THE DYNAMICS OF PUBLICATION AND CITATION NETWORKS IN ACADEMIA

by

Burak Eskici

B.S., Industrial Engineering, Boğaziçi University, 2006

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering

Boğaziçi University

2009

THE DYNAMICS OF PUBLICATION AND CITATION NETWORKS IN ACADEMIA

APPROVED BY:

Prof. Yaman Barlas (Thesis Supervisor)	…………………
Assist. Prof. Aybek Korugan	………………...
Prof. Orhan Yenigün	………………...

DATE OF APPROVAL: 05.10.2009

to my parents İsmail and Münevver Eskici

ACKNOWLEDGEMENTS

First of all I would like to express my deepest gratitude to Prof. Yaman Barlas, my thesis supervisor, for his invaluable guidance and advisory throughout this study. Yet, supervising this thesis a just a small portion of his contribution to my academic career, knowledge and worldview. As a professor, as an advisor, as a friend, as a public intellectual he has added lots of values to my life. I would not have been myself, if we had not met.

I would like to thank to Assist. Prof. Aybek Korugan and Prof Orhan Yenigün, the members of my thesis committee, for their valuable comments and suggestions.

Moreover, as an industrial engineer, I would like to thank to professors at the Industrial Engineering Department of the Bogazici University. They have contributed a lot to my life in the most important time. Additionally, I would like to thank to all my teachers in the high school, secondary school and elementary school, for enlightening my path with the most powerful light, modern scientific education.

I am also very grateful to Bilge Küçük, Nisa Güler and Selin Tamçakır, who did a graduation project on the same topic of this thesis. Their work, model and analyses have formed the base for this study.

As a member of the group for two years, I would like to thank to Sesdyn Research Group friends, namely Güven Demirel, Emre Demirezen, Emre Keskin, Nisa Güler and Onur Özgün. All of them have been very tolerable, helpful and friendly. Among them I would like to exprees my special thank to Güven Demirel, who has helped me in the initial stages of the agent based model construction, and Nisa Güler, with whom we have written a conference paper from this project.

Finally I wish to express my deepest gratidute to parents, İsmail and Münevver Eskici, for eveything they did for me in all aspects of my life. They have been always like lighthouses to find my path in the life course. I would not be able to find it without their exıstense. I would also thank to my brother, Bahadir Eskici, who has always been a wonderful friend and guide.

I have been supported by TUBİTAK domestic masters degree fellowship program – 2228 – for two years. The ideas, views and conclusions expressed herein do not represent the official view of TUBİTAK.

ABSTRACT

THE DYNAMICS OF PUBLICATION AND CITATION NETWORKS IN ACADEMIA

In academia, one of the problems with publications can be summarized as “90% of the citations are received by 10% of the publications”. Ineffective scientific policies drive problematic initiatives for doing spurious publications and sustaining the existing situation. Operationalizing award and punishment systems by counting publications and citations, at least needs closer attention. In this thesis I analyze three aspects of scientific publication and citation dynamics. Firstly, historical data analysis shows that average publication statistics and tendencies are different in different fields. In terms of co-authorship and network structures, at least in two exemplary fields, system dynamics and operations research, it is shown that they are substantially different. Moreover, publication and citation network data show certain sub-group formations in citation networks, but further investigations on whether there is an effect of social relationships on publication-citation networks show no evidence on the existence of such an effect in exemplary fields. Second part of the thesis, a system dynamic model, describes how academicians in a complex environment behave under two opposite pressures, publication and citation pressures. Intellectual skill levels and reputation are two important decisive factors on the success of publications. Last part of the analysis, agent based model, shows that different initial conditions of a field have a substantial effect on the emerging network structure among academicians. Young scholars, who try to enter into fields which are filled by senior authors, have difficulties in receiving enough citation, compared to the ones who are trying to enter into fields with junior academicians. This suggests that young scholars in different fields with different seniorities are not in the same situation and comparing the research performances of young researchers by just publication and citation numbers may not be valid and fair.

ÖZET

AKADEMİK DÜNYADA YAYIN VE ATIF AĞLARININ DİNAMİKLERİ

Akademik yayınların aldıkları toplam atıfın yüzde 90'ının yayınların yüzde 10'u tarafından paylaşıldığı gerçeği, yayın ve atıf dünyasında bilinen bir sorunu özetlemektedir. Akademik politikalar iyi tasarlanmadıkları takdirde yüzeysel ve çabuk yayın yapılmasına teşvik edip, bu durumun devam etmesine neden olabilirler. Bu yüzden akademik ödül ve yaptırımların, yayın ve atıf sayılarına bağlanması en azından daha dikkatli bir incelemeyi gerekli hale getiriyor. Bu tezde, bu ihtiyaca yönelik olarak, akademik yayın ve atıf dinamiklerini üç farklı açıdan incelenmiştirç İlk olarak, tarihsel veri analizi, farklı alanlarda ortalama yayın istatistiklerinin ve bu değerlerdeki eğilimlerin farklı olduğunu göstermektedir. Özellikle ortak yayın yapma alışkanlıkları ve ağ yapıları açısından, en azından örnek alanlar olan sistem dinamiği ve yöneylem araştırması alanları birbirlerinden oldukça farklı çıkmaktadır. Diğer taraftan yayın ve atıf ağ verileri, atıf ağlarında küçük alt grup oluşumlarını göstermektedir. Ancak bu ağ oluşumlarının sosyal ilişkilerle bir bağlantısı olup olmadığı konusundaki inceleme, en azından mevcut verilerde ve seçilen alanlarda böyle bir ilişkiyi ortaya çıkarmamıştır. Tezin ikinci kısmı olan sistem dinamiği modeli ise bilim adamlarının oldukça karmaşık bir ortamda ve birbirine zıt şekilde hareket eden yayın yapma ve atıf alma baskıları altında karar vermelerini incelemektedir. Ayrıca akademisyenlerin yetenekleri ve kazandıkları ün, yayın ve atıf davranışları üzerinde oldukça etkili iki etken olarak kendilerini göstermektedir. Tezin son kısmı olan etmen temelli model ise herhangi bir alanın şartlarının, alanın ilerideki ağ yapısı üzerinde önemli etkileri olduğu ortaya koymuştur. Daha deneyimli ve başarılı bilim adamlarıyla dolu bir alana girmeye çalışan genç akademisyenlerin, daha genç ve az yayın yapmış bilim adamları ile dolu bir alana girmeye çalışan genc akademisyenlere göre atıf almakta oldukça zorlandıkları görülmektedir. Oluşan ağ yapısının bu özelliği, farklı olgunluk düzeyinde olan farklı alanlardaki genç akademisyenlerin aynı durumda olmadığını ve onları sadece yayın ve atıf sayılarla karşılaştırmanın çok geçerli ve adil olmadığını göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION AND RESEARCH OBJECTIVES

The word *"science"* is derived from the Latin word *"scientia"* for knowledge, which in turns comes from *"scio"*, I know. The word is defined in the Encyclopedia Britannica as

> "*any system of knowledge that is concerned with the physical world and its phenomena and that entails unbiased observations and systematic experimentation"* (Britannica_Encyclopedia).

In ancient time the science may have referred to a humble effort to know; however, today's science is a complex system which contains a lot of components in addition to desire to know. The question of how this system has emerged; managed and should be in the future, is the main issue of history and philosophy of science (Price, 1975).

Academic knowledge proceeds by accumulation and an academician tries to make contribution to this accumulation by publication. He / she aims to be the part of the common knowledge and shares his/her work by publishing, but it does not mean that every publication makes a good contribution. To get citation in others' papers has that indicator role. Citation can be defined as the glue that binds a research paper to the body of knowledge in a particular field (Hamilton, 1990). The journey of a certain research is that somebody does research and publishes it, and then others use it to do new research. It can be conceptualized as generating, distributing and consuming the scientific knowledge (Price, 1964). In this journey the key factor is to measure the contribution of a paper to the field, because it is very important to decide what to be read, especially in the expansion of scientific literature (Margolis, 1967).

The question is how the quality of a paper and an academician can be measured. For this question *"scientometrics"*, the science of measuring and analyzing science, has emerged. Today, scientometrics has different kinds of measures all of which have been built on number of publications and citations (Garfield, 1970).

The rapid expansion of the academic work has required an objective measure, and first, the number of publications is considered adequate to decide whether a researcher is successful or not. However, since the number of papers does not say anything about the quality of the work, indexing – citation issues have entered the measurement of science (Garfield, 1970). Then there has emerged some publication citation balance measures (Like h-index, g-index). Today, h-index[1] and average number of citations per paper are widely used measures for an academician. However, a better measure is a subject of future work and a necessity for a healthy progression of science, because there is a question mark on the effects of existing measures to the evolution of science and the picture of today is a little bit problematic.

An article by David P. Hamilton published in 1990 at Science Magazine calls attention to the statistics about the contribution of new publications to the knowledge (Hamilton, 1990). The name of the article is *"Publishing by – and for? – the Numbers"* and Hamilton puts the following sentence under the heading: *"New evidence raises the possibility that a majority of scientific papers make negligible contributions to knowledge."* The evidence is some statistics from Institute for Scientific Institution (ISI – web of science) which shows that 55% of the papers published between 1981 and 1985 in journals indexed by ISI received no citations at all in the 5 years after they were published.

David Pendlebury, an ISI statistician, who derived the figure above, argues that the conventional wisdom in the field is that *"10% of the journals get 90% of the citations"*. In other words 90% of the journals get just 10% of the total citations. This end result of publications seems very shocking and controversial, because it is exactly the opposite to the aim of academic publication. Why are we trying to publish something? The answer should be to make other researchers be aware of our findings and by this way to invoke new research. Shortly the aim should be to contribute to the literature. However, if more than half – perhaps more than three quarters – of the scientific literature is essentially not read by anybody (someone calls this as 'worthless' or 'trash'), isn't there a problem? At this point words of Allen Bard (editor of the Journal of the American Chemical Society, [in 1990]) give some clue about the cause:

> *"In many ways, publication no longer represents a way of communication with your scientific peers, but a way to enhance your status and accumulate points for promotion and grants"*.

---

[1] An academician with an H-index of *h*, has at least *h* number of papers each of which has earned at least *h* citations.

This raises the suspicion that the academic culture encourages the spurious publication.

The academic policies put some pressure on scholars to publish. The famous saying "Publish or perish" shows the amount of this pressure. The question is how this pressure – or academic policies – affects the researchers? It is obvious that the number of publications has been increasing – i.e. quantity of production; however, what about the quality?

Moreover if this pressure surpasses the academic desires of individuals, in other words if researchers are not making research for their curiosity (their own desire), but their need for grants, keeping the position (i.e. for padding the resume), then in Marxian terms, they are like the alienated labor who is forced to sell his labor force which is his essence of existence. If this is true, we face researchers who is alienated to his work (research area), to his research process, to his colleagues and to himself/herself.

Related to this, what might be the long term consequences of these academic policies? Trying to measure the success with numbers generate the situation in which there are too many people who are willing to publish and too few people who are willing to read, because of time constraint. So the question coming to mind is whether the academic policies might limit our creativity.

All in all, the perception of publication by academicians and the quality of publications are the outcomes of a complex system in which researchers, policy makers and institutions have certain roles. In order to grasp the situation better, I aim to analyze the tendencies in publication and citation behaviors by using the historical data, by constructing a system dynamics model and an agent-based model.

This thesis has three analytical chapters all of which try to analyze a different aspect of the problem. The first analytic chapter (Chapter 3) tries to grasp the network structure relying on real data. On the other the second one (Chapter 4) tries to focus on individual behavior with a system dynamics model. The last chapter (Chapter 5) then turns back to networks and tries to generate fictitious data of inter-citation between authors, real data of which is not available.

The analysis begins with Chapter 3, which is the analysis of real network data. Operations research and system dynamics are chosen as the exemplary fields. The former one is the representative of a big, heterogeneous field and the latter one corresponds to a relatively compact and homogeneous field. In these two fields the collected data of different network types, such as co-citation, collaboration, inter-citation, has been analyzed. Especially any possible effect of shared institution or being in closer region is the main issue of analysis.

The aim of the following chapter is to examine the behaviors of researchers in response to the dynamics of publication and citation pressures. In Chapter 4 a model[2] including faculty members in a department, and their publication and citation dynamics has been constructed by using system dynamics methodology. In addition to running the base model some scenarios, like different levels of pressure or intellectual skills have been tested.

The historical data analyses chapter and system dynamics model chapter lead to an agent based model (i.e. process oriented) which is the content of the Chapter 5. The inter-citation[3] data between authors are not available and the agent based model tries to simulate the real publication citation behavior in order to generate the missing data.

All in all, this entire thesis aims to analyze the dynamics behind the everyday publication activities of scholars. It doesn't argue to be able to solve the entire puzzle but I hope it will contribute to the existing knowledge and stimulate new research.

[2] Model is built upon the model in the graduation project of Güler,Tamçakır and Küçük (Küçük, et al., 2007)
[3] The data of who cites who and how many times.

2. LITERATURE REVIEW

The problem that this thesis is dealing with has different aspects. For that reason in order to give a comprehensive background the literature review chapter has three components. First part is about the idea of counting publication and citation, and indexing them. I try to summarize the development of scientometrics which defined as the *"science of measuring and analyzing science"*. The second part of this chapter deals with the critical arguments against citation-indexing tradition and problems that I try to handle in this thesis. Lastly I summarize some of the related studies on similar research questions.

The origins of citation indexing go back to 14th century conceptually. In practice it is widely accepted that 1950 is the turning point with Science Citation Index (SCI) which has the catalytic effect on the expansion and popularity of the bibliometric research. Although one century before Shaphard's citation index in law is similar to that we have today, SCI has been accepted as the starting point of research on this issue with massive datasets. (Cronin, 2001)

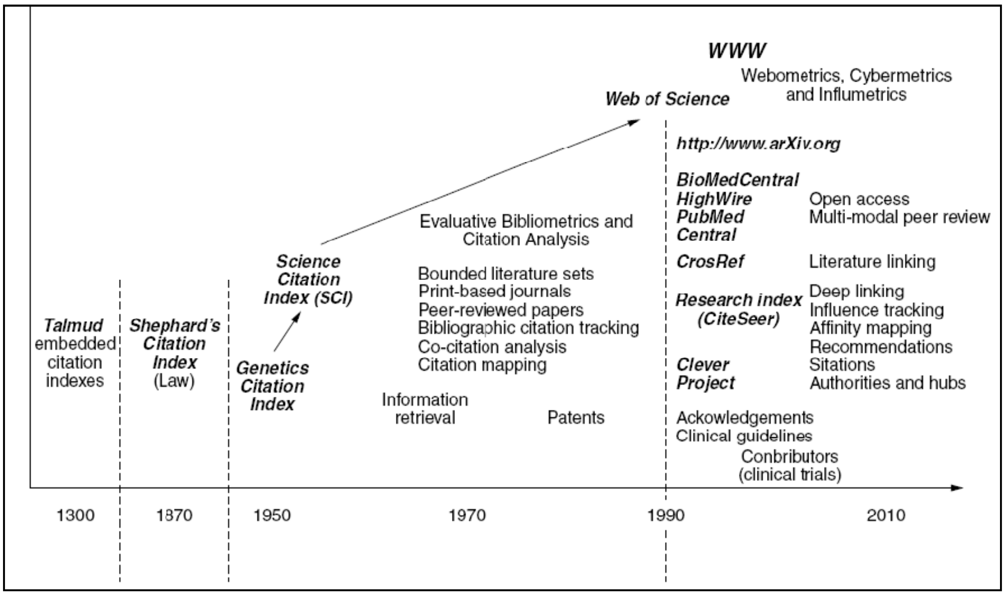

Figure 2.1. The Evolution of citation indexing[4]

[4] The figure is taken from (Cronin, 2001)

Eugene Garfield and Derek De Solla Price are identified as the two founders of the scientometrics field, which has been accepted as initiated with science citation index (SCI) in 1950's.

Derek De Solla Price (1922-1983) who is an English physicist, historian of science and information scientist, worked on citation links among articles in *Philosophical Transactions of the Royal Society of London*[5] and tried to identify some patterns. He is seen as the conceptual and practical father of the field. Moreover in his book, '*Science since Babylon'* (Price, 1975), he has discussed the disease of science and first mentioned the exponential growth of scientific literature. In his seminal book *'Little Science, Big Science'* (Price, 1963) Price has discussed not only philosophical aspects, but findings on empirical research. Details of the book are out of the scope of this review but I think, the articles by Jonathan Furner on Price's book and contribution might be helpful. (Furner, 2003a) (Furner, 2003b).

The other founding father of the field is Eugene Garfield (1925-) who is the founder of *Science Citation Index* (SCI) (Garfield, 1955) and *Institute of Scientific Information* (ISI). For more than 50 years he has published several articles (more than 400) on citation indexing. (Garfield, 1970). The span of his writings is very wide[6].

Lastly about the field of scientometrics I would like to mention some review articles. The first one is by Hood and Wilson (Hood, et al., 2001) which covers the topic until the end of 20th century. The second one deals with the more recent developments in the field and mostly on web related measures (Thelwall, et al., 2005). The last and most comprehensive review paper deals with the developments between 2000 and 2006 (Bar-Ilan, 2008).

On the other hand there are some critiques to the citation analysts. The critical review by Macroberts covers the vast majority of the critical point of views which starts with the underlying assumption: *"references cited by an author are roughly valid indicators of influence on his work"* (MacRoberts, et al., 1989). The problem of the validity of the main

[5] It is established in 1665 and commonly regarded as the world's first scientific journal (Furner, 2003)
[6] All of publications by Eugene Garfield can be reached through http://www.garfield.library.upenn.edu

assumption has many aspects and each aspect has its own problem: Some of the problems are: a) Formal influences might not be cited (due to lack of awareness or oversight by the author). b) Biased citing. c) Informal influences are not cited. d) Self-citation (10-30 % of all citations) e) Different types of citation (such as affirmative or negational) f) Variation in citation rate with type of publication, nationality, time period and size-type of the specialty. g)Technical limitations of citation indices and bibliographies (like multiple authorship, synonyms, clerical errors, coverage of indices). (MacRoberts, et al., 1989). Almost all the problems mentioned above have their empirical evidence.

In addition to conceptual and practical problems, the end result of the citation indexing and science policies might be argued as being not desirable. As mentioned in the introduction chapter, the conventional wisdom is that 90% of the citation is received by the 10% of the articles (Hamilton, 1990). This means that majority of the papers do not contribute to science a lot. If the vast majority of the products of the academia do not contribute to science at all, then there is a problem. In another study, Hamilton shows that the uncitedness ratio (percentage of the papers which do not get any citation in four years after publishing) differs from field to field (Hamilton, 1991). For example engineering – general– and political science are the leading fields with the highest uncitedness ratio (86.9% and 90.1% respectively). These high ratios point some structural, systemic problems.

Problems caused by the citation indexing policies are not only structural, but also related to individual characteristics of scholars. If all the entire academic community had behaved ideally and ethically, the problems mentioned above would not have been that much serious. In an early paper, while there was not a problematic situation, Price has warned about the ethical issues (Price, 1964). First the perception of academic work should be in this way: *"the scientific publication should be considered a privilege consequent upon the finding of something which people may need to read"* (Price, 1964). Additionally scholars should be aware of the problems mentioned above and be very careful in awarding credit and citing. So academicians have an individual responsibility to have a proper development of science.

Lastly in the literature review, I would like to summarize some of the applications related to my topic. Two early examples are by Margolis (Margolis, 1967) and by Price (Price, 1965). The former one deals with measuring the impact of an early paper and historical data, whereas the later is one of the very early examples of the network analysis in the field. The study which is the closest to this thesis in terms of the underlying question is that of White et al. (White, et al., 2004). The authors have checked the relationship between social relationship and citation relationship among 16 scholars of a closed network and this micro-focused study could not find a strong support for the relationship.

Moreover, the problem has also been studied by complex system scholars and statistical physicists. Redner has figured out the citation statistics of *Physical Review* in 110 years, and proposed different citation models (Redner, 2005). Newman who is another researcher in these fields, have studied collaboration networks in different fields, such as physics, biomedical and computer science (Newman, 2001a) (Newman, 2001b). Apart from these data analyses and mathematical models, there are also simulation model attempts. Two of them are by Menczer and Borner et al. The former one has modeled the evolution of the paper networks in World Wide Web (Menczer, 2004), whereas the latter one has proposed a model which simulates the evolution of author and paper networks (Borner, et al., 2004).

All in all the literature review basically covers three different domains. Firstly it summarizes some of the literature on how the citation indexing system has developed. Then It covers some critical reviews on citation indexing and problems related to issue. Lastly it illustrates some studies similar to my thesis. Although there is some literature on complex studies or network theory, their literature review is beyond the scope of this thesis.

3. DATA AND FIELD ANALYSIS

The aim of this chapter is to analyze the publication-citation data in exemplar fields and to draw a general picture of the historical trends and network structures. Basically the analysis is descriptive and far from drawing causal conclusions. However, I believe in that the descriptive power of network methodology is enough to give an insight about the structural properties of the problem.

This chapter consists of three main parts. In the first part I discuss the data, sample and some certain aspects of social network analysis. In addition to these, I also give brief information about the data analysis software packages that I have used. In the second part of this chapter, I give some brief information on the historical trends and evolution of different types of fields. And in the last section of this chapter the author networks are shown which is one of the most important parts of the analyses.

3.1. Data And Sample

All the data are collected form Web of Knowledge – ISI data base, which is the most comprehensive scientific database and covers more than 23.000 scientific journals from 1900's. I choose operations research (OR) and system dynamics (SD) fields as the exemplary. The former one represents a general, broad engineering field which has different sub-fields and collaborations with other literatures; whereas the latter one represents a relatively small and closed field. In order to analyze the structure of OR literature I pick two of the most prestigious journals of the field, namely Operations Research (ORJ) and Management Science (MS). Both of the journals are type-A journals with an impact factor of 1,47 and 2,35 in 2007, respectively. In order to analyze the structure of the System Dynamics community I chose the journal System Dynamics Review (SDR) which is the most prestigious journal of the field with the impact factor of 1.415 in 2008[7]. Table 3.1 summarizes the main characteristics of the sample which covers 25 years of ORJ and MS and 16 years of SDR including only articles. I do not include last 4 years on purpose, because

[7] The source is Wiley InterScience.

usually an article needs some time to reach the audience. In addition to that I only include articles as the document type, because some other types, such as editorial or letters, are not reflecting the citation behavior that I want to study.

Table 3.1. Summary of sample drawn from ORJ, MS and SDR

	Time	Type	Number of Articles
SYSTEM DYNAMICS REVIEW	1991 - 2006	Articles	225
OPERATIONS RESEARCH	1981 - 2005	Articles	1983
MANAGEMENT SCIENCE	1981 - 2005	Articles	2778

3.2. Social Network Analysis Methodology

Before going on to the analysis, this section is aimed to give brief information about social network analysis. Social network analysis (SNA) is defined in (Otte, et al., 2002) as:

> "Most broadly, social network analysis (a) conceptualizes social structure as a network with ties connecting members and channeling resources, (b) focuses on the characteristics of ties rather than on the characteristic of the individual members, and (c) views the communities as 'personal communities' that is, as networks of individual relations that people foster, maintain, and use in the course of their lives."

Why do I use social networks to study citation behavior? The answer is highly related to the nature of the citation data which is relational. When we try to test hypotheses focused on relational data, using standard statistical analysis, which assumes independent observations, is not appropriate; because relational hypotheses deal with interdependencies (relations) among actors. Assuming interdependence between the number of citations from A to B and number of citations form B to A does not seem to be realistic. For that reason social network analysis seems appropriate for analyzing the citation relations.

3.2.1. Some Notions of Social Network Analysis

Networks are defined by two different elements: nodes and links (or actors and edges respectively). Networks are divided into two groups with respect to whether the link is directed or not. If link from i^{th} actor to j^{th} actor is identical with the link from j^{th} actor to i^{th} actor for all set of (i,j) actor pairs, then the network is called undirected network. Moreover networks are divided into two groups in another dimension which is whether the links are valued or dichotomous. If the links show just the absence or presence of a relationship then, the network is called dichotomous network, otherwise it is called valued (or weighted network).

In addition to types of networks, some basic measurements to compare networks are also worth to mention here. First one is density which is the indicator for the level of connectedness of a network. Density takes values between 0 and 1, and if all possible links are realized it has the value of one and this is a measure of whole network. Second measure is degree centrality which is an actor property and equal to the total number of connections that an actor has. In directed networks degree centrality has two components as in-degree centrality and out-degree centrality. While degree centrality measures the centrality of an actor as counting the number of links, there are other centrality measures, such as closeness and betweenness which measure centrality by looking at the position of an actor in the entire network. Lastly cliques are the subgroups in networks which consist of highly connected three or more nodes.

3.2.2. Network Types in Scientometrics

In scientometrics literature, three types of networks have been analyzed up to now. These are co-citation networks, co-authorship networks and inter-citation networks.

3.2.2.1. Co-citation Networks: Co-citation networks of both authors and papers can be constructed. In the literature mostly co-citation networks among authors have been analyzed. In this network, the nodes (actors) are authors and the link between two nodes is formed if both the authors are cited by a third party. Co-citation networks are undirected, and they might be dichotomous or valued, with respect to whether number of co-citedness is counted or not. Co-citation networks are important in understanding the evolution of the fields, because it shows which two authors are cited together. Moreover, the formation of link between two authors is independent from the two authors, because it is decided by a third author. For that reason it is totally structural. I use co-citation networks of fields with time dimension, to show how the OR and SD fields have been evolved.

3.2.2.2. Co-Authorship Networks: Co-authorship networks are the other most widely analyzed type of networks in scientometrics literature. In a co-authorship network, nodes are authors and the link between two authors is formed if two authors collaborate (write paper together). Co-authorship networks are undirected and again, due to the analyst's choosing, the network can be constructed either dichotomous or valued. In the literature mostly dichotomous co-citation networks are studied.

3.2.2.3 Inter-citation Networks: Inter-citation networks are not widely analyzed in the literature. In an inter-citation network the nodes might be authors or papers and the link is formed when one cites the other one (paper or author). Inter-citation networks are directed and might be dichotomous or weighted. Inter-citation networks are the hearth of this thesis. The major problem about the inter-citation analysis is the scarcity of the data. Some software packages bring out the co-authorship and co-citation data from the Web of Science database which is the main source of scientometrics data. However, there is no software package available that provides the information of who cites who, and how many times. I use the paper inter-citation data to draw a picture of the fields and I also use the author inter-citation data (which I collect) to analyze the structure more closely.

3.2.2. Software Packages

I use several specific software packages in the process of data collection and social network analysis. Apart from MS Excel, some scientometrics and network analysis softwares are HistCite, CiteSpace II, UCINET and Cytoscape.

Firstly, HistCite has been developed by Eugene Garfield who is the founder of the Institute of Scientific Information and the inventor of the Science Citation Index. HistCite is a very useful tool, to handle massy Web of Science Data. It can categorize data and give historical, descriptive measures such as average citation per paper, average number of references…etc. More important than these descriptive statistics, it can also make it possible to analyze the data in terms of authors, institutions, journals…etc. throughout the historical process. Moreover, HisCite draws histographs, graphical representation of the historical development of a research field, which I have used in the field analysis section. The feature that has helped me a lot, is that HistCite transfers the data into format which other software packages such as Excel, UCINET, CiteSpace II require.

Second software package I use is CiteSpace which is a non-commercial product developed by Chaomei Chen, associate professor in Drexel University. CiteSpace is a very powerful visualization tool, for co-citation and co-authorship networks of scientometrics data. The most important feature of the software is its ability to include historical dimension into the network visualization (Chen, 2006). I have used CiteSpace II heavily in the field analysis section to draw the historical development of OR and SD fields.

Another software package is UCINET (University of California, Irvine Network Tool) which has been basically developed in academia by Steve Borgatti and Roberta Chase and has turned out to be a commercial product which is the most widely used social network analysis software. Basically it does every kind of analysis from descriptive measures to random networks and visualization. I have used UCINET to obtain basic descriptive measures and for data format transformation.

Last but not least Cytoscape is an open-source network visualization tool which has been developed and widely used in biomedical area. The most important features are that it is very user-friendly and the quality of the visualization is far better than the visualization tool of UCINET. I have used Cytoscape heavily in the author inter-citation networks of 80 authors in OR and 50 authors in SD fields.

3.3. Field Analysis

3.3.1. Historical Trends

Table 3.2 shows the general descriptive measures. An article in SDR gets 8.6 citations on average, whereas an article in ORJ and MS gets 23.13 and 34.23 citations on average, respectively. This implies that in the long run, on the average an article in the MS or

ORJ is cited more than that an article in SDR is. This might be due to the different scope of system dynamics and operations research fields.

Table 3.2. Descriptive measures of the journals in different fields

	#Papers	#Cited	#Cited /Paper	#Cited in this Journal	#Cited in this Journal/Paper	#Refer-ence	#Refer-ence/paper	#Author	#Author/ paper
DR	225	1935	8.60	319	1.42	5864	26.06	427	1.90
R	1983	45867	23.13	2372	1.20	44553	22.47	4114	2.07
S	2778	95090	34.23	4222	1.52	78997	28.44	5710	2.06

Another important point is that 1.42 citations out of 8.6 citations are received by the papers in SDR are within SDR, which implies that on average 16% (1.42/8.6) of the citations received by a paper in SDR is actually received in the SDR. On the other hand this ratio is 5.2% (1.2 / 23.13) in ORJ and 4.4 % in MS. This implies that system dynamics field is much more closed and smaller compared to operations research field, because it seems that there are other journals in which papers in ORJ and MS are highly cited, whereas it is not the case for system dynamics field.

Lastly, on the average the number of authors per paper is around 2 in all of the journals, and SDR seems to have slightly lower number of authors per paper. On the other hand average number of authors per paper in the same period is 5.48 in PNAS[8] (Borner, et al., 2004). This number is highly related to the characteristics of the field and journal.

Although the situation of authors per paper differs among different fields, the trend is the same; it increases. In MS the average number of authors per paper was 1.78 in 1981 and it has increased to 2.31 in 2005 which is a 30 % increase, and in ORJ the average number of authors per paper was 1.79 in 1981 and has increased up to 2.38 in 2005, which is a 33% increase. Moreover, in SDR the average author per paper was 1.71 in 1991 (which is the date of the first issue) and has increased up to 2.13 in 2006 (25% increase). On the other hand the increase of average number of authors per paper in PNAS was more

[8] Proceedings of National Academy of Sciences

dramatically and from 3.91 in 1982 to 8.4 in 2001 which is a 115 % increase (Borner, et al., 2004).

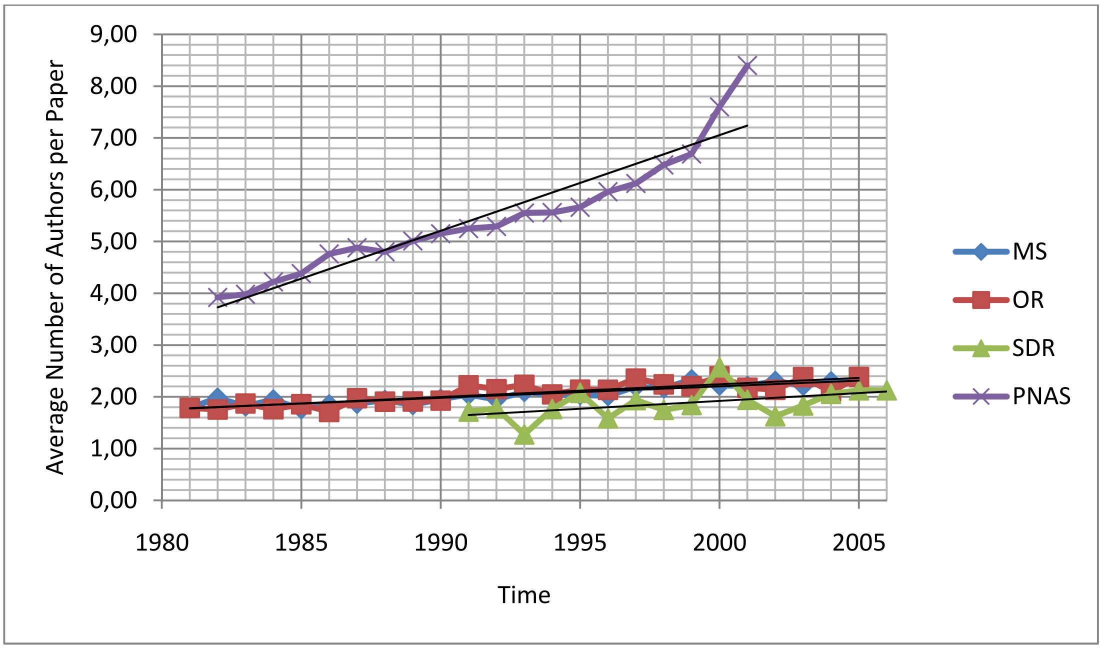

Figure 3.1. Trends in number of Co-authors per paper

The positive trend in all fields is seen in the Figure 3.1 . However the increase in PNAS is much sharper compared to ORJ, MS and SDR.

3.3.2. Development of Literatures

Figures 3.2, 3.3, and 3.4, can be named as the genealogical trees of the SDR, ORJ and MS journals respectively, if a cited paper is identified as an ancestor. In the figures below, the nodes are papers and the links go from citing article to cited article. The size of nodes is proportional to the number of citation that the paper received and nodes are positioned vertically in which older one is at the top.

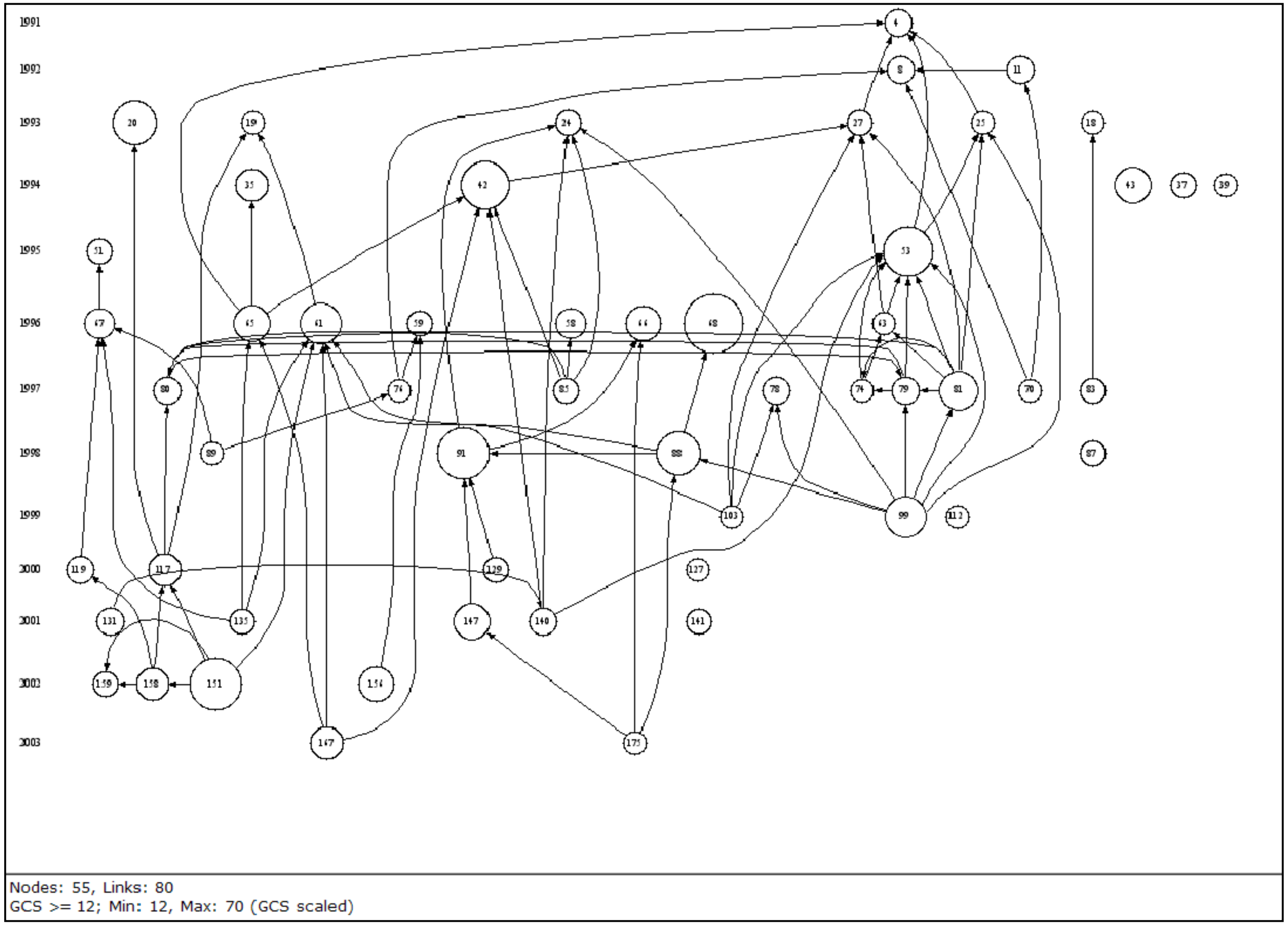

Figure 3.2. SDR Genealogical tree

Figure 3.2 includes the papers in SDR which have been published between 1991 and 2006 and gained more than 12 citations at total. As seen on the figure, there are many citations given to the papers which have been published at the same year or closer and those papers are from the different segments. This implies that system dynamics field has evolved horizontally as well as vertically throughout the years. It is very hard to distinguish different subfields which barely have citations from other subfields.

On the other hand Figures 3.3 and 3.4 show the genealogical maps of ORJ and MS respectively. Compared to system dynamics field, the operations research field has divided into many sub-fields which can be seen horizontally in figures below. The citation ancestor trees reveal those different subfields and there are certain chains of papers which has no relation with other chains. Papers usually cite their previous ancestors in the same subfield. Figures 3.2, 3.3 and 3.4 show the different structures and ways of development of the system dynamics and operations research fields.

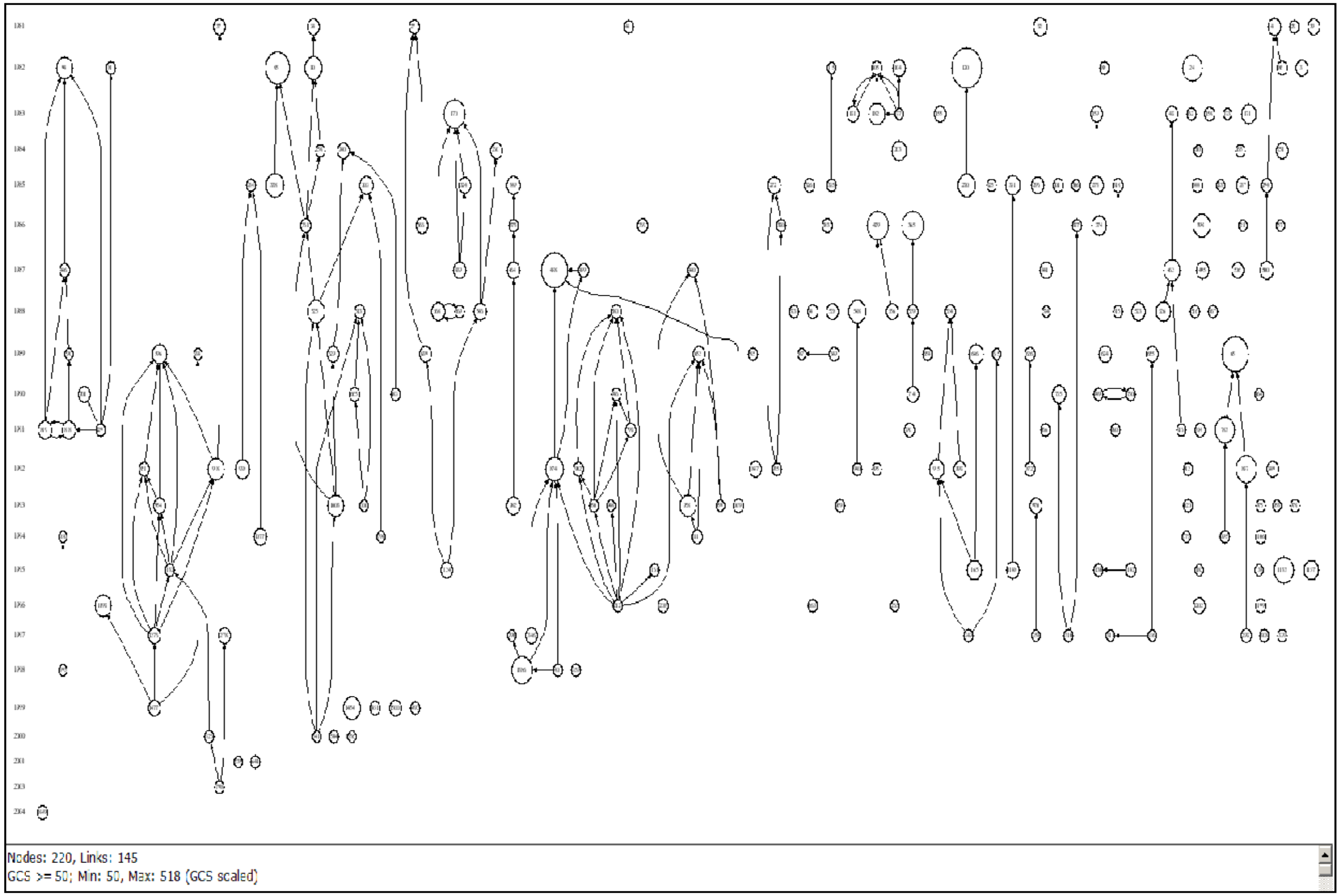

Figure 3.3. ORJ Genealogical tree

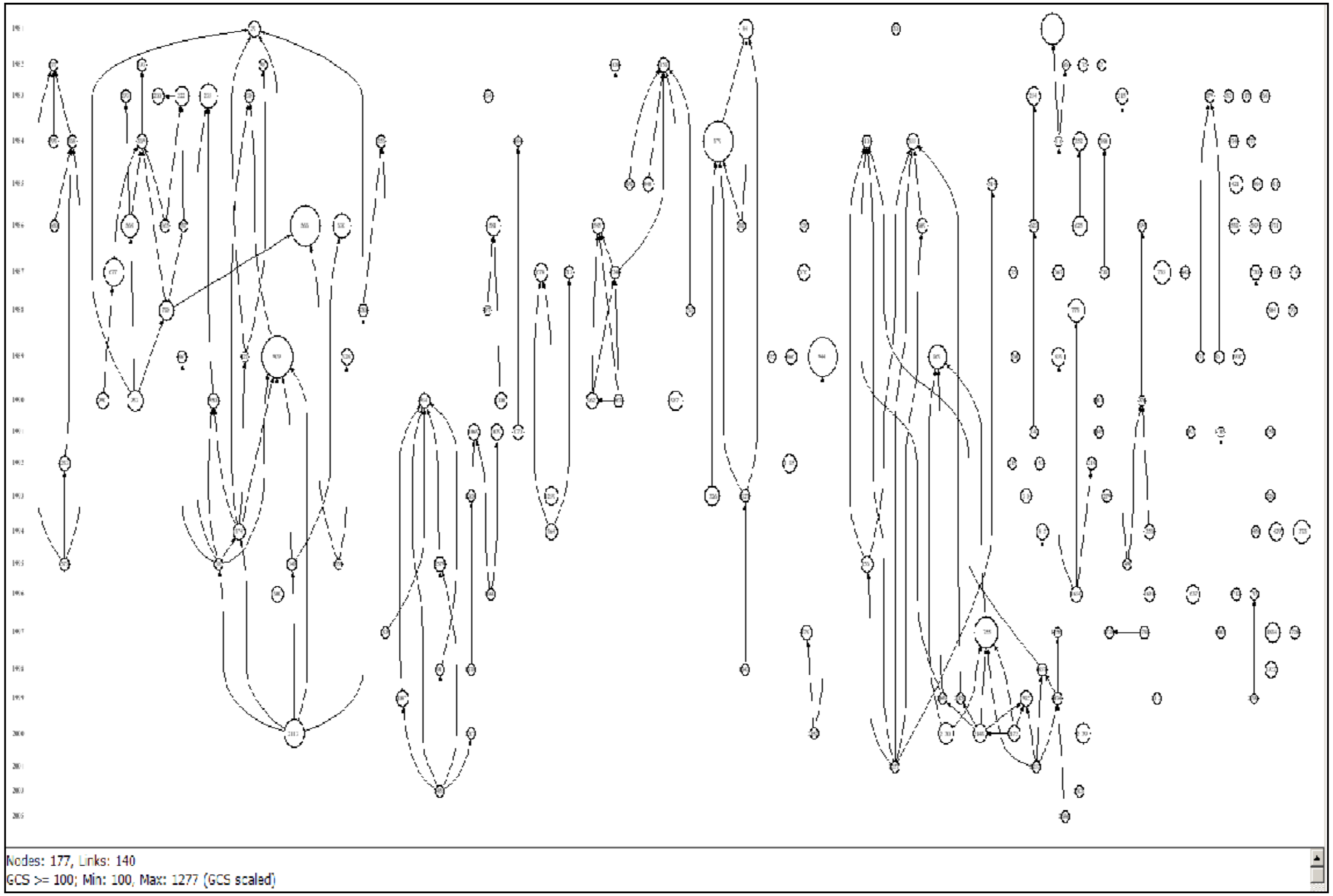

Figure 3.4. MS Genealogical tree

3.3.3. Author Co-citation Maps

As mentioned previously co-citation networks are important in terms of their power of reflecting the structure, because the formation of link between two authors is determined by a third party. The networks on Figures 3.5, 3.6, 3.7 and 3.8 are drawn by CiteSpace II which includes the time dimension into the networks. The spectrum from blue to red reflects the time of co-citedness from older to newer. For example if a link between two authors is green that means that they have been cited together in a late 70's paper. Additionally, the size of nodes is proportional to the number of co-citedness and colors in circles represent the time of co-citedness. Purple ones are nodes which exceed a threshold and are called as turning points of the field.

In Figure 3.5 it is seen that in SDR there are three turning points which can be identified as the author who gets citation with many different authors. These are Forrester, Sterman and Richardson and whereas Forrester and Richardson were prominent in 60's and 70's, Sterman has taken their place in 80's as the turning point of the field.

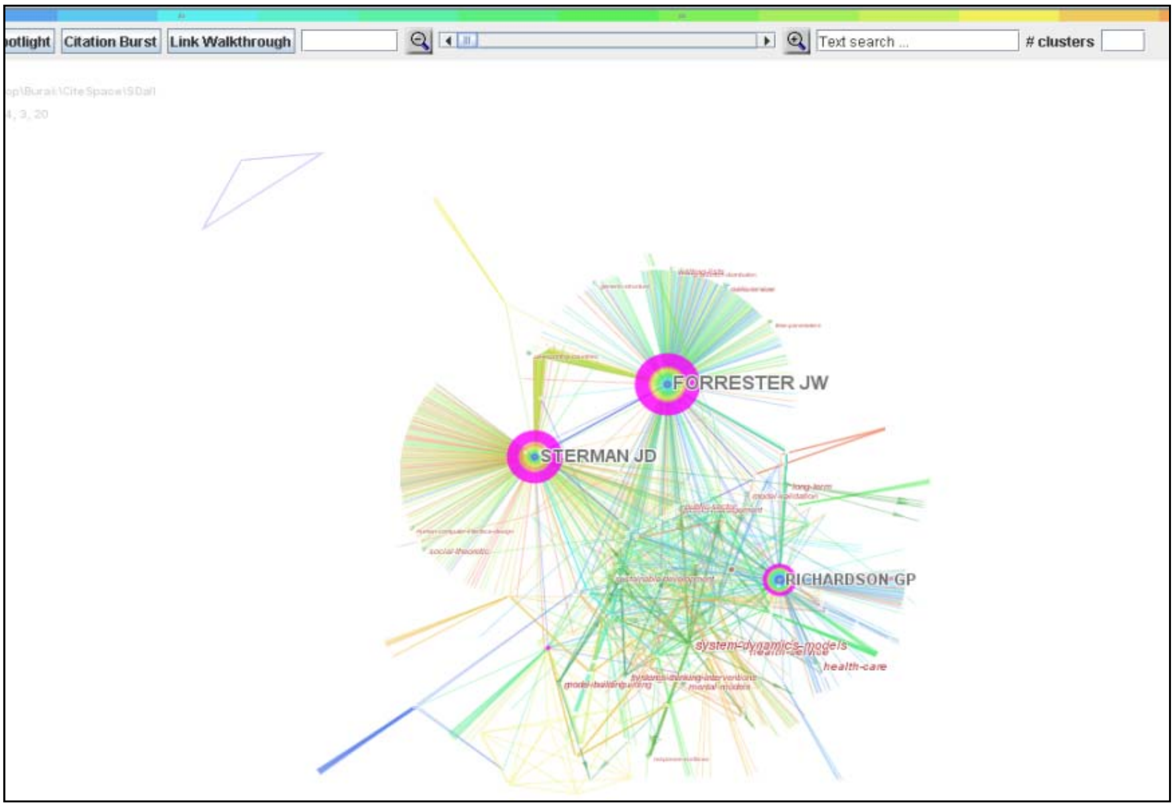

Figure 3.5. SDR Co-Citation network of authors (most salient links)

On the other hand Figure 3.6 shows the author co-citation map of the MS journal (or operations research field). The major difference from system dynamics field is the number of turning points, which implies more diversified field. There are six authors who get cited with many authors, and these are Garey, Federgruen, Wagner, Dantzig, Whitt and Geoffrion. These networks are the big pictures of the fields. It is also possible to analyze individual authors' networks, with whom they have been cited, but since the aim of this chapter is to draw a picture of the fields, individual (ego-centric) networks are beyond the scope of this part.

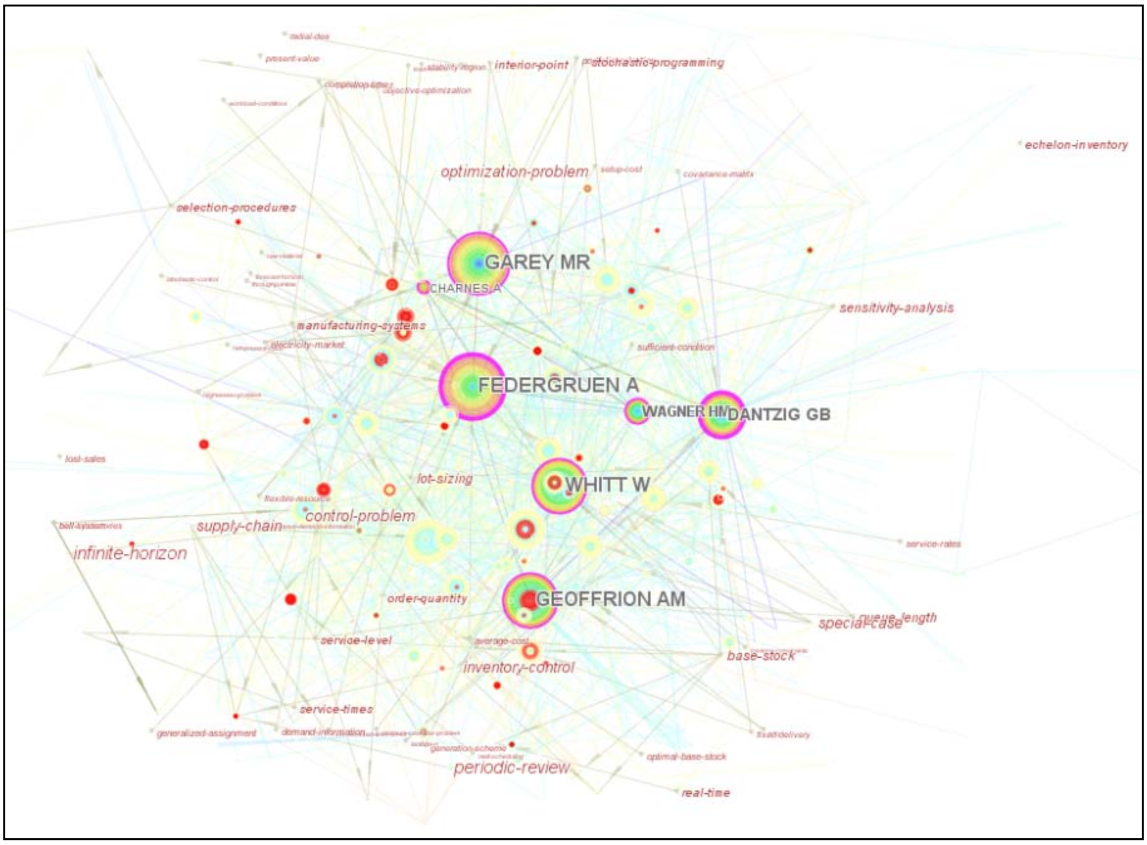

Figure 3.6. MS Co-Citation network of authors (most salient links) (1981-2005)

In order to grasp which other fields have contributed in the development of system dynamics and operations research, I have taken the 30 most productive[9] authors from each field and drawn the co-citation maps of those authors in a wider historical range from 1950 to 2005.

[9] Productivity is measured with the number of papers published.

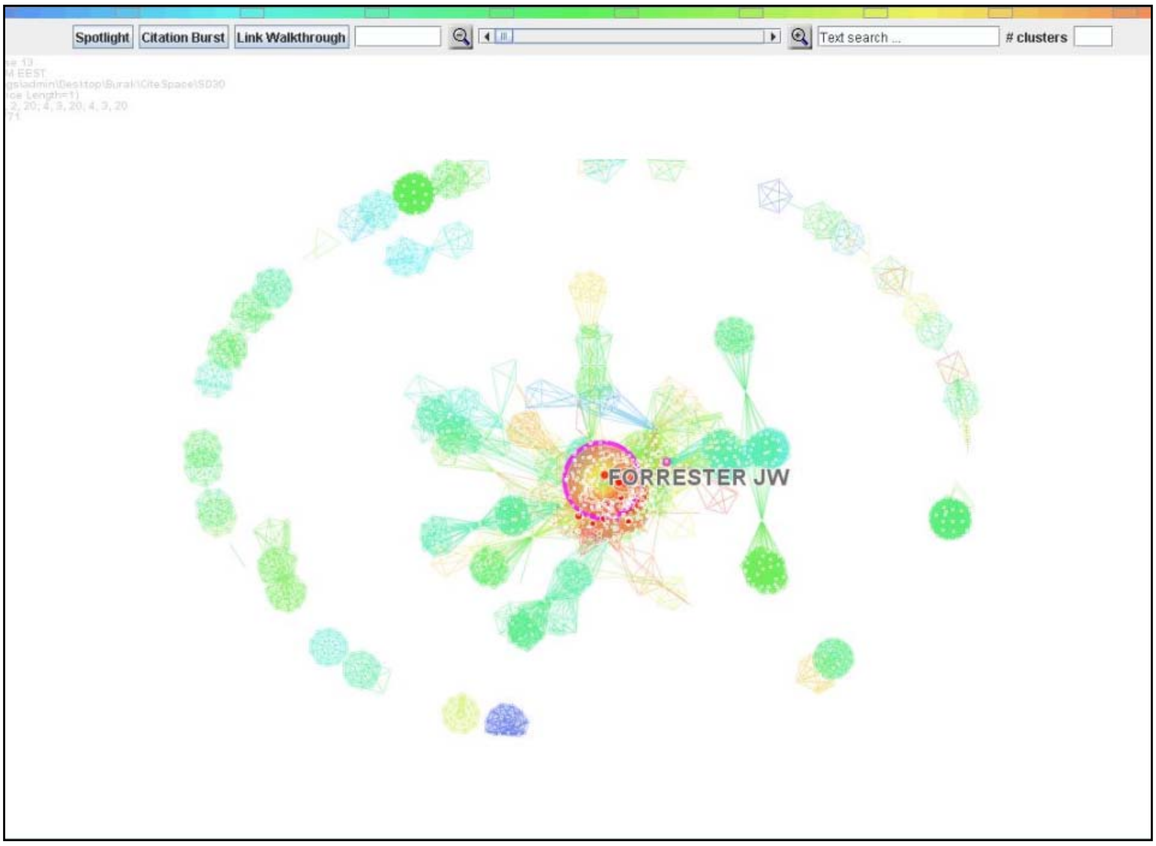

Figure 3.7. SD Top 30 author co-citation network (1950-2005)

In Figure 3.7 the green clusters show that late 70's were the era that system dynamics literature has been developed most rapidly and took the advantage or some other fields. For example some of the green (which implies late 70's) clusters are papers from the American state reports on sustainability and development. Those relations correspond to the expansion of the field with studies of Forrester, Meadows and others. On the other hand close to the center there are some yellow-orange (mid 90's) clusters which implies that there are some expansion in the field in those years. Lastly, it can be said that since there are not that much reddish clusters, the system dynamics field does not have a substantial dynamism in recent years.

On the other hand Figure 3.8 shows the co-citation network of top 30 authors in operations research field. In order to have a clearer picture I put the nodes in a temporal order from left to right. First distinction from the system dynamics field is that operations research field almost has the same level of productivity in all decades. In other words in Figure 3.8 the transition from blue to green and from green to red is very smooth and there is not any interruption. Moreover there are red dots which imply recently reputed authors as well as green dominated nodes which imply previously reputed scholars. In short operations research field is an ongoing field and the level of productivity seems more or less the same in all decades, whereas the field of system dynamics is kind of very popular in 70's, 80's and 90's, and doesn't have highly reputed junior academicians in the last decade that much. This can be related to the size of the field, because operations research consists of several sub-fields and even if some of them obsoletes, some other sub-fields might take the place of previous ones. However, system dynamics is very compact and doesn't have further division as operations research field has. For that reason decrease in the expansion of system dynamics field is observable on the Figure 3.8

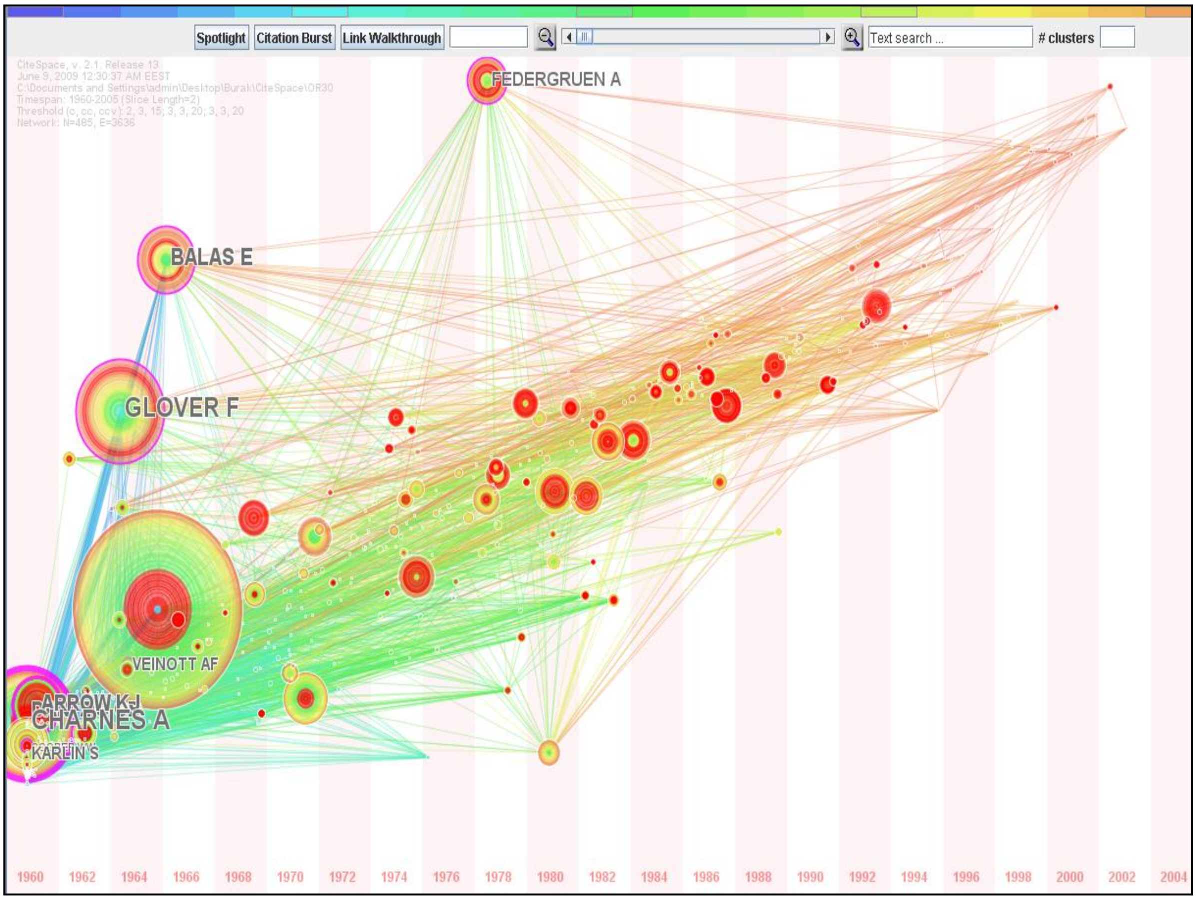

Figure 3.8. OR Top30 authors co-citation network 1960-2005

3.3.4. Co-Authorship Network

Co-authorship networks are collaboration networks in which institutional or regional cliques are very obvious. Figure 3.9 shows the author collaboration network in SDR. Unfortunately, the data management tool of the CiteSpace II couldn't convert the OR data due to its size. I have only the co-authorship analysis of system dynamics fields. However I expect more or less the same kind of patterns in operations research field, because the average number of author per paper is almost the same in both of the fields.

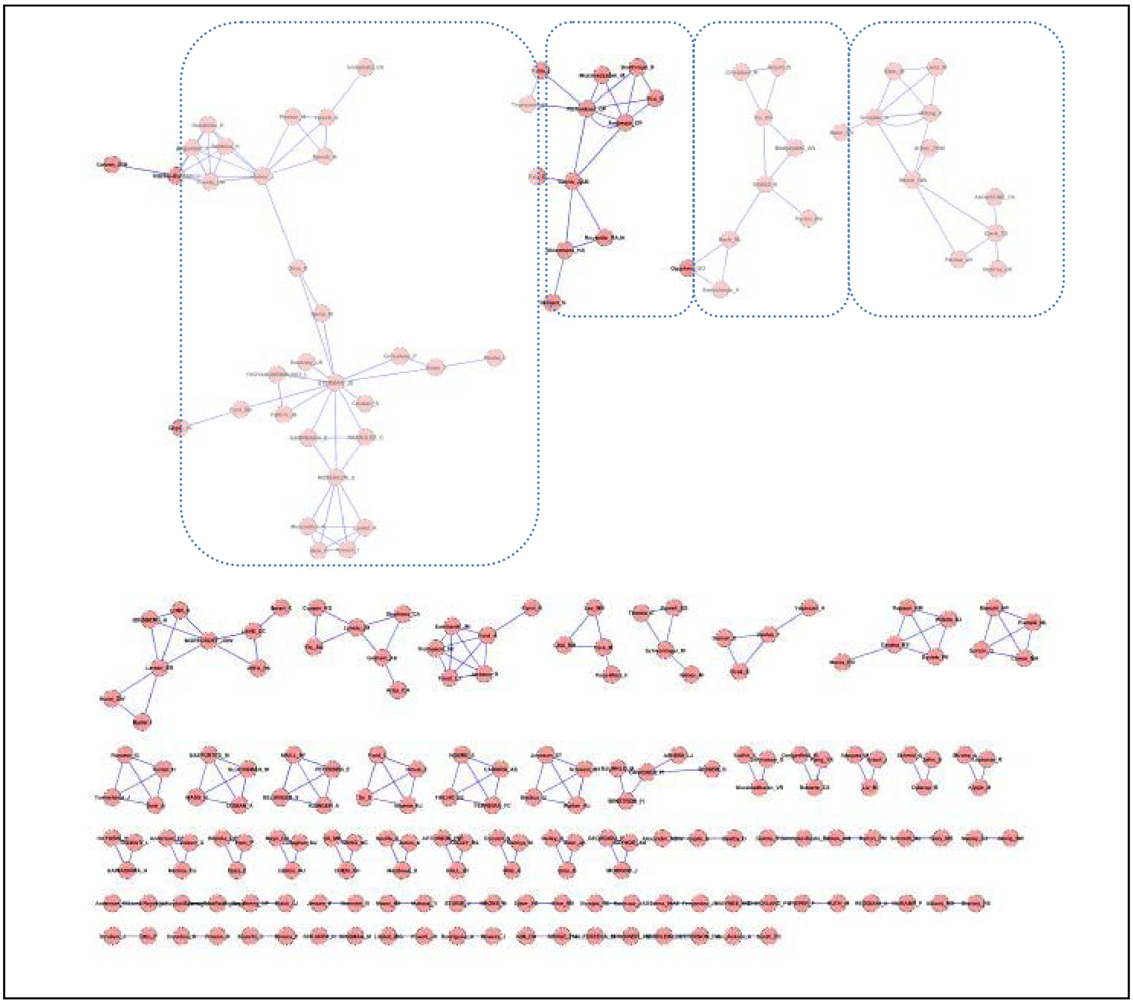

Figure 3.9. SDR Author co-authorship network

In Figure 3.9 there are seven different collaboration groups and several small cliques in which there are 2 to 3 people. The major four cliques are shown in detail in Figures 3.10 and 3.11. Firstly, in Figure 3.10a MIT people are seen around John Sterman which is the node at the center. On the other hand, Figure 3.10b shows the collaboration network of the people from SUNY, Albany which is one of the most important places having system dynamics research group. Moreover in Figure 3.11a there is the network around Khalid Saeed, which can be identified as the network of WPI and people related to institutional economics. Lastly Figure 3.11b focuses on the part which can be identified as the European clique with scholars such as Grossler, Milling and Winch.

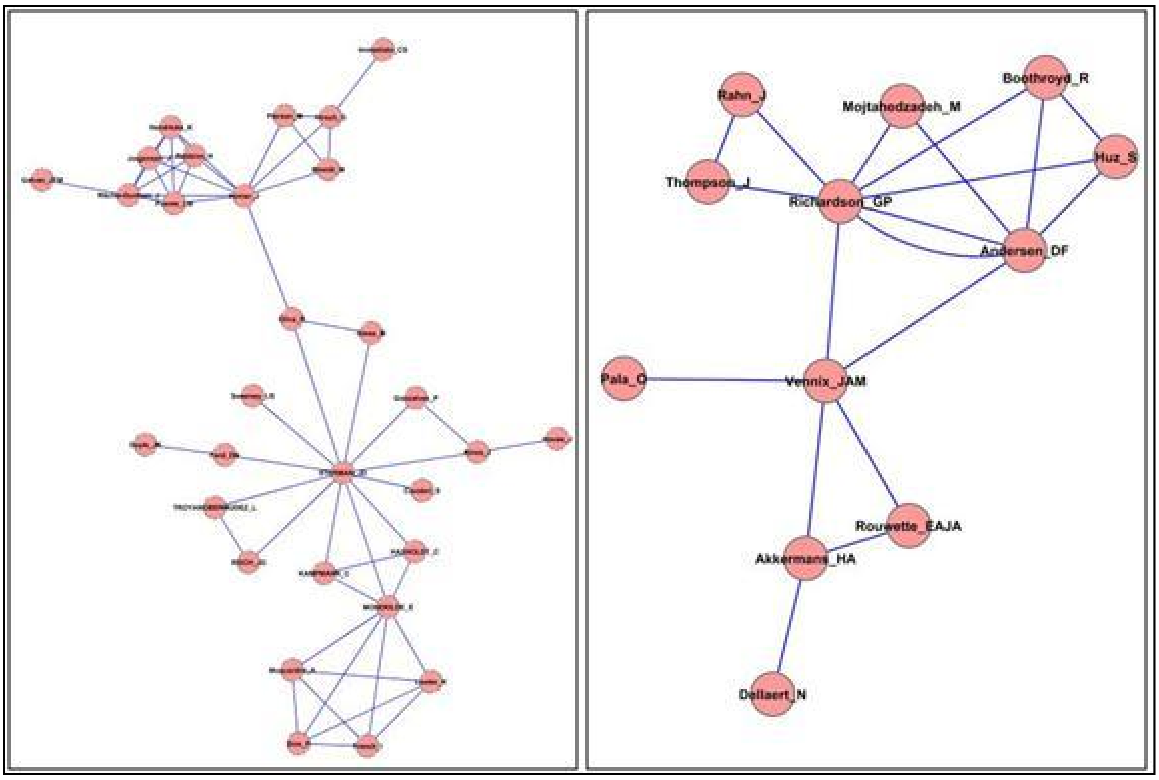

Figure 3.10 a. & b. SDR Co-authorship network cliques 1 and 2

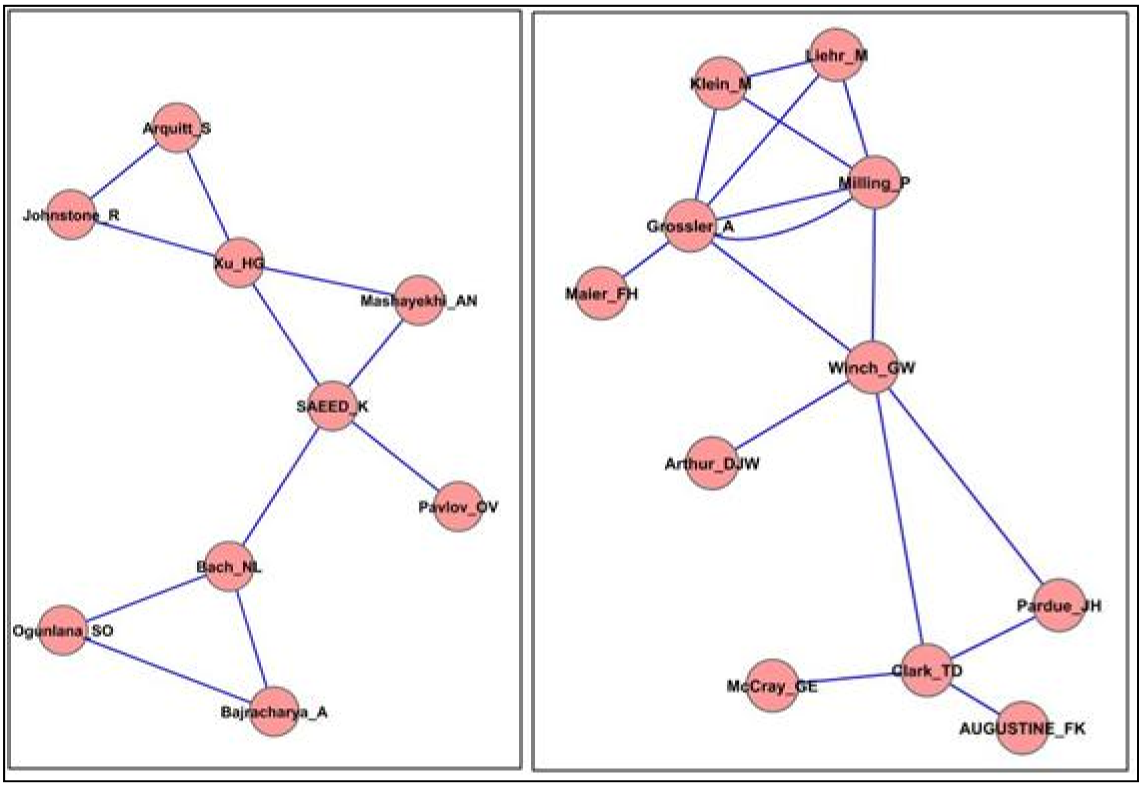

Figure 3.11 a. & b. SDR Co-authorship network, cliques 3 and 4

3.4. Author Inter-Citation Networks

In the previous section the collaboration network of the system dynamics field has been analyzed and some regional and institutional patterns have been shown. Having these kinds of patterns might be seen as natural, because usually people collaborate with people whom they know and regional and institutional affinity is a very legitimate rationale for collaboration in this sense. The crucial question is whether or not inter-citation networks are look like those of collaboration. Since the dynamics of finding someone to write a paper and finding someone to cite are totally different, one doesn't expect to have at least the effect of regional or institutional affinity on inter-citation networks, in an ideal world.

In terms of methodology, as mentioned before, the major problem is the scarcity of the inter-citation data. I have collected all the data from Web of Science. Since it is impossible to collect individual database of all authors, I have operationalized a two-phase analysis. At the first phase I have collected the data of 30 most productive scholars of the fields and analyze their inter-citation networks. Then at the second phase I have collected the individual database of 50 more scholars, who are chosen randomly, in OR field; and those of 20 more scholars, who are chose randomly, in system dynamics field and try to represent the whole network by these selected sample networks.

3.4.1. Top 30 Authors

The authors are ranked in terms of number of papers they have published in ORJ and MS for the operations research field and in SDR for the system dynamics field. And then first 30 authors have been chosen as the most productive authors. Then I have downloaded the raw data from the Web of Science database one by one for each author. The most important point is that the individual data consist of all articles published by authors in any journal not just in MS, ORJ or SDR. In order to obtain weighted network structures, these raw data files have been processed by using first HistCite and then an Excel Macro code. The networks are visualized by using Cytoscape. The size of the nodes is proportional to the number of citations won, and the thickness of the links is proportional to the number of citations from one scholar to other.

Before going on the network structures in different fields, a short note about the different lay-outs is important. There are two basic layout algorithms that I use. First one, seen in Figure 3.12, is spring embedded layout which is identified as the best to see the sub-grouping structures. The spring algorithm basically defines the nodes as similar charged particles and tries to minimize the total potential energy of the network. By this way spring embedded algorithm makes cliques very visible. On the other hand the layout in the Figure 3.13 is the circular layout which puts the highly connected nodes in a circular way and puts subgroups separately. Figures 3.12 and 3.13 actually carry the same information and for this reason I illustrate both algorithms only in first data set and confine with spring embedded layout in following sets.

3.4.1.1. System Dynamics Field: Firstly, in Figures 3.12, 3.13 and 3.14 the inter-citation network of the top 30 authors in system dynamics field has been shown. Obviously (compared to author-collaboration networks) there is not any clear distinct sub-groupings. More or less the structure is one compact network. However, to check whether there are any patterns due to institutional or regional affinity, I have colored the nodes.

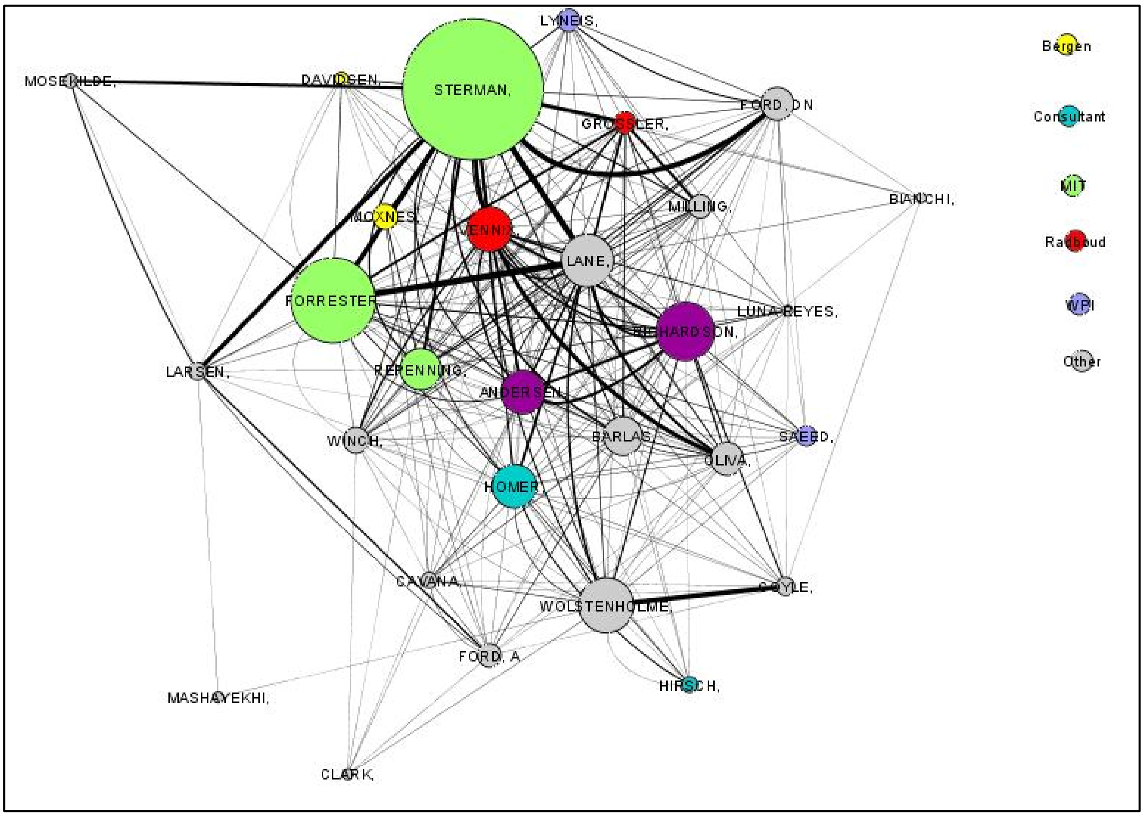

Figure 3.12. SD30 Inter-citation – affiliation – in spring embedded layout

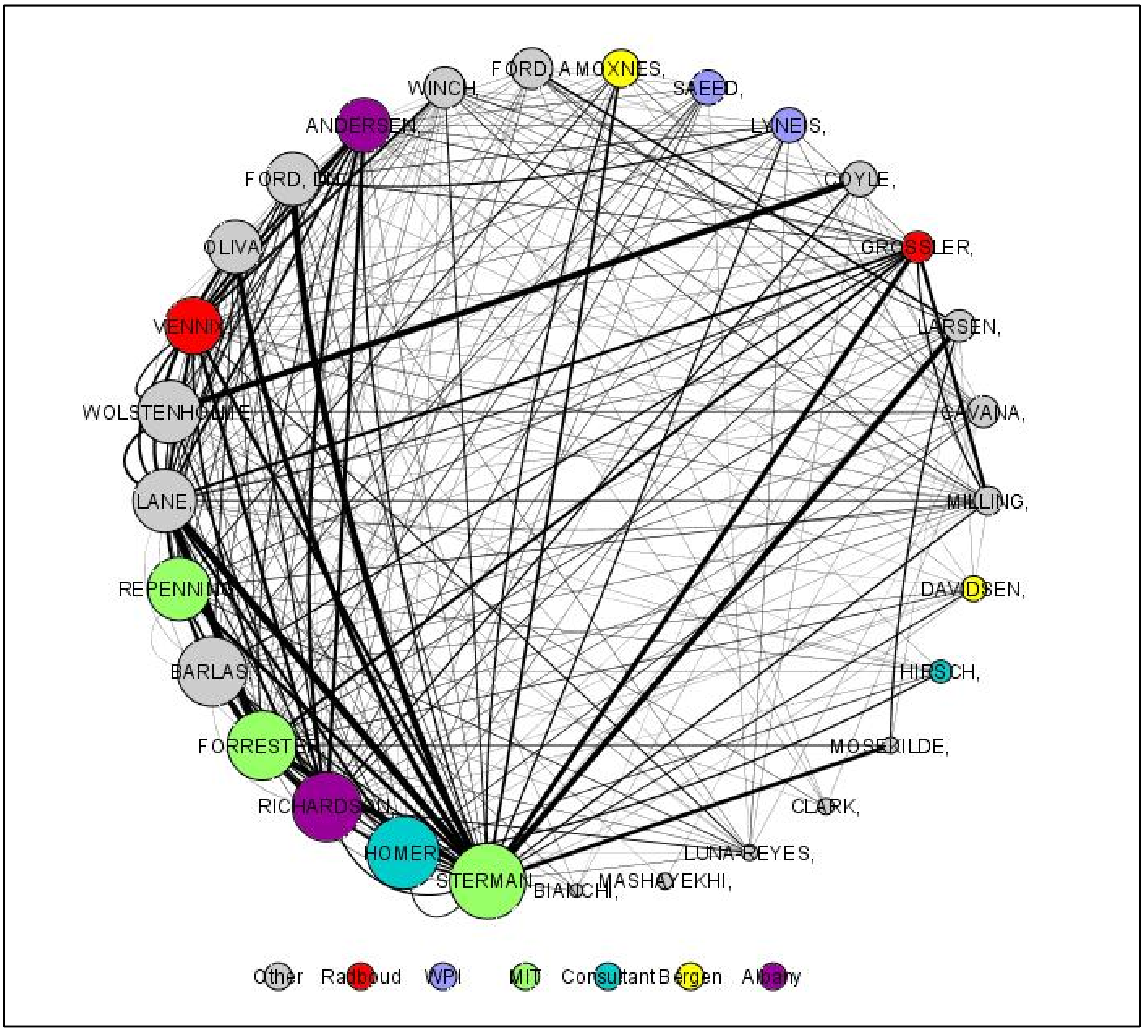

Figure 3.13. SD30 Inter-citation – affiliation – in circular layout

In Figures 3.12 and 3.13 the nodes are colored due to the institutions where scholars are appointed. The institutions with one scholar all have been categorized as ‘other’ and colored with grey. MIT, Albany and WPI and Bergen are the leading institutions, but there is not any obvious pattern of effect of institutional affinity on inter-citation structure. Even if there is an effect, this database and network pictures doesn’t catch at all.

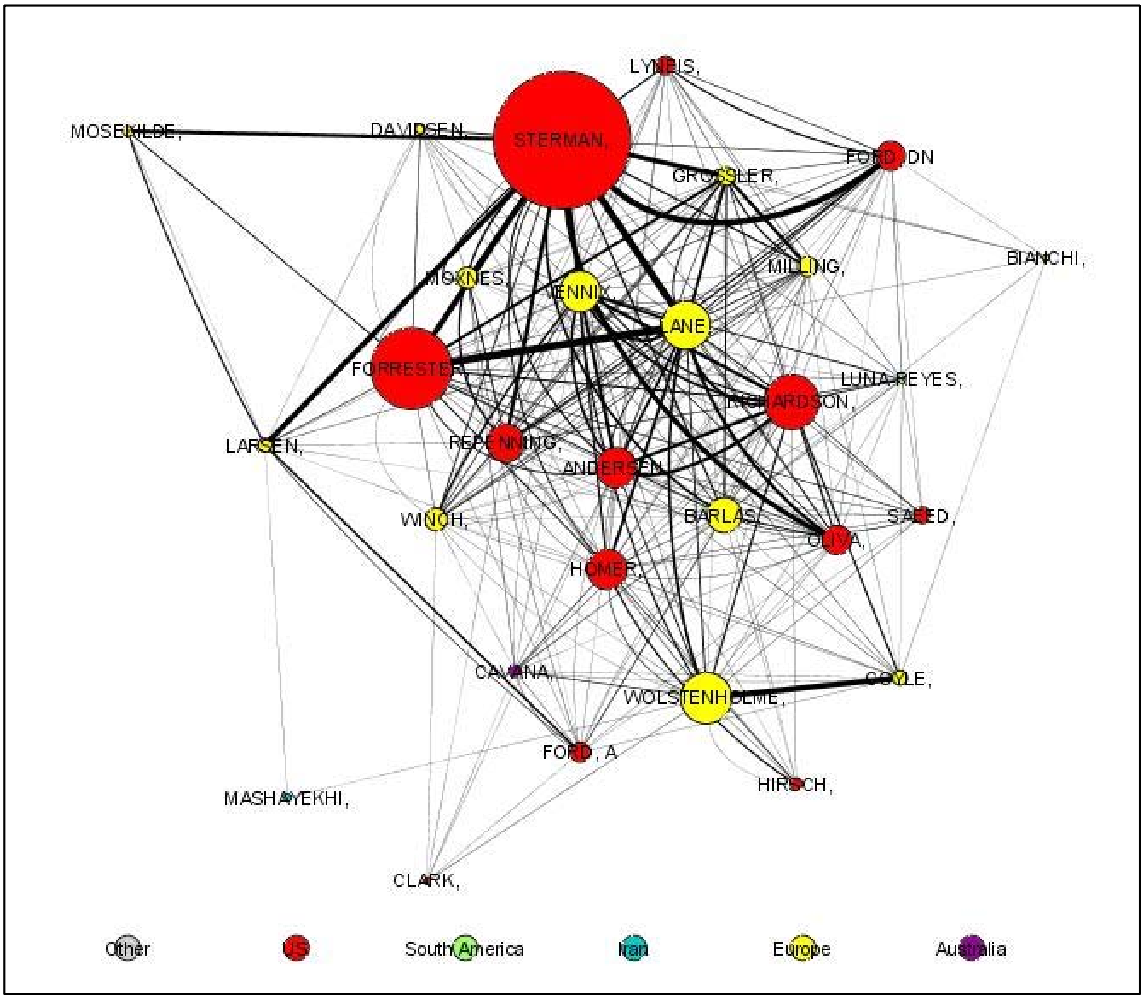

Figure 3.14. SD30 Inter-citation – region

Lastly in this section, I have colored the nodes due to the geographical regions. As seen in Figure 3.14, although the scholars in the US have larger nodes (highly cited), in terms of number of scholars there is a balanced view between Europeans and Americans. Again, from this picture it is very hard to argue that being in America has an effect on getting more citation. This dataset and network visualizations didn't catch any effect of regional affinity on citation network structure. This does not mean that there is no effect on individual behavior, but it is fair to say that in aggregate level I fail to see this kind of a pattern.

3.4.1.2. Operations Research Field: Inter-citation networks among the top 30 scholars of the operations research field have been shown in Figures 3.15, 3.16 and 3.17. Different sub-groups are seen in those figures. However, these subgroups seem to be related to the different sub-fields of the operations research. Before looking at the different subgroups, firstly in Figure 3.15, I have looked at the effect of institutional affiliation and the picture doesn't propose any pattern of institutional effect on the inter-citation networks at the aggregate level.

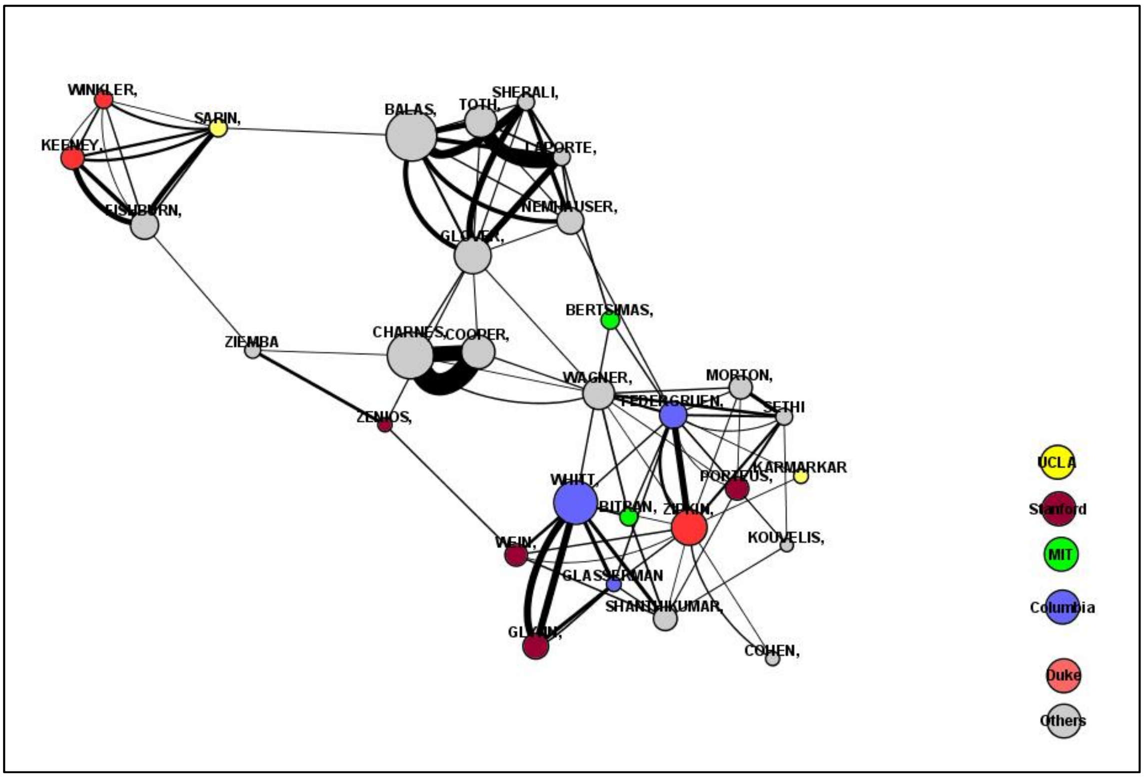

Figure 3.15. OR30 - Intercalation - affiliation

Moreover in the Figure 3.16, I have colored the scholars in the same color if they have PhD from the same department. Although there are some universities being more prominent, the distribution of the colors is more or less balanced throughout the network. Lastly in Figure 3.17, I have colored the nodes whether they are in Europe, America or Canada and in this picture again there is not any regional effect seen on the aggregate network structure.

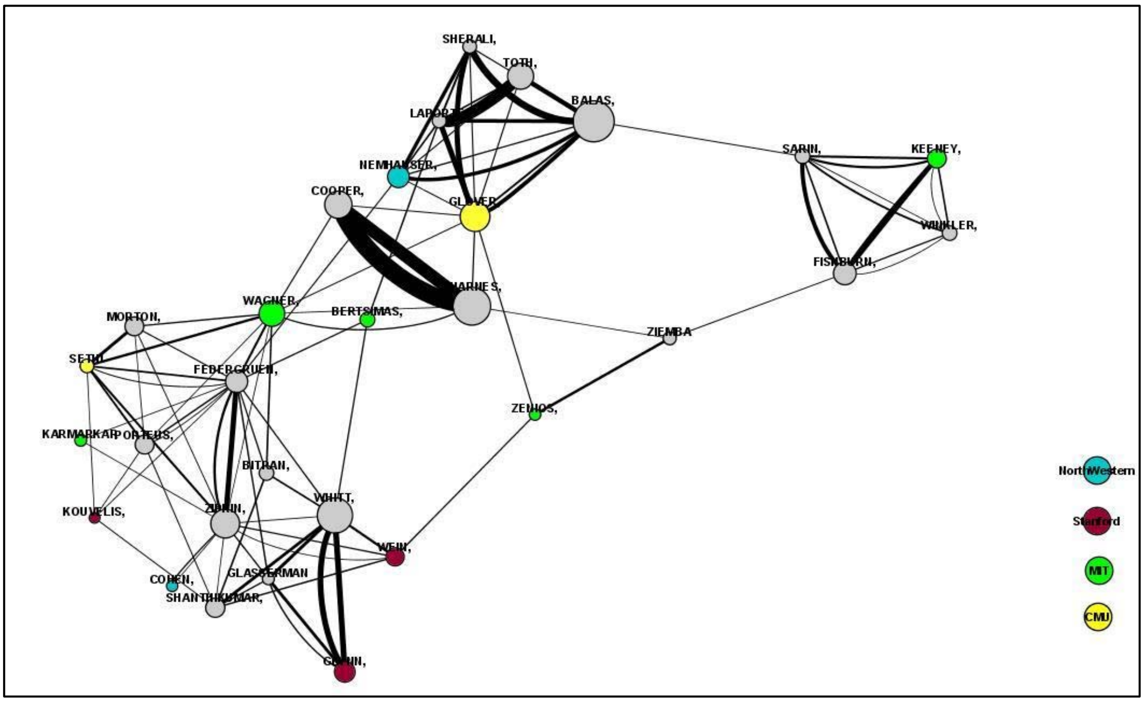

Figure 3.16. OR30 Inter-citation - PhD

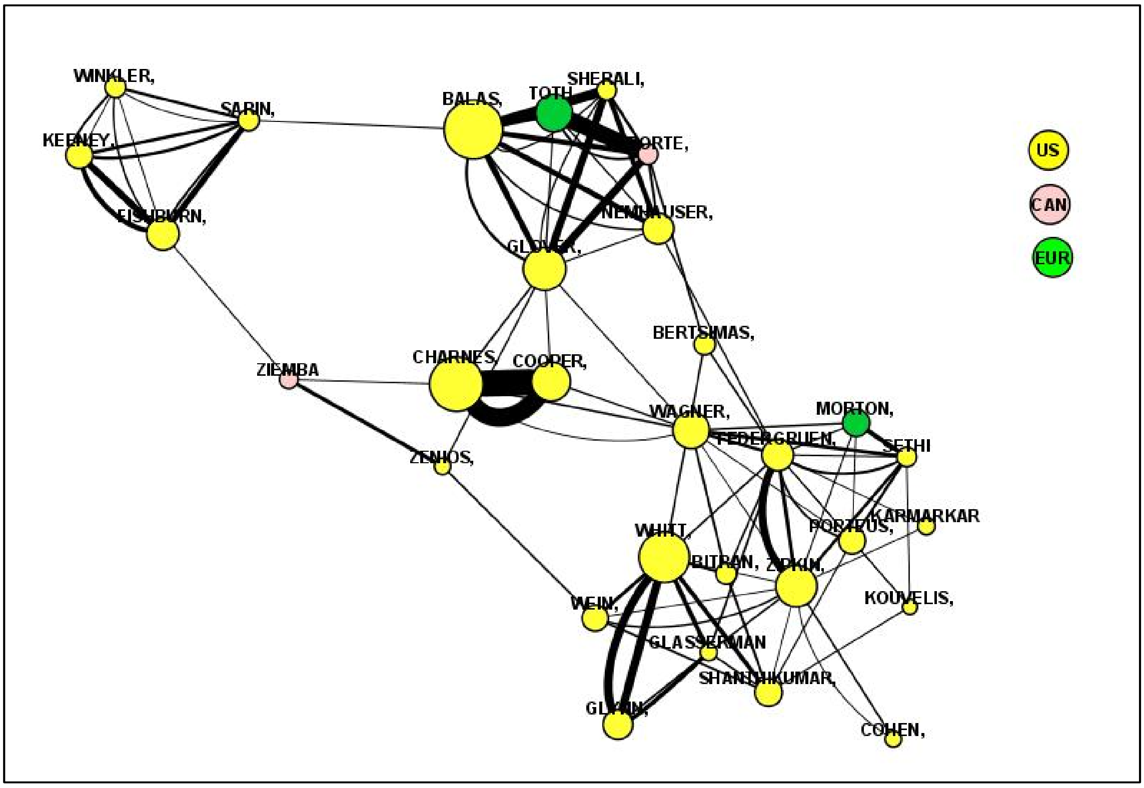

Figure 3.17. OR30 Inter citation - region

One important point related to the Figure 3.17 is that there is domination in the field by the scholars in the US. The case is more balanced in the system dynamics field, but in the operations research it is not. There might be two reasons for this outcome: one is the selection of authors. Since I have chosen the most productive scholars in MS and OR, one may argue that those journals are highly US oriented and not including European Journal of Operations Research (EJOR) might have caused a biased result. I cannot accept or falsify this argument due to lack of data. The second explanation might be that there is a real domination of scholars in the US. This is an open ended question for now.

Lastly, I have coded the authors in terms of their subfields. By looking OR related associations and talking to Ph.D. students in OR field, I have divided the operations research field into ten subfields, namely, optimization, computing, stochastic methods, manufacturing-supply change operations, production, decision and risk analysis, financial engineering, policy modeling and public sector, information theory-management-marketing and telecommunication networks. Then I have coded the authors as 1 one in each field, if they have published at least an article in that area. Although the assignment of authors to the subfields sounds like subjective, I paid a great effort to be objective as much as possible. I have checked the final assignments with three different PhD students in OR field, independently. After the assignment process I have formed the sub-field sharing network (Figure 3.18) in which nodes are authors and a link is formed between A and B if both have at least one common sub-field, and the thickness of the links is proportional to the number of shared sub-fields. The more the number of shared sub-fields, the thicker the link is.

The sub-field sharing network has more or less similar sub-groupings of inter-citation networks in a certain extent. And this similarity feeds the idea that inter-citation behavior reflects the subfield affinity in the aggregate level. This is important that in the aggregate level the data I used and networks I have drawn don't show any pattern contrasting this idea.

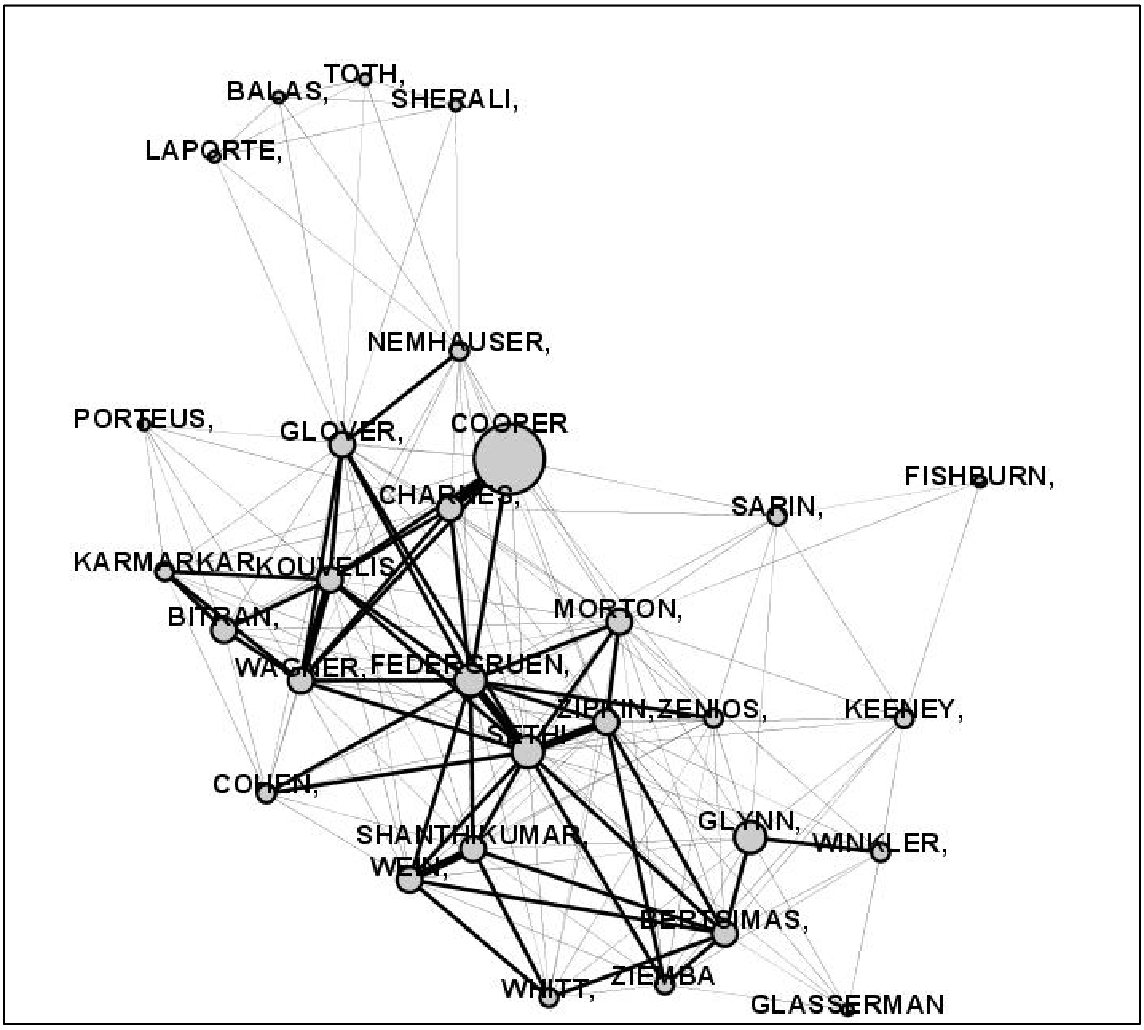

Figure 3.18. OR30 Field sharing network

3.4.2. Sample Authors Set

The previous section is about all the top authors. However the network analysis requires the examination of the entire network. Even if it is impossible to collect the inter-citation data of all the authors, I have constructed samples which consist of top authors and randomly chosen scholars. The network is analyzed in the same way as in the previous section.

3.4.2.1 System Dynamics Field: In addition to top 30 authors in the previous section, I have collected the individual data of randomly selected 20 more scholars who has published at least one paper in SDR. Then I used the same data processing techniques that I have used in the previous section to the sample of 50 authors, which resulted in the following network in Figure 3.19. (Links with a value of less than 3 citations haven't been shown in the network visualizations)

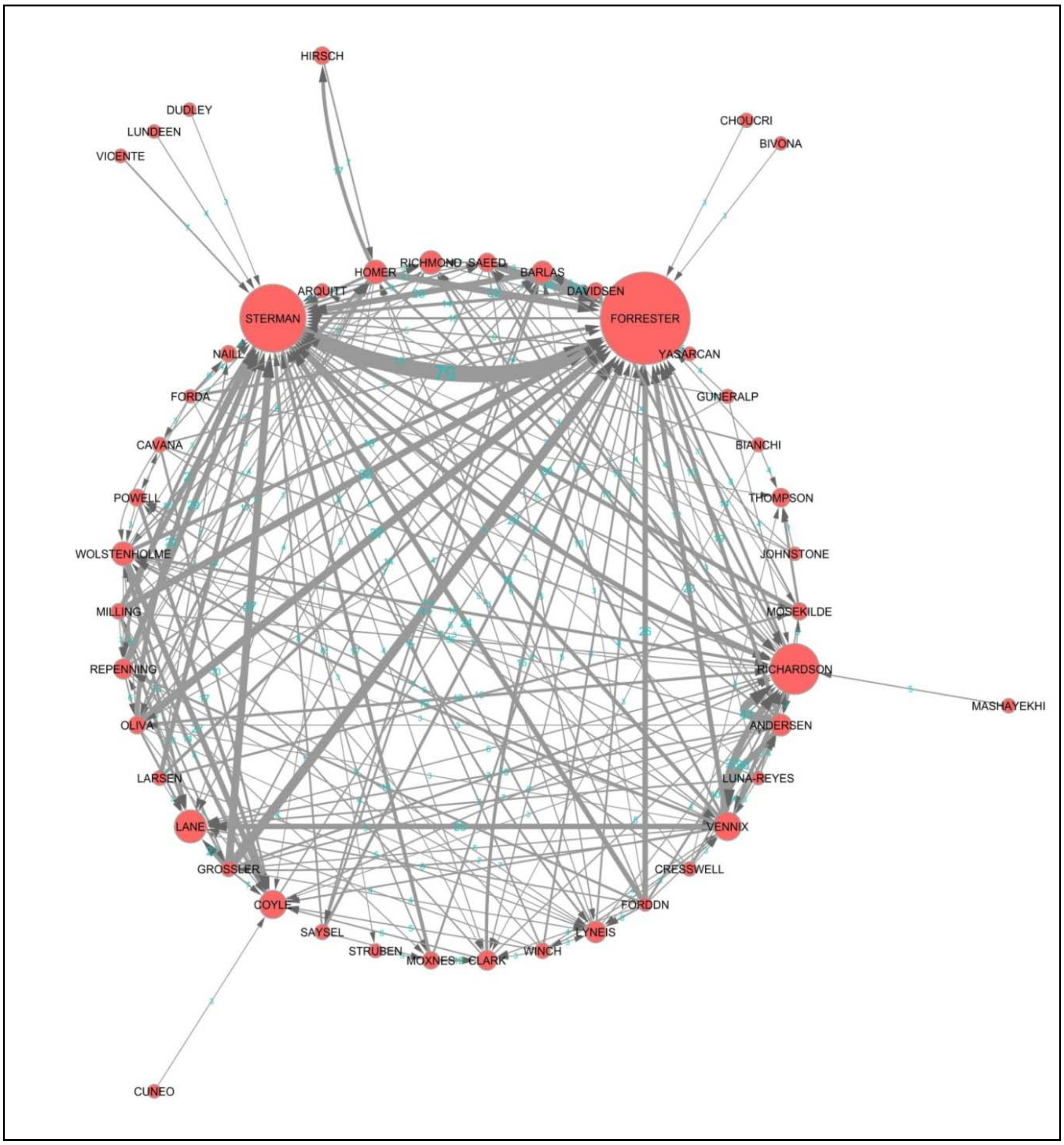

Figure 3.19. SD Inter citation network of 50 authors

Inclusion of 20 random scholars has changed the network structure a little bit. Firstly the size of the Forrester has increased which implies that Forrester is cited more among the authors who are not at the top 30. Secondly some satellite authors who has very weak connections to the system dynamics field have been emerged. Lastly the pattern between Homer and Hirsch is quite interesting. Although both are very famous people in the field, Homer cited Hirsch for 17 times and Hirsch cited Homer for 7 times. This imbalance is one of the first things attracting attention at the first sight.

3.4.2.2. Operations Research Field: As for randomly chosen scholars to the data of 30 top authors. Since the operations research field is bigger and more complex than the system dynamics field, I have collected the data of 50 randomly chosen authors. At the end of the data processing procedures I have obtained the following network structure of 80 authors in figure 3.20. (Links with a value less than 4 citations haven't been shown in the network visualization.)

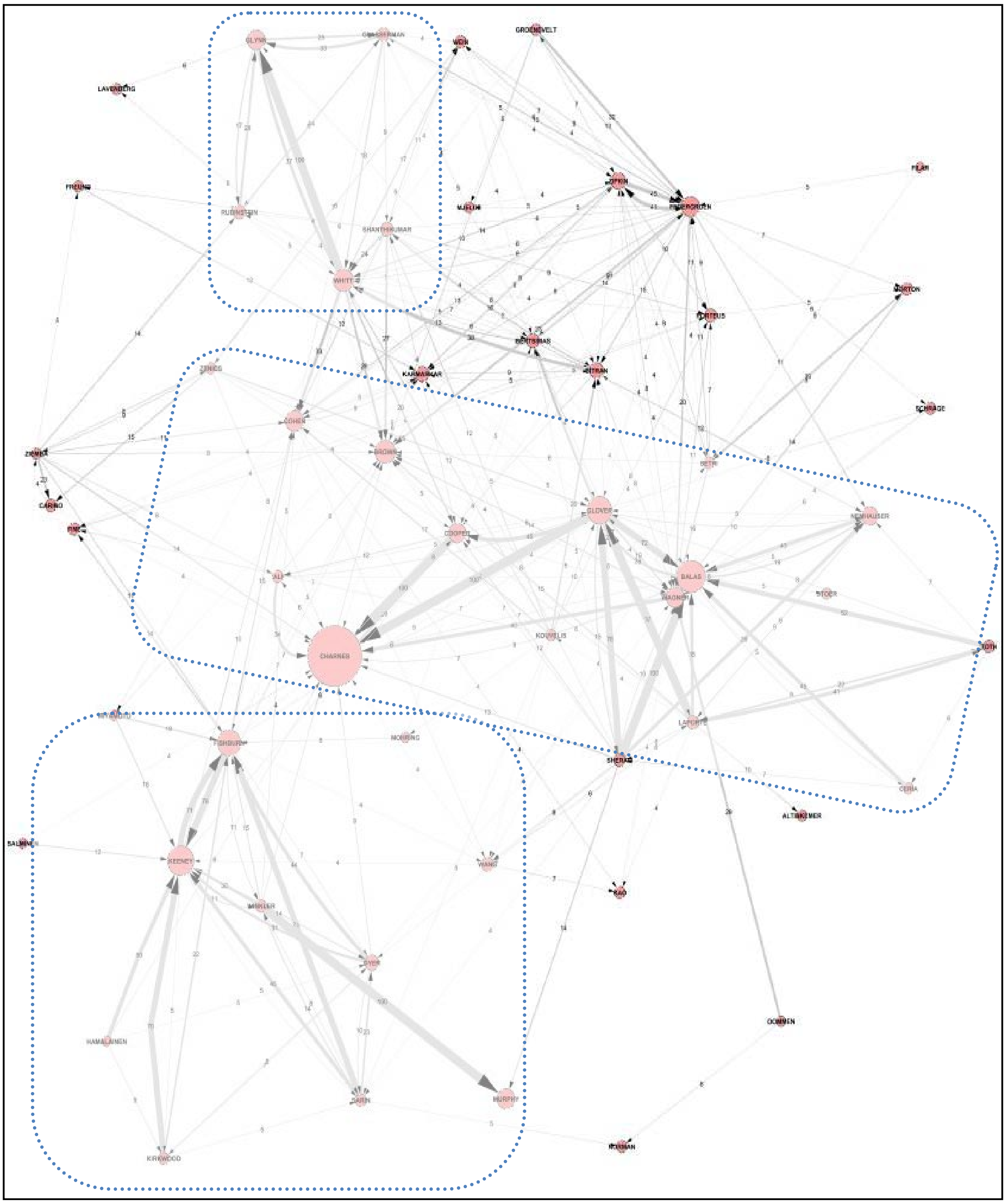

Figure 3.20. OR Inter-citation network of 80 authors

In this network structure, the big picture is similar to the structures in Figures3.15, 3.16 and 3.17 in terms of sub-groupings. Different segments in the field are obvious and these sub-structures are shown in Figures 3.21, 3.22 and 3.23. Of course these sub-groups do not separate from each other at all, but the links between groups are very weak.

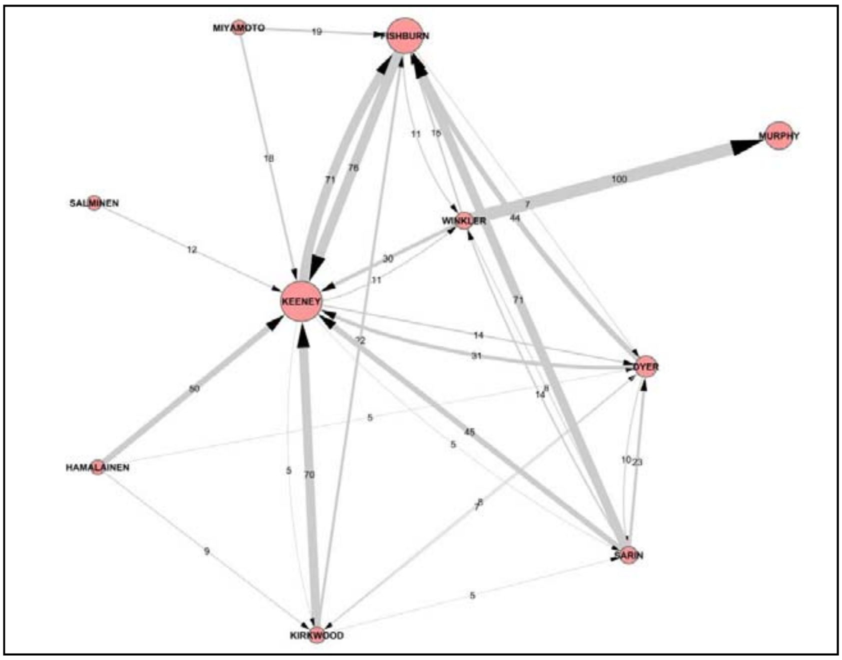

Figure 3.21. OR Inter-citation network - subgroup 1

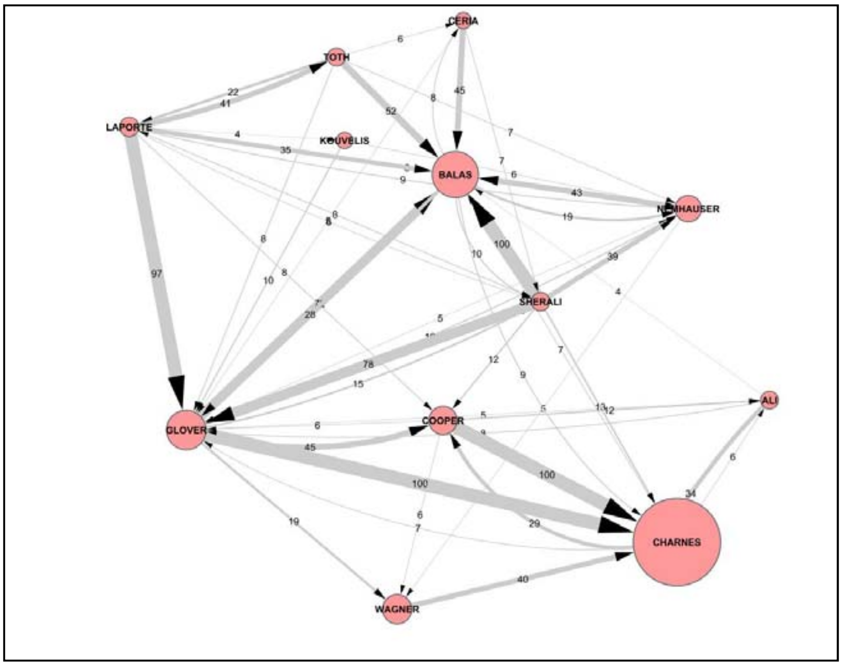

Figure 3.22. OR Inter-citation network - subgroup 2

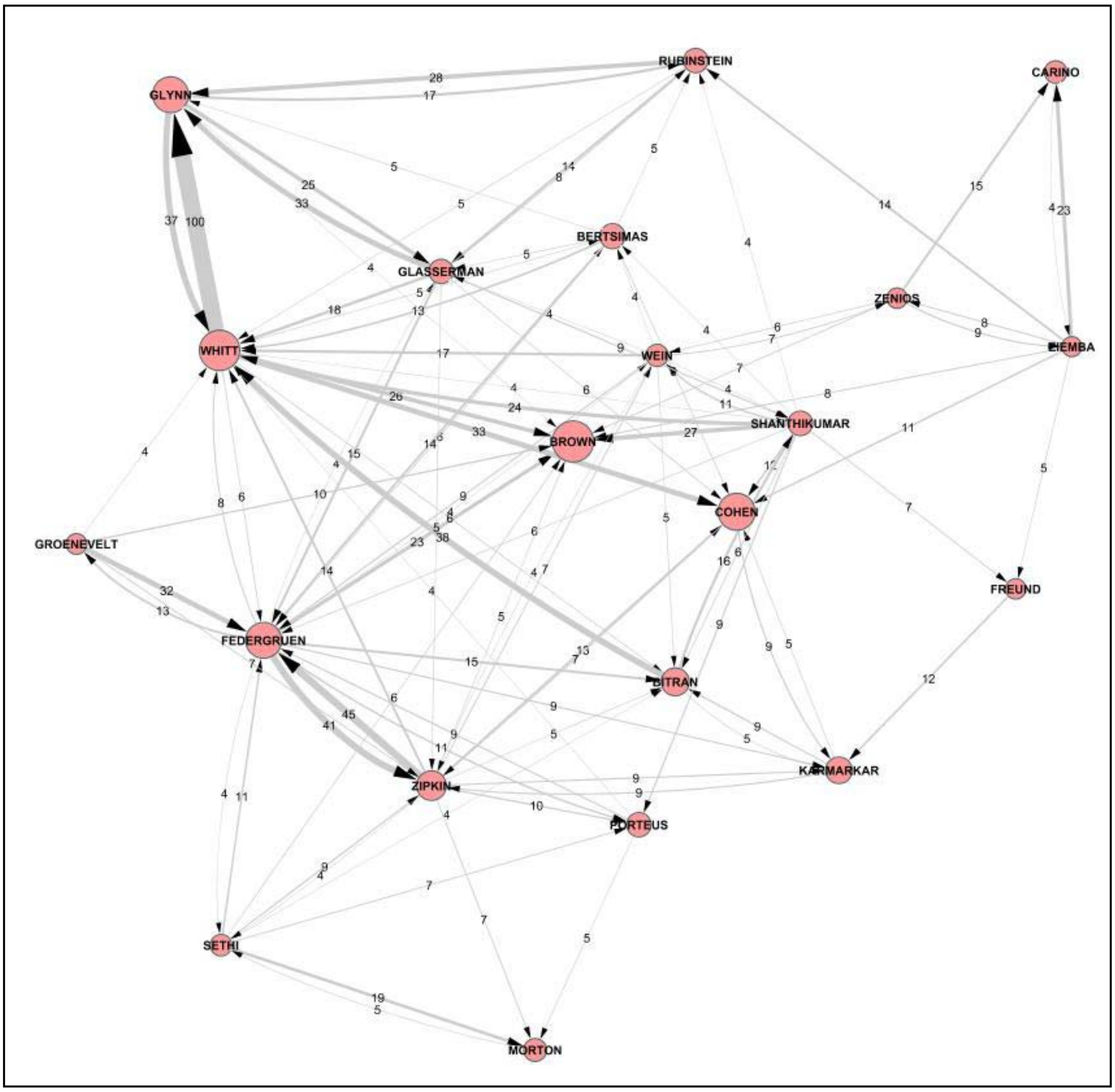

Figure 3.23. OR Inter-citation network - subgroup 3

In Figures 3.21, 3.22 and 3.23 the sub-groups are shown and some balanced and un-balanced relations are remarkable. For example number of citations between Keeney – Fishburn (71 vs. 76) and Winkler – Fishburn (11 vs. 15) and Kirkwood – Dyer (7 vs. 8) are quite balanced. On the other hand there certain number of pairs between which the number of citations are very unusual. Some of them are Keeney – Kirkwood (5 vs. 70) and Keeney – Sarin (5 vs. 45). Table 3 shows the list of some eye catching balanced and unbalanced citation relations.

Table 3.3. Eye catching balanced and unbalanced citation relations in figures 3.21, 3.22 and 3.23

A	B	A to B	B to A
Keeney	Kirkwood	5	70
Keeney	Sarin	5	45
Keeney	Dyer	14	31
Keeney	Winkler	11	30
Keeney	Fishburn	71	76
Fishburn	Dyer	7	44
Dyer	Sarin	10	23
Dyer	Kirkwood	7	8
Winkler	Sarin	8	14
Charnes	Cooper	29	100
Charnes	Glover	7	100
Charnes	Ali	6	34
Glover	Sherali	15	78
Glover	Balas	72	28
Balas	Sherali	10	100
Balas	Nemhauser	19	43
Balas	Ceria	8	45
Laporte	Toth	41	22
Glynn	Whitt	37	100
Glynn	Glasserman	25	33
Glynn	Rubinstein	17	28
Glasserman	Rubinstein	8	14
Carino	Eiemba	4	23
Zenios	Eiemba	9	8
Federgruen	Zipkin	41	45
Whitt	Bitran	4	38
Federgruen	Groenevelt	13	32

There are certain numbers of citation cliques or pairs, but the reason behind this picture is not very obvious. One may argue that people are citing reciprocally or at least they are doing a favor to scholars who have cited them before. On the other hand a counter argument may be that these people (people in above table with many citations) are the leading people of the field and it is natural for them to get cited a lot reciprocally. From the aggregate data above it is not fair to argue that the intention behind these groups is either natural or non-scientific. A micro data analysis which deals with scholars one by one would be necessary to understand the mechanisms behind the structure.

3.5. Summary of Findings

All in all, this chapter aims to show a general picture of historical trends and the network structures in two exemplary fields. Historical data shows that in both fields (operations research and system dynamics) collaboration increases, although the rate of increase is much lower than that in natural sciences (PNAS is taken as the indicator journal for natural sciences). I do not do any analyses on the reason of this increase but increase in the importance of inter-disciplinary work facilitated by the development of the certain technologies (such as internet, video-conference…etc.) seems to be one of the reasons.

In co-citation analyses, the difference between SD and OR is emerged. SD as a compact field has lower rate of interaction with different fields, whereas OR has many connections with different literatures. Moreover in SD field, the people who have been cited most, usually published their papers in late 70's and 80's, whereas in OR field, we highly cited authors who have a balanced distribution in time domain. This implies that OR is more dynamical than SD field.

Co-authorship networks in the SD field show an obvious picture of the collaboration groups. In addition to four major groups, several small groups exist in the field. The scarcity of the data and the limitation of the software packages prevent us to have the collaboration networks of OR field, but I expect a similar structure (with perhaps larger sub-groups), because the micro behavior is similar in two fields. However the co-authorship structure of the PNAS would be much more complex since the papers in PNAS have 8.1 authors on average.

In the inter-citation networks, some patterns of sub-group formation are observed. These sub-groups are more visible in OR field than it is in SD field, because of the different characteristics of two fields. Operations research have many separated different sub-fields, whereas system dynamics is compact compared to OR. In the top 30 author networks, labeling nodes according to some attributes, such as affiliation or region, does not take out a certain pattern, except in OR field the domination of scholars in US is seen. Moreover adding more authors (50 in OR, 20 in SD) changes the network structures a little bit. In the sample author set networks, balanced and unbalanced reciprocity in citation behavior between authors are remarkable. The mechanism behind pair structures might be scientific or non-scientific, but it is hard to argue something by the analysis in this chapter. A micro analysis might be necessary, which is beyond the scope of this thesis.

Another issue related with the trend in academic publications is increase in the number of journals. In recent years – last 20 years – the number of academic journals has been increased substantially. This of course affected the behavior of academicians. Acceptance probabilities have been increased and people tend to submit low quality papers to the low quality journals, in order to make the system count them. Further analysis of this aspect of the situation is promising. Moreover any simulation attempt which mimics this trend in journals might be very informative.

The major problem in the data analyses chapter is the scarcity of data and limitations of software packages for inter-citation data. It is impossible to gather the entire author's inter-citation data of a field. Chapter Five is dealing with this issue, and is about an agent based model which tries to generate a simulated inter-citation data.

4. SYSTEM DYNAMICS MODEL

4.1. Modeling Objective

The previous chapter mostly focuses on the network structure of academic interactions which includes collaboration, co-citedness and giving-receiving citations. After having a picture of the network structures, the aim of this chapter is more about the individual behavior and its aggregate projection at the departmental level.

In academia, the two main measures of research performance are the number of publications and citations. These two measures in a sense quantify the research success of scientists and academic units. Yet, perception of these performance measures can create pressures on researchers and cause different behaviors in different conditions. The objective of this modeling effort is to examine the behaviors of researchers in response to the dynamics of publication and citation pressures. Publishing pressure and citation pressure affect the scholars in different ways. While the former one pushes academicians to publish high number of papers, the latter makes them to worry about the quality. In addition to these different causal mechanisms, factors like reputation, institutional dimensions, and time constraints make the problem more complicated and necessitate a comprehensive systemic analysis of the issue.

4.2. Model

Quality is the most important factor (at least it is supposed to be) that determines number of citations received. As one publishes papers in good quality he/she gets more citations which increases the reputation of him/her. At this point there is an important feedback loop so that the reputation in turn influences the citations the author will receive. During this feedback mechanism a researcher, who has citation pressure on him, would be forced to produce higher quality papers for getting more citations and this increases his/her reputation. On the other hand, there is another mechanism, namely publication pressure. Since the number of citations is not the only measure, and since the quantity of the scientific work matters, there is an inevitable publication pressure which would push the researcher to produce lower-quality papers in higher numbers, in shorter times. At this point I assume that time devoted to a paper is an important determinant of the quality and this assumption brings the idea that there is a tradeoff between quality and quantity. The main decision of researchers is related to this balance and they decide on how to devote their time between quantity and quality, which is modeled through the allocation of researchers' time in research activities and time devoted on each research.

The dynamic simulation model includes faculty members in a department, their publications and citations. System dynamics methodology is used in constructing the model. Reserved time per paper, total research time, reputation and quality of papers are included in the model as the main factors affecting the behavior. Complete stock-flow diagram can be seen in Figure 4.1.

For the base run School of Engineering at the Boğaziçi University is chosen to be examined. The initial conditions, the number of faculty and grand average values are obtained in the data analysis part of the paper by (Eskici, et al., 2008). Time unit in the model is quarters and 200 quarters (50 years) is examined in the simulation. Time step (DT) analysis is done and DT is chosen as 1/8.

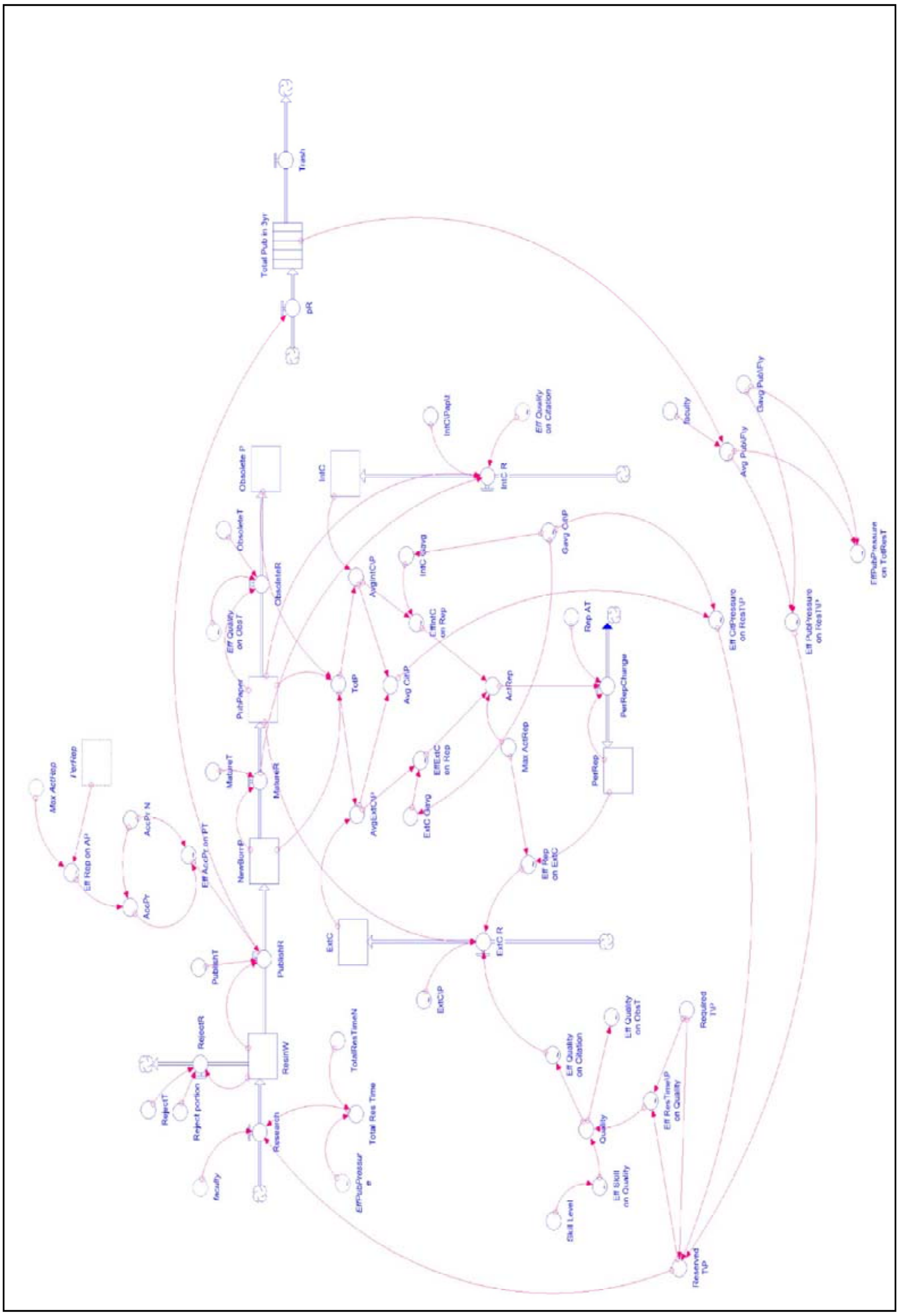

Figure 4.2. Stock - Flow diagram of the complete model

The model has the aging structure of the papers as the core part and citation – reputation relations as the supplementary part to this structure. The stocks in the model can be seen in an abstract view in Figure 4.2. The four stocks seen in the upper line represent the papers in different stages. The first stock in this aging structure is ResinW (Research in Writing) stock and shows the papers which are in the research stage yet. The work in ResinW stock either turns into a paper or goes into trash. Papers which are published are started to be cited after approximately 3 years and this transition stage is represented by the NewBornP (New Born Papers) stock which shows the published but not citable papers. The stock PubPaper (Published Papers) represents the papers which are published and being cited. After staying in PubPaper for a long time depending on their quality, the papers become old. The stock Obsolete P (Obsolete Papers) represents the papers which are published a long time ago and do not get citation any more.

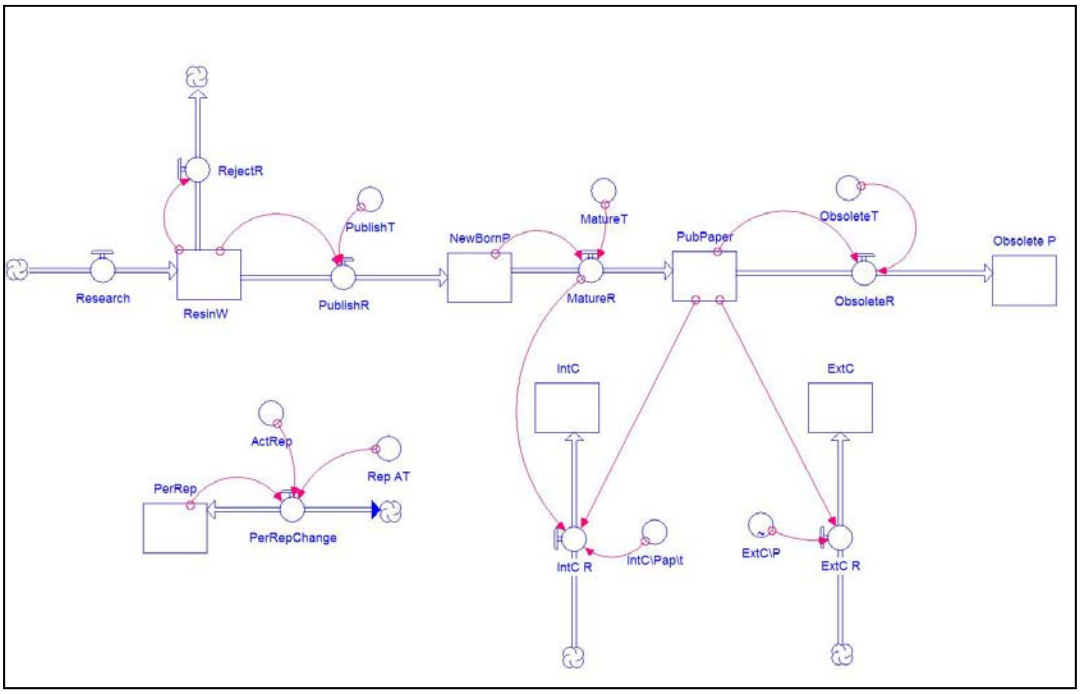

Figure 4.3. Simplified stock-flow potion of the model

IntC accumulates internal citations (citations from the same department of the author) and ExtC accumulates external citations (from other departments). Finally, PerRep is the perceived reputation of the department all over the world. Since the reputation doesn't change immediately, in the modeling of reputation, there is a first order information delay.

The most important effect variables which are not shown in Figure 4 are shown with a causal-loop diagram in Figure 4.3.

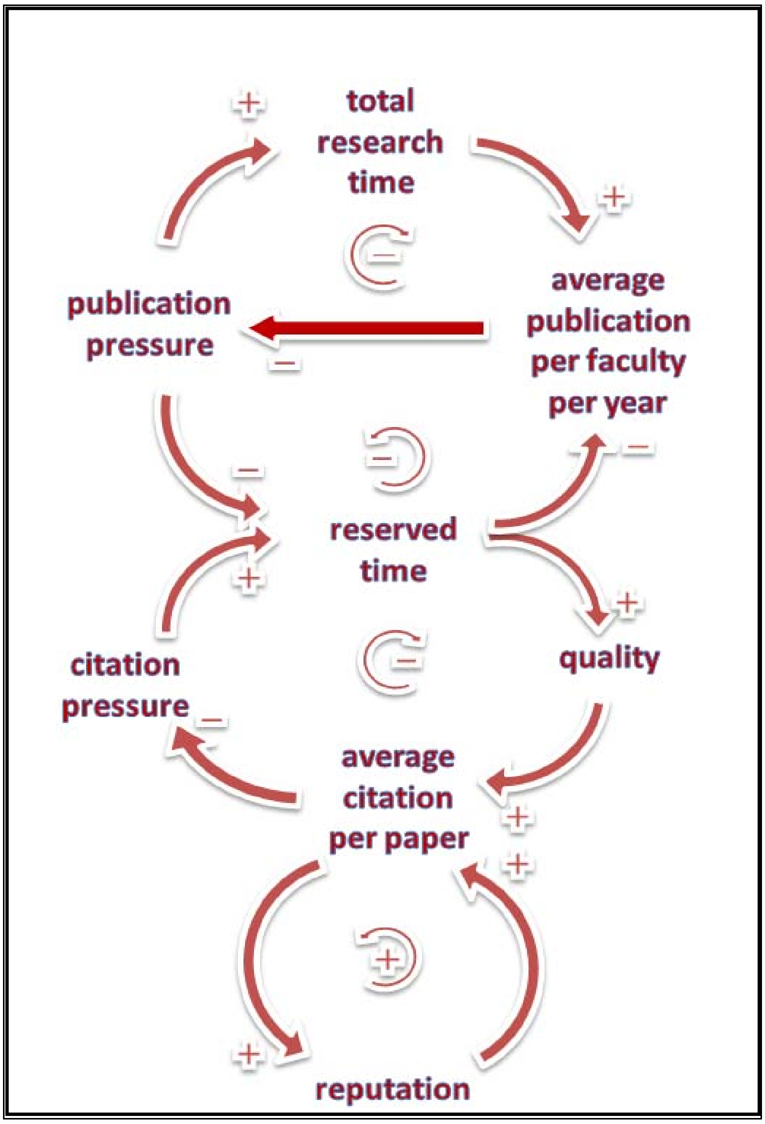

Figure 4.4. Causal-Loop diagram

In Figure 4.3 “total research time” is the researchers’ time in research activities and “reserved time” is the time devoted on each paper. As seen in the causal-loop diagram, reserved time per paper is affected by publication and citation pressures. When citation pressure increases, a researcher would try to increase his citations by producing higher quality papers. So, he would increase reserved time, i.e. he would spend more time on each research. This would reduce average publication per faculty per year; which would increase the publication pressure. When publication pressure increases, a researcher would be forced to increase his publications and so he would produce more publications in shorter times which would decrease the quality. There are two negative feedback loops regarding reserved time per paper and pressures. So the faculty members will come over these pressures by deciding on the amount of time spent per paper. Besides, total research time is affected by publication pressure in that if there is pressure, amount of time allocated to research activities would increase. This negative feedback loop will try also to overcome publication pressure. Apart from these, there is an important and well-known positive feedback loop between reputation and citations. If average number of citations per paper increases the faculty will increase its reputation and if its reputation is high it will get more citations.

Before discussing model formulations, the list of basic variables and their dimensions are shown in Table 4.1.

Table 4.1. Basic variables and their types and dimensions

ResinW	Research in Writing	Stock	Paper
NewBorn P	New Born Papers	Stock	Paper
PubPaper	Published Papers	Stock	Paper
ObsoleteP	Obsolete Papers	Stock	Paper
ExtC	External Citation	Stock	Citation
IntC	Internal Citation	Stock	Citation
Research	Research per Quarter	Flow	Paper / Quarter
PublishR	Published Papers per Quarter	Flow	Paper / Quarter
MatureR	Maturation Rate of Papers per Quarter	Flow	Paper / Quarter
ExtC R	External Citation Rate	Flow	Citation / Quarter
IntC R	Internal Citation Rate	Flow	Citation / Quarter
PublistT	Publishing Time	Auxilary	Quarter
MatureT	Maturation Time	Auxilary	Quarter
ObsoleteT	Obsolescence Time	Auxilary	Quarter
ExtC /P	External Citation Per Paper Per Quarter	Auxilary	Citation / Paper /Quarter
Res T /P	Reserved time per Paper	Auxilary	Time / Paper
Total Res T	Total Research Time Per Quarter Per Author	Auxilary	Time / Paper / Quarter / Author

4.3. Model Formulations

4.3.1. Citation Pressure and Publication Pressure

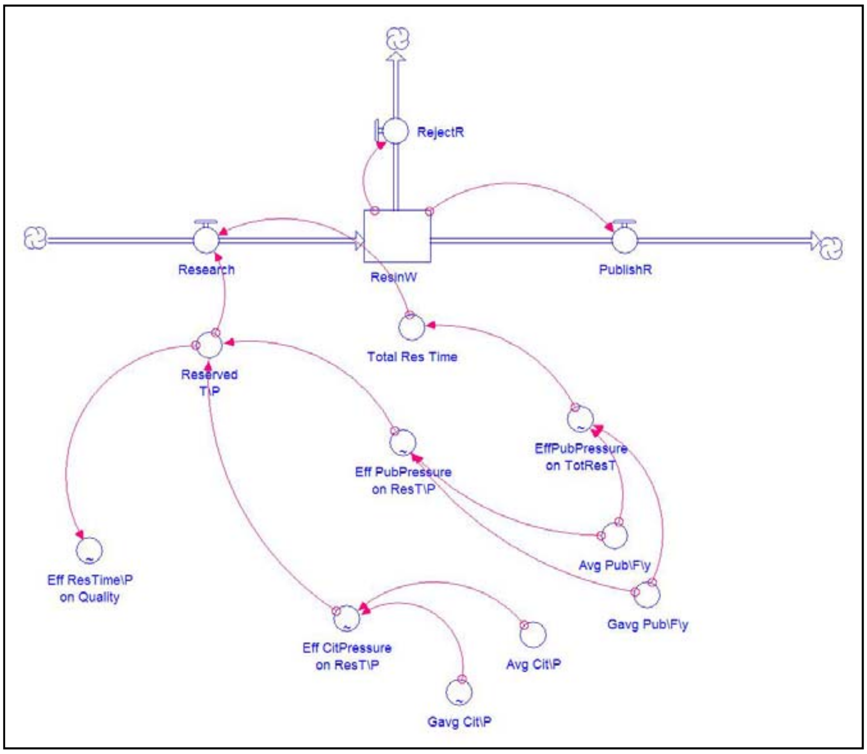

Figure 4.4. Citation pressure and publication pressure in the model

One of the main effects in the model is the balance between citation pressure and publication pressure.

$$\text{Citation Pressure} = \frac{\text{Average Citation per Paper}}{\text{Grand Average Citation per Paper}}$$

$$\text{Publication Pressure} = \frac{\text{Average Publication per Faculty per Year}}{\text{Grand Average Publication per Faculty per Year}}$$

where grand average publication per faculty per year and grand average citation per paper are the world averages in the engineering field[10].

[10] The data are taken from (Eskici, et al., 2008)

It is assumed in the model that, the time the researchers spend on each research depends on the pressures on them. If the average number of citations per paper is lower than the world average, than the researchers will feel a pressure to produce higher quality papers to get more citations. For higher-quality papers, the researchers will need to spend more time on each paper. The formulation of reserved time in the model is as follows;

Reserved time per paper= Required time per paper*Effect of Citation Pressure
*Effect of Publication Pressure

There are two effects which are inserted to the equation in multiplicative form. These are effect of citation pressure and effect of publication pressure, and they affect the reserved time per paper in opposite ways. If the number of publications of the researchers is much lower than the world average in the particular field, then the researchers will feel a pressure to produce more publications. In order to increase the number of publications, they will decrease time devoted on each paper. This will provide more publications with a lower-quality. On the other hand citation pressure increases if the received citations fell behind the world average and citation pressure increases the reserved time per paper in order to increase quality and citation. The shape of effect functions can be seen in Figure 4.5. Since there is a saturation effect in the lower and upper ends, S-shaped function has been chosen.

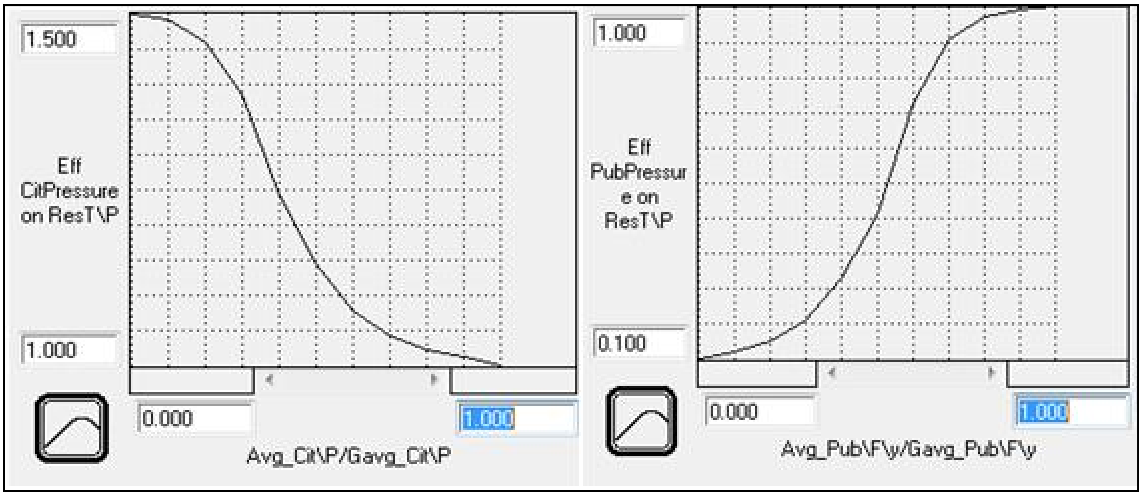

Figure 5.5. Effect of citation pressure and publication pressure on reserved time per paper

4.3.2. Total Research Time

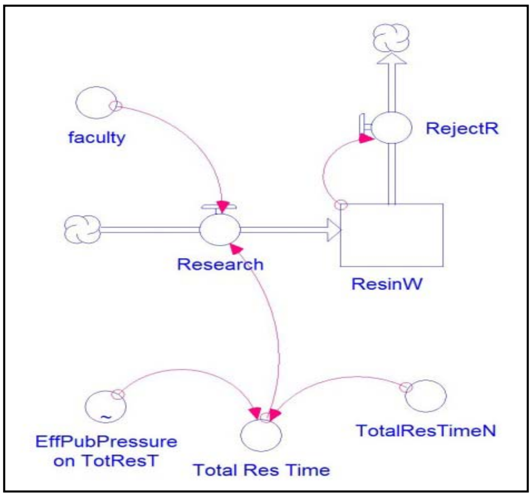

Figure 4.6. Total research time section

It is assumed that, if there is publication pressure, then in addition to decreasing the reserved time, the researchers will also try to increase the total time they spend on research activities. This could be possible by decreasing administrative or teaching load, or working at the weekends.

$$\text{Research}=\text{Faculty}*\frac{\text{Total Research Time}}{\text{Reserved Time per Paper}}$$

Total Research time=Effect of PubPressure on Res Time*Total Res Time Normal

4.3.3. Quality

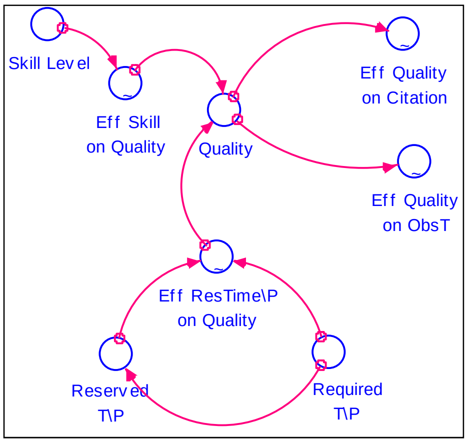

Figure 4.7. Quality section

Quality is one of the key effects in the model. Skill level of members of the department is very important in the quality of the papers. It is assumed that, good-quality universities hire researchers who produce good-quality papers. So, one of the indicators of the quality is the overall skill level of the members of the department. In the model, skill level is an exogenous variable. Required T/P (required time per paper) is the time that is needed to produce a paper in normal quality level.

The other indicator of quality is reserved time. If reserved time is lower than the required time, which is the time that is needed to produce a paper in normal quality level, then the paper will be a low-quality one. As the time spent on a paper increases, its quality level increases. The formulation of quality in the model is seen below.

Quality=Normal Quality *Effect of Research Time per Paper * Effect of Skill

A good-quality paper gets more citations than the others. So there is a positive relationship between quality and number of citations. Additionally, if a paper is in good-quality, its obsolete time is longer. Obsolete time is the time that how long a paper stays in the stock of published papers (PubPaper) after being published.

4.3.4. Reputation

Reputation of the department is determined by the average number of citations received by the papers of the department's members. The average external citations and the average internal citations of the faculty are compared with the grand average (world average) values. Since there is a time delay in perceiving any change in reputation, first order information delay structure is used in modeling this part

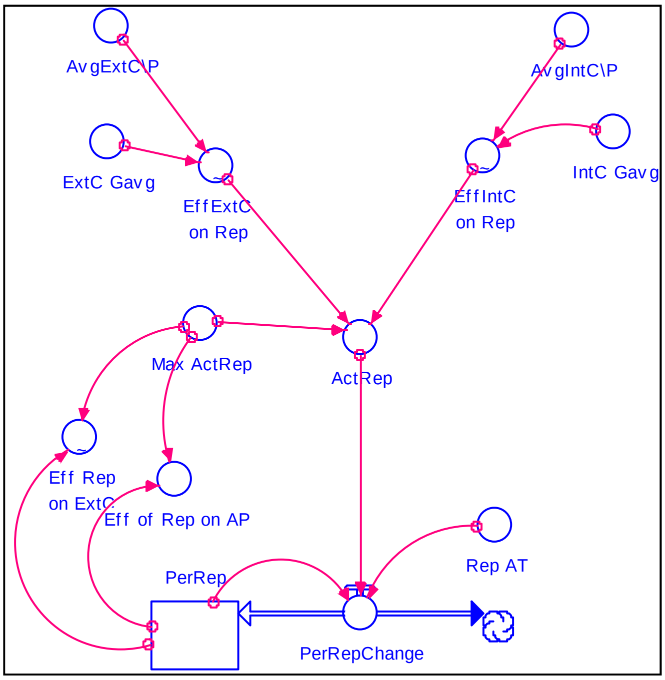

Figure 4.8. Reputation section

$$\text{Perceived Reputation Change} = \frac{\text{Actual Reputation} - \text{Perceived Reputation}}{\text{Reputation Adjustment Time}}$$

Reputation is directly affecting the external citations because reputation means being known by the other academicians. Other academicians prefer to cite from the one whom they know more than from anyone On the other hand, reputation does not affect internal citations since internal citations come from the colleagues of the researcher. Reputation has also a positive effect on the acceptance probability of a paper to a journal.

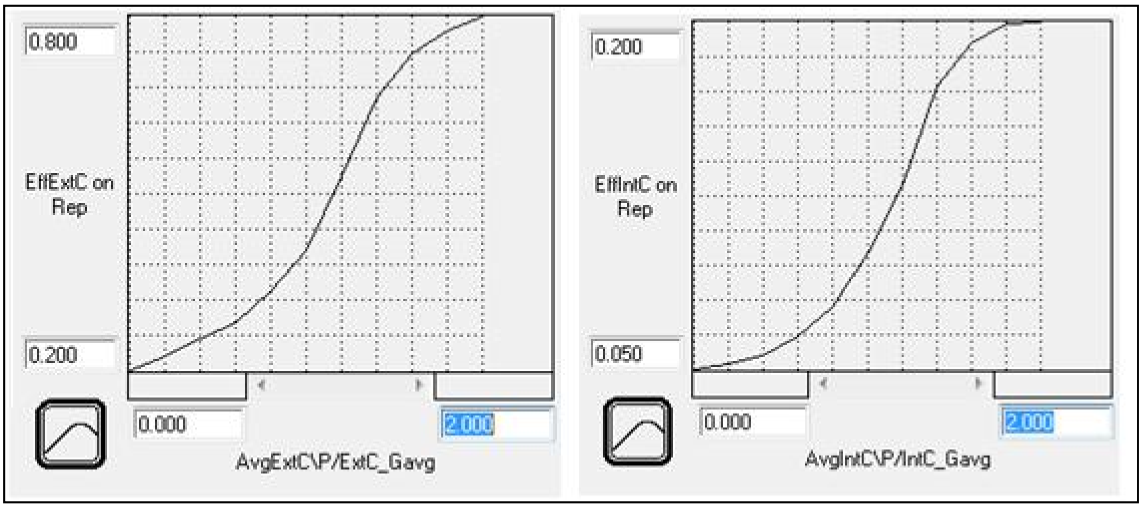

Figure 4.9. Effect functions of external and internal citations on reputation

4.3.5. Publication Flow

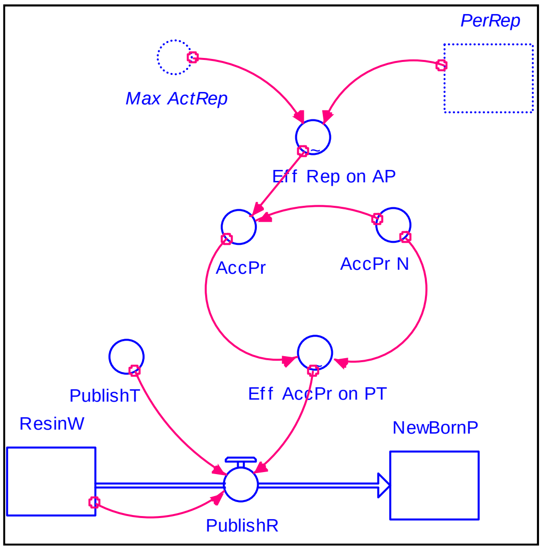

Figure 4.10. Publication Flow section

AccPr (Acceptance Probability) is the probability that a paper is accepted by a journal. PublishT (Publish time) is the time that a paper waits before being published. When PerRep (Perceived Reputation) is close to MaxActRep (Maximum Actual Reputation) AccPr (Actual Probability) is high. In the same manner, when the perceived reputation is low, AccPr is low. AccPr has a negative effect on PublishT (Publish time). I.e., more reputation means more acceptance probability and more acceptance probability means less waiting time for the paper before being published in a journal.

$$\text{Publish Rate}=\frac{\text{Research in Writing}}{\text{Publish time*Effect of Acceptance Pr on Publish Time}}$$

4.4. Model Validation

The purpose of model validation is to assure that the model is an acceptable description of the real system behavior with respect to the dynamic problem (Barlas, 1996). Model validation is carried out in two steps.

4.5.1. Structure Validity

Structure test is to check whether the structure of a model is a meaningful description of the real relations that exists in the problem or not. There are two types of structure tests: direct structure tests and structure-oriented behavior tests (Barlas, 1996).

Direct structure tests assess the validity of the model structure by direct comparison with knowledge about real system structure. Parameter and variable confirmation, dimensional consistency and extreme condition tests are included in direct structure testing (Barlas, 1996). In the model, all parameters and variables have real life counterparts, there is no dimensional inconsistency in equations and the model passes the extreme condition tests.

One of the tests in indirect structure testing is extreme-condition test via simulation. In order to validate the model some extreme conditions are simulated. One of our external input variables is skill level. The upper extreme for skill level is 100. When we start the simulation with a skill level of 100, reputation climbs up to the maximum value of 100. It is consistent in that, if a school consists of the most skilled faculty members in the world it becomes the most reputed one in the world.

Another extreme-condition test is applied to the number of faculty. When there is 1 faculty member, all publication stock levels decrease as expected. On the other hand, when we start with a faculty of 300 members publication stock levels come to equilibrium at high levels. Additionally, extreme-condition test is done with the total research time parameter. If faculty members allocate a very small portion of their time to research (for example 5% of a quarter) then publication stock levels decrease as expected. On the other hand, if very high portion of available time is devoted to research, publication stocks reach their equilibrium at high levels. These entire extreme-condition tests are consistent with the construction of the model. Results of some of the extreme conditions are in Appendix.

4.4.2. Behavior Validity

Behavior pattern tests are designed to measure how accurately the model can reproduce the major behavior patterns of the real system (Barlas, 1996). Real data is not available for our case; however we can judge the resulting behavior of the system. According to our assumptions, there should be a balance between the pressures and the actions of the faculty. When the behavior is examined it is seen that time reserved for a paper reaches its equilibrium after a set of decisions according to publication and citation pressures. This is kind of seeking a balance between number of papers published and citations received. This main behavior is consistent with our assumptions.

4.5. Output Analysis

4.5.1. Base Run

As seen in Figure 4.11, new-born papers and research in writing stocks reach their equilibrium after oscillation. This is a result of negative feedback loops of the model. Mainly, publication and citation pressures govern these oscillations. Published papers stock has also a kind of oscillation before it settles down. Because of the fact that there is not an outflow of obsolete papers, this stock continues to grow.

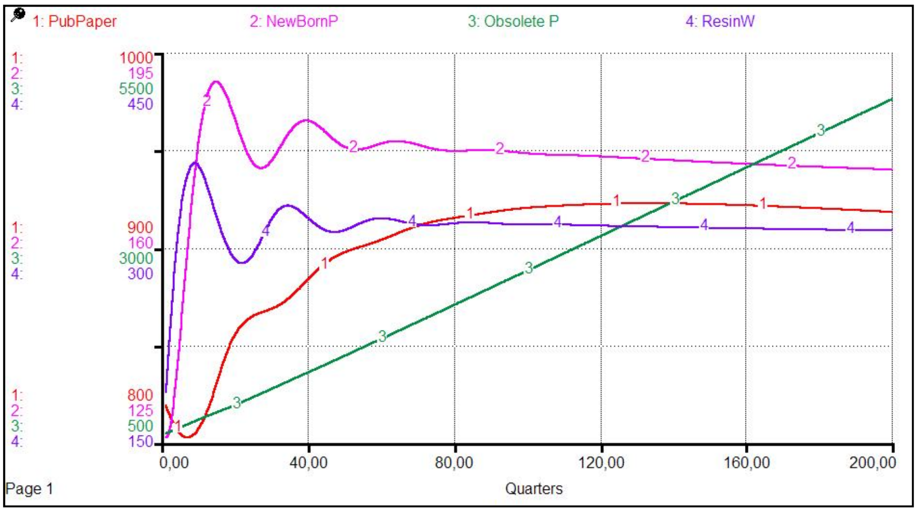

Figure 4.11. Output of paper stocks - Base run

In Figure 4.12 pressure effects can be seen. Publication-pressure increases the total research time while it decreases the reserved time per paper. In the figure, these opposite effects can be seen easily. Effects of publication pressure reach equilibrium after oscillations.

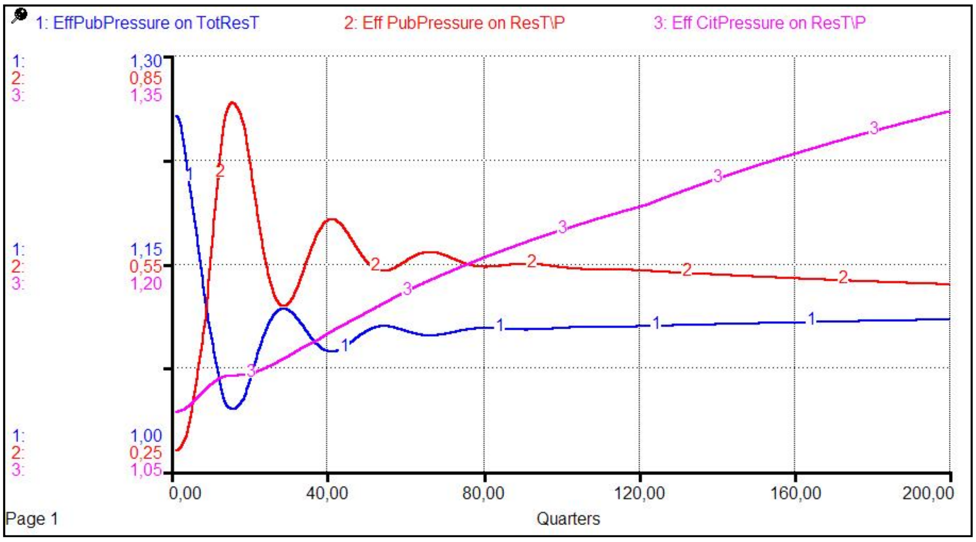

Figure 4.12. Pressure effects in the base model

In Figure 4.13, behaviors of reserved time per paper and total research time per faculty are shown. Reserved time per paper is the decision of the faculty on the average time devoted to a paper under the effects of citation and publication pressures. Faculty seeks equilibrium for the reserved time per paper and it results with a damped oscillation. Total research time per faculty is also the decision of the faculty in terms of the time devoted to research per faculty member per semester. Faculty seeks equilibrium for it and it results with a damped oscillation. It is seen that behavior of these two variables are in the opposite direction. This is as expected because if there is a publication pressure, then the total research time per faculty will increase; however reserved time per paper will decrease to be able to publish more papers.

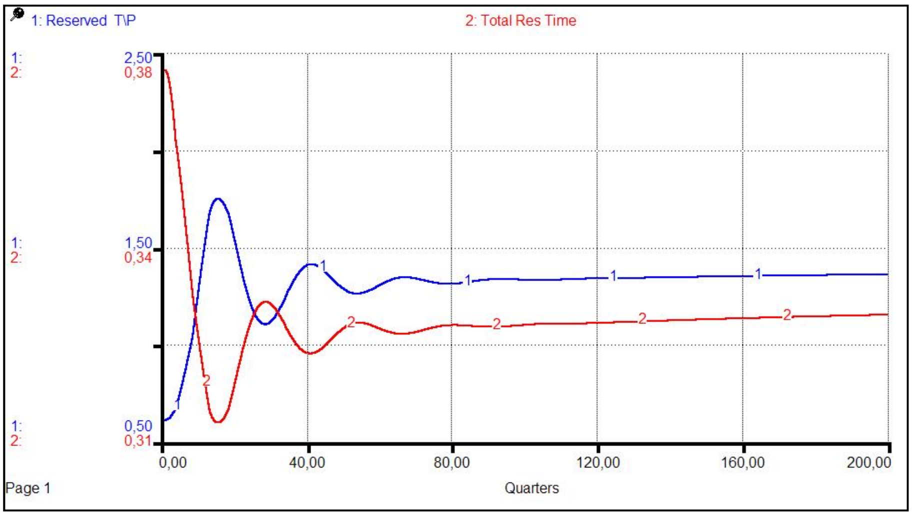

Figure 4.13. Reserved time per paper and total research time in the base model

Figure 4.14 shows the behaviors of average citation per paper, perceived reputation and quality. As I have mentioned before, when citation increases reputation increases, too. From the figure this relation can be seen easily. Quality has an oscillation because it is mainly related to reserved time per paper. Because of the fact that quality is below 1 it effects citation negatively.

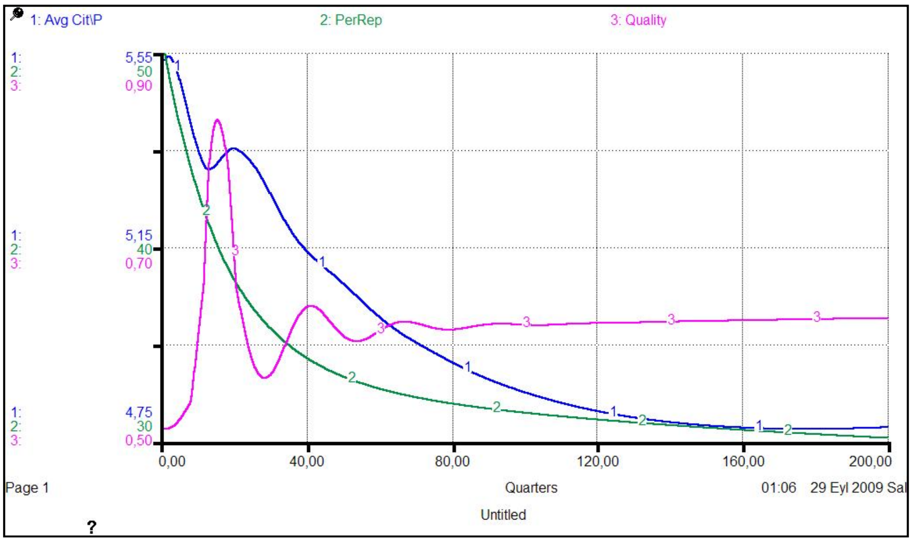

Figure 4.14. Average citation per paper, reputation, and quality in the base model

Last two figures are about the behavior of cumulative number of papers and paper publication rates. In Figure 4.15, Cumpapers represents the sum of new born papers and published papers. Cumulative number of papers has the behavior shown in blue line.

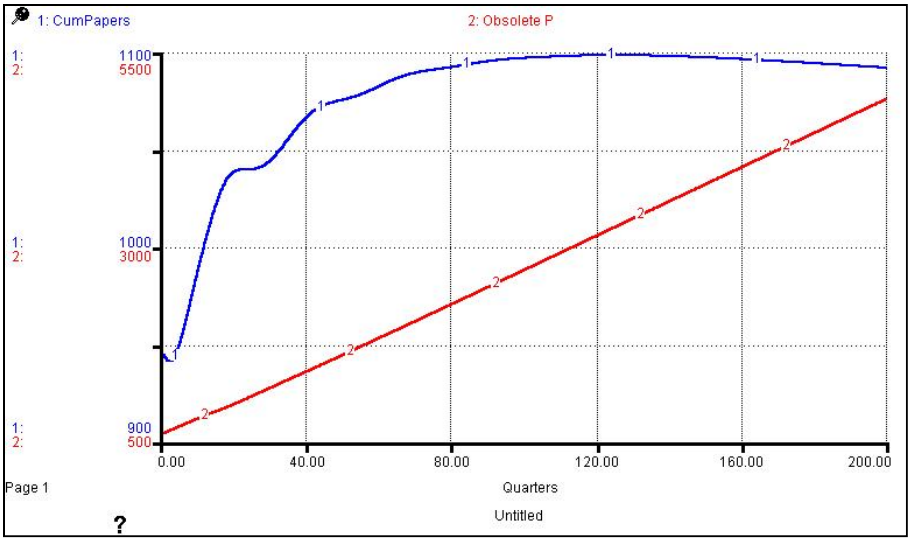

Figure 4.16. Cumulative paper stocks in the base run

On the other hand in Figure 4.16, paper flows are shown separately. Blue line represents publishing rate which is publication rate per quarter of the department. Since the number of faculty is constant, the behavior of the average publication rate per faculty is the same. After some oscillations the publication rate of the department stabilizes around 20 papers per quarters, which means roughly 0.8 papers per academician per year. Maturation rate has similar behavior with publication rate, except the oscillations are smaller. Lastly green line represents the obsolete rate which stabilizes around 20 after transition period has finished.

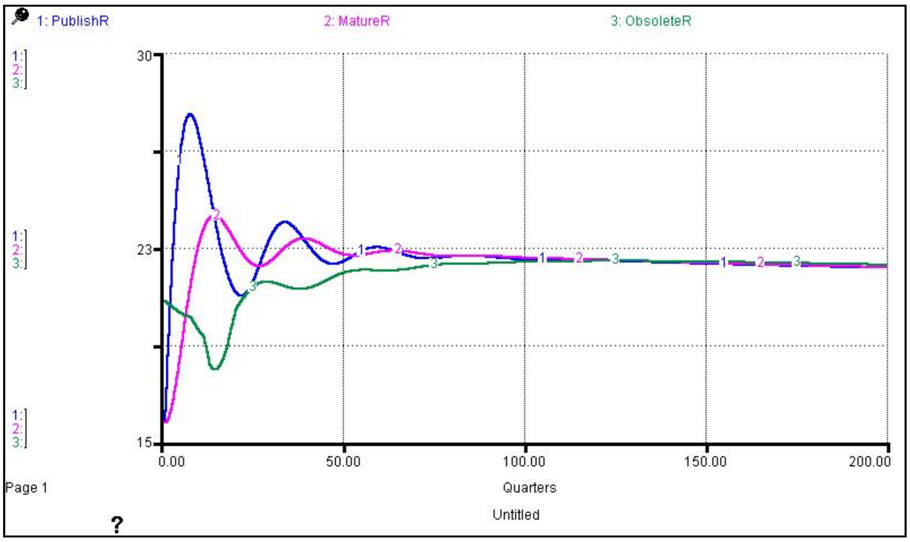

Figure 4.17. Paper flows - publication, maturation and obsolete rates - in the base model

4.5.2. Scenario Analysis

<u>4.5.2.1. No Publication Pressure Effect on Total Research Time:</u> If there is no pressure effect on total research time, published paper stock reaches equilibrium at a lower level than it does in the base model. It is expected because faculty will not be able to increase number of publications as much as that in the base model. The behavior of the paper stocks can be seen in Figure 4.17.

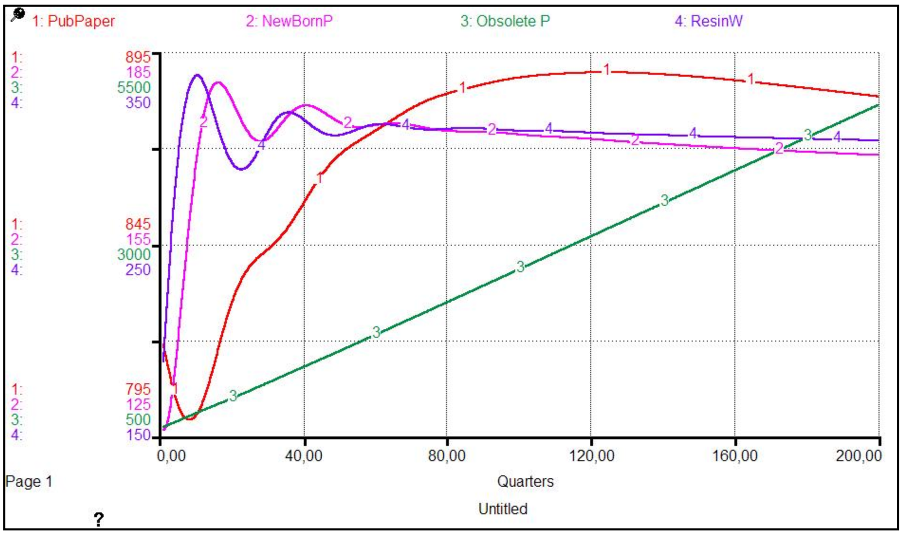

Figure 4.17. Paper stocks in scenario 1

As seen in Figure 4.18, reserved time per paper reaches equilibrium at a lower value in the scenario of no publication pressure compared to the base model. This is expected; faculty cannot increase total research time and to be able to catch the world average of the publication performance, faculty should decrease the amount of time devoted to each paper.

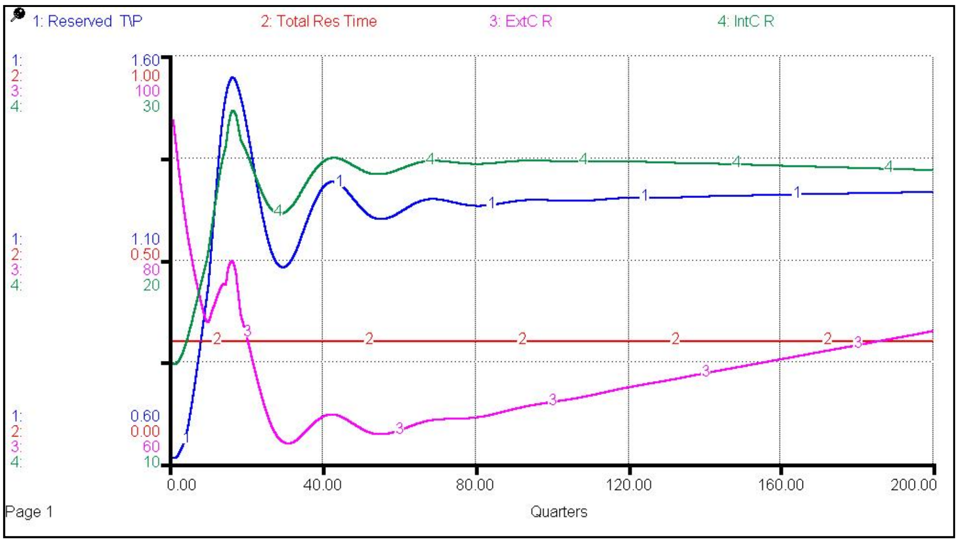

Figure 4.18. Reserved time per paper, total research time and citation flows in scenario 1

4.5.2.2. No Citation Pressure: Everything being equal, if there is no citation pressure, the faculty does not keep track of the citations received and so does not care about quality. The main effect is on reserved time per paper and on total research time per faculty which can be seen in Figure 4.20. As expected, reserved time per paper reaches equilibrium at a lower level than that does in the base model. Besides, total research time's equilibrium value is lower than its value in the base model, because of the same reason. In Figure 4.19, it is seen that paper stock values are higher compared to the base model as a result of devoting less time to each paper in the absence of citation pressure.

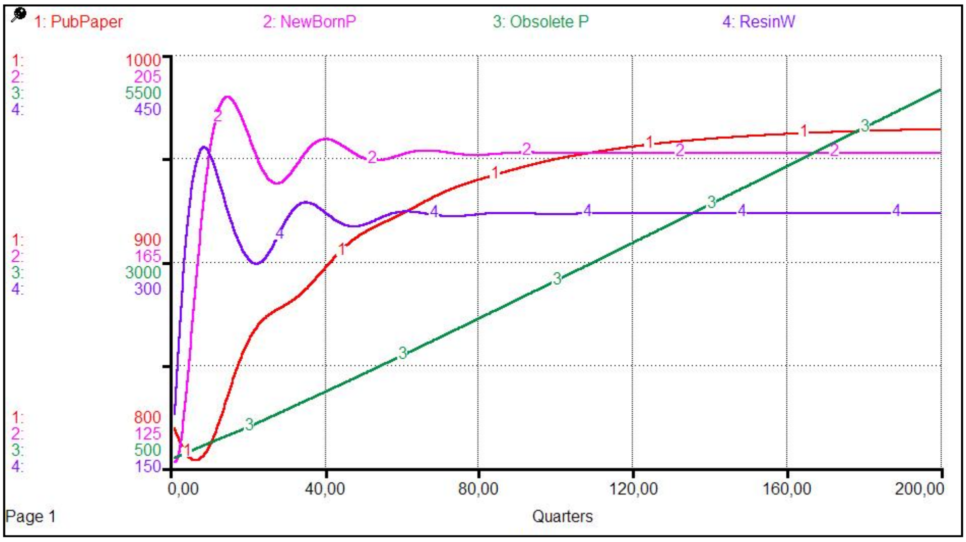

Figure 4.19. Paper stocks in scenario 2

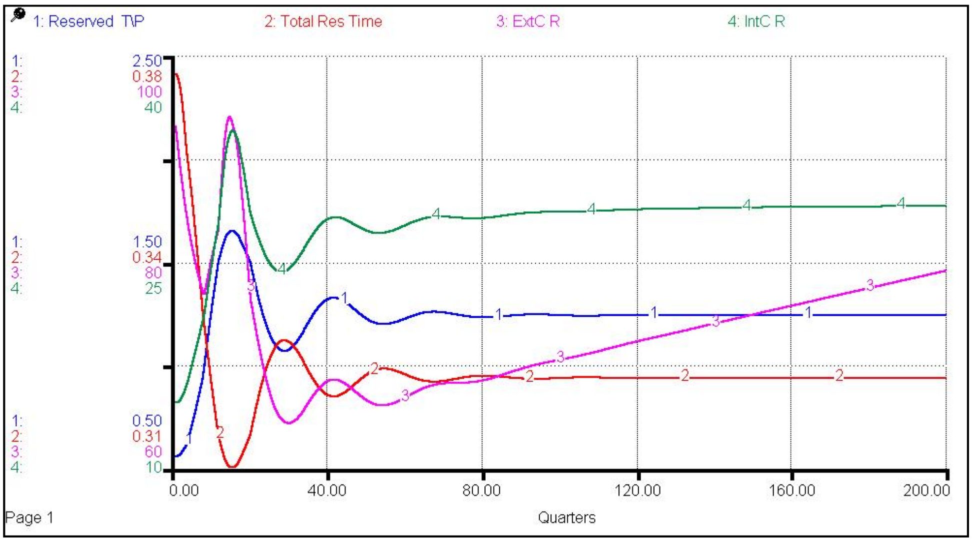

Figure 4.20. Reserved time per paper and total research time in scenario 2

<u>4.5.2.3. Low Level of Skill:</u> Skill level is an input for the quality, as stated before. In the base model it was 50 (normal value for the quality). Different scenarios are created with different values of skill level. One of them is carried out with a skill level of 20. In this

case paper stock levels decrease as expected (Figure 4.21). Besides, because of the fact that quality of the papers is low, citation pressure occurs. Compared to the base model reserved time per paper reaches equilibrium at a higher level (Figure 4.22). Quality decreases and this effects the citation and reputation negatively (Figure 4.23).

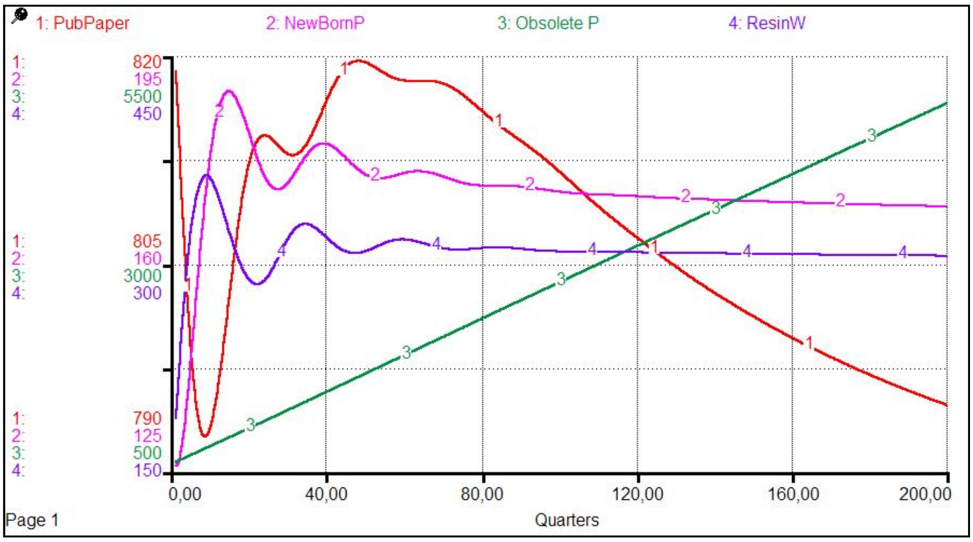

Figure 4.21. Paper stocks in scenario 3

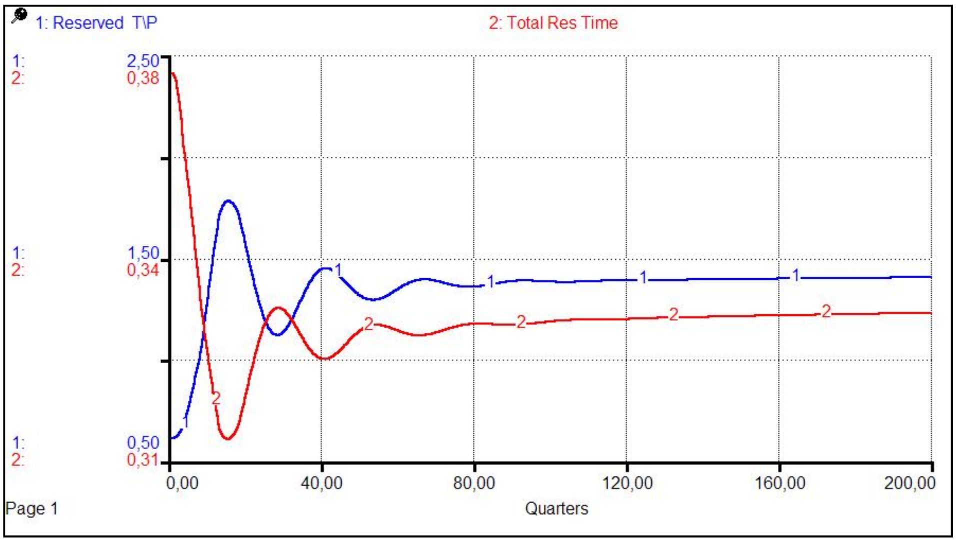

Figure 4.22. Reserved time per paper and total research time in scenario 3

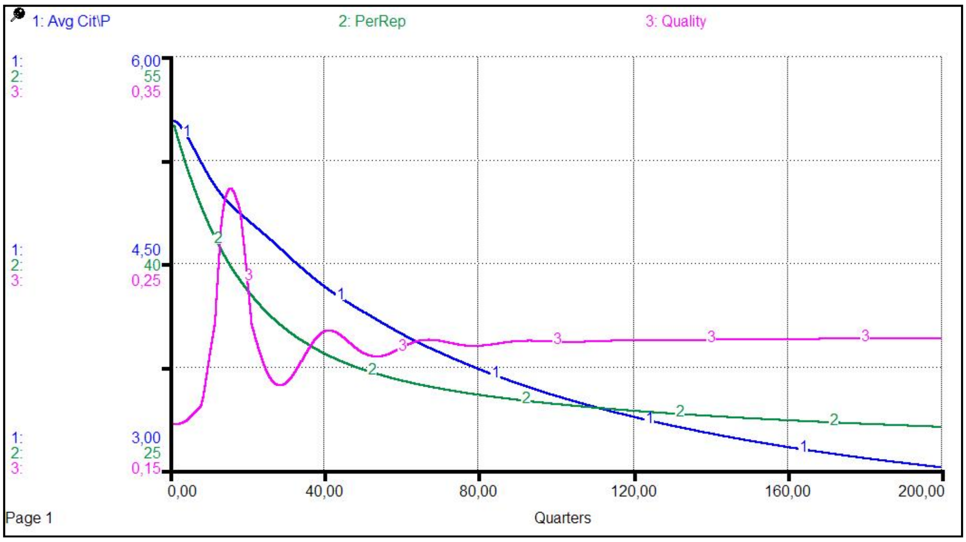

Figure 4.23. Average citation per paper, reputation, and quality in scenario 3

<u>4.5.2.4. High Level of Skill:</u> In this case skill level is increased to 80 and as a result paper stock levels increase as seen in Figure 4.24. Because of the fact that quality of the papers is high, citation pressure is not effective. As a result compared to the base model reserved time per paper reaches equilibrium at a lower level. Parallel to that total research time has a lower equilibrium than that of the base model. These behaviors can be seen in Figure 4.25.

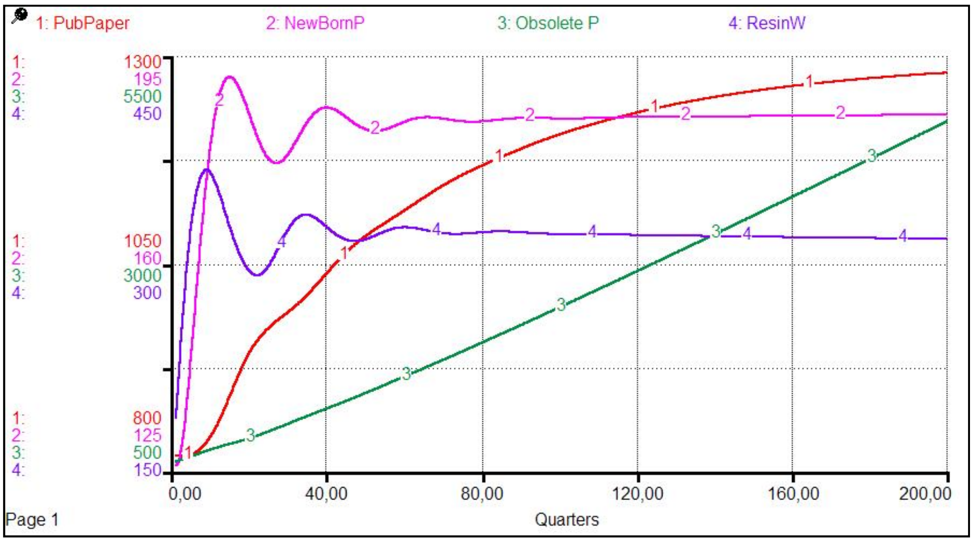

Figure 4.24. Paper stocks in scenario 4

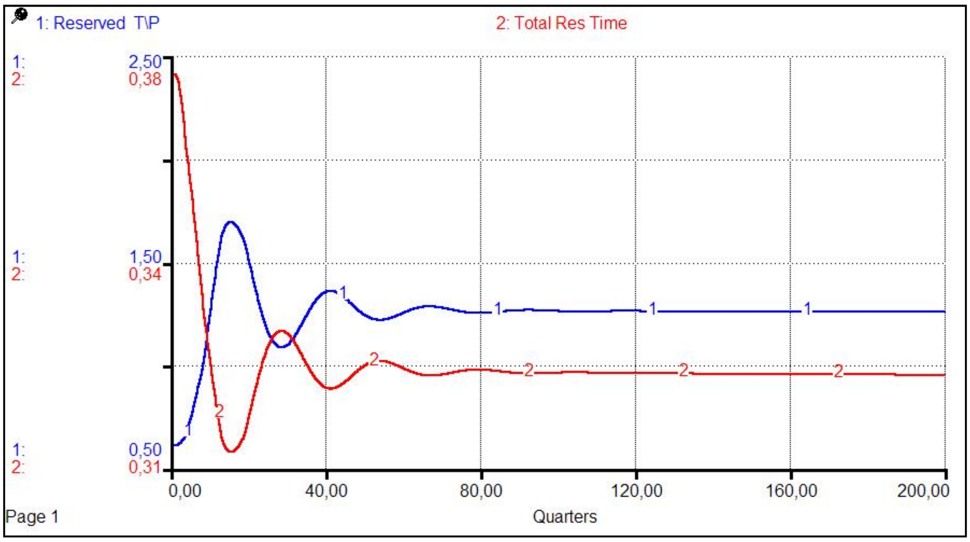

Figure 4.25. Reserved time per paper and total research time in scenario 4

In Figure 4.26, it is seen that quality increases with the high skill level. As a result of higher quality compared to the base model, average citation values and reputation increase.

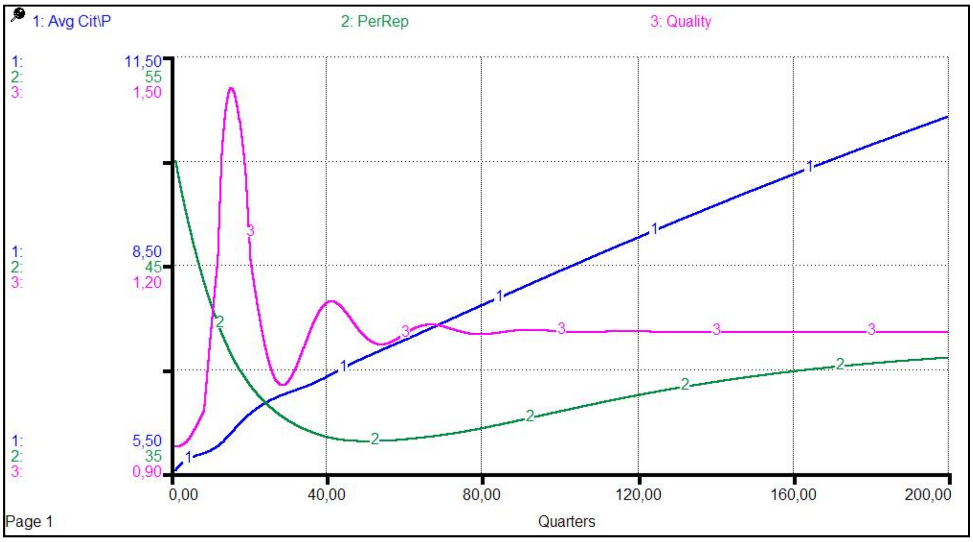

Figure 4.26. Average citation per paper, reputation, and quality in scenario 4

<u>4.5.2.5. Low Initial Reputation and High Level of Skill:</u> To be able to show the effect of the initial reputation and the skill level together, these last two scenarios are created. In our base model we have taken initial reputation as 50. In this case initial reputation is 20 and the skill level is 80. As seen in Figure 4.27, quality is high, and reputation climbs up together with citation, because skill value is quite high.

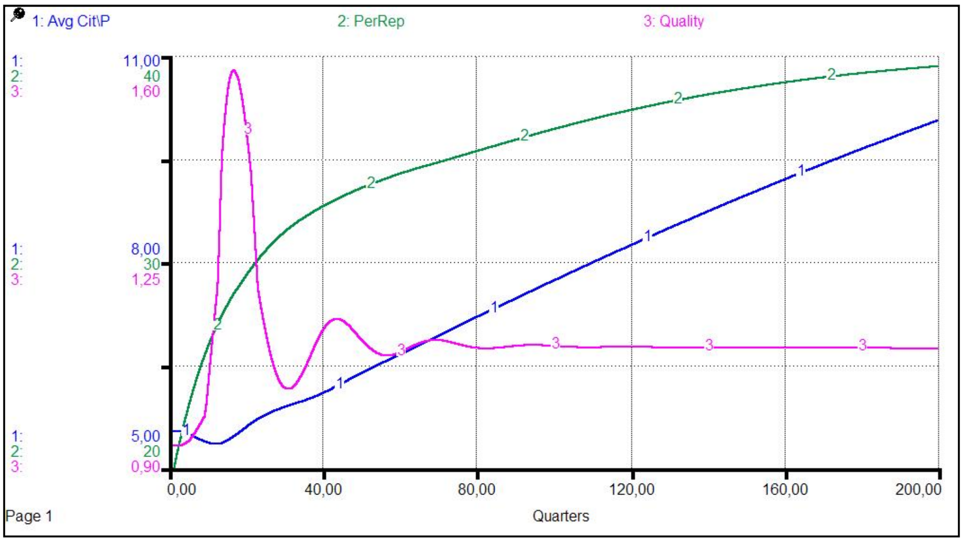

Figure 4.27. Average citation per paper, reputation, and quality in scenario 5

<u>4.5.2.6. High Initial Reputation and Low Level of Skill:</u> In this last scenario, initial reputation is taken as 100 and the skill level is 20. As seen in Figure 4.26, quality is low, and reputation goes down together with citation, because the skill value is quite low.

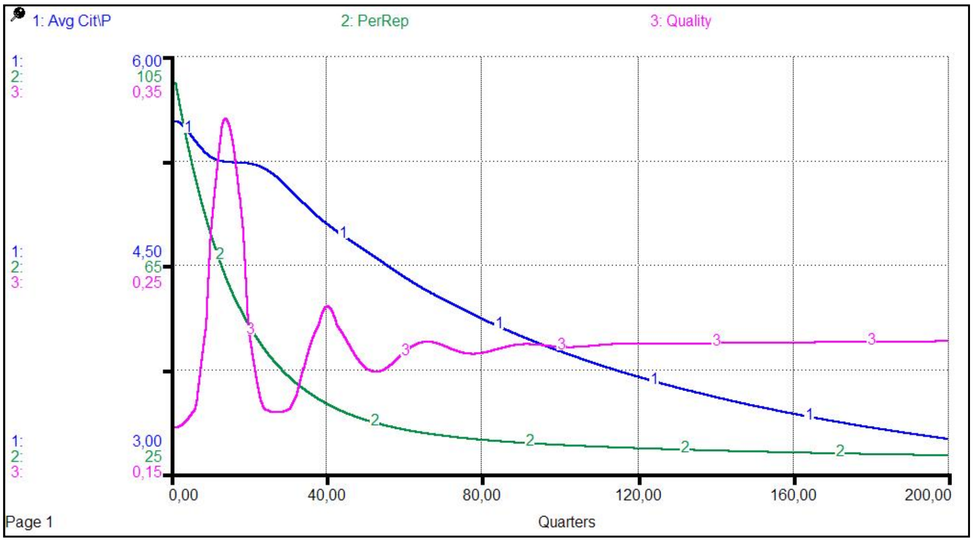

Figure 4.28. Average citation per paper, reputation, and quality in scenario 6

4.6. Chapter Conclusion And Summary Of Findings

The aim of the forth chapter which is to examine the behaviors of researchers in response to dynamics of publication and citation pressures is achieved by using a model including researchers in a department, their publications, citations and the factors such as reputation, quality, pressures on researchers and their skill levels.

The main decision of the department (faculty members) is the allocation of time to produce in high quality / number or low quality / number of papers and this decision creates the dynamics. In the base run, quality of publications, reputation of the department, publication and citation pressures are seen as main factors. High level of intellectual skills and devoting more time result in high quality papers which get more citations. As papers get more cited, reputation increases and this results in more citations. The positive feedback loop between reputation and citation is very strong but other feedback mechanisms balance it. Publication and citation pressures act in opposite way. While former causes producing more papers in shorter times (low quality); the latter tries to make high quality papers in longer times (few paper). These opposite effects cause model to reach equilibrium after some oscillations.

In the scenario analyses, when citation pressure is removed the paper stocks reach equilibrium at higher levels with low quality, less-cited papers. Additionally, the system is sensitive to the skill level which is modeled as an exogenous factor. And lastly, when the reputation and skill level analyzed together, it is seen that the skill level is a decisive factor. A skilled department obtains reputation, regardless of the initial level of reputation.

As further research, in order to better grasp the decision mechanism of the researchers, this model can be widened by including other pressures (such as career, financial…etc). Making skill level an endogenous variable could make the model more complex and realistic. Or by adding an agent based approach, interrelations between multiple departments can be analyzed with a multi-level model.

5. AGENT-BASED MODEL

5.1.Modeling Objective

As mentioned in the introduction chapter, my aim in this thesis is to grasp the structural problems of the academic publication – citation relationships and behavior dynamics under the science policies which are heavily relied on counting numbers of publications and citations. For this purpose, Chapter Four – system dynamics model – is an attempt to shed a light on behavior dynamics, whereas Chapter Three – historical data analyses section – summarizes the structural picture in two exemplary fields with empirical data. This last chapter is related to network structures and generates simulated inter–citation data among authors.

ISI web of science is one of the main data sources in the publication – citation indexing. It provides a wide range of data in a massy way. Since ISI web of science is a commercial institution, it limits the size of the data, which can be downloaded at each time, with 500 papers. This makes collecting data of more than 20k papers almost impossible. In addition to the scarcity of the data, there is also a limitation in obtaining the inter–citation data among authors. Although there are some software packages (such as CiteSpace, Histcite) which enable taking out inter-citation data of papers and co-citation data of both papers and authors, there is not any software which can obtain inter-citation data of authors, from the raw data of ISI Web of science. A proper simulation model which can generate artificial data is required to analyze the network structures of authors.

This chapter is about a process – oriented model which generates artificial data of inter-citation networks of authors. Analyzing inter-citation data is important to see the emerging network dynamics and whether or not there are patterns of citation cliques.

5.2. Agent-Based Modeling Methodology

Agent-based is a relatively new modeling technique compared to system dynamics theory. Agent based models are based on individual actions of agents, their interaction with other agents and with environment. Agent which is the building block of the models is defined as “a system component that has autonomy in its actions and has a social ability to interact with other agents in the system through some patterns like cooperation, coordination, and negotiation.” (Wooldridge, 2001). By modeling the smallest component of the system, agent based modeling is a micro approach, and usually the agent structures are simple, because a powerful aspect of the approach is that the source of the complex dynamic behavior is the interactions. The aim of the agent based modeling is to look at the global consequences of local actions (Demirel, 2007). In this sense system dynamics and agent based methodologies look to the problem from opposite sides. While the former one approaches from top to bottom, the latter one does from bottom to top.

Since the problem is lack of inter-citation data which is the direct consequence of the individual behavior of the scholars in real life, agent based approach seems appropriate for this model. As a last reminder agent based models are also called as process-based models, because actually what is modeled is simple rules of processes led by individual actors. I use both interchangeably.

5.3. Model

The model basically consists of “academicians” as the main set of agents and “papers” as the secondary set of objects, and the publication – citation processes followed by the academicians and papers. Agent structures are quite simple, because in the agent based models complexity emerges from the simple interaction rules which are modeled very carefully. Basically the model simulates a set of academicians who publish new papers, in a multi year period.

I have used NetLogo which is a Java based multi-agent programmable modeling environment. Although agent based modeling does not require specific software, it is easier to use programmable environments which have short cuts and features designed for multi-agent programming.

In the following section I explain the academician and paper structures, which will be followed by discussion of the process flow. After that, details of the mechanism, the rules of co-authorship formation and citing will be explained. I will conclude this section by listing the assumptions of the base model.

5.3.1. Structure of Academicians

The model has academicians as the building block. Academician object has been designed as to be able to carry all the information required in the publication – citation processes. The attributes of the academician object are the followings:

- Age: age of the academician; is used to check retirement.
- Active: the flag variable indicating whether the academician is retired.
- Aquality: academician-quality, the score of intellectual skill which is randomly assigned in the creation of each academician.
- Afield {}: academician-field, the field list of the academician, each academician has one major field and two minors.
- Publication {}: list of the publications done by the academician.
- Co-authors_people {}: list of academicians with whom was done collaboration at least once.
- Co-authors_times {}: number of co-authorship done with the people in co-authors_people list. These two lists are parallel; the second one keeps the data of frequency in collaboration.
- Citing-authors_people {}: list of academicians who have cited this academician at least once.
- Citing-authors_times {}: number of citations received from the authors in citing-authors_list. Like the co-author lists, citing-authors lists are parallel, and the second one keeps the data of frequency in receiving citation.

5.3.2. Structure of Papers

Papers are the second set of agents (or objects). Since the processes run over academicians, papers are actually auxiliary objects. Important attributes of papers are the followings:

- Year: the year in which the paper is published.
- Pquality: paper quality, it is the average quality of the authors of the paper. It is fixed at the paper creation and does not change during the simulation. This is the counterpart of the intellectual skill measure in academicians.
- Pfield: paper field, it is determined by the major field of its authors.
- Authors {}: authors of the paper.
- References {}: reference list of the paper; keeps only references given to the papers in the simulation.
- Citing-papers {}: list of papers which have cited in this paper.
- Totcit: total number of citations received by this paper.

5.3.3. Process Flow

There are two main procedures of the model, one of which is “setup” and the other one is “start”. “Setup” procedure is the initialization of the model with the initial values which are taken from user through the interface. The interface of the model can be seen in Figure 5.1. In addition to the initial number of papers and academicians, interface enables user to determine some parameters, such as number of subfields and time horizon of the simulation.

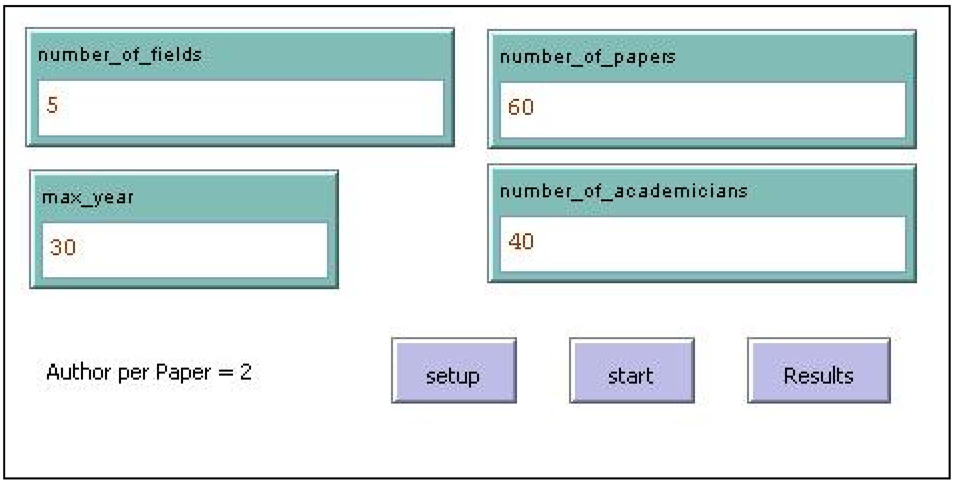

Figure 5.8. Interface of the model

The buttons in the lower right hand side of the Figure 5.1 correspond to different procedures. Actually the model has two major processes in order and an output generator function. Basically “setup” button initializes the model and “start” button runs the simulation. At the end of the simulation, “Results” button generates the proper .csv (comma separated values) files. Before discussing the details of the processes, I would like to mention the assumptions and leading rules of the simulation:

- Every year pairs of scholars form co-author (or collaboration) groups.
- Every year, each co-author group publishes certain number of papers.
- Model tries to simulate a closed field and there are certain numbers of sub-fields.
- The authors and papers are not actually in a closed system, but the model deals with the citations and publications within the simulated world. In other words every citation and co-authorship link in the model is formed within authors and within papers in the model, but these are not necessarily the entire co-authorships or citations.
- Base model initializes with a set of fresh starting scholars. Since the number of references given to existing set of academicians (all fresh starting) should be small, length of reference list starts with a small number and increases and saturates as time passes.
- Academicians predominantly publish in their major field; however, rarely they can also publish in their secondary fields.
- People become retired (inactive) at the age of 65.
- Each year new scholars join the academicians’ pool, with one paper, random quality and age around 25.
- Academicians form co-author groups with the people in the same field, randomly.
- The likelihood of giving citation to a paper is a function of field closeness, quality of the paper, total citation that the paper received, previous co-authorship between the authors of the papers, previous citation relation between two sets of authors and age of the paper.
- The time step and DT are equal to one year.

Under these assumptions the model goes year by year, and constructs the academicians – papers networks. The pseudo code is shown below:

Table 5.2. Pseudo Code

```
//Setup (initialization)
        Create #n of authors
        Create #m of papers
        Set authors' attributes randomly
        Link authors to papers in their field
        Update paper attributes.
//Start (simulation)
For each time step (year) do
{
        Check retirement – deactivate retired ones, but they remain in the pool.
        Add #new authors to the list, randomly assign them age (25+rn (5)), field,
quality...
        Co-authorship partition →#t groups
        For each group do
        {
                Create #p paper
                Update paper attributes
                        Quality
                        Authors
                        References = f(relevance ~ field, quality, total citation
won, previous co-authorship between authors, previous citing between authors, age of pa-
per)
                Update author attributes
        }
}
```

The process flow of the model is quite straightforward. After the simulation is initialized, the rest of it is a loop which repeats itself each year. For each year, coauthor groups are formed and for each coauthor group new papers are added to the paper pool. For each new paper, all papers are sorted with respect to a citation score (which is discussed in the following section) and papers with high citation score get the citation. The number of references is a function of time. After citations are given, the remaining part of the simulation is the data update for both academicians and papers.

<u>5.3.3.1. Citation Formation:</u> One of the key points of the model is the mechanism how papers select the reference list. In order to model the real behavior, I include some different factors and they are modeled as an effect formulation in multiplicative form. By multiplying effect of each factor simulation calculates a score for each paper.

> Score (Likelihood of being cited) = f (field relevance, quality, total citation won, previous co-authorship between authors, previous citing behavior between authors , age of the paper)

Let's say that NP is the new paper, and CP is a candidate paper to get cited. Firstly the quality of the paper is the normal likelihood of the citation. I assume that if the skill level of the author is high, the paper of that author would be high. Then first effect is the field closeness (f). If NP and CP are in the same field a takes the value of 1, if they are in close fields it takes value of 0.5 and 0.1 due to the closeness[11]. If the fields of NP and CP are not related than a takes value of 0, which implies the likelihood of NP's citing CP is zero. Although quality and total citations measure similar phenomenon, I would like to include the effect of total citation in order to separate different effects of inherent quality and gained fame. Second factor is the gained fame which is measured by total citations (tc) of CP. I divide tc by maximum citation (mc) which is the total citation of the most cited paper. Third factor is the previous co-authorship (ca) between the authors, which is operationalized by the portion of the co-authorship between the authors of NP and CP in total co-authorships. Fourth factor is the previous citation (c) between the authors of NP and CP, and it is measured as the percentage in order to have a normalized value. The last but not the least effect is the affect of aging which has an inverted U-shape as time increase. A paper needs to be visible to being cited, and this requires certain amount of time, for this reason, likelihood of being cited increases with time to a certain point. At a point around 10 years the characteristics of aging effect changes, because after a certain amount of time papers start to be obsolete. The shape of the effect function is seen on Figure 5.2.

---

[11] I define the field closeness as such: If there are 5 fields and NP is in the 1st field, than any paper in the same field has an *f* value of 1; any paper in 2nd filed takes an *f* value of 0.5; any paper in 3rd field takes an *f* value of 0.1; and any paper in fields 4 or 5 takes an *f* value of 0. The relation is the same for NP's in other fields: 2→2,3,4; 3→3,4,5; 4→4,5,1; 5→5,1,2.

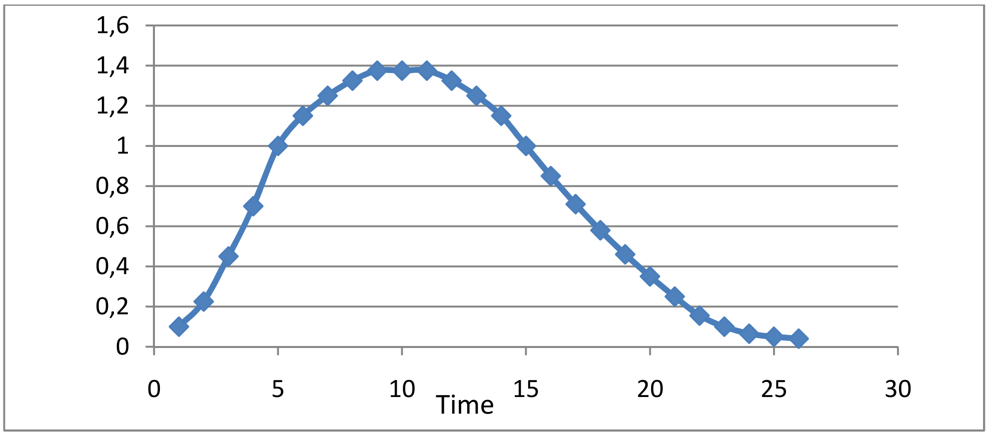

Figure 5.9. Effect of Aging on likelihood of being cited

Hence the likelihood score is calculated as:

$$\text{Score(Likelihood of being cited of CP by NP)} = \text{Quality (cp)} * \text{f} * \frac{\text{tc}}{\text{mc}} * \frac{\text{ca}}{\text{total(ca)}} * \frac{\text{c}}{\text{total(c)}} ** \text{Eff (Aging)}$$

After scores are calculated for each paper in the paper pool, the papers are sorted with respect to the scores. And first N (N = f (time) + Random r (time)) of papers are chosen as the references of the new paper. The shape of the function of the time varying component (f (time)) is shown in Figure 5.3. Since initially there are only a few papers around, each paper can only have a few papers in its reference list in the first years.

Moreover random part also changes with time. When time is less than 5 years, it is a random number between 0 and 2; when time is between 5 and 10 years, it is a random number between 0 and 4; when time is between 10 and 20 years, it is a random number between 0 and 6; and when time is more than 20 years, it is a random number between 0 and 8.

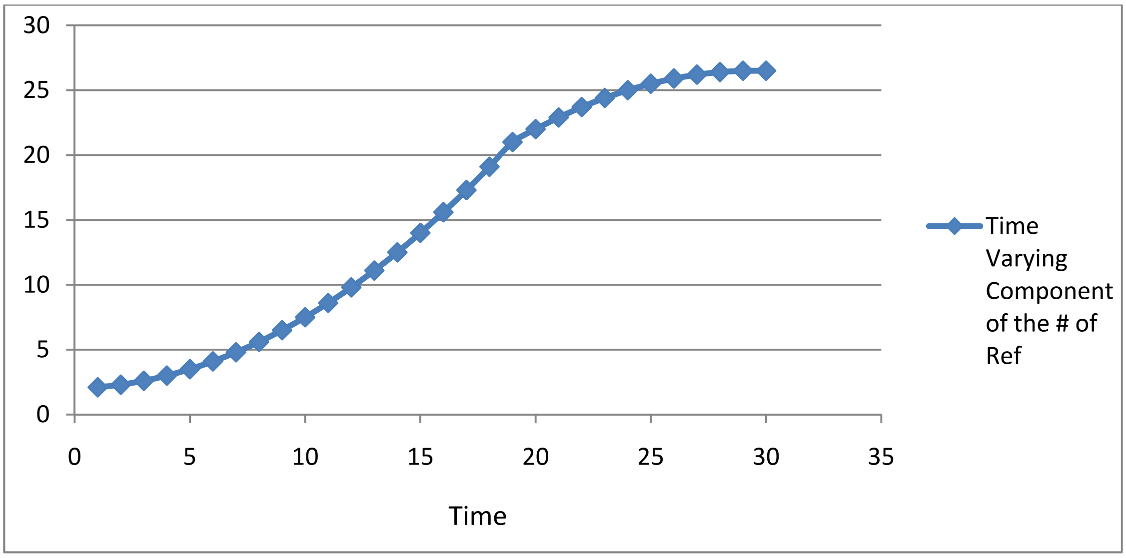

Figure 5.10. Time varying component of the number of references per paper in the base run

5.4. Model Verification and Validation

In order to verify the model, I have debugged the entire model step by step, with low level of initial conditions, and the lists of each period are compared with those of the hand simulation. By this way I have verified that there is a consistency between the conceptual model and the formal one. In other words the model does what it supposes to do.

On the other hand, behavior validation of the model is not very easy due to the scarcity of the real data. In terms of structural validity, I have done direct structural tests while building the model. Additionally I have run the model in some extreme conditions such as zero initial papers and zero initial academicians, in order to make indirect validity tests, and the model behaved in the expected way. I can say that the applied tests have not indicated any invalidity, which implies that the model is a good enough representation of the real situation.

5.5. Output Analysis

The output analysis part shows the network structures of authors. The raw data on citing-authors lists are taken out from Netlogo and then processed with MS Excel to obtain the suitable network data format for Cytoscape. The network figures are drawn by Cytoscape.

5.5.1. Base Model

The base model is the simulation of a field which has five different subfields with 40 fresh starting scholars and 60 papers written by two authors each. Initial authors are created as having 3 authorships[12] on the average. Initial papers have random number of references between 1 and 3, because the model keeps the record of references given to the papers in simulation pool. I assume that on the average 5% of the citations are distributed within the initial paper set of fresh starting authors.

The simulation runs for 30 years and at the end of the simulation, number of authors reaches 205. Among these authors there occur 5654[13] different citation links. There are two ways of explaining this quantity. If all the possible links are considered, which is 42025[14], in this network 13% of all possible connections is realized (the density of the network is 0.13). On the other hand, all possible links are less than 42025, because papers can only give reference to papers in three close fields. By this mechanism the actual number of possible links is about 25215[15] and in the network 24% of possible links are realized. The network structure is shown in Figure 5.4.

---

[12] 60 Papers * 2 Authors per Paper = 120 Authorship
120 Authorship / 40 Auhtors = 3 Authorship per Author
[13] The number of dichotomous links.
[14] 205 * 205
[15] 205* (205/5)*3

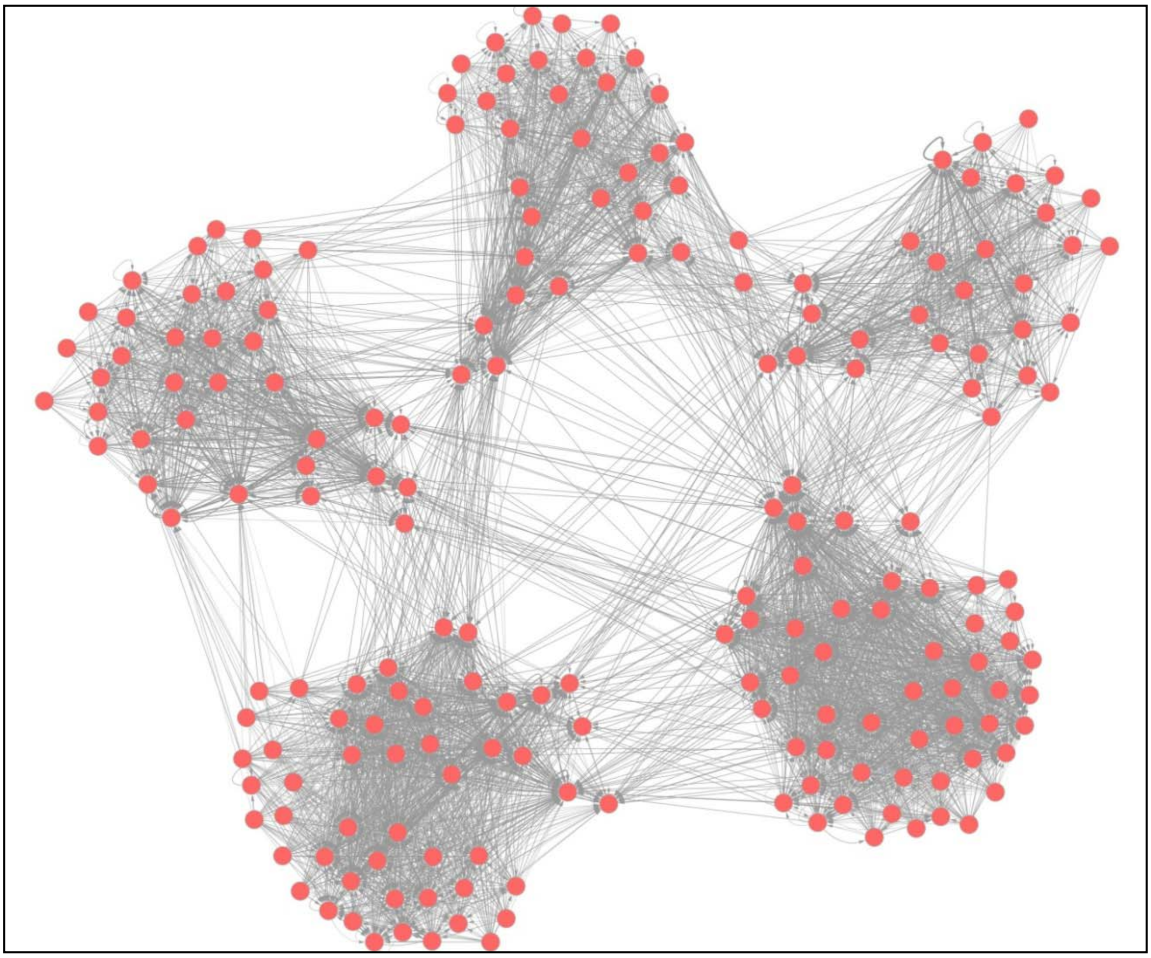

Figure 5.11. Author inter-citation network, base run, nodes=205, links=5654

The author network structure shows 5 clusters each of which is related to a different subfield. This behavior is consistent with the real data shown in Chapter Three. Since the model does not include author properties such as friendship, geographical region or institution, the picture above does not explain possible causes related to those personal attributes. Instead, it is the structure emerged when a group of fresh starting scholars in different subfields are left to a closed supra-field.

5.5.2. Scenario Analyses

I run the model in different initial settings which imply different academic societies. In the base model a society consisting of 40 fresh starting authors are simulated. However, in different scenarios, different academic societies with 40 junior authors, and 40 senior authors, and 40 advanced ones are simulated.

5.5.2.1. Initial Authors Are Junior: The first scenario is the simulation of a field which has five different subfields with 40 junior scholars and 150 papers written by two authors each. Initial authors are created as having 7.5 authorships[16] on the average and this quantity is plausible for assuming the initial authors are junior. Initial papers have random number of references between 3 and 5, because the model keeps the record of references given to the papers in simulation pool. I assume that on the average 16% of the citations are distributed within the initial paper set of junior authors.

The simulation runs for 30 years and at the end of the simulation number of authors reaches 196. Among these authors there occur 5530[17] different citation links. There are two ways of explaining this quantity. If all the possible links are considered, which is 38416[18], in this network 14% of all possible connections is realized (the density of the network is 0.14). On the other hand, the actual number of possible links is about 23049[19] and in the network 24% of possible links are realized. The network structure is shows in Figure 5.5.

The network structure in this scenario is similar to that of base model in terms of clusters due to sub-fields. In the base run within the fields the authors are very closed to each other, however in the scenario analysis the distance between them is increased. This implies that if the initial authors are experienced and have more papers, it is hard for the new authors to enter into the citation cliques.

---

[16] 150 Papers * 2 Authors per Paper = 300 Authorship
300 Authorship / 40 Auhtors = 7.5 Authorship per Author

[17] The number of dichotomous links.

[18] 196*196

[19] 196* (196/5)*3

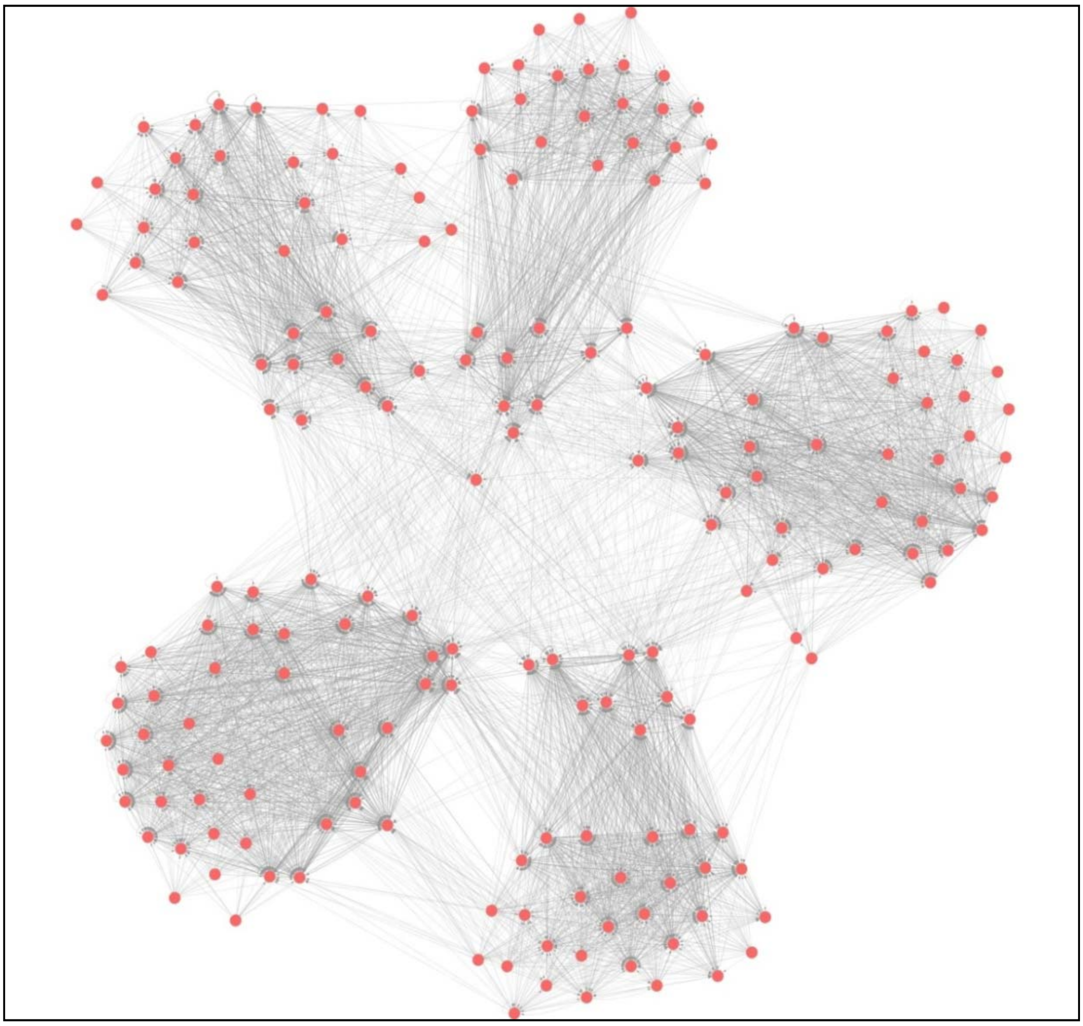

Figure 5.12. Author inter-citation network, initially junior authors and 150 papers,nodes=196, links=5530

5.5.2.2. Initial Authors Are Senior: The second scenario is the simulation of a field which has five different subfields with 40 senior scholars and 300 papers written by two authors each. Initial authors are created as having 15 authorships[20] on the average and this quantity is plausible for assuming the initial authors are senior. Initial papers have random number of references between 5 and 10. I assume that on the average 30% of the citations are distributed within the initial paper set of authors.

[20] 300 Papers * 2 Authors per Paper = 600 Authorship
600 Authorship / 40 Auhtors = 15 Authorship per Author

The simulation runs for 30 years and at the end of the simulation number of authors reaches 199. Among these authors there occur 6477[21] different citation links. Again, there are two ways of explaining this quantity. If all the possible links are considered, which is 39601[22], in this network 16% of all possible connections is realized (the density of the network is 0.16). On the other hand, the actual number of possible links is about 23760[23] and in the network 27% of possible links are realized. The network structure is shows in Figure 5.6.

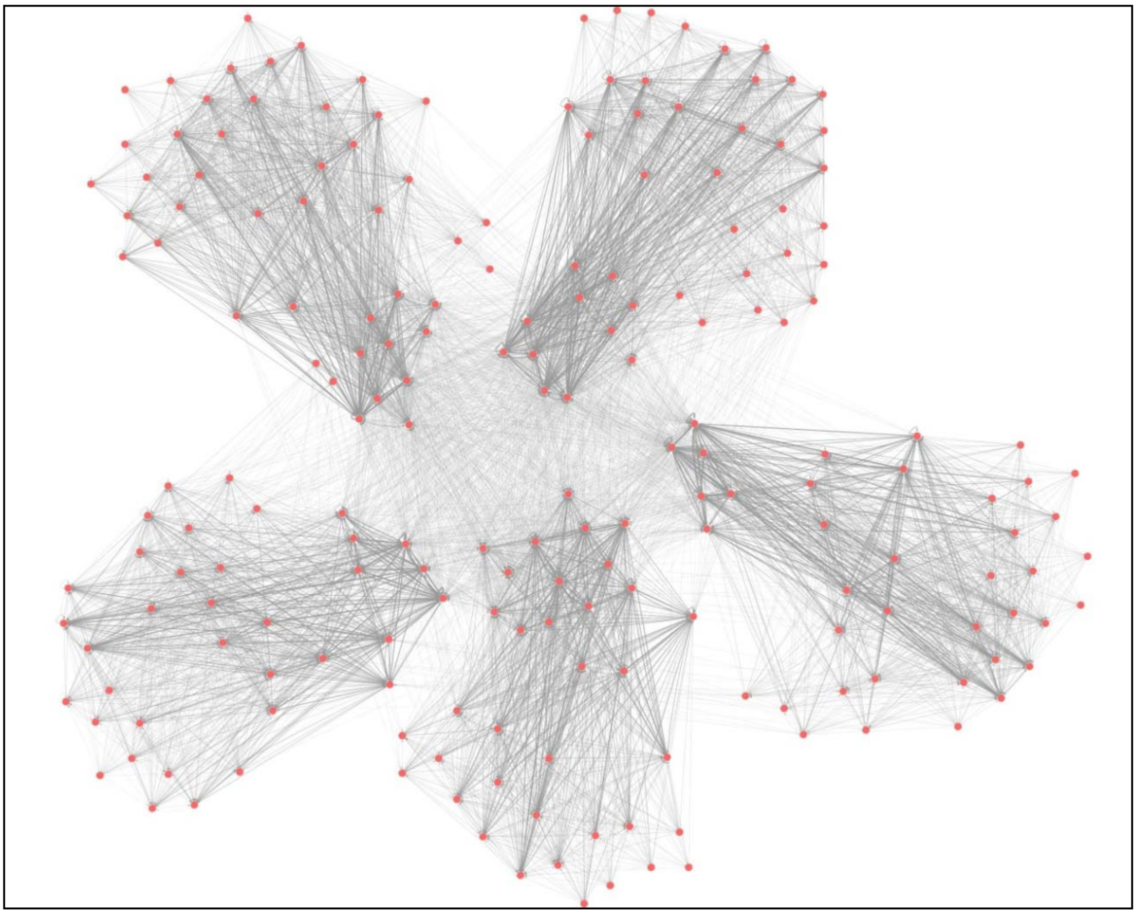

Figure 5.13. Author inter-citation network, initially senior authors and 300 papers, nodes=199, links=6477

Compared to the previous network, scholars who come later, have more difficulty to enter into the core cliques and to find themselves in the peripheries of the network.

[21] The number of dichotomous links.

[22] 199*199

[23] 199* (199/5)*3

5.5.2.3. Initial Authors Are Advanced: The third scenario is the simulation of a field which has five different subfields with 40 advanced authors and 1000 papers written by two authors each. Initial authors are created as having 50 authorships[24] on the average and this quantity is plausible for assuming the initial authors are senior. Initial papers have random number of references between 10 and 20, because I assume that on the average 60% of the citations are distributed within the initial paper set of tenured authors.

The simulation runs for 30 years and at the end of the simulation number of authors reach to 216. Among these authors there occur 5728[25] different citation links. If all the possible links are considered, which is 46656[26], in this network 12% of all possible connections is realized (the density of the network is 0.12). On the other hand, the actual number of possible links is about 27993[27] and in the network 20% of possible links are realized. The network structure is shows in Figure 5.7.

---

[24] 1000 Papers * 2 Authors per Paper = 2000 Authorship
2000 Authorship / 40 Auhtors = 50 Authorship per Author

[25] The number of dichotomous links.

[26] 216*216

[27] 216* (216/5)*3

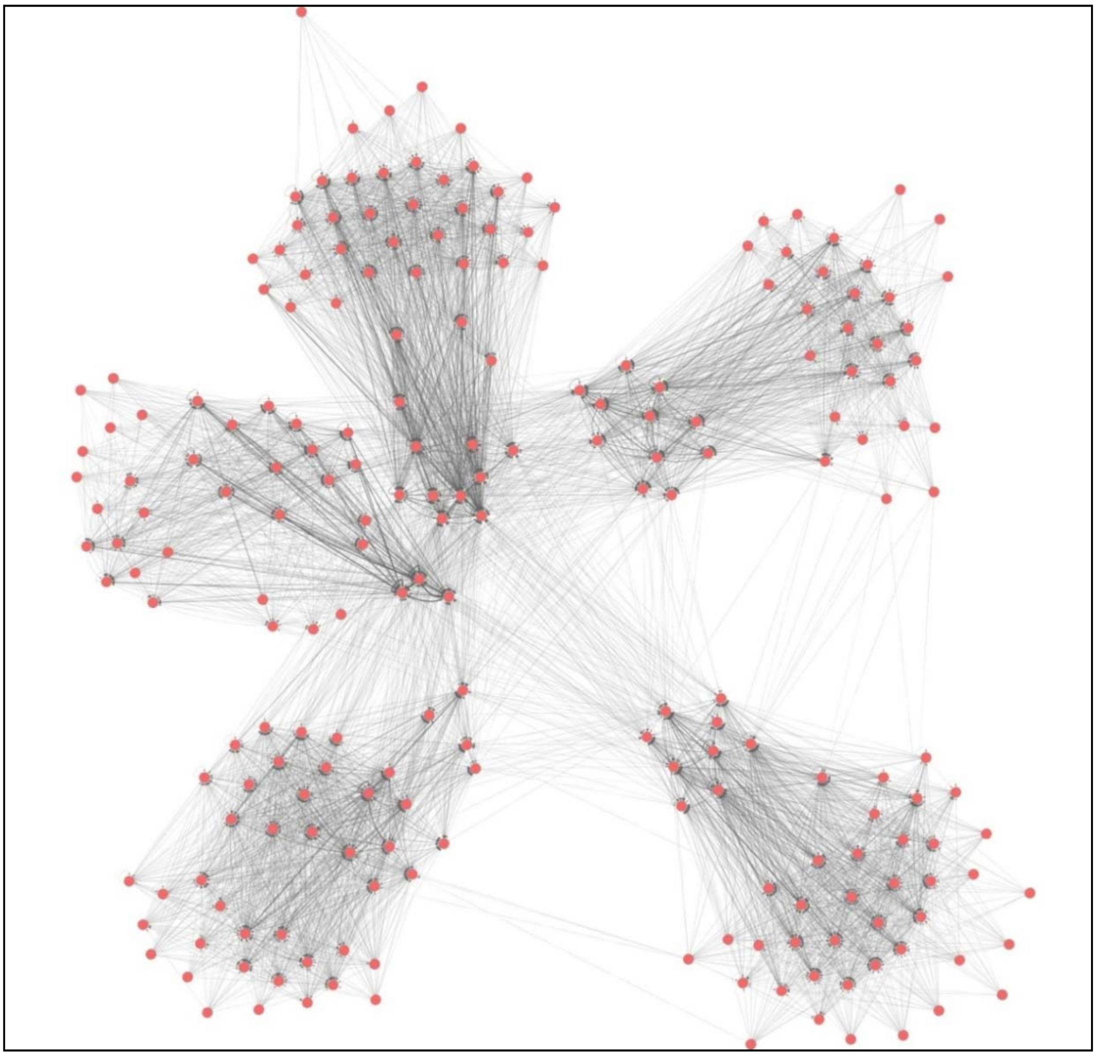

Figure 5.14. Author inter-citation network, initially 1000 papers, nodes=216, links=5728

Number of nodes and links, and density values are similar in base run and scenario analyses, because those values are descriptive measures related to the size of the networks. However, how the initial structure of the field affects the emerging network structure is seen in Figures 5.4, 5.5, 5,6 and 5.7. Additionally, standard deviations of total citations received by authors differ from base run to scenarios. In the base run standard deviation is 13723, in scenario one with junior authors, standard deviation is 13824; in scenario two, standard deviation is 18309; and in scenario three, standard deviation is 37835. As the productivity of the authors in initial settings increases, the variation among authors in eventual network increases. It would be very difficult for junior scholars to enter into the citation clusters of a field with existing 40 productive authors. Standard deviation and average values are listed in Table 5.1.

Table 5.3. Basic Statistics on Citation per Author

	Standard Deviation of TOTAL CITATION per AUTHOR	Average TOTAL CITATION per AUTHOR
Base Run (fresh starting)	13723	10408
Initial scholars are junior	13824	10742
Initial scholars are senior	18309	14551
Initial scholars are advanced	37835	26538

5.6.Chapter Conclusion And Discussion

As the last chapter of this thesis, Chapter Three is dedicated to the agent based simulation model which generates simulated inter-citation data among authors. By this way the aim of the chapter is to see whether some patterns on the network structures emerge or not.

First of all data and the networks generated by the model are consistent with the real data. The cliques emerge in the network structures and it is because of the existence of sub-field divisions. Since the model does not include personal attributes such as affiliation, social links or geographical locations, the model excludes effects of personal attributes.

In the base run, the model mimics an academic society which has a number of fresh starting academicians, each with 3 authorships on the average. In this setting, since new coming scholars are similar to the existing ones in the field, it is easy for them to enter into the center of citation clusters. The eventual network structure is homogenous and distances between nodes are small.

On the other hand, different scenarios mimic different academic societies. In the first scenario analysis, the model simulates an academic society with faculties each of which has 7.5 authorships on the average. In this initial setting, the difference between new coming scholars and the existing ones increases and it becomes harder for the new comers to enter into the center of the citation clusters. The eventual network structure is less homogenous and distances between nodes are larger, compared to the result of the base model.

In the second scenario, the model aims to mimic the network evolution of a field which has authors with 15 authorships on the average. This model starts with scholars in higher caliber and this initial setting of the field results in a more heterogeneous network structure. In the last scenario, the model simulates a field which is full of advanced scholars initially. In this setting, new coming scholars cannot enter into the core of the citation clusters, and the radius of the network increases.

All in all, the model shows that initial conditions of a field are very decisive on the evolution and the eventual network structure of the field.

6. CONCLUSION

The conventional observation that 10% of the journals receive 90% of the citations, is a striking signal of a problem in science policy. The evolution of scientometrics which uses numbers of publications and citations as a measure, should change its direction and find new policies in order to address this problem. For such a paradigm shift and in order the grasp the roots of the problem, a systemic analysis of the situation is needed. This thesis is an initial effort of such an analysis, showing different aspects of the problem and suggesting different ways of studying it.

First way of studying the problem is empirical data analysis which is provided in Chapter Three. A general picture of the historical trends and network structures are shown in two exemplary fields, namely system dynamics and operations research. In terms of historical trends, it is shown that in both fields collaboration increases over time, although the rate of increase is much lower than that in natural sciences. Moreover, historical data on co-citation networks show the different characteristics and different levels of activity in system dynamics and operations research fields. Co-authorship networks in system dynamics illustrate the collaboration teams which can be matched to certain institutions very easily. Lastly but most importantly, inter-citation networks in both fields show the different characteristics of the system dynamics and operations research fields. The former one is a compact and highly connected field, whereas the latter one has different clusters. Further investigations on whether there is an effect of affiliation or geographical position on authorship and citations fail to prove any connection. It does not mean that there is no effect, but it was not possible to see it in the aggregate level data.

The second way is focusing on the aggregate level behavior dynamics. Chapter Four examines the behaviors of researchers of a department in response to dynamics of publication and citation pressures, by a system dynamics model including researchers in a department, their publications, citations and the factors such as reputation, quality, pressures on researchers and their skill levels. The positive feedback mechanism between reputation-citation is very strong and the balancing loops lead by citation pressure and publication pressures are the core underlying mechanisms behind the complex situation. While publication pressure causes producing more papers (in low quality) in shorter times, citation pressure pushes academicians to write high quality papers (in fewer numbers) in longer times. Opposite effects cause the model to reach equilibrium after some oscillations.

Another way of studying the problem is focusing on individual agents and their interactions. This is done in the last chapter by an agent based model. By this way, structural factors are analyzed on the simulated author inter-citation data and it is seen that the initial setting of a field is quite decisive on the evolution of network structures. In a field which has less productive scholars initially, new coming authors can easily receive citation and this results in a more homogenous and highly connected inter-citation network. On the other hand it is not very easy to receive citation for new coming scholars in fields which are full of senior, productive authors. In these kinds of fields the eventual network structure is more heterogeneous and loosely connected.

To conclude, the aim of this thesis is opening a path in studying the problematic situation of science policies which heavily rely on numbers of citations and publications. Future research is very promising in all three directions. In terms of further descriptive data analysis, data of different networks, such as friendship, institutional links, can be collected and their relationship to inter-citation or co-authorship networks can be studied. In terms of modeling, more effort is needed to build richer and empirically validated models in both aggregate modeling and micro modeling approaches. The agent based software used in this thesis took prohibitively long times to analyze more than two hundred authors or a few thousand publications. Stronger software is needed to analyze such large data sets.

APPENDIX A

EQUATIONS OF SYSTEM DYNAMICS MODEL

```
ExtC(t) = ExtC(t - dt) + (ExtC_R) * dt
INIT ExtC = 8000
INFLOWS:
ExtC_R = PubPaper*ExtC\P*Eff_Quality_on_Citation*Eff_Rep_on_ExtC
IntC(t) = IntC(t - dt) + (IntC_R) * dt
INIT IntC = 650
INFLOWS:
IntC_R = PubPaper*Eff_Quality_on_Citation*IntC\Pap\t*MatureR
NewBornP(t) = NewBornP(t - dt) + (PublishR - MatureR) * dt
INIT NewBornP = 126
INFLOWS:
PublishR = ResinW/(PublishT*Eff_AccPr_on_PT)
OUTFLOWS:
MatureR = NewBornP/MatureT
Obsolete_P(t) = Obsolete_P(t - dt) + (ObsoleteR) * dt
INIT Obsolete_P = 617
INFLOWS:
ObsoleteR = PubPaper/(Eff_Quality_on_ObsT*ObsoleteT)
PerRep(t) = PerRep(t - dt) + (PerRepChange) * dt
INIT PerRep = 100
INFLOWS:
PerRepChange = (ActRep-PerRep)/Rep_AT
PubPaper(t) = PubPaper(t - dt) + (MatureR - ObsoleteR) * dt
INIT PubPaper = 819
INFLOWS:
MatureR = NewBornP/MatureT
OUTFLOWS:
ObsoleteR = PubPaper/(Eff_Quality_on_ObsT*ObsoleteT)
ResinW(t) = ResinW(t - dt) + (Research - PublishR - RejectR) * dt
INIT ResinW = 189
INFLOWS:
Research = faculty*Total_Res_Time/Reserved__T\P
OUTFLOWS:
PublishR = ResinW/(PublishT*Eff_AccPr_on_PT)
RejectR = ResinW/RejectT*Reject_portion
Total_Pub_in_3yr(t) = Total_Pub_in_3yr(t - dt) + (pR - Trash) * dt
INIT Total_Pub_in_3yr = 189
 TRANSIT TIME = 12
 INFLOW LIMIT = INF
 CAPACITY = INF
INFLOWS:
pR = PublishR
OUTFLOWS:
Trash = CONVEYOR OUTFLOW
AccPr = AccPr_N*Eff_Rep_on_AP
AccPr_N = 0.5
ActRep = Max_ActRep*(EffExtC_on_Rep+EffIntC_on_Rep)
AvgExtC\P = ExtC/TotP
AvgIntC\P = IntC/TotP
Avg_Cit\P = AvgExtC\P+AvgIntC\P
```

```
Avg_Pub\F\y = Total_Pub_in_3yr/(3*faculty)
EffExtC_on_Rep = GRAPH(AvgExtC\P/ExtC_Gavg)
(0.00, 0.208), (0.2, 0.224), (0.4, 0.252), (0.6, 0.28), (0.8, 0.32), (1.00, 0.4), (1.20, 0.528), (1.40, 0.66), (1.60, 0.74), (1.80, 0.776), (2.00, 0.8)
EffIntC_on_Rep = GRAPH(AvgIntC\P/IntC_Gavg)
(0.00, 0.05), (0.2, 0.052), (0.4, 0.056), (0.6, 0.064), (0.8, 0.077), (1.00, 0.1), (1.20, 0.13), (1.40, 0.172), (1.60, 0.191), (1.80, 0.199), (2.00, 0.2)
EffPubPressure_on_TotResT = GRAPH(Avg_Pub\F\y/Gavg_Pub\F\y)
(0.00, 1.50), (0.1, 1.48), (0.2, 1.44), (0.3, 1.38), (0.4, 1.22), (0.5, 1.12), (0.6, 1.05), (0.7, 1.01), (0.8, 1.01), (0.9, 1.00), (1, 1.00)
Eff_AccPr_on_PT = GRAPH(AccPr/AccPr_N)
(0.00, 1.23), (0.25, 1.21), (0.5, 1.20), (0.75, 1.13), (1.00, 1.00), (1.25, 0.893), (1.50, 0.797)
Eff_CitPressure_on_ResT\P = GRAPH(Avg_Cit\P/Gavg_Cit\P)
(0.00, 1.50), (0.1, 1.50), (0.2, 1.48), (0.3, 1.38), (0.4, 1.24), (0.5, 1.15), (0.6, 1.08), (0.7, 1.04), (0.8, 1.02), (0.9, 1.01), (1, 1.00)
Eff_PubPressure_on_ResT\P = GRAPH(Avg_Pub\F\y/Gavg_Pub\F\y)
(0.00, 0.1), (0.1, 0.118), (0.2, 0.145), (0.3, 0.199), (0.4, 0.307), (0.5, 0.469), (0.6, 0.753), (0.7, 0.915), (0.8, 0.973), (0.9, 0.991), (1, 1.00)
Eff_Quality_on_Citation = GRAPH(Quality)
(0.00, 0.209), (0.333, 0.371), (0.667, 0.614), (1.00, 1.00), (1.33, 1.28), (1.67, 1.53), (2.00, 1.71), (2.33, 1.86), (2.67, 1.96), (3.00, 2.00)
Eff_Quality_on_ObsT = GRAPH(Quality)
(0.00, 0.7), (0.25, 0.719), (0.5, 0.765), (0.75, 0.849), (1.00, 1.00), (1.25, 1.19), (1.50, 1.42), (1.75, 1.75), (2.00, 1.89), (2.25, 1.95), (2.50, 1.99), (2.75, 1.99), (3.00, 2.00)
Eff_Rep_on_ExtC = GRAPH(PerRep/Max_ActRep)
(0.00, 0.101), (0.1, 0.204), (0.2, 0.35), (0.3, 0.52), (0.4, 0.74), (0.5, 1.14), (0.6, 1.55), (0.7, 1.78), (0.8, 1.92), (0.9, 1.97), (1, 1.99)
Eff_Rep_on_AP = GRAPH(PerRep/Max_ActRep)
(0.00, 0.01), (0.1, 0.15), (0.2, 0.29), (0.3, 0.46), (0.4, 0.68), (0.5, 1.00), (0.6, 1.52), (0.7, 1.81), (0.8, 1.92), (0.9, 1.96), (1, 1.99)
Eff_ResTime\P_on_Quality = GRAPH(Reserved__T\P/Required_T\P)
(0.25, 0.505), (0.5, 0.54), (0.75, 0.66), (1.00, 1.00), (1.25, 1.27), (1.50, 1.40), (1.75, 1.46), (2.00, 1.49)
Eff_Skill_on_Quality = GRAPH(Skill_Level)
(10.0, 0.2), (20.0, 0.335), (30.0, 0.515), (40.0, 0.731), (50.0, 1.00), (60.0, 1.37), (70.0, 1.65), (80.0, 1.83), (90.0, 1.94), (100, 2.00)
ExtC\P = GRAPH(TIME)
(0.00, 0.2), (100, 0.24), (200, 0.28), (300, 0.32), (400, 0.36), (500, 0.4), (600, 0.44), (700, 0.48), (800, 0.52), (900, 0.56), (1000, 0.6)
ExtC_Gavg = Gavg_Cit\P*5/6
faculty = 105
Gavg_Cit\P = GRAPH(TIME)
(0.00, 9.60), (100, 11.6), (200, 13.6), (300, 15.6), (400, 17.6), (500, 19.6), (600, 21.6), (700, 23.6), (800, 25.6), (900, 27.6), (1000, 29.6)
Gavg_Pub\F\y = 1.6
IntC\Pap\t = 0.0023
IntC_Gavg = Gavg_Cit\P/6
MatureT = 8
Max_ActRep = 100
ObsoleteT = 52
PublishT = 12
Quality = Eff_ResTime\P_on_Quality*Eff_Skill_on_Quality
RejectT = 12
Reject_portion = 0.15
Rep_AT = 16
Required_T\P = 2
```

```
Reserved__T\P = Re-
quired_T\P*Eff_PubPressure_on_ResT\P*Eff_CitPressure_on_ResT\P
Skill_Level = 20
TotalResTimeN = 0.3
Total_Res_Time = TotalResTimeN*EffPubPressure_on_TotResT
TotP = Obsolete_P+PubPaper+NewBornP
```

APPENDIX B
SOME EXAMPLES OF EXTREME CONDITION RUNS OF SYSTEM DYNAMICS MODEL

Run with only 1 faculty number:

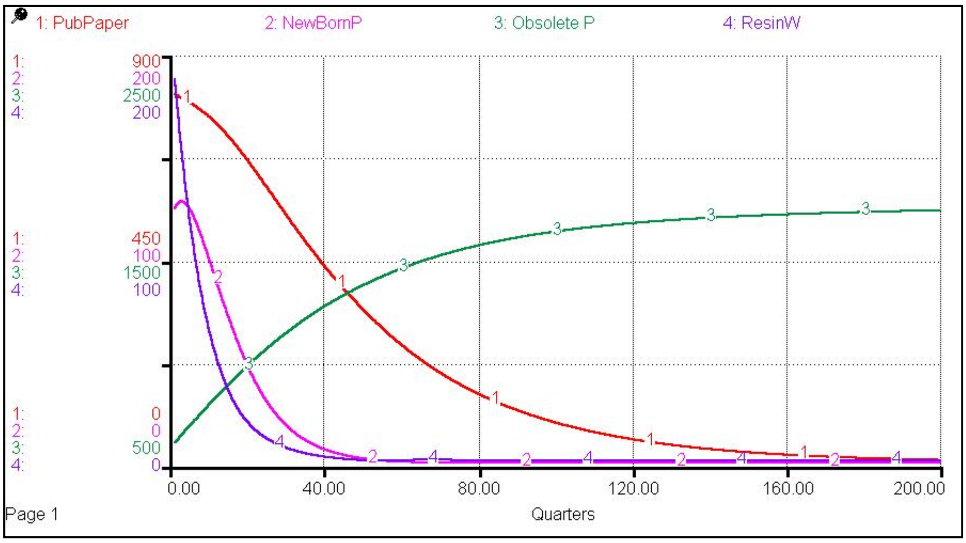

Figure 7.15. Paper Stocks in extreme condition 1 - With 1 faculty member

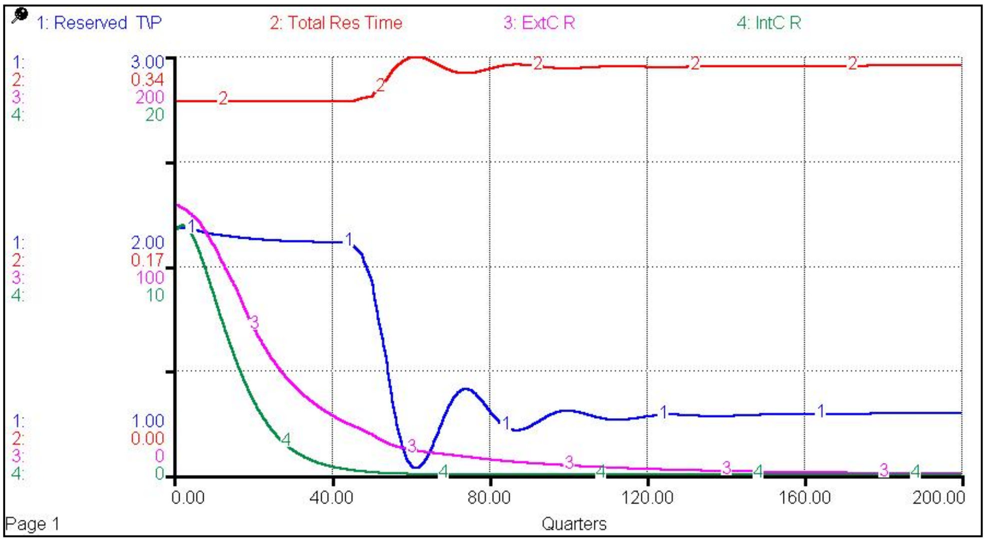

Figure 7.2. Reserved time and citation flows in extreme condition 1 - With 1 faculty member

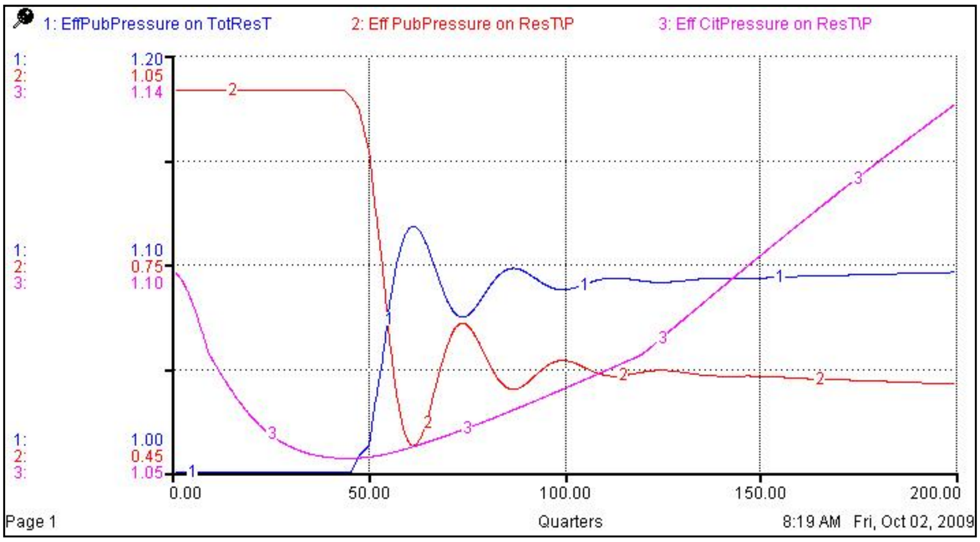

Figure 7.3. Pressures in extreme condition 1 - With 1 faculty member

Skill level is set 100 – absolutely perfect faculties.

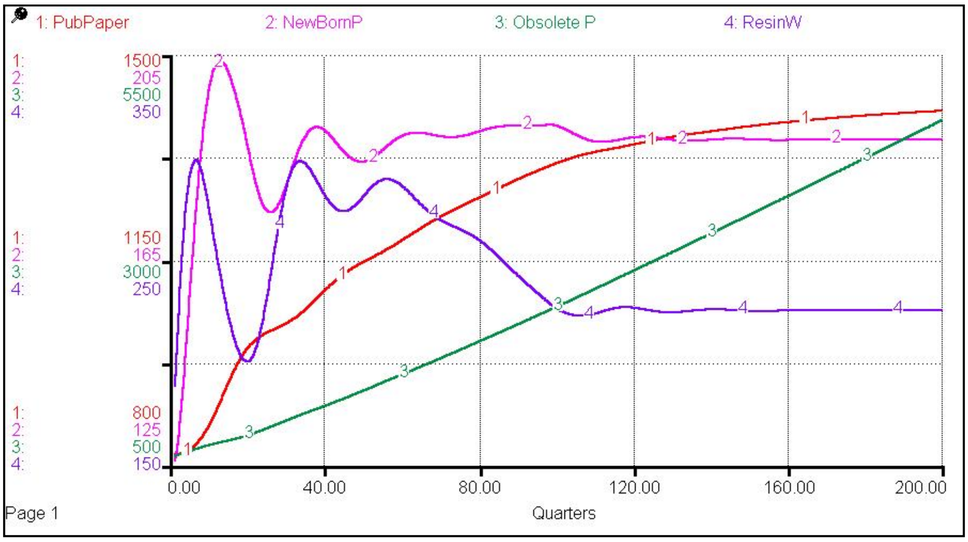

Figure 7.4. Paper Stocks in extreme condition 2 - With perfect faculty members

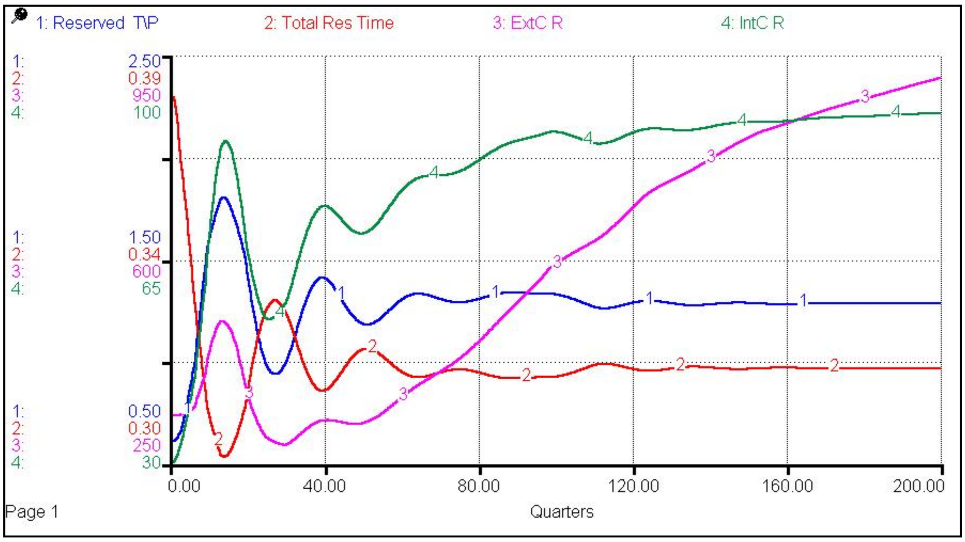

Figure 7.5. Reserved time and citation flows in extreme condition 2 - With perfect faculty members

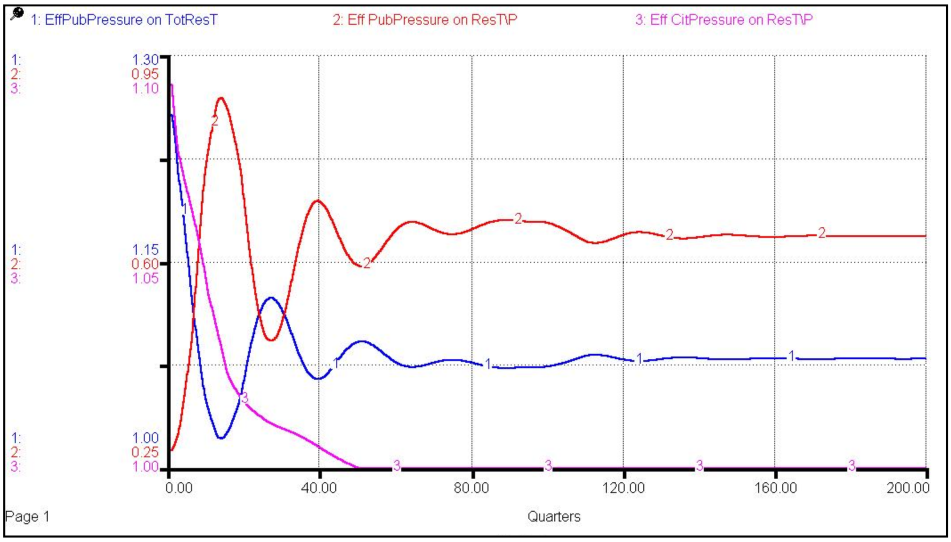

Figure 7.6. Pressures in extreme condition 2 - With perfect faculty members

APPENDIX C

CODE OF AGENT BASED MODEL

```
breed [academicians academician]
breed [papers paper]
globals [coauthorslist plist time scorelist maxcite tpaper taut1 taut2 rpa-
per citeda grnumref rndref EffAge]


academicians-own
[
 age
 active
 afield
 publications
 coauthors
 citingauthors
 aquality
 coauthors_people
 coauthors_times
 citingauthors_people
 citingauthors_times
]

papers-own
[
 authors
 year
 pfield
 pquality
 references
 citingpapers
 totcit
 score
 rw
]
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
to setup
  let i 0
  set coauthorslist []
  clear-all
  create-papers number_of_papers
  create-academicians number_of_academicians
  set-default-shape academicians "person"
  set-default-shape papers "dot"
  set grnumref 2
  set EffAge [0 0.1 0.225 0.45 0.7 1 1.15 1.25 1.325 1.375 1.375 1.375 1.325
1.25 1.15 1 0.85 0.71 0.58 0.46 0.35 0.25 0.155 0.1 0.065 0.05 0.04 0.03 0.02
0.01]
  ask academicians
   [setup-academicians]
  ask papers
   [setup-papers]
  setup-patches
  set plist papers
```

```
        ask academicians
        [set publications [who] of papers with [member? [who] of myself authors =
true] ]

       ask papers
        [
          let nofref 1 + random 2
           set references [who] of n-of nofref papers with [pfield = [pfield] of
myself and who != [who] of myself] ]

       ask papers
        [
        set citingpapers [who] of papers with [member? [who] of myself references
= true]
        set totcit length citingpapers
        ]

       ask academicians
        [set coauthors [item 0 filter [? != [who] of myself] authors] of papers
with [member? [who] of myself authors = true] ]

       ask academicians
        [
        set i 0
        while [i < length publications ]
         [
         let cpc [citingpapers] of papers with [who = item i [publications] of my-
self]
         let cpac []
         let j 0
         while [j < length cpc ]
           [
           let k 0
           while [k < length item j cpc ]
             [
             set cpac fput item k item j cpc cpac
             set k k + 1
             ]
           set j j + 1
           ]
         set j 0
         while [j < length cpac ]
           [
           let k 0
           while [k < length item 0 [authors] of papers with [who = item j cpac] ]
             [
             set citingauthors sentence citingauthors (item k item 0 [authors] of
papers with [who = item j cpac])
             set k k + 1
             ]

           set j j + 1
           ]
         set i i + 1
         ]
        ]

       ask academicians
        [
        set coauthors_people remove-duplicates coauthors
        set coauthors_times n-values length coauthors_people [0]
        set i 0
        while [i < length coauthors_people ]
         [
         let j 0
         while [j < length coauthors ]
```

```
          [
          if item j coauthors = item i coauthors_people [set coauthors_times re-
place-item i coauthors_times (item i coauthors_times + 1)]
          set j j + 1
          ]
        set i i + 1
        ]
       ]

      ask academicians
       [
       set citingauthors_people remove-duplicates citingauthors
       set citingauthors_times n-values length citingauthors_people [0]
       set i 0
       while [i < length citingauthors_people ]
        [
        let j 0
        while [j < length citingauthors ]
          [
          if item j citingauthors = item i citingauthors_people [set citingau-
thors_times replace-item i citingauthors_times (item i citingauthors_times + 1)]
          set j j + 1
          ]
        set i i + 1
        ]
       ]

     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
     to setup-papers
         let i 1
         hide-turtle
         set pquality random-normal 50 (50 / 3)
         set year 0
         set pfield random number_of_fields
         set references []
         set citingpapers []
         set authors [who] of n-of authorperpaper academicians with [item 0
afield = [pfield] of myself]
         set rw [0 1 1 0 0 0]
     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
     to setup-academicians
         let i 1
         set color one-of base-colors
         set SIZE 0.5
         setxy random-xcor random-ycor
         set aquality random-normal 50 (50 / 3)
         set age 25 + random 40
         set active 1
         set afield [1 1 1]
         set afield replace-item 0 afield random number_of_fields
         set afield replace-item 1 afield (random number_of_fields)
         set afield replace-item 2 afield (random number_of_fields)
         set coauthors []
         set citingauthors []
         set publications []
         set coauthors_people []
         set coauthors_times []
         set citingauthors_people []
         set citingauthors_times []
     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
```

```
to setup-patches
   ask patches [set pcolor black]
end
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

to start
  set time 0
  loop
  [
    ifelse time < 20 [set grnumref grnumref + (time / 10)] [set grnumref grnumref + ((max_year - time) / 10)]
    ifelse time < 5 [ set rndref 2] [ifelse time < 10 [set rndref 4] [ifelse time < 20 [set rndref 6][set rndref 8]]]
    ;show "burak"
    ;show time
    ;show grnumref
    ;show rndref
    if time >= max_year [stop]
    let nofnewacad 4 + random 4
    create-academicians nofnewacad        ;yearly new academician
     [
      setup-academicians
      set age 22 + random 8
     ]
    ask academicians
    [
     set age age + 1
     check_retirement
    ]
    coauthorspartition
    foreach coauthorslist
     [
     let m 0
     while [m < 1] ; yearly new publication per authorgroup
      [
      create-papers 1            ; yearly new publication per authorgroup
       [
        setup-papers
        set authors sentence (item 0 ?) (item 1 ?)
        set taut1 one-of academicians with [who = item 0 ?]
        set taut2 one-of academicians with [who = item 1 ?]
        set year time
        set pfield [item 0 afield] of taut1
        set pquality ([aquality] of taut1 + [aquality] of taut2) / 2
        set tpaper who
        ;set authors remove-duplicates authors
        selectreferences
        updatecitingpapers
       ]
       ;show "burak"
       updateacademicions;
       set m m + 1
       ]
     ]
    set time time + 1
  ]
end
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to Results
  ;file-delete "output.csv"
  file-open "poutput.csv"
  ;show "burak"
  ask papers
  [file-show references]
```

```
    file-close
    file-open "aoutput.csv"
    ask academicians
     [file-show citingauthors_people
      file-show citingauthors_times
     ]
    file-close
    file-open "co-aut.csv"
    ask academicians
    [
     file-show coauthors_people
     file-show coauthors_times
    ]
    file-close
  end
  ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
  ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
  to check_retirement
    if age > 65 [set active 0]
  end
  ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
  to coauthorspartition
    let tempacadlist academicians with [active = 1]
    set coauthorslist []
    while [any? tempacadlist = true]
      [
      let aut1 one-of tempacadlist
      let aut2 one-of tempacadlist
      let candlist tempacadlist with [who != [who] of aut1 and item 0 afield =
[item 0 afield] of aut1 and active = 1]
      ifelse any? candlist = true [set aut2 one-of candlist] [set aut2 aut1]
      set coauthorslist lput sentence ([who] of aut1) ([who] of aut2) coau-
thorslist
      set tempacadlist tempacadlist with [who != [who] of aut1 and who !=
[who] of aut2]
      ]
  end
  ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
  to selectreferences
    set scorelist n-values ((length [who] of papers) - 1) [0]
    set plist [who] of papers with [who != [who] of myself]
    set plist sort plist
    let i 0
    set maxcite max [totcit] of papers
    ;show "burak"
    ;show who
    while [i < length plist]
      [
      let cpaper one-of papers with [who = item i plist]
      set rw replace-item 0 rw 0              ;field
      set rw replace-item 1 rw 0              ;pquality
      set rw replace-item 2 rw 0.01           ;total citation
      set rw replace-item 3 rw 0              ;co-authorship
      set rw replace-item 4 rw 0              ;previous citation
      set rw replace-item 5 rw (item (time - [year] of cpaper) EffAge)

      ;show cpaper
      ;show [year] of cpaper
      ;show rw
      if pfield = [pfield] of cpaper
         [set rw replace-item 0 rw 1]
      if (pfield + 1) mod number_of_fields = [pfield] of cpaper
```

```
            [set rw replace-item 0 rw 0.5]
         if (pfield + 2) mod number_of_fields = [pfield] of cpaper
            [set rw replace-item 0 rw 0.3]
         set rw replace-item 1 rw [pquality] of cpaper
         if [totcit] of cpaper != 0
           [set rw replace-item 2 rw ([totcit] of cpaper / maxcite)]
         let aut1 academicians with [who = item 0 [authors] of myself]
         let aut2 academicians with [who = item 1 [authors] of myself ]
         let aut3 academicians with [who = item 0 [authors] of cpaper ]
         let aut4 academicians with [who = item 1 [authors] of cpaper ]
         let ctpacoauth sentence [coauthors_people] of aut1 [coauthors_people] of
aut2
         let ctpacoauthtimes sentence [coauthors_times] of aut1 [coauthors_times]
of aut2
         let tpacoauth []
         let tpacoauthtimes []
         let l 0
         while[l < length item 0 ctpacoauth]
           [
             set tpacoauth lput item l (item 0 ctpacoauth) tpacoauth
             set tpacoauthtimes lput item l (item 0 ctpacoauthtimes) tpacoauth-
times
             set l l + 1
           ]
         set l 0
         while[l < length item 1 ctpacoauth]
           [
             set tpacoauth lput item l (item 1 ctpacoauth) tpacoauth
             set tpacoauthtimes lput item l (item 1 ctpacoauthtimes) tpacoauth-
times
             set l l + 1
           ]
         let pacoauth remove-duplicates tpacoauth
         let pacoauthtimes n-values length pacoauth [0]
         let j 0
         let k 0
         while [j < length pacoauth]
           [
           set k 0
           while [k < length tpacoauth]
             [
             if item j pacoauth = item k tpacoauth [set pacoauthtimes replace-
item j pacoauthtimes ((item j pacoauthtimes) + (item k tpacoauthtimes))]
             set k k + 1
            ]
            if (item j pacoauth = one-of [who] of aut1) or (item j pacoauth =
one-of [who] of aut2) [set pacoauthtimes replace-item j pacoauthtimes (2 * (item
j pacoauthtimes))]
            set j j + 1
         ]

         set k 0
         while [k < length pacoauth]
             [
             if (item k pacoauth = one-of [who] of aut3 or item k pacoauth = one-
of [who] of aut4)
               [set rw replace-item 3 rw (item 3 rw + item k pacoauthtimes)]
             set k k + 1
             ]
         ifelse item 3 rw = 0
           [set rw replace-item 3 rw 0.01]
           [set rw replace-item 3 rw (item 3 rw / sum pacoauthtimes)]


         let ctpaciting sentence ([citingauthors_people] of aut1) ([citingau-
thors_people] of aut2)
```

```
          let ctpacitingtimes sentence ([citingauthors_times] of aut1) ([citingau-
thors_times] of aut2)
          let tpaciting []
          let tpacitingtimes []
          set l 0
          while[l < length item 0 ctpaciting]
            [
              set tpaciting lput item l (item 0 ctpaciting) tpaciting
              set tpacitingtimes lput item l (item 0 ctpacitingtimes) tpaciting-
times
              set l l + 1
            ]
          set l 0
          while[l < length item 1 ctpaciting]
            [
              set tpaciting lput item l (item 1 ctpaciting) tpaciting
              set tpacitingtimes lput item l (item 1 ctpacitingtimes) tpaciting-
times
              set l l + 1
            ]

          let paciting remove-duplicates tpaciting
          let pacitingtimes n-values length paciting [0]

          set j 0
          set k 0
          while [j < length paciting]
            [
            set k 0
            while [k < length tpaciting]
              [
              if item j paciting = item k tpaciting [set pacitingtimes replace-
item j pacitingtimes ((item j pacitingtimes) + (item k tpacitingtimes))]
              set k k + 1
             ]
             if item j paciting = one-of [who] of aut1 or item j paciting = one-of
[who] of aut2 [set pacitingtimes replace-item j pacitingtimes (2 * (item j paci-
tingtimes))]
             set j j + 1
          ]

          set k 0
          while [k < length paciting]
              [
              if (item k paciting = one-of [who] of aut3 or item k paciting = one-
of [who] of aut4)
                 [set rw replace-item 4 rw (item 4 rw + item k pacitingtimes)]
              set k k + 1
              ]

         ifelse item 4 rw = 0
            [set rw replace-item 4 rw 0.01]
            [set rw replace-item 4 rw (item 4 rw / sum pacitingtimes)]

          set scorelist replace-item i scorelist (item 0 rw * item 1 rw * item 2
rw * item 3 rw * item 4 rw * item 5 rw)
          set i i + 1
          ]


      let nofref grnumref + random rndref
      set i 0
      while [i < nofref]
        [
        let tmax 0
        let tmaxindex 0
```

```
        let k 0
        while [k < length plist]
          [
            if item k scorelist >= tmax
              [
              set tmax item k scorelist
              set tmaxindex k
              ]
            set k k + 1
          ]
        set references lput (item tmaxindex plist) references
        set scorelist remove-item tmaxindex scorelist
        set plist remove-item tmaxindex plist

        set i i + 1
        ]
       set references sort references

     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
     to updatecitingpapers
      ask papers
       [
         set citingpapers [who] of papers with [member? [who] of myself refer-
ences = true]
         set totcit length citingpapers
       ]

     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;

     to updateacademicions

       updatePubList
       updateCoAuthors
       updateCitingAuthors
     end
     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
     to updatePubList
       ask academicians
        [
         set publications [who] of papers with [member? [who] of myself authors =
true]
        ]
     end

     ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
     to updateCoAuthors

      set rpaper papers with [who = tpaper]
       if (taut1 != taut2)
       [
        let j 0
        let fl 0
        while [j < length [coauthors_people] of taut1 and fl = 0]
          [
           if (item j [coauthors_people] of taut1 = [who] of taut2)
            [
             set [coauthors_times] of taut1 replace-item j [coauthors_times] of
taut1 (item j [coauthors_times] of taut1 + 1)
             set fl 1
```

```
          ]
         set j j + 1
        ]
     if (fl = 0)
        [set [coauthors_people] of taut1 lput [who] of taut2 [coauthors_people]
of taut1
         set [coauthors_times] of taut1 lput 1 [coauthors_times] of taut1
         set [coauthors_people] of taut2 lput [who] of taut1 [coauthors_people]
of taut2
         set [coauthors_times] of taut2 lput 1 [coauthors_times] of taut2
        ]
    set j  0
    while [j < length [coauthors_people] of taut2 and fl = 1]
      [
       if (item j [coauthors_people] of taut2 = [who] of taut1)
        [
         set [coauthors_times] of taut2 replace-item j [coauthors_times] of
taut2 (item j [coauthors_times] of taut2 + 1)
         set fl 0
        ]
       set j j + 1
      ]
    ]

  end
  ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
  to updateCitingAuthors
   set rpaper one-of papers with [who = tpaper]
   let reflist [references] of rpaper
   let z 0
   while[z < length reflist]
   [
    let t 0
    let citedp one-of papers with [who = item z reflist]
    while [t < length [authors] of citedp]
    [
     let fl1 0
     let fl2 0
     let j 0
     set citeda one-of academicians with [who = item t [authors] of citedp]
     while [j < length [citingauthors_people] of citeda and (fl1 = 0 or fl2 =
0)]
        [
         if item j [citingauthors_people] of citeda = [who] of taut1 and fl1 =
0
          [
           set [citingauthors_times] of citeda replace-item j [citingau-
thors_times] of citeda (item j [citingauthors_times] of citeda + 1)
           set fl1 1
          ]
         if item j [citingauthors_people] of citeda = [who] of taut2 and fl2 =
0 and taut1 != taut2
          [
           set [citingauthors_times] of citeda replace-item j [citingau-
thors_times] of citeda (item j [citingauthors_times] of citeda + 1)
           set fl2 1
          ]
         set j j + 1
        ]
       if fl1 = 0              ;taut1 yok
          [
           set [citingauthors_people] of citeda lput [who] of taut1 [citingau-
thors_people] of citeda
           set [citingauthors_times] of citeda lput 1 [citingauthors_times] of
citeda
```

```
            ]
          if fl2 = 0  and taut1 != taut2          ;taut2 yok
           [
            set [citingauthors_people] of citeda lput [who] of taut2 [citingau-
thors_people] of citeda
            set [citingauthors_times] of citeda lput 1 [citingauthors_times] of
citeda
           ]

         set t t + 1
        ]
        set z z + 1
      ]
      end
      ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;
```

REFERENCES

Bar-Ilan, J. (2008), "Informatics at the Beginning of the 21s Century - A Review". *Journal of Informetrics* , 2 (1), pp. 1-52.

Barlas, Y. (1996), "Formal Aspects of Model Validity and Validation in System Dynamics", *System Dynamics Review* , 12 (3), pp. 183-210.

Borner, K., Maru, J. T., & Goldstone, R. L. (2004), "The Simultaneous Evolution of Author and Paper Networks", *Procedings of the National Academy of Science,* 101 (6), pp. 5266-5273.

Britannica Online Encyclopedia. (Retrieved 01 21, 2008, from Britannica OnlinE Encyclopedia: http://www.britannica.com/eb/article-9066286/science)

Chen, C. (2006), "CiteSpace II: Detecting and Visualizing Emerging trends and Transinent Patterns in Scientific Literature", *Journal of the American Society for Information Science and Technology* , 57 (3), pp. 359-377.

Contractor, N. S., Wasserman, S., & Faust, K. (2006), "Testing Multitheoretical, Multilevel Hypotheses about Organizatoinal Networks: An Analytic Framework and Emprical Example", *Acamdemy of Management Review* , 31 (3), pp. 681-703.

Cronin, B. (2001), "Bibliometrics and beyond: Some thoughts on web-based citation analysis", *Journal of Information Science* , 27 (1), pp. 1-7.

Demirel, G. (2007), " Aggregated and Disaggregated Modeling Approaches to Multi Agent Dynamics. *International Conference of System Dynamics Society.* Boston, USA.

DiPrete, T. A., & Eirich, G. M. (2006), "Cumulative Advantage as a Mechanism for Inequality: A review of Theoretical and Emprical Developments", *Annual Review of Sociology* , 32 (1), pp. 271-297.

Eskici, B., Küçük, B., & Güler, N. (2008). "A Dynamic Simulation Model of Academic Publications and Citations", *International Conference of System Dynamics Society,Athens, Greece.*

Furner, J. (2003), "Little book, big book: before and after Little science, big science: a review article, Part I", *Journal of Librarianship and Information Science*, 35 (2), pp. 115-125.

Furner, J. (2003), " Little book, big book: before and after Little science, big science: a review article, Part II", *Journal of Librarianship and Information Science*, 35 (3), pp.189-201.

Garfield, E. (1955), "Citation indexes for science", *Science* , 122 (1), pp. 108-111.

Garfield, E. (1970), "Citation Indexing for Studying Science", *Nature*, 669 (1), pp. 669-671.

Garfield, E. (1967), "*Diagonal Display",* Rossly, VA: ISI-US Air Force.

Garfield, E. (1998), "I had a Dream...about Uncitedness", *The Scientist* , 12 (1), pp. 10-11.

Hamilton, D. P. (1990), "Publishing by -and for?- the Numbers", *Science,* 250(4986),pp.1331-32.

Hamilton, D. (1991), "Research Papers: Who's Uncited Now?", *Science* , 251 (4989), pp 25.

Hood, W. W., & Wilson, C. S. (2001), "The literature of bibliometrics, scientometrics,and informetrics", *Scientometrics* , 52 (2), pp. 291-314.

Küçük, B., Güler, N., & Tamçakır, S. (2007), "*Measurement Analysis and Evaluation of Academic Publications and Citations".* BS Senior Thesis, Bogazici University, Istanbul.

MacRoberts, M. H., & MacRoberts, B. R. (1989), "Problems of Citation Analysis: A Critical Review", *Journal of the American Society for Information Science* , 40 (5), pp. 342-349.

Margolis, J. (1967), "Citation Indexing and Evaluation of Scientific Papers", *Science* , 155 (1), pp. 1213-1219.

Menczer, F. (2004), "Evolution of document networks", *Procedings of the National Academy of Science* , 101 (1), pp. 5261-5265.

Moed, H. F. (2005), "*Citation Analysis in Research Evaluation",* Dordrecht: Springer.

Moed, H. F., Glanzel, W., & Schmoch, U. (2004), "*Handbook of Quantitative Science and Technology Research"* Dordrecht: Kluwer Academic Publishers.

Newman, M. E. (2001), "Scientific Collaboration Networks. I.Network construction and fundamental results", *Physical Review* , 64 (1), pp. 161031/1-8.

Newman, M. E. (2001), "Scientific Collaboration Networks. II.Shortest paths, weighted networks, and centrality", *Physical Review* , 64 (2) , pp. 161032/1-7.

Nicolaisen, J. (2007), " Citation Analysis", *Annual Review of Information Science and Technology* , 41 (1), pp. 609-641.

Otte, E., & Rousseau, R. (2002), "Social Network Analysis: A Powerful Strategy, also for the Information Sciences", *Journal of Information Science* , 28 (6), pp. 441 - 453.

Pendlebury, D. A. (1991), "Science, Citation and Funding", *Science* , 251 (1), pp. 1410 1411.

Price, D. J. (1964), "Ethics of Scientific Publication", *Science,* 144 (1), pp. 655-657.

Price, D. J. (1963), "*Little Science Big Science",* New York: Columbia University Press.

Price, D. J. (1965), "Networks of Scientific Papers", *Science*, 149 (3683), pp. 510-515.

Price, D. J. (1975), "*Science since Babylon"*, New Heaven: Yale University.

Redner, S. (2005), "Citation Statistics from 110 Years of Pyhsical Review", *Physics Today,* 58 (1), p. 49.

Small, H. (1973, July -August), "Co-citation in the Scientific Literature: A new Measure of the Relationship Between Two Documents", *Journal of the American Society for Information Science* , 265-269.

Thelwall, M., Vaughan, L., & Bjorneborn, L. (2005), "Webometrics. In Editor", *Annual Review of Information Science and Technology*, Book Chapter.

Van Raan, A. F. (1997), "Scientometrics: State of art. *Scientometrics*", *38* (1), pp. 205-218.

White, H. D., Wellman, B., & Nazer, N. (2004), "Does citation reflect social structure", *Journal of American Society for Information Science and Technology* , 55 (2), pp. 111-126.

Wooldridge, M. J. (2001), *"An Introduction to Multiagent Systems"*, University of Liverpool.