

SITKIS: Software for Bibliometric Data Management and Analysis

Version 1.0

USER'S GUIDE

Henri Schildt

Helsinki University of Technology

Institute of Strategy and International Business,

Department of Industrial and Engineering Management

E-mail: henri.schildt@hut.fi

Latest update: March 10th, 2004

Table of Contents

1	Introduction	3
2	Networks: Exporting and Analysis	3
2.1	Articles to Citations (2-dimensional network).....	3
2.1.1	2-Mode Factor Analysis	5
2.1.2	2-Mode Categorical Core/Periphery Model	5
2.1.3	Calculating article similarity based on common references.....	5
2.1.4	Calculating co-citation networks from article-to-reference data.....	5
2.2	Co-Citation Networks.....	6
2.3	Authors to Articles.....	6
2.4	Author Co-Authorship	7
2.5	Institutes to Articles.....	7
2.6	University / Company Collaboration	7
2.7	Countries to Articles	7
2.8	Cross-border Research Collaboration	7
2.9	Directed Article Cross-Citations	7
3	Exporting Statistics	8
3.1	Reference Statistics.....	8
3.2	Yearly Citations Statistics.....	8
3.3	Article Statistics.....	9
3.4	Calculating Article / Reference Centrality.....	9
4	Additional Topics	10
4.1	Longitudinal Data Sets.....	10
5	Additional material	10
5.1	Bibliometric Basics	10
5.2	More recent methodological perspectives	10
5.3	Network Presentation	11
5.4	Example studies.....	11

1 Introduction

This document provides guidance for doing various analyses with Sitkis. The document is by no means final, and it will be updated constantly when I have time. The main goal is to provide a perspective on what can be done with Sitkis, and also help the reader to use UCINET in conjunction with Sitkis.

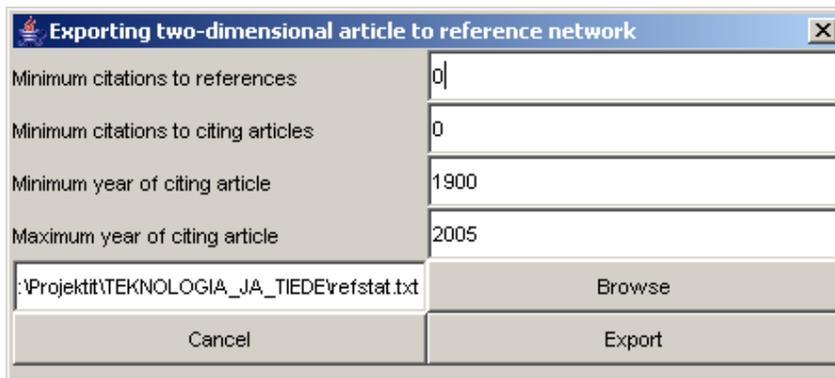
This document is ordered in the following way. First, the network export commands in Sitkis are introduced along with the related use of UCINET. Second, the commands to export different statistics from Sitkis to Excel are introduced and some possible analyses are discussed. Third, some suggestions on the use of Access to create new queries and match data are presented.

2 Networks: Exporting and Analysis

Creating networks from the data and exporting them to UCINET is relatively easy. The different networks are discussed each in turn. Each operation will utilize the database that has been defined with the command “*Options / Set database*”, and that is displayed in the main window. Please notice that only unfiltered articles are included, and that after importing new articles you need to run “*Options / Fix database*” command.

2.1 Articles to Citations (2-dimensional network)

To export a matrix that links the articles downloaded from ISI to the prior referenced articles, select from the pull-down menu “*Citations / Articles to References (2d)*”. The program will display a dialog for specifying the articles to export.



Exporting two-dimensional article to reference network	
Minimum citations to references	0
Minimum citations to citing articles	0
Minimum year of citing article	1900
Maximum year of citing article	2005
:\\Projektit\\TEKNOLOGIA_JA_TIEDE\\refstat.txt	Browse
Cancel	Export

The higher the number of minimum citations to references, the less referred articles will be included in the analysis and the faster the analysis will run (all numbers lower than 2 will include all articles). Similarly, “minimum citations to citing articles” will exclude all articles in your database that have lower than given number of received citations. This number is from all citations within ISI database at the time article was exported. You can also set the year range for citing articles, which allows a comparison of referred articles across time. It is advisable to test analyses with various minimum reference numbers. I have found that searches that include more than 30 most cited references provide are somewhat difficult to make sense of. Typically one would like to include references that are cited by at least 1% - 10% of all articles.

The analysis will run for some time. Afterwards, the file you have specified will include the network. You can now import the network to UCINET. Start UCINET 6 and press Ctrl+I (or choose “*Data / Import / DL ...*”). Select the file you created with Sitkis as the “input text file in DL format” and then specify the output file name. You have here now the 2-dimensional network that links ISI articles to cited articles.

Although there are various analyses that can be made with 2-mode data, most uses for the network will come from the co-citation (Citations-to-Citations) matrix that can be generated from this network file. However, below are some possible analyses with the 2-mode data.

2.1.1 2-Mode Factor Analysis

You can create a factor analysis out of the downloaded ISI articles and the cited articles by selecting “*Tools / 2-Mode Scaling / Factor Analysis*”. While I have not found the results very enlightening, I suppose they could be used to distinguish certain clear dimensions within the material.

2.1.2 2-Mode Categorical Core/Periphery Model

This command “*Network / 2-Mode / Categorical Core/Periphery ...*” distinguishes between a dense “core” of the articles and cited references and the less dense periphery. The appropriateness of the analysis depends on the existence of a dense core in the material. To my experience, the bibliometric data I have utilized has not provided a meaningful core/periphery distinction.

2.1.3 Calculating article similarity based on common references

You can create a network that describes the similarity of articles based on the number of common previous articles they cite. First, select “*Data / Affiliations ...*”, then choose the Articles to Citations network file you have imported to UCINET (see 2.1) as the input dataset, and the mode “Row”. In the resulting network, the strength of ties between articles is relative to the number of common references. You can use NetDraw to draw the resulting network. Just launch NetDraw and open the network. You can select “*Properties / Lines / Size / Tie Strength*” to have the line widths correspond with the relatedness of the articles.

2.1.4 Calculating co-citation networks from article-to-reference data

Please note that Sitkis has proper Co-Citation network export function (*Citations / Co-Citation Network*), which makes the process described in this chapter useless in most cases.

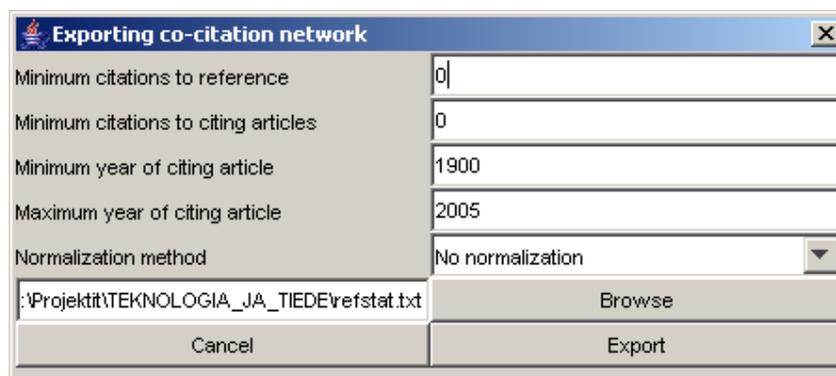
Co-citation networks are constructed in the similar manner as article similarity network, except after you have chosen “*Data / Affiliations ...*” you need to set the

mode to “Column”. In the resulting network the strength of ties corresponds to the number of articles that co-cite the given two references.

You can draw the network in Netdraw as described earlier, or you can run various analyses within the UCINET. I advice you to experiment and read some books on social network theory methods. One initial analysis you can run are “*Tools / Cluster / Hierarchical ...*”. This analysis creates a dendrogram or a tree, depicting the similarity of articles based on co-citations.

2.2 Co-Citation Networks

Like 2-mode network export, co-citation networks provide a dialog. In addition to selections described in chapter 2.1, you can set the normalization of the network. There are four different ways to normalize data.



Normalization divides the count of common citations to references a & b ($C[a,b]$) by indices calculated from total citations to references a and b individually ($C[a]$, $C[b]$). The standard measure $(C[a,b] / (C[a] + C[b] - C[a,b]))$ gives network ties with strength between 0 and 1, depending on the percentage of citations to each article that cite the other as well.

2.3 Authors to Articles

You can create a network export (.DL) file in Sikis by choosing “*Authors / Authors to Articles (2d)*” from the menu. You may then import it to UCINET with ctrl+I or by choosing “*Data / Import*”.

2.4 Author Co-Authorship

To create a co-authorship network, select “*Data / Affiliation*” and choose the author to articles file you have imported (see chapter 2.3). Select the mode “Row” for affiliations graph. Column mode creates a network of articles connected through matching author(s).

Co-authorship network may be interesting to look at in Netdraw, but remember that if your articles are from a very short time period most authors are unlikely to have published multiple articles and the density of the network will be very low.

2.5 Institutes to Articles

Just as 2.3 except you choose “*Authors / Institutes to Articles (2d)*”.

2.6 University / Company Collaboration

Just as 2.4 but you choose the file exported as described in chapter 2.5. Please note that you have to manually distinguish universities, companies and other institutes if your research question requires that. For example, you can calculate centralities for each institute and then mark them as company / university in Excel.

2.7 Countries to Articles

Just as in 2.3 except you choose “*Authors / Countries to Articles (2d)*”.

2.8 Cross-border Research Collaboration

Just as in 2.4 except you choose the file created as described in chapter 2.7. The network ties indicate the number of articles that have been created in cross-border collaboration between the two countries.

2.9 Directed Article Cross-Citations

Select “*Articles / Cross-citations (1d)*” and export the network file. Then import the file in UCINET (ctrl+I). The resulting network can be viewed with Netdraw. The network is directed, so the arrows point from a citing article to the cited article. I am

unfamiliar with suitable quantitative analysis methods for such cross-citation networks.

3 Exporting Statistics

Statistics can be exported to Excel-friendly tab-delimited text files.

3.1 Reference Statistics

This command is available from *Citations / Reference Statistics*, and exports a .txt file that can be opened with Excel. First the program asks for the name of the file, and then for the minimum number of citations references to be included should have.

Once you open the reference statistics in Excel, you can copy-paste the data to UCINET to form an attribute file. In Excel, mark and copy data in columns A & E, then go to Ucinet, open the editor (ctrl+S), and paste the data. Save as a Ucinet file. Now go to Netdraw where you have the related co-citation network open, select *Open / Ucinet dataset / Attribute data*, then select *Properties / Nodes / Size / Attribute based*, and your reference-nodes in the graph have diameter relative to the number of times they have been cited!

The file has also some potentially interesting information regarding the most cited references:

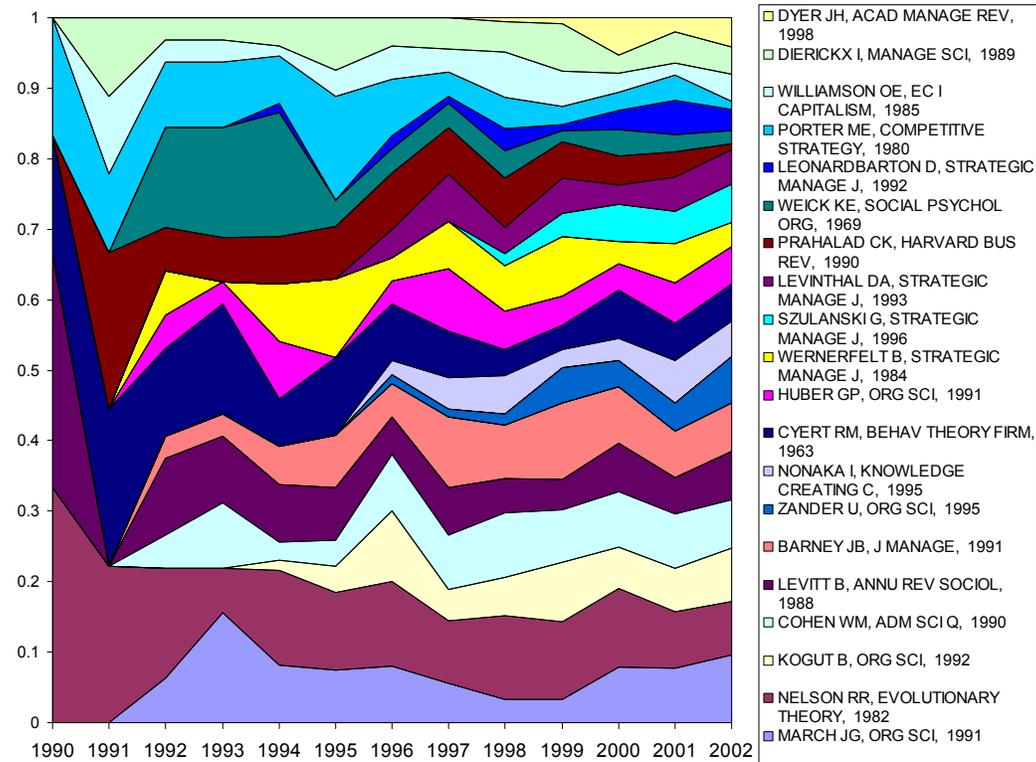
- Average, minimum and maximum year in which they are cited
- Average, minimum and maximum number of citations those articles that cite the reference have received

Basically, those references with a later average year of citation are “emerging” works, while those with earlier average year of citation may be “fading” from the conversation. Interpret with caution!

3.2 Yearly Citations Statistics

Basically here you have a set of data regarding citations from a range of years. After you have exported the .txt file, open it with Excel. You can sum() the number of

citations across years, and sort the references according to their total received citations. With some basic knowledge of Excel, you should be able to create graphs showing the relative “popularity” of various classics in different years, just like the one below.



3.3 Article Statistics

Article statistics are similar to reference statistics, but they are naturally about the stuff you have imported in the database. Nothing much to say about these, except that you can get the Times Cited count here and use it as attribute data in Netdraw / Ucinet.

3.4 Calculating Article / Reference Centrality

You can calculate various centrality scores for citing or cited articles in Ucinet. I will add some references for related work here later, since the meaning of centrality scores in bibliometric research are not trivial, and they have been debated.

4 Additional Topics

4.1 Longitudinal Data Sets

Although some operations already support year-range selection, all of them do not. For the time being one of the easiest ways to accomplish different time windows is to export various (yearly) files from ISI and creating multiple databases. For example you would then import files from a set of years to each database and calculate centrality scores. You could then for example compare the average times authors are cited with the centrality of previous publications or such...

5 Additional material

The following books, documents, and articles may prove useful resources. This list is partial, but will be updated when I have more time.

5.1 Bibliometric Basics

Small, H.G. (1973): Co-citation in the Scientific Literature: A New Measure of the Relationship Between Documents, Journal of American Society for Information Science.

- The seminal article for co-citation analysis

For an excellent collection of classic material, see “the Farther of Bibliometric Analysis” Eugene Garfield’s collection: <http://www.garfield.library.upenn.edu/>

5.2 More recent methodological perspectives

Osareh, F. (1996a, b): Bibliometrics, citation analysis, and co-citation analysis – a review of literature I and II, Libri.

Gmur, M (2003): Co-citation analysis and the search for invisible colleges: A methodological evaluation, Scientometrics.

- Compares clustering algorithms with data regarding organization theory

5.3 Network Presentation

Everton, S. F. 2004. A Guide for the Visually Perplexed: Visually Representing Social Networks. Available online at:

<http://www.stanford.edu/group/esrg/siliconvalley/documents/networkmemo.doc>

- A guide to visually representing networks with UCINET and related software

5.4 Example studies

Oliver, A.L., Ebers M. 1998. Networking Network Studies: An Analysis of Conceptual Configurations in the Study of Inter-organizational Relationships. *Organization Studies*, 19(4): 549-583.

- A network-based bibliometric study of organizational networks literature