


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 0.3

Vladimir Batagelj

FMF, matematika

University of Ljubljana

Manual

Ljubljana, December 2007

Index

1	Searching on the Web of Science	1
2	Saving the records	2
4	Structure of a WoS record	4
5	Names of works	5
7	Program WoS2Pajek	7
15	Analyses	15
21	References	21

Searching on the Web of Science

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®

Results Topic=("social network*")
Timespan=All Years. Databases=SCI-EXPANDED, SSCI, A&HCI.

Results: **6.936** Page 1 of 694 Go Sort by: Latest Date

Refine Results

Search within results for

Subject Areas Refine

- SOCIOLOGY (966)
- PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH (779)
- PSYCHIATRY (559)
- ANTHROPOLOGY (383)
- PSYCHOLOGY, MULTIDISCIPLINARY (349)
- more...

Document Types Refine

- ARTICLE (6,036)
- REVIEW (328)
- BOOK REVIEW (252)
- MEETING ABSTRACT (151)
- EDITORIAL MATERIAL (95)
- more...

Authors

Source Titles

1. Title: **Managerial social capital, strategic orientation, and organizational performance in an emerging economy**
Author(s): Acquah 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 1 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**

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.

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

```

We decided to treat in short names the ARTN and UNSP values as BP values. We also remove the DOI parts.

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 \times discipline, works \times countries, ...

Program **WoS2Pajek** can be run as an executable program by double-clicking on its icon – see slide 9.

The source code can be executed in different ways using the Python interpreter. See slides 10, 11 and 12.

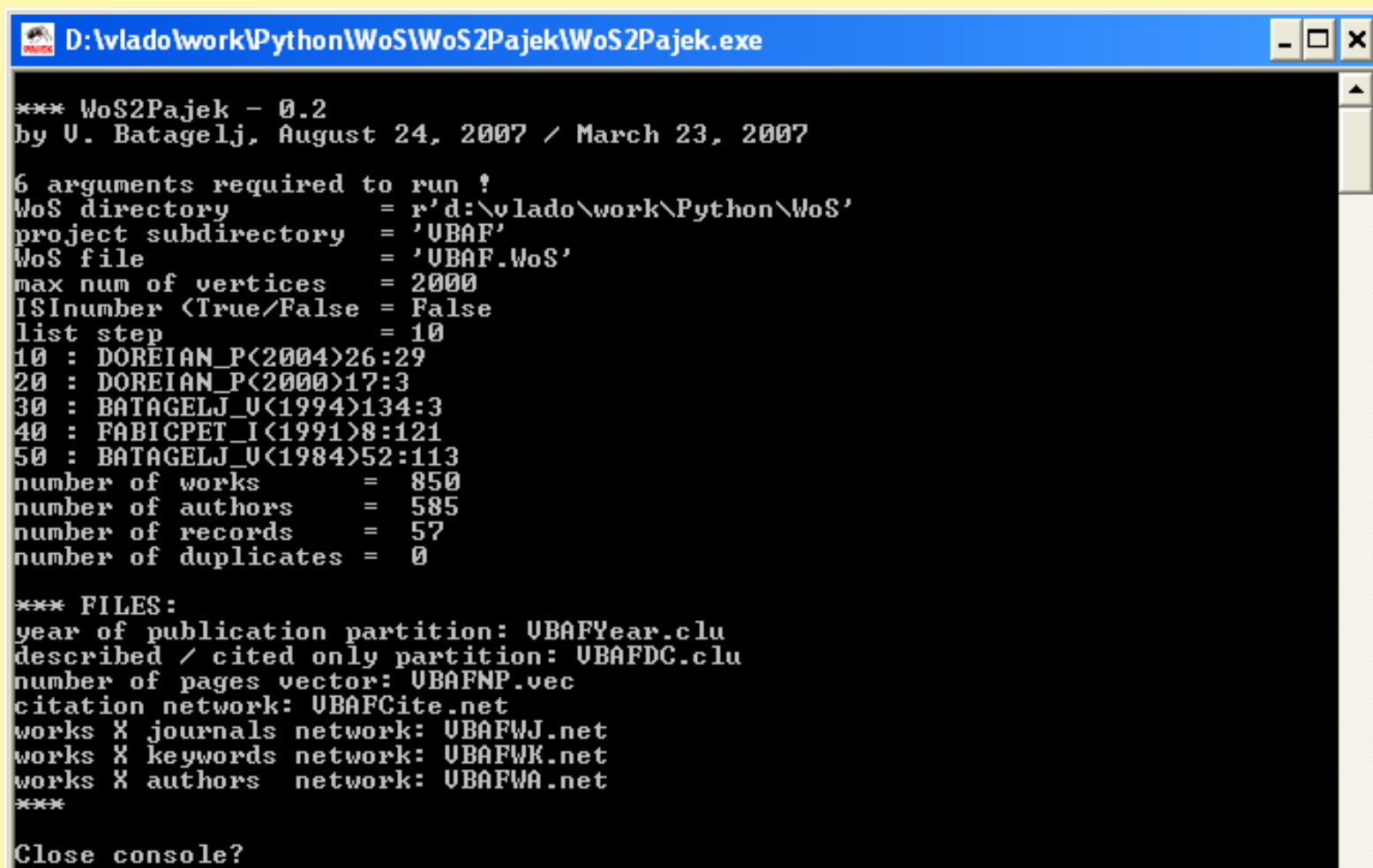
Some slides still to be updated to version 0.3 !!!

Program **WoS2Pajek**

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

- **WoS directory**: path to the directory in which the project subdirectory with the **WoS** file is located;
- **MontyLingua directory**: path to the directory in which the **MontyLingua** package is installed (put it also in the `PATH` env-variable);
- **project subdirectory name**; it is used also as a part of the name of output files;
- **WoS file**;
- **maxnum** – estimate of the number of all vertices (number of records + number of cited Works) – $30 \times$ number of records;
- **use ISI name / short name**;
- **step** – prints info about each $k \times \text{step}$ record as a trace; $\text{step} = 0$ – no trace.

Running WoS2Pajek / EXE version



```
D:\vlado\work\Python\WoS\WoS2Pajek\WoS2Pajek.exe

*** WoS2Pajek - 0.2
by U. Batagelj, August 24, 2007 / March 23, 2007

6 arguments required to run !
WoS directory      = r'd:\vlado\work\Python\WoS'
project subdirectory = 'UBAF'
WoS file           = 'UBAF.WoS'
max num of vertices = 2000
ISInumber (True/False) = False
list step          = 10
10 : DOREIAN_P<2004>26:29
20 : DOREIAN_P<2000>17:3
30 : BATAGELJ_U<1994>134:3
40 : FABICPET_I<1991>8:121
50 : BATAGELJ_U<1984>52:113
number of works      = 850
number of authors    = 585
number of records    = 57
number of duplicates = 0

*** FILES:
year of publication partition: UBAFYear.clu
described / cited only partition: UBAFDC.clu
number of pages vector: UBAFNP.vec
citation network: UBAFCite.net
works X journals network: UBAFWJ.net
works X keywords network: UBAFWK.net
works X authors network: UBAFWA.net
***

Close console?
```

Running WoS2Pajek 0.3 / 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.
```

To run, type:

```
Wos2Pajek.run(wdir,MLdir,project,bibfile,maxvert,ISIn,listep)
```

for example:

```
Wos2Pajek.run(r'D:\Vlado\work\Python\WoS',
              r'c:\Python25\Lib\site-packages\MontyLingua-2.1\Python',
              'SN', 'SN.WoS', 200000, False, 500)
```

```
>>> WoS2Pajek.run(wdir,MLdir,'joc','JoC.WoS',250000,False,500)
```

```
***** 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 - 0.3
by V. Batagelj, December 27, 2007 / March 23, 2007
```

```
started: Thu Dec 27 15:15:24 2007
```

```
WoS query: SO=(Journal of classification)
*** duplicate: STEINLEY D, 2007, J CLASSIF, V24, P144
*** duplicate: STEINLEY D, 2007, J CLASSIF, V24, P144
*** duplicate: STEINLEY D, 2007, J CLASSIF, V24, P144
*** duplicate: STEINLEY D, 2007, J CLASSIF, V24, P144
*** duplicate: STEINLEY D, 2007, J CLASSIF, V24, P144
Common sense violated! Correcting...
```

...Running WoS2Pajek 0.3 / from Python interpreter

```
Common sense violated! Correcting...
*** duplicate: KIERS HAL, 1996, J CLASSIF, V13, P175
*** duplicate: ROUX M, 1991, J CLASSIF, V8, P274
*** duplicate: MURTAGH F, 1991, J CLASSIF, V8, P115
Common sense violated! Correcting...
*** duplicate: WARTENBERG D, 1988, J CLASSIF, V5, P145
*** duplicate: BAUM BR, 1987, J CLASSIF, V4, P135
500 : NEUMANN_D(1986)3:281
*** duplicate: MILLIGAN GW, 1985, J CLASSIF, V2, P133
number of works      = 5865
number of authors    = 2893
number of journals    = 892
number of keywords   = 3494
number of records    = 569
number of duplicates = 11

*** FILES:
year of publication partition: jocYear.clu
described / cited only partition: jocDC.clu
number of pages vector: jocNP.vec
citation network: jocCite.net
works X journals network: jocWJ.net
works X keywords network: jocWK.net
works X authors network: jocWA.net
finished: Thu Dec 27 15:26:51 2007
time used: 0:11:26.303000
***
>>>
```

Running WoS2Pajek / Python from command line using parameters

```
D:\>python D:\vlado\work\Python\WoS\WoS2Pajek.py D:\vlado\work\Python\WoS UBAF U
BAF.WoS 2000 False 10

*** WoS2Pajek - 0.1
by U. Batagelj, April 3, 2007 / March 23, 2007

D:\vlado\work\Python\WoS
UBAF
UBAF.WoS
2000
False
10
-----
10 : DOREIAN_P<2004>26:29
20 : DOREIAN_P<2000>17:3
30 : BATAGELJ_U<1994>134:3
40 : FABICPET_I<1991>8:121
50 : BATAGELJ_U<1984>52:113
number of works      = 845
number of authors    = 585

*** FILES:
year of publication partition: UBAFYear.clu
described / cited only - partition: UBAFDC.clu
citation network: UBAFCite.net
works X authors network: UBAFWA.net
***

D:\>
```

Running WoS2Pajek / Python from command line

```
D:\>python D:\vlado\work\Python\WoS\WoS2Pajek.py

*** WoS2Pajek - 0.1
by V. Batagelj, April 3, 2007 / March 23, 2007

6 arguments required to run !
WoS directory      = r'D:\vlado\work\Python\WoS'
project subdirectory = 'UBAF'
WoS file           = 'UBAF.WoS'
max num of vertices = 2000
ISInumber (True/False = False
list step         = 10
10 : DOREIAN_P(2004)26:29
20 : DOREIAN_P(2000)17:3
30 : BATAGELJ_U(1994)134:3
40 : FABICPET_I(1991)8:121
50 : BATAGELJ_U(1984)52:113
number of works    = 845
number of authors  = 585

*** FILES:
year of publication partition: UBAFYear.clu
described / cited only - partition: UBAFDC.clu
citation network: UBAFCite.net
works X authors network: UBAFWA.net
***

D:\>
```


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

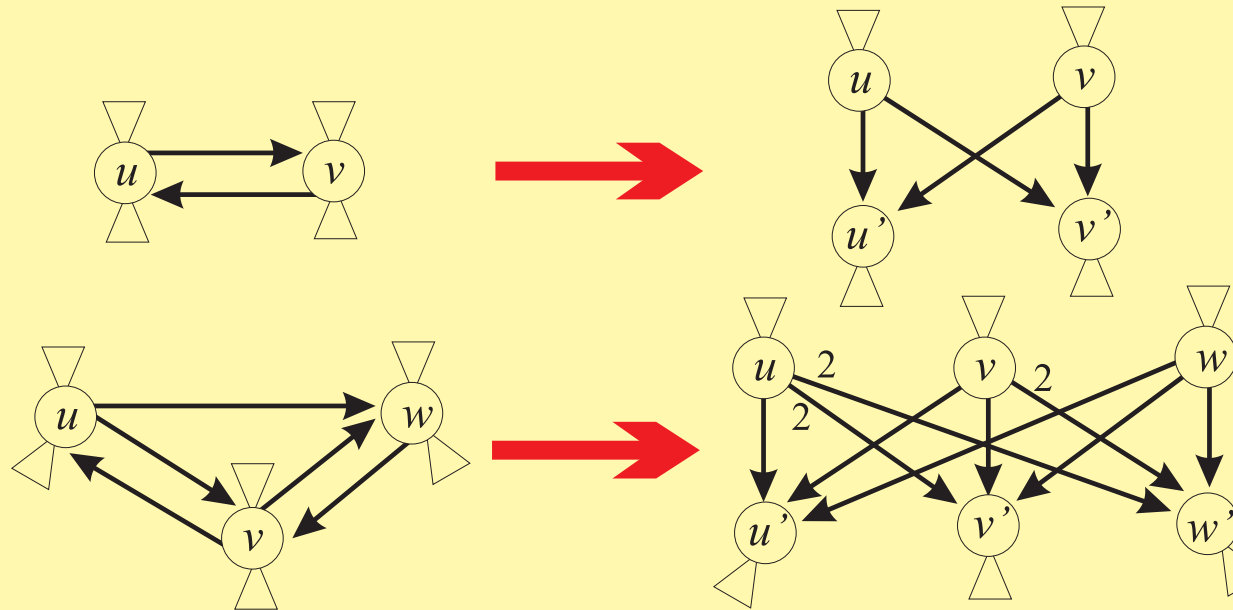
```
Info/Network/General  
Net/Transform/Remove/Loops  
Net/Transform/Remove lines/Single line
```

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')
```

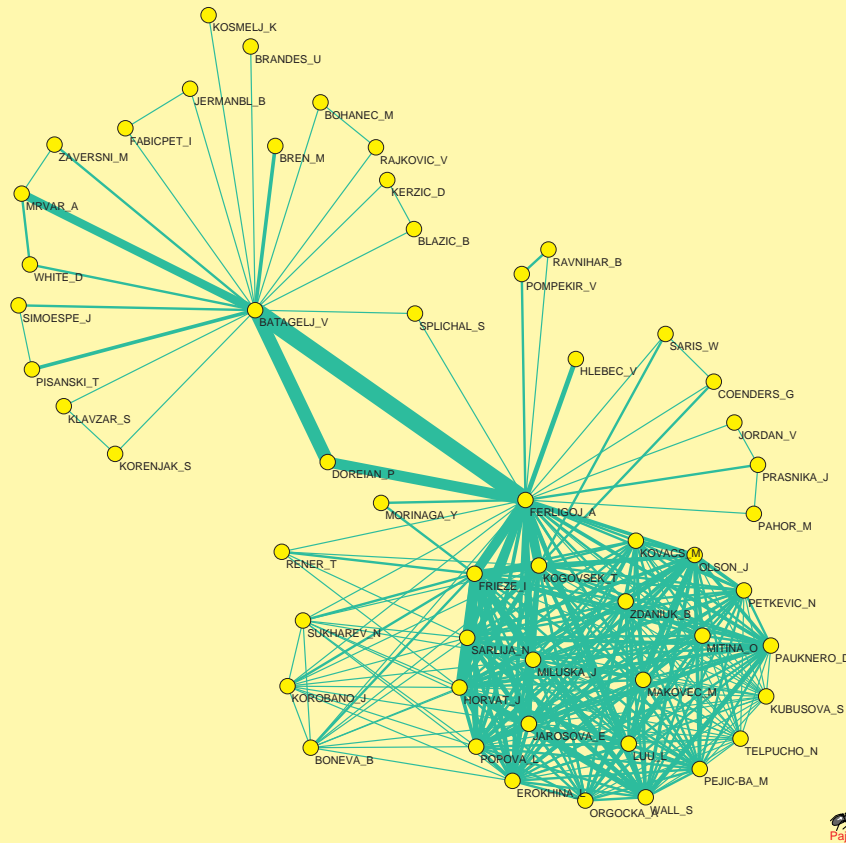
...Analyses: network boundary problem

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} ,
and the authorship network with \mathbf{A} . Then
 $\mathbf{A}^T * \mathbf{A}$ 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 $\mathbf{A}^T * \mathbf{C} * \mathbf{A}$ is a network of citations
between authors.



Pajek

...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

WoS2Pajek:

<http://vlado.fmf.uni-lj.si/pub/networks/pajek/WoS2Pajek/>

Web of Science – WoS (ISI/Thomson):

<http://portal.isiknowledge.com/portal.cgi>

Garfield E.: HISTCITE. <http://www.histcite.com/>;
[HISTCITE/index](#); [Social networks](#)

Python: <http://www.python.org/>

Py2Exe: <http://www.py2exe.org/>

MontyLingua package: <http://web.media.mit.edu/~hugo/montylingua/>