


Sign in | My EndNote Web | My Citation Alerts

ISI Web of KnowledgeSM Take the next step 

Web of Science Additional Resources

Search | Cited Reference Search | Advanced Search | Search History | Marked List (0)

Web of Science®

Search for:

in

Example: oil spill AND "North Sea"*

AND in

Example: O'Brian C OR OBrian C**
Need help finding papers by an author? Use [Author Finder](#).

AND in

Example: Cancer OR Journal of Cancer Research and Clinical Oncology*

[Add Another Field >>](#)

Limit to: [\(Change Limits\)](#)
Timespan=All Years . Databases=Science Citation Index Expanded (SCI-EXPANDED); Social Sciences Citation Index (SSCI); Arts & Humanities Citation Index (A&HCI)

WoS

WoS2Pajek networks from Web of Science version 0.7

Vladimir Batagelj

FMF, matematika

University of Ljubljana

Manual

Ljubljana, August 2009 / December 2007

Index

1	Searching on the Web of Science	1
4	Using the Advanced Search	4
8	The list of citing articles	8
11	Structure of a WoS record	11
12	Names of works	12
14	Program WoS2Pajek	14
25	Types on DC file	25
26	Analyses	26

Searching on the Web of Science



The screenshot shows the ISI Web of Knowledge search results page. The search query is "social network*" and the results are sorted by "Latest Date". The page displays 6,936 results. The "Refine Results" sidebar on the left shows filters for Subject Areas and Document Types. The main results list shows three entries:

- 1. Title: **Managerial social capital, strategic orientation, and organizational performance in an emerging economy**
Author(s): Acquaaah M
Source: **STRATEGIC MANAGEMENT JOURNAL** Volume: 28
Issue: 12 Pages: 1235-1255 Published: DEC 2007
Times Cited: 0
[Full Text](#)
- 2. Title: **"Freeter" selection and a social network: From the course consideration investigation of the third grade of high-school**
Author(s): Uchida R
Source: **SOCIOLOGICAL THEORY AND METHODS** Volume: 22
Issue: 2 Pages: 139-153 Published: 2007
Times Cited: 0
- 3. Title: **Up close and personal: Employee networks and job satisfaction in a human service context**
Author(s): Haley-Lock A
Source: **SOCIAL SERVICE REVIEW** Volume: 81 Issue: 4 Pages: 683-707 Published: DEC 2007
Times Cited: 0
[Full Text](#)

The Web of Science – WoS (ISI/Thomson) allows us to save on a file the records corresponding to our queries.

For example, using **General search** with a query "social network*" we get 6936 hits (27. December 2007).

Trying to save them we are informed that we can save at once at most 500 records. We have to save the records by parts on separate files. At the end we concatenate all these files into a single file.

Saving the records

Results: **6.936** Show 10 per page Page of 694 Sort by: Latest Date

Output Records

Step 1:

- Selected Records on page
- All records on page
- Records to

Step 2:

- Authors, Title, Source
 - plus Abstract
- Full Record
 - plus Cited Reference

Step 3:

At the bottom of the page in the **Output Records** select **Records** and enter the interval bounds *firstRec* to *lastRec* on record numbers that you want to save.

Select **Full Record + Cited Reference**.

Select also **- as Plain Text** and click on the **Save** button.

...Saving the records

ISI Web of KnowledgeSM Take the next step

Processing Records

Please wait while your request is processed.
(Note: Depending on the number of records, this may take a few moments.)

Product: Web of Science
Selected action: Save to File
Processing 500 records:
10...20...30...40...50...60...70...80...90...100...
110...120...130...140...150...160...170...180...190...200...
210...220...230...240...250...260...270...280...290...300...
310...320...330...340...350...360...370...380...390...400...
410...420...430...440...450...460...470...480...490...500...Done.

Save If the "Save As" dialog does not appear automatically, then click "Save" to download the records.

Back to Results When you are done saving the file, click "Back to Results."

Opening savedrecs.txt

The site has suggested that "savedrecs.txt" be handled as an attachment. It is of type application/octet-stream (Text Document) and located at:
http://pcs.newisiknowledge.com

What should SeaMonkey do with this file?

Open it with the default application (txtfile)

Open it with **Choose...**

Save it to disk

Always perform this action when handling files of this type

OK **Cancel**

THOMSON

In a new window the export process starts ...it takes some time ...wait until done. Select **Save it to disk** and click **OK**. When the file-chooser appears determine the file on which the records are saved.

Clicking on the **Back to Results** button you return back to the results window.

Repeat these steps until all the records are saved on files.

Using the Advanced Search

ISI Web of KnowledgeSM Take the next step

Web of Science **Additional Resources**

Search | Cited Reference Search | **Advanced Search** | Search History | Marked List (0)

Web of Science®

Advanced Search. Use 2-character tags, Boolean operators, parentheses, and set references to create your query. Results appear in the Search History at the bottom of the page.

Example: TS=(nanotub* SAME carbon) NOT AU=Smalley RE #1 NOT #2 [more examples](#) | [view the tutorial](#)

TS=(centrali* AND (network* OR graph))

Current Limits: [Hide Limits and Settings](#)

Timespan:

All Years (updated 2008-12-20)

From 1970 to 2008 (default is all years)

Search History

Set	Results		Combine Sets	Delete Sets
# 1	3,199	TS=(centrali* AND (network* OR graph)) <small>Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years</small>	<input type="radio"/> AND <input type="radio"/> OR <input type="button" value="Combine"/>	<input type="button" value="Select All"/> <input type="button" value="Delete"/>

At the computer with access to Web of Science (at Uni-LJ you can use the **IZUM** and select the option **ISI Web of Knowledge (Web of Science) - na strežniku Thomson Reuters**).

Once on the WoS we select the folder **Advanced Search** and enter our query – for example:

TS=(centrali* AND
(network* OR graph))

If necessary we can set also the time bounds (WoS allows only up to 100000 hits in a query).

We obtain the information about the number of hits at the bottom of the page.

Get the list of hits and save selected on file

The screenshot shows the Web of Science search results page. At the top, it displays the search query: **Results** TS=(central* AND (network* OR graph)) with a timespan of All Years and databases including SCI-EXPANDED, SSCI, and A&HCI. The results count is 3,199, and the page is 1 of 320. The search is sorted by Latest Date. On the left, there is a 'Refine Results' section with a search box and subject areas like ENGINEERING, ELECTRICAL & ELECTRONIC (813), TELECOMMUNICATIONS (641), and COMPUTER SCIENCE, THEORY &. The main results list shows the first entry: 'Saturation Throughput Analysis of a Cluster-based Medium Access Control Protocol for Single-hop Ad Hoc Wireless Networks' by Alonso-Zarate J, Kartsakli E, Skianis C, et al. The source is SIMULATION-TRANSACTIONS OF THE SOCIETY FOR MODELING AND SIMULATION INTERNATIONAL, Volume 84, Issue 12, Pages 619-633, published in DEC 2008, with 0 times cited. Below the results, there are three steps for output records: Step 1 (Selected Records on page, All records on page, or Records 1501 to 2000), Step 2 (Authors, Title, Source; plus Abstract; Full Record; plus Cited Reference), and Step 3 (Print, E-mail, Add to Marked List, Save to EndNote@Web, Save to EndNote@RefMan, ProCite, Save to Plain Text, and Save).

To get the list of hits we click to their number (blue 3,199 in our case).

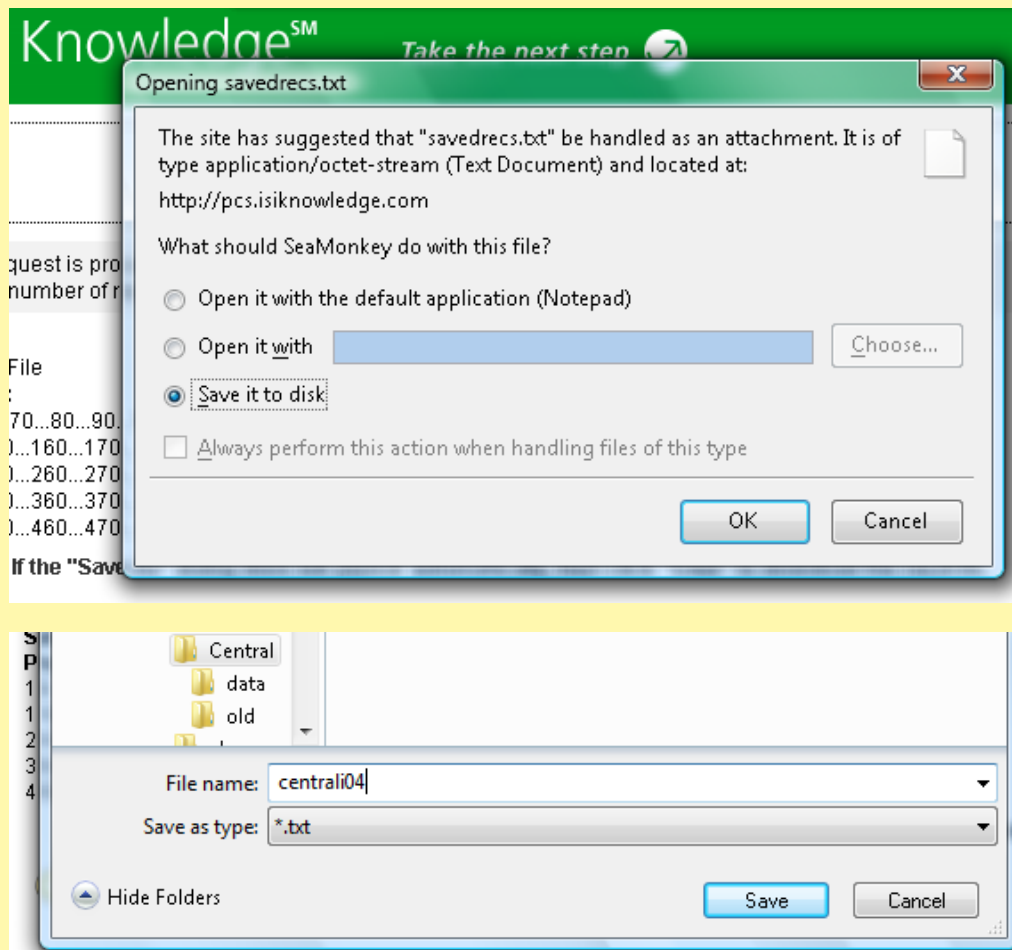
At the bottom of this page we can request that some of the hits are saved to the file. For longer lists we have to do this by parts - WoS allows only 500 hits to be saved at once.

To save selected hits we proceed as follows:

- * **step 1:** determine the range of hits to be saved (1-500, 501-1000, 1001-1500, ...);
- * **step 2:** select **Full Record** and **plus Cited Reference**;
- * **step 3:** select **Save to Plain Text**.

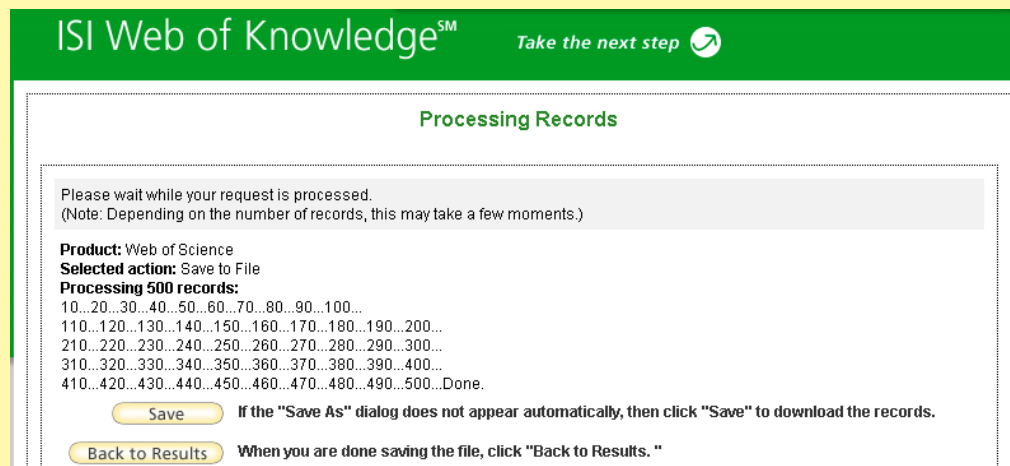
Finally we click on the **Save** button.

... saving



A new page **Processing Records** appears. We have to wait until the selected records are processed and written to the file. In the window that appears we select the option **Save to Disk** and click **OK**. In a new window that appears we select the directory and enter the name of the file on which the selected hits are saved, for example `Centrali004.txt`. Finally we click on the **Save** button.

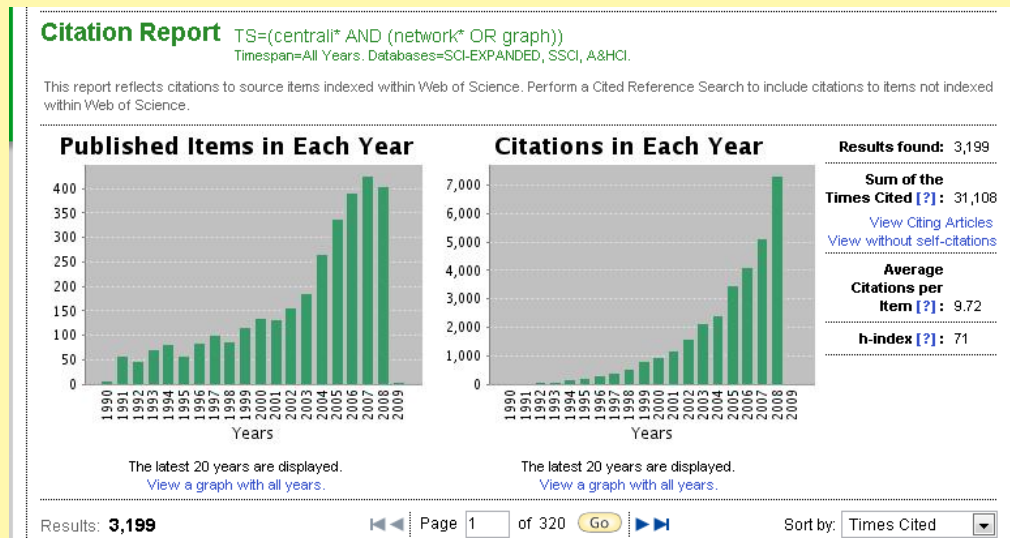
... saving



To return back to the saving of selected hits we click on **Back to Results**.

We repeat the procedure described in this subsection until all the hits are saved.

The list of citing articles



We return to the top of the page with list of hits - see the picture in the subsection **Get the list of hits**. In the upper right corner we click on the option **Create Citation Report**. We obtain a new page with histograms.

To obtain the list of citing articles we click on the option **View Citing Articles**.

To save them we repeat the procedure described in subsection **Save the selected hits to file**.

Web of Science®

<< [Back to Citation Report](#)

Total Citing Articles TS=(central* AND (network* OR graph))
Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

Results: **24,977** Page 1 of 2,498 Go Sort by: Latest Date

[Print](#) [E-mail](#) [Add to Marked List](#) [Save to EndNote® Web](#) [Analyze Results](#)
[Save to EndNote, RefMan, ProCite](#) more options

Refine Results
Search within results for: [Search](#)

▼ **Subject Areas** [Refine](#)
 ENGINEERING, ELECTRICAL &

1. Title: **A mobility model for classical swine fever in feral pig populations**
Author(s): Milne G, Fermanis C, Johnston P
Source: **VETERINARY RESEARCH** Volume: 39 Issue: 6 Article Number: 53
Published: **NOV-DEC 2008**
Times Cited: 0

Additional records

The screenshot shows the Web of Science interface. At the top, there is a navigation bar with 'Web of Science' and 'Additional Resources'. Below this, there are tabs for 'Search', 'Cited Reference Search', 'Advanced Search', 'Search History', and 'Marked List (26)'. The 'Advanced Search' tab is active. The main content area is titled 'Web of Science®' and contains an 'Advanced Search' section. It provides instructions on using 2-character tags, Boolean operators, parentheses, and set references. An example query is given: 'TS=(nanotub* SAME carbon) NOT AU=Smalley RE #1 NOT #2'. Below the instructions is a search input field containing the query 'au=(TAKANE Y*) and py=(1977)' and a 'Search' button. Below the search input is a 'Search History' section. It contains a table with columns for 'Set', 'Results', 'Combine Sets', and 'Delete Sets'. The table lists four search sets with their respective queries and results.

Set	Results		Combine Sets	Delete Sets
# 10	1	au=(TAKANE Y*) and py=(1977) <i>Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years</i>	<input type="checkbox"/>	<input type="checkbox"/>
# 9	2	au=(FOWLKES E*) and py=(1983) <i>Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years</i>	<input type="checkbox"/>	<input type="checkbox"/>
# 8	2	au=(CALINSKI T*) and py=(1974) <i>Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years</i>	<input type="checkbox"/>	<input type="checkbox"/>
# 7	3	au=(HARTIGAN J*) and py=(1979) <i>Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=All Years</i>	<input type="checkbox"/>	<input type="checkbox"/>

At WoS we enter the advanced search and for an entry from the list, for example

97

"FELSENST_J (1985) 39:783"

we enter a query

au=(FELSENST* J*) and

py=1985

In the list of hits at the bottom of the page click the blue number of hits to obtain the list of their basic descriptions.

Using the information about the volume and the first page, 39 and 783 in our example, identify the corresponding work (if it exists), check the box in front of it and then click the button **Add to Marked List** at the beginning of the list. After addition of the work to the Marked list the red check mark will appear in front of the work (see picture). Repeat the described procedure for other entries.

...Additional records

The screenshot shows the Web of Science interface. At the top, there are navigation tabs: Search, Cited Reference Search, Advanced Search, Search History, and Marked List (27). Below this, the search results are displayed for the query 'au=(TAKANE Y*) and py=(1977)'. A single result is shown with the title 'NONMETRIC INDIVIDUAL-DIFFERENCES MULTIDIMENSIONAL-SCALING - ALTERNATING LEAST-SQUARES METHOD WITH OPTIMAL SCALING FEATURES' by TAKANE Y, YOUNG FW, DELEEUW J. The result is marked with a checkmark.

Below the search results, there is a 'Web of Science Marked Records - 43 Articles' section. This section contains a configuration window for selecting fields and options for the marked records. The window is divided into two steps:

Step 1. Select the fields to include in the output. This step includes a grid of checkboxes for various fields:

<input checked="" type="checkbox"/> Author(s)	<input checked="" type="checkbox"/> Title	<input checked="" type="checkbox"/> Source	<input checked="" type="checkbox"/> language
<input checked="" type="checkbox"/> abstract*	<input checked="" type="checkbox"/> cited references*	<input checked="" type="checkbox"/> document type	<input checked="" type="checkbox"/> subject category
<input checked="" type="checkbox"/> addresses	<input checked="" type="checkbox"/> times cited	<input checked="" type="checkbox"/> keywords	<input checked="" type="checkbox"/> publisher information
<input checked="" type="checkbox"/> ISSN	<input checked="" type="checkbox"/> cited reference count	<input checked="" type="checkbox"/> source abbrev.	<input checked="" type="checkbox"/> page count
<input checked="" type="checkbox"/> IDS number	<input checked="" type="checkbox"/> funding information		

*Selecting these items will increase the processing time.

Step 2. Select an option. This step includes a dropdown menu for 'Field Tagged' (set to 'Plain Text'), buttons for 'Format for Print', 'Save to My EndNote Web', 'Save to EndNote, RefMan, ProCite', and 'Save to File'. There are also input fields for 'E-mail records to:', 'Return e-mail (optional):', and 'Notes (optional):', along with a dropdown for 'Plain Text' and an 'E-mail' button.

At the bottom of the configuration window, there is a checkbox: Automatically delete selected records from the Marked List after output is complete.

When the list of hits becomes too long click the **Select All** button in its Delete Sets column and after it the **Delete** button. The list of hits will empty.

To save the works from the Marked List click on Marked List at the top of the page. In the new window select all options in **Step 1** and in **Step 2** select the **Plain Text** option in front of **Save to File** button and click on this button.

Structure of a WoS record

```

PT J
AU KOSMELJ, K
   BATAGELJ, V
TI CROSS-SECTIONAL APPROACH FOR CLUSTERING TIME-VARYING DATA
SO JOURNAL OF CLASSIFICATION
DT Article
CR *UN, 1979, STAT YB
   *UN, 1981, STAT YB
   *UN, 1982, STAT YB
   ANDERBERG MR, 1973, CLUSTER ANAL APPLICA
   BATAGELJ V, 1981, CLUSE CLUSTERING PRO
   BATAGELJ V, 1988, 2ND M YUG SECT CLASS
   BATAGELJ V, 1988, CLASSIFICATION RELAT, P67
   GORDON AD, 1981, CLASSIFICATION
   KOSMELJ K, 1983, REV STAT APPL, V31, P5
   KOSMELJ K, 1986, J MATH SOCIOL, V12, P315
TC 7
SN 0176-4268
J9 J CLASSIF
JI J. Classif.
PY 1990
VL 7
IS 1
BP 99
EP 109
SC Mathematics, Interdisciplinary Applications; Psychology, ...
UT ISI:A1990DE57600006
ER

```

Names of works

The usual *ISI name* of a work (field CR)

LEFKOVITCH LP, 1985, THEOR APPL GENET, V70, P585

has the following structure

AU + ', ' + PY + ', ' + SO[:20] + ', V' + VL + ', P' + BP

All its elements are in upper case.

In WoS the same work can have different ISI names. To improve the precision the program **WoS2Pajek** supports also *short names* (similar to the names used in HISTCITE output). They have the format:

LastNm[:8] + ' _' + FirstNm[0] + ' (' + PY + ') ' + VL + ' :' + BP

For example: LEFKOVIT_L (1985) 70 : 585

From the last names with prefixes VAN, DE, ... the space is deleted.

Unusual names start with character * or \$.

...Names of works

In the CR field other forms of ISI names and several errors and inconsistencies can be found:

```

NEWMAN MEJ, 2004, PHYS REV E 2, V69, ARTN 066133
PALLA G, 2005, NATURE, V435, P814, DOI 10.1038/nature03607
PAPIN JA, 2004, TRENDS BIOCHEM SCI, V29, P641, DOI
10.1016/j.tibs.2004.10.001
DOLCINI MM, 2005, J ADOLESCENT HEALTH, V36, UNSP 267.E6-15
EVANS JD, 2001, GENOME BIOL, V2, UNSP RESEARCH0001
NEWMAN MEJ, 2001, IN PRESS COMPLEX NETUNSP 215239
GRANOVET.MS, 1973, AM J SOCIOL, V78, P1360
GRANOVETTER M, 1983, SOCIOLOGICAL THEORY, V1, P203
BORGATTI SP, 2002, UGINET WINDOWS SOFTW
BORGATTI S, 1999, UCINET V USERS GUIDE
CANTANZARO M, 2005, PHYS REV E, V71, UNSP 027103
CANTAZARO M, 2005, PHYS REV E, V71, UNSP 056104
CATANZARO M, 2005, PHYS REV E 2, V71, ARTN 056104
BRICKER PD, 1968, OCT M PSYCH SOC ST L : BRICKER

```

We decided to treat in short names the ARTN and UNSP values as BP values. We also remove the DOI parts. There are also irregular names in AU field:

```

AU BENSON, , C
    KULHAVY, , W
AU SCHONEMA.PH

```

The user can correct the typing errors and nonuniformities on the WoS file.

Program WoS2Pajek

For converting WoS file into networks in **Pajek**'s format a program **WoS2Pajek** was developed (in Python). It produces the following files:

- citation network: works \times works;
- authorship (two-mode) network: works \times authors, for works without complete description only the first author is known;
- keywords (two-mode) network: works \times keywords, only for works with complete description;
- journals (two-mode) network: works \times journals, field J9;
- partition of works by the publication year;
- partition of works – complete description (1) / ISI name only (0);
- vector number of pages, PG or EP – BP +1.

Program **WoS2Pajek**

The keywords are obtained from the fields **TI** (title), **ID**, **DE** and **AB** (abstract). From the text the **stopwords** are removed and a list of words is produced. The words are lemmatized using **MontyLingua** package.

In future versions additional networks can be derived: works × discipline, works × countries, ...

In version 0.7 a GUI support (based on Tkinter) for specifying the program parameters was implemented.

Program **WoS2Pajek** can be run as an executable program by double-clicking on its icon – see slide 21.

The source code can be executed in different ways using the Python interpreter. See slides 19, 22 and 23.

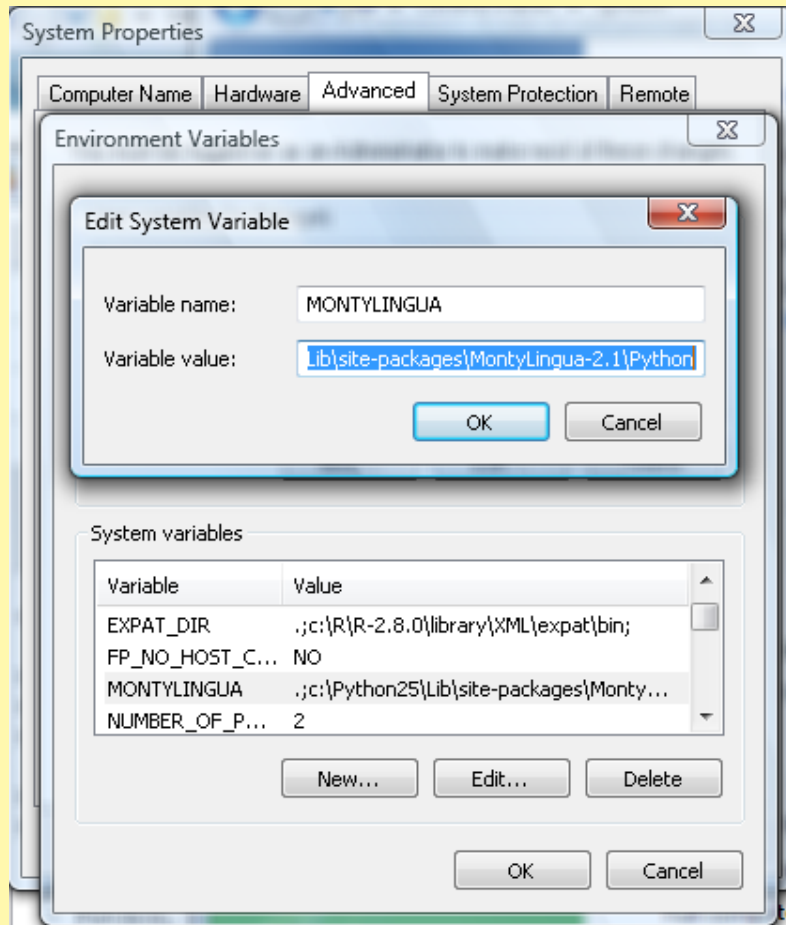
Program **WoS2Pajek**

The current version of **WoS2Pajek** requires 7 parameters to be given by the user:

- MontyLingua directory: path to the directory in which the MontyLingua package is installed (put it also in the `PATH` env-variable);
- project directory: where the output files are saved;
- WoS file;
- maxnum – estimate of the number of all vertices (number of records + number of cited Works) – $30 \times$ number of records;
- step – prints info about each $k \times$ step record as a trace; step = 0 – no trace.
- use ISI name / short name;
- make a clean WoS file without duplicates;
- boolean list [DE, ID, TI, AB] specifying which fields are sources of keywords.

Program **WoS2Pa jek**– details

To use **WoS2Pa jek** program you need to install at your computer:



- Python, version 2.5
- download **WoS2Pa jek 0.7** (latest version)
- **MontyLingua package**
- Copy the MontyLingua package into directory `Python25\Lib\site-packages\montylingua-2.1\`
- add to the environment variable MONTYLINGUA (or PATH) the path to MontyLingua (see the picture):
Control Panel/ System/ Advanced System Settings/ Environment Variables/ New

...Program WoS2Pajek– details

- **WoS2Pajek** expects in the subdirectory `resources` (of directory in which it is located) the files `StopWords.dat` and `Pajek.ico`;
- run Python and use the commands similar to the following:

```
>>> import sys; wdir = r'c:\users\Batagelj\work\Python\WoS'  
>>> sys.path.append(wdir)  
>>> MLdir = r'c:\Python25\Lib\site-packages\MontyLingua-2.1\Python'  
>>> sys.path.append(MLdir)  
>>> import WoS2Pajek
```

A dialog box will appear in which we specify required parameters and press the RUN button.

WoS2Pajek 0.6 works nicely also on 64-bit machines with `python-2.5.4.amd64`.

Running WoS2Pajek 0.7 / from Python interpreter

```
>>> import sys; wdir = r'c:\users\Batagelj\work\Python\WoS'; sys.path.append(wdir)
>>> MLdir = r'c:\Python25\Lib\site-packages\MontyLingua-2.1\Python'
>>> sys.path.append(MLdir)
>>> import WoS2Pajek
Module Wos2Pajek imported.

*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007

WoS2Pajek parameters
WoS dir: c:\users\Batagelj\work\Python\WoS
ML dir: c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
Proj dir: C:/Users/Batagelj/work/Python/WoS/batagelj
WoS file: C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
MaxNum : 1000
step : 10
ISI name: False
clean : True
keywords: [True, True, False, False]

***** MontyLingua v.2.1 *****
***** by hugo@media.mit.edu *****
Lemmatiser OK!
Custom Lexicon Found! Now Loading!
Fast Lexicon Found! Now Loading!
Lexicon OK!
LexicalRuleParser OK!
ContextualRuleParser OK!
Commonsense OK!
Semantic Interpreter OK!
Loading Morph Dictionary!
*****

*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007

started: Mon Aug 24 03:19:29 2009

10 : DOREIAN_P(2000)17:3 - 2009-08-24 03:19:29.614000
20 : BATAGELJ_V(1994)11:93 - 2009-08-24 03:19:30.134000
30 : BATAGELJ_V(1984)52:113 - 2009-08-24 03:19:30.426000
36 : BATAGELJ_V(1975)18:216 - 2009-08-24 03:19:30.640000
>>> End of processing of WoS file
```

...Running WoS2Pajek 0.7 / from Python interpreter

```
number of works      = 371
number of authors    = 230
number of journals    = 102
number of keywords    = 82
number of records    = 36
number of duplicates = 0
clean WoS data: clean.WoS
```

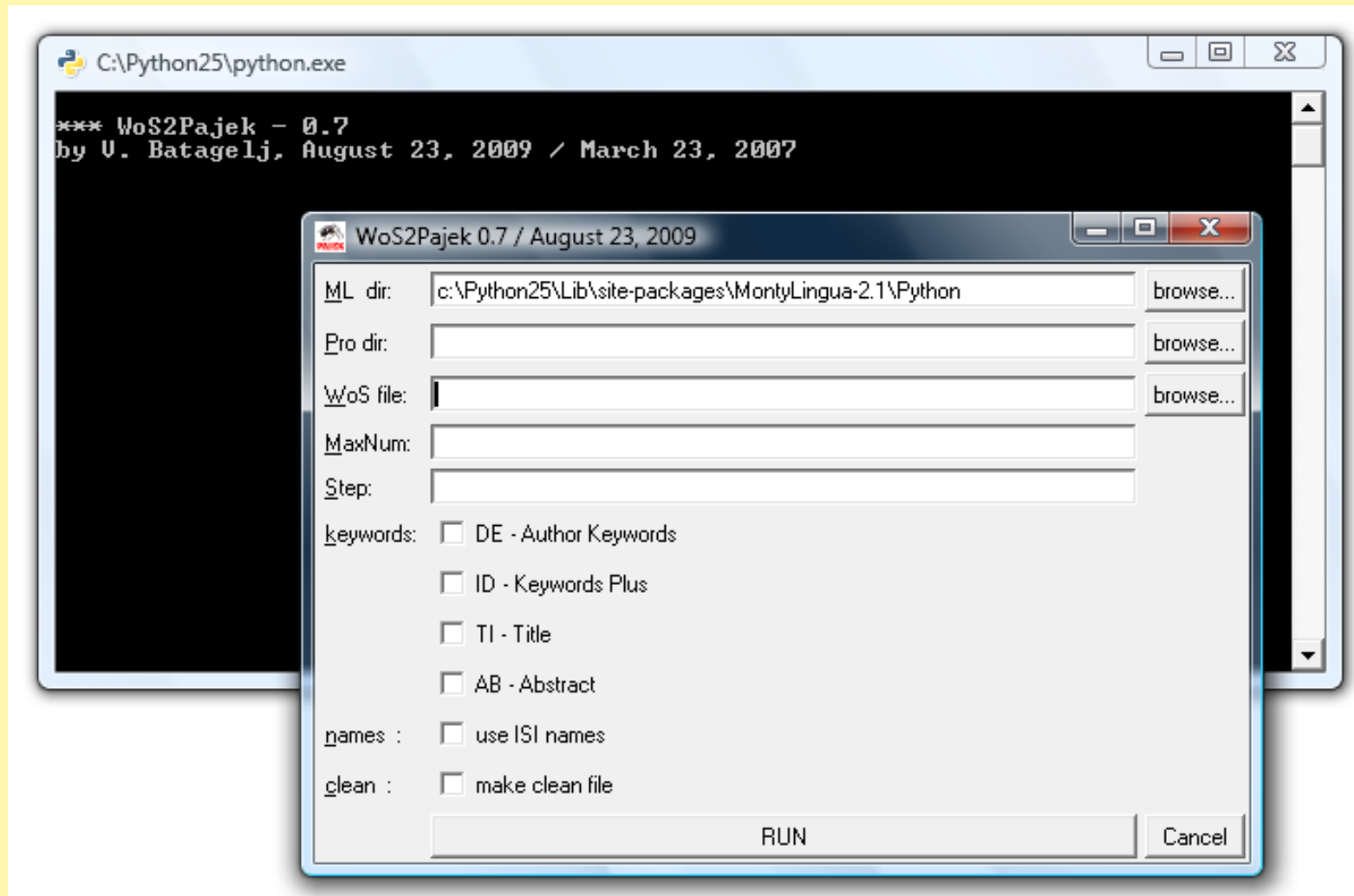
*** FILES:

```
year of publication partition: C:/Users/Batagelj/work/Python/WoS/batagelj\Year.clu
described / cited only partition: C:/Users/Batagelj/work/Python/WoS/batagelj\DC.clu
number of pages vector: C:/Users/Batagelj/work/Python/WoS/batagelj\NP.vec
citation network: C:/Users/Batagelj/work/Python/WoS/batagelj\Cite.net
works X journals network: C:/Users/Batagelj/work/Python/WoS/batagelj\WJ.net
works X keywords network: C:/Users/Batagelj/work/Python/WoS/batagelj\WK.net
works X authors network: C:/Users/Batagelj/work/Python/WoS/batagelj\WA.net
finished: Mon Aug 24 03:19:30 2009
time used: 0:00:01.770000
```

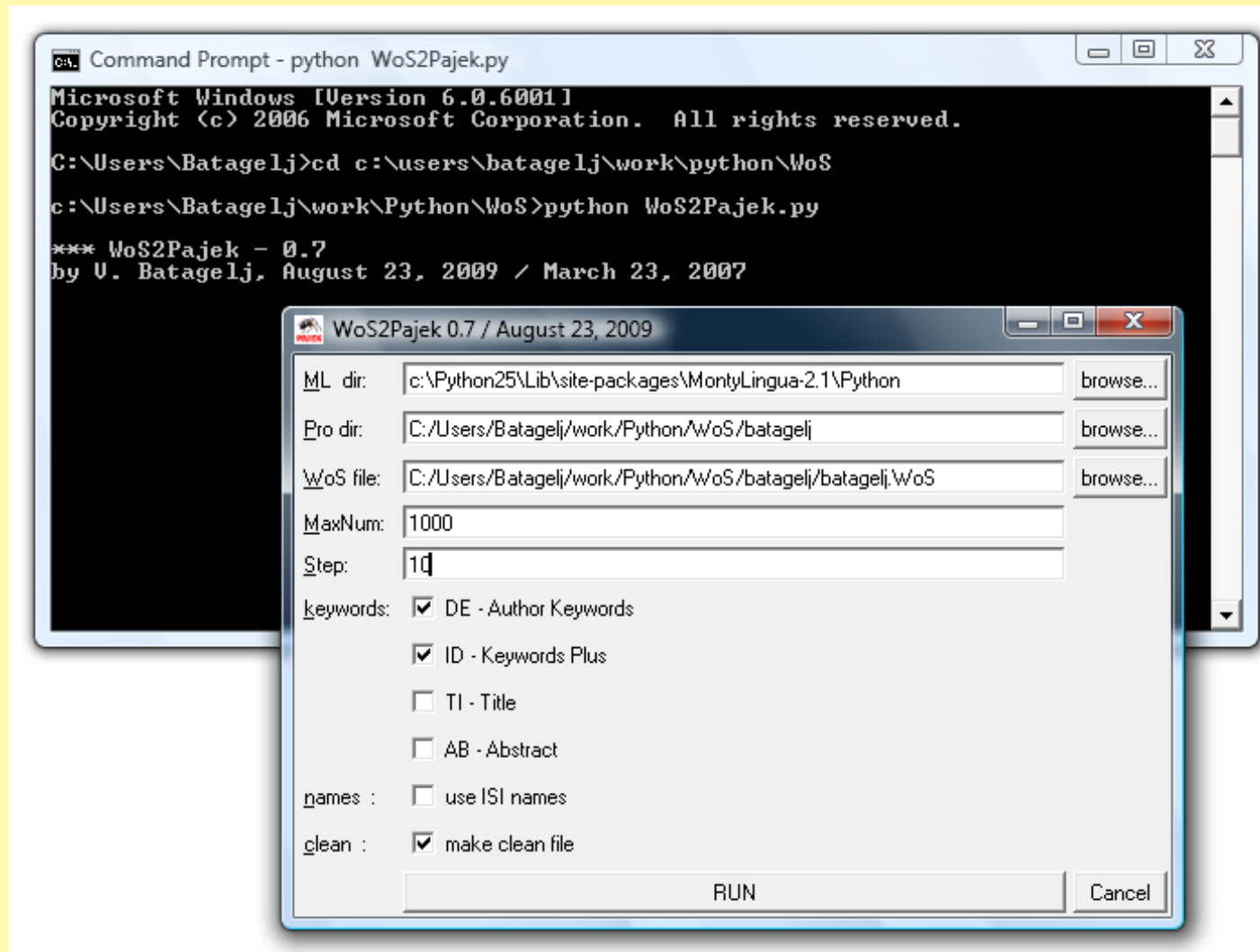
To rerun, type:

```
    reload(WoS2Pajek)
<module 'WoS2Pajek' from 'c:\users\Batagelj\work\Python\WoS\WoS2Pajek.py'>
>>>
```

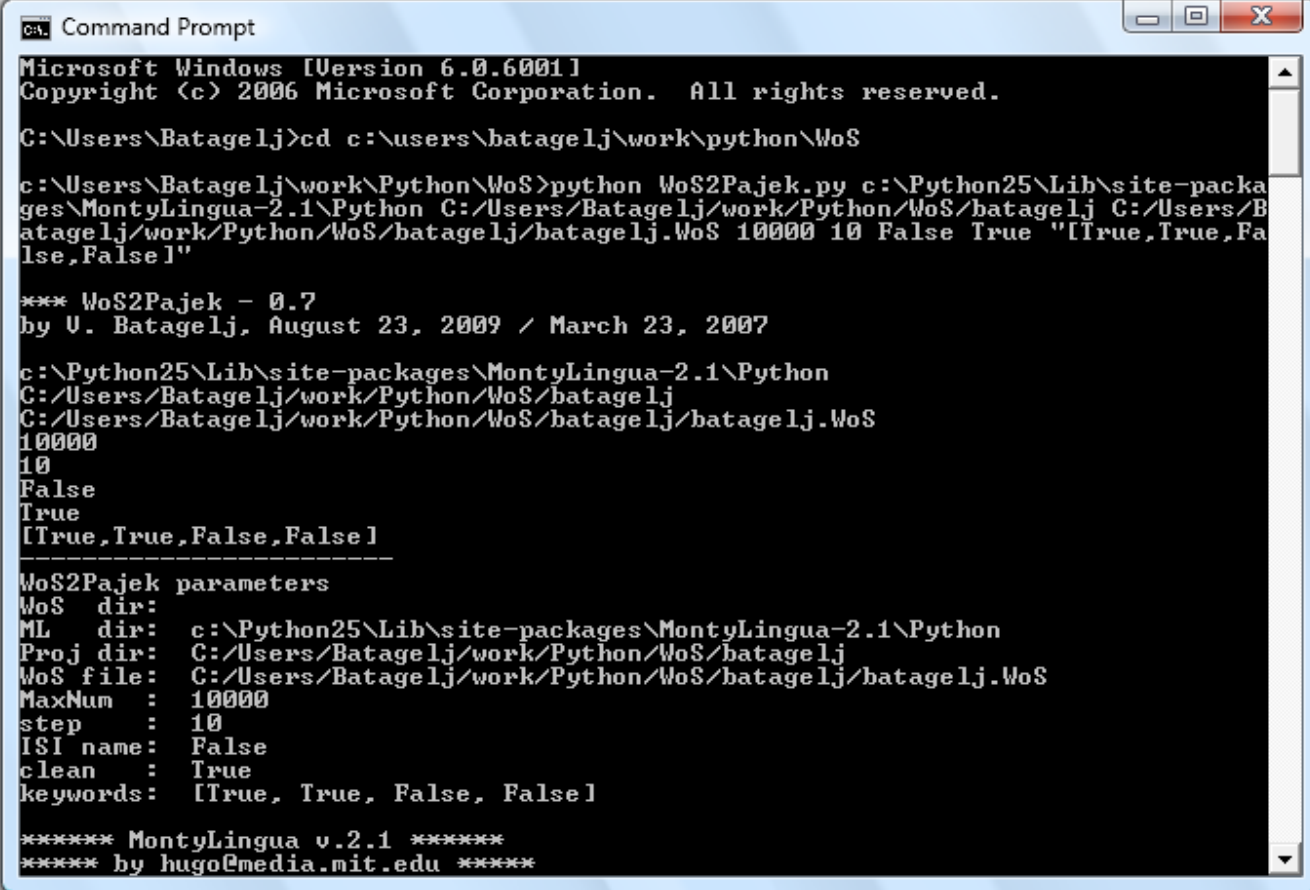
Running WoS2Pajek / Python by double-clicking it



Running WoS2Pajek / Python from Dos window



Running WoS2Pajek / Python from Dos window using parameters



```
ca. Command Prompt
Microsoft Windows [Version 6.0.6001]
Copyright (c) 2006 Microsoft Corporation. All rights reserved.

C:\Users\Batagelj>cd c:\users\batagelj\work\python\WoS

c:\Users\Batagelj\work\Python\WoS>python WoS2Pajek.py c:\Python25\Lib\site-packa
ges\MontyLingua-2.1\Python C:/Users/Batagelj/work/Python/WoS/batagelj C:/Users/B
atagelj/work/Python/WoS/batagelj/batagelj.WoS 10000 10 False True "[True,True,fa
lse,False]"

*** WoS2Pajek - 0.7
by V. Batagelj, August 23, 2009 / March 23, 2007

c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
C:/Users/Batagelj/work/Python/WoS/batagelj
C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
10000
10
False
True
[True,True,False,False]

-----
WoS2Pajek parameters
WoS dir:
ML dir: c:\Python25\Lib\site-packages\MontyLingua-2.1\Python
Proj dir: C:/Users/Batagelj/work/Python/WoS/batagelj
WoS file: C:/Users/Batagelj/work/Python/WoS/batagelj/batagelj.WoS
MaxNum : 10000
step : 10
ISI name: False
clean : True
keywords: [True, True, False, False]

***** MontyLingua v.2.1 *****
***** by hugo@media.mit.edu *****
```

Types on DC file

When we combine partial files with saved records from WoS into a single file required by the program **WoS2Pajek** we can include into this file some additional lines: Comments have the form

```
** comment
```

Besides this we can specify different types of input records using the lines of the form

```
*T n
```

where n is a type number (1, 2, ...). Since the same record can appear in different parts of the file its class is determined as the set of all corresponding types transformed in integer. For example: $\{3, 1\} \rightarrow 5$.

Analyses

The saved records from WoS can still contain some inconsistencies:

- different names for the same person;
- same name for different persons;
- duplicated entries;
- ...

Some of them are detected as results of the analyses. The simplest way to deal with them is to correct them in the saved WoS file and rerun the creation of **Pajek**'s files and analyses.

To improve the quality of the data some tools for detecting (possible) inconsistencies could be developed.

Check (in **Pajek**) the obtained networks for multiple lines and remove them, if they exist. Remove also the loops from the citation network.

Preparing the citation network

Using on *PRcite.net* the commands

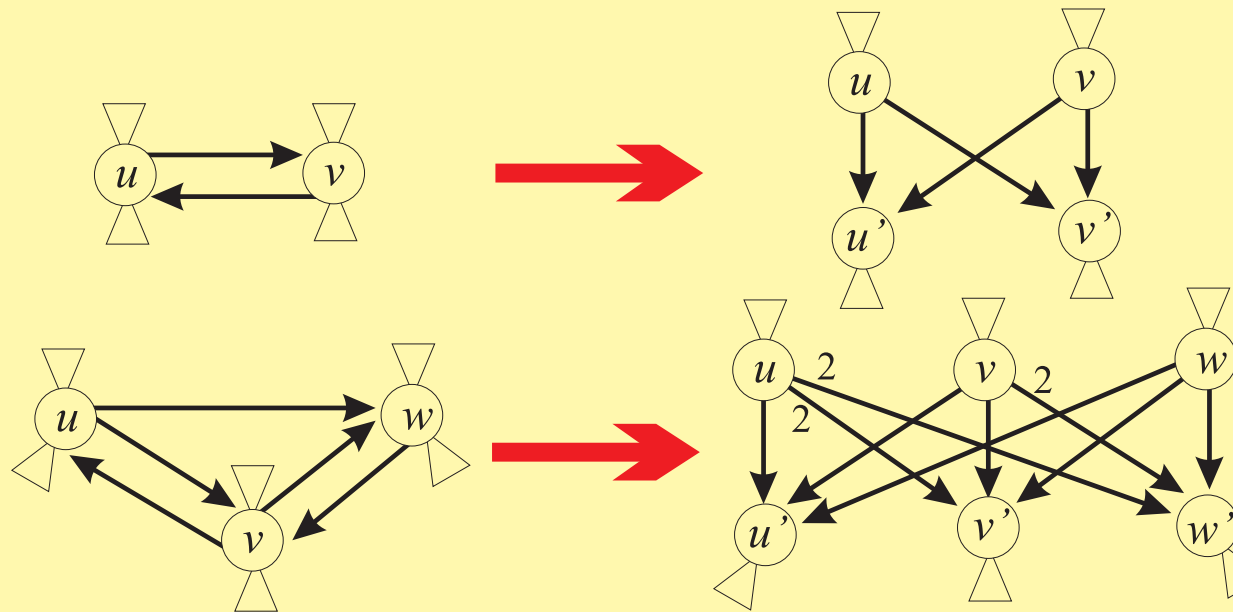
```
Info/Network/General  
Net/Transform/Remove/Loops  
Net/Transform/Remove lines/Single line
```

we get the information about the number of loops and multiple lines, remove loops, and replace multiple lines with single lines. The obtained network we save (Options – Save coordinates [OFF]) to file *PRciteR.net*. For further analysis the citation network has to be acyclic – has no nontrivial strong component. To identify nontrivial strong component and extract them use the commands:

```
Net/Components/Strong [2]  
Operations/Extract from Network/Partition [1-*]  
Operations/Transform/Remove Lines/Between Clusters
```

Save the obtained network to file *PRstrong.net*.

...Preparing the citation network



To transform the network `PRciteR.net` into acyclic network using the **preprint** transformation use the program **Preprint**

```
import Preprint;
Preprint.run(wdir, 'PR', 'PRciteR.net', 'PRstrong.net')
```

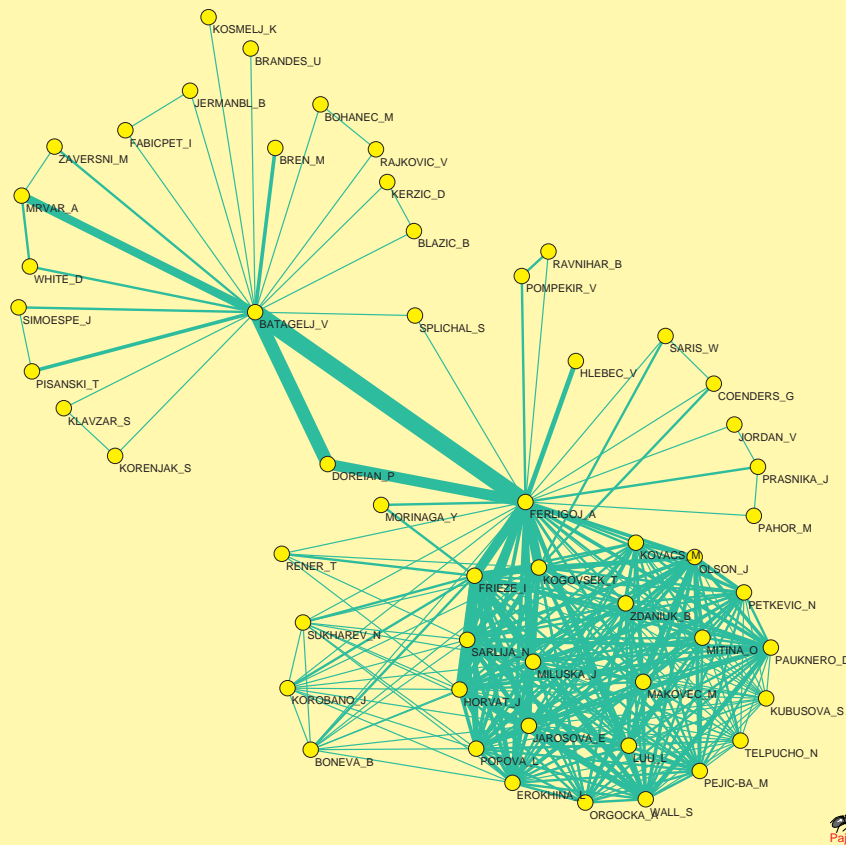
...Analyses: network boundary problem

Networks obtained from the WoS file using the program **WoS2Pajek** are in the 'raw' form. We still have to resolve in some way the *network boundary problem*. The first option is to limit the network to the works with complete descriptions – records from the WoS file. We can get a richer network if we decide to include also some referenced (only) works that are referenced often – at least k times; we delete vertices for which it holds

$$(0 < \text{indeg}(v) < k) \wedge (\text{outdeg}(v) = 0)$$

```
Net/Partition/Degree/Input
Partition/Binarize [1-(k-1)]
Net/Partition/Degree/Output
Partition/Binarize [0]
[select partition 1]
[select partition 2]
Partitions/Min(V1,V2)
Operations/Extract from Network/Partition [0]
```

...Analyses: collaboration network



Let us denote the citation network with \mathbf{C}_i , and the authorship network with \mathbf{WA} . Then $\mathbf{C}_o = \mathbf{WA}^T * \mathbf{WA}$ is the *collaboration network*

[Read xyzWA.net]
 Net/Transform/2-mode to 1-mode
 /Columns
 Net/Components/Weak [2]
 Operations/Extract from Network
 /Partition [1-*]
 Net/Transform/Remove/Loops

and $\mathbf{C}_a = \mathbf{WA}^T * \mathbf{C}_i * \mathbf{WA}$ is a network of citations between authors. [3]



...Analyses: Bibliographic Coupling and Co-Citation

In **WoS2Pajek** the citation relation means $u\mathbf{Ci}v \equiv ucitesv$. Therefore the *bibliographic coupling* network **biCo** can be determined as

$$\mathbf{biCo} = \mathbf{Ci} * \mathbf{Ci}^T$$

```
[Read xyzCite.net]
Net/Transform/1-mode to 2-mode
Net/Transform/2-mode to 1-mode/Rows
Net/Components/Weak [2]
Operations/Extract from Network/Partition [1-*]
```

and the *co-citation* network **coCi** can be determined as

$$\mathbf{coCi} = \mathbf{Ci}^T * \mathbf{Ci}$$

Since the network can be quite large we first eliminate the only-cited works.

```
[Read xyzCite.net]
Net/Partitions/Degree/Output
Operations/Extract from Network/Partition [1-*]
Net/Transform/1-mode to 2-mode
Net/Transform/2-mode to 1-mode/Columns
Net/Components/Weak [2]
Operations/Extract from Network/Partition [1-*]
```

In the analysis of the obtained networks the comparability of units could/should be considered [1].

... Analyses: other derived networks

The weights $w(a, p)$ in the *author citation* network

$$\mathbf{ACi} = \mathbf{WA}^T * \mathbf{Ci}$$

counts the number of times author a cited work p .

```
[Read xyzWA.net]
Net/Transform/Transpose/2-mode
[Read xyzCite.net]
Nets/Multiply First * Second
Net/Components/Weak [2]
Operations/Extract from Network/Partition [1-*]
```

Let $b(\mathbf{A})$ denotes the binarized version of \mathbf{A} . The *author co-citation* network can be obtained as

$$\mathbf{ACo} = b(\mathbf{ACi}) * b(\mathbf{ACi})^T$$

...Analyses: temporal network

We can also transform the citation network into temporal network using the partition of works by publication year:

```
[Read xyzCite.net]
[Read xyzYear.clu]
Vector/Create Identity Vector
Vector/Transform/Multiply by [2008]
Vector/Make Partition/by Truncating
[select as partition 1: xyzYear]
[select as partition 2: obtained from vector]
Operations/Transform/Add/Time intervals determined by Partitions
```

References

- [1] Batagelj V., Mrvar A.: Density based approaches to network analysis – Analysis of Reuters terror news network. Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003, Washington, DC, USA) August 27, 2003.
<http://www.cs.cmu.edu/~dunja/LinkKDD2003/papers/Batagelj.pdf>
- [2] Garfield E.: HISTCITE. <http://www.histcite.com/>;
[HISTCITE/index](#); [Social networks](#)
- [3] Kejžar N., Korenjak-Černe, Batagelj V.: Network Analysis of Works on Clustering and Classification from Web of Science. Submitted to Proceedings of IFCS'09 (Dresden, Germany, March 2009).
http://pajek.imfm.si/lib/exe/fetch.php?media=dl:gfk1_305.pdf
- [4] Kessler, M. M.: Bibliographic Coupling between Scientific Papers. American Documentation, 14(1963)1, 10-25.
- [5] Small H.: Co-citation In Scientific Literature – New Measure Of Relationship Between 2 Documents. Journal Of The American Society For Information Science, 24(1973)4, 265-269.
- [6] **WoS2Pajek**:
<http://pajek.imfm.si/doku.php?id=wos2pajek>

- [7] Web of Science – WoS (ISI/Thomson):
<http://portal.isiknowledge.com/portal.cgi>
- [8] Python: <http://www.python.org/>
- [9] Py2Exe: <http://www.py2exe.org/>
- [10] MontyLingua package: <http://web.media.mit.edu/~hugo/montylingua/>