

Social Ties Within The Bibliography Of An Accounting Research Paper.

**Written by: Alexander Evangelos Toufexis
USP nr.: 5534791
Course: Inteligência Competitiva Aplicada aos
Sistemas de Informações Empresariais
Teachers: Edson Luiz Riccio, Luc Quoniam
Date: 22/11/2005**

1. Content

1. Content.....	2
2. Introduction	3
3. Literature Overview.....	4
<i>3.1 Zipf's law.....</i>	4
<i>3.2 Granovetter's strong and weak ties.....</i>	4
4. Methodology.....	5
<i>4.1 Sample Data.....</i>	5
<i>4.2 Analyzing program.....</i>	5
5. Analysis.....	6
<i>5.1 Histogram by Frequency of Authors.....</i>	6
<i>5.2 Histogram by Frequency of Keywords.....</i>	7
<i>5.3 Cross-matrix of Authors and Keywords.....</i>	7
<i>5.4 Relations between Authors.....</i>	10
<i>5.5 Relations between Authors and Keywords.....</i>	12
6. Limitations and Suggestions	14
7. Conclusion.....	15
8. Literature.....	16

2. Introduction

Within the topic of value relevance of financial statements, I will investigate relations between authors of related articles. The relations (or ties) between authors will be investigated as well as through co-authorship as by means of an asymmetrical network between authors and keywords. Granovetter (1973) defines the strength of a tie the following: “The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.” As may be obvious, due to the nature of this research, there will not be a complete assessment of the strengths of the ties between authors. However, we can still make a distinction between the strength of the ties between authors based on how many times an authors cooperated with one specific other author, as well as existing links between authors established through another (mutual co-) author, or based on a link established through mutual keywords used by the authors to describe the content of their articles. I will base the strength of the ties/links between the authors on their relationship with each other; that is, for example multiple co-authorship, single co-authorship, a relation through a mutual co-author, or a relation through co-authorship of co-authors.

Within the subject of value relevance of financial statements, different subtopics can be distinguished; for example, asymmetry, conservatism, timeliness and earnings persistence. I expect to find clusters/linked authors around this subtopics. Besides that, I expect to observe ties through keywords between non-co-authors. Obviously, next to the visible networks, there will also be a number of authors that have not cooperated within the sample of articles used in this research. The sample used in this research consists of 120 bibliographical references that are highly relevant to the subject of my thesis in value relevance of financial statements. Due to the requirement that an article (and thus the reference to it) had to be relevant to this specific topic in accounting research, we have to recognize that not all links between authors can be visualized in the analysis. However, one could question to what extent these absent links would have value for the analysis of a bibliography in the current subject, as they would be established through other (irrelevant) subjects.

In chapter three, I will provide a brief overview of the relevant literature regarding bibliometrical analysis. Following the literature overview, I will discuss the methodology and data acquirement. In the fifth chapter, I will analyze and discuss the data and outcomes. Chapter six will contain the limitations of the paper and suggestions for further research. Chapter seven will conclude the research.

3. Literature Overview

3.1 Zipf's law

The Harvard linguist George Kingsley Zipf (1949) found that in the body of an (English) tekst, the frequency of any word is roughly inversely proportional to its rank in the frequency table. In other words, the word with the highest frequency will occur twice as often as the word with the second highest frequency, which on its turn will occur twice as often as the word with the fourth highest frequency. Resulting; the n th most common frequency will occur $1/n$ as often as the first. Do note however, that this will never be exactly true, as the items that are measured must will occur only in an integer number of times. This type of probability distribution is ever since referred to as Zipf's law.

The easiest way to see whether a distribution follows Zipf's law is to scatter plot the data, with the axis x as $\log(\text{rank order})$ and axis y as $\log(\text{frequency})$. If the points are close to a single straight line, the distribution follows Zipf's law.

3.2 Granovetter's strong and weak ties

Granovetter (1973) inquires into the analysis of social interaction. He defines the strength of a tie as follows: "The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie." He finds that weak ties are more important for personal advancement, such as good jobs, than strong ties, such as family and friendship.

4. Methodology

4.1 Sample Data

I gathered the references relating to the topic “value relevance of financial statements” from EBSCO host. In other words, from the ISI database. I chose this database because it is American, and within this topic, most has been written within American journals. For the purpose of this research, the resulting references stemming from this database should suffice. There are a few issues that still arise even though I use only one database. Because even within the same databases, authors are cited in different ways. For example, the one moment the middle initial of a author is included, and the other time not. This may result in missing links within the social network. I removed double entries and also corrected the different author names so that the same author is actually recognized as the same author.

4.2 Analyzing program

In order to analyze the networks and ties between authors (and keywords), I have used the program Matheo Analyzer. Using Matheo Analyzer, I was able to plot the different histograms, cross-matrices, and do the different analyses between co-authors, and authors with keywords.

5. Analysis

This chapter is subdivided into the different parts of analyzing the bibliographical data. In the first part of this chapter, I will discuss the mere histograms of both the authors and the keywords, and the combined matrix of authors to keywords. The second part of the chapter will discuss the relations between the authors and also the relations between the authors and keywords.

5.1 Histogram by Frequency of Authors

The histogram of authors shows how many times an author has written an article that is part of our sample. The histogram is sorted by how many times an author has written an article; that is, the more articles an author has written, the more to the left he will be positioned. The total amount of different authors is 195.

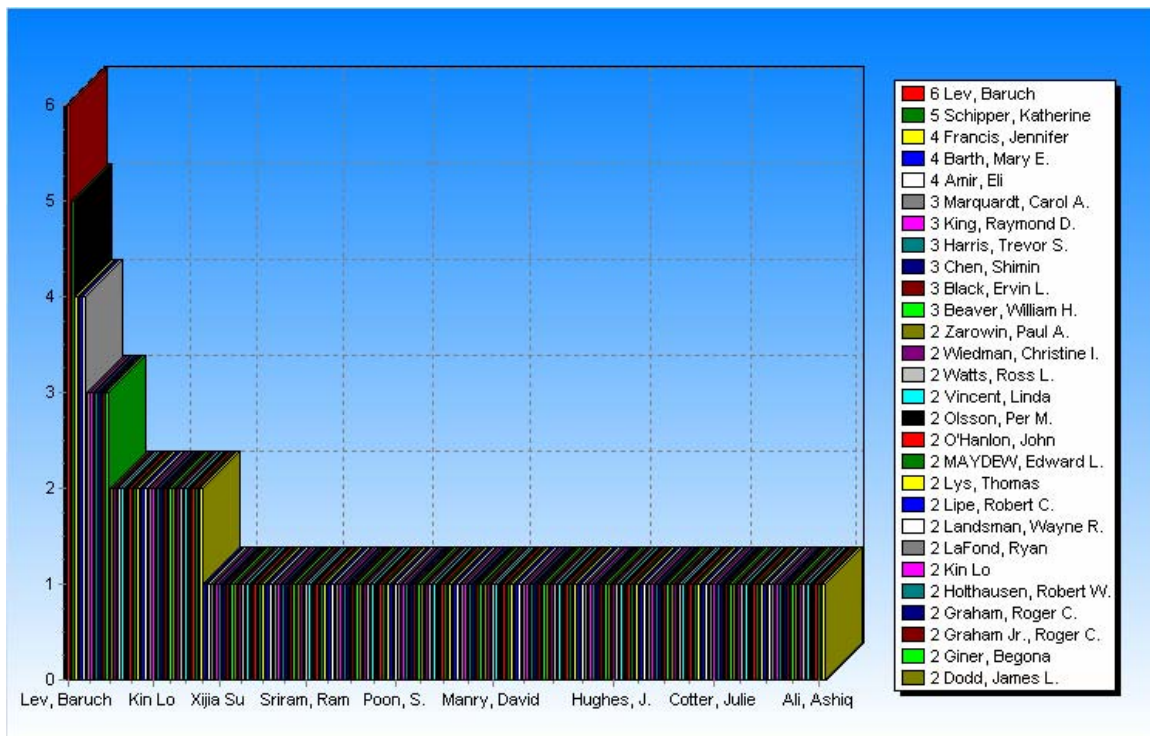


Figure 1. Histogram by frequency of 195 Authors

When we take a look at the histogram by frequency of authors, the graph looks initially like a zipfian distribution. However, when we look a bit closer to the graph, we see that Zipf's law is not the case. The second rank (in this case Schipper) did not write half the amount of articles Lev (rank 1) wrote. Also the following number of ranks do not comply with Zipf's law. Only the tail would follow Zipf's law (as there the amount of articles is 1). I consider this not enough to regard the distribution as a zipfian one. I do think that the reason that this distribution cannot be considered now as a zipfian distribution is due to the low number of occurrences. If the occurrences had been higher, in other words, the

sample bigger, I expect that we would have been able to distinguish a distribution that follows Zipf's law.

5.2 Histogram by Frequency of Keywords

The histogram by frequency of keywords shows how many articles use the same keyword. This histogram is also sorted by the frequency a keyword occurs as related to an article. It is sorted in such a way, that the more a keyword occurs, the more that keyword is situated to the left. The total amount of keywords referred to is 324.

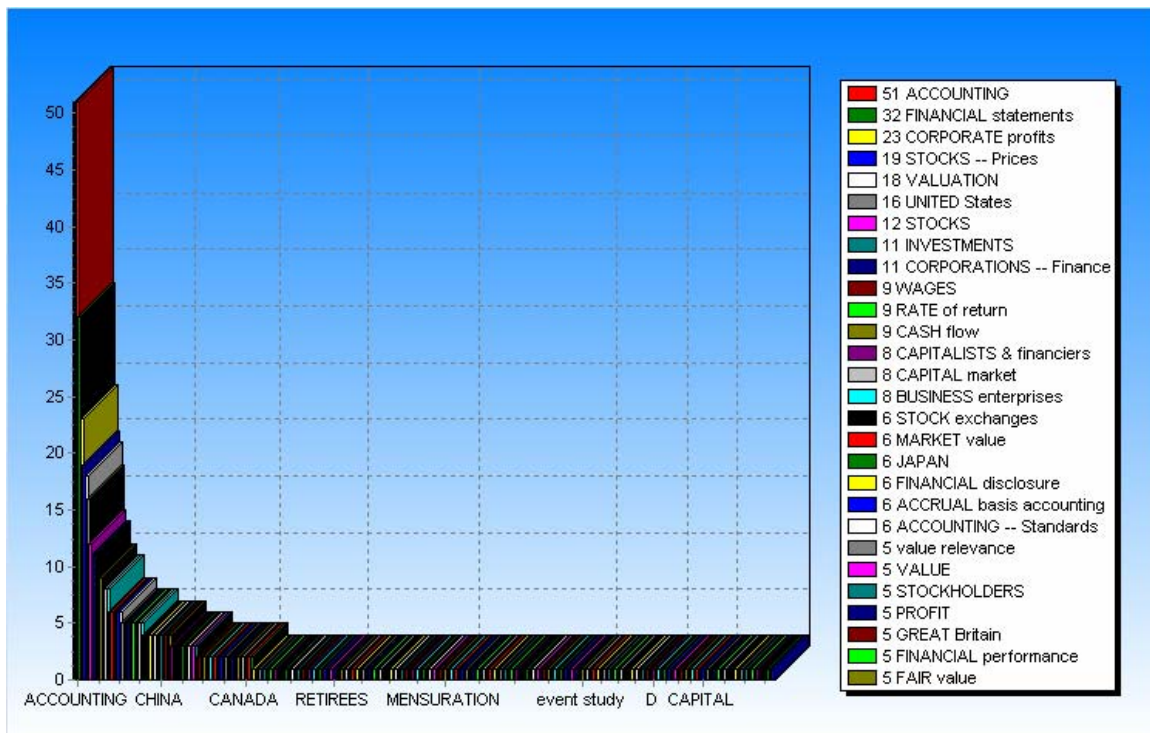


Figure 2. Histogram by frequency of keywords

In this histogram, we are also able to observe that Zipf's law is approximately correct. For example, 32 is not exactly half of 51, but it comes close. Most of the 324 keywords are also only used once in the articles, and only few keywords are mentioned more than ten times in the articles. However, there is one quite remarkable fact in this histogram; the main search term I used to obtain the references ("value relevance") only occurs five times as a keyword in the articles as mentioned by the authors.

5.3 Cross-matrix of Authors and Keywords

In the following matrix, I set out the authors against the keywords, and in the separate cells, the amount of times an author used a certain keyword is displayed. To create a

better overview of the distribution I assigned different colors to those cells that had a value of 1 or higher.

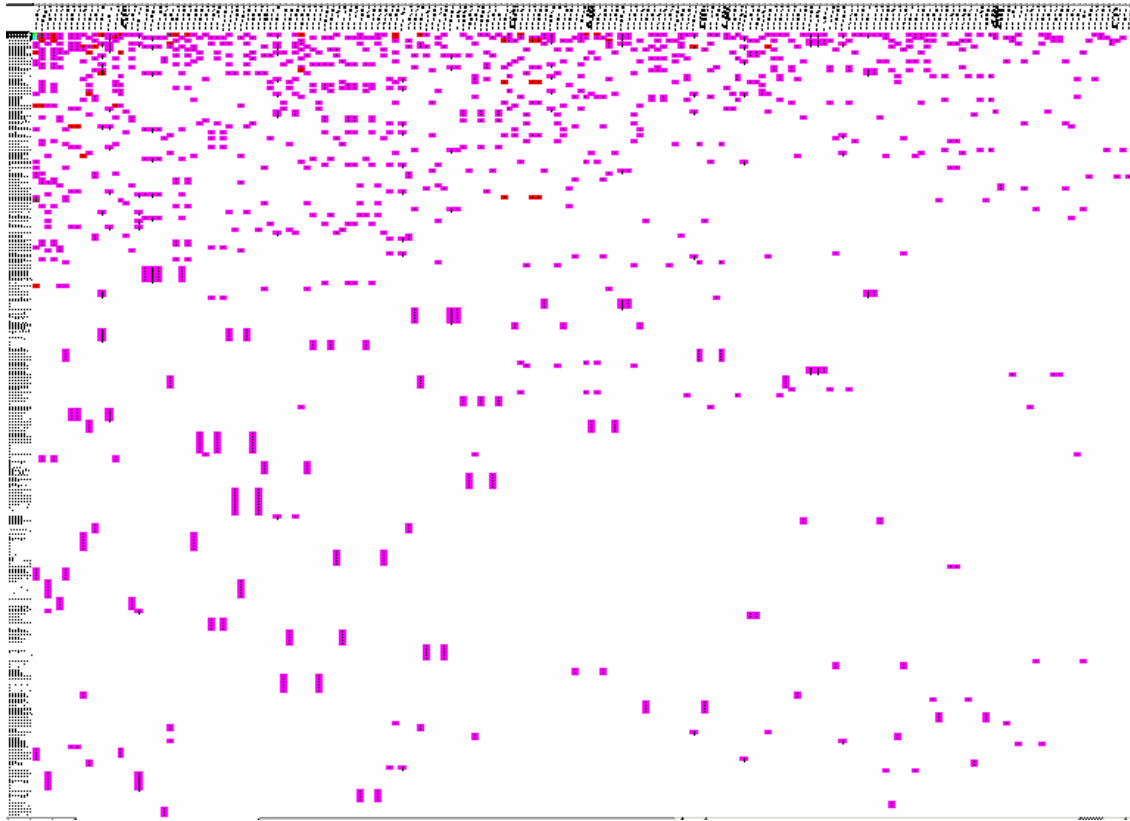


Figure 3. Cross-Matrix of authors against keywords

Even though one might expect that two distributions of which one follows Zipf's law and the other looks like a distribution that Zipf's law - yet is lacking a sufficient large sample - set out against each other in a matrix would result in a new distribution that follows the Zipf's law, this is not the case. Yet, the outcome makes sense. Realizing that the distribution is still based on the amount of times a keyword is used. We see that there is a high concentration of colored field in the left top (for a zoom, see the figure below), including cells that indicate that an author has used a specific keyword up to 3 times. Also, the rest of the few top keywords is highly populated among the complete sample population of authors. After in and about 25% of the keywords, we see that the distribution has become much thinner populated and spread over the authors without any specific pattern.

	Lev, Baruch	Schipper, Katherine	Marquardt, Carol A.	Francis, Jennifer	Zarowin, Paul A.	Abodoy, David	Barth, Mary E.	Beaver, William H.	Chen, Shimin	Harris, Trevor S.	Ben-Hsien Bao	Graham Jr., Roger C.	Landsman, Wayne R.	Vincent, Linda	Amir, Eli	Boone, Jeff P.	Ryan, Stephen G.	Wiedman, Christine I.	Chao-Shin Liu	Davis-Friday, Paquita Y.	Folami, L. Buky	Hand, John R. M.	King, Raymond D.	LaFond, Ryan
ACCOUNTING	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FINANCIAL statements	2	3	1	2	1	0	1	1	0	1	0	3	1	1	1	0	1	1	0	1	1	1	1	2
CORPORATE profits	0	2	0	2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1
STOCKS -- Prices	1	0	1	0	1	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	1
VALUATION	1	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0
STOCKS	0	1	0	1	1	0	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
UNITED States	2	1	0	1	1	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INVESTMENTS	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
CAPITAL market	0	1	0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0
CORPORATIONS -- Finance	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
CAPITALISTS & financiers	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
RATE of return	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
BUSINESS enterprises	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
MARKET value	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	1	0	0	0
STOCK exchanges	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
WAGES	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASH flow	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
FINANCIAL performance	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
ACCURAL basis accounting	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CORPORATION reports	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0
ACCOUNTING -- Standards	1	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0
EQUITY	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
CHINA	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
JAPAN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FINANCIAL disclosure	2	2	1	1	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0
STOCKHOLDERS	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
VALUE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
BOOK value	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FINANCE	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
capital markets	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
INCOME accounting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAIR value	0	0	0	0	0	0	2	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
DISCLOSURE	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
reliability	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
GREAT Britain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
value relevance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
BUSINESS enterprises -- Valuation	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASH management	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ASSETS (Accounting)	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CORPORATIONS -- Valuation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GERMANY	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
PROFIT	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BUSINESS records	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
RESEARCH	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Earnings persistence	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RESEARCH & development	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Value-relevance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INCOME	1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
BUSINESS enterprises -- Finance	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COST accounting	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
EARNINGS per share	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
PRICE-earnings ratio	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
INVESTMENT analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
INTEREST rates	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
LIABILITIES (Accounting)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
SECURITIES	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DISCLOSURE of information	3	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
REGRESSION analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
EARNINGS trends	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
INFORMATION technology	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EQUITY method (Accounting)	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Stock returns	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DISCLOSURE in accounting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0

Figure 4. Cross-Matrix of authors against keywords, zoomed in at the left corner.

5.4 Relations between Authors

When we plot the diagram of the links between the different authors with Matheo analyzer, we get the following:

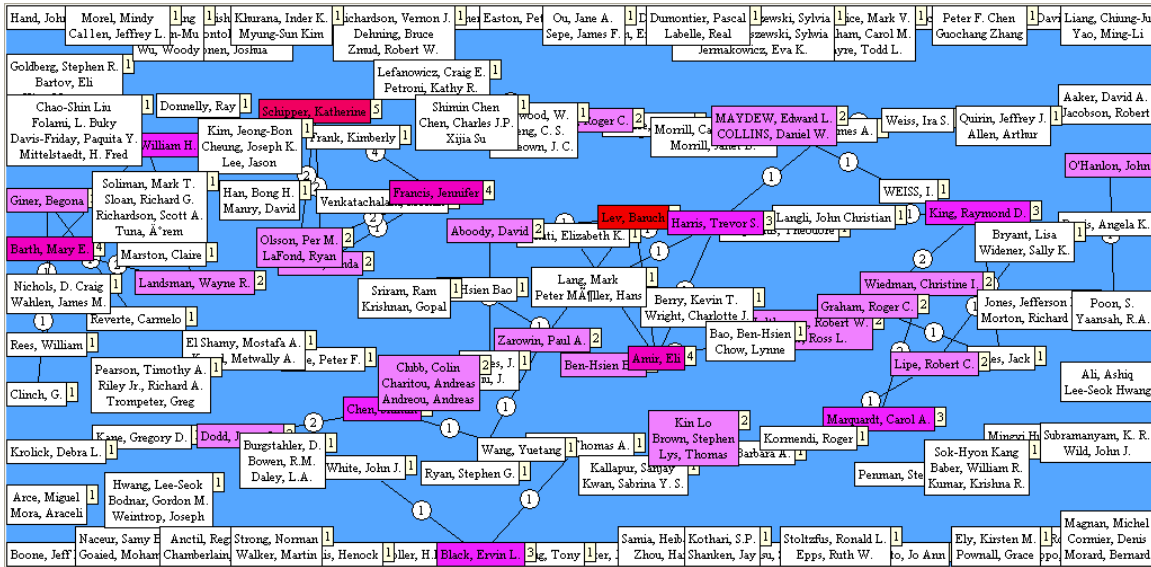


Figure 5.

The image we get lacks a clear overview. And considering our interest in the links between different authors, we removed those authors who do not have any links with other authors. We do this by setting the constraint of minimum amount of links to 1. In that situation, we get the following image:

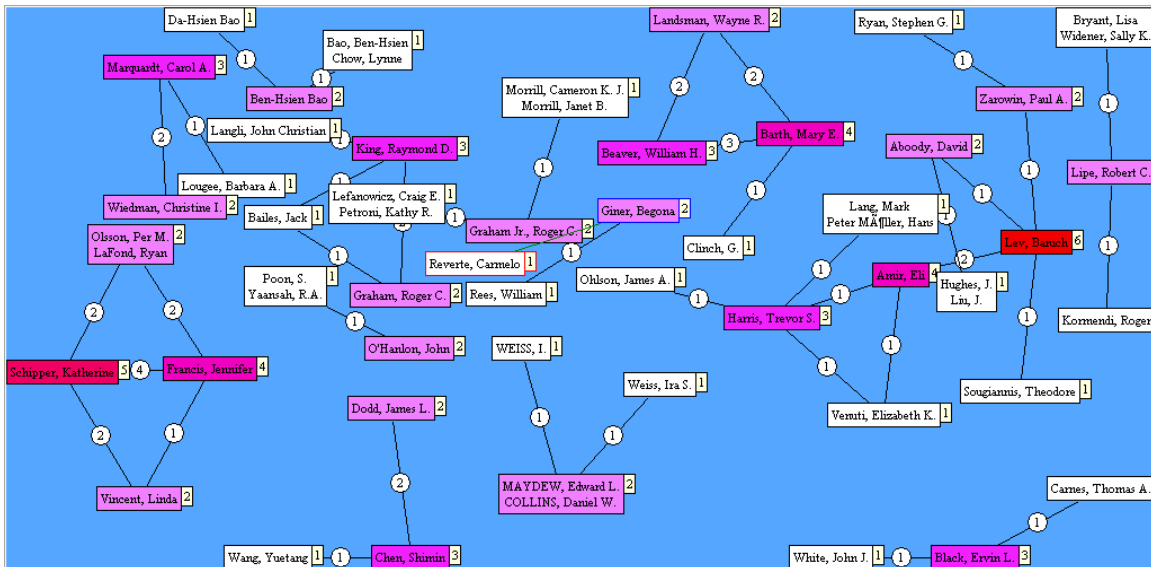


Figure 6.

We can see that most networks contain only a few members (3 to 4 members). These networks most often exist of authors that have written a few articles together (notice for example the very strong link (that is; they were co-authors 4 times) between Schipper and Francis). We are also able to distinguish one larger network. This network is situated around three authors that have written at least 3 articles within the topic used for our analysis. These authors are Harris, Amir, and Lev. The total network with these authors consists of 11 authors. Especially Lev initiates many ties. This was not completely unforeseen, as Lev is also the author with the largest amount of articles in the sample (he wrote 6 articles that are included in this sample).

As for the strength of the ties between the authors, we can distinguish four different categories of strengths. The first category would be very weak ties between authors. These ties can be seen in figure 6. Very weak ties would be those ties between authors, where authors both have co-authors, that are co-authors. For example, Elizabeth Venuti has a co-authorship with Eli Amir, and Theodore Sougiannis has a co-authorship with Baruch Lev; Baruch Lev and Eli Amir are also co-authors. Therefore, there exists a very weak tie between Elizabeth Venuti and Theodore Sougiannis. The second category would be weak ties between authors. This would refer to those authors that do not have a direct link to another author, but yet are still linked to each other through a mutual co-author. For example, Baruch Lev has co-authorship with both Eli Amir and Theodore Sougiannis; therefore, a weak link exists between Eli Amir and Theodore Sougiannis. The third category regarding the strength of ties would be strong ties. Strong ties imply that authors have actually written an article together. For these ties, we will have to look at both figure 6 as figure 5, as figure 5 also displays those authors that have cooperated, but only once within the sample and have no other types of links to other authors. This type of tie seems to be the most occurring. The fact that this type seems to be the most occurring is due to the fact that the networks that do exist are small in size, so the amount of weak or very weak ties is restricted. This in contrast with many authors that only once wrote an article together, and had only one contribution to this area of interest, or at least in this sample. The last category would be very strong, or at least stronger ties. By this, I mean that authors have written more than one article together. An example of this kind of tie is Baruch Lev that has written two articles with Amir Eli. Another, a lot “stronger” example is Katherine Schipper and Jennifer Francis, that have written four articles together within this sample.

Do notice that I am using different definitions regarding the strengths of ties than Granovetter (1972) does. This is due to the nature of the ties within this research. Where we differ from Granovetter mostly is in the definition of weak and very weak ties.

5.5 Relations between Authors and Keywords

In order to look at the relations between authors and keywords, I plotted in Matheo analyzer an asymmetrical network diagram of authors to keywords. The following image came out, while I had the most occurring keyword selected:

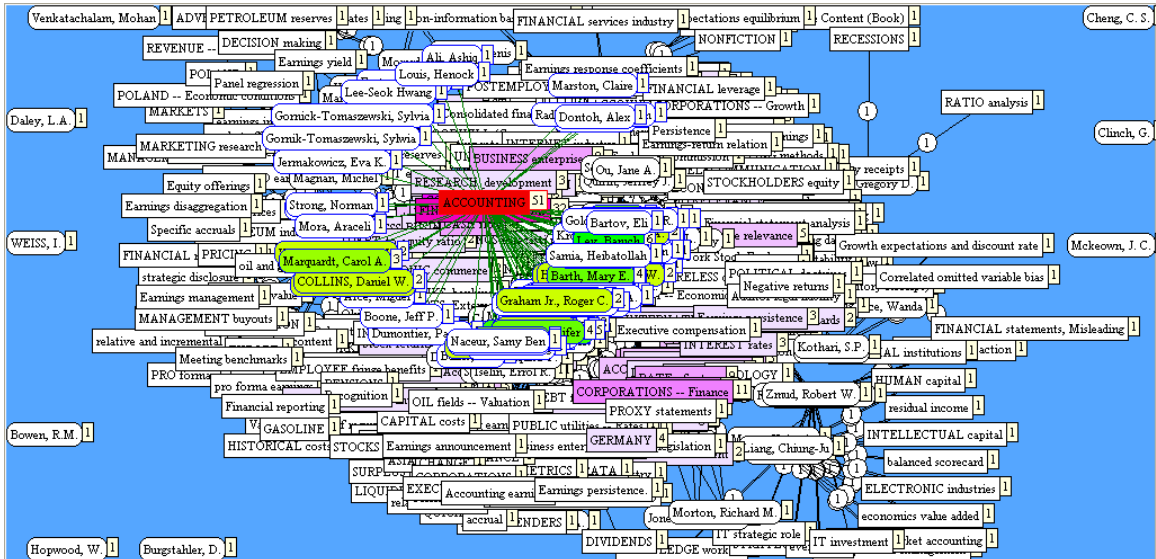


Figure 7. Relations between authors and keywords without any constraints

As can be seen in the diagram, almost everything is linked to each other in one or another way. We cannot deduce much from this diagram, as it is too much of a chaos to distinguish any specific links. Therefore, I have made two more diagrams. In the first diagram, I added the constraint that a form (that is: an author or keyword) should be present in the sample for at least 3 times.

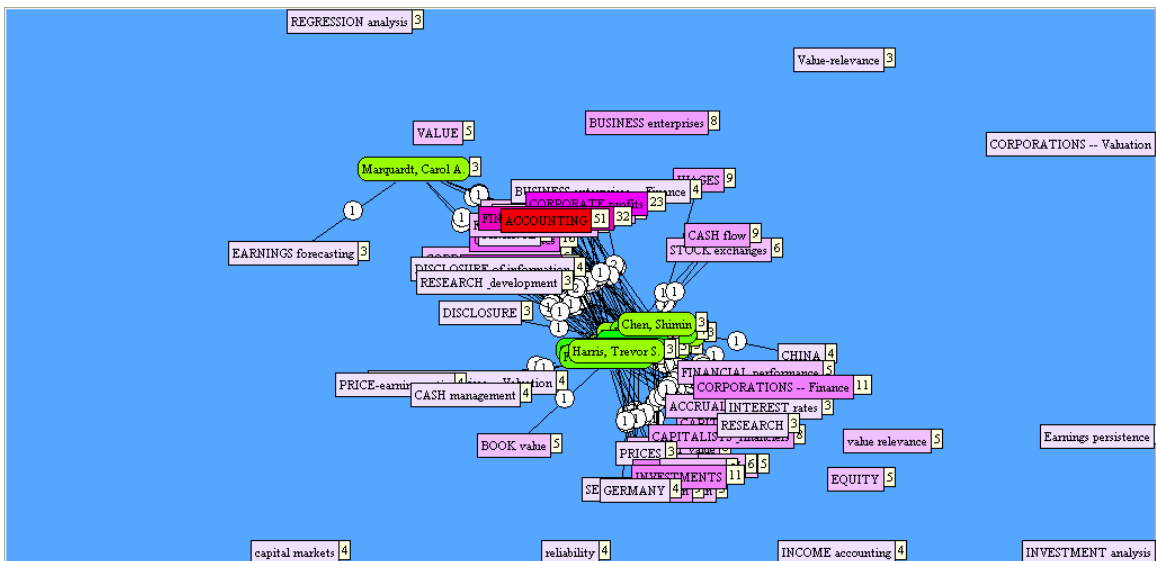


Figure 8. Relations between authors and keywords with a minimum form frequency of 3

With the constraint of a minimum form frequency of 3, many authors have been removed, yet still many keywords remain. What we see in the diagram, is that there is a very high concentration of authors in the middle of two groups of keywords. One remarkable case is Marquardt, that seems to be only linked to one of the two groups of keywords, and one other keyword, and is therefore not in the place where all the other authors are concentrated.

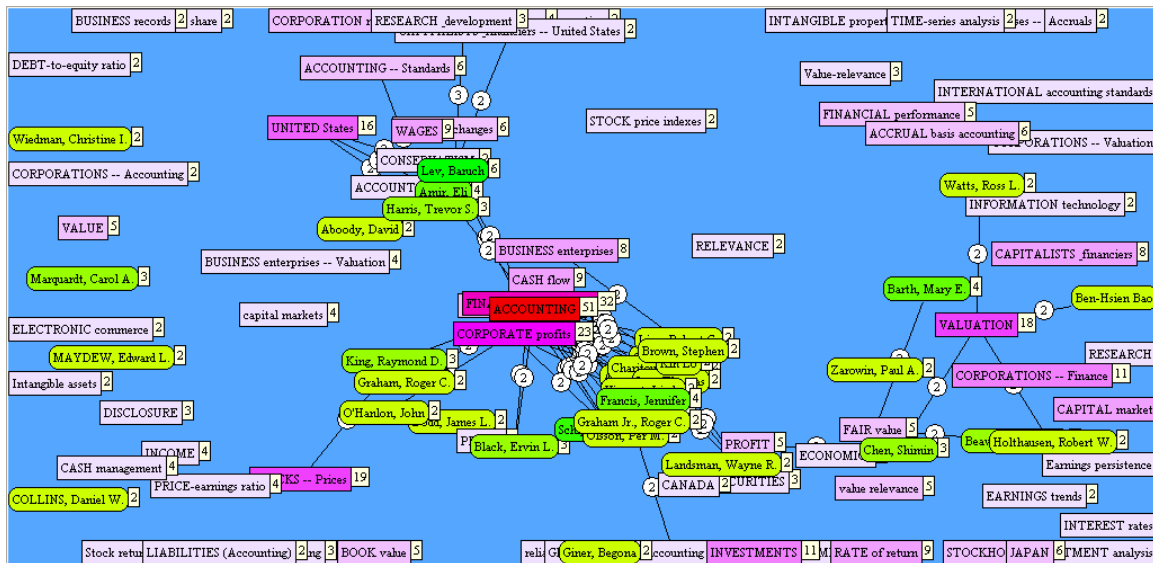


Figure 9. Relation between authors and keywords with a minimum form frequency of 2 and a minimum pair frequency of 2.

Figure 9 displays the relation between authors and keywords with a minimum form frequency of two and a minimum pair frequency of two. The minimum form frequency of two implies that every author or keyword has to occur in the sample at least two times. The minimum pair frequency of two implies that the form has to have a link with another form that occurs at least two times. We see a strong concentration of authors around the keywords with the highest occurrence, as is exactly what was to be expected. In this graph, we see that many authors are linked through the most common keywords. Only one other group can be distinguished, next to the group of accounting, corporate profits, and financial statements, and that is the group around valuation.

6. Limitations and Suggestions

There are a number of limitations to this research. First of all, even though I use a relatively large sample of references regarding a single topic, they still all stem from the same database. This could be extended to different databases. Second of all, the ties have only been investigated on co-authorship. This implies a very limited inquiry on the strength of the existing ties. If possible, an interesting research could be to look into the different references made in and made to paper. This way one might be able to analyze the importance of a paper, based on the number of times it is referred to, and also where it would be positioned when conducting a network analysis.

7. Conclusion

In our analysis, we have found that the existing distribution of both the keywords, as well as (even though not in the same degree) the authors, follows Zipf's law, or in other words is a zipfian distribution. It is harder to distinguish a zipfian distribution in the cross-matrix of keywords to authors, as the keywords are more spread over the authors. When we look at ties between authors, we were able to find a number of mostly small networks, and one somewhat larger network. We were able though, to find a number of interesting links between authors. Especially in the network where three more prominent writers are connected. The analysis between keywords and authors did not result in very enlightening outcomes, as due to a number of highly occurring keywords, almost all authors were connected. Even when applying restrictions, only one separate group could be distinguished next to the group around the most frequently occurring keywords.

8. Literature

George K. Zipf, *Human Behaviour and the Principle of Least-Effort*, Addison-Wesley, Cambridge MA, 1949

Granovetter, Mark S.; The Strength of Weak Ties; *American Journal of Sociology*, Vol. 78, No. 6., May 1973, pp 1360-1380