ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Τμήμα Ψηφιακών Συστημάτων

Μεταπτυχιακή Διπλωματική Εργασία

ΜΕΛΕΤΗ ΜΕΘΟΔΩΝ ΑΝΑΛΥΣΗΣ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΤΟ ΠΕΔΙΟ ΤΗΣ ΤΕΧΝΟΛΟΓΙΚΑ ΥΠΟΣΤΗΡΙΖΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

Τσιτμηδέλλη Ασημένια

Α.Μ.: 11045

Επιβλέπων: Δημήτριος Σάμψων, Καθηγητής

Πειραιάς, Ιούνιος 2014

UNIVERSITY OF PIRAEUS

Department of Digital Systems

Master of Science Dissertation

STUDYING CO-AUTHORSHIP NETWORKS IN THE FIELD OF TECHNOLOGY ENHANCED LEARNING

Tsitmidelli Asimenia

Supervisor: Demetrios G. Sampson, Professor

Piraeus, June 2014

Abstract

Research collaboration is studied in different research areas, so as to provide useful insights on how researchers combine existing distributed scientific knowledge and transform it into new knowledge. Additionally, research analytics attempt to provide meaningful data for researchers by analyzing the way research and research collaboration takes place. A substantial input for research analytics, is, thereby, the analysis of research collaboration. Within this context, in this thesis, we utilize the co-authorship network of researchers who collaborate in Technology-enhanced Learning (TeL), in order to gain meaningful insights about the research collaboration in this specific field. This is achieved through the case of the Educational Technology & Society (ETS) Journal, where Social Network Analysis (SNA) metrics are applied for analyzing its co-authorship network over a period of 15 years. The results of our analysis provided us with useful insights about the ETS Journal co-authorship network's identity and evolution. Our main findings provided evidence that the network consists of a main core of prolific authors and of a large number of non-connected subgroups. At the same time, it was found that the majority of the key authors appear to have a Taiwanese national background and that they have established a strongly connected group that collaborates frequently, diversely and widely. Both authors and collaborations of the ETS co-authorship network have been polynomially increasing during the past 15 years and the co-authorship network has probably arrived at its "phase transition", which means that it is expected to start becoming more connected in the coming years.

Περίληψη

Η ερευνητική συνεργασία έχει μελετηθεί σε ποικίλους τομείς της επιστημονικής έρευνας, έτσι ώστε να αποκτηθούν χρήσιμες πληροφορίες για το πώς οι ερευνητές συνδυάζουν την υφιστάμενη επιστημονική γνώση, η οποία, διαμοιράζεται μεταξύ ερευνητών, ώστε να παράγουν νέα ερευνητικά αποτελέσματα και κατ' επέκταση νέα γνώση, μέσω της ερευνητικής συνεργασίας. Από την άλλη πλευρά, η «Ανάλυση Ερευνητικών Δεδομένων» (research analytics), χρησιμοποιείται με σκοπό την παροχή πληροφοριών, δεδομένων και συμπερασμάτων, αναλύοντας τον τρόπο που η επιστημονική έρευνα λαμβάνει χώρα σε διαφορετικά πεδία. Συνεπώς, μια από τις σημαντικότερες πηγές δεδομένων για την Ανάλυση Ερευνητικών Δεδομένων, είναι η ανάλυση της ερευνητικής συνεργασίας. Στο πλαίσιο αυτό, στην παρούσα Μεταπτυχιακή Διπλωματική Εργασία, αξιοποιούμε το δίκτυο ερευνητικής συνεργασίας που δημιουργείται μέσω της συνεργατικής συγγραφής επιστημονικών άρθρων των ερευνητών που συνεργάζονται στον τομέα της Τεχνολογικά υποστηριζόμενης μάθησης (TeL). Σκοπός μας είναι η εξαγωγή συμπερασμάτων τα οποία θα συμβάλλουν στην κατανόηση της ερευνητικής συνεργασίας στο συγκεκριμένο τομέα. Για το σκοπό αυτό, εφαρμόζουμε μετρικές Ανάλυσης Κοινωνικών Δικτύων (Social Network Aanalysis) στο έγκριτο επιστημονικό περιοδικό Educational Technology & Society (ETS) Journal. Τα αποτελέσματα της ανάλυσης, μας παρέχουν ενδιαφέρουσες πληροφορίες τόσο για την ταυτότητα του δικτύου – σε ποια μορφή βρίσκεται σήμερα- όσο και για την εξέλιξή του με τα χρόνια. Από τα ευρήματά μας είναι εμφανές ότι στον κεντρικό πυρήνα του δικτύου βρίσκεται ένας σημαντικός αριθμός συγγραφέων με μεγάλο αριθμό δημοσιεύσεων. Ταυτόχρονα, στο δίκτυο υπάρχει και ένας μεγάλος αριθμός μη συνδεδεμένων, μικρών σε μέγεθος, ομάδων συγγραφέων. Επιπλέον, διαπιστώθηκε ότι η πλειοψηφία των κεντρικών συγγραφέων προέρχεται από την Ταϊβάν και ότι έχουν δημιουργήσει

μια καλά συνδεδεμένη ομάδα που συνεργάζεται με μεγάλη συχνότητα, ποικιλομορφία και εύρος. Ακόμα, παρατηρήθηκε ότι τόσο οι συγγραφείς όσο και οι συνεργασίες του δικτύου, κατά τα τελευταία 15 χρόνια, παρουσιάζουν αυξητικές τάσεις, οι οποίες είναι πολυωνυμικού χαρακτήρα. Τέλος, διαπιστώθηκε από την ανάλυση, ότι το δίκτυο συνεργασίας πιθανότατα έχει φτάσει στη «μεταβατική φάση» της εξέλιξής του, πράγμα που σημαίνει ότι αναμένεται αύξηση των συνδέσεων, τα επόμενα χρόνια.

Acknowledgements

I would like to express my deepest appreciation to all those who provided me the opportunity to complete this thesis. Without the support, patience and guidance of the following people, this study would not have been completed.

Firstly, I would like to express my deepest gratitude to Professor Demetrios G. Sampson who undertook to act as my supervisor despite his many academic commitments. His wisdom, knowledge and encouragement truly motivated and inspired me.

I would also like to thank Dr. Panagiotis Zervas for the overall contribution he made to my thesis. I especially appreciate his constructive comments and support throughout this process.

I have also received insightful comments and suggestions by Professor Nian-Shing Chen and Professor Kinshuk, on the grounds of our common publications, for which I am grateful.

Finally, my friends' and family's support and encouragement were invaluable.

Table of Contents

Table of Figures

Table of Tables

Table of Abbreviations

ETS	Educational Technology and Society
SNA	Social Network Analysis
TeL	Technology enhanced Learning

Chapter 1

Introduction

1.1 Research Topic and Motivation

One of the most fundamentally common elements of our everyday lives is our human to human interactions. Those interactions may typically vary in duration, type, purpose and context but they are built into every aspect of our life and, in fact, they are a very significant part of it.

The case, in which an interaction has more impact and produces the greatest results, is, undoubtedly, when an interaction turns into collaboration. Collaborations are complex but rather important procedures, especially since they have been proved to be one of the most effective ways of producing new, creative and significant work. Particularly in today's society, where the flow of information alone is too demanding to be handled by one person, expertise and division of information into small pieces, is considered necessary.

In the aforementioned context, where collaboration has become a necessity, analysis is also a basic requirement. Analysis helps us to break down in small parts, and consequently, relatively easily grasp, all the available information. Serving this principle, the analytics approach is a contemporary view of any major process. The analytics' contribution is crucial to making the information understandable and easily manageable, as well as monitoring a behavior or a sequence of events and predicting a future outcome.

In our study, the common denominator that binds those two concepts- collaboration and analytics- is scientific research. We are fascinated with scientific research as a process and we analyze the place of collaboration and analytics in this said process.

Consequently, we focus our analysis on research collaboration and research analytics, respectively.

The means we utilize in our study, in order to focus on the research process and, ultimately, on the researcher behind the research results, is co-authorship network analysis. This approach allows us to face the research collaboration analysis, in the context of social networks. We also use relevant tools, which in our case, are the social network analysis (SNA) metrics, namely graph metrics and vertex-specific metrics.

Our research scope is to analyze the co-authorship process that takes place in the field of Technology Enhanced Learning (TeL). Our study has revealed that although there are some studies addressing this issue, they are focused on the communities created around scientific conferences. Attempting to take a different approach, in this study we analyze an accredited and mature scientific journal, namely the Educational Technology and Society (ETS) Journal. The subject of our analysis is also open access and covers multiple research topics in the field of TeL, for over 15 years.

To sum up, in our study we focus on the research process, and specifically we analyze the collaboration, involved into producing the research results which are published in ETS Journal. In order to do that, we use a co-authorship network, which is based on the community of the ETS Journal. On the aforementioned network we thereafter apply a series of SNA metrics in order to gain insights which are aligned with the input that is necessary for research analytics.

The importance of our study can be initially found on the significance of the insights and patterns that we acquire. Through the aforementioned analysis, we become capable of providing satisfactory answers no only to the question “what?” but also to the questions “how?” and even “why?” in relevance to the way researchers collaborate in order to produce research results. This indicates that through such types of analysis we are able to gain really useful information concerning not only the research results but the researcher and the overall research process, as well.

Our study is, moreover, aligned, with the major trends in research analysis. What we notice in recent years, in particular, is an accumulative interest on research analysis and a rise of the bibliometrics field. We, thereby, follow that pattern, but by introducing some news elements, as well. For instance, for the type of community that we have chosen to analyze, namely the researchers who have collaboratively published their work in a scientific journal, there has never been conducted any similar kind of analysis in this particular TeL field.

1.2 Thesis Contribution

The main outcomes of this thesis have resulted to the following publications:

P. Zervas, A. Tsitmidelli, D. Sampson, N.S. Chen and Kinshuk, "Identifying Research Collaboration Patterns via Co-Authorship Network Analysis of the Educational Technology & Society Journal", Educational Technology & Society (submitted for publication – under review)

P. Zervas, A. Tsitmidelli, D. Sampson, N.S. Chen and Kinshuk, "Studying Research Collaboration via Co-authorship Analysis in the field of TeL: The Case of Educational Technology & Society Journal", In Proc. of the 14th IEEE International Conference on Advanced Learning Technologies (ICALT 2014), Athens, Greece, IEEE Computer Society, 7-10, July 2014

1.3 Thesis Structure

The current thesis is structured as follows: Following the introduction, the next chapter presents the background and problem definition of our study. The background includes four fundamental elements, namely research collaboration, research analytics, co-authorship networks and SNA metrics that will be utilized in the present work, as well as related works from the literature in the field of TeL. Those works include citation and co-authorship network studies. In the subsequent section, we clearly introduce the research problem which has emerged from our previous analysis. In the next chapter, we describe how we address the said problem ourselves. Thus, we present the sample and method of analysis conducted on the co-authorship network of the ETS Journal. Following that, we study the results of the analysis. Our findings are divided into two conceptual subcategories, describing (a) the network at its present state and (b) its evolution through time, from 1999 to 2012. Within these subcategories, we present the main findings and the conclusions that can be made based on our analysis. Finally, we discuss the meaning and impact of our results and present ideas for future work.

Chapter 2
Background and Problem Definition

In this chapter we will explore the background of our study and we will present the research question that we attempt to answer in the following chapters. Firstly, we discuss the most significant elements that are involved in our study. Every element is clearly defined and meticulously described. Next, we will explore relative works in our research field of interest. From the presentation of background concepts and relevant studies, the research problem will, thereafter, emerge. This chapter sets the basis for understanding and interpreting our scope and our research findings.

2.1 Research Collaboration

Research collaboration is one of the main concepts that we discuss in our study. We find this topic particularly interesting, since it provides us with insights on the scientific methodology, and specifically the interactions between researchers. However, because of its vast and irregular characteristics, research collaboration is hard to be grasped, described or studied using the traditional research methods. This fact has dictated several research limitations which we took into consideration in our study.

In this section we will explore the nature of collaboration that takes place between researchers in various scientific fields. Our scope is focused on the definition and main characteristics of research collaboration.

2.1.1 Definition

Research collaboration is a rather complex concept. That is the main reason why there have been several attempts in literature (e.g., Hu & Racherla, 2008; Katz & Martin, 1997; Subramanyam, 1983) to define what research collaboration consists of, but most of them are limited by research collaboration's vast nature. However, research collaboration could be defined as: "the working together of researchers to achieve the common goal of producing new scientific knowledge" (Katz & Martin, 1997). Research collaboration is seen, in a general manner, as a special form of collaboration, undertaken for the purpose of scientific research (Amabile et al., 2001).

As it was previously mentioned, research collaboration is a rather complex concept, since it involves the participation of various factors in the context of collaboration that promotes scientific research. Research collaboration inherits its complexity from the complexity involved in almost any form of human interaction. For example, all the researchers participating in a common work are not expected to produce similar results, in quantity or even quality, do not necessarily devote the same amount of time or energy, have usually different methods and follow various approaches in relevance not only to their research techniques and cognitive strategies, but also to their collaborating patterns and routines. Certain difficulties in defining research collaboration in an absolute manner are, therefore, anticipated.

According to a different definition for research collaboration, (Katz & Martin, 1997), "only those researchers who contributed directly to all the main research tasks over the duration of the project can be counted as collaborators". This immediately brings us to a research collaboration categorization. For that purpose, we can categorize research collaboration into two subcategories, namely direct and indirect.

Direct collaboration takes place between two or more researchers who directly interact to serve a common goal or solve a shared problem, and ultimately, produce jointly shaped results. There are cases, though, when some researchers may contribute indirectly to a research project, e.g. by providing just initial guidelines or final feedback. Then, the collaboration should be characterized as indirect, since those researchers do not directly contribute to the main core of the research tasks.

Another notable aspect is the time frame of research collaboration. Research collaboration does not necessarily fall into the category of synchronous or asynchronous interaction. It usually consists of both types of elements tied together in a common process. What is more, research collaboration is a scalar process which often spreads out to many weeks, months or years, incrementally. It might also be continuous or sporadic. Even the contribution of different collaborators may vary or change through time. For those reasons, the exact dimensions of research collaboration are difficult to be grasped by a researcher using traditional means such as interviews or questionnaires.

2.1.2 Influencing Factors

A rather important point that can be made by inspecting relative literature is that research collaboration can be associated with a number of influencing factors (e.g. Bukvova, 2010; Glänzel, 2002; Katz& Martin, 1997). Those factors have been analyzed and organized in several studies. For example, Bukvova categorizes those factors as internal, utilized by researchers, or external, which present interest for decision makers. Next, we will present four of the most important influencing factors that we will further explore in this study.

Research collaboration and spatial distribution of collaborating parts

Firstly, it has been shown (Hagstrom, 1965; Kraut & Egido, 1988) that research collaboration can be influenced by the spatial distribution of collaborating parts. It appears, from the aforementioned studies, that the shorter the distance between two researchers is, the bigger the possibility to engage in research collaboration activities. Another study researching collaboration through co-authorship (Katz, 1993), has, as well, shown that research collaboration decreases exponentially with the distance separating pairs of institutional partners. The reason why geographical location is associated with research collaboration can be grasped rather easily. Factors such as common speaking language, background or common cultural influences appear to assist and nurture research collaboration. Additional proof for the importance of the geographical distribution is the usual categorization of research collaboration as presented in Table 2.1.

Table 2.1: Different Levels of Collaboration and Distinction between Inter and Intra Forms, Katz & Martin (1997)

	Intra	**Inter**
Individual	-	Between individuals
Group	Between individuals in the same research group	Between groups (e.g. in the same department)
Department	Between individuals or groups in the same department	Between departments (in the same institution)
Institution	Between individuals or departments in the same institution	Between institutions
Sector	Between institutions in the same sector	Between institutions in different sectors
Nation	Between institutions in the same country	Between institutions in different countries

As we notice from Table 2.1, the type of collaboration that benefits the most from spatial proximity is intramural collaboration, as presented in the middle column. That type of collaboration thrives within the same department, research group or institute. Although intra-national and inter-national collaboration activities produce results of comparable quality, inter-national collaborations appear to have a positive influence on the future output of collaborations (He, Geng & Hunt, 2009).

Research collaboration and social proximity of collaborating parts

It has also been shown (Hagstrom, 1965) that research collaboration can be favored by social proximity. In his study, Hagstrom has observed that collaboration occurs in higher rate between peers, compared to individuals who belong to different social or academic ranks. This indicates that two individuals are far more likely to collaborate if, for example, they share a common academic rank, e.g. professor, lecturer etc., since they are much more likely to have common interests, share experiences and background.

Research collaboration and scientific productivity rate of collaborating parts

In addition to the previous factors, scientific productivity can be associated with the research collaboration rate. Studies have shown (Beaver & Rozen, 1979; Braun et al., 2002; Katz & Martin 1997) that the productivity rate of a researcher correlates with his collaboration rate. The previous finding indicates that the more intensely a researcher publishes his work, the wider the network of his collaborations is and vice versa; a researcher with many collaborations is very likely to have a high productivity rate. It has also been shown that there is a preferential attachment to researchers with higher productivity rates (Pravdic & Oluic-Vukovic, 1986). Thus, we can claim the fact that researchers with many published papers enjoy a wider range of research collaborators.

Research collaboration and impact of the end result

Finally, research collaboration has been indicated as an influential factor for the impact of a paper, based on citation number (Gomez et al., 1995; Hicks & Katz, 1997; Hicks et al., 1994; Hurley et al., 2013; Katz & Martin, 1997). More specifically, adding an author to an article has a positive impact on the number of citations the article receives. Glänzel (2002) reports similar findings in a large number of related studies; it was found that intense research collaboration is associated with high citation number. This could be partly explained by the fact that a high number of researchers, who collaborate on a common research project, potentially raise the quantity, quality and variety of the end result (Rigby & Edler, 2005). The produced work consequently has a higher impact, since it enjoys a great number of citations.

In conclusion, research collaboration is one of the most fundamental aspects of research. It is an excellent way to expand and promote research attempts, as well as to encourage and spread new research methods and techniques not only across institutes or countries but also across scientific fields or even across the academic world. It comes thereby as no surprise the fact that, over the past two decades, bibliometric studies have shown an increase of collaborative activity in almost every single scientific field (e.g., Hicks & Katz, 1997; Sonnenwald, 2008).

For its great importance, when studying research collaboration, the previously listed factors should undoubtedly be taken into consideration, i.e., spatial distribution, social position, scientific productivity rate and impact through citation number. Furthermore, we should not overlook the necessity of utilizing special means and tools in order to describe and accurately visualize this intricate concept, since, as it emerged from our study, traditional means appear to be inadequate. (Bukvova, 2010; Katz & Martin, 1997). These issues need to be addressed by any researcher wishing to study research collaboration both sufficiently and effectively.

2.2 Research Analytics

Research analytics is one of the main concepts that run through our entire study. We find this topic particularly interesting because of its unique characteristics and extremely useful results. Research analytics provides us with insights on the scientific methodology, and particularly on the research methods followed by various scientists.

Thus, in this section we will discuss research analytics as a contemporary method of studying the research process. We are going to define analytics by introducing a sequence of separate analysis steps and we are going to specifically focus on the scientific research field. Additionally, we will explore research analytics' main characteristics and basic requirements for proper implementation.

2.2.1 Definition

The analytics approach is the use of mathematical and algorithmic methods to describe part of the real world (Harmelen & Workman, 2012). Analytics can therefore be used as a tool which reduces the real world complexity to a more easily understandable form. Analytics is more often than not used as an "umbrella term" (Watson, 2011) to describe the way a provided set of data is analyzed by using various mathematical and statistical techniques.

The analytics approach has the following separate conceptual steps:

- Firstly, analysts gather information about a process, a situation or a cause. The information can be gathered through various means namely, customer questionnaires, feedback surveys or analysis software. Consequently, the data gathering can be either synchronous with the user experience or it can be implemented after the process is over, as an overview or assessment point for the user.
- Secondly, the data gathered are analyzed. The analysis can be further divided into small steps, with each serving a different purpose or answering a different question. These steps can be carried out by different types of software, specifically designed for each purpose. Since analytics can require extensive

computation, the algorithms and software used for analytics take advantage of the most current methods in computer science, statistics, and mathematics (Kohavi, Rothleder & Simoudis, 2002).

- The main goal for analytics is to provide insights on a process, situation or cause, as we have mentioned beforehand. Therefore, the third and final step is the presentation of the analysis results in a meaningful way for the end user. In order to optimize the end result presentation, analytics often favor data visualization. This fact indicates that a visual, schematic approach is most of the times preferable, since analytics combine the process of visualizing large amounts of data in one final, multi-layered result (Harmelen & Workman, 2012). This visualization is often achieved through various means such as multiple filters or detailed charts that cover issues from growth or progression to user behavior.

The analytics approach is also usually associated with concepts such as predictive modeling, since one of its most fundamental purposes is to provide insights which can be utilized in order to make accurate predictions. In other words, the data gathered and analyzed by the analytics approach, if provided in the proper way, can be a first basis for making safe assumptions and, consequently, modifying one's behavior. In conclusion, analytics can be used in creative ways in order to optimize a current situation, or predict the possible circumstances, needs or obstacles that may appear in the horizon.

Analytics are proved particularly valuable and efficient in areas rich with recorded information, and that is one of the main reasons that they are so popular with a wide range of fields. Thus, there are many forms of analytics, namely business analytics, web analytics, learning analytics, research analytics etc.

Research analytics is the brunch of analytics which is addressed to and designed for the specific area of scientific research. Research analytics is a research field that utilizes mathematical and algorithmic methods for analysing the way research takes place and it can provide meaningful data for researchers (Harmelen & Workman, 2012). In the context of research analytics, we can identify a number of particular characteristics which contribute to the desirable result, which more often than not, is to make suggestions and provide insights to researchers in order to optimize the scientific research process.

2.2.2 Main Characteristics

The research analytics approach is, beyond any doubt, extremely useful for researchers and scientists. Because of its unique nature, this approach concentrates a series of specific components. Some of the principal components of the analytics approach are the following (Harmelen & Workman, 2012):

- Application of analytics in the real world.
- Visualization of the results of analysis.
- Methods, which include the implementation of statistics, SNA etc.
- Data, which is the raw material for analytics.
- Technological infrastructure which is a prerequisite for the realization of the analytics approach.

From the aforementioned components, maybe the most challenging is the acquisition of proper data. This dictates that the data need to be accurate and valid, in order to provide meaningful and multi-functional analytics. This also reveals the necessity for constant comparisons with other available sources, in order to come to accurate conclusions. Finally, it is very important that the extracted data are not only specifically focused on certain areas and accurate, but also that they are in an easily understandable, practical and useful form.

Although the implementation of research analytics can provide important insights and suggest solutions to many problems, some researchers suggest that generalizing results should be done cautiously (De Bellis, 2009). That is not to say that generalizing is impossible, rather than there should be rules and measures to avoid oversimplification or misinterpretation. One such solution is scaling (Iglesias & Pecharromán, 2007).

In conclusion, as we mentioned beforehand, the research analytics approach is an approach that attempts to provide meaningful data for researchers by analyzing the way research takes place. For that purpose, the fact that there are particular pieces of information that have been proven to be especially important for researchers in the past, should be taken into consideration. Such data involve (a) key research areas in various research fields, (b) prolific authors and some of their background information, e.g. geographical distribution or (c) the impact of specific research papers which can be counted through citations. This information appears to be particularly useful in the research field and it is thereby provided by most major research analytics sources.

2.3 Co-authorship Networks

Co-authorship networks are social networks that are formulated through strong social ties. A social bond in a co-authorship network implies a social bond that has been cultivated through the process of producing common research work. For that reason, we think co-authorship networks are a unique indicator of the research process.

In this section, we present the features and characteristics of co-authorship networks, which are a specific class of social networks. The special properties and abilities of co-authorship networks are utilized in our study.

2.3.1 Definition

A social network is a social science tool which consists of a structure made up of a set of social actors (such as individuals or organizations) and a set of dyadic ties between these actors (Wasserman & Faust, 1994). The sum of those actors, along with their ties, the circumstances under which these are built and the impacts which they may cause, are all parts of a social network. A social actor can be not only a person, but also a group, a company, a research team, a lab or an organization (de Nooy, Mrvar & Batagelj, 2005). A social network can therefore be created between any kinds of actors, which are connected through various socials interactions.

The fundamental difference between a social network approach and a non-network approach is the inclusion of data which are associated with the relationships among actors (Wasseman & Faust, 1994). Traditional sociological approaches appear to not be addressing sufficiently the relational information, while the social network perspective places the relational information in the center of the research process. As a result, researchers who wish to study relations of any kind between some given actors should utilize the social network approach, as one of the most effective solutions to the problem of analyzing in depth relational information.

In general, social networks are self-organizing, emergent, and fragmented, mostly because of their origin, which favors the rapid, irregular, spontaneous and even chaotic growth. At the same time, social networks are rather complex by nature, an exact imitation of social interactions. This fact can raise several obstacles to researchers and it turns social network research in a particularly demanding, yet highly useful, task.

Although social networks are particularly complex, in most social networks we can notice a distinct structural organization, which is more often than not, found in all sorts of networks. That structure typically contains many small subgroups that are highly interconnected. In these subgroups interactions appear mostly in a local level and seem to concentrate around highly influential actors. Those smaller interactions form the bigger patterns that we spot macroscopically. These patterns can become more apparent as network size increases or as new actors are added over time. On the other hand, the importance of a local system or even a prestigious actor may be lost or lessened in a large network. The study of these structures extends from identifying local and global patterns, locating influential entities, to examining network dynamics.

The growing tendencies of the field of social networks are shown in numerous studies. Knoke and Yang confirm a significant increase in the number of social science publications with “social network” as a key concept (Knoke & Yang, 2008). Their research has shown the increasing rate of publications with “social network” found in the Title or the Abstract of papers. Another innovating study was published by Mislove et al. in 2007. The study is covering the issue of online social networks exclusively. It involved a data set of over 11.3 million users as social actors. The data was retrieved from just 4 popular online social networks. It is obvious from the above that the interest in social networks grows rapidly, as the role social networks play in our lives, becomes more important.

In our study, we focus on co-authorship networks. Authorship is a primary bibliometric descriptor of a scientific publication (Glänzel, 2002). Analyzing authorship facilitates the understanding of the underlying scientific, cognitive and social structure of a scientific work. Since our scope is the analysis of research collaboration, we utilize co-authorship networks.

Co-authorship networks fall under an important class of social networks, which are structured through the process of co-authoring a paper between two or more authors

(Liu et al., 2005). The typical actors of co-authorship networks are researchers who co-author scientific papers.

The idea of co-authorship networks has its origins, among other places, with the Hungarian mathematician Paul Erdös. Erdös had published at least 1512 papers in his field and mathematicians often tried to calculate their geodesic distance from him. For example, somebody who had directly co-authored a paper with Erdös, would get an Erdös number of 1, somebody who collaborated with him, an Erdös number of 2 etc. Erdös has a number of 0. Eventually, it reflected an individual researcher's level in terms of being well connected with a prestigious scientific society centered by Paul Erdös. Many great scientists and noble prize winners have small Erdös numbers. It has been found that the average Erdös number is 4.7 (Grossman, 2002). The paradigm of Erdös number, not only partly covers the origin of co-authorship networks, but it also offers an insight on how co-authorship networks are structured. The specific way actors are tied through connections in co-authorship networks is considered as functional for the transfer of data across the network (Hansen, Smith & Shneiderman, 2011).

As it can be expected, there is a growing interest for studying co-authorship networks, as well. That increasing tendency was first noted around the beginning of the century, with three important studies (Barabási et al., 2002; Moody, 2004; Newman, 2001).

In his study, Newman applied methods of distribution analysis such as number of papers per author and number of collaborators per paper, and techniques to measure the strength of the collaboration between pairs of individuals, shortest scientific distance between a pair of authors in a given co-authorship network, and connectivity (e.g., giant component and clustering) among the participating authors in a co-authorship network.

A similar study was performed for the disciplines of mathematics and neuro-science by mapping the co-authorship networks in the aforementioned fields, through all relative scientific journals (Barabási et al., 2002). In that study, a new model is proposed; it covers the mapping and the visualization of the network evolution from 1991 to 1998.

James Moody, accordingly, studied the sociological community around the journal articles listed in Sociological Abstracts. His effort to relate collaboration with the research specialty of every author presents great interest.

The growing trend in co-authorship networks is well continued until today with added interest. New research fields are analyzed, while variant means are used as a context for the analysis, e.g., scientific conferences (Liu et al., 2005, Erman & Todorovski, 2010, Erfanmanesh, Rohani & Abrizah, 2012), literature available from the Web of Science (Velden, Haque & Lagose, 2010), journals (Barabási et al., 2002, Yan & Ding, 2010) and most cited/popular papers (Lee et al., 2010).

2.3.2 Main Categories

Co-authorship networks follow the categorization of most other social networks, in relevance to their structure. The two main categories under which a co-authorship network can fall are (a) directed and (b) undirected. That characterization is dependent on the way the vertices of the network are tied with each other.

More specifically, if the formation of the connections has a specific direction, then, we characterize the network as directed. The networks, which present an obvious inclination in relevance to the direction of the connections, are the ones where the actors do not simultaneously receive the same role in a social relationship. For example, in a network which is formed based on the exchange of emails between social actors, the actors are tied through a connection where there can be one sender and one or more receivers. Although in our example the actors are interconnected, they do not hold the same roles, or in other words, their relationship is directional. That type of network, and the relevant graph, is characterized as directed.

On the other hand, in the case of a network based on the co-authorship of scientific papers, the actors obtain the role of co-authors simultaneously. As a result, the graph created from this type of network, will be undirected. Consequently, for the co-authorship networks the standard structure is undirected. There are, though, some examples in literature on how a co-authorship network can correspond to a directed graph (Liu et al., 2005).

2.4 Social Network Analysis

2.4.1 Overview and Definition

Social Network Analysis (SNA) is the analysis of social networks. SNA views social relationships in terms of network theory, consisting of nodes and ties. SNA provides us with a set of rules and metrics that measure those social relationships. In this section we will discuss SNA as one of the most common and effective means of social analysis.

One of the most fundamental characteristics of SNA, is the fact that it is not a formal theory, but rather a perspective or a broad strategy. SNA is, in essence, a sociological approach for analyzing patterns of relationships and interactions between actors, in the context of social networks (Otte & Rousseau, 2002). SNA can support the identification of leaders within a social network, as well as highly connected groups and patterns of interactions between these groups. In general, SNA conceptualizes social structure as a network, with meaningful connections between actors and it focuses on the characteristics of the ties, rather than the characteristics of the individual member.

SNA is a relatively new academic field, since just in 1954, J. A. Barnes was the first who started using the term systematically (Barnes, 1954). The progress of the field was slow and linear, until around 1970, when major methodological breakthroughs occurred. In 1990, the field became particularly popular, when the interest in social networks and social network analysis methodology grew at a much more rapid rate. The range of the field's applications has grown exponentially since then, given that SNA can visualize the invisible patterns of social interaction.

As we mentioned beforehand, SNA originated in social sciences and was primarily used as a sociology tool. During the 20th century though, the field acquired elements from social psychology, statistics, graph theory and computer science. The field of SNA is today clearly multidisciplinary. What is more, its concepts are applied in order to analyze and study various other fields, such as biomedical research, mathematics, physics and neuroscience (e.g. Newman, 2001; Barabási et al., 2002).

SNA has its own scientific methods, tools and concepts that are widely used and highly acceptable. Some of the most commonly used concepts and theories of SNA can assist us understand the metrics and principles that are introduced in order to study a network. Those concepts reveal the majority of factors that a researcher investigates when studying a social network, and could be called common denominators in several similar studies that include the said topic.

- Bridge: A vertex which provides the only link between two individuals or even between two clusters. It is therefore called a bridge because it ties together different parts of the network that could not be connected in any other way. Bridges also include the case when the connection that they create is by far the shortest route. Shortest paths are always preferred due to possible message distortion or delivery failure (Granovetter, 1973).

- Distance: The minimum number of steps required to connect two particular vertices. As a step, we can imagine the edge that connects two vertices. The distance between two residents of the USA, in a network with all USA residents, was researched by Stanley Milgram in his "small- world" experiments. It was then, that the idea of 'six degrees of separation' was introduced for the first time. That idea indicated that two people, living in the USA, would be connected in mostly six steps of other vertices (Watts & Strogatz, 1998).
- Triadic Closure: Triadic closure is a concept in social network theory, first suggested by Georg Simmel in the early 1900s. The triadic closure concept includes three separate vertices. If those vertices are connected through edges two by two, there is a very large possibility the non-connected edges will connect eventually. Let's take for example three edges A, B, and C. If a tie exists between A-B and A-C, there is a tie between B-C, as well.
- Clique: Groups are identified as 'cliques' if every individual is directly tied to every other individual.

SNA clearly focuses on the interpretation of the ties between individuals and it uses a variety of statistical and visual means to achieve that. SNA also makes a variety of assumptions about actors, relations and the resulting structure (Wasseman & Faust, 1994):

- Actors are viewed as interdependent rather that independent.
- The ties between actors are considered as paths or channels for transfer of messages or resources.
- Network models conceptualize structure as enduring patterns of relations among actors.

There are three typical levels of analysis into which SNA may fall: micro-level, meso-level, or macro-level.

- Micro level

At the micro-level, social network research typically begins with an individual, or may begin with a small group of individuals in a particular social context.

- Meso level

In general, meso-level theories begin with a population size that falls between the micro- and macro-levels. However, meso-level may also refer to analyses that are specifically designed to reveal connections between micro- and macro-levels. Meso-level networks are low density and may exhibit causal processes distinct from interpersonal micro-level networks.

- Macro level

Rather than tracing interpersonal interactions, macro-level analyses generally trace the outcomes of interactions, such as economic or other resource transfer interactions over a large population.

2.4.2 Social Network Analysis Metrics

In SNA there is usually applied a series of measures, which aim to measure and describe the network, as an entity. Those are the graph metrics which are addressed to the whole structure. In SNA there is, as well, a great interest for vertices that are important for the structure of the network and can dramatically affect its evolution (Kaza & Chen, 2010). So, the thing that usually interests researchers is the positioning of certain authors. This leads to applying a series of measures that are specifically addressed to vertices, the vertex specific metrics.

As we mentioned in the previous section, co-authorship networks based on their directionality, can be categorized in two main subcategories, namely directed and undirected. This dictates that the latter category of metrics, namely the vertex specific metrics, can also be divided in two subcategories (Erman & Todorovski, 2009): (a) metrics for directed graphs, which measure the prestige and (b) metrics for undirected graphs, which measure the centrality.

2.4.2.1 *Graph Metrics*

In a social network can be applied a series of measures, which aim to measure and describe the whole network. Those are the graph metrics and they are the following (Hansen, Smith & Shneiderman, 2011):

- Graph Type: This metric defines the type of graph of the social network. The graph could either directed or undirected based on the nature of the connections (namely, the edges between vertices) that the graph represents.
- Unique Edges: This metric is used to count the number of connections or edges, which occur at least once between two vertices.
- Edges with Duplicates: This metric is used to measure edges that occur more than once between two vertices. If a connection occurs more than once, then, this it is recorded with this metric.
- Total Edges: This metric is the sum of unique and duplicate edges and counts the total number of edges that exist in the graph.
- Connected Components: This metric identifies and measures closed subgroups of vertices that are interconnected, but they do not connect with other vertices or groups outside of the specific subgroups.

- Maximum Vertices in a Connected Component: This metric measures the maximum number of vertices that are connected to each other in a component. This component is called the giant component.
- Maximum Edges in a Connected Component: This metric measures the number of edges that exist inside the giant component.
- Maximum Geodesic Distance (Diameter): This metric is used to measure the shortest possible distance between the two farthest vertices in the network. This metric is also called the "diameter" of the graph of the social network.
- Average Geodesic Distance: This metric is used to measure how close in average two vertices are placed, inside the graph of the social network. To measure this distance, steps are used; every vertex is a step. The higher the value of this metric is, the bigger the overall distance between vertices.
- Graph Density: This metric represents the completeness of vertices' interconnections. The maximal density is 1 (for complete graphs) and the minimal density is 0 (Coleman & Moré, 1983).
- Modularity: This metric is used to measure the degree of clustering that appears in the graph. This means that if a graph consists of distinct closed subgroups, then it has bigger modularity. Modularity is measured by comparing the number of edges inside a subgroup with the number of edges in the whole network. The values of the modularity lie in the range (−1/2, 1).

2.4.2.2 *Vertex-Specific Metrics*

As mentioned above, in a social network, researchers often try to identify some specific vertices which have an important role. In order to identify those vertices, a series of metrics is used. Those are the vertex-specific metrics and they are listed and described below (Hansen, Smith & Shneiderman, 2011):

Degree Centrality

The degree or degree centrality (Freeman, 1979) is one of the most common and useful metrics in SNA. It is used to measure the "popularity" of a vertex i.e., to count the number of unique edges that create a connection between this specific vertex and others. Each connection is only counted once, even if it occurs multiple times. In a directed graph, there are also two types of degree, the in-degree and out-degree. Since the graph under study in this case is undirected, in-degree and out-degree do not apply.

Betweenness Centrality

It is not only the number of edges of a vertex that is interesting, but also the positioning of a vertex. Betweenness centrality is based on the number of shortest paths passing through a vertex. Vertices with high betweenness centrality play the role of connecting different subgroups of the social network. They are also called bridges.

A good example of a vertex that is important for its location is a case when a vertex is the only connection between an isolated part of a graph and the rest of the vertices. When a message is to be communicated to all vertices in the network, this vertex will be of unquestionable significance. Thus betweenness centrality (Freeman, 1979) measures the "role" of a vertex in a network during the spread of a message. During this process, the message travels though the shortest path possible (Borgatti, 2004). Therefore, if a vertex is located on many such paths, then it will have a high value for betweenness centrality.

Closeness Centrality

Another metric that Freeman introduced is closeness centrality (Freeman, 1979). This metric is used to demonstrate how close a vertex is to others. It, specifically, represents the distance of a vertex to all other in the network by focusing on the geodesic distance from each vertex to all others. Closeness centrality can be interpreted as a metric of how long it will take information to spread from a given vertex to others in the network (Freeman, 1979).

For this specific metric, it needs to be pointed out that higher values indicate a more central role. For the value of this metric to be calculated, we use the opposite number of the sum of the distance (in steps) of the vertex under study, from all the other vertices. Also, the metric can be normalized, taking values between 0 and 1.

Eigenvector Centrality

It is used to measure the degree of a vertex and the degree of the vertices with which this vertex is connected (Bonacich, 1972). Eigenvector centrality can identify vertices that are connected with other "popular" vertices. The Eigenvector centrality metric has, nonetheless, a strong connection with the degree metric. It can range from 0 to 1.

2.5 Literature Review: Research Collaboration in Technology Enhanced Learning

There are a number of related studies presented in relevant literature, which provide us with fragments of data on how research collaboration takes place in the TeL research field. Because of the special nature of research collaboration that we have previously presented, these studies use exclusively social science tools.

The existing studies can be categorized into two subcategories, namely, citation/content and co-authorship analysis. The studies that fall under the former category provide their findings by conducting citation or content analysis on data from the TeL research field while the latter explore co-authorship analysis.

2.5.1 Citation and Content Analysis

In the field of Educational enhanced Learning, we come across a number of papers studying scientific collaboration through citation analysis (Cho et al., 2013; Kinshuk et. al., 2013) or content analysis (Bozkaya, Aydin & Kumtepe, 2012). Citation analysis generally studies who cites who in a scientific community or field, using social networks, while content analysis studies the major topics that emerge. Thus, the purpose of the following papers may not always be an exclusive study of research collaboration as it is formed in the field of TeL, but, nonetheless, they provide important and useful data on that particular topic.

Bozkaya, Aydin & Kumtepe (2012), have studied the Turkish Online Journal of Educational Technology (TOJET) using content analysis. The findings in that current study were filed under three sections, namely general characteristics, research themes and issues and research design. In reference to research collaboration, it was found that articles had either one or two authors. This fact is mainly caused by the Council of Higher Education in Turkey, which stipulates it as a prerequisite for academic promotions and appointments to have single author-based-studies, according to Bozkaya, Aydin & Kumtepe. They also claim that preferring individual studies may also be considered as a cultural phenomenon of the Turkish educational system.

Cho et al., 2013, conducted a study on the Educational Technology Research and Development (ETR&D) journal, in order to examine the trends and issues of the educational technology field's scholarly community that have evolved in the past two decades. The authors have applied SNA metrics to the citation network of the ETR&D journal. Their scope was to identify influential papers and scholars in the field of TeL. They noticed that among frequently co-cited papers, there was a strong relationship between a few influential researchers. They also detected five major research subgroups, which have generated and promoted five distinct key research areas in the field.

Kinshuk et al., 2013, have conducted an analysis of the most highly cited empirical studies in the Educational Technology & Society Journal for the time period 2003-2010. They found that the vast majority of the highly cited articles were co-authored with one or more collaborations. Their findings on research collaboration specifically addressed the topic of international collaboration. They found that of the top highly cited empirical studies published in the ETS Journal, 20% involved international co-authorship. However they concluded that the numbers of internationally co-authored articles in different time intervals were equal and they did not present the expected increase. This is mainly caused, according to the authors, by practical reasons such as different characteristics and English proficiencies.

2.5.2 Co-authorship Networks Analysis

As we have previously mentioned, there is a growing tendency of co-authorship network studies in various research fields, involving a variety of analyzed means. Nonetheless, in relevance to the TeL research field, studies are limited in number. In this section, we provide an overview of existing studies that focus on co-authorship network analysis.

Kienle & Wessner (2005), who have studied the co-authorship network of the researchers that publish their papers to the proceedings of the conference on "Computer Supported Collaborative Learning (CSCL)". Within their study, the authors have focused on the evolution of the co-authorship network and the identification of key authors. Their findings revealed a relatively small number of recurring participants and a growing tendency for international participation.

Ochoa et al. (2009) have analyzed the co-authorship network of the researchers that publish their papers to the proceedings of the "World Conference on Educational Media and Technology (ED-MEDIA)". The study focused on (a) the identification of influential authors by applying vertex specific centrality metrics for author ranking and (b) the identification of important groups of the network. The analysis revealed that ED-MEDIA is a vibrant and collaborative community. The authors also proposed a Personalized Recommender System for future participants.

Reinhardt et al. (2011) have analyzed the co-authorship network of the researchers that publish their papers to the proceedings of the European conference on technology enhanced learning (EC-TEL). Within their study, the authors have identified important groups of the network and they have indicated that EC-TEL is a highly fragmented conference which consists of a large number of weakly connected subgroups. In addition to that, the findings revealed a small number of central and influential authors.

Pham et al. (2012) studied evolution patterns of co-authorship networks for five TeL conferences. Thus, they proposed a development model with different stages, which can be implemented in order to determine in which stage a conference is. The authors have also applied SNA metrics to the co-authorship networks of these five TeL conferences and they have compared the results with four conferences in the database research field.

As we can notice from the aforementioned studies, analysis of co-authorship networks can identify influential authors, as well as important groups of authors that work on specific research areas in the field of TeL. Nevertheless, the existing studies have been focused on co-authorship networks developed by researchers, who publish their papers to the proceedings of international conferences.

2.6 Problem Definition

In our study we propose a solution for the problem of the acquisition of patterns and insights on the research collaboration that takes place in the field of TeL. In order to gain a clear picture of the research collaboration in this specific scientific field and to make our results usable by others in a meaningful way, we need to answer this problem by considering every possible aspect and difficulty that we might face.

One main difficulty that we are facing is that there is a number of research limitations that we need to consider before analyzing research collaboration. These limitations are dictating, as we mentioned above, that we need to use non-traditional means in our research, i.e. networks. Commonly used metrics for measuring research collaborative activity include, among others, the co-authored publications (concerned with who works with whom) and the citations (concerned with who cites whom). From the literature review, we also noticed that some conclusions can be drawn from content analysis, as well.

Nonetheless, we find co-authorship networks to be the most detailed and direct approach available for measuring research collaboration for a number of very specific reasons. Firstly, it can be argued that co-authorship networks capture scientific collaboration in more detail than the much studied citation networks, given that co-authorship indicates a much stronger social bond, considered the time and effort that is put into publishing research results. Secondly, studies have shown a significant positive correlation between collaboration and co-authorship (Patel, 1973), a fact which amplifies our claim. Furthermore, many researchers, (e.g. Newman, 2001; Zare-Farashbandi, Geraei, & Siamaki, 2014) find co-authorship to be one of the most tangible forms of research collaboration, in relevance to other available methods. It is thereby concluded that co-authorship is the currently suggested method that should be used in order to map research collaboration, especially since it presents a number of important advantages.

The benefits of co-authorship networks are numerous. Firstly, we are presented with an accurate method of mapping scientific collaboration, which is thoroughly defined through the co-authorship of a paper, and can be considered solid and relatively stable. What is more, it frees the researcher from any geographical or temporal limitations, which are normally present, when studying collaboration; a usually localized and strictly time-defined phenomenon. But, maybe, the major contribution of a co-authorship analysis approach is that we can easily study the detailed end-result, in order to gain meaningful insights on the way scientists collaborate in a specific context, e.g. we can identify most influential authors and we can study the progress of collaboration or do comparisons with other cases. We can, thereby, conclude that for all their abilities and advantages, co-authorship networks should be used in our study as a primary indicator of how research collaboration takes place.

At the same time, another difficulty is that our research results need to be aligned with the research requirements that we have presented for analytics. For that purpose, we will utilize various SNA metrics, in order to extract suitable results.

As we previously mentioned there are three basic data requirements for an analysis that aims to provide input for research analytics, namely (a) specific types of data, (b) proper data and (c) easily understandable, practical and useful data. For that purpose and according to the relevant literature (Otte & Rousseau, 2002), co-authorship network analysis can be achieved by using SNA metrics, which cover the aforementioned prerequisites. Another fact that supports our claim is that SNA is a main element of the method component of analytics (Harmelen & Workman, 2012). Thus, the SNA metrics are used, in our study, in order to identify features and patterns of the typology of the co-authorship network.

A third element that needs to be taken into consideration, in the problem solving process, is the context around the network which we study. For example, there should be obvious and easily explainable differences between the formulation of patterns of collaboration between members of different communities, such as an annual international conference and an international journal, even if they both address to and share the same scientific field. What we concluded from the analysis of the relevant literature, is, that the co-authorship studies all focus on one or multiple scientific conferences. Thus, one issue to investigate is to study the co-authorship networks developed by scientific journals in the field of TeL. Next, we address this issue by applying SNA metrics for the analysis of the co-authorship network of the Educational Technology & Society (ETS) Journal.

2.7 Synopsis

In this chapter, we have firstly presented the four principal background elements of our study, namely, research collaboration, research analytics, co-authorship networks and social network analysis. We have provided definitions and descriptive information for each of these concepts, in order to set the context and scope of our study in the right limits.

Furthermore, we have listed and described related works that explore the aforementioned concepts. The studies are specifically conducted in the field of TeL. The presentation of those studies serves as an indicator of the principle methods that are followed for the analysis of the presented concepts. The studies provide us, as well, with details on the typology and content the most significant results that can be extracted by analyzing relevant data.

In the last section, we have thoroughly described the research problem that has emerged from the previous analysis. We specifically justified the use of co-authorship networks, SNA metrics and of the ETS Journal, as a sample, for solving the problem of acquisition of information about research collaboration in the field of TeL. We have, thus, set the theoretical justification for the methodology that we will discuss in the subsequent chapter.

Chapter 3
Research Methodology

The present chapter provides details on how we attempt to solve our research problem. Thus, in the first section we will describe in detail the sample of our study. We will next present the software we used for the analysis and the process that we followed. Finally, we will present our interpretation of the used SNA metrics, specifically adjusted for co-authorship networks. This chapter provides a detailed overview of our research methods.

3.1 Sample

The ETS Journal was selected because of the following three reasons: (a) it is open access and consequently we were able to have access to all published papers; (b) it is an accredited journal because it has an impact factor of 1.171 according to Thomson Scientific 2012 Journal Citations Report and it is currently ranked 4th in the top-20 publications for Educational Technology in Google Scholar[1]; and, (c) it is published for more than 15 consecutive years, since 1998.

Table 3.1: The ETS Journal Sample

Journal	**Journal of Educational Technology and Society**
First Publication Year	1998
No. of Volumes	15

[1] http://scholar.google.com/citations?view_op=top_venues&hl=en&vq=soc_educationaltechnology

No. of Issues	57
Publication Frequency (issues/year)	4
Abstracted/Indexed in	16 Databases
Publication Volumes and Years included in this study	1999-2012
Types of articles studied	Full Length Articles and Special Issue Articles

The papers were collected from ETS Journal website (http://www.ifets.info/). More specifically, papers from 15 volumes were collected that correspond to 57 issues (Table 3.1). These issues span a time period of 14 years, from January 1999 to December 2012, which is translated into 1041 papers, namely 589 regular papers and 452 special issues papers. All single author papers (249 papers) were excluded from the sample, since the scope of our study is to study the co-authorship network of the ETS Journal. Consequently, we considered in our sample 455 regular papers and 337 special issues papers, which were co-authored by two or more authors. No papers from 1998 were valid for this study, since they did not match the above criteria.

3.2 Research Instrument: NodeXL

From the types of papers previously mentioned, a list of authors who have jointly authored papers was created with the use of NodeXL (http://nodexl.codeplex.com/). NodeXL is a free, open-source template for Microsoft Excel. Its purpose is to enable users to create, edit and visualize network graphs. NodeXL is intended for users with little or no programming experience to allow them to collect and analyze a variety of networks.

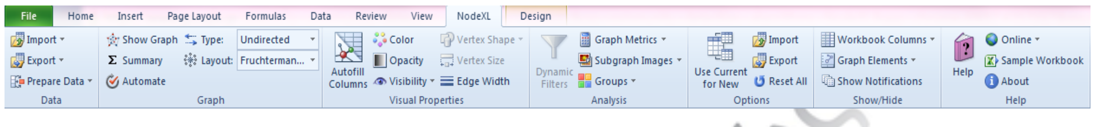

Figure 3.1: The NodeXL User Interface

NodeXL has the following features:

- ✓ Structured interface that coincides with the social network structure

  NodeXL offers a user interface that guides the user in his steps through the creation and analysis of a network. It includes multiple worksheets that represent the vertices, edges and groups of the network.

- ✓ Immediate and accurate visualization

  NodeXL also incorporates a visualisation pane which contains the network graph. The network graph is built based on the data the user provides. For the visualization of a graph, NodeXL allows users to pick from several Force-directed graph drawing layout algorithms such as Fruchterman-Reingold and Harel-Koren. Finally, NodeXL allows the user to change the visual properties (size, color, and opacity) of the graphs elements, based on the metrics that correspond to those elements. This feature enables a more accurate and directed representation.

- ✓ Easy and complete data analysis toolset

  NodeXL contains a library of commonly used graph metrics. Those metrics include centrality and diameter. The aforementioned graph metrics and various visual properties appear as additional columns in the respective worksheets.

- ✓ Various grouping options

  NodeXL offers a variety of grouping options, namely, group by vertex attribute, by connected component, by cluster or motif. Each of these options contains a different set of rules which are implemented into the structure of the different subgroups, resulting in various, equally useful, layouts.

- ✓ Automatic clustering

  NodeXL incorporates a variety of community detection algorithms to allow the user to automatically discover clusters in the social networks. Those algorithms include the Wakita and Tsurumi algorithm, the Girvan-Newman and the Clauset-Newman-Moore algorithm.

- ✓ Differentiation between directed and undirected networks

  Usually the type of network, namely directed and undirected, is taken into consideration in order to calculate in-degree and out-degree. This metric does not apply to undirected networks, thus, it is calculated only when the network relationships are directional. NodeXL applies the differentiation between the two types of networks to other metrics, as well. For example, for the betweenness centrality metric- as it was normalised in previous NodeXL versions- the two different formulas used by NodeXL were:

  $(n-1)(n-2)$ for directed graphs

  $(n-1)(n-2)/2$ for undirected graphs

- ✓ Active support community

NodeXL is supported by a lively community. There is rich documentation, a very active discussion forum and a number of printed works that provide information on the specific tool. All this substantial support makes the tool easily usable for researchers.

3.3 Process

In this section we will present the process which has been followed for our analysis. The process has been divided into two subcategories, namely data import and data preparation.

3.3.1 Data Import

The list was filled in the following way: $author_a$ and $author_b$ who collaborated, were set on the same row. For three authors, $author_a$, $author_b$ and $author_c$ in a joint paper, the following structure was made (Table 3.2).

Table 3.2: First step of data import

Author1	**Author2**
$author_a$	$author_b$
$author_a$	$author_c$
$author_b$	$author_c$

In our effort to make not only the data, but also our understanding of them, more complete, we made a few additions to our table. First of all, the year of publication was added, as a separate column. That way, it was easy to keep track of any positive or negative fluctuation on a yearly basis. To enrich our insights on co-authorship networks in TeL, we also added columns for the keywords of every article studied, as seen in Table 3.3.

Table 3.3: The second step of data import

Author1	**Author2**	**Year**	**Keyword**
$author_a$	$author_b$	xxxx	keyword1
$author_a$	$author_c$	xxxx	keyword1
$author_b$	$author_c$	xxxx	keyword1

By following the above pattern, we synthesized a complex network of scientific collaboration through co-authorship, in which authors are represented by vertices and the co-authorship connections between them are shown in the graph, as edges. There are a total of 3352 collaborations and 1944 unique collaborators, who participate in the network. The network is, therefore, highly complex with a high number of vertices and edges.

Moreover, it is important to mention that if a paper had only one author, the paper was not included into our sample, since the scope of our study was to study the co-authorship network of the ETS Journal.

3.3.2 Data Preparation

After the entry of data, the next step was consequently, its proper and right preparation for the analysis. A number of filters and techniques were implemented on the dataset. For example, the logical function "If" was used in order to find and erase any self-loops, in the following form:

IF([@[Vertex 1]]=[@[Vertex 2]]; 1; 0)

By following the above process, we created the ETS Journal co-authorship network, where authors are represented as vertices and co-authorship connections appear as edges. **Figure** 3.2 provides an overview of the ETS Journal co-authorship graph.

Figure 3.2: The ETS Journal Co-Authorship Graph

As we can notice from **Figure** 3.2, we have used a clustering algorithm, namely, the Clauset-Newman-Moore algorithm. Consequently, in our graph there are various subgroups of authors which are represented by the same color and shape. The smaller subgroups can be found at the bottom part of the graph. This is due to our effort to optimize the visual representation of the graph by placing small groups on the bottom, so as to focus on bigger groups which formulate the core of the co-authorship network, e.g. the giant component (Figure 3.3).

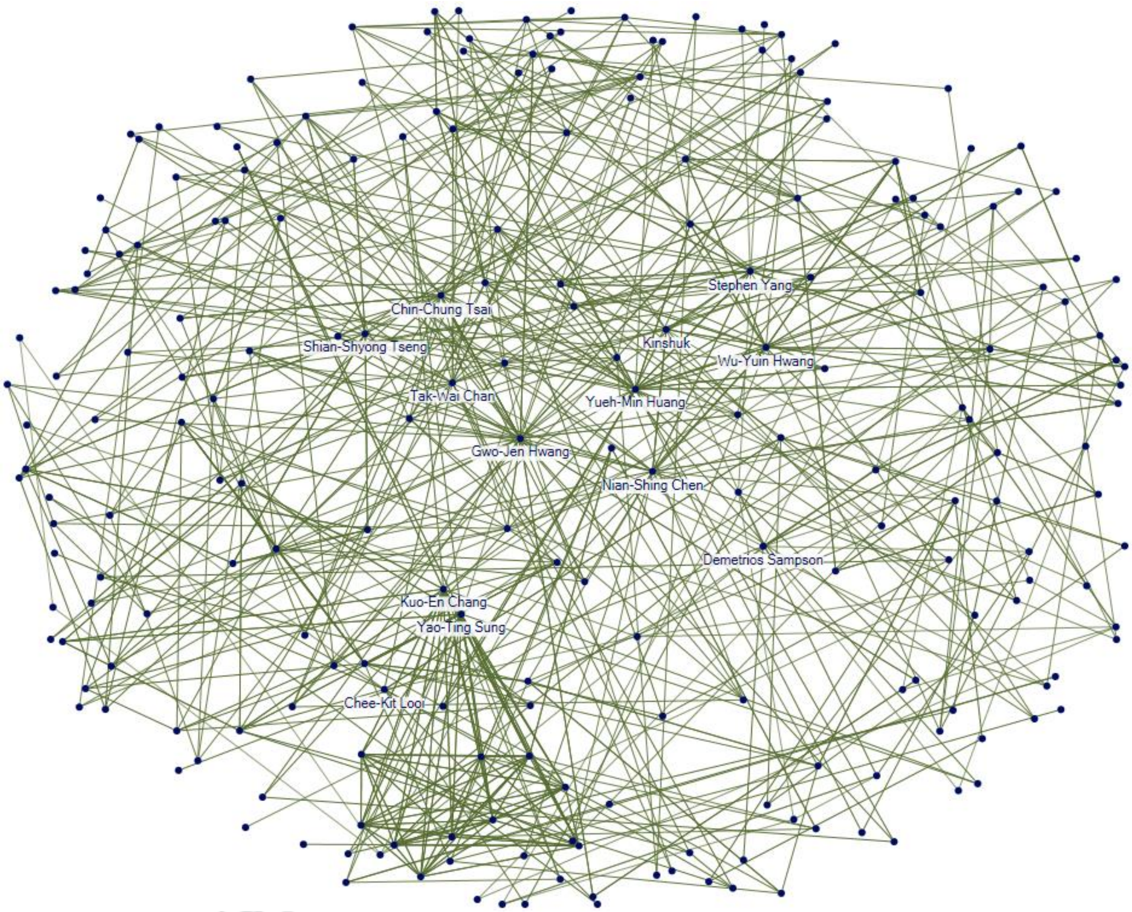

Figure 3.3: The giant component graph

As we notice from Figure 3.3, all authors are represented with a common symbol, i.e., blue disks, since they participate in the same subgroup. This is by far the largest group of our network, which in SNA is called the giant component. The giant component is one of the most vital parts of a network and it presents great interest for researchers.

The typology of the smaller subgroups of our network, created by the clustering algorithm, is presented next through specific examples, in order to provide additional information on the structure of the graph (Figure 3.4, Figure 3.5 and Figure 3.7).

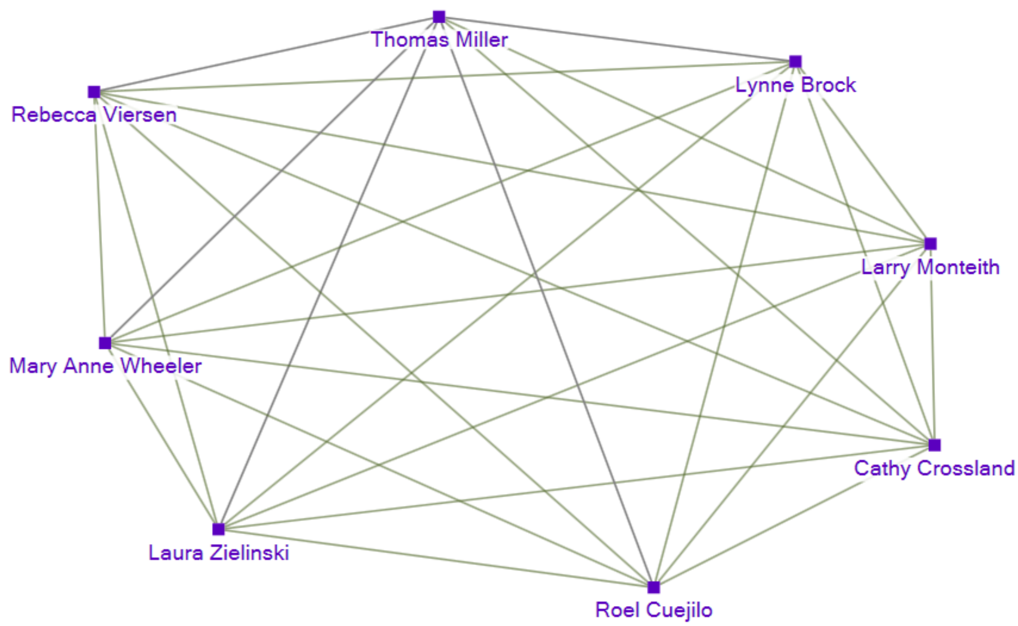

Figure 3.4: Group 23 in expanded form

On one hand, a good example of a small group is group 23. This group, although, it is consisted of several vertices, it only involves one paper. As a result, all the authors are tied with each other, creating a clique, which can be observed in Figure 3.4.

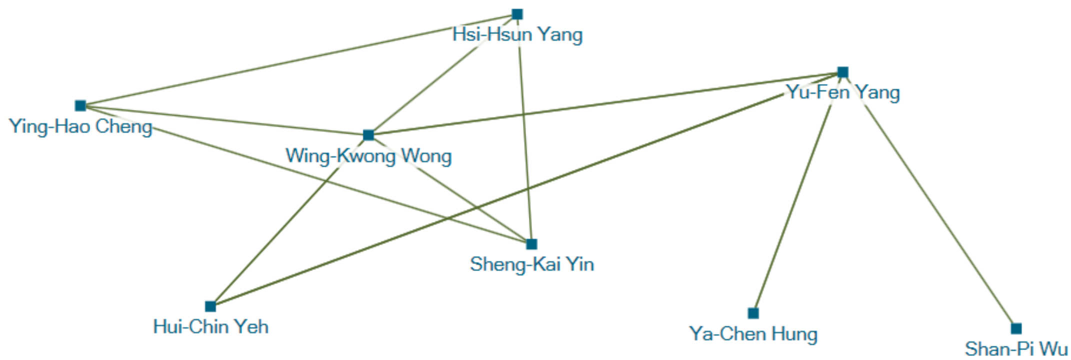

Figure 3.5: Group 24 in expanded form

Another popular structure is that of group 24. This group, which involves 6 papers, has the same amount of authors as group 23 but has followed a totally different formation process. As we notice from Figure 3.5, not all authors have been connected through collaboration. However, because of the concept of triadic closure in social networks, which we presented beforehand, we expect the vertices which are not connected, to connect soon through collaboration.

The smallest subgroups of our network, though, consist of just two vertices. As it can be expected, that is the lowest number of authors that can be involved in a co-authorship network subgroup. As we have already mentioned, we have placed the smaller groups at the bottom of our graph (Figure 3.6).

Figure 3.6: The smallest clusters of the network

In our network those kinds of subgroups are relatively many in number. Specifically, there are 196 such groups that only contain authors tied to each other through their common papers, and have no external connections in our network (Figure 3.6).

An example of such a cluster, group 314, is presented below (Figure 3.7). This cluster contains exactly two authors, who are isolated and present no external connections.

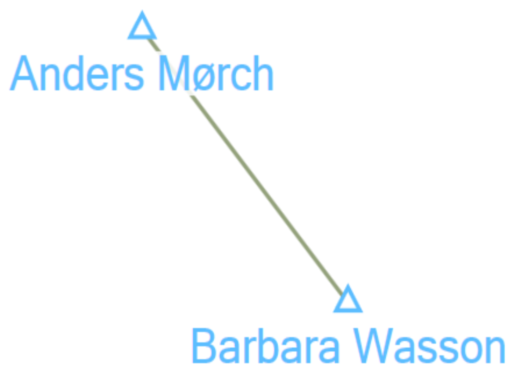

Figure 3.7: Group 314 in expanded form

The aforementioned formations are examples of the main typology of subgroups in the network. We find that the subgroups' internal structure is particularly interesting, especially in the case of the larger components, which we will further analyze.

The type of analysis that we have conducted is exploratory, i.e. there is not a specific hypothesis to test but the conclusions emerge from the analysis itself (de Nooy et al., 2005). It was also taken into consideration, that the collaboration is inter-individual. The analysis is conducted at a micro-level, as our starting and focus points are individual connections.

3.4 SNA Metrics in Co-authorship Networks

Co-authorship network analysis can be achieved by using SNA metrics (Otte & Rousseau, 2002). These metrics are used to identify features and patterns of: (a) the typology of the co-authorship network; and, (b) the centralities of the authors of the network. The former can be addressed by graph metrics (as presented in Table 3.4), whereas the latter can be addressed by vertex specific metrics (as presented in Table 3.5).

Table 3.4: Graph Metrics in Co-Authorship Networks

SNA Metric	Definition (Hansen et al., 2011)	Application in co-authorship networks
Unique Edges	It is used to measure the number of edges, which occur exactly once between two vertices	If combined with the number of vertices (namely, authors), it can denote the average number of co-authors of the publications modelled by the network
Edges with Duplicates	It is used to measure edges that occur more than once between two vertices	It counts the number of repeating collaborations between authors
Total Edges	It is the sum of unique and duplicate edges in the graph	It counts all collaborations that exist in the network
Connected Components	It identifies and measures closed subgroups of vertices	It denotes the number of closed subgroups of

	that are interconnected, but they do not connect with other vertices or groups outside of the specific subgroups	authors, which are not collaborating with other specific subgroups
Maximum Vertices in a Connected Component	It measures the maximum number of vertices that are connected to each other in a component. This component is called "Giant component"	It denotes the number of authors that can be found in the biggest subgroup
Maximum Edges in a Connected Component	It measures the number of edges that exist inside the "Giant component"	It counts the number of collaborations that can be found in the biggest subgroup
Maximum Geodesic Distance (Diameter)	It is used to measure the shortest possible distance between the two farthest vertices in the network.	It denotes the maximum distance that a research idea will need, in order to be spread between the two farthest authors in the network
Average Geodesic Distance	It is used to measure how close in average two vertices are placed, inside the graph of the social network	It denotes the average distance that a research idea will need, in order to be spread between any two authors in the network
Graph Density	It represents the completeness	It denotes the degree of

	of vertices' interconnections. The maximal density is 1 (for complete graphs) and the minimal density is 0	collaboration that takes place within the network
Modularity	It is used to measure the degree of clustering that appears in the graph. This means that if a graph consists of distinct closed subgroups, then it has bigger modularity. Modularity is measured by comparing the number of edges inside a subgroup with the number of edges in the whole network	It denotes the degree of fragmentation of the network to closed subgroups of authors

Table 3.5: Vertex Specific Metrics in Co-Authorship Networks

SNA Metric	**Definition** (Hansen et al., 2011)	Application in co-authorship networks
Degree Centrality	It is used to measure the "popularity" of a vertex. This means that it measures the number of unique edges that	It denotes the existence of authors that collaborate very often with many other authors

	create a connection between this specific vertex and others. Each connection is only counted once, even if it occurs multiple times	
Betweenness Centrality	It is based on the number of shortest paths passing through a vertex. Vertices with high betweenness centrality play the role of connecting different subgroups of the social network	It denotes the existence of authors that connect and strengthen the collaboration between different subgroups of authors
Closeness Centrality	It represents the distance of a vertex to all other in the network by focusing on the geodesic distance from each vertex to all others. Closeness centrality can be interpreted as a metric of how long it will take information to spread from a given vertex to others in the network	It denotes the existence of authors that have collaborated with researchers from a wide variety of research areas and they can quickly spread research ideas across the network.
Eigenvector Centrality	It is used to measure the degree of a vertex and the degree of the vertices with which this vertex is connected. Eigenvector	It denotes the existence of authors that are more likely to receive first new research ideas that are

	centrality can identify vertices that are connected with other "popular" vertices.	spread across the network

As we notice from the previous Tables, the interpretation of every specific SNA metric can be influenced by the special context of each network. Consequently, the description of the meaning of SNA metrics in the context of co-authorship networks is vital to a fuller and more accurate understanding of their measurements.

3.5 Synopsis

In this chapter, we have presented the research methodology that we have followed in order to solve our research problem. The present chapter has been, thus, divided into four consecutive sections.

The first section introduced the sample of our analysis, which is the data extracted from publications of the ETS Journal. We list the reason that prompted us to select the specific sample and present the number of issues and articles studied. The following section describes our research instrument, which is the NodeXL software. We name important features of NodeXL that support and facilitate the research process, as reasons for choosing that particular software.

Next, we present the process of data import and data preparation. We also give an overview of the most representative network structures, in the co-authorship network that we have created. In the last section, we present the particular meaning of SNA metrics in the context of co-authorship networks.

In the subsequent chapter we will utilize these said metrics in order to extract meaningful results and interesting patterns on research collaboration in the TeL field. Those elements will support and demonstrate the effectiveness of the previously proposed solution to our research problem.

Chapter 4
Experiments and Data Analysis

In this chapter, we apply the SNA metrics that have been previously presented to the ETS Journal co-authorship graph. Our analysis consists of two stages, namely (a) network identity: at this stage, we analyze the identity of the current state of the network by applying graph metrics and we perform a ranking and a clustering of authors by applying vertex specific centrality metrics and (b) network evolution: at this stage, we analyze the annual evolution of the network by monitoring the changes of specific graph metrics. This chapter presents the most important findings of the current study.

4.1 Identity of the Network

In this section we will discuss the state of the network in 2012, its main attributes and characteristics, as we reveal them through our graph metrics and vertex-specific metrics analysis. The former addresses the whole network, while the latter provides insights on specific prolific individuals.

4.1.1 Graph Metrics Analysis

Table 4.1 presents the values of the graph metrics that have been applied to the ETS Journal co-authorship network. These values have been calculated by importing our data to NodeXL.

Table 4.1: Graph Metrics for the ETS Journal Co-Authorship Network

Graph Metrics	Values
Graph Type	Undirected

Vertices (Authors)	1944
Unique Edges (Unique Collaborations)	2776 (82,82%)
Edges With Duplicates (Duplicated Collaborations)	576 (17,18%)
Total Edges (Total Collaborations)	3352
Connected Components	468
Maximum Vertices (Authors) in a Connected Component (Giant Component)	226 (11,63%)
Maximum Edges (Collaborations) in a Connected Component (Giant Component)	770 (22,97%)
Maximum Geodesic Distance (Diameter)	9
Average Geodesic Distance	3,766
Graph Density (max value: 1,00)	0,002
Modularity (max value: 1,00)	0,868

As we can notice from Table 4.1, there are 1944 collaborating authors in this network and 2776 unique collaborations. This means that on average an author collaborates with 1,43 other authors. Thus, having excluded the single author papers, we can mention that, on average, the authoring team of an ETS Journal paper consists of 2 or 3 authors. These are slightly lower values comparing to co-authorship networks from other research fields such as biology and physics constructed by Newman (2001), who has found that authors per paper for biology co-authorship network are 3,75 and for physics co-authorship network, authors per paper are 2,53. This provides us with indications that TeL researchers/authors are collaborating in smaller groups compared to other research fields.

Additionally, we can notice that there are 576 duplicated collaborations that correspond to 17,18% of the total collaborations. This means that only 1 out of 6 collaborations have occurred twice or more times. An important issue is the high number of connected components (468) in a network of 1944 vertices. This means that this network includes many subgroups that are collaborating independently from the main core of collaborations (giant component). This claim is also supported by the modularity metric, which is 0,868, since it is relatively elevated. An additional metric that supports this argument is the graph density, which is very low (0,002). This also means that the authors included in the various subgroups of the network are rarely collaborating with authors outside these subgroups. Our previous findings are, finally, consistent with the typology of various subgroups, which was previously presented.

Another interesting element is the fact that the giant component of the network includes 226 authors and 770 collaborations. These values correspond to 11,63% of the total authors and to 22,97% of the total collaborations. This means that authors included in the giant component collaborate more often than the authors who are not part of the giant component.

Other important metrics are the diameter of the network and the average distance of the network. More precisely, the diameter of the network is 9. This value appears to be the same with the diameter of the mathematics co-authorship networks studied by Barabási et al. (2002), as well as with the library and information science co-authorship networks studied by Yan & Ding (2010). This means that the co-authorship network's typology in TeL follows similar patterns with other research fields. Finally, the average distance of the network highlights that 3,766 steps are needed to reach an author within the network.

4.1.2 Vertex-Specific Metrics Analysis

Vertex-specific centrality metrics are the way to locate and study, unique, highly important vertices. In order to extract a first, abstract view of the identity of the networkIn order to analyze insights of the central authors of the ETS co-authorship, Table 4.3 presents the top 30 authors based on degree centrality, betweeness centrality, closeness centrality and eigenvector centrality. Authors appear in the top 30 of all centrality metrics are marked in bold and italic fonts., we have calculated the average vertex-specific metrics for it.

4.1.2.1 *Vertex-specific metrics for the network and the giant component*

However, as it is clearly identified from **Figure** 3.2, as well as from the graph metrics presented in the previous paragraph, the ETS Journal co-authorship network includes many subgroups that are not connected to each other. As a result, centrality metrics (such as betweenness centrality and closeness centrality) might not take accurate values, since these metrics are calculated by taking into account the distance between vertices, which is infinite for disconnected subgroups (Opsahl et al., 2010). In order to examine and overcome this problem, we firstly compare the average centrality metrics of the network to those specifically drawn from the giant component (Table 4.2).

Table 4.2: Vertex-specific metrics for the network and the giant component

Network		**Giant Component**	
Vertex-Specific Metrics	**Value**	**Vertex-Specific Metrics**	**Value**

Average Degree (Degree Centrality)	3,126	Average Degree (Degree Centrality)	5,204
Average Betweenness Centrality	48,275	Average Betweenness Centrality	369,704
Average Closeness Centrality	0,375	Average Closeness Centrality	0,001
Average Eigenvector Centrality	0,001	Average Eigenvector Centrality	0,004

The first row in Table 4.2 shows the average value for the degree centrality metric. It indicates that the average number of connections for the researchers in our network is approximately 3. The average degree value for the giant component, though, is approximately 5. Consequently, we can conclude that there are a greater number of collaborations found in the giant component.

In relevance to the betweenness centrality, the difference between the two average values is significant. Specifically, we have a 48,275 value for the network average betweenness, compared to the 369,704 value for the corresponding giant component metric. It appears that in the giant component network there a lot more bridges that connect and strengthen the collaboration between different subgroups.

In the third row we come across the corresponding values for the closeness centrality metric. The difference of the two values found in this row, is also important. In this case though, the average closeness centrality of the network is much higher. This indicates that the authors participating in the complete network are much closer to each other, on average, than the ones participating just in the giant component, which is a logical paradox.

Finally, the average eigenvector centrality is slightly elevated in the giant component column, a fact that can be easily associated with the previously mentioned differences in the degree centrality values.

As mentioned beforehand, in Table 4.2, there can be observed some divergences. These divergences are more obvious for the betweeneess and closeness centrality metrics. Our findings can be easily explained by the method which is used to calculate these two metrics, as discussed in the previous paragraphs. Due to the major differences between the network and the giant component values, caused by the big number of disconnected subgroups (Opsahl et al., 2010) we next discuss the application of the vertex- specific centrality metrics only on the giant component.

4.1.2.2 *Distribution of vertex-specific metrics for the giant component*

Figure 4.1 presents the distribution of the calculated centrality metrics for the authors of the giant component.

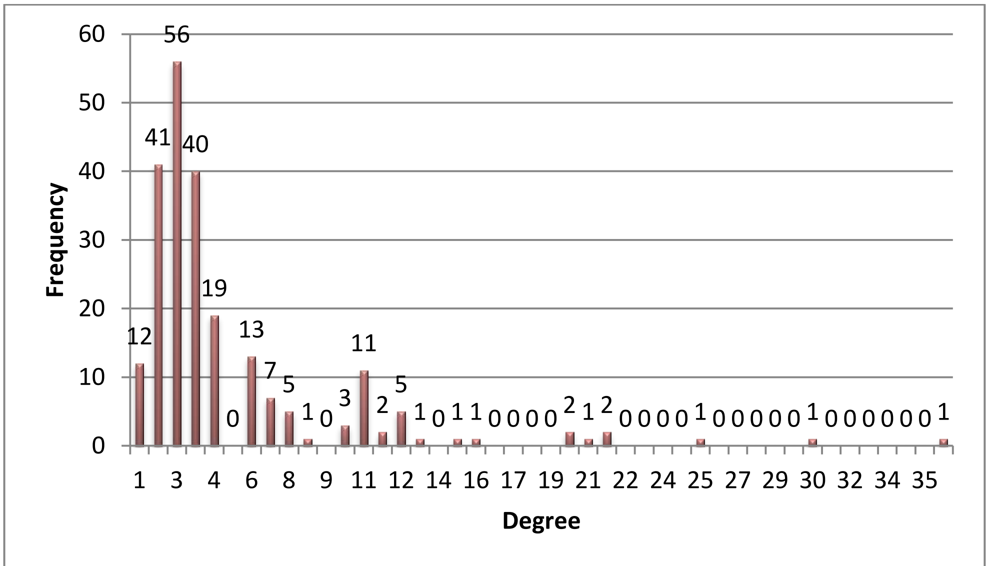

Figure 4.1: Frequency distribution of degree centrality

The degree centrality, whose distribution has been illustrated in

Figure 4.1, is one of the most important vertex specific centrality metrics. As we mentioned beforehand, it denotes authors that collaborate very often with many other authors. As we notice from

Figure 4.1 the frequency of degree centrality follows power-law distribution, where most authors have low centrality value while a few authors have high degree centrality value.

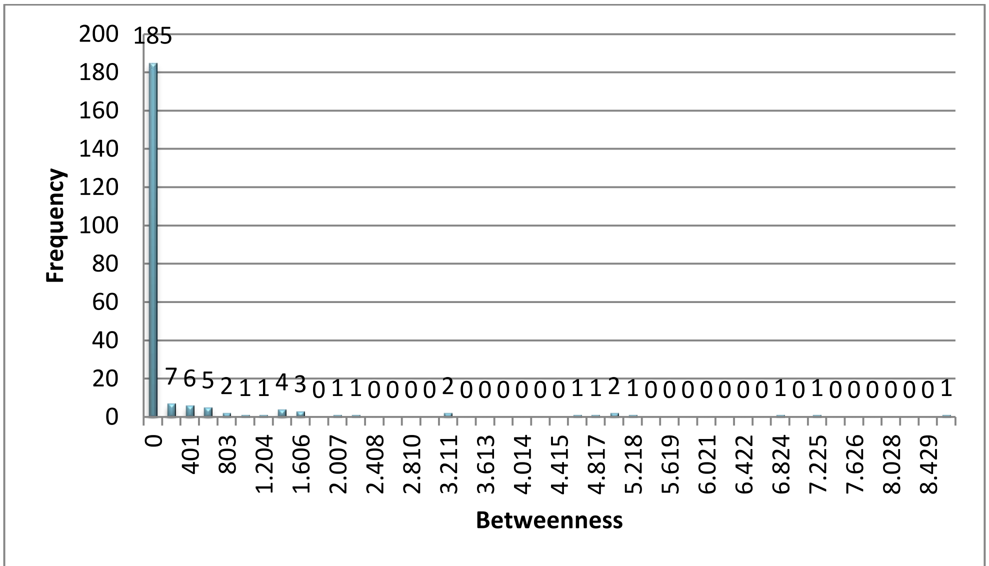

Figure 4.2: Frequency distribution of betweenness centrality

In our previous analysis we have concluded that betweenness centrality denotes authors that connect and strengthen the collaboration between different subgroups. The frequency of betweeness centrality, in our network, follows power-law distribution where most authors have low centrality value while a few authors have high betweenness centrality values (Figure 4.2).

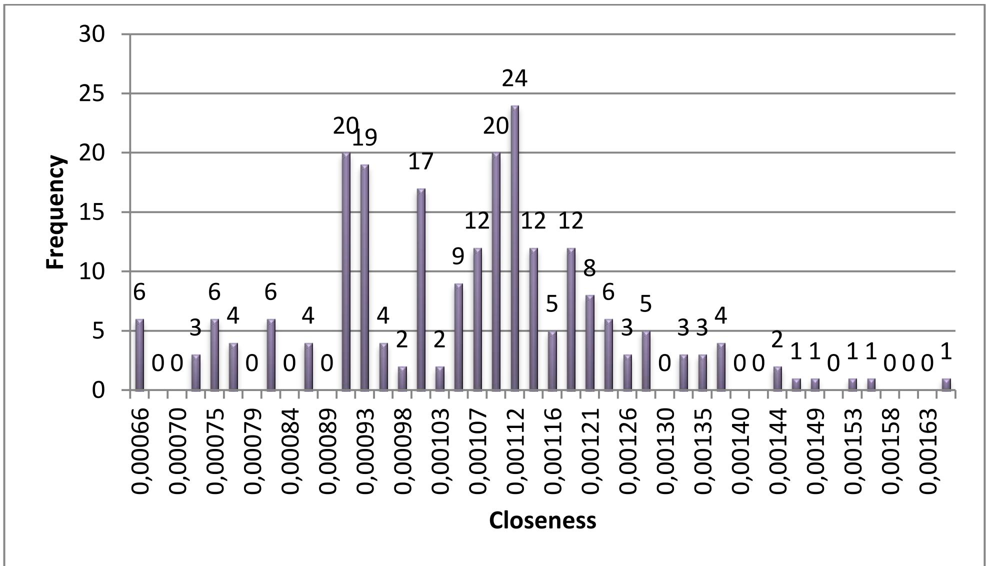

Figure 4.3: Frequency distribution of closeness centrality

The closeness centrality denotes authors that have collaborated with researchers from a wide variety of research areas and they can quickly spread research ideas across the network. The distribution of closeness centrality in our network (Figure 4.3) follows the normal curve distribution.

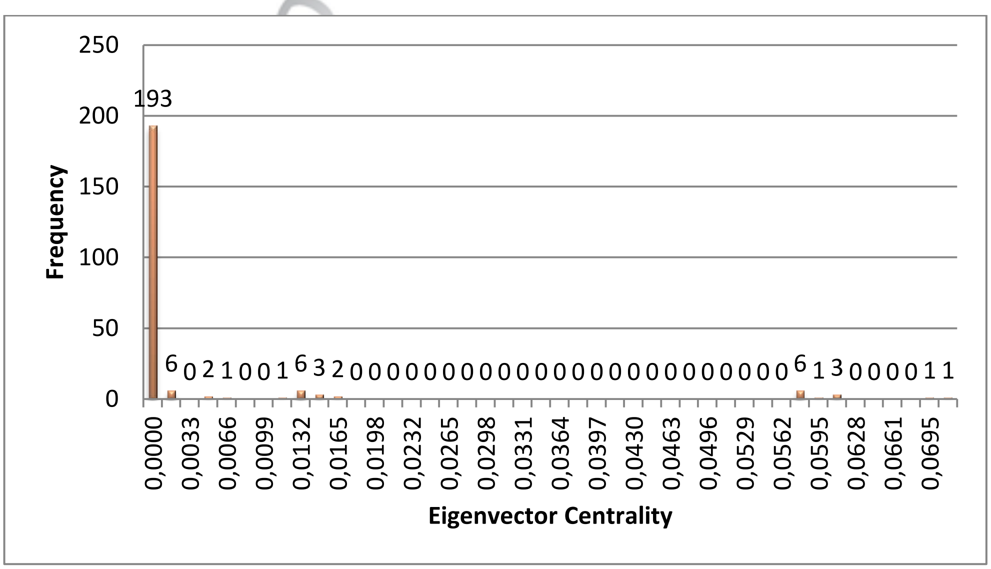

Figure 4.4: Frequency distribution of eigenvector centrality

Finally, eigenvector centrality denotes authors that more likely to receive first new research ideas that are spread across the network. In our case (Figure 4.4), the frequency of eigenvector centrality follows power-law distribution where most authors have low centrality values while a few authors have high eigenvector centrality values.

These distributions appear to be similar with frequency distributions of centralities in mathematics and neuro-science co-authorship networks studied by Barabási et al. (2002), in library and information science co-authorship networks studied by Yan & Ding (2010) and in e-Government co-authorship networks studied by Ernan & Todorovski (2010). This provides us with indications that the authors/researchers of the ETS Journal collaborate by following similar patterns with other research fields.

That claim can also be supported by the reversed degree centrality (Figure 4.5).

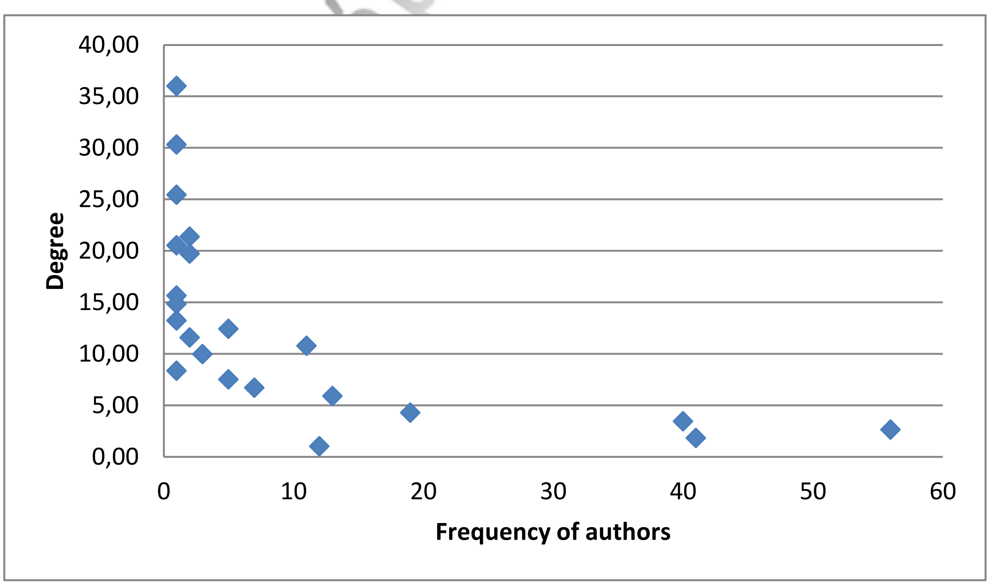

Figure 4.5: Degree to Frequency of authors

What is easily noticeable in Figure 4.5, is the vast amount of authors that are located in positions with degree value lower than 1. At the same time, there are very few authors who are measured to have degree values higher than 15. Those really significant individuals have many connections among the authors and strongly influence the structure of collaborations.

The two elements of the graph, called a “long tail” and a “short head” respectively, are combined to form this unbalanced distribution. That distribution, as illustrated in Figure 4.5, is called preferential attachment and has been studied as a common element of social networks in the past (Price, 1976; Barabási & Albert, 1999). The long-tailed degree distribution also seems to be common in other journal analysis studies (Barabási et al., 2002).

4.1.2.3 *The top 30 authors based on vertex-specific metrics*

In order to analyze insights of the central authors of the ETS co-authorship, Table 4.3 presents the top 30 authors based on degree centrality, betweeness centrality, closeness centrality and eigenvector centrality. Authors appear in the top 30 of all centrality metrics are marked in bold and italic fonts.

Table 4.3: Top 30 Authors based on Centrality Metrics

Rank	Degree Centrality	Value	Betweenness Centrality	Value	Closeness Centrality	Value	Eigenvector Centrality	Value
1	Gwo-Jen Hwang	36	Kinshuk	8629,879	Nian-Shing Chen	0,00165	Kuo-En Chang	0,071
2	Yueh-Min Huang	31	Tak-Wai Chan	7230,966	Stephen Yang	0,00157	Yao-Ting Sung	0,070
3	Chin-Chung Tsai	26	Nian-Shing Chen	6965,082	Kinshuk	0,00155	Chun-Hua Chen	0,062
4	Nian-Shing Chen	22	Stephen Yang	5293,760	Gwo-Jen	0,0015	Yun-Cheng	0,062

					Hwang		Luo	
5	Kuo-En Chang	22	Gwo-Jen Hwang	5145,543	Yueh-Min Huang	0,00148	Chenn-Jung Huang	0,062
6	Yao-Ting Sung	21	Wu-Yuin Hwang	5115,723	Chin-Chung Tsai	0,00146	Ming-Chou Liu	0,060
7	Tak-Wai Chan	20	Kuo-En Chang	4838,570	Wu-Yuin Hwang	0,00145	Tz-Hau Huang	0,059
8	Wu-Yuin Hwang	20	Yueh-Min Huang	4678,167	Tak-Wai Chan	0,00139	Hung-Yen Shen	0,059
9	Stephen Yang	16	Chin-Chung Tsai	3354,892	Tzu-Chien Liu	0,00139	Kuo-Liang Huang	0,059
10	Shian-Shyong Tseng	15	Chee-Kit Looi	3223,000	Irene Chen	0,00138	Jia-Jian Liao	0,059
11	Kinshuk	14	Chiu-Pin Lin	2349,809	Chun-Wang Wei	0,00138	Kai-Wen Hu	0,059
12	Chun-Hua Chen	13	Tzu-Chien Liu	2131,545	Ting-Ting Wu	0,00137	Tun-Yu Chang	0,059
13	Yun-Cheng Luo	13	Demetrios Sampson	1760,500	Hui-Chun Chu	0,00137	Yi-Ta Chuang	0,018
14	Chenn-Jung Huang	13	Michael Spector	1736,000	Chin-Yeh Wang	0,00136	Hong-Xin Chen	0,018
15	Ming-Chou Liu	13	Shian-Shyong Tseng	1637,500	Hsin Ning Jessie Ho	0,00134	Chiu-Pin Lin	0,016
16	Chin-Yeh Wang	13	Chin-Yeh Wang	1579,189	Yu-Chen Hsu	0,00134	Yu-Ju Lan	0,015
17	Hui-Chun Chu	12	Minhong Wang	1538,000	Hsinyi Peng	0,00133	Ning-chun Tan	0,015
18	Ting-Ting Wu	12	Ben Chang	1532,000	Rustam Shadiev	0,0013	Yi-Hsuan Lee	0,014
19	Tz-Hau Huang	11	Toshio Okamoto	1526,000	Tony Kuo	0,0013	Wen-Cheng Yu	0,013
20	Hung-Yen Shen	11	Hsinyi Peng	1401,532	Jian-Jie Dung	0,0013	Rey-Bin Chang	0,013

21	Kuo-Liang Huang	11	David Hung	1100,000	Yi-Lun Yang	0,0013	Shui-Cheng Lin	0,013
22	Jia-Jian Liao	11	Alexandra Cristea	887,000	Kuo-En Chang	0,00129	Yu-Ru Hong	0,013
23	Kai-Wen Hu	11	Yao-Ting Sung	871,267	Yen-Hung Kuo	0,00128	Yu-Lung Chen	0,013
24	Tun-Yu Chang	11	Maiga Chang	775,803	Yu-Lin Jeng	0,00126	Huei-Tse Hou	0,012
25	Ben Chang	11	Ting-Ting Wu	710,495	Qing Tan	0,00126	Wu-Yuin Hwang	0,008
26	Gwo-Dong Chen	11	Ying-Shao Hsu	668,000	Jeff Huang	0,00125	Jhen-Yu Wang	0,005
27	Tzu-Chien Liu	11	Chu-Sing Yang	666,000	Indy Hsiao	0,00125	Ming-Hsiao Chi	0,005
28	Chee-Kit Looi	11	Lung-Hsiang Wong	666,000	Taiyu Lin	0,00125	Tak-Wai Chan	0,002
29	Demetrios Sampson	11	Jun-Ming Su	536,000	Yi-Chan Deng	0,00125	Yueh-Min Huang	0,002
30	Jun-Ming Su	10	Ming-Chou Liu	447,000	Baw-Jhiune Liu	0,00124	Chee-Kit Looi	0,002

The first column of Table 4.3 shows the ranking of the top 30 authors based on the degree score. The top 3 authors in terms of degree centrality are: Gwo-Jen Hwang (15 papers, National Taiwan University of Science and Technology, Taiwan), Yueh-Min Huang (15 papers, National Cheng Kung University, Taiwan) and Chin-Chung Tsai (13 papers, National Taiwan University of Science and Technology, Taiwan). These authors are collaborating frequently with many other authors and they can be viewed as popular authors.

With regard to betweeness centrality (second column of Table 4.3), the top 3 authors are: Kinshuk (8 papers, Athabasca University, Canada) followed by Tak-Wai Chan (5 papers, National Central University, Taiwan) and Nian-Shing Chen (11 papers, National Sun Yat-sen University, Taiwan). Thus, these authors can be viewed as central in the ETS Journal co-authorship network, since, being on the shortest collaboration paths between other authors, means that they (a) collaborate directly with a relatively high number of researchers and (b) they connect secluded groups or authors with the giant component, acting as bridges.

In terms of closeness centrality (third column of Table 4.3), the scores of the top 30 authors were very close with the leaders being Nian-Shing Chen (11 papers, National Sun Yat-sen University, Taiwan) followed by Stephen Yang (7 papers, National Central University, Taiwan) and Kinshuk (8 papers, Athabasca University, Canada). These authors have collaborated with researchers from a wide range of research areas in the field of TeL and this means that they can be viewed as influential authors, who can quickly spread research ideas across the network.

Finally, the last column of Table 4.3 shows the ranking of the top 30 authors based on the eigenvector score. The top 3 authors in terms of eigenvector centrality are: Kuo-En Chang (7 papers, National Taiwan Normal University, Taiwan), Yao-Ting Sung (7 papers, National Taiwan Normal University, Taiwan) and Chun-Hua Chen (2 papers, National Dong Hwa University, Taiwan). These authors are connected to many other authors, who are well connected and popular, and, thus, they are more likely to co-author new papers with those authors and are directly introduced to new concepts or research ideas.

What we can notice from the previous analysis is that two influential factors for research collaboration, namely rate of productivity and social proximity, appear to affect our network. Firstly, we can notice that indeed the productivity rate correlates with the collaboration rate of researchers. The high numbers of papers co-authored by the top 3 authors for each metric indicate that, in our network, as well, research collaboration can be associated with high productivity when it comes to the numbers of papers produced. The social proximity factor of research collaboration plays also an important role in our network. Specifically, we notice that almost all the authors that appear in the top 3 ranks hold an associate professor or higher academic position.

It is also important to mention that two of the top three positions regarding betweenness centrality and closeness centrality have been taken by the two out of three editors of the ETS Journal, namely Kinshuk and Nian-Shing Chen. This means that ETS editors play an important role to the collaborations that are developed within the ETS Journal co-authorship network, as they connect various sub-groups (as identified by the betweenness centrality) and they can receive the information spread within the network very quickly (as identified by the closeness centrality). Moreover, another interesting fact of the top 3 authors according to the different centrality metrics is that all of them are from Taiwan except from Kinshuk, who is from Canada. This means that the ETS co-authorship network includes a strongly connected subgroup of authors, who are from Taiwan and they collaborate frequently (as indicated by degree centrality), diversely (as indicated by betweenness centrality) and widely (as indicated by closeness and eigenvector centrality). This finding is also consistent with the spatial proximity factor that influences research collaboration, in a manner that we have previously presented.

This claim is also supported by the country of origin of the authors that appear in the top 30 of all centrality metrics and they are marked in bold and italic fonts. These authors are: Yueh-Min Huang (National Cheng Kung University, Taiwan), Kuo-En Chang (National Taiwan Normal University, Taiwan), Tak-Wai Chan (National Central University, Taiwan) and Wu-Yuin Hwang (National Central University, Taiwan). As we can notice, all these authors are from Taiwan but come from various universities. We also notice that these authors can be found on the top 8 rankings for both degree and betweenness centralities. This gives an indication that research collaboration thrives between Taiwanese researchers, especially since national research policy in Taiwan promotes and facilitates research collaboration between researchers from different universities and/or research centers.

4.1.2.4 Correlation between Number of Citation and Centrality Metrics for the top 30 authors

In order to identify whether authors' ranking based on the centrality metrics can play an important role to the recognition of their research work, we studied the correlation of the centrality metrics of the top 30 authors with the number of citations to their publications. Table 4.4 presents the top 30 author per centrality metric, as well as their citations. It is worth mentioning that the citations of top 30 authors were calculated form Google Scholar service (http://scholar.google.gr/) and these are the numbers that were available on December 23, 2013.

Table 4.4: Top 30 Authors based on Centrality Ranking along with Number of Citations

Rank	Degree Centrality	Value	Citation Number	Betweenness Centrality	Value	Citation Number	Closeness Centrality	Value	Citation Number	Eigenvector Centrality	Value	Citation Number
1	Gwo-Jen Hwang	36	3393	Kinshuk	8629,879	3676	Nian-Shing Chen	0,00165	1761	Kuo-En Chang	0,071	1620
2	Yueh-Min Huang	31	4999	Tak-Wai Chan	7230,966	2848	Stephen Yang	0,00157	3226	Yao-Ting Sung	0,070	1375
3	Chin-Chung Tsai	26	7321	Nian-Shing Chen	6965,082	1761	Kinshuk	0,00155	3676	Chun-Hua Chen	0,062	1272
4	Nian-Shing Chen	22	1761	Stephen Yang	5293,760	3393	Gwo-Jen Hwang	0,0015	3393	Yun-Cheng Luo	0,062	200
5	Kuo-En Chang	22	1620	Gwo-Jen Hwang	5145,543	510	Yueh-Min Huang	0,00148	4999	Chenn-Jung Huang	0,062	614

6	Yao-Ting Sung	21	1375	Wu-Yuin Hwang	5115,723	3226	Chin-Chung Tsai	0,00146	7321	Ming-Chou Liu	0,060	225
7	Tak-Wai Chan	20	2848	Kuo-En Chang	4838,570	1620	Wu-Yuin Hwang	0,00145	510	Tz-Hau Huang	0,059	23
8	Wu-Yuin Hwang	20	510	Yueh-Min Huang	4678,167	4999	Tak-Wai Chan	0,00139	2848	Hung-Yen Shen	0,059	31
9	Stephen Yang	16	3226	Chin-Chung Tsai	3354,892	7321	Tzu-Chien Liu	0,00139	808	Kuo-Liang Huang	0,059	393
10	Shian-Shyong Tseng	15	2185	Chee-Kit Looi	3223,000	1829	Irene Chen	0,00138	507	Jia-Jian Liao	0,059	37
11	Kinshuk	14	3676	Chiu-Pin Lin	2349,809	496	Chun-Wang Wei	0,00138	217	Kai-Wen Hu	0,059	107
12	Chun-Hua Chen	13	1272	Tzu-Chien Liu	2131,545	808	Ting-Ting Wu	0,00137	2776	Tun-Yu Chang	0,059	29
13	Yun-Cheng Luo	13	44	Demetrios Sampson	1760,500	2121	Hui-Chun Chu	0,00137	774	Yi-Ta Chuang	0,018	191
14	Chenn-Jung Huang	13	614	Michael Spector	1736,000	4353	Chin-Yeh Wang	0,00136	225	Hong-Xin Chen	0,018	162
15	Ming-Chou Liu	13	225	Shian-Shyong Tseng	1637,500	2185	Hsin Ning Jessie Ho	0,00134	32	Chiu-Pin Lin	0,016	496
16	Chin-Yeh Wang	13	225	Chin-Yeh Wang	1579,189	225	Yu-Chen Hsu	0,00134	357	Yu-Ju Lan	0,015	156
17	Hui-Chun Chu	12	774	Minhong Wang	1538,000	709	Hsinyi Peng	0,00133	406	Ning-chun Tan	0,015	3
18	Ting-Ting Wu	12	2776	Ben Chang	1532,000	2171	Rustam Shadiev	0,0013	54	Yi-Hsuan Lee	0,014	1207
19	Tz-Hau Huang	11	23	Toshio Okamoto	1526,000	2006	Tony Kuo	0,0013	598	Wen-Cheng Yu	0,013	1138
20	Hung-Yen Shen	11	31	Hsinyi Peng	1401,532	406	Jian-Jie Dung	0,0013	39	Rey-Bin Chang	0,013	28
21	Kuo-Liang	11	393	David Hung	1100,000	132	Yi-Lun	0,0013	45	Shui-Cheng	0,013	68

	Huang						Yang			Lin		
22	Jia-Jian Liao	11	37	Alexandra Cristea	887,000	2355	Kuo-En Chang	0,00129	1620	Yu-Ru Hong	0,013	26
23	Kai-Wen Hu	11	107	Yao-Ting Sung	871,267	1375	Yen-Hung Kuo	0,00128	370	Yu-Lung Chen	0,013	201
24	Tun-Yu Chang	11	29	Maiga Chang	775,803	446	Yu-Lin Jeng	0,00126	737	Huei-Tse Hou	0,012	493
25	Ben Chang	11	2171	Ting-Ting Wu	710,495	2776	Qing Tan	0,00126	200	Wu-Yuin Hwang	0,008	510
26	Gwo-Dong Chen	11	3669	Ying-Shao Hsu	668,000	325	Jeff Huang	0,00125	441	Jhen-Yu Wang	0,005	9
27	Tzu-Chien Liu	11	808	Chu-Sing Yang	666,000	845	Indy Hsiao	0,00125	34	Ming-Hsiao Chi	0,005	29
28	Chee-Kit Looi	11	1829	Lung-Hsiang Wong	666,000	809	Taiyu Lin	0,00125	282	Tak-Wai Chan	0,002	2848
29	Demetrios Sampson	11	2121	Jun-Ming Su	536,000	373	Yi-Chan Deng	0,00125	206	Yueh-Min Huang	0,002	4999
30	Jun-Ming Su	10	373	Ming-Chou Liu	447,000	225	Baw-Jhiune Liu	0,00124	597	Chee-Kit Looi	0,002	1829

Based on the data presented in Table 4.4, we calculated the Pearson's correlation coefficient between number of citations and centrality metrics. The results are presented in Table 4.5.

Table 4.5: Correlation between Number of Citation and Centrality Metrics

Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
0,60*	**0,44***	**0,62***	**-0,20***

*** Correlation is significant at the 0,01 level.**

As we can notice from Table 4.5, three centrality metrics (namely, degree centrality, betweenness centrality and closeness centrality) have significant correlation with the number of citations at the 0,01 level, with closeness centrality as the highest. On the other hand, there is no correlation between the number of citations and eigenvector centrality. As a result, we can claim that only degree centrality, betweenness centrality and closeness centrality can possibly be used as supplementary indicators for assessing author's scientific recognition, providing alternative perspectives to the current methods, such as h-index and i10-index.

But even between these three metrics, we can detect a variation of values. The correlation values are categorized in three subcategories (Cohen, 1988; Weinberg & Abramowitz, 2008). Those subcategories include weak, moderate and strong correlation, and measure the strength of a relationship. Every subcategory corresponds to a specific range of values. According to the values presented in Table 4.5, degree and closeness values fall under the subcategory of "strong" correlation, while the betweenness correlation is characterized as moderate. The absolute value of the correlation coefficient is an effect size that summarizes the strength of the relationship, and although correlation is significant at the 0,01 level, this scale contributes to the understanding of the relationship that is being described. Given the above results, we can verify that high collaboration rate correlates - with some limitations - with citation number, as proposed in the relative literature.

Still, as it was proved by Anscombe (1973) with examples of scatters that measured the same degree value but presented a totally different end result in relevance to the positioning of the scatters, the value of the correlation coefficient (r), is not the only factor that should be taken into consideration. What we should also consider as a valid factor of influence for the correlation coefficient is the visual placement and representation of those elements in the form of a scatter (Figure 4.6- Figure 4.9).

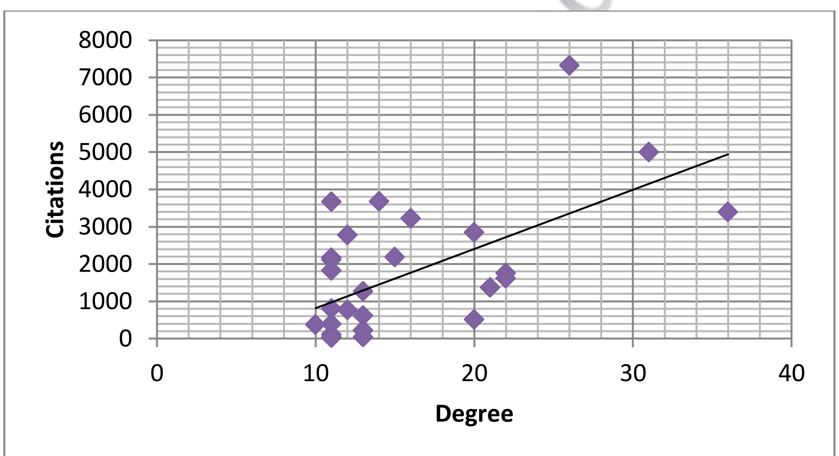

Figure 4.6: Citations to Degree

As we notice from Figure 4.6, we are presented with a clear positive correlation between degree and citation measures. It is a relatively focused scatter when we compare it to the linear trend line which is presented in the graph.

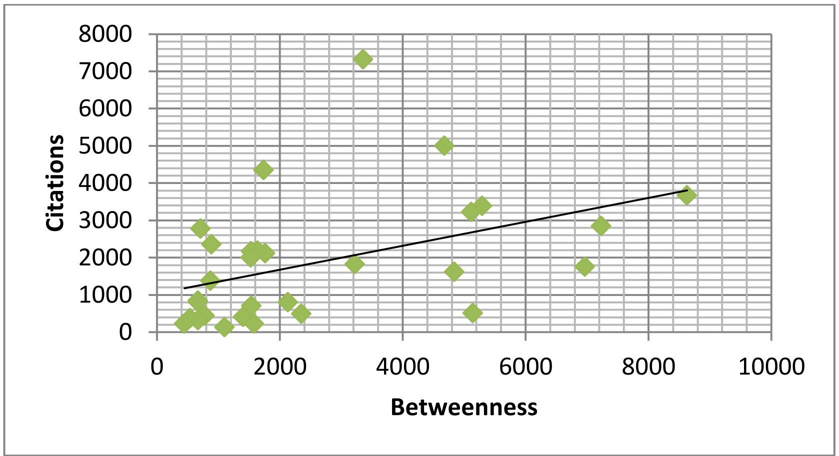

Figure 4.7: Citations to Betweenness

As we can notice from Figure 4.7 there is a moderate correlation between the number of citations and betweenness centrality. We notice that in this case the points spread wider from the linear trend line. Consequently, in this case the correlation is smaller.

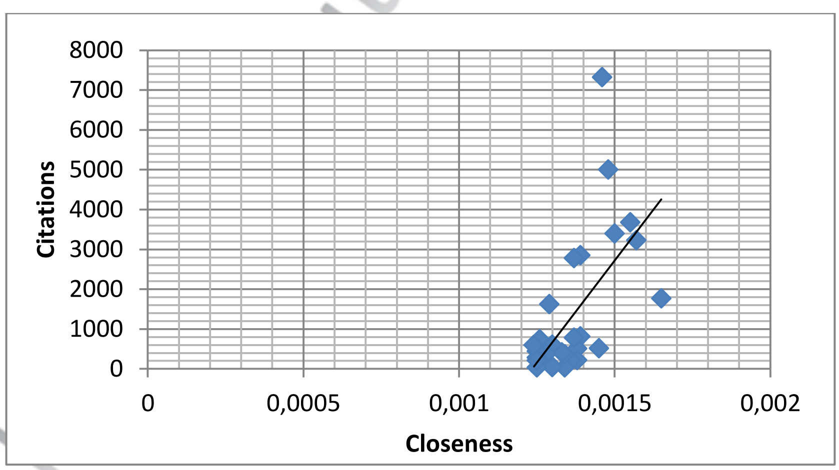

Figure 4.8: Citations to Closeness

As we can notice from Figure 4.8 there is a clear positive correlation between the number of citations and closeness centrality. Most points are strictly focused on the trend line, with two exceptions.

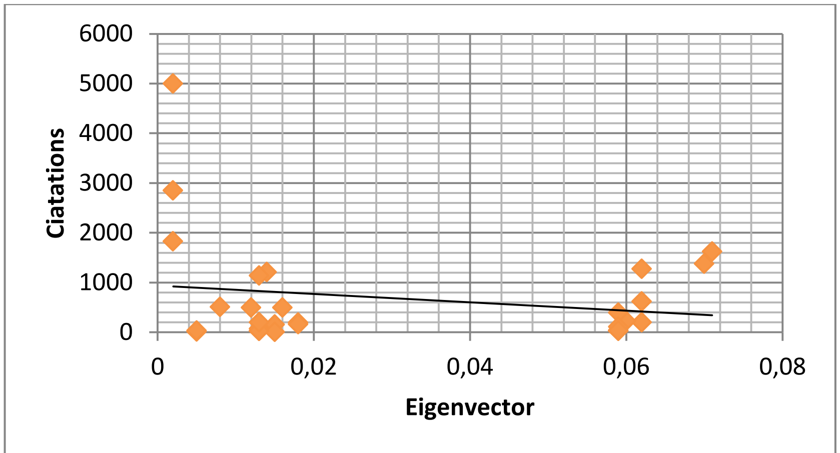

Figure 4.9: Citations to Eigenvector

As we can notice from Figure 4.9 there is no strong correlation between the number of citations and eigenvector centrality. However, we can notice a weak negative correlation.

4.1.3 Author's Clustering

In this section and by using NodeXL's clustering algorithm, which clusters the vertices of a social network based on how they are connected, we perform a clustering of the authors included in the ETS Journal co-authorship network. The scope of this process is to identify major clusters of authors, as well as the papers that they have co-authored towards identifying key research areas of collaboration, of the papers published to ETS Journal. In order to achieve this, we focused on the papers' keywords of each cluster and we have used a word count tool (http://www.writewords.org.uk/word_count.asp) for measuring the frequency of keywords in the papers of each cluster. This will facilitate us to identify relevant research areas for each cluster. Table 4.6 presents the four (4) major authors' clusters that have been identified. It is worth mentioning that we have focused on those clusters that include at least 20 papers (approximately 2,50 % of our sample), in order to have a substantial number of keywords to analyze. Clusters after cluster 4 include less than 6 papers and for this reason, they were not analyzed.

Table 4.6: ETS Journal Major Authors' Clusters

No	Authors (total authors: 1944)	Papers (total papers:792)	Top 3 Frequent Papers' Keywords	Total Counts
Cluster 1	162 (8%)	85 (11%)	mobile learning	24
			ubiquitous learning	17
			collaborative learning	7
Cluster 2	83 (4%)	30 (4%)	learning design / educational modelling / IMS learning design	9
			learning object metadata	7
			open specifications/learning environments	5
Cluster 3	46 (2%)	22 (3%)	computer supported collaborative learning	4
			problem based learning	3
			web based learning	3
Cluster 4	16 (0,8%)	21 (3%)	learning objects	6
			adaptive learning	5
			personalized learning	4

As we can notice from Table 4.6, cluster 1 appears to focus on "*Wireless, Mobile and Ubiquitous Technologies for Learning*" according to the top 3 frequent keywords identified in its papers with 48/85 papers related to this research area. Cluster 2 appears to focus on "*International Standards and Specifications for Learning Technologies*" with 21/30 papers related to this research area. Cluster 3 appears to focus on "*Computer Supported Collaborative Learning*" and Cluster 4 on "*Adaptive and Personalized Technology-Enhanced Learning*", with 10/22 and 15/21 papers respectively.

These research areas have also been identified by Kinshuk et al. (2013) as the main research areas, when analyzing the research topics of highly cited papers published in the ETS Journal for the period from 2003 to 2010. This means that established clusters within the ETS Journal co-authorship network with an adequate number of published papers can attract high numbers of citations and thus become influential in the research community.

Hsu et al., 2012 have also come to similar conclusions from their content analysis of five Social Sciences Citation Index (SSCI) journals, namely the British Journal of Educational Technology, Computers & Education, Educational Technology Research & Development, Educational Technology & Society, the Journal of Computer Assisted Learning. Their study included 2,976 articles published in the technology-based learning (TBL) field, from 2000 to 2009. Similar themes between this study and our findings are "Mobile and Ubiquitous Learning", "Computer Supported Collaborative Learning" and "Adaptive and Personalized Technology-Enhanced Learning: Knowledge and Competencies Management".

4.2 Co-authorship Network Evolution

In this section, we study the evolution of the ETS Journal co-authorship network. With this process, we can understand the current status of the network and we can also make predictions about the future status of the network.

4.2.1 Annual distribution of authors and collaborations

In Figure 4.10 and Figure 4.11 we are presented with the annual accumulative evolution of ETS authors and collaborations from 1999 to 2012. These Figures provide us with an important first impression of the network evolution.

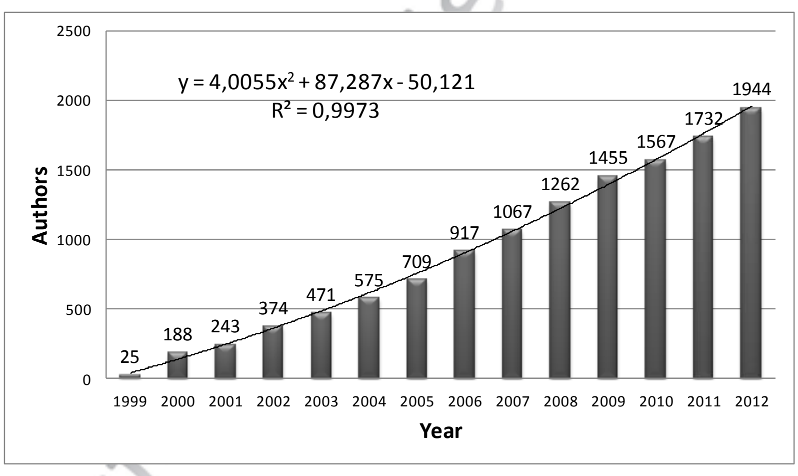

Figure 4.10: Annual accumulative evolution of authors

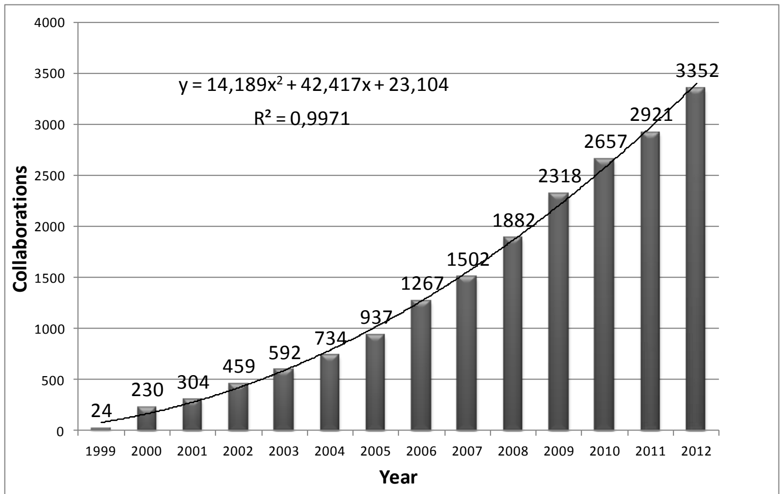

Figure 4.11: Annual accumulative evolution of collaborations

As we can notice from Figure 4.10 and Figure 4.11 the number of authors and collaborations increases gradually. The two curves fit y = 4,0055x^2 + 87,287x - 50,121 and y = 14,189x^2 + 42,417x + 23,104 respectively, with coefficient of determination R^2=0,9973 and R^2=0,9971. This result indicates that there is a growing tendency for authors and collaborations is expected to increase approximately with these curves in the coming years.

Table 4.7 presents the properties of the evolving ETS co-authorship network from 1999 to 2012. It includes information about the annual accumulative distribution of authors and collaborations for the ETS co-authorship network, as well as for its giant component. Moreover, for the giant component, we have calculated the annual ratio of its authors per annual total number of authors, as well as the annual ratio of its collaborations per annual total number of collaborations. Finally, we have calculated the annual ratio of collaborations for the entire network and the annual average distance for the ETS co-authorship network and its giant component.

Table 4.7: Properties of the evolving ETS co-authorship network from 1999 to 2012

Year	Accum. Number of Authors	Accum. Number of Collab.	Ratio of Collab.	Avg. Distance	Giant Component				
					Accum. Number of Authors	Ratio per Total Authors (%)	Accum. Number of Collab.	Ratio per Total Collab. %	Avg. Distance
1999	25	24	0,96	0,76	0	0,00%	0	0,00%	0,00
2000	188	230	1,22	0,76	6	3,19%	9	3,91%	0,67
2001	243	304	1,25	0,76	12	4,94%	14	4,61%	0,72
2002	374	459	1,23	0,79	16	4,28%	31	6,75%	0,97
2003	471	592	1,26	0,84	21	4,46%	37	6,25%	1,01
2004	575	734	1,28	0,93	33	5,74%	61	8,31%	1,62
2005	709	937	1,32	1,05	41	5,78%	74	7,90%	1,61
2006	917	1267	1,38	1,26	56	6,11%	108	8,52%	1,64
2007	1067	1502	1,41	1,46	71	6,65%	131	8,72%	2,09
2008	1262	1882	1,49	2,69	125	9,90%	328	17,43%	3,46
2009	1455	2318	1,59	2,97	153	10,52%	438	18,90%	3,74
2010	1567	2657	1,70	3,63	170	10,85%	595	22,39%	4,37
2011	1732	2921	1,69	3,72	194	11,20%	653	22,36%	4,41
2012	1944	3352	1,72	3,76	226	11,63%	769	22,94%	4,27

In Table 4.7, there is an increase of the ratio of collaborations per year. This means that as the journal and the field, in general, matures, authors collaborate more widely. Moreover, ratio of giant component's authors per total authors, as well as the ratio of giant component's collaborations per total collaborations increases per year. This provides us with indications that small groups of authors start collaborating more widely in recent years and they enter the giant component. As a result, we could expect in the coming years that the co-authorship network will transform from fragmented to adequate connected. Finally, we can notice that the average distance of both the network and the giant component increases per year. This is due to the fact that more new authors are publishing papers to the ETS Journal each year. However, if we focus on the time slice from 2010 to 2012, we can notice that there is a slight increase of network's average distance. On the contrary, the giant component's average distance was slightly increased from 2010 to 2011 and slightly decreased from 2011 to 2012. This provides us with indications that the network might have arrived at its "phase transition" (Lee et al., 2010) in 2010, where authors collaborate with each other much more frequently and more widely. We can also expect, from the perspective of SNA, that the average distance will start decreasing in the coming years.

4.2.2 The ETS Co-authorship Development model

Another issue that we investigated was the development model of the ETS Journal co-authorship network. As it has been identified by the Pharm et al. (2012) co-authorship networks follow a specific development pattern that consists of four stages, as follows:

- **Born Phase:** this is the initial phase of the co-authorship network, where there are only a few connections between authors.
- **Bonding Phase:** this is the phase after some years, when author groups start becoming visible in the network.
- **Emergence Phase:** this is the phase, when authors groups created during the bonding phase are gradually integrated through publications that involve authors from more than one group.

- **Stable Phase:** this is the phase, when the co-authorships network has formed a specific type of topology. There different types of topologies that include: (a) focused topology, that is for networks featuring a strongly connected core group of authors that is connected to other smaller groups, (b) interdisciplinary topology, that is for networks featuring several groups connected via some gatekeepers, but where there is no core group and (c) hierarchical topology, that is for networks featuring some “super gatekeepers” who connect a hierarchy of groups.

Figure 4.12 presents the development of the ETS Journal co-authorship network in four time periods, namely 1999 to 2003 (5 years), 1999 to 2007 (9 years), 1999 to 2010 (12 years) and 1999 to 2012 (14 years). As we can notice from Figure 4.12, in 2003 there were many small groups formed by authors who have co-authored one paper with at least one co-author. Four years later, in 2007, we can notice the existence of some larger co-authorship groups, and also some *bonding* among author groups. In 2010, some larger co-authorship groups are already clearly discrete, and the network also starts to form a clearly visible large component of core authors in the *emergence* phase. By 2012, the largest component is beginning to actually deserve the label “giant component” and we can notice that many members of the giant component have co-authorship ties to other authors but author groups at the periphery remain isolated. Although it is evident at the bottom of each network snapshot that the network includes a large pool of non-connected co-authorship groups, the ETS Journal co-authorship network tends toward developing a *focused* topology.

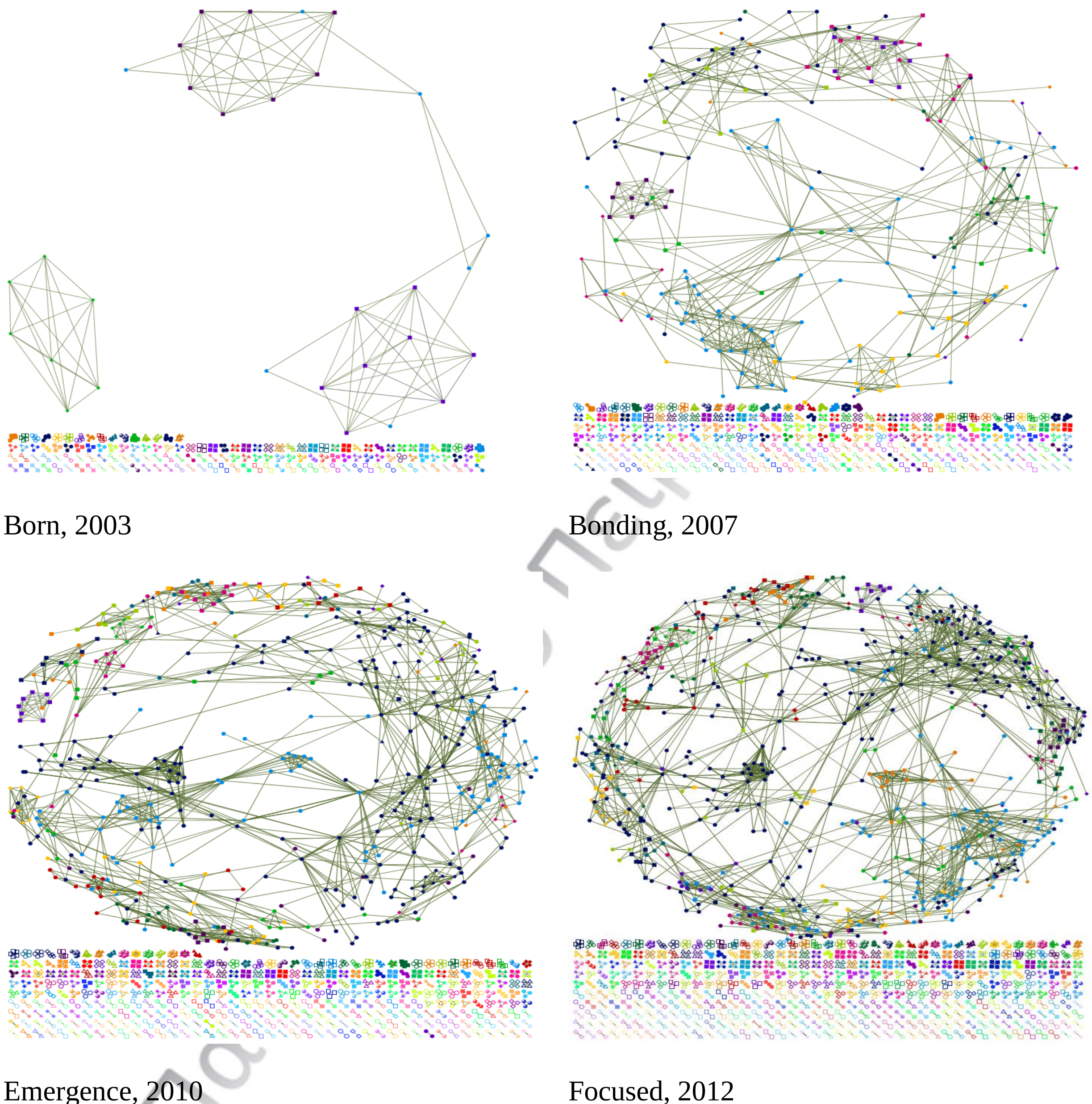

Figure 4.12: ETS Journal Co-Authorship Network Development

4.3 Synopsis

In this chapter, we applied the SNA metrics that have been previously presented to the ETS Journal co-authorship network. We divided our analysis in two stages, namely, (a) network identity and (b) network evolution.

At the former stage, we analyzed the current state of the network by applying graph metrics and we performed ranking and clustering of authors by applying vertex specific centrality metrics. This process enabled us to extract results that are aligned with the research analytics requirements. We have, thereby, made conclusions about the highly influential authors and major clusters of authors collaborating on specific research areas.

In relevance to the graph metrics, our main findings include that:

- on average, the authoring team of an ETS Journal paper consists of 2 or 3 authors.
- only 1 out of 6 collaborations have occurred twice or more times.
- authors included in the giant component collaborate more often than the authors who are not part of the giant component.
- our network is consisted of a main core of collaborations and many small disconnected peripheral groups.
- the co-authorship network's typology in TeL follows similar patterns with other research fields.

In relevance to the vertex-specific metrics, we have concluded that:

- the distribution of degree centrality, betweeness centrality and eigenvector centrality among vertices follows power-law distribution. The distribution of closeness centrality follows the normal curve distribution.

- in relevance to the degree centrality, there are very few authors who are measured to have realtively very high values. Those really significant individuals have many connections among the authors and strongly influence the structure of collaborations.
- the authors/researchers of the ETS Journal collaborate by following similar patterns with other research fields.
- the ETS co-authorship network includes a strongly connected subgroup of authors, who are from Taiwan and they collaborate frequently, diversely and widely.
- for the top 30 vertex-specific metrics authors, there is a significant correlation between three centrality metrics (namely, degree centrality, betweenness centrality and closeness centrality) and the number of their citations, with closeness centrality as the highest.
- only degree centrality, betweenness centrality and closeness centrality can possibly be used as supplementary indicators for assessing author's scientific recognition.

We have also come to some conclusions about the four influential factors of research collaboration. In Table 4.8, we have summarized for clarity the research collaboration literature claims next to our own findings.

Table 4.8: Conclusions about the influential factors of research collaboration

Factor	Research Collaboration Literature	Our findings
Spatial distribution of collaborating parts	The shorter the distance between two researchers is, the bigger the possibility to engage in research collaboration.	The top 3 authors according to the different centrality metrics are from Taiwan except from Kinshuk, who is from Canada. This means that the ETS co-authorship network includes a strongly connected subgroup of authors, who are from Taiwan and they collaborate frequently (as indicated by degree centrality), diversely (as indicated by betweenness centrality) and widely (as indicated by closeness and eigenvector centrality).
Social proximity of collaborating parts	Research collaboration can be favored by social proximity.	The social proximity factor of research collaboration plays also an important role in our network. Specifically, we notice that almost – except for one- all the authors that appear in the top 3 ranks hold an Associate professor or higher academic position.
Scientific productivity rate of collaborating parts	The productivity rate of a researcher correlates with his collaboration rate.	The high numbers of papers co-authored by the top 3 authors for each metric- 4,54 papers on average- indicate that, in our network, as well, research collaboration can be associated with high productivity in relevance to the numbers of papers produced.
Impact of the end result	Adding an author to an article has a positive impact on the number of citations the article receives.	In our network, the high collaboration rate correlates with citation number. Significant correlation lies, specifically, between the degree centrality (r= 0,60), betweenness centrality (r= 0,44) and closeness centrality (r=0,62) with the citation number.

In relevance to the authors clustering, we have concluded that:

- there are four (4) major authors' clusters.
- some of the emerging niche areas of research include:
    - Wireless, Mobile and Ubiquitous Technologies for Learning
    - International Standards and Specifications for Learning Technologies
    - Computer Supported Collaborative Learning
    - Adaptive and Personalized Technology-Enhanced Learning.
- those topics are aligned with relevant research works for the field of TeL.

In the latter section of network evolution, we focused our analysis on the formation of the ETS Journal co-authorship network through time. We studied (a) the annual distribution of authors and collaborations and (b) the development model that the network has followed.

Our main conclusions about the annual distribution of authors and collaborations from 1999 to 2012 are that:

- more new authors are publishing papers to the ETS Journal each year.
- there is a growing tendency for authors and collaborations which is expected to continue in the coming years.
- the co-authorship network will transform from fragmented to adequate connected.
- the network might have arrived at a literature defined phase in 2010, where authors collaborate with each other much more frequently and more widely.

Our main conclusions about the development model from 1999 to 2012 are that:

- there are four separate, literature defined, phases which our network follows: (a)Born, (b)Bonding, (c)Emergence and (d)Focused .
- the ETS Journal co-authorship network tends toward developing a focused topology.

In the subsequent chapter we will synthesize our conclusions and we will present ideas for future work.

Chapter 5
Conclusions and Future Work

In this chapter we will revisit the main aspects of this study. We will present our main conclusions and we will set the basis for the next steps that need to be taken in order to expand, generalize or utilize our findings. This chapter synthesizes the interpretation and significance of our research results.

5.1 Conclusions

This study was set out to explore the research collaboration that takes place in the field of TeL. Research collaboration is one of the most important aspects through which we can gain insights on the process of producing scientific knowledge. But, for that knowledge to be most useful, we took into consideration the research analytics approach, which dictates specific requirements for our research methods. In order to achieve that, co-authorship networks and SNA analysis where chosen as instruments, since they can provide substantial insights and patterns about how research communities are connecting through collaboration and how they gradually evolve, over time.

Although co-authorship analysis is the most suitable method to study research collaboration, nevertheless, as we have concluded from the relevant literature review, there are no co-authorship network analysis studies for networks formed by the authors of scientific journals in the field of TeL. Thus, in this thesis, we studied the co-authorship network formed by the authors of the ETS Journal, an open access, accredited academic journal dedicated to TeL research.

Our study involved 792 co-authored, regular and special issue, papers from 1999 to 2012. Our co-authorship analysis software was NodeXL, a free, open-source template for Microsoft Excel. We used NodeXL which not only enabled us to import, filter and edit our data but also to create, edit and visualize the corresponding network graph. In addition to that, we utilized NodeXL in order to apply a variety of SNA metrics, namely graph metrics and vertex-specific metrics, to our co-authorship network, which enabled us to extract results that are aligned with the research analytics requirements. One problem that we had to address was the fact that the network is fragmented, with many non-connected subgroups, which raised some obstacles in the application of the vertex-specific metrics. The said problem was solved by applying those metrics only on the giant component, following relevant literature. This process has enabled us to draw conclusions about the highly influential authors, the major clusters of authors collaborating on specific research areas and the interesting characteristics of the network evolution.

Within this context, our first important conclusion about research collaboration is its growth over time. As new authors come into the network, the rate of collaborative activity increases, too. It is, in fact, expected to sustain its augmenting tendencies for the forthcoming years, as well.

In addition to that, research collaboration activity is fairly associated with influencing factors, such as geographical distribution and productivity rate. Our findings confirmed a strong relationship between collaboration and the aforementioned influencing factors.

According to our findings, research collaboration is also correctly associated with the overall impact of authors and papers, in the scientific community. In fact, the strong correlation that was found between degree and closeness centrality metrics with respective citation numbers, suggest that the said metrics could be used as supplementary indicators for assessing author's scientific recognition. Those metrics could, thereby, provide alternative perspectives to the current methods, such as h-index and i10-index.

We have also concluded that external motives and the overall background and culture can boost research collaboration. In this specific case, based on our findings about the very strong and central presence of Taiwanese authors from various institutions, we have but to conclude that the national research policy in Taiwan promotes and facilitates research collaboration. This claim is also supported by the opposite effects of the Council of Higher Education in Turkey (Bozkaya, Aydin & Kumtepe, 2012).

One additional insight is that the research collaboration topics of interest in the ETS Journal, which we have underlined, appear to be aligned with the general trends in the TeL field (Hsu et al., 2012; Kinshuk et al., 2013). We can thereby conclude that the field has some specific emerging research topics, which are independent from the means or context of the scientific research.

Overall, we notice a general consistency between our research findings and relevant studies, analyzing research collaboration in various fields (e.g. Barabási et al., 2002; Newman, 2001; Yan & Ding, 2010). This fact indicates that research collaboration is a concept with some interdisciplinary, global characteristics. Within the specific research field of TeL, there are also some very strong similarities, e.g. the development model proposed by Pham et al. (2012), is generally followed by our network too, or the relative small number of prolific authors (Reinhardt et al., 2011).

Ultimately, interesting recommendations can be extracted as follows:

- For the *ETS Journal editors*, their efforts in retaining the key authors (namely, those with high centrality metrics) of the journal active, are very important for the future evolution of the co-authorship network. More specifically, ETS editors can approach key authors for organizing special issues in specific research topics or they can offer them roles in the editorial board of the journal.

- The *key authors* of the network should be active in finding, suggesting and setting up new collaborations with members in different sub-groups, particularly from the journal network's periphery, which will make the entire network more integrated and will promote research collaboration by contributing to new collaborations. Key members also play an important role in engaging new authors and connecting them to the core of the co-authorship network. This can also lead to the introduction of emerging research topics to the current main research areas.
- *For the potential authors/ TeL researchers,* they should take into consideration the current social structure of the ETS community, before publishing a paper, since they can benefit from the knowledge of research collaboration patterns and key research areas of collaboration.

5.2 Future work

In future work, we are planning to augment the findings drawn from this study by considering data from other journals in the field of TeL. This will enable us to generalize our findings and compare them to other research fields towards identifying similarities or differences, as well as to have a complete view of the research collaboration that takes place in the TeL landscape. Moreover, another dimension for our future work will be the spatial proximity factor of research collaboration, which can be studied by identifying collaboration patterns between authors from different countries. Finally, our future efforts will also focus on building research analytics tools and/or services that will be able to visualize and further analyze the data that have been extracted from this study.

References

Amabile, T. M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P. W., Marsh, M. et al. (2001). Academic- practitioner collaboration in management research: A case of cross-profession collaboration. The Academy of Management Journal, 44(2), 418-431.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27, 17–21.

Barabási, A. L., & Albert R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*, 509-512.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek T. (2002). Evolution of the social network of scientific collaborations *Physica A, 311*, 590-614.

Barnes, J. A. (1954). Class and Committees in a Norwegian Island parish, *Human Relations*, 7, 39-58, doi:10.1177/001872675400700102.

Beaver, D. & Rosen, R. (1979). Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship. *Scientometrics*, 1, 133-149.

Bonacich, P. (1972). Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology*, *2*, 113-120.

Bozkaya, M., Aydin, İ. & Kumtepe, E. (2012). Research Trends and Issues in Educational Technology: A content Analysis of TOJET (2008–2011). *The Turkish Online Journal of Educational Technology*, 11(2), 264- 277.

Braun, T.; Glinzel, W.; & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, 51, 499-510.

Bukvova, H. (2010). Studying Research Collaboration: A Literature Review. Sprouts: Working Papers on Information Systems, 10(3). http://sprouts.aisnet.org/10-3.

Cho, Y., Park, S. Jo S. J. & Suh, S. (2013). The landscape of educational technology viewed from the ETR&D journal, *British Journal of Educational Technology, 44*(5), 677-694

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.), New Jersey: Lawrence Erlbaum Associates.

Coleman, T. F., & Moré, J. J. (1983). "Estimation of sparse Jacobian matrices and graph coloring Problems", SIAM *Journal on Numerical Analysis, 20,* 187–209.

Erfanmanesh, M., Rohani, V. A., & Abrizah. (2012). A. Co-authorship Network of Scientometrics Research Collaboration. *Malaysian Journal of Library & Information Science, 17*, 73-93.

Erman, N., & Todorovski L. (2010). Analyzing the structure of the EGOV conference community. EGOV'10 *Proceedings of the 9th IFIP WG 8.5 international conference on Electronic government.* 73-84.

Freeman, L. C. (1979). Centrality in Social Networks. Conceptual Clarification. Lausanne *:* Elsevier Sequoia S.A.

Garfield, E. (1979). Citation Indexing-Its Theory and Application in Science, Technology, and Humanities. John Wiley & Sons, New York, NY.

Glänzel, W.(2002). Coauthorship Patterns and Trends in the Sciences (1980-1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies. Library Trends, 50(3), 461-473.

Gomez, I., Fernindez, M. T. & MCndez, A. (1995). Collaboration patterns of Spanish scien- tific publications in different research areas and disciplines.

Proceedings of the Biennial Conference of the International Society for Scientometrics and Informatics 187-196.

Granovetter, M. (1973). "The strength of weak ties". *American Journal of Sociology*, 78 (6), 1360–1380.

De Bellis, N. (2009). Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Scarecrow Press Inc.

Harmelen, M. & Workman, D. (2012). Analytics for Learning and Teaching. CETIS Analytics Series, 1(3), 1-41.

Hagstrom, W.O. (1965). The Scientific Community, Basic Books: New York. Hansen, D. L., Smith, M. A., Shneiderman, B. (2011). Analysing Social Media Networks with NodeXL: Insights from a Connected World, Burlington: Morgan Kaufmann.

He, Z. L., Geng, X. S. & Hunt, C. C. (2009). Research collaboration dn research output: A longitudinal study of 65 biomedical scientists in New Zealand university. *Research Policy*, 38(2), 306-317.

Hicks, D., Ishizuka,T., Keen, P. & Sweet, S. (1994). Japanese corporations, scientific research, and globalization. *Research Policy*, 23, 375-384.

Hicks, D. & Katz, J. S. (1997). The changing shape of British science. STEEP special report No. 6, SPRU.

Hsu, Y.-C., Ho, H. N. J., Tsai, C.-C., Hwang, G.-J., Chu, H.-C., Wang, C.-Y. & Chen, N.-S. (2012). Research Trends in Technology-based Learning from 2000 to 2009: A content Analysis of Publications in Selected Journals. *Educational Technology & Society*, 15 (2), 354–370.

Hu, C. & Racherla, P. (2008). Visual representation of knowledge networks: A social network analysis of hospitality research domain. *International Journal of Hospitality Management*, 27 (2), 302- 312.

Hurley, L., Ogier, A., & Torvik, V. (2013). Deconstructing the collaborative impact: Article and author characteristics that influence citation count. *ASIS&T 2013 Conference Proceedings, Peer-reviewed paper presented at ASIS&T 2013 Annual Meeting in Montreal,* Canada.

Iglesias, J. E. & Pecharromán, C. (2007). Scaling the h-index for different scientific ISI fields*, Scientometrics*, 73(3), 303-320.

Katz, J. S. (1993). Bibliometric assessment of intranational university-university collaboration. doctoral dissertation, Science Policy Research Unit, University of Sussex, Brighton.

Katz, J. S. & Martin, B. R. (1997). What is research collaboration?, *Research Policy,* 26, 1-18.

Kaza, S., & Chen, H. (2010). Identifying High-status Nodes in Knowledge Networks, *Annals of Information Systems*, 12, 91-108.

Kienle, A., & Wessner, M. (2005). Our way to Taipei: an analysis of the first ten years of the CSCL community. Proceedings of the 2005 *Conference on Computer Support for Collaborative Learning*, CSCL '05. 262–271.

Kinshuk, Huang, H.W., Sampson, D., & Chen, N. S. (2013). Trends in Educational Technology through the Lens of the Highly Cited Article. *Educational Technology & Society*, *16(2)*, 3–20.

Knoke, D. & Yang, S. (2008). Social network analysis (2nd ed). Thousand Oaks: Sage Publications.

Kohavi, R., Rothleder N. J. & Simoudis E. (2002). Emerging Trends in Business Analytics. *Communications of the ACM*, 45 (8), 45–48.

Kraut, R. & Egido, C. (1988). Patterns of contact and communication in scientific research collaboration, *Proceedings of the Conference on Computer-Supported Cooperative Work*, 1-12.

Lee, D., Goh K. I., Kahng B., & Kim D. (2010). Complete trails of co-authorship network evolution, *Physics and Society*. doi: 10.1103/PhysRevE.82.026112 .

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-Authorship Networks in the Digital Library Research Community, *Information Processing and Management*, *41*, 1462–1480.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P. & Bhattacharjee, B. (2007). Measurement and analysis of social networks. *Proceedings of the IMC '07*, 29-42.

Moody, J. (2004). "The Structure of a Social Science Collaboration Network" *American Sociological Review*. *69*, 213-238.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 404-409.

de Nooy, W.,Mrvar, A. & Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek, *Cambridge University Press*. Retrieved from: http://courses.arch.ntua.gr/fsr%2F144992/Pajek-Manual.pdf.

Ochoa, X., Méndez, G., & Duval, E. (2009). Who We Are: Analysis of 10 Years of the ED-MEDIA Conference. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* 2009. 189-200.

Opsahl, T., Agneessens, F., & Skvoretz, J., (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32 (3), 245-251.

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science, 28(6)*, 441–453.

Patel, N. (1973). Collaboration in the professional growth of American sociology, *Social Science Information*, *vol.12, 77-92.*

Pham, M. C., Dentrl, M., & Klamma, R. (2012). Development Patterns of Scientific Communities in Technology Enhanced Learning. *Educational Technology & Society*, *15(3*). 323-335.

Pravdic, N. & Oluic-Vukovic, V. (1986). Dual approach to multiple authorship in the study of collaboration/scientific output relationship. *Scientometrics, 10(5-6), 259-280.*

Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes*. Journal of the American Society for Information Science,* 27, 292–306.

Reinhardt, W., Meier, C., Drachsler, H. & Sloep, P. (2011). Analyzing 5 Years of EC-TEL Proceedings. *Towards Ubiquitous Learning - 6th European Conference of Technology Enhanced Learning, EC-TEL 2011*. 531–536.

Rigby, J. & Edler, J. (2005). Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. *Research Policy,* 34(6), 784-794.

Sonnenwald, D. H. (2007). Scientific collaboration. Ann. Rev. Info. Sci. Tech., 41: 643–681. doi: 10.1002/aris.2007.1440410121 .

Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review, *Journal of Information Science, vol. 6(1), 33-38,* doi:10.1177/016555158300600105 .

Velden, T., Haque, A. & Lagoze, C. (2010). New Approach to Analyzing Patterns of Collaboration in Co-authorship Networks - Mesoscopic Analysis and Interpretation, *Scientometrics*, *85*, 219-242.

Wasserman, S. & Faust, K. (1994) Social Network Analysis: Methods and Applications. Cambridge University Press.

Watson, H. J. (2011). Business analytics insight: Hype or here to stay? . *Business Intelligence Journal*, 16, 4-8.

Watts, D. J., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393 (6684)*, 440–442.

Weinberg, S.L. & Abramowitz, S. (2008). Statistics Using SPSS: An Integrative Approach (second edition). London: Cambridge University Press.

Wetherell, C., Plakans, A. & Wellman, B. (1994). Social Networks, Kinship, and Community in Eastern Europe. *Journal of Interdisciplinary History 24 (Spring): 639-663*.

Yan, E. & Ding Y. (2010). Applying centrality measures to impact analysis: A coauthorship network analysis. Retrieved from: http://arxiv.org/ftp/arxiv/papers/1012/1012.4862.pdf

Zare-Farashbandi F., Geraei E. & Siamaki S. (2014). Study of co-authorship network of papers in the Journal of Research in Medical Sciences using social network analysis. *Journal of Research in Medical Science*, 19 (1), 41–46.