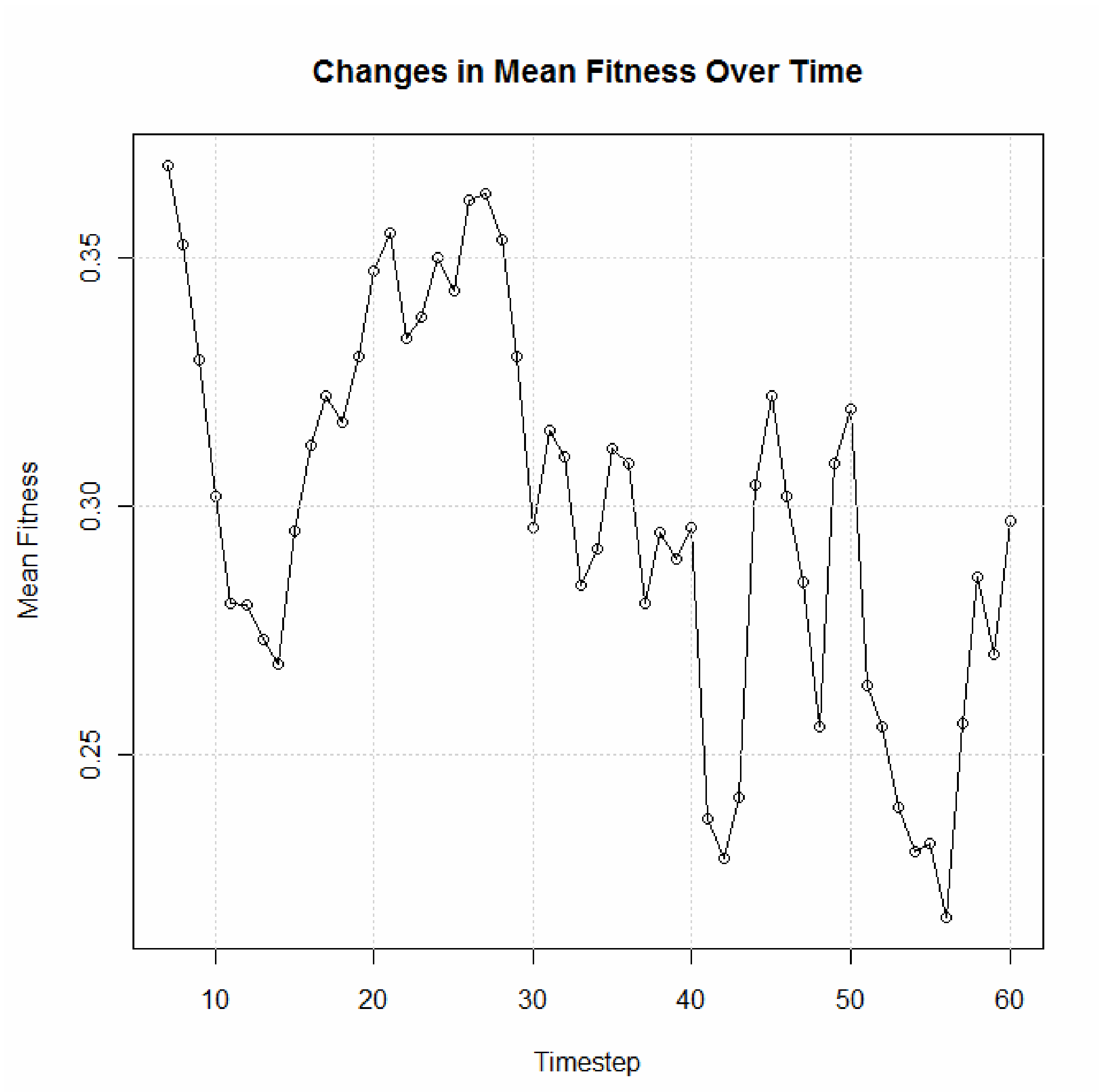


Figure 20. Changes in mean fitness of nodes within the network, $\langle\eta\rangle_t$.

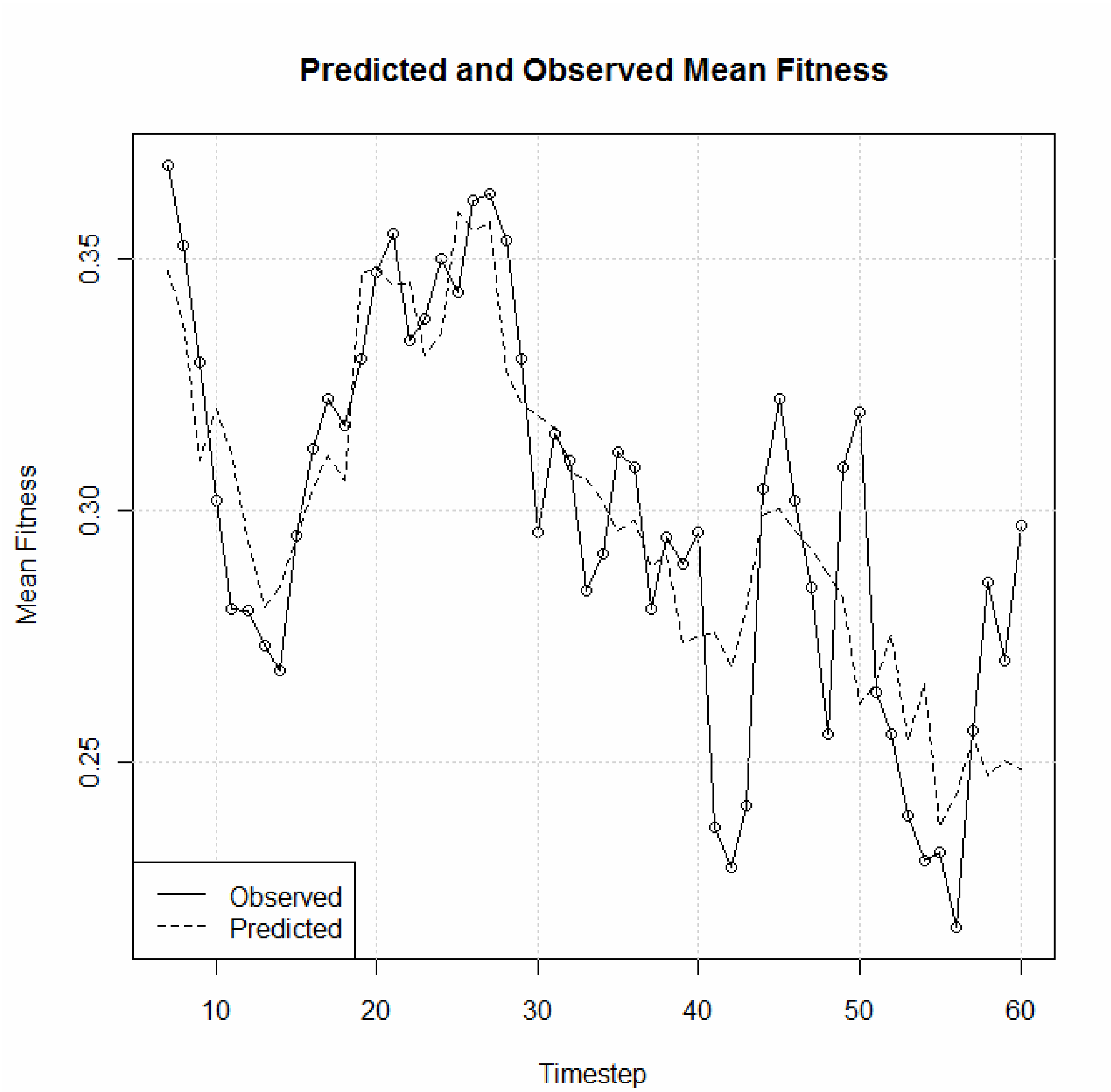
Although the shift in quality measures in 1994 could be due to artifacts of the early construction of the network, in both 2004 and 2007 there is an influx of hub articles into the network. This suggests that adding hubs can appear to depress the measured quality of the network temporarily, even though other factors (e.g. $\Delta\langle k_{in}\rangle$) are undisturbed.

The final regression model for $\langle\eta\rangle_t$ employed five predictors: change in number of nodes over the time window Δt, change in number of edges over the time window Δt, average path length *L*, power law exponent γ, and probability of attachment to a pre-existing node ζ. The predictive model is:

$$\langle\eta\rangle_t = -1.155\zeta - 0.477L + 0.054\,\gamma + 1.3^{e-4}\,(dn/dt) - 1^{e-4}\,(de/dt) + 2.726$$

All predictors were significant at the $\alpha = 0.05$ level except for average path length, which was significant at $\alpha = 0.001$. The multiple R^2 was 0.6940, indicating that 69.4% of the variation in the data could be explained by this model. The value of R^2 adjusted for the number of terms in the model still suggested a reasonable level of correspondence at 0.6621. A plot of the observed mean fitness compared to the predictive model is shown in Figure 21.

A linear model using only the power law exponent as the predictor for the mean fitness indicated that the power law exponent was significant as a predictor at $\alpha = 0.05$. However, the model could not account for more than 10% of the variance in the data which is not acceptable. ($R^2 = 0.1013$, Adj. $R^2 = 0.0839$, $p = 0.019$) The single-predictor model thus did not present an effective relationship to capture the fluctuations in mean fitness.



F*igure 21*. Comparison of predicted and observed mean fitness $<\eta>_t$.

Predicting the Attachment-Weighted Quality of the Network

Multiple linear regression was also used to test the null hypothesis that a linear model would be appropriate to predict the attachment-weighted quality of the network at a given time. For this second test case, the dependent variable was the attachment-weighted quality of the network at time t, $<\eta>_t\zeta_t$. The suite of independent variables was identical to those used for the first test case, with the exception of ζ which is a part of the

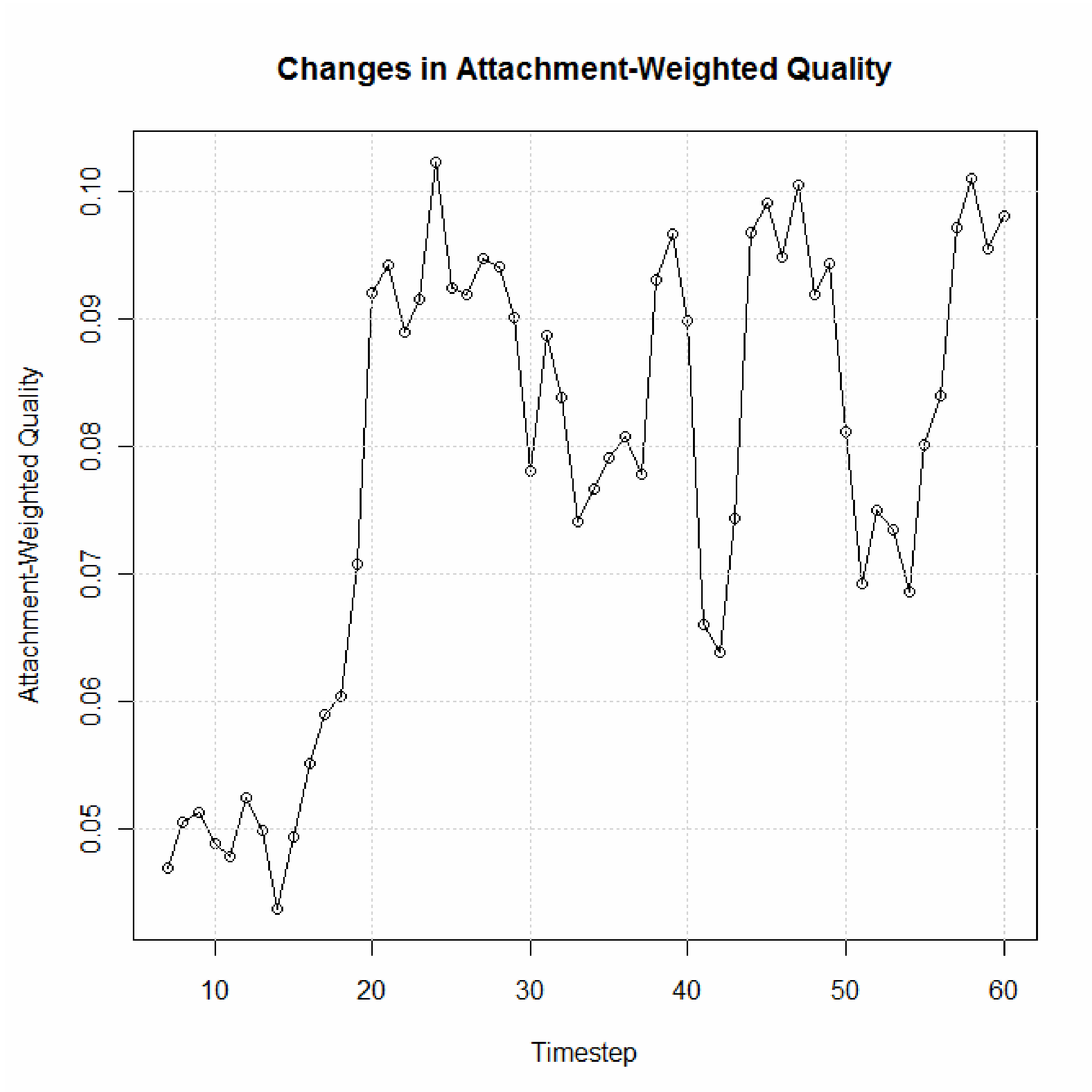


Figure 22. Changes in attachment-weighted quality of the network, $<\eta>_t\zeta_t$

Regression models were tested using stepwise elimination of independent variables. The final regression model for $<\eta>_t\zeta_t$ employed three predictors: change in number of nodes over the time window Δt, change in number of edges over the time window Δt, and average path length *L*. The predictive model is:

$$<\eta>_t\zeta_t = 0.681 - 0.126L - 3.06^{e-4}\,(dn/dt) + 2.36^{e-4}\,(de/dt)$$

All predictors were significant at the $\alpha = 0.001$ level except for power law exponent γ, which was significant at $\alpha = 0.12$. The multiple R^2 was 0.8668, indicating that 86.7% of the variation in the data could be explained by this model. The value of R^2 adjusted for the number of terms in the model still suggested a reasonable level of correspondence at 0.8588. A plot of the observed attachment-weighted quality of the network compared to the predictions is shown in Figure 23.

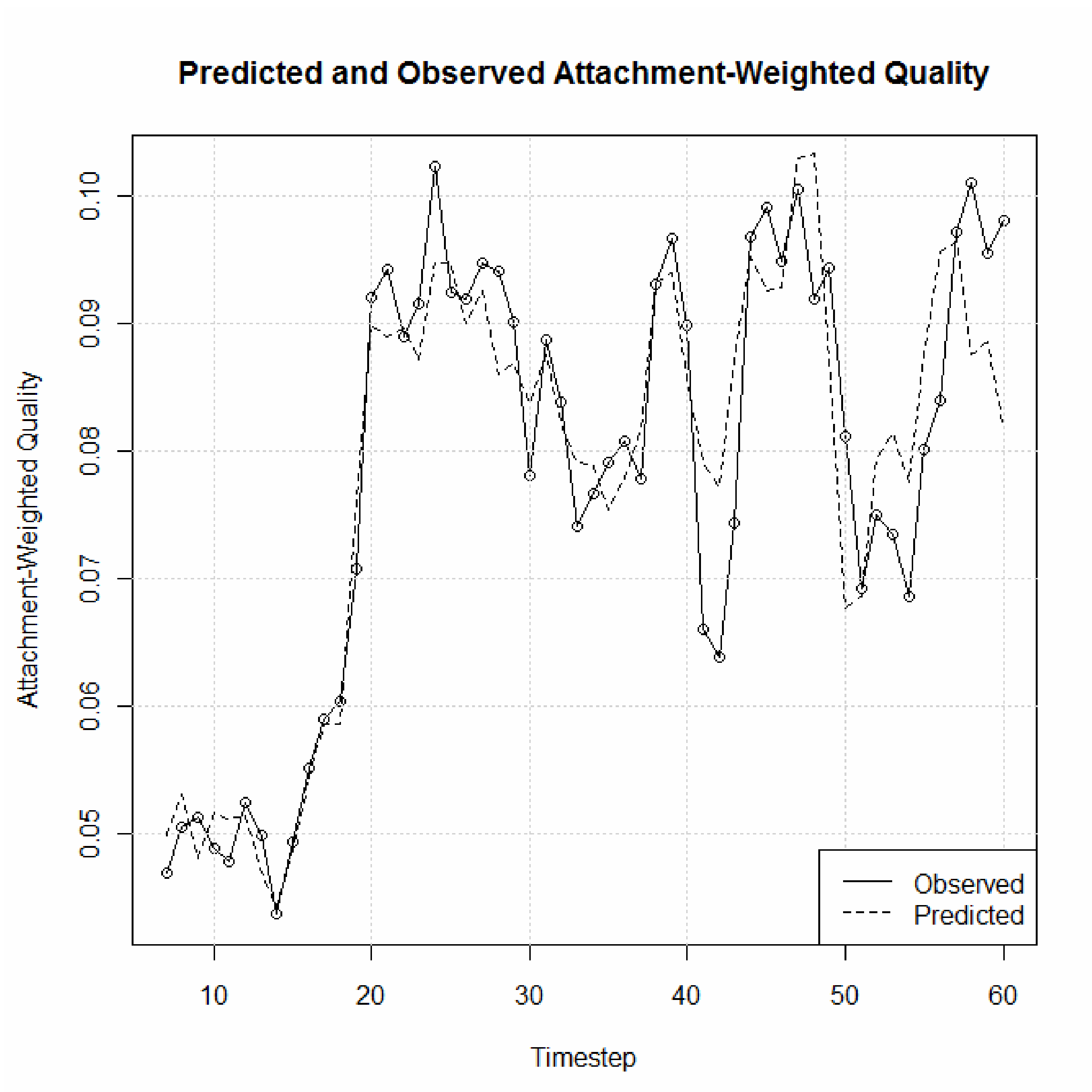


Figure 23. Comparison of predicted and observed attachment-weighted quality $<\eta>_t\zeta_t$

Interpretation of the Quality Measures

In this section, two potential quality indicators (each measured at the systemic level of the network) were proposed: the mean fitness of the network in terms of the average mean fitness across all of the nodes $<\eta>$, and the attachment-weighted quality $<\eta>\zeta$ which takes into consideration the probability of a new edge connecting to a pre-existing node. Over the observed evolution of the *QMJ* citation network, the two measures were expressed in remarkably different ways. The mean fitness, a measure of how active the force of preferential attachment is within the network at a given time, dropped from the time the network started growing until $t = 14$, increased continually until $t = 27$, and then returned to a highly variable downward trend. The attachment-weighted quality, a measure of the attractiveness of the pre-existing nodes coupled with the force of preferential attachment over the time window Δt, started low and gradually hit its peak at $t = 24$, after which it dropped, but remained consistently higher than the measurements from $t = 0$ to $t = 19$. The attachment-weighted quality dropped from $t = 41$ to $t = 43$ and $t = 51$ to $t = 55$. This roughly corresponds to the two periods where the mean fitness of the nodes within the network was lower, calendar years 2004 and 2007.

Quality Improvement in the QMJ Citation Network

To determine whether there was a difference in quality between the early years of the *QMJ* citation network and the later years, a two-sample t-test was conducted. The first group of measurements, the mean fitness values from 1993 to 2000, is contrasted against the mean fitness values from 2001 to 2008. The t-test shows a statistically significant difference with respect to the mean fitness ($t = 5.685$; $p < 0.001$) as well as the attachment-weighted quality ($t = -5.105$; $p < 0.001$). The mean fitness shows a statistically significant decrease, indicating that preferential attachment is more diffuse in the network in its later years (that is, more nodes are receiving connections preferentially). The attachment-weighted quality shows a statistically significant increase, indicating that even though preferential attachment is less active in the network as a whole over time, the pre-existing nodes are receiving more recognition.

It is interesting to note that both the mean fitness and the attachment-weighted quality undergo a reversal around t = 24. This is precisely the time step at which small world characteristics became evident in the network, and also the time step at which the rate of increase of the mean in-degree shifted downward. Continually increasing clustering, limited growth in the average path length compared to the size of the network, and formation and maintenance of a small world starting at t = 24 indicate that the citation network achieved a level of maturity, and thus demonstrated an increase in quality over time. Because all of the other topological characteristics suggest that the quality of the citation network increased from t= 0 to t = 60, this suggests that the attachment-weighted quality is a more appropriate measure of the quality of the network than the mean fitness of the nodes.

Summary of Key Findings

The analyses presented in this chapter revealed that the *QMJ* citation network is not random, that it is effectively modeled by a power law with degree distribution exponent $\gamma = 2.96$, and that it is a small world. This is consistent with other citation networks. Evaluation of the time evolution of the topological characteristics indicated an increase in quality of time, maturation of knowledge flows, and a transition around t = 24 (October 1999). These results are summarized in Table 16.

Table 16

Summary of Results for Research Question 1

Item	Description
Question	What does the citation network of articles in the *QMJ* reveal about knowledge flows in the study of quality management and the maturity of the discipline?
Hypotheses	H_{01A}: The citation network of articles in *QMJ* is random. H_{01B}: The citation network of articles in *QMJ* is scale-free. H_{01C}: The citation network of articles in *QMJ* exhibits the small world property of short average path length and high clustering.
Tests	1A: Shapiro-Wilk test of normality on the degree distribution of the *QMJ* citation network, assuming a mean of $<k_{in}>$ 1B: Kolmogorov-Smirnov goodness of fit test on the power law fit to degree distribution with given exponent derived from MLE 1C: Small world heuristic test, examining whether the ratio of the average path length in the network to the average path length of an equivalent random network was greater than 2, or that the ratio of the observed clustering coefficient to the clustering coefficient from an equivalent random network was less than 5
Results	Results from the Shapiro-Wilk test indicated that the null hypothesis For 1a should be rejected ($W = 0.1135$; $p < 0.001$). Kolmogorov-Smirnov goodness of fit test indicated that the null hypothesis for 1b could not be rejected at the $\alpha = 0.05$ level ($D = 0.194$; $p = 0.978$) for $\gamma = 2.96$. The small world test revealed that the null hypothesis for 1c could not be rejected ($L/L_r = 0.557$; $C/C_r = 11$).
Notes	The *QMJ* citation network is a scale-free small world with a degree distribution following $P(k) \propto k^{-\gamma}$ with $\gamma = 2.96$. The time evolution of the topology indicated the onset of small world features at $t = 24$, a distinct shift in mean in-degree at $t = 24$, increasing clustering, slowly increasing average path length (very small as compared to network size), a transition in the power law exponent from > 3 to < 3 around $t = 24$, and regular densification with a growth exponent of $\alpha = 1.115$. These results show a transition near $t = 24$ (Jan. 2000), and indicate maturity of knowledge flows after the transition and a general increase in the overall quality of the network over time.

Centrality analysis was then applied to determine the top 20 most central articles, the top 20 hubs and authorities, the top 5 methodology references, and the most frequently cited academic journals. This revealed that quality management is more dependent upon the marketing and strategic management literature than the production and operations management literature. A summary of results is shown in Table 17.

Table 17

Summary of Results for Research Question 2

Item	Description
Question	What are the most influential resources in the *QMJ* citation network?
Subquestions	2a: What are the Top 20 most frequently cited references? 2b: What are the Top 20 most central references? 2c: What are the Top 20 most authoritative references? 2d: What are the Top 20 hubs/review articles in the *QMJ*? 2e: What academic journals provided the most references for *QMJ* articles? 2f: What methodology resources were most frequently referenced by QMJ articles?
Tests	2a: Identify 20 nodes with largest degree centrality, k_{in} 2b: Identify the 20 nodes with largest Google PageRank scores 2c: Identify the 20 nodes with largest Kleinberg authority score 2d: Identify the 20 nodes with largest Kleinberg hub score 2e: Identify the 20 journals with the largest aggregate k_{in} 2f: Identify the 5 methodology references with the largest k_{in}
Results	Results from the centrality analysis are presented in Table 10 through Table 15 on pages 83 to 95. These lists of citations represent the most central articles in the *QMJ* citation network.
Notes	The most connected portion of the *QMJ* citation network (the 3-core) was inspected manually to uncover key concepts. Seven categories were noted: business results, strategy/Baldrige criteria, quality tools, international aspects, service quality, quality culture, and validating quality practices. The last four are not emphasized in either the ASQ or Juran Center quality management body of knowledge.

Two potential systemic quality measures were then explored: the mean fitness of nodes within the network, and the attachment-weighted quality, a measure defined in this research to weight the mean fitness in terms of the attractiveness of the pre-existing nodes in the network. Both of these measures were effectively assessed by regression models with multiple predictors, which included the change in nodes over a time window, the change in edges over a window, and the average path length. Neither quality measure could be predicted by the degree distribution exponent alone (see Table 18).

Table 18

Summary of Results for Research Question 3

Item	Description
Question	Can the mean quality of the network $<\eta>$ or attachment-weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics at that time? Can $<\eta>$ or $<\eta>\zeta$ be predicted using only the exponent of the degree distribution at that time, γ_t?
Hypotheses	H_{03}: The regression model is appropriate.
Test	Multiple linear regression was used to predict the dependent variables $<\eta>$ and $<\eta>\zeta$, using stepwise elimination at $\alpha = 0.05$ of independent variables: number of nodes in the network at time t, number of edges in the network at time t, number of new edges over the window Δt, number of new edges over the window Δt, ratio of new edges to new nodes over the window Δt, average path length, clustering coefficient, power law exponent γ, network density, diameter, number of growing nodes, number of stagnant nodes that acquired no links over the window Δt, and for the prediction of $<\eta>$, the probability of attachment to a pre-existing node, $<\eta>\zeta$.
Results	H_{03} could not be rejected for model with multiple predictors for $<\eta>$ on page 101 (Adj. $R^2 = 0.6621$) or for $<\eta>\zeta$ on page 103 (Adj. $R^2 = 0.8588$). H_{03} rejected using only γ_t for both $<\eta>$ (Adj. $R^2 = 0.0839$; $p = 0.019$) and $<\eta>\zeta$ (Adj. $R^2 = 0.3831$; $p < 0.001$) due to low R^2.

Using these measures, it was determined that there was a statistically significant decrease in the mean fitness of nodes within the network over time, whereas there was a statistically significant increase in the attachment-weighted quality (as indicated in Table 19). Both measures expressed a reversal around the time when the network became a

Table 19

Summary of Results for Research Question 4

Item	Description
Question	Are there significant differences in the mean quality of articles in *QMJ* between 1993-2000 and 2000-2008 in terms of either $<\eta>$ or $<\eta>\zeta$? Which measure is more appropriate as a quality indicator?
Hypotheses	H_{04}: $\mu_{1(1993\text{-}2000)} - \mu_{2(2000\text{-}2008)} = 0$
Test	The t-test shows a statistically significant decrease in $<\eta>$ ($t = -5.685$; $p < 0.001$) and a statistically significant increase in $<\eta>\zeta$ ($t = 5.105$; $p < 0.001$) between the periods 1993-2000 and 2000-2008.
Results	H_{03} was rejected for both cases.
Notes	Because the results from Research Question 1 (see Table 17) indicate qualitatively that the quality of the network did increase between the periods 1993-2000 and 2000-2008, the attachment-weighted quality $<\eta>\zeta$ is more appropriate as a quality indicator than the mean fitness of the nodes $<\eta>$.

small world. All predictive models indicated that the strategy for increasing the quality of a network is twofold: ensure that there are always more new edges than new nodes added at each time step, and that the average path length is decreased by adding new short cuts between articles within the network. One way to achieve both simultaneously is to encourage the regular production of review articles on various cross-cutting topics, which may also stimulate further research innovation.

CHAPTER 5

CONCLUSIONS

This study modeled knowledge flows in quality management, using the construct of citation networks, to determine how the discipline has grown and evolved and how it can be continually improved. The key goals were to 1) survey the literature in quality management using the Quality Management Journal (*QMJ*) citation network, constructed from 1993 to 2008, 2) determine the most central and influential references over this period, and 3) identify how to use information about the topology of the network to measure quality and effect continuous improvement at the systemic level. The study investigated the problems of how to assess the maturity and cohesiveness of the quality management discipline, how to identify the most influential references, and how to continually improve the citation network.

This research initially required constructing a network of journal articles in quality management. Then, applying graph theoretical methods to analyze the network, this project set out to answer the following research questions:

RQ1. What does the citation network of articles in the *QMJ* reveal about knowledge flows in the study of quality management and the maturity of the discipline?

RQ2. What are the most influential resources in the *QMJ* citation network?

RQ3. Can the mean quality of the network $<\eta>$ or attachment-weighted quality $<\eta>\zeta$ at time t be predicted from the network's topological characteristics at that time? Can $<\eta>$ or $<\eta>\zeta$ be predicted using only the exponent of the degree distribution at that time, γ_t?

RQ4. Are there significant differences in the mean quality of articles in *QMJ* between 1993-2000 and 2000-2008 in terms of either $<\eta>$ or $<\eta>\zeta$? Which measure is more appropriate as a quality indicator?

The first research question was explored by studying the global structure, or *topology*, of the citation network in quality management. The second research question examined one specific feature of the network topology, namely the *centrality* of nodes within the network, to determine the most central and influential references. To satisfy the third research question, the study examined the *dynamics* of topological and quality measures on the citation network as it evolved over the 15 years of observations. The mean fitness of the nodes in the network, and a newly derived measure called *attachment-weighted quality*, were predicted from the topological characteristics of the network to accomplish this. To understand the fourth research question, the population of quality measures on the citation network was examined to determine whether there had been a quality improvement from the early to the later stages of the network.

This final chapter presents the results of the investigation within the context of the research questions, and illustrates how the outcomes from this study can be applied to promote rigor and relevance in quality management research. Mechanisms for estimating the quality of a citation network, and continually improving the quality of a citation network, are proposed. The chapter concludes with a discussion of the implications of these results, and recommendations for further research.

Discussion and Implications of the Results

The purpose of this research was to survey the quality management research within the *QMJ* to extract the most influential resources, and to use quality measures based on the network's topological characteristics to identify how to continually improve the citation network. The research methodology was accomplished by constructing a citation network from all articles within the *QMJ*, and using it to extract the most central and influential resources. This research also evaluated the role of two quality metrics, the mean fitness of the nodes within the network and the attachment-weighted quality, to determine whether either could be predicted from the network's topology and used to assess and continually improve the quality of the network.

Research Question 1, Part 1: Knowledge Flows and Network Topology

As indicated in the results summary in Table 17, the *QMJ* citation network is not random, but instead is a scale-free network influenced by the preferential attachment of new connections to nodes with higher intrinsic levels of quality which appear to vary over time. The network became a small world around the time its 24th issue was published in October 1999. The exponent of the degree distribution of the *QMJ* citation network, measured using a maximum likelihood estimate method, has varied between γ = 2.6 and γ = 3.0 between the time at which the network matured into a small world and the present. Clustering within the network has continually increased since its inception, coupled with a slowly increasing average path length across the network, indicating that the level of self-organization has also continually expanded over time and that knowledge flows effectively through the system.

This study also examined the 3-core decomposition of the *QMJ* citation network to determine which topics formed the most connected component within the network. The seven topics dominating the network were: validation of quality improvement methods, quality tools, international aspects of quality, service quality, strategy development and Baldrige criteria, quality and business results, and quality culture. Although three of these areas are categories in the Body of Knowledge taxonomies from the ASQ and the Juran Center, four categories are not part of the established BOKs: validation, international aspects, service quality, and quality culture. The prominence of these categories in the citation network analysis indicates that they may be more important than initially realized in terms of contemporary research. Perhaps more significantly, these results show the utility of examining a citation network as an input to the process of formally constructing a Body of Knowledge.

Research Question 1, Part 2: Discipline Maturity and Network Dynamics

By examining the topological changes in the network over time, it was determined that the network experienced a topological transition around October 1999 ($t = 24$). At this time, the small world property became observable within the network. The network continued to demonstrate small world characteristics throughout the period of observation. The rate of increase of mean in-degree, which had been increasing linearly at a rate of 7.18×10^{-3} additional in-degrees per timestep, also shifted radically at this time to linearly increase at a rate of 3.11×10^{-3} new in-degrees per timestep thereafter. Between $t = 23$ and $t = 24$, the exponent of the degree distribution crossed the threshold of $\gamma = 3$, indicating that preferential attachment began playing a role around the same time that the network became a small world. There was also a distinct reversal in both quality measures at this time: the mean fitness of the nodes within the network decreased significantly, indicating that preferential attachment had become more diffuse and spread out over more nodes, whereas the attachment-weighted quality rose sharply to indicate that the ability of the pre-existing nodes to capture links had become enhanced. This outcome lends further credence to the significance of the small world onset in a citation network.

These results are consistent with the observations of Li-Chun et al. (2006), who explain that citation networks tend towards small worlds with short average path lengths (relative to the size of the network), a high level of clustering, and exponential or power-law degree distributions. These authors note that such features are "important characteristics of scientific communication patterns because they indicate an overall coherence and integration… coupled with intensely connected local and specialized communities." As a result, this study can conclude that the *QMJ* citation network achieved a level of coherence and integration characteristic of the maturation of the research discipline around October 1999.

Research Question 2: Resource Influence and Node Centrality

The second research question assessed the relative influence of nodes within the network in terms of degree centrality (that is, counting incoming citations), Google's random walk-based PageRank algorithm, and Kleinberg's authority and hub scores. The summary of results is presented in Table 18. The hub scores, as expected, revealed which articles within the *QMJ* can be considered the most influential review articles. The other measures extracted top performers across the entire network. As expected, several of the "quality gurus" including Deming, Juran and Crosby are ranked highly. But many of the most central articles in the quality management citation network, in terms of Google PageRank or Kleinberg's authority score, had demonstrably fewer incoming citations and would not be recognized as influential if only degree centrality had been considered. Although there was much overlap between the top ranked articles using each of these measures, some infrequently cited resources were associated with high authority scores and PageRanks. These include, for example, Buzzell & Gale's influential 1987 study on the Profit Impact of Market Studies (PIMS) principles and Black & Porter's (1995) study on the most critical aspects of Total Quality Management (TQM) from the perspective of business results.

The authority score tended to extract the most influential academic journal articles from the citation network, and for this reason it is recommended as a measure of prominence within the network over Google PageRank, which tended to include seemingly unimportant articles in its top 20 (e.g. Heineke & Meile's 1995 text on games in operations research, which was only cited by two other articles within the network). This finding is consistent with Maslov & Redner (2009) who first reported this tendency of the PageRank algorithm. When PageRank is used for finding and presenting search results, this could be considered a beneficial "innovative factor". However, when centrality measures are used for quality estimation purposes, this random aspect could negatively impact the interpretation of quality measures by numerically inflating articles that are not truly influential.

Research Question 3: Quality Estimation and Predictive Models

The predictive models described in the results summary of Table 19 show that either the mean fitness of the nodes $\langle\eta\rangle$ or the attachment-weighted quality $\langle\eta\rangle\zeta$ can be predicted from the change in nodes over the timestep, the change in edges over the timestep, the average path length of the network, and (with lesser significance) the degree distribution exponent. This model assumes that the changes in nodes and edges can be approximated as linear, and that the power law fit is also statistically significant.

Despite the presence of the average path length in the predictive model, the ability to predict either of the quality metrics did not depend on the clustering, which suggests that the quality of the network may not be as strongly influenced by clustering. The predictive models also revealed the counterintuitive conclusion that an increase in average path length may actually result in an improvement in quality for the whole network, suggesting that perhaps not all short cuts in a small world are created equal. As an alternative, it was considered that a better independent variable for a predictive model might be a ratio of the average path length to a sizing parameter for the network, but when this was attempted in practice, this predictor was not significant in any models.

Despite the predictive capabilities of the models with multiple predictors, neither the mean fitness of the nodes or the attachment-weighted quality could reliably be predicted from only the degree distribution exponent at any given time. This suggests that the scale-free nature of the network may not be the distinguishing factor for a higher quality citation network. It leaves open the possibility that even a random network that exhibits the small world effect could reflect a more effective and efficient citation network than a scale-free network which is not a small world.

Research Question 4: Predicting Quality Using Exponent γ

Using the two proposed systemic measures of the quality of a network, the data also indicated a statistically significant decrease in the mean fitness of the nodes between 1993-2000 and 2001-2008, and a statistically significant increase in attachment-weighted quality between the same two periods. The shift in the mean fitness of nodes indicates that dynamic centrality is indeed a powerful observed characteristic of the evolving citation network. The shift in attachment-weighted quality shows that the most central and influential resources within the network continued to retain their significance to future articles, and furthermore supports the evidence of maturation as the network became a small world in October 1999.

Synthesis: Continuous Improvement of the Research Process

What do these results reveal about how to automate quality improvement of a knowledge base from a citation network? First, it is important to note that the difference between common cause and special cause variation, in terms of the quality measures, has not yet been explored. It would be critical to investigate this factor before the quality metrics explored in this study are used for practical and actionable quality improvement. However, there remain several possibilities for creatively employing automation and techniques for applying artificial intelligence based methods. For example, a genetic algorithm to mutate parts of the citation network and propose ideas for linking disparate topics to create innovative new research is possible. The automated text mining of the citation data in the context of the network topology, using an expert system, might also help to determine emerging leaders and to link related yet previously unconnected topics. Finally, there may be some correlation between Kleinberg's hub scores and the hidden layer in a feedforward neural network, as well as a correspondence between the authority scores and the output from a neural network. Exploring this possible relationship could uncover methods for applying neural networks in automated citation network analysis.

The discussion of results in the previous sections raise the question of what an automated system for continuous improvement on the *QMJ* citation network, and potentially other citation networks, might look like in practice. The present research study reveals four characteristics of such quality improvement systems. First, a continuous improvement system would give authors visibility into the topological and centrality information on the network as it evolves over time. By being able to see new features emerge and new connections being made, the researchers could use the system to stimulate innovation by recognizing links between topic areas that may not be prominent just by searching through references. Second, the system would provide the capability to interactively browse the network. This could change the behavior of researchers as they conduct literature reviews from a random walk to a guided walk through the resources, potentially finding more appropriate resources more quickly as well as covering the literature more comprehensively. Third, the quality improvement system would be constructed to detect shifts in both the mean fitness and the attachment-weighted quality. This would flag times at which the incoming nodes and edges should be more critically examined to determine how they are different than the typical nodes and edges that are added to the system. In some cases, this could reveal an influx in new innovations, and at others it could illuminate when the editorial process has integrated references that do not support the overall mission of the journal. Either way, the information would provide valuable feedback to journal editors and perhaps the entire research community. Fourth, an automated system would enable the researcher or the journal editor to simulate the effects of adding new articles in terms of their citations. This could be an effective vetting mechanism, or it could help editors and authors recognize when it might be useful to prepare a new review article, or hub, on a particular topic.

Actionable Recommendations for QMJ Editors

The presence of multiple indicators for maturation and an increase in attachment-weighted quality around t = 24 (October 1999) prompted inquiries into what substantive or policy change related to the *QMJ* might have impacted the changes. A representative for ASQ confirmed that around this time, there were indeed two major changes: a) a new Editor started at the *QMJ*, and 2) the journal's policies were changed to accept fewer practitioner articles and more academic articles in an effort to boost the quality of the publication. (Wilson, 2009) Not only did the data from this study indicate that some substantive change had occurred, but the data also confirm that the change in policies had a positive impact on the quality of the journal itself.

Using the results, it is possible to extract actionable recommendations for the editors of the *QMJ* that can be used to improve the journal in terms of the attachment-weighted quality of its citation network. These recommendations can be implemented independent of, or in conjunction with, the guidance for an automated system for continuous improvement described above. For example, the collection of citations made by a potential new addition to the *QMJ* should be examined in the context of the full citation network. If the new article has no overlapping references with the pre-existing network, this suggests that it may be off topic for the *QMJ* and not meet the needs of the readership. If the new article only references the most central articles (those with high authority or PageRank scores) this suggests that the author may not have explored the available resources in quality management, and may be missing key linkages to existing research. Reviewers or editors could request that these authors conduct a more extensive literature search prior to publication. If a new article links together two previously unrelated concepts, then editors can recognize this in their introductory remarks, potentially stimulating additional innovation that otherwise would go unrecognized by the readers. Finally, trend analysis can also be conducted to effect quality improvement. If the attachment-weighted quality of the network indicates a downward trend, the editors can explore options such as selectively publishing hub or review articles to draw certain concepts together more tightly. Alternatively, if an upward trend is noted, the editors can explore their publication decisions over the time window that the quality of the network was measured. Using this information, it may be possible to link specific decisions regarding strategy or publishing policy to the impacts of those choices, which could provide valuable insight into how editorial decision making can be enhanced.

General Implications

The results of this study support the view that quality measures based on the topology of the underlying citation network can be useful determinants of network quality, and measures for continually improving the citation network. Furthermore, upon examination of the results of the centrality analysis in this study, as presented in *Table 9* through *Table 14* on pages 83 through 93, this study uncovered lists of the most influential references within quality management research that can be used for course and curriculum development. By helping scholars of quality management become more informed about the resources that have most strongly shaped the research discipline, possibilities for new research will be more effectively and rapidly illuminated and links between critical ideas will be more forthcoming. The results of this study thus contribute to the growing body of research on quality management, and provide needed direction for understanding and improving the research process in that particular area of investigation.

Furthermore, the results indicate that the topological characteristics of citation networks may be useful tools for estimating the collective quality of the research process underlying a discipline. The study demonstrated that attachment-weighted quality may provide an effective metric for monitoring and controlling the quality of the citation network in quality management, an outcome that could help researchers and journal editors in this area proactively shape the discipline. These metrics have similar potential for business managers that oversee corporate knowledge repositories, and members of organizations that supply the knowledge stored in these systems, who can explore using these measures to evaluate and improve the quality of their enterprise systems.

Recommendations for Future Research

The appropriateness and generalizability of this method for quality assessment using a citation network is an area that demands further research. Although the methods in this study show promise for understanding the evolution and improvement of the research discipline in quality management, additional studies would be needed to evaluate consistency of the methods across multiple disciplines and the effects of choosing different sampling strategies. Despite these limitations, the results from the present study conclusively indicate that there is potential in using the mean fitness of the nodes and the attachment-weighted quality as indicators to evaluate the quality of a citation network for any discipline.

There are a myriad of other possibilities for new research that are made achievable as a result of this study. Most significantly, the same techniques could be applied to understand the genesis, evolution, and conceptual structure of other fields. The industrial impact of these methods would benefit from research using the techniques presented in this study to examine the growth, evolution, and quality of a corporate body of knowledge. An actionable method for improving a knowledge base in industry could yield tremendous returns on investment for any company that manages a corporate knowledge repository.

There are several aspects of the current study which could be performed in slightly different ways as well to improve the overall capability of the model. For example, the links within the citation network were all considered to carry the same weight, and the tie strength could be weighted in terms of the academic rigor of the cited material (in terms of academic literature, the “gray” literature of trade magazines and whitepapers, and other materials). This may provide a more pure representation of the relationship between quality management and other disciplines. Additionally, the present study did not correct for self-citation, so the authority of an article could have been unrealistically inflated if an author referenced several of his or her own prior articles. The effect of self-citation in quality management could be examined to determine whether this practice is prevalent in this field. There are several directions for future research involving the algorithms used in this study and the understanding of how centrality impacts quality. For example, an 18-month window was used in this study, but this window could be decreased to see if there is a minimum threshold for detecting quality variability in a citation network, or increased to see how little sampling must be done to detect trends. Determination of an optimal sampling window may be useful to identify how the quality measures could be transformed from lagging indicators to leading indicators. Other mechanisms for defining the timesteps could also be attempted. For example, instead of batching the addition of new nodes, a timestep could be considered any time a new article enters the citation network. Alternatively, the timesteps could be defined in terms of months, weeks, quarters or years to accommodate multi-journal studies, or to support research that also explores the influences of trade journals and business magazines.

Note that the current study did not attempt to determine whether Google PageRank or Kleinberg's authority score is a more effective indicator of quality, although the results suggested that the authority score is more appropriate because it does not randomly give preference to articles that have not been cited frequently. This is a potential topic for future research. Also, this study assumed the existence of dynamic centrality and measured it in terms of the change of in-degrees on a node over time, but did not attempt to uncover *why* that dynamic centrality was operating. By studying the evolution of authority, PageRank, and degree centrality over time, the relative significance of each measure would not only be uncovered, but this would open up the potential to investigate whether physical processes (such as percolation or Bose-Einstein condensation) are evident within the *QMJ* citation network.

Effective coordinated research (and coordinated production of knowledge in general) depends on reliable and useful interactions between human agents and the enabling technologies that help them locate relevant information. Often, these interactions occur in a widely distributed operating environment with no central coordination. Human performance is a critical resource that may be improved by new technologies for facilitating individual interactions with a knowledge repository. This study suggests that topological measures on a citation network may be useful to facilitate systemic quality improvement in these cases.

REFERENCES

Abbott, A. (1988). *The system of professions: An essay for the division of expert labor*. Chicago: University of Chicago Press.

Albert, R. & Barabasi, A.L. (2002). Statistical mechanics of complex networks. *Rev. Modern Physics A.*

Airoldi, E. M. & Carley, K. M. (2005). Sampling algorithms for pure network topologies: a study on the stability and separability of metric embeddings. ACM SIGKDD Explorations Newsletter, 7(2), 13-22.

American Society for Quality (ASQ) (2008a). Quality transforms the world: 2008 ASQ state of the society. American Society for Quality: Milwaukee, WI, p. 11.

American Society for Quality (ASQ) (2008b). Q-BoK Framework: minutes from internal Q-BoK advisory meetings, March 17-19, 2008. Internal Working Paper, American Society for Quality: Milwaukee, WI.

Barabasi, A.-L. & Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5, 101-113.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv:cs/0309023v1.

Benson, P. G. (1993). Letters to the Editor. *Quality Management Journal*, 1(1), 5.

Biehl, M., Kim, H., & Wade, M. (2006). Relationships among the academic business disciplines: a multi-method citation analysis. *Omega*, 34, p. 359-371.

Bilke, S. & Peterson, C. (2001). Topological properties of citation and metabolic networks. *Physical Review E.*

Bollacker, K.D., Lawrence, S., & Giles, C.L. (1998). CiteSeer: an autonomous web agent for automatic retrieval and identification of interesting publications. *2nd International ACM Conference on Autonomous Agents*, 116-123.

Campanario, J.M. (1995). Using neural networks to study networks of scientific journals. *Scientometrics*, 33(1), 23-40.

Carley, K. & Kamneva, N. (2004). A network optimization approach for improving organizational design. *Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-102*

Carolan, B. & Natriello, G. (2006). Data mining journals and books: using the science of networks to uncover the structure of the educational research community. Columbia University, New York, NY.

Clauset, A. , Shalizi, C.R., Newman, M.E.J. (2007). Power-law distributions in empirical data. Preprint, arXiv:0706.1062.

Davenport, T.H., DeLong, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review*, 39(2), 43-57.

Day, A. & Peters, J. (1994). Quality indicators in academic publishing. *Library Review*, 43.

Denning, P.J. (2004). Network laws. *Communications of the ACM*, 47(11), November 2004, 15-20.

Dorogovtsev, S.N. & Mendes, J.F.F. (2002). Evolution of networks. *Advances in Physics*, 51

Fleisher, C.S. (2007). Developing the public affairs body of knowledge. *Journal of Public Affairs*, 7, 281-290.

Ford, D. (2001). Trust and knowledge management: the seed of success. Working Paper WP 01-08, Queen's KBE Center for Knowledge-Based Enterprises, Queen's

School of Business: Kingston, Ontario. Retrieved on October 25, 2006 from http://business.queensu.ca/knowledge/workingpapers/working/working_01-08.pdf.

Fowler, J.H. & Jeon, S. (2008). The authority of Supreme Court precedent. *Social Networks*, 30(1), January 2008.

Giles, C.L. (n.d.) Modern (computational) scientometrics and next generation CiteSeer. Retrieved from http://clgiles.ist.psu.edu on August 30, 2008.

Golomski, W. A. (1993). Introduction. *Quality Management Journal*, 1(1), 4.

Hajra, K. B. & Sen, P. (2004). Aging in citation networks. Preprint submitted to Elsevier Science. arXiv:cond-mat/0409017v1.

Hagan, L. (2000). Improving journal quality with process improvement methods. Whitepaper, The Sheridan Press.

Hamacher, K. (2008). Incomplete information in citation networks. Preprint, arXiv:cond-mat/0411246v3 [cond-mat.other].

Huberman, B.A. & Adamic, L.A. (1999). Growth dynamics of the World Wide Web. *Nature,* 410, 131.

Jezek, K., Fiala, D. & Steinberger, J. (2008). Exploration and evaluation of citation networks. ELPUB 2008 Conference on Electronic Publishing, Toronto, Canada, June 2008.

Jha, S., Noori, H., & Michela, J.L. (1996). The dynamics of continuous improvement: aligning organizational attributes and activities for quality and productivity. *International Journal of Quality Science*, 1(1), 19-47.

Lawrence, S., Giles, C.L. & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer,* 32(6), 67-71.

Le, M. H., Ho, T. B. & Nakamori, Y. (2005). Detecting emerging trends from scientific corpora. *International Journal of Knowledge and Systems Science*, 2(2), June.

Lehmann, S., Lautrup, B., & Jackson, A.D. (2008). Citation networks in high energy physics.

Lehmann, S., Jackson, A.D. & Lautrup, B. (2008). Life, death and preferential attachment.

Leskovec, J., Kleinberg, J. & Faloutsos, C. (2006). Graph evolution: densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data, Vol. 1*, 1-40.

Leskovec, J. & Horvitz, E. (2008). Planetary scale views on a large instant messaging network. Proceedings of WWW2008, Beijing, China, April 2008.

Li-Chun, Y., Kretschmer, H., Hanneman, R.A. & Ze-Yuan, L. (2006). The evolution of a citation network topology: the development of the journal *Scientometrics*. Proc. of the Intl. Workshop on Webometrics, Infometrics & Scientometrics and Seventh COLLNET meeting, Nancy, France.

Lillrank, P., Shani, A.B., & Lindberg, P. (2001). Continuous improvement: exploring alternative organizational designs. *Total Quality Management*, 12(1), 41-55.

Luger, G. & Stubblefield, W. (2004). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (5th ed.), The Benjamin/Cummings Publishing Company, Inc.

Memmesheimer, R-M. & Timme, M. (2006). Designing complex networks. *Physica D*, 224, 182-201.

Menczer, F. (2004). Evolution of document networks. *Proceedings of the National Academies of Science*, 101(1), April 6, 2004, 5261-5265.

Milgram, S. (1967). The small world problem. *Psychology Today*, 2, p. 60-67.

Moore, S., Shiell, A., Hawe, P. & Haines, V.A. (2005). The privileging of communitarian ideas: citation practices and the translation of social capital into public health research. *American Journal of Public Health*, 95(8), 1330-1337.

Newman, M.E.J. (2002). Random graphs as models of networks.

Newman, M.E.J. (2003a). The structure and function of complex networks.

Newman, M.E.J. (2003b). The mathematics of networks. Unpublished manuscript, Center for the Study of Complex Systems, University of Michigan.

Pfeffer, J. (2007). A modest proposal: how we might change the process and product of managerial research. *Academy of Management Journal*, 50(6), 1334-1345.

Pilkington, A. & Teichert, T. (2006). Management of technology: themes, concepts and relationships. *Technovation*, 26, p. 288-299.

Polites, G.L. & Watson, R.T. (2008). The centrality and prestige of CACM. *Communications of the ACM*, 51(1), January 2008.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sanyal, S. (2006). Effect of citation patterns on networks structure. *Physical Review D.*

Scharnhorst, A. & Thelwall, M. (2005). Citation and hyperlink networks. *Current Science*, 89(9), November 10, 2005, p. 1518-1523.

Shibata, N., Kajikawa, Y. & Matsushima, K. (2007). Topological analysis of citation networks to discover the future core articles. *Journal of the American Society for Information Science and Technology*, 58(6), 872-882.

Stumpf, M.P.H. & Wiuf, C. (2005). Sampling properties of random graphs: the degree distribution. *Physical Review E,* 036118.

Van der Veer Martens, B. (2001). Do citation systems represent theories of truth? *Information Research*, 6(2), January 2001.

Vazquez, A. (2001). Statistics of citation networks. Preprint, arXiv:cond-mat/010503v1.

Vazquez, A. (2003). Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations. *Physical Review E*, 67.

Wallace, M. L., Gingras, Y. & Duhon, R. (2008). A new approach for detecting scientific specialties from raw cocitation networks. arXiv:0807.4903v1[soc-ph]

Watts, D. & Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440-442.

Whitworth, B. & Friedman, R. (2008). The challenge of modern academic knowledge exchange. SIGITE Newsletter, 5(2), 4-10.

Wilson, N. (2009). Personal communication.

Yook, S.H., Jeong, H. & Barabasi, A.L. (2001). Weighted evolving networks. *Physical Review Letters*, 86(25), p. 5835.

Zhuge, H. (2006). Discovery of knowledge flow in science. *Communications of the ACM*. 49(5), 101-107.

APPENDIX A: LIST OF ACRONYMS

ANSI: American National Standards Institute

ASQ: American Society for Quality

BA: Barabasi-Albert Model of Preferential Attachment

BOK: Body of Knowledge

EFQM: European Foundation for Quality Management

ISO: International Organization for Standardization

MBNQA: Malcolm Baldrige National Quality Award

MLE: Maximum Likelihood Estimation

PIMS: Profit Impact of Market Studies

Q-BOK: Quality Body of Knowledge (term used internal to the ASQ)

QIC: ASQ Quality Information Center at http://www.asq.org/qic

QICID: Resource identifier used by the ASQ Quality Information Center

QMJ: Quality Management Journal

TQM: Total Quality Management

APPENDIX B: ASQ CERTIFIED QUALITY MANAGER BODY OF KNOWLEDGE

Leadership
  A. Organizational Structures and Culture
  B. Leadership Challenges
  C. Teams and Team Processes
  D. ASQ Code of Ethics

Strategic Plan Development and Deployment
  A. Strategic Planning Models
  B. Business Environment Analysis
  C. Strategic Plan Deployment

Management Elements and Methods
  A. Management Skills and Abilities
  B. Communication Skills and Abilities
  C. Project Management
  D. Quality System
  E. Quality Models and Theories

Quality Management Tools
  A. Problem-Solving Tools
  B. Process Management
  C. Measurement: Assessment and Metrics

Customer-Focused Organizations
  A. Customer Identification and Segmentation
  B. Customer Relationship Management

Supply Chain Management
  A. Supplier Selection
  B. Supplier Communications
  C. Supplier Performance
  D. Supplier Improvement
  E. Supplier Certification, Partnerships, and Alliances
  F. Supplier Logistics

Training and Development
  A. Training Plans
  B. Needs Analysis
  C. Training Materials/Curriculum Development and Delivery
  D. Training Effectiveness and Evaluation

APPENDIX C: EXAMPLE OF A STANDARDIZED CITATION FILE

QICID: 12070

Out Degrees: 7

Title: TQM and Organization of the Firm: Theoretical and Empirical Perspectives

Copyright: 1994, ASQC

Author: *Becker, Selwyn W.; Golomski, William A. J.; Lory, Daniel C.*

Organization: *The University of Chicago Graduate School of Business, Chicago, IL*

Subject: Empowerment,Organizational design,Incentive systems,Problem solving,Supervisors,Implementation,Total Quality Management (TQM);

Series: *Quality Management Journal*, Vol. 1, No. 2, January 1994, pp. 18-24

Becker, S. W. 1993. TQM does work: Ten reasons why misguided attempts fail. *Management Review* (May): 30–33.

General Accounting Office. 1991. U.S. companies improve performance through quality efforts. *Management Practices* (May): 2–42.

Harari, O. 1993. Ten reasons why TQM doesn’t work. *Management Review* (January): 33–38.

Mintzberg, H. 1979. *The structuring of organizations.* Englewood Cliffs, N.J.: Prentice-Hall.

Procter & Gamble. 1992. *Report to the total quality leadership steering committee and working councils.* Cincinnati, Oh.: Procter & Gamble.

U.S. Department of Commerce. 1986–1992. *Survey of current business.* Vols. 66–72.

Walton, M. 1986. *The Deming management method.* New York: Dodd, Mead & Company.

APPENDIX D: `process.pl`: A PERL PROGRAM FOR DATA ASSIMILATION

```
#! /usr/bin/perl5.10.0 -w

# Usage: ./process.pl $QICID_ID_OF_SOURCE_ARTICLE
#               $NEXT_AVAILABLE_NODE_ID

$qicid = $ARGV[0];       # the ASQ identifier for the QMJ article
$node_id = $ARGV[1];     # the next available node_id to use
$dir = "/usr/QMJ2/";
$fn = "$qicid.txt";

$last_node = `wc -l /usr/phdanalysis/node.master`;
$source_node = $last_node + 2;

open(SRC,"$dir$fn") || die "Can't open $dir$fn $!n";
while() {
        $line = $_;
        $pattern = /[12]ddd/;
        $year = $&;

        $search_char = substr($line,0,4);
        print "\n\n\n*******************************************\n";
        print "Will be searching using the substring: $search_char\n\n";
        @matches = `grep "\*$search_char" node.master`;

        if (!@matches) {
                print "No Matches \n";

# CASE 1: Since there are no matches, we KNOW the new reference is
# unique and so we can add it into the @nodemaster and create an
# $edge from $node_id to our new number, $source_node.

# Create an Edge Pointing from this $node_id to NEW$source_node
                $edge = "$node_id $source_node\n";

                # Add New Node Details to @nodemaster holding queue
                $new_entry = "$source_node**$year**$line";
                push(@nodemaster,$new_entry);

                # Now we used a node so we have to increment!
                ++$source_node;
                push(@edges,$edge);

        } else {

# CASE 2: There are some matches, so we want to display those to the
# command line. If the user wants the program to create a new entry in
# the @nodemaster, then she hits the Enter key and no !$target_node is
# specified; we want to create a new one using $source_node.
```

```
                print @matches;
        print "\n>>>>>Source Reference Is>>>>>$line\n\n";
        $prompt = "Enter Node ID to Link To or ENTER for No Match: ";
                $target_node = &promptUser($prompt);
                if (!$target_node) {

                   # Add Node to node.master and Increment Node ID
                   $new_entry = "$source_node**$year**$line";
                   push(@nodemaster,$new_entry);

                   $edge = "$node_id $source_node\n";
                   push(@edges,$edge);
                   ++$source_node;
                }

# CASE 3: The user has detected that there is a pre-existing match. We
# don't need to touch @nodemaster, but we do need to add an $edge from
# $node_id to the $target_node the user entered at the $prompt.

                if ($target_node) {

                   # This means our match is already in the database and
                   # we want to link the current paper, $node_id, to the
                   # $target_id that we just entered:

                   $edge = "$node_id $target_noden";
                   push(@edges,$edge);

                }
        }
}
close(SRC);

# Now Append All New Nodes to node.master

print "\n\n\nNEW NODES:\n";
print(@nodemaster);
open(NODEMASTER, ">>node.master") || die "Can't open node.master $!n";
   print NODEMASTER @nodemaster;
close(NODEMASTER);

# Now Update edge.master

print "\n\n\nEDGELIST:\n";
print(@edges);
open(EDGEMASTER, ">>edge.master") || die "Can't open edge.master $!n";
   print EDGEMASTER @edges;
close(EDGEMASTER);

#--------------------------(  promptUser  )---------------------------#
#  FUNCTION:   promptUser
#
#  PURPOSE:    Prompt the user for some type of input, and return the
#
```

```
#               input back to the calling program.
#
#  ARGS:        $promptString - what you want to prompt the user with
#
#               $defaultValue - (optional) a default value for the
prompt #
#----------------------------------------------------------------------#

sub promptUser {

   #-----------------------------------------------------------------
   #  two possible input arguments - $promptString, and $defaultValue
   #  make the input arguments local variables.
   #------------------------------------------------------------------

   local($promptString,$defaultValue) = @_;

   #-------------------------------------------------------------------
   #  if there is a default value, use the first print statement; if
   #  no default is provided, print the second string.
   #-------------------------------------------------------------------

   if ($defaultValue) {
      print $promptString, "[", $defaultValue, "]: ";
   } else {
      print $promptString, ": ";
   }

   $| = 1;               # force a flush after our print
   $_ = ;         # get the input from STDIN (presumably the keyboard)

   #-------------------------------------------------------------------
   # remove the newline character from the end of the input the user
   # gave us.
   #-------------------------------------------------------------------
   chomp;
   #------------------------------------------------------------------
   #  if we
had a $default value, and the user gave us input, then
   #  return the input; if we had a default, and they gave us no
   #  no input, return the $defaultValue.
   #
   #  if we did not have a default value, then just return whatever
   #  the user gave us.  if they just hit the  key,
   #  the calling routine will have to deal with that.
   #------------------------------------------------------------------

  if ("$defaultValue") {
      return $_ ? $_ : $defaultValue;    # return $_ if it has a value
   } else {
      return $_;
   }
}
```

APPENDIX E: SELECTED R SCRIPTS FOR DATA ANALYSIS

```
# Describe Network

library(igraph)
library(netmodels)
source("C:\\Users\\nicole\\Documents\\classes.R")
source("C:\\Users\\nicole\\Documents\\plfit.r")

g <- read.graph("finished/edge-master-29")
d <- degree(g,mode=c("in"))
dno0 <- d[which(d>0)]
dd <- degree.distribution(g,mode=c("in"))

do.small.world(g)
diameter(g)
graph.density(g)
plfit(dno0)

# Clusters

# Number of components
clusters(g)$no

# Size of the largest component
max(clusters(g)$csize)

# Proportion of the network occupied by the largest component
max(clusters(g)$csize)/sum(clusters(g)$csize)

# Proportion of Isolates
dd[1]


# Take the subgraph that is the largest component
# This step determines the "membership" number for the largest
component
gcn.group <- which(clusters(g)$csize==max(clusters(g)$csize))-1
gcn.members <- which(clusters(g)$membership==gcn.group)
gcn <- subgraph(g,(gcn.members-1))

# Average path lengths: TRUE is path length for a directed network
# This is gcn.path.undir
average.path.length(gcn,FALSE)
# This is gcn.path.dir
average.path.length(gcn,TRUE)
# This is just to compare the average path length in g vs. gcn
(directed)
average.path.length(g,TRUE)
average.path.length(gcn,TRUE)-average.path.length(g,TRUE)
```

```
# Centrality Script

library(igraph)
library(netmodels)
g <- read.graph("ALLMYEDGES")
d <- degree(g,mode=c("in"))
outd <- degree(g,mode=c("out"))

# Assign Attributes
V(g)$name <- seq(vcount(g))-1
V(g)$core <- graph.coreness(g)

dno0 <- d[which(d>0)] # this is the 6889-element vector of degrees that
has no zeroes
outdno0 <- outd[which(outd>0)]

stem(d)
cendf <- data.frame(V(g)$name,V(g)$core,d,outd)

# Make a data frame containing (node_ids, indegree), order it and take
the top ones
indf <- data.frame(V(g)$name,d)
indfsorted <- indf[rev(order(indf[,2])), c(1,2)]
indfsorted[1:20,] # THIS IS THE TOP 20

# Do the same thing with Google PageRank
pr <- page.rank(g,directed=TRUE)
prbig <- (pr$vector*10000)
prdf <- data.frame(V(g)$name,prbig)
prdfsorted <- prdf[rev(order(prdf[,2])), c(1,2)]
prdfsorted[1:20,]

cendf$pr <- prbig

# Do the same thing with authority.score
au <- authority.score(g)
aubig <- (au$vector*100)
audf <- data.frame(V(g)$name,aubig)
ausorted <- audf[rev(order(audf[,2])), c(1,2)]
ausorted[1:20,]

cendf$auth <- aubig

# Do the same thing with hub.score
hu <- hub.score(g)
hubig <- (hu$vector*100)
hudf <- data.frame(V(g)$name,hubig)
husorted <- hudf[rev(order(hudf[,2])), c(1,2)]
husorted[1:20,]

cendf$hub <- hubig

# Do the same thing with Bonacich Power
bon <- bonpow(g)
bonbig <- bon*100
```

```
bondf <- data.frame(V(g)$name,bonbig)
bonsorted <- bondf[rev(order(bondf[,2])), c(1,2)]
bonsorted[1:20,]

cendf$bonpow <- bonbig
cendf[1:10,]

# Power Law Fitting
# Some of this code adapted from
# http://www.bigre.ulb.ac.be/Users/jvanheld/BMOL-F-#
#501/practicals/r_scripts

source("C:\\Users\\nicole\\Documents\\classes.R")
library(igraph)
library(netmodels)


g <- read.graph("all-edges/edge-master-50")
d <- degree(g,mode=c("in"))
outd <- degree(g,mode=c("out"))


# Assign Attributes
V(g)$name <- seq(vcount(g))-1
V(g)$core <- graph.coreness(g,mode=c("in"))


# Keeping my Variables Straight
dno0 <- d[which(d>0)] # this is the 6889-element vector of degrees that
has no zeroes
outdno0 <- outd[which(outd>0)]


pk <- tabulate(dno0)/sum(dno0)
pkout <- tabulate(outdno0)/sum(outdno0)
k <- seq(1:80)
plot(k,pk,log="xy") # is traditional degree distribution



degree.freq <- data.frame(
degree=as.numeric(row.names(table(d))),
frequency=as.vector(table(d))
)


mean.degree <- mean(d)
no0 <- degree.freq$degree > 0
node.nb <- length(d) # number of nodes
fit.domain <- no0
dataset <- "Network at t=60"
do.dd <- "no"

# # # # #
```

```
degree.freq$iCDF <- rev(cumsum(rev(degree.freq$frequency)))
degree.freq$rel.freq=degree.freq$frequency/sum(degree.freq$frequency)
degree.freq$rel.iCDF=degree.freq$iCDF/sum(degree.freq$frequency)
degree.freq$degree.log <- log(degree.freq$degree,base = 10);
degree.freq$frequency.log <- log(degree.freq$frequency,base = 10);
print(degree.freq[1:10,]) # Make sure all the nunmbers calculated right


degree.freq$poisson.density <- dpois(x=degree.freq$degree,
lambda=mean.degree)
degree.freq$poisson.exp <- degree.freq$poisson.density * node.nb


# CALCULATE VALUES FOR ALL DATA POINTS
degree.freq$power.exp <-  NA
fit.domain <- no0
distrib.fitting <- lm(frequency.log ~ degree.log, data =
degree.freq[fit.domain,])
degree.freq[fit.domain,'power.exp'] <- 10^(predict.lm(distrib.fitting))
print(degree.freq[1:10,])




# CALCULATE VALUES FOR FIRST $first NUMBER OF DATA POINTS
first <- 20
degree.freq$power.exp.first <-  NA
fit.domain <- 2:first
distrib.fitting.first <- lm(frequency.log ~ degree.log, data =
degree.freq[fit.domain,])
degree.freq[fit.domain,'power.exp.first'] <-
10^(predict.lm(distrib.fitting.first))



if (do.dd == "yes") {
# THIS PRINTS DEGREE DISTRIBUTION ONLY
plot(degree.freq[no0,'degree'],
     degree.freq[no0,'frequency'],
     col='darkgreen',
     type = 'p',
     main = paste('Fitting of Poisson and power law distribution\n
on',dataset),
     xlab = 'Degree',
     ylab = 'Number of nodes',
     panel.first=grid (col = 'black',equilog=F),
     log='xy'
     );
}

# THIS IS THE ONE THAT PRODUCES CHART ON A LOG LOG PLOT
plot(degree.freq[no0,'degree'],
     degree.freq[no0,'frequency'],
     col='darkgreen',
     type = 'p', pch=16,
```

```
     main = paste('Fitting of Poisson and power law distribution\n
on',dataset),
     xlab = 'Degree (k)',
     ylab = 'Number of Nodes with Degree k',
     panel.first=grid (col = 'black',equilog=F),
     log="xy"
     );
lines(degree.freq$degree, degree.freq$poisson.exp, col = 'blue')
lines(degree.freq$degree, degree.freq$power.exp, col = 'red')
lines(degree.freq$degree, degree.freq$power.exp.first, col = 'green')
legend ('topright', col = c('darkgreen', 'blue', 'red', 'green'),
        legend = c('Observed Degree Frequency', 'Poisson fit (all data,
k > 1)', 'Power law fit (all data, k > 1)',paste('Power law fit
(first',first,'degrees)')), lwd = 2,bg='white',bty='o')

ks.test(degree.freq$power.exp.first,degree.freq$frequency,alternative=c
("two.sided"))
summary(lm(distrib.fitting.first))

ks.test(degree.freq$power.exp,degree.freq$frequency,alternative=c("two.
sided"))
summary(lm(distrib.fitting))

plfit(dno0)



# THIS IS THE ONE THAT PRODUCES CHART ON A NON LOG LOG PLOT
plot(degree.freq[no0,'degree'],
     degree.freq[no0,'frequency'],
     col='darkgreen',
     type = 'o',
     main = paste('Fitting of Poisson and power law distribution\n
on',dataset),
     xlab = 'Degree (k)',
     ylab = 'Number of Nodes with Degree k',
     panel.first=grid (col = 'black',equilog=F),
     );
lines(degree.freq$degree, degree.freq$poisson.exp, col = 'blue')
lines(degree.freq$degree, degree.freq$power.exp, col = 'red')
lines(degree.freq$degree, degree.freq$power.exp.first, col = 'green')
legend ('topright', col = c('darkgreen', 'blue', 'red', 'green'),
        legend = c('Observed Degree Frequency', 'Poisson fit (all data,
k > 1)', 'Power law fit (all data, k > 1)',paste('Power law fit
(first',first,'points)')), lwd = 2,bg='white',bty='o')
```