

ISI Web of KnowledgeSM Web of Science

Web of Science[®]

Select a search option:

GENERAL SEARCH CITED REF SEARCH ADVANCED SEARCH

Quick search: SEARCH Example: tuberculosis vaccine*

[Author Finder](#): Need help finding papers by an author? Use Author Finder. **NEW!**

OPEN SAVED SEARCH Open a previously saved search history.

NEW SESSION Clear all search forms, your marked list,

WoS

WoS2Pajek networks from Web of Science version 0.2

Vladimir Batagelj
FMF, matematika
University of Ljubljana

Manual
Ljubljana, August 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
6	Program WoS2Pajek	6
13	Analyses	13
17	References	17

Searching on the Web of Science

ISI Web of KnowledgeSM Web of Science GO ✓ Signed In HOME LOG OUT

TS=('social network*')
DocType=All document types; Language=All languages; Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=1970-2007

CrossSearch: [View additional results in other databases](#) Search within results: SEARCH

Refine your results
[Subject Categories](#) | [Source Titles](#) | [Document Types](#) | [Authors](#) | [Publication Years](#) [more choices](#)

6,257 results found (Set #3) Go to Page: 1 of 626 GO
Records 6,251 -- 6,257 Show 10 per page [621 | 622 | 623 | 624 | 625 | 626]

Use the checkboxes to select records for output. See the sidebar for options.

- 6,251. ARONSON DR
[SOCIAL NETWORKS - EDITORS PREFACE](#)
CANADIAN REVIEW OF SOCIOLOGY AND ANTHROPOLOGY 7 (4):
221-225 1970
Times Cited: 3
- 6,252. CHRISMAN NJ
[SITUATION AND SOCIAL NETWORK IN CITIES](#)
CANADIAN REVIEW OF SOCIOLOGY AND ANTHROPOLOGY 7 (4):
245-257 1970
Times Cited: 4
- 6,253. ARONSON DR
[SOCIAL NETWORKS - TOWARDS STRUCTURE OR PROCESS](#)
CANADIAN REVIEW OF SOCIOLOGY AND ANTHROPOLOGY 7 (4):
258-268 1970
Times Cited: 5
- 6,254. DOREIAN P
[SOCIAL NETWORKS IN URBAN SITUATIONS - MITCHELL,JC](#)
SOCIOLOGICAL REVIEW 18 (1): 137-138 1970
Times Cited: 0
- 6,255. BERNARD G
[ANALYSIS OF SOCIAL NETWORK IN URBAN SETTING](#)
CAHIERS D ETUDES AFRICAINES 10 (4): 632-639 1970
Times Cited: 0
- 6,256. TOOMEY DM
[IMPORTANCE OF SOCIAL NETWORKS IN WORKING CLASS AREAS](#)
URBAN STUDIES 7 (3): 259-270 1970
Times Cited: 2

VIEW FULL TEXT

Sort by:
Latest date SORT

Analyze Results:
ANALYZE
View rankings of the authors, journals, etc. for these records.

Citation Report:
CITATION REPORT
View detailed citation counts and the h-index value for the results.

Output Records:
 Selected records on page
 All records on page
Records to
Bibliographic Fields
PRINT E-MAIL SAVE
EXPORT TO REFERENCE SOFTWARE
SAVE TO MY EndNote Web
[Go to my EndNote Web library]
Or add them to the Marked List for later output and more options.
ADD TO MARKED LIST
[0 articles marked]

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 6257 hits (April 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

Sort by:

Analyze Results:

 View rankings of the authors, journals, etc. for these records.

Citation Report:

 View detailed citation counts and the h-index value for the results.

Output Records:

Selected records on page
 All records on page
 Records to

[\[Go to my EndNote Web library\]](#)

Or add them to the Marked List for later output and more options.

[\[500 articles marked\]](#)

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

Select **Full Record + Cited Refs**.

Click on **ADD TO MARKED LIST** and afterward also on **[*** articles marked]**.

In the new window (see next slide) check (all) the interesting fields and click on **SAVE TO FILE** and ... it takes some time ... wait until done.

When the file-chooser appears determine the file on which the records are saved.

In the web browser click on **Back**. You return to the window **Web of Science Marked Records**. Click on **DELETE THIS LIST** and afterward on **RETURN**.

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

...Saving the records

ISI Web of KnowledgeSM
Web of Science
Signed In

Web of Science Marked Records

Output Options

Display marked list:

Web of Science Marked Records - 500 Articles

Scroll down to view records

Step 1. Select the fields to include in the output.

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

**Selecting these items will increase the processing time.*

Step 2. Select an option.

<input type="text" value="Field Tagged"/> <input type="button" value="FORMAT FOR PRINT"/> <input type="text" value="Field Tagged"/> <input type="button" value="SAVE TO FILE"/> <input type="button" value="EXPORT TO REFERENCE SOFTWARE"/> <input type="button" value="SAVE TO MY EndNote@Web"/>	E-mail records to: <input type="text"/> Return e-mail (optional): <input type="text"/> Notes(optional): <input type="text"/> <input type="text" value="Plain Text"/> <input type="button" value="E-MAIL"/>
--	---

Automatically delete selected records from the Marked List after output is complete.

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 \$.

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, EP – BP + 1.

Program **WoS2Pajek**

The keywords are obtained from the fields **TI** (title), **ID** and **DE**. From the title the **stopwords** are removed and a list of words is produced.

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.

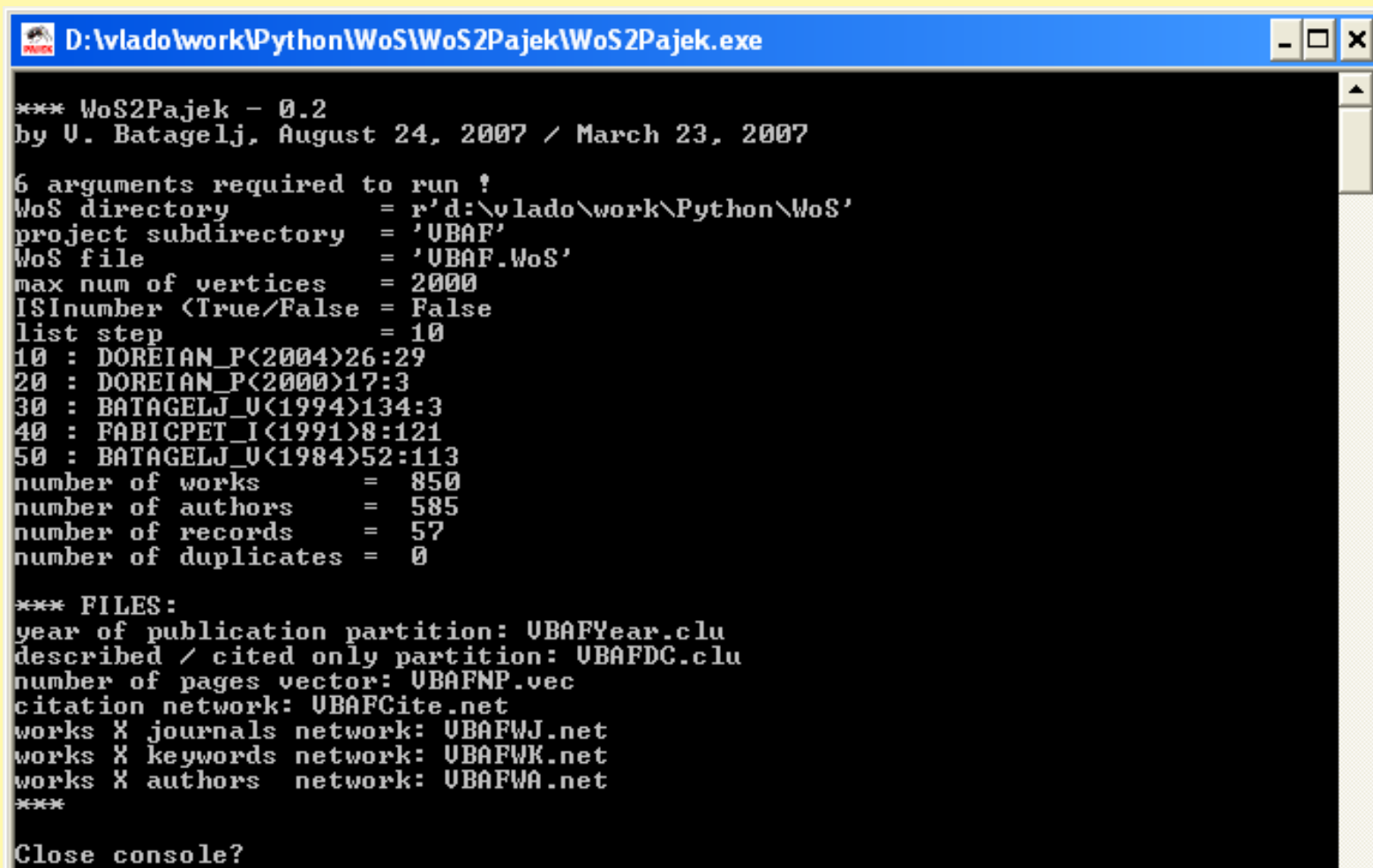
The source code can be executed in different ways using the Python interpreter. See slides 10, 11 and 12.

Program **WoS2Pajek**

The current version of **WoS2Pajek** requires 6 parameters to be given by the user:

- **WoS directory**: path to the directory in which the project subdirectory with the **WoS** file is located;
- **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* number of records;
- **use ISI name / short name**;
- **step** – prints info about each $k \cdot \text{step}$ 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: UBAYear.clu
described / cited only partition: UBAYDC.clu
number of pages vector: UBAYNP.vec
citation network: UBAYCite.net
works X journals network: UBAYWJ.net
works X keywords network: UBAYWK.net
works X authors network: UBAYWA.net
***

Close console?
```

Running WoS2Pajek / from Python interpreter

```
>>> import sys
>>> wdir = r'D:\Vlado\work\Python\WoS'; sys.path.append(wdir)
>>> import WoS2Pajek
Module WoS2Pajek imported.

To run, type:
  WoS2Pajek.run(wdir,project,bibfile,maxvert,ISIn,listep)
for example:
  WoS2Pajek.run(r'D:\Vlado\work\Python\WoS','SN','SN.WoS',200000,False,500)

>>> WoS2Pajek.run(wdir,'VBAF','VBAF.WoS',2000,False,10)

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

10 : DOREIAN_P(2004)26:29
20 : DOREIAN_P(2000)17:3
30 : BATAGELJ_V(1994)134:3
40 : FABICPET_I(1991)8:121
50 : BATAGELJ_V(1984)52:113
number of works      = 850
number of authors    = 585
number of records    = 57
number of duplicates = 0

*** FILES:
year of publication partition: VBAFYear.clu
described / cited only partition: VBAFDC.clu
number of pages vector: VBAFNP.vec
citation network: VBAFCite.net
works X journals network: VBAFWJ.net
works X keywords network: VBAFWK.net
works X authors network: VBAFWA.net
***
>>>
```

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 authorship network works \times authors and works \times keywords for multiple lines and remove them, if they exist.

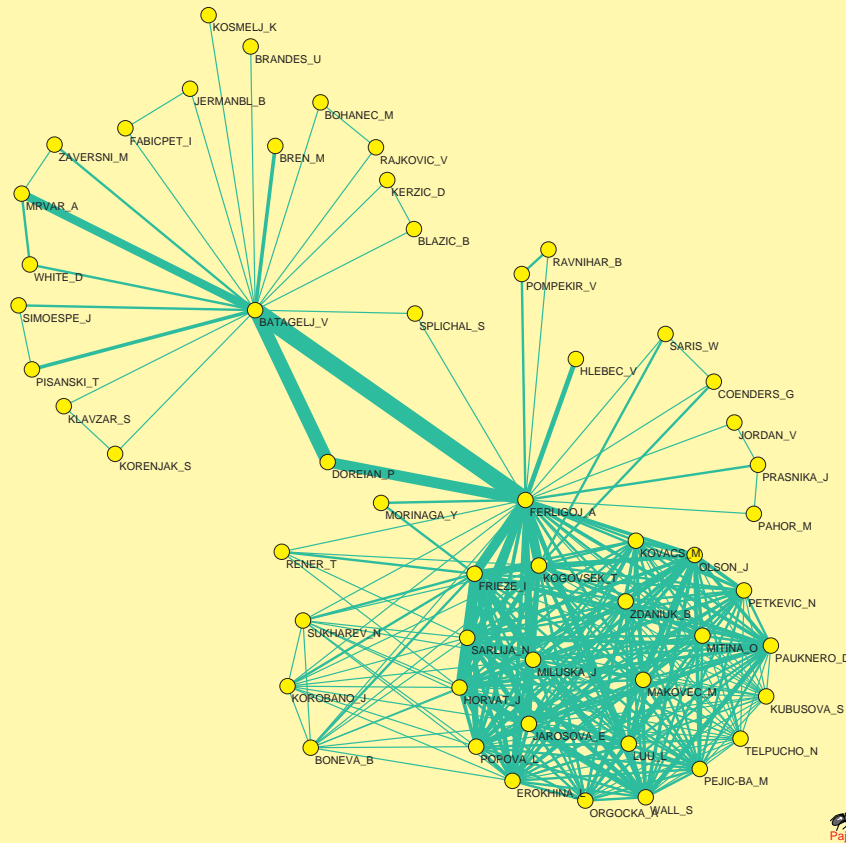
...Analyses: network boundary problem

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.



...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](http://www.histcite.com/); Social networks

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

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