Research Trail Creation

Manual for an efficient workflow using VBA and Pajek

Carla Ventzke, Jonas Höftig & Markus Hoffmann

Content

0. Introduction 1

1. Preparing the publication data 2

1.1. Getting the publication data from the Web of Science 2

1.1.1. Searching for the publications 2

1.1.2. Saving the publication data 4

1.2. Adding publications manually 5

1.2.1. Automated reference extraction from pdfs 6

1.2.2. Preparing the publication data 7

2. Processing the publication data with the Excel Macro 10

2.1. Creating the .xls file 10

2.2. Creating the .net and .vec files 12

3. Using Pajek to initially draw and alter the network pictures 15

3.1. Drawing the initial network picture with Pajek 15

3.2. Working with the network picture using Pajek 18

3.2.1. Changing the size of network picture 18

3.2.2. Changing the appearance of the connecting lines 20

3.2.3. Rearranging the clustered publications 20

3.2.4. Using Export Options to further modify the network picture 22

3.2.4.1. Add a radial 3D-Effect to Circles (optional) 23

3.2.4.2. Adjust the position of the number labels 23

3.2.4.3. Adjust font colour and size 24

3.2.4.4. Adjust the border colour of the circles 24

3.2.4.5. Adjust line width 24

3.2.5. Export Pajek picture for further processing 24

4. Further refinement of Research Trails using external programs 25

4.1. Finishing Research Trails using Visio 25

4.1.1. Align network picture and add text 25

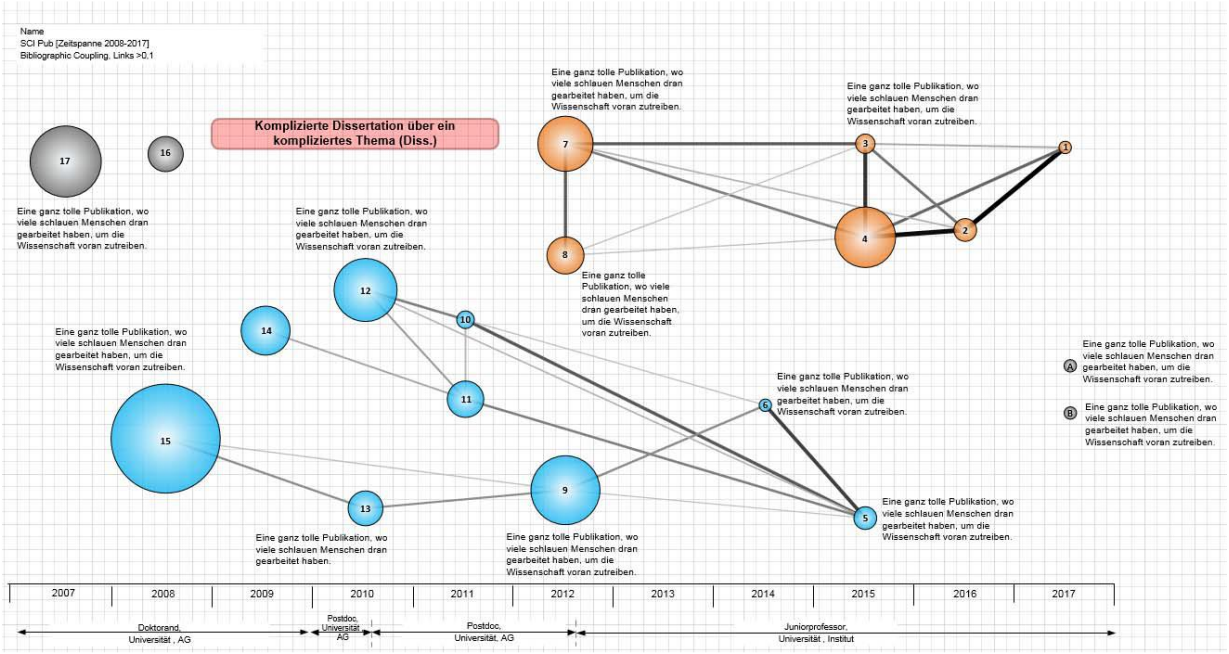
4.1.2. Add a radial gradient and move the numbers 27
4.2. Finishing the Research Trail using Inkscape and draw.io 28
4.2.1. Basic edits with Inkscape 28
4.2.2. Adding a radial gradient with Inkscape 29
4.2.3. Basic edits with draw.io 30
4.3. Creating Research Trails for internal and external use 31
5. Keeping track of publications and clusters 33
5.1. Publication list sorted by clusters 33
5.2. Publication list for internal and external use 34
6. End 34
Appendix A – Decision tree for reference matching in “Extended_Trail” (“WoS_Trail” only checks exits 1 & 2): 35

0. Introduction

In this document, we collected everything you need to know about the creation of so-called Research Trails (RT) which are visualizations researchers' cognitive careers using bibliometric tools. They can be used for example in qualitative interviews with these researchers as a guide to reconstruct their careers.[1] To create a RT with Visual Basic for Applications (VBA), you best have access to the Web of Science (WoS) and need the programs Microsoft (MS) Excel and Pajek. For further refinement either MS Visio or Inkscape and draw.io are useful[2]. Furthermore, the Excel-Macro-package "Research trail macros.xlsm" is needed.[3] The general steps which you need to follow are:

1. downloading the publication data from WoS and/or processing pdfs;
2. preparing the input files with the Excel-Macros;
3. create the network-picture with Pajek;
4. finalize the network-picture with Visio, Inkscape, and/or draw.io;
5. check if the cluster created in the process can be mapped to possible topics.

The result of this will be a visualization of the different topics a researcher works on in their career as seen below.



[1] The following is based on a collection of documents provided by Grit Laudel and Jochen Gläser, who developed this method. For further information on the theoretical background, see Gläser and Laudel (2015).

[2] Excel is always part, Visio is part of some versions of Microsoft Office – and thus not for free. Inkscape (https://inkscape.org/) and draw.io (https://app.diagrams.net/) are open-source freeware and can be downloaded.

[3] You can download this Macro-package here from Grit Laudel's Website.

1. Preparing the publication data

For this first step, you best have a complete and accurate publication list of the researcher in question. In most cases, this won't be a problem, because many researcher have a complete publication list on their or their group's website. However, sometimes this is not the case, so you may only get an unconfirmed publication list from ResearchGate or Google Scholar or even a few conflicting lists. In these cases, you need to be cautious and doublecheck the list(s) that you are using. When you start researching the researcher's publication list, you should create a folder in which you collect everything concerning this researcher. Additionally, you should create a central document in which you gather all important information such as personal data (e.g., name, birthdate, languages, nationality, family status), some short description of their research which is often provided on their websites, and their CV. All of this is going to be needed in an interview situation and it helps in the creation of the RT.

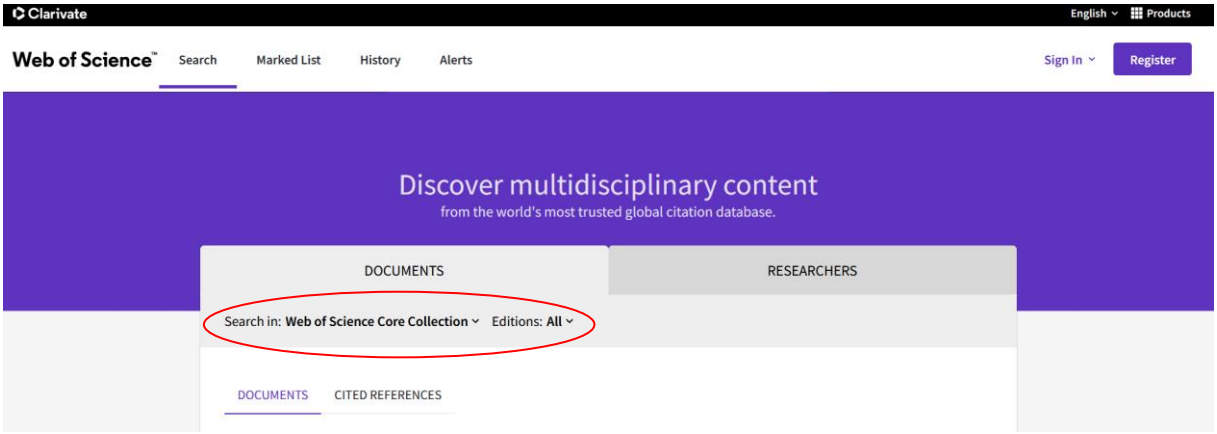
Once you have a list of publications there are two ways to proceed, which can be combined:

1.1. Getting the publication data from the Web of Science

The first way consists of downloading publication data from the WoS. This option is faster and has more accurate reference lists for the publications (which are used to create links between publications). But it is not always possible to use this way, depending on how completely the publications from the researcher in question are covered in the WoS. This will very likely depend on the field of research and type of publication. The natural and biosciences are usually covered very well, as long as journal articles are the main type of publication. How to look for the publications of a researcher in the WoS and how to download the needed fields will be shown in the next sections:

1.1.1. Searching for the publications

1. Open the Web of Science (https://www.webofscience.com/wos/woscc/basic-search). For this you need access to the Web of Science, usually through your organisation.

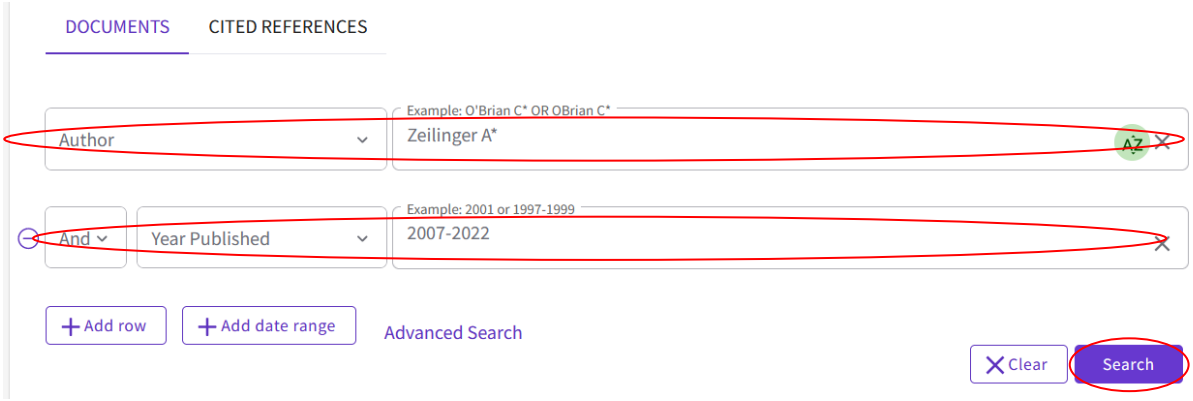


2. Select the database "Web of Science Core Collection" & "All" editions.

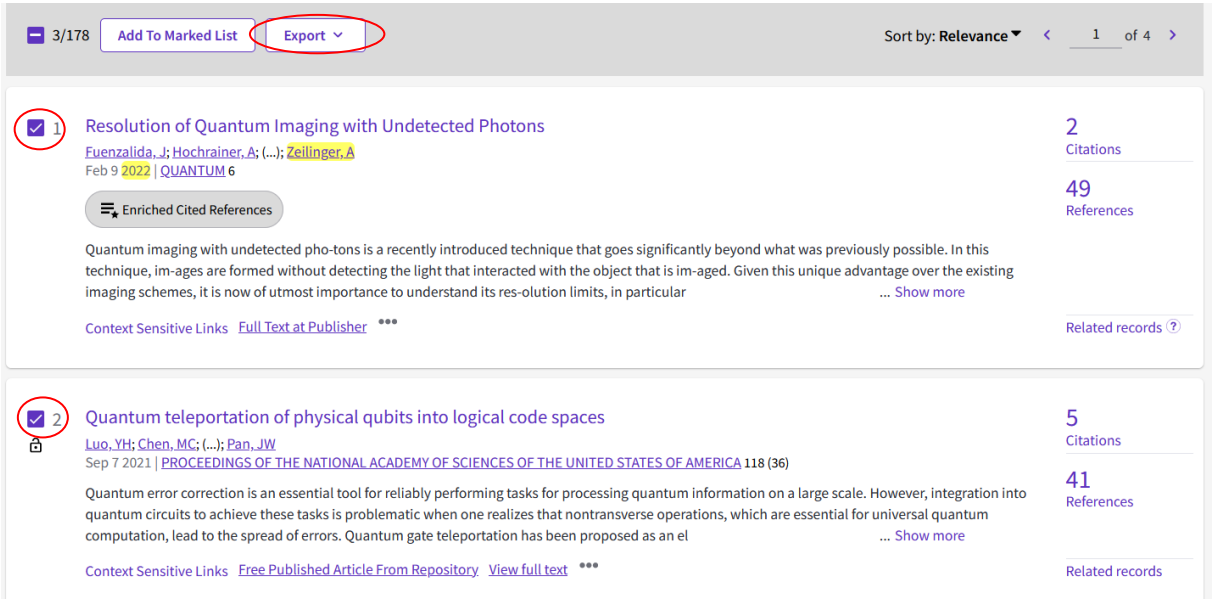
3. Write the last name and initial of the first name plus an "*" of the researcher in the first search field and select "Author" on the left-hand site of this search field.

4. Limit years. If no list of publications of the author is available from other sources, the first year should be ca. five years prior to the time at which the PhD was granted. The last year should usually be the current year.

5. Click "search".



6. In the list of results there may be some non-fitting publications (caused by error or homonyms), so you have to doublecheck it (see the box below). Click the checkbox for all the fitting publications and click on "Export" (alternatively you can also save them as a "Marked List" and export them later[4], which might be useful for longer research trails, which cannot be done as quickly).



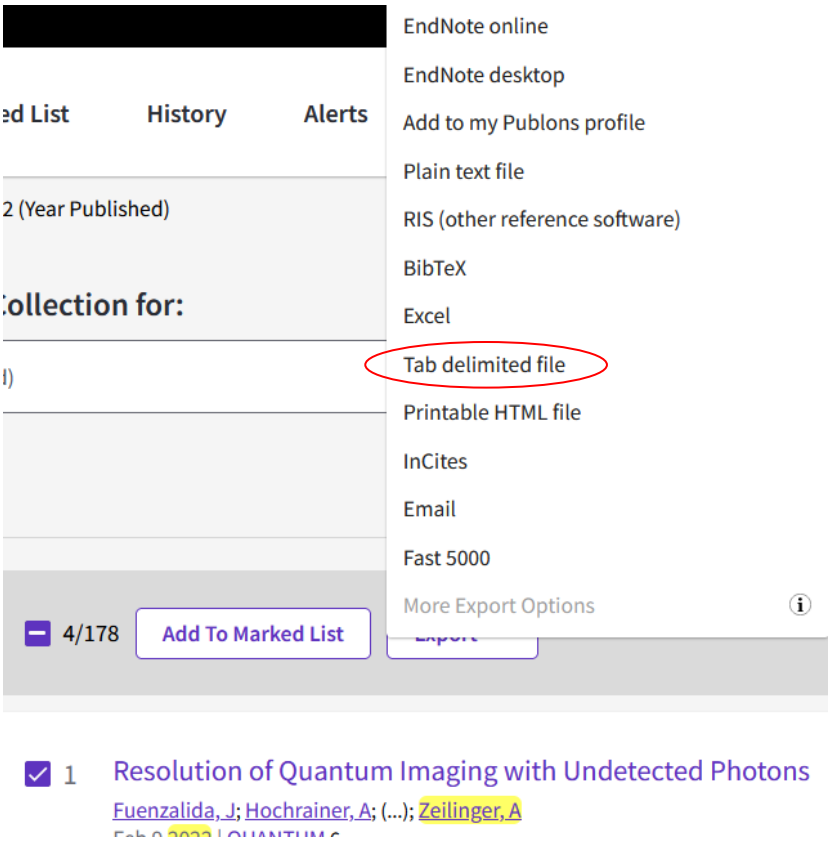
[4] If you have an account for the Web of Science, you can also save your marked lists. We recommend creating an account if possible.

How to cope with problems of **homonyms** and **non-fitting publications**:

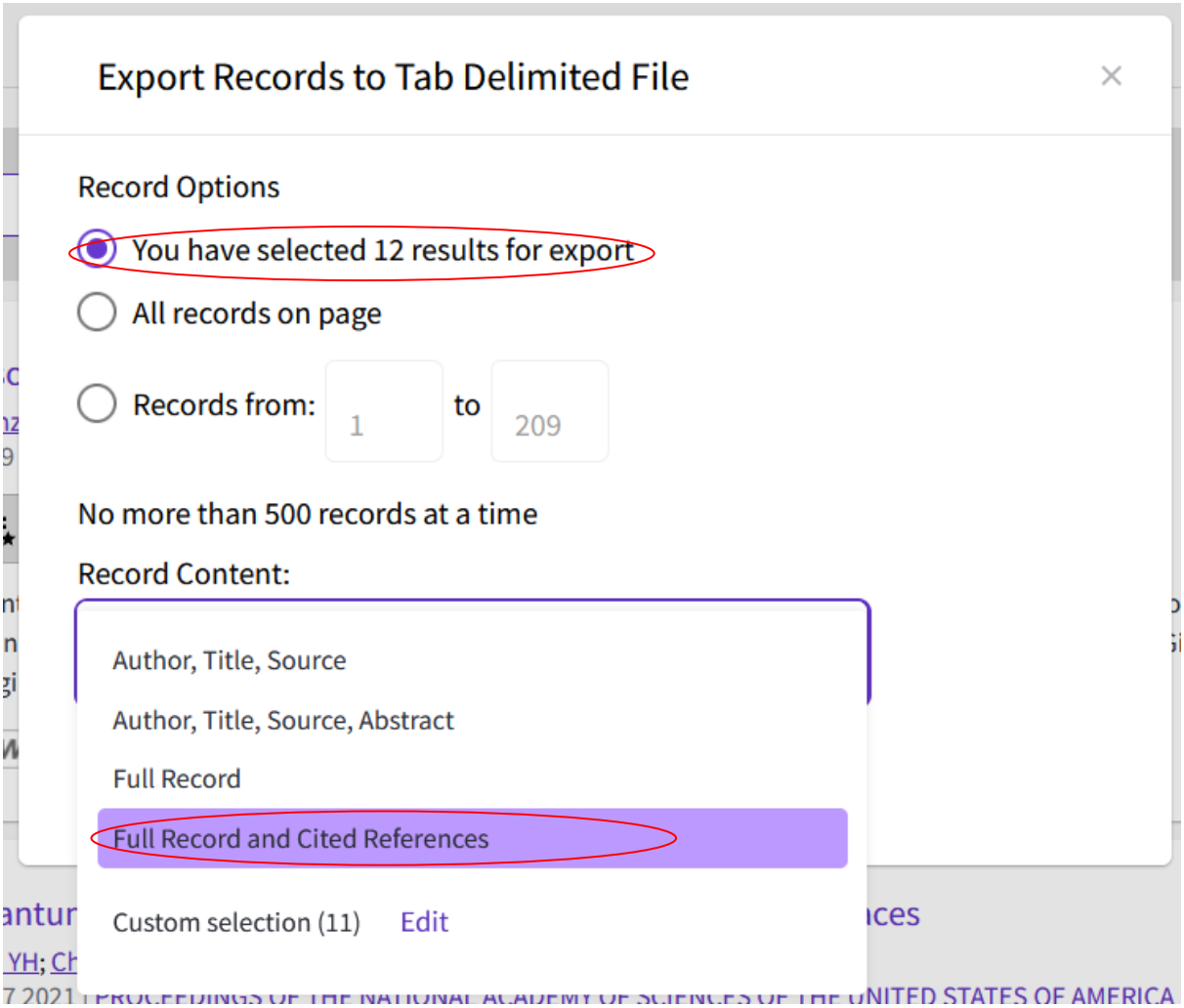
- get a full publication list of the researcher.
- if there are publications of the researcher, which could not be found through the initial Web of Science search it is often helpful to search for them individually by title.
- get other background information about the researcher (particularly about their career), their co-authors, the researcher's research field and topics etc.
- Use the "Refine results" area on the left: exclude under "Document Types" types of publications (e.g., meeting abstracts, corrections, news items; they usually do not have references and are therefore not interesting for bibliographic coupling).
- the "Refine results" area can also be used for "Authors" and "Web of Science Categories" to further narrow down the search results.
- look up if earlier papers are cited later by the same author.
- look up full paper online versions of articles to track institutional affiliations.
- some publication types like dissertations are usually not listed in the WoS.

1.1.2. Saving the publication data

1. After clicking on "Export" a field with export options will appear. Choose "Tab delimited file".



2. Check that you have selected all your results and then choose under “Record Content”: “Full Record and Cited References”. It is important that the cited references are included, so the “Full Record” is not ‘full’ enough. Press “Export”.



3. You should have downloaded a .txt file titled “savedrecs”. Rename it to your researcher’s name and copy it to your researcher’s folder on your computer.

1.2. Adding publications manually

In addition to the WoS output file, it is also possible to manually add publications that are not indexed in the Web of Science or to create a research trail without the WoS at all. Some cases where this might be useful are:

- **Preprints** and **working papers** are usually not indexed in the WoS but usually represent the more recent content of a researcher in written form. They may be especially important in interview situations where the current research will be talked about. In some cases, you will find older working papers or preprints, which have been

published somewhere else in the meantime. If that is the case, then you should exclude them and only go for the eventual publication.

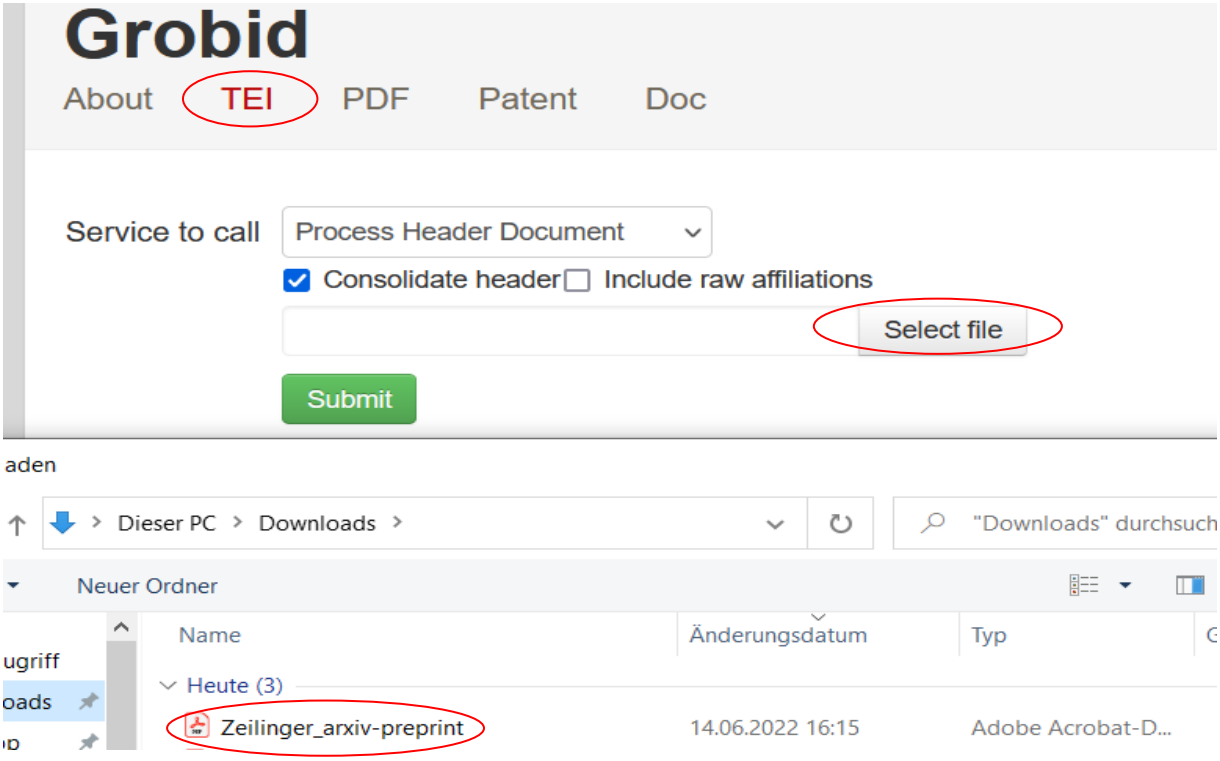
- **Non-article publications** are very weakly covered by the Web of Science. Depending on the research field, they may just be incidental side publications or comprise the main channel of formal communication. A research trail in the humanities without book chapters would be useless. Examples would be book chapters, books and dissertations.
- Depending on the researcher, **non-indexed journal articles** might also be an important part of their publications. The WoS and its policy for journal inclusion is not very transparent. And depending on research field, and more importantly, the language of publications, many to even all publications may not be found in the WoS.

Regardless of the reason for inclusion, if you want to add non-indexed publications in a research trail, the following steps should be followed:

1.2.1. Automated reference extraction from pdfs

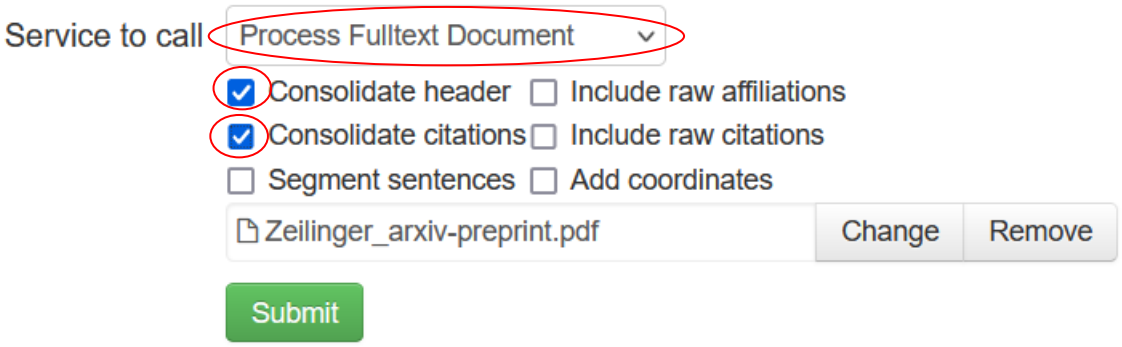
1. Get the wanted publications in an OCR-readable .pdf format. This step can be more or less problematic, depending on the age of the publication and issues of access.

2. Go to https://cloud.science-miner.com/grobid/[5] and upload your pdf under the "TEI" rubric. Here the references from the publication will be automatically extracted.



[5] Shoutout to Patrice Lopez, who manages the online application!

3. Change in the dropdown menu from “Process Header Document” → “Process Fulltext Document” and check the two boxes “Consolidate header” and “Consolidate citations”.



Tips for (and common problems with) the automated reference extraction via **Grobid**:

- It is necessary to have a bibliography or some kind of reference list in the document. References that are only in footnotes do not work!
- Some citation styles are problematic and result in dropped or flawed references (for instance when the author’s name gets replaced by “------” in the case of multiple publications by the same author).
- It works best with English references. German ones work reasonably well, all other languages should be expected to produce worse results.
- For whole books and chapters in edited volumes it is necessary to trim the pdf. The Grobid web application will return an error for pdfs that are too long and for book chapters you need to limit the pdf to the chapter in question.
- Sometimes other text will be mistakenly interpreted as references, which needs to be cut out. Examples are lists of ‘other publication in this series’ or Author information at the end.
- The best solution for solving the preceding two points is to limit the pdfs to 1) a title page (for author, title and year information – if available) and 2) the reference list.

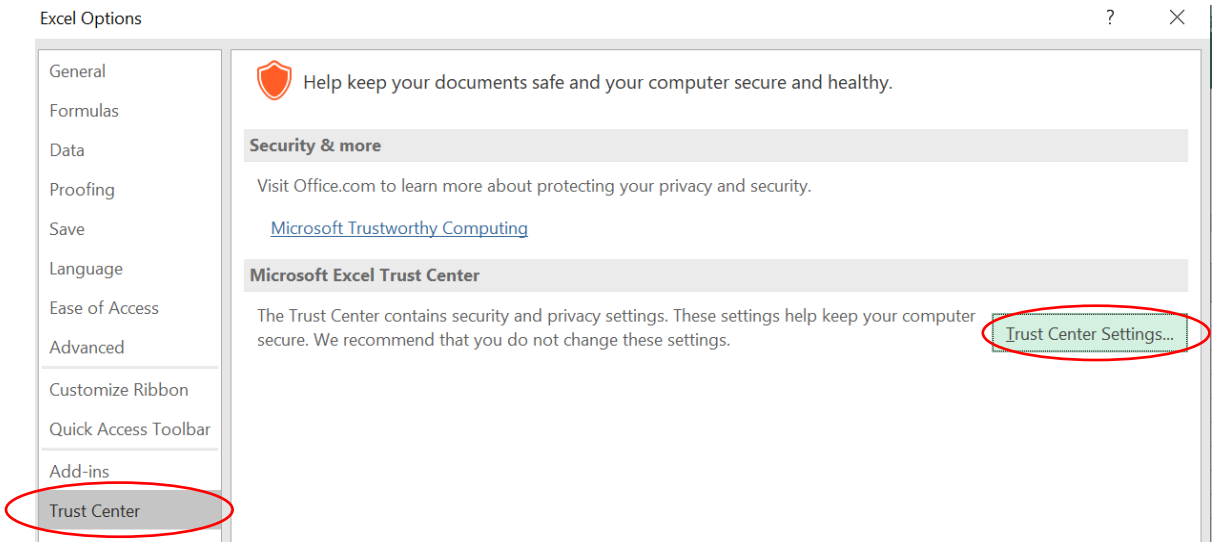
4. Press Submit.

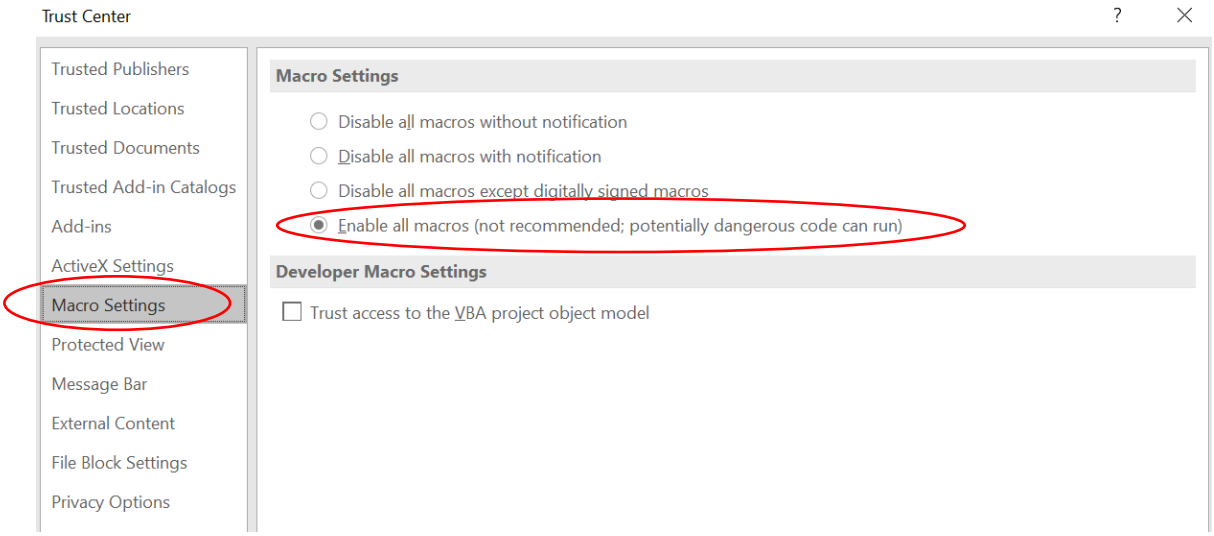
5. Depending on the size of the file the program will work a while and then present you with an xml file. Download the TEI result.

1.2.2. Preparing the publication data

The TEI result must be put into the same form as the WoS output file using the macro collection Grit Laudel provides on her website. You will either create a new txt file with the data or add to an already existing download from the WoS.

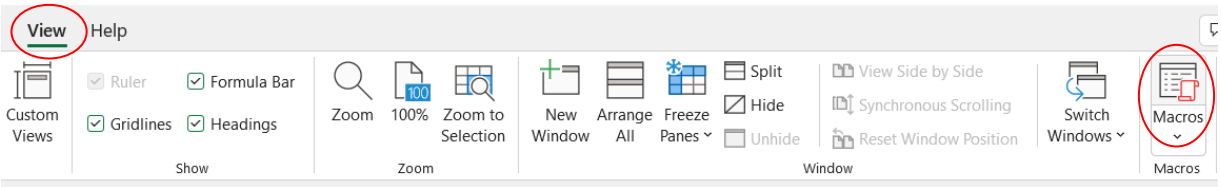
1. First you need to make sure that the macros can run: Open Excel and set Macro security to “medium” by going to “File –> Options –> Trust Centre –> Trust Centre Settings –> Macro Settings –> Enable all macros").

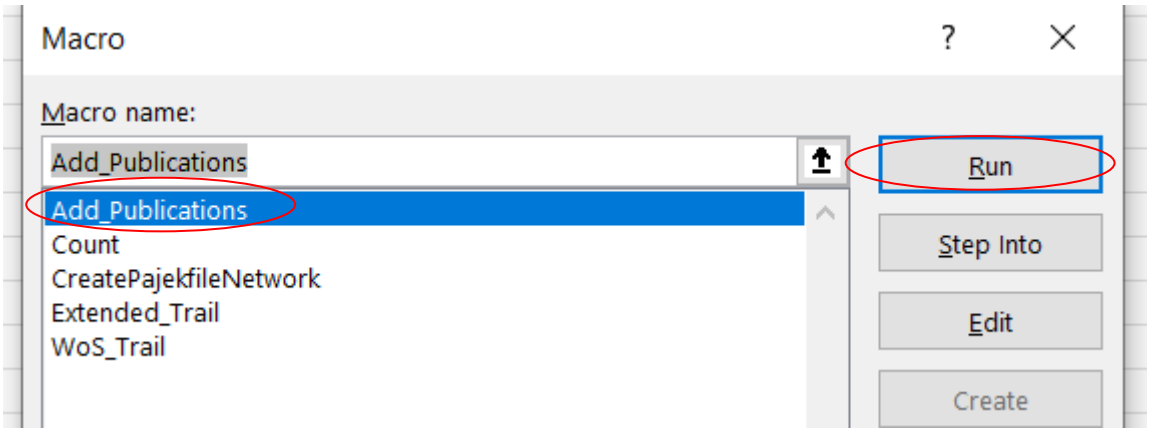




2. Open “Research trail macros.xlsm”. For easier handling we recommend that you put all files you are using in the same folder (the one you have created before for you researcher).

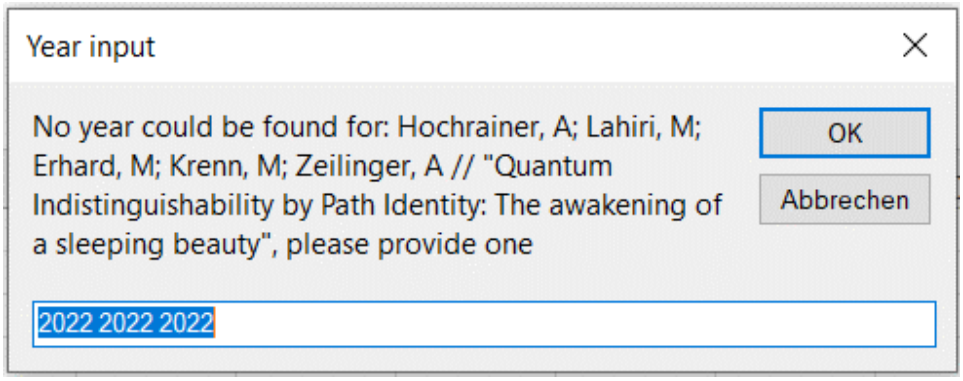
3. Start the macro “Add_Publications” by going to View -> Macros, selecting it from the list and clicking “Run”.





The macro will

- ask you to choose any number of xml-files, where you should choose the downloaded xml(s) from Grobid,
- open those files; and look for author name, title and year of the respective publication in question and read the extracted references,
    - If author/title/year cannot be found it will ask you to provide one. We recommend that you keep the researcher's publication list at hand to provide this information:



- ask you if you want to add the publications to an existing Web of Science output or not. If you choose "yes", you will be asked to choose a txt-file. If you choose "no", the macro will create a new txt file with the last name of the first author of the first publication as file name.

Now you have the data basis to create research trails.

2. Processing the publication data with the Excel Macro

The prepared data will now be further processed using the same macro package. It will first create an .xls file with the processed data, which is then used to create input files for the network picture with Pajek. If you have not used the macro package already to add publications, you may still need to set the macro security to a low enough level.

2.1. Creating the .xls file

First you have to choose between two macros that can process your data: "Extended_Trail" and "WoS_Trail". **If you have added additional publications to your data, you need to use Extended_Trail to produce a sensible research trail.** If you use only publications from the Web of Science and especially if most of the references in these publications are also pointing towards WoS-indexed publications, you can also use WoS_Trail.

	"Extended_Trail"	**"WoS_Trail"**
Speed	Slow (will take a lot of minutes if too many publications are included)	Fast
Reference matching (Appendix)	Complex	Simple
Manual checks	Yes	No
Inclusion of non-indexed pubs	Possible	Not possible

The workflow for both versions is described in what follows. Some steps only apply for the "Extended_Trail" macro and are indicated by a red border:

1. Open "Research trail macros.xlsm".

2. Start either the macro "Extended_Trail" or "WoS_Trail" by going to View->Macros, selecting it from the list, and clicking "Run".

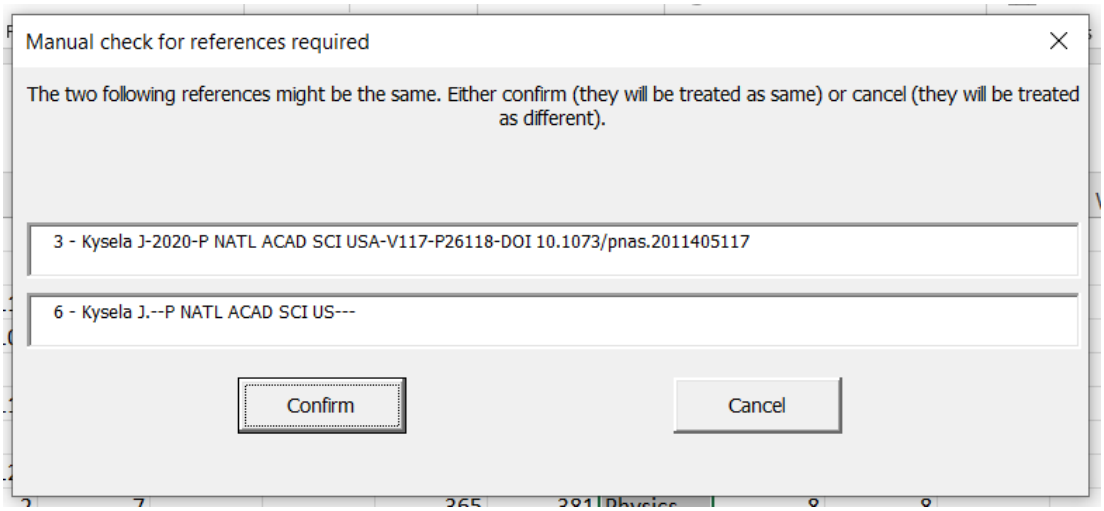
The macro will

- ask you to open a file, for which you must select the researcher's .txt file

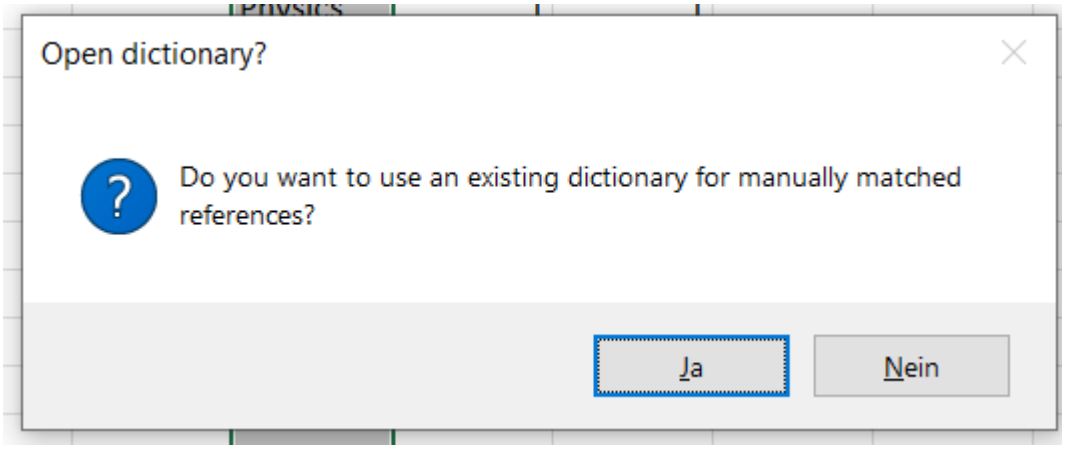
- read and analyse this file,
- create and save a file “[name of researcher].xls” with the results of the analysis.

In the case of “Extended_Trail” it will additionally

- ask you to make decisions about similar references. You will have to decide if the two references in question will be treated as different or the same. The program will remember your choices and not ask for the same two references again.

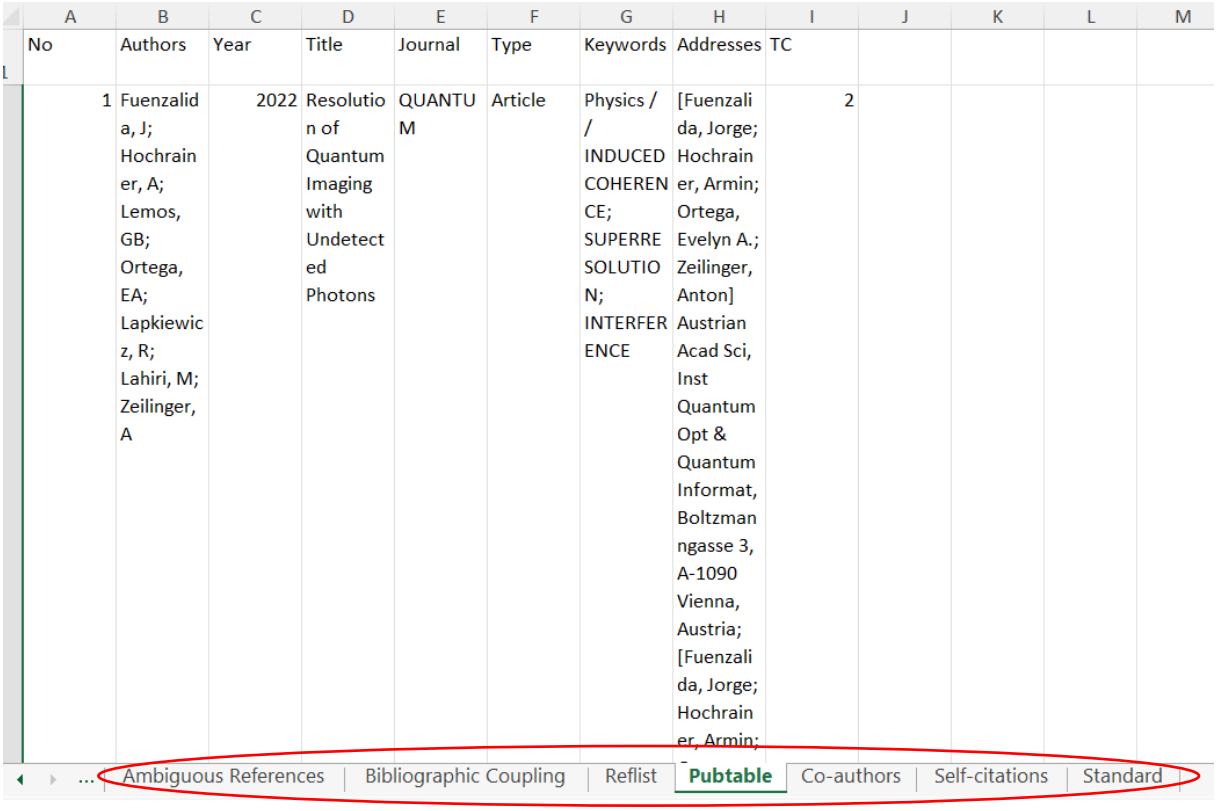


- ask you at the end twice if you want to save your decisions about similar references in a “dictionary”. If you press “yes”, a .csv file will be created with your decisions about matching and non-matching references respectively. If you press “no” you need to make the same manual checks again if you create another research trail with the same data.
- ask you at the beginning if you want to import a dictionary for manually matched references. Then it will ask you the same for a dictionary for non-matching references. This is useful if you run research trails with the same data but also if you do research trails of researchers in the same field (because they will have some overlap in their references). In that case, choose a fitting .csv file you already created.



- determine what the most referenced sources for the publications from the research trail are. This list will be shown in the spreadsheet “Reflist”.

The **result of this analysis is an .xls file**, which contains different spreadsheets like in the example below. One of them is a publication list which you can use later and may copy into your researcher’s information document.

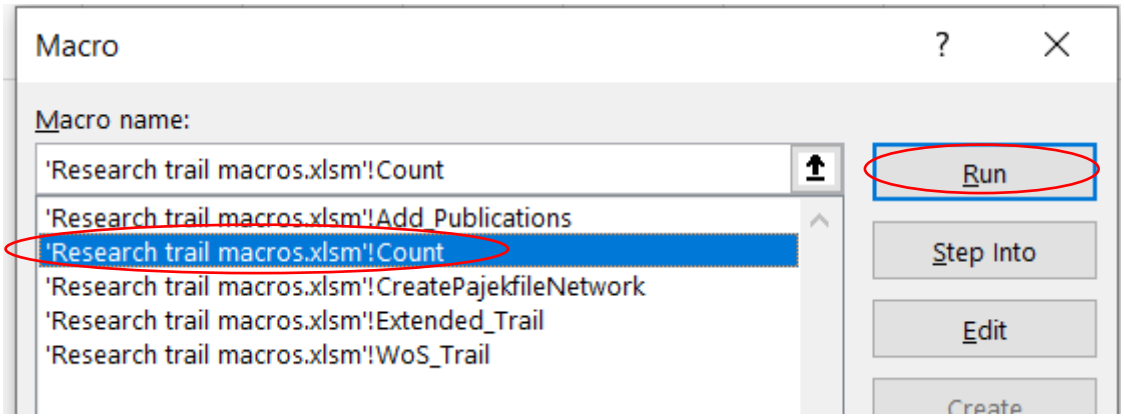


No	Authors	Year	Title	Journal	Type	Keywords	Addresses	TC
1	Fuenzalida, J; Hochrainer, A; Lemos, GB; Ortega, EA; Lapkiewicz, R; Lahiri, M; Zeilinger, A	2022	Resolution of Quantum Imaging with Undetected Photons	QUANTUM	Article	Physics / / INDUCED COHERENCE; SUPERRESOLUTION; INTERFERENCE	[Fuenzalida, Jorge; Hochrainer, Armin; Ortega, Evelyn A.; Zeilinger, Anton] Austrian Acad Sci, Inst Quantum Opt & Quantum Informat, Boltzmangasse 3, A-1090 Vienna, Austria; [Fuenzalida, Jorge; Hochrainer, Armin;	2

... | Ambiguous References | Bibliographic Coupling | Reflist | **Pubtable** | Co-authors | Self-citations | Standard

2.2. Creating the .net and .vec files

With this newly created .xls file open and visible, go to View->Macros again. **For the macros to show up now it is necessary to have the “Research trail macros.xlsm” file open as well!** Before you create the Pajek files we strongly recommend that you run the macro “Count” first:



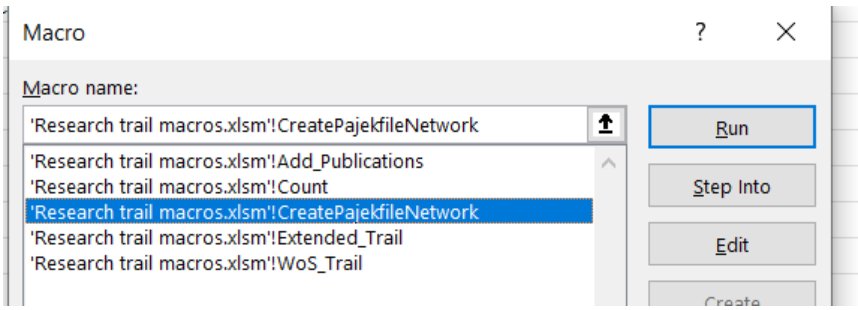
This macro will:

- count how many **clusters** (publications connected by a strong enough bibliographic coupling) and **orphans** (publications without any strong enough coupling) each value from Salton's cosine will produce (starting by 0,01 up to 0,5).
- give you a list of the 10 most cited references in the research trail (if you used "Extended_Trail").

			Threshold	Orphans	# of Cluste	Cluster					
Most referenced papers			0,01	0	1	1, 3, 5, 6, 12, 8, 2, 7, 9, 10, 4, 11,					
PubInfo	ID of Citin	Amount of Citing pap	0,02	0	1	1, 3, 5, 12, 6, 8, 2, 7, 9, 10, 4, 11,					
			0,03	0	2	1, 3, 12,	9, 10,				
krenn-201	5, 6, 7, 8,	5	0,04	2	1	1, 3, 12,	9,	10,			
krenn-201	5, 6, 7, 8,	5	0,05	2	3	1, 3, 12,	2, 7, 8, 5	4, 11,	9,	10,	
krenn-201	5, 7, 8, 9,	5	0,06	2	3	1, 3, 12,	2, 7, 8, 5	4, 11,	9,	10,	
wang-201	2, 5, 6, 7,	5	0,07	2	3	1, 3, 12,	2, 7, 8, 5	4, 11,	9,	10,	
luo-2019-I	2, 5, 6, 7,	5	0,08	2	3	1, 3, 12,	2, 7, 8, 5	4, 11,	9,	10,	
erhard-20	5, 6, 7, 8,	4	0,09	3	3	1, 3, 12,	2,	4, 11,	5, 6, 7, 8	9,	10,
malik-201	5, 6, 7, 8,	4	0,1	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
wang-201	5, 6, 7, 8,	4	0,11	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
zou-1991-I	1, 3, 6, 12,	4	0,12	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
wang-199	1, 3, 5, 12,	4	0,13	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
			0,14	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
			0,15	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
			0,16	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
			0,17	3	3	1, 3, 12,	2,	4, 11,	5, 7, 8, 6	9,	10,
			0,18	5	2	1, 3, 12,	2,	4,	5, 7, 8, 6	9,	10,

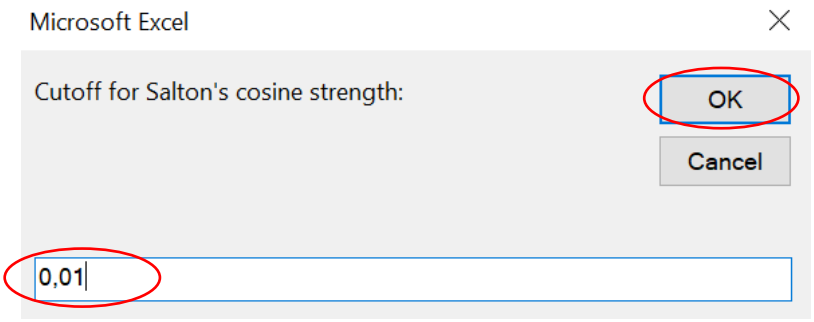
This macro should help you to go for a good value for Salton's cosine (for the next step). As a general rule of thumb, you should find a good balance of: 1) few orphans, 2) many clusters and 3) low threshold for Salton's cosine.

Based on what the count macro tells you, the next step is to create the input files for Pajek (the program that will visualize the research trail). Start the macro "CreatePajekfileNetwork". It will ask you for the "Cutoff for Salton's cosine strength:". Type in your desired value und confirm. You may want to do this with 2-3 values, depending on what the count macro tells you.[6] Please note that depending on your version of office and the primary language of your operating system you either need a "." or a "," for the decimal position.



[6] This macro produces two input files for Pajek. The .net file contains information about
- the number of vertices (= the number of publications),
- the label assigned to each vertex (= the number of each publication)
- the initial coordinates of each vertex in two-dimensional space (the x-coordinates represent the years),
- the edges between vertices (= the strength of the bibliographic coupling).

The .vec file contains information about the size of the vertices (normalized number of citations).
Both files are stored in the same directory as the .xls file for the researcher you are investigating and will be named like this: "[researcher name][cosine].net/vec". For further information on the Salton's cosine, see Gläser and Laudel (2015:305).

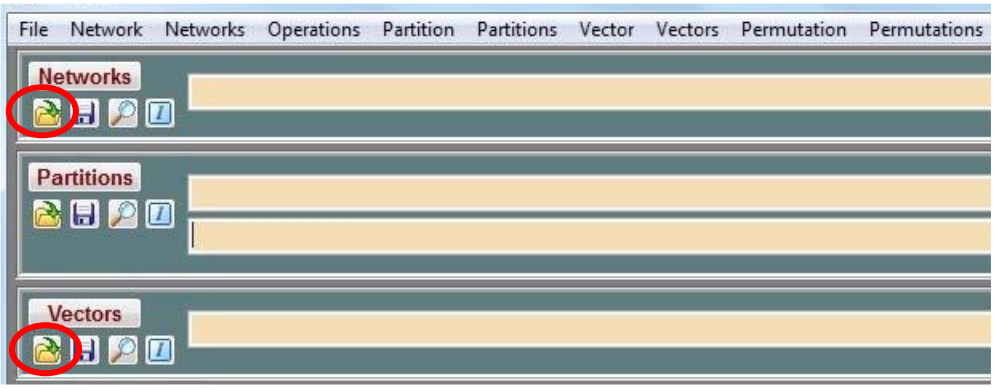

Microsoft Excel
Cutoff for Salton's cosine strength:
OK
Cancel
0,01

3. Using Pajek to initially draw and alter the network pictures

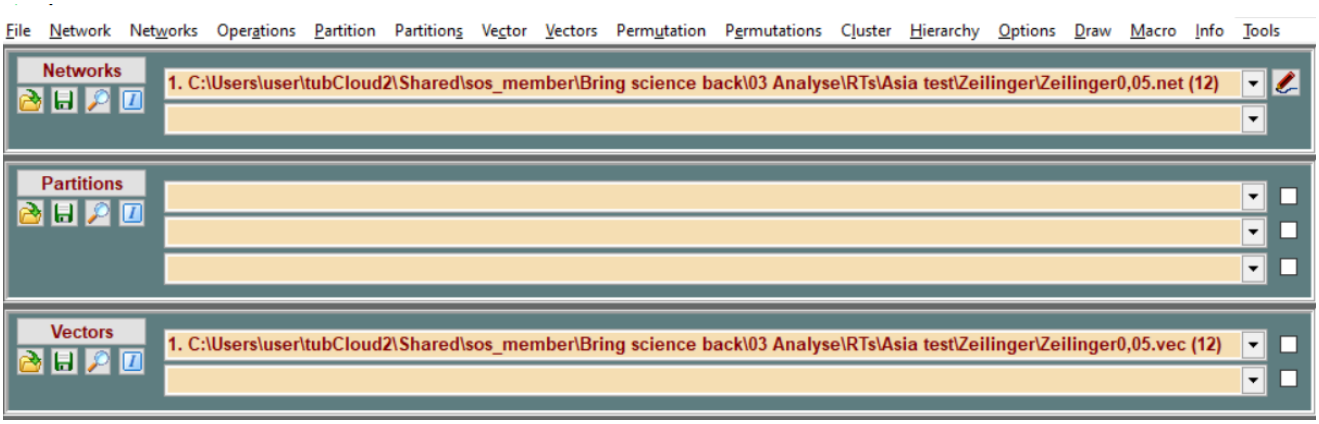
Pajek is a free network analysis and visualization package that *runs under Windows*.[7] Pajek must be downloaded and can be installed on your computer.

3.1. Drawing the initial network picture with Pajek

1. Start Pajek.
2. On the screen, click on the open folder under "Networks".
3. Navigate to the folder in which your input files are stored.
4. Open the file [researcher name][cosine].net.
5. Click on the open folder under "Vectors".
6. Open the file [researcher name][cosine].vec.

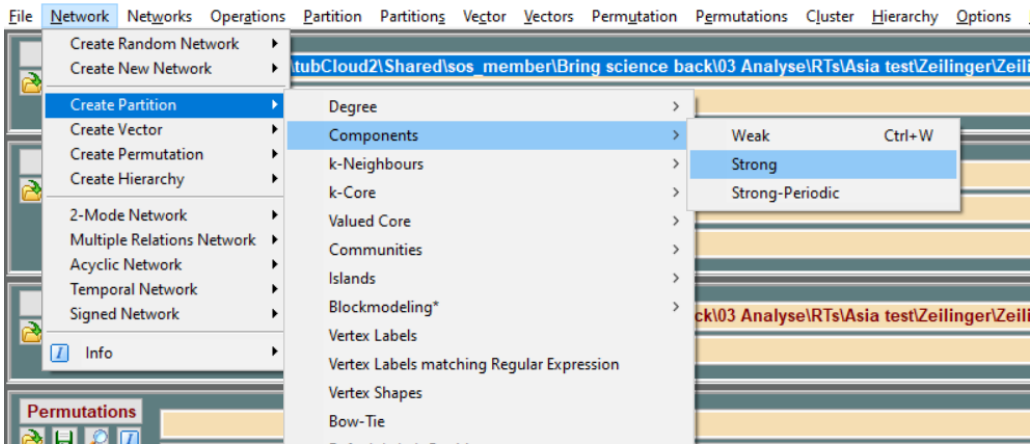


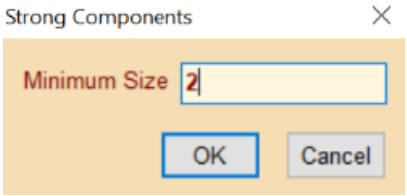
The "Networks" and "Vectors" fields should now show the input files.

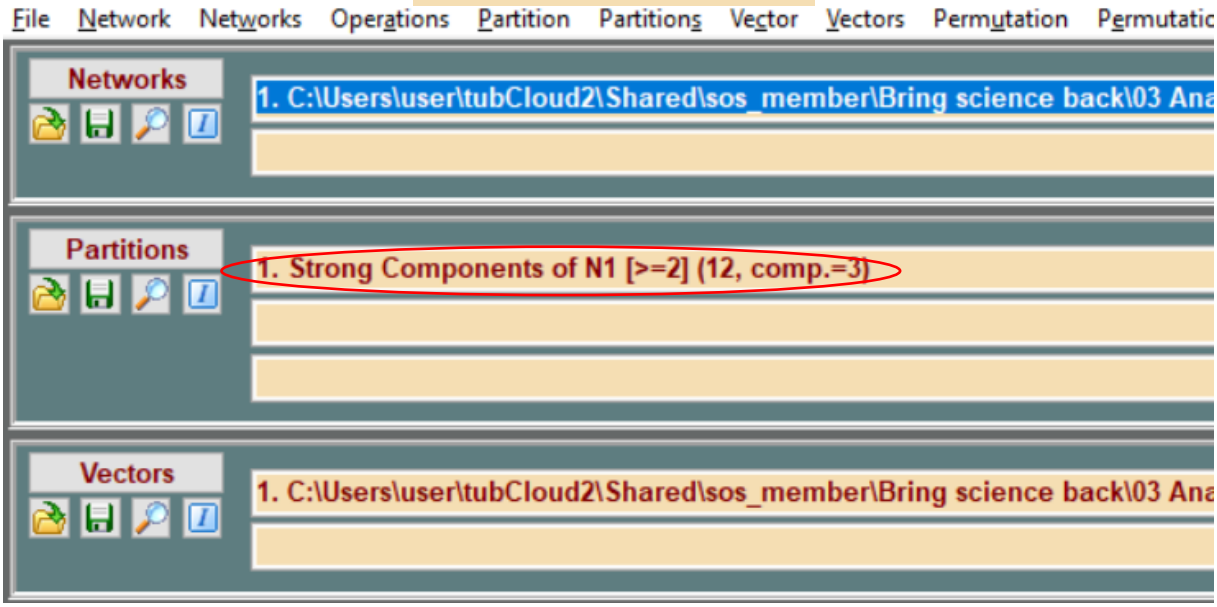


[7] Pajek can be downloaded from http://pajek.imfm.si/doku.php?id=download. There is also a manual that can be downloaded but this is not really helpful because it just explains the menus and functions without any 'how to' – advice. Laudel and Gläser have worked with the book Exploratory Network Analysis with Pajek by Wouter de Nooy, Andrej Mrvar and Vladimir Batagelj (2009).
Pajek can also be run on MAC and Linux, but with some additional steps needed (see the homepage).

7. Select Network->Create Partition->Components->Strong. Pajek will ask you about the minimum size of a component. Put in “2”.

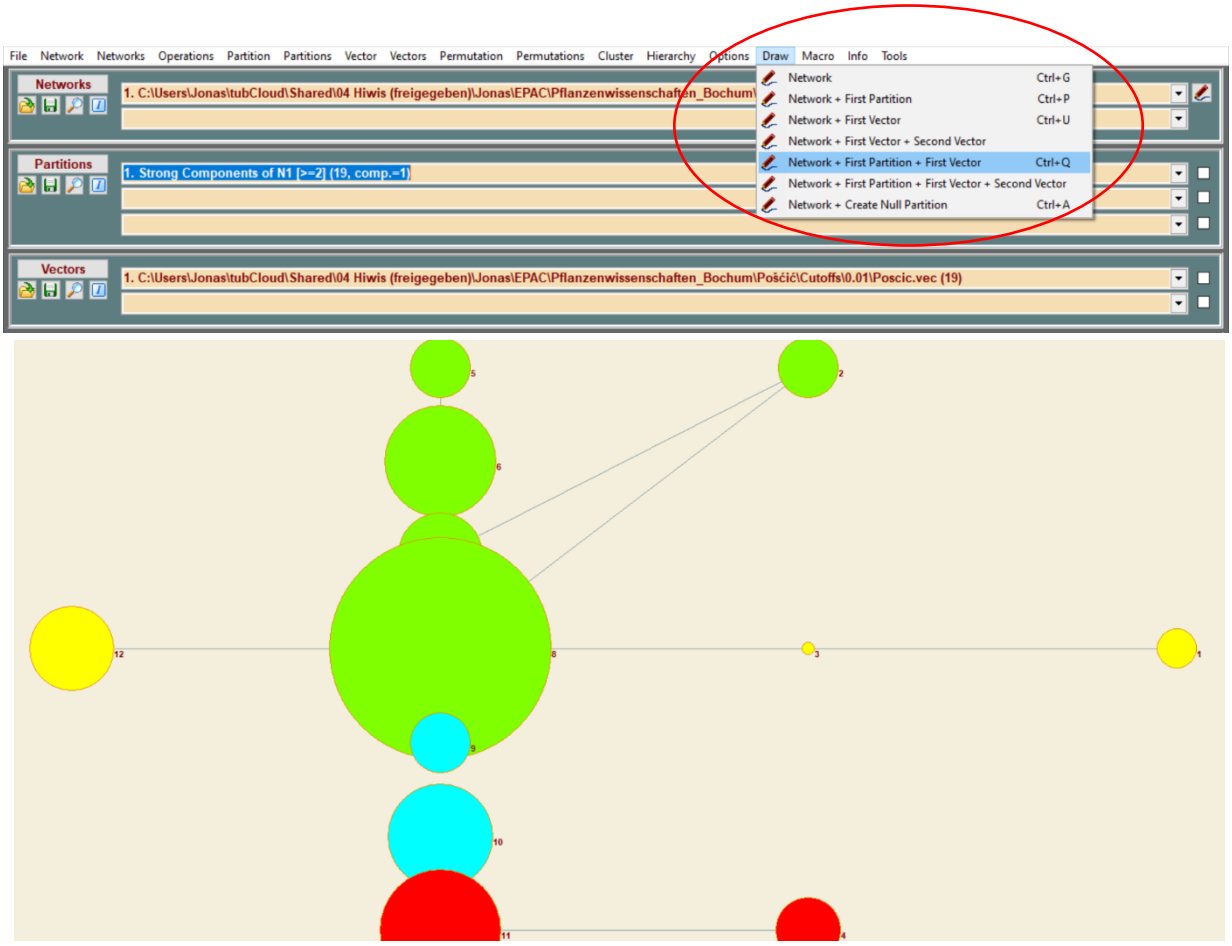




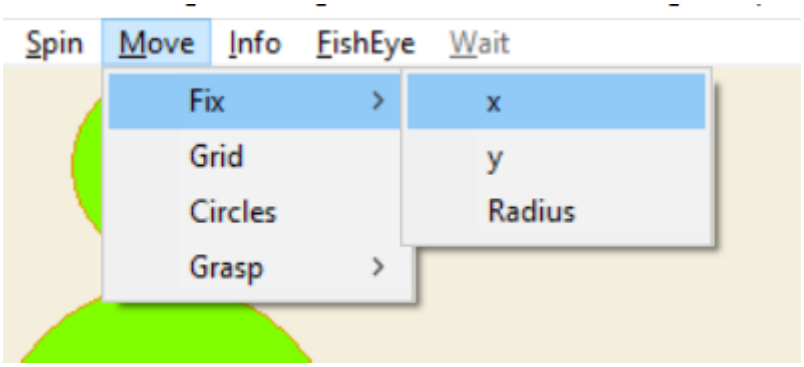


Pajek will now analyse your network and write the clusters it found in a partition file, which will appear under “Partitions”.

8. Now go to Draw->”Network + First Partition + First Vector” (Shortkey: Ctrl+Q). Pajek will draw a picture in which the publications belonging to different clusters have different colours. The picture appears in a separate window and should look something like the picture below.



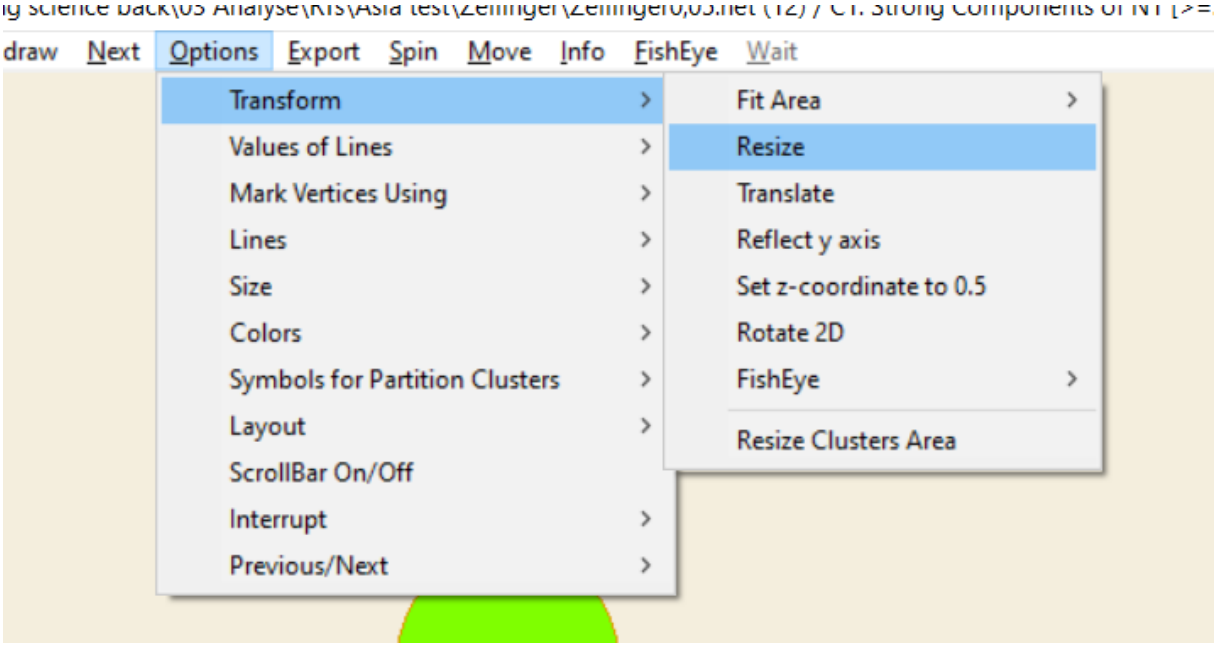
9. From the menu bar of the picture, select Move->Fix->”x”. This holds the publications in the position assigned to their year and restricts movements to the y-direction (so up and down). This is important because Pajek has no time axis and no grids, and it is very easy to accidentally move a publication to a different year. If you feel safe later, you can ‘free’ the x-direction (by again selecting Move->Fix->”x”). and slightly vary the x-coordinate in order to see all connections of a vertex (publication).

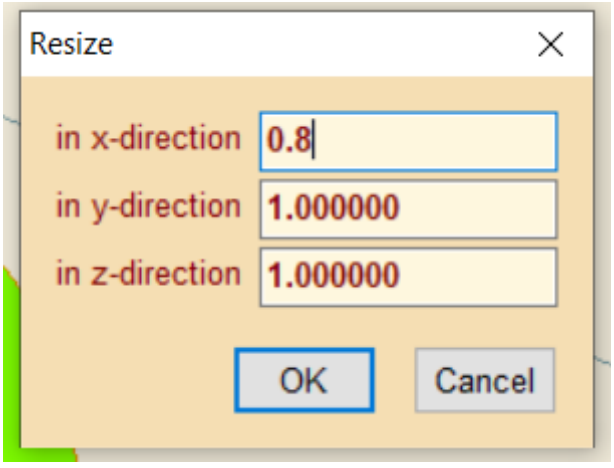


3.2. Working with the network picture using Pajek

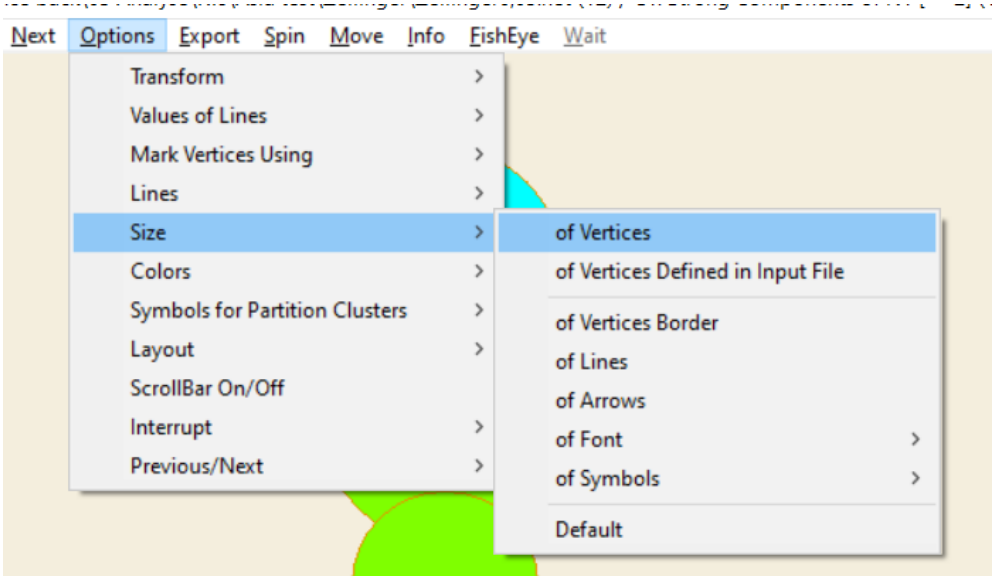
3.2.1. Changing the size of network picture

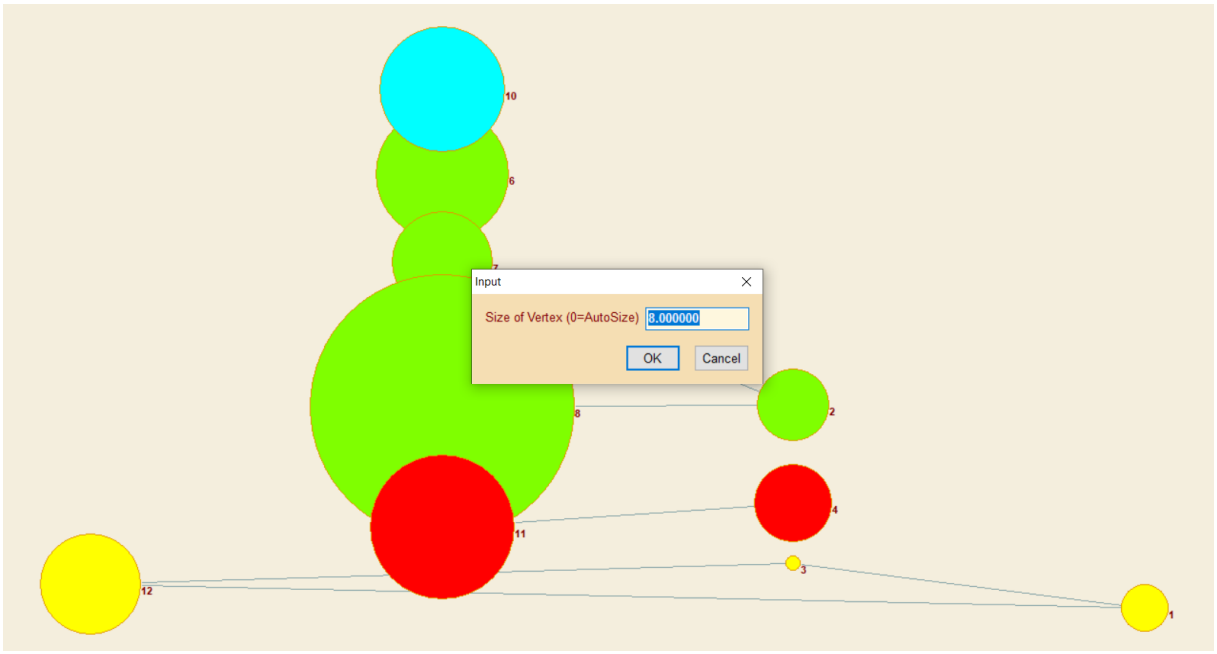
Move the circles around (preferably only along the y-axis) to produce the nicest picture. By clicking on a line of a cluster, you can move the whole cluster around. You can change the proportions of the picture by resizing in X or Y direction. For example, for long research trails (often older researchers) you might need to have a lower value. Simply put a new value in, e.g., “0.8”.

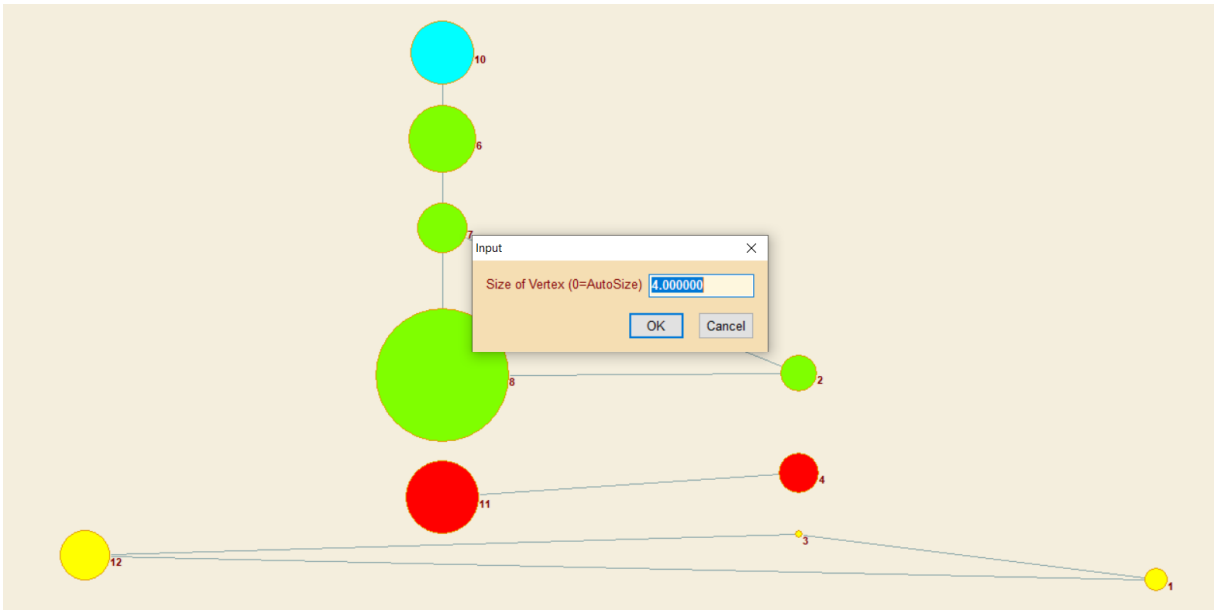




But this only changes the size of the picture, not of the vertices in it. If they are too big and start overlapping too much you can make the circles smaller by going to Options->Size->”of Vertices”. Depending on which value you find in there, choose any lower value and you will get more room:

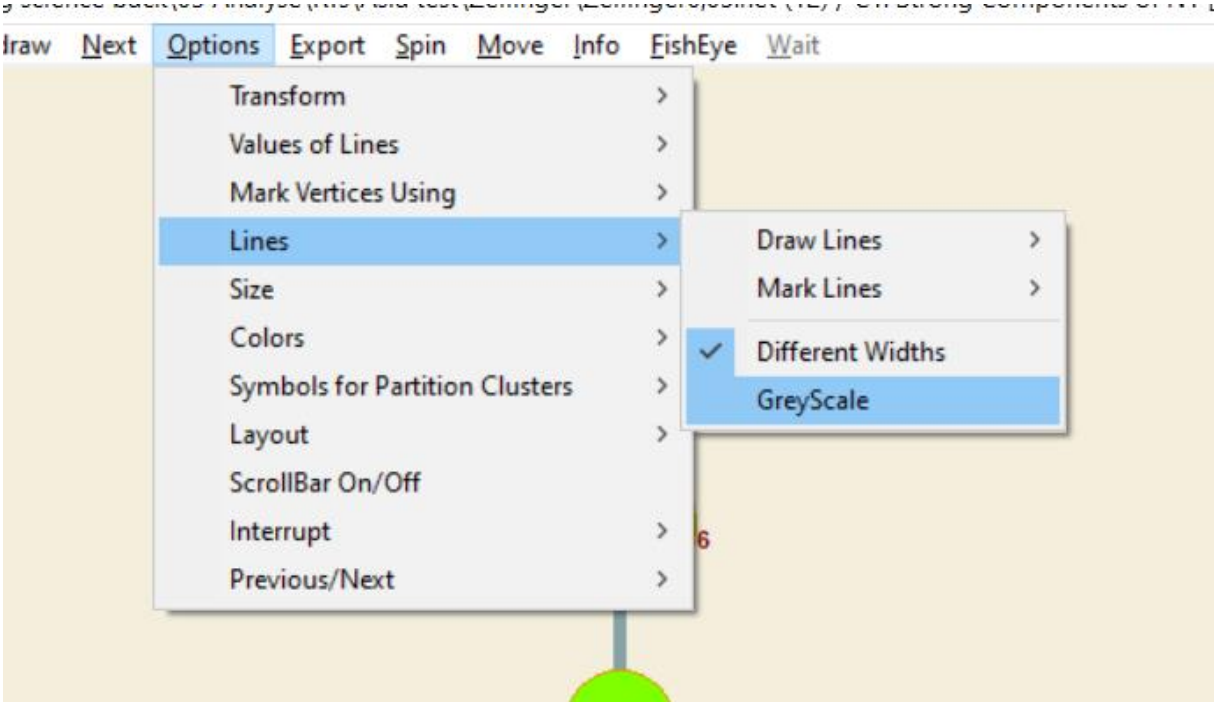

Next
Options
Export
Spin
Move
Info
FishEye
Wait
Transform
Values of Lines
Mark Vertices Using
Lines
Size
Colors
Symbols for Partition Clusters
Layout
ScrollBar On/Off
Interrupt
Previous/Next
of Vertices
of Vertices Defined in Input File
of Vertices Border
of Lines
of Arrows
of Font
of Symbols
Default


Input
Size of Vertex (0=AutoSize)
8.000000
OK
Cancel


Input
Size of Vertex (0=AutoSize)
4.000000
OK
Cancel

3.2.2. Changing the appearance of the connecting lines

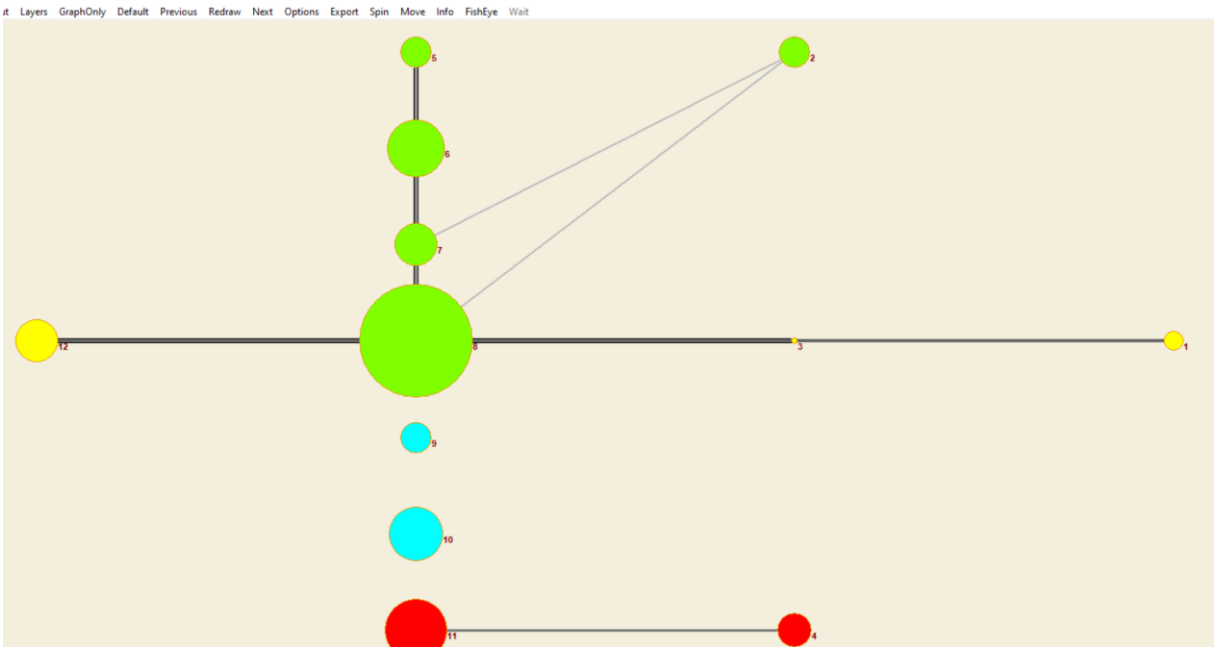
To add relevant information to your Research Trail, you can alter the way Pajek is rendering the lines between the circles. Go to Options->Lines and click on ”Different Widths” as well as on “GreyScale”. Now lines will be displayed with different width according to the strength of the connection between two publications and in lighter or darker shades of gray, depending on the width of the line. This is particularly useful for research trails with many publications.



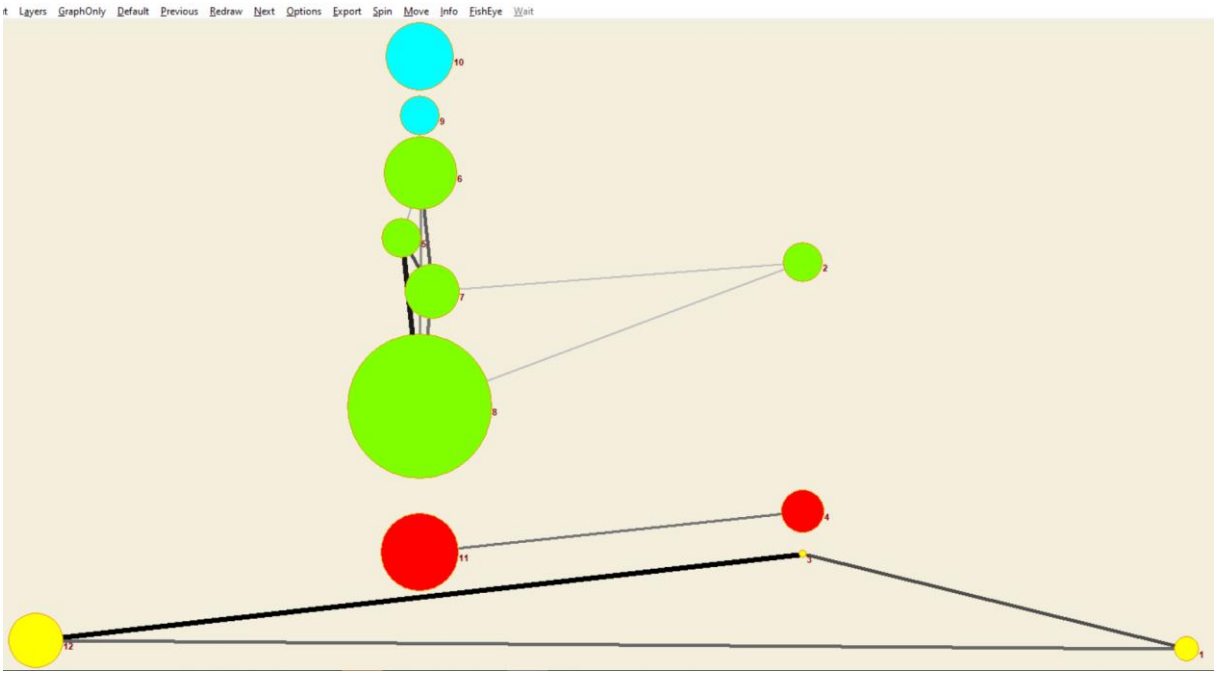
3.2.3. Rearranging the clustered publications

Circles can be rearranged via drag and drop. Remember to fix the x-axis (From the Drawing menu bar, select Move->Fix->”x”).

- Orphans (circles not connected to anything else) are getting pushed to the edges.
- Circles belonging to the same cluster (connected with lines and of the same colour) should be grouped in the same area.
- As many connections between publications as possible should be visible and distinguishable. You may need to cautiously move the circles on the x-axis if there are more than three publications in one year.



/\ Before & After \/



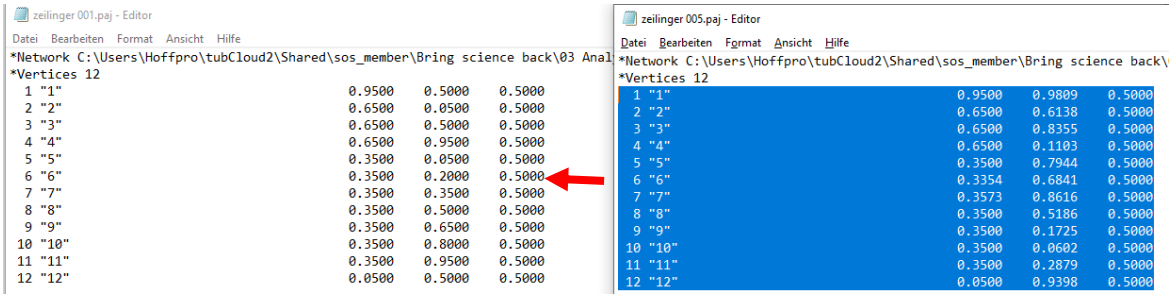
Saving and transferring your progress through Pajek-project files:

- Before you export your network as a picture, we recommend that you save your progress as a Pajek-project file. This is especially useful for: 1) long research trails; and 2) research trails of the same data with different thresholds.

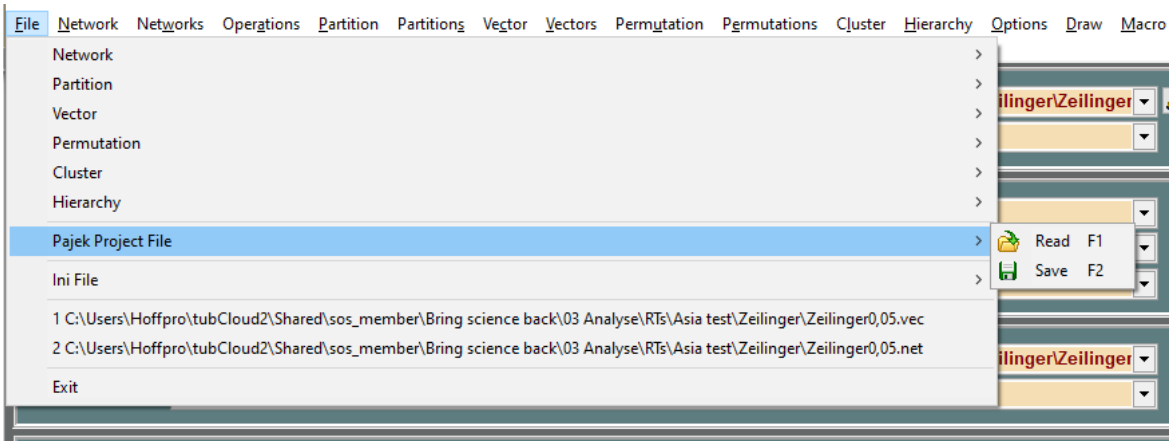
- For longer research trails, it is useful to not have to do all the positioning in one go. And if you realize that you would want to reposition some publications, this is the easiest way.

- For different thresholds, it is possible to rearrange your clusters as you want for one threshold, save the project file, and then copy-paste the positions (and if you changed it, also the size) of your publications into the project file of a different threshold-version. This will make it easier to produce comparable research trails.

It is generally recommended to try out some different thresholds for the Salton's cosine because it will show you different options of clustering publications that might help you to get an idea of which publication might belong with which topic.

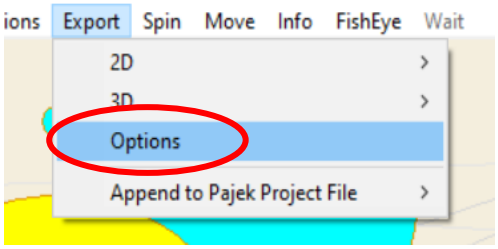


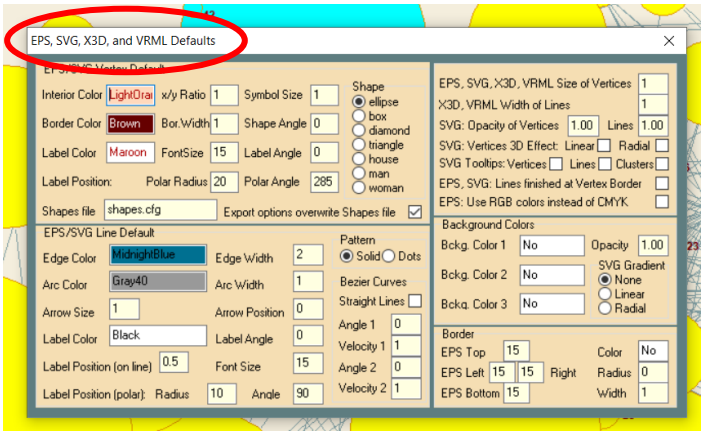
- To save (and import) a Pajek project file go to File->Pajek Project File and choose the respective option.



3.2.4. Using Export Options to further modify the network picture

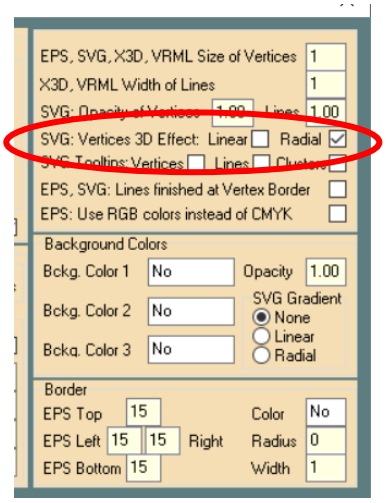
The Export Options in Pajek can act as a shortcut to modify the network picture without using external programs. Depending on your needs, these options can save a lot of time which would be spent tinkering with external programs otherwise. To open the Export Options in Pajek go to Export->"Options". An overlay called "EPS, SVG, X3D, and VRML Defaults" opens. Settings made here will only be visible in the exported SVG file. **They will not be displayed in the Pajek picture itself!**





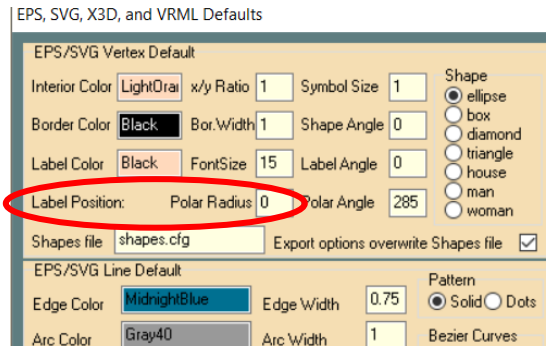
3.2.4.1. Add a radial 3D-Effect to Circles (optional)

Activate the option **Radial** next to **SVG: Vertices 3D Effect**. In the exported SVG file, the circles will now have a 3D effect. This step can also be done in Visio and InkScape (where it will look considerably nicer). If you want a more granular control of the visual effect you should consider using Visio or InkScape as explained in chapter 4.



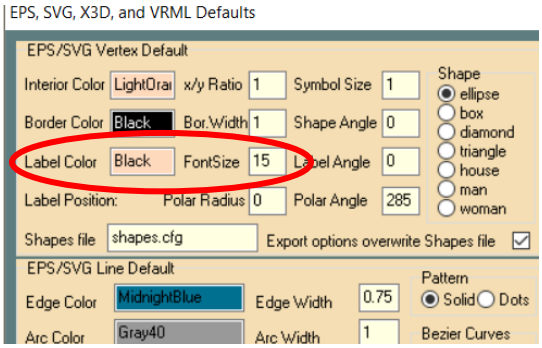
3.2.4.2. Adjust the position of the number labels

Change the value of the **Polar Radius** (default value 20) and/or Polar Angle. They define the distance and direction at which the numbers will be displayed in relation to the centre of the circle. 0 will centre the numbers on the circles. This step can also be done in Visio and InkScape once the network picture is exported. Depending on the desired outcome doing it in Pajek can safe you some time later on.



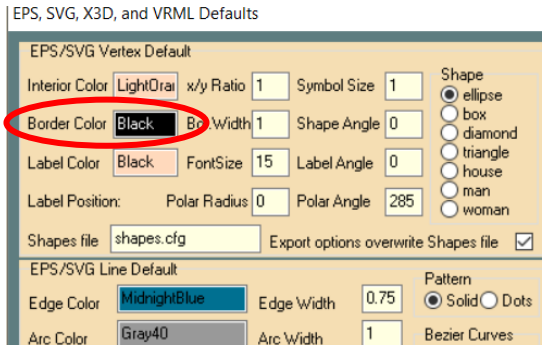
3.2.4.3. Adjust font colour and size

Click on **Label Colour**, select the preferred colour and close the colour selection. The font size can be changed via the value for **FontSize**.



3.2.4.4. Adjust the border colour of the circles

Click **Border Colour**, select the preferred colour and close the colour selector.

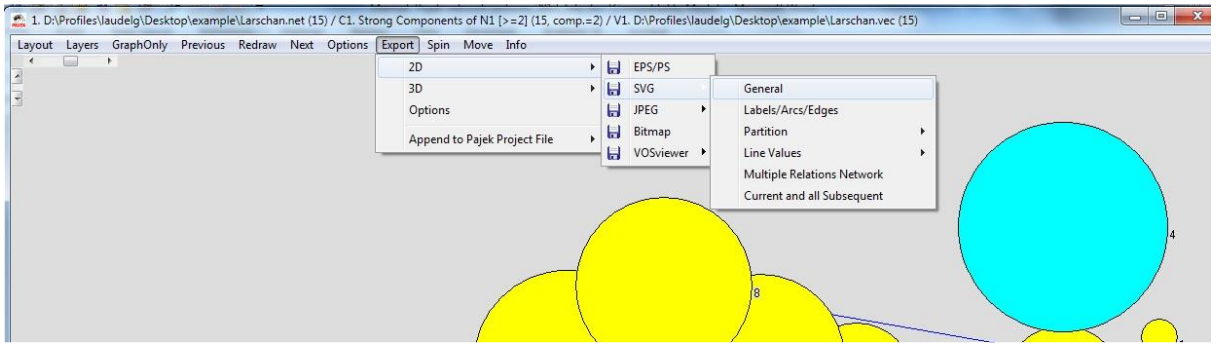


3.2.4.5. Adjust line width

Adjust the value of **Edge Width**. The default value is 2. Changing the value will make all lines thinner (<2) or thicker (>2). For research trails with a lot of connections, a value between 0.5 and 1.5 is recommended for better readability.

3.2.5. Export the Pajek picture for further processing

Export the Pajek picture as a 2D .svg file by going to Export->2D->SVG->"General". You will be asked for a filename.



It is easiest to use the researcher's name (+ Salton's Cosine) again. Further adjustments can now be made using Visio, Inkscape or draw.io (or any other fitting program).

4. Further refinement of Research Trails using external programs

The following processing steps are shown using Microsoft Visio. However, Inkscape can also be used to achieve the same goals. Draw.io is another program that can be used for diagram creation and is very easy to work with. However, with draw.io you will not be able to manipulate the network picture itself as draw.io can only load the network picture as a picture file.

Depending on your workflow you can skip one program or the other. We will show the following workflows with different examples.

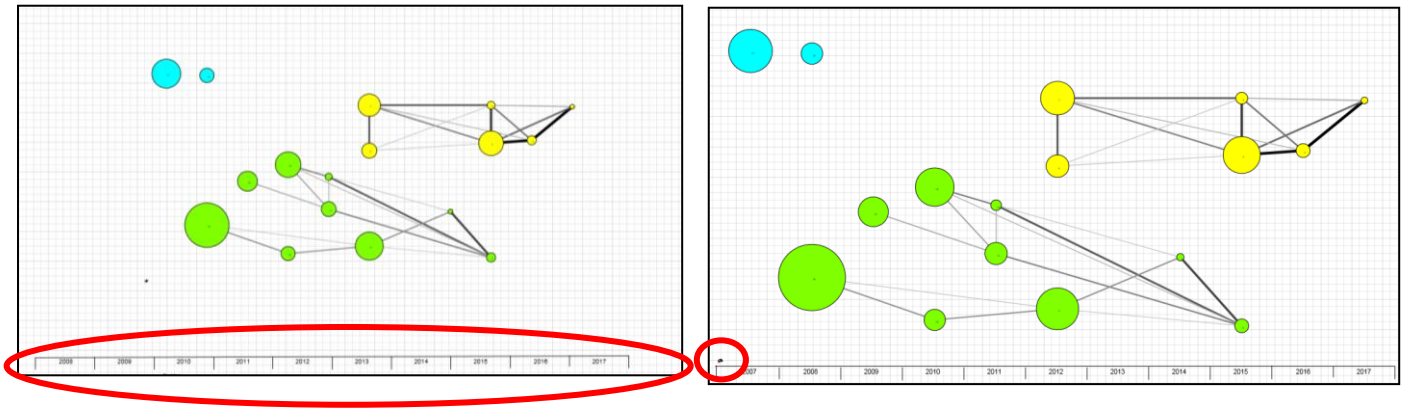
- Pajek (drawing your network picture) → Visio (finishing your Research Trail)
- Pajek (drawing your network picture) → Inkscape (refining your Research Trail) → draw.io (finalizing your Research Trail)

4.1. Finishing Research Trails using Visio

To open your network in Visio open Visio and go to Insert → Picture → From File. Locate your .svg file and insert it (the pictures shown now are from a different example).

4.1.1. Align network picture and add text

It is best to have a template for a timeline at hand. Adjust your timeline to match the years of the publications in the network picture. Align the network picture to the lower left edge. Pull apart at the upper right corner until each publication is correctly assigned to its respective year. It is important not to change the proportions of the network picture.

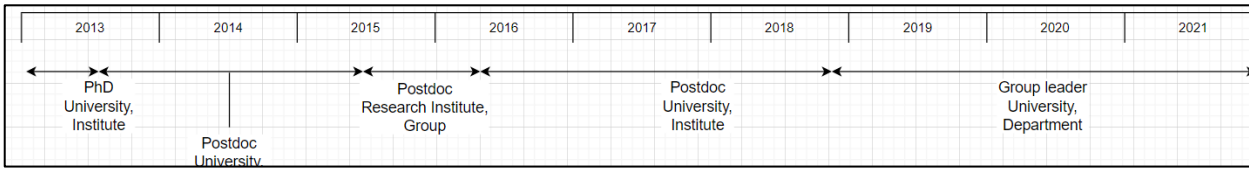


Also now is a good time to remove the little Pajek-Spider (which automatically gets inserted into every Pajek export) in the lower left corner of the network picture. Simply select it and press the delete key on your keyboard.

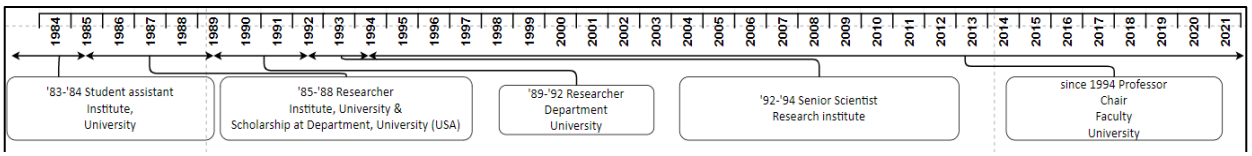
Add the author's career data (institutional affiliations) to the timeline.

Creating a **template for timelines** can save a good amount of time. At best you have two templates at hand depending on the length of your Research Trail. Two examples are shown below.

Spread out design for covering shorter periods of time:



Compact design for long timelines:

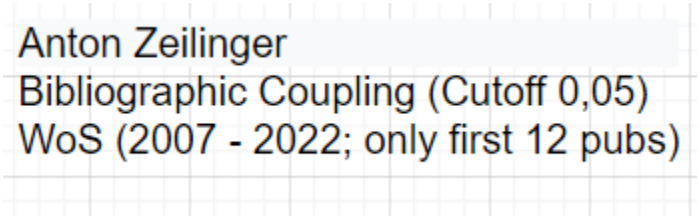


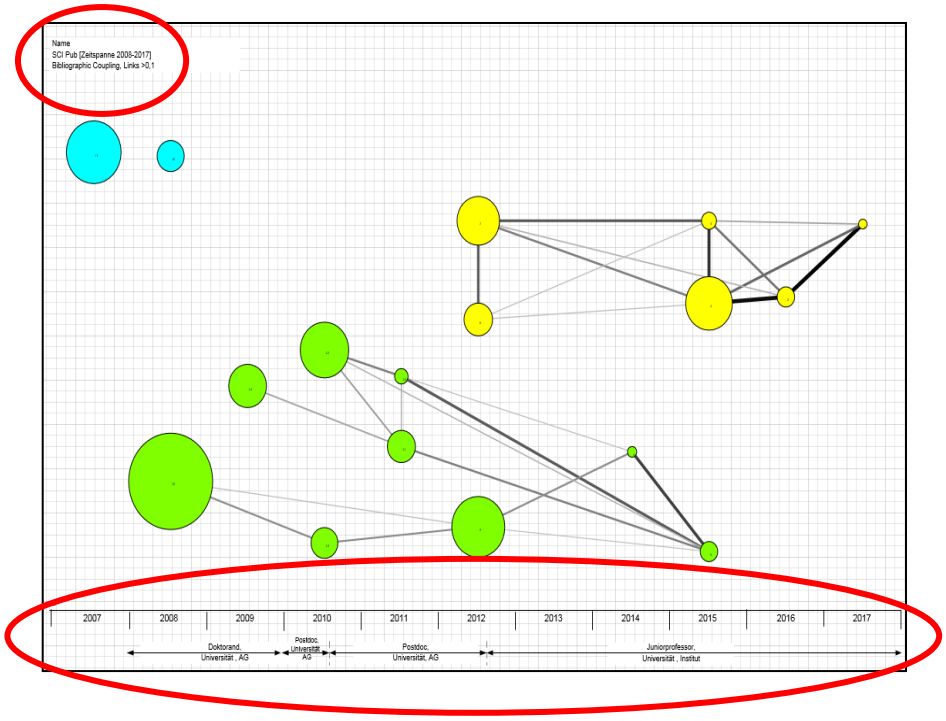
Necessary **meta data**, describing the research trail, should be added in the upper left corner as shown here:

[Name of Researcher]

Bibliographic Coupling (Cutoff [value]) →

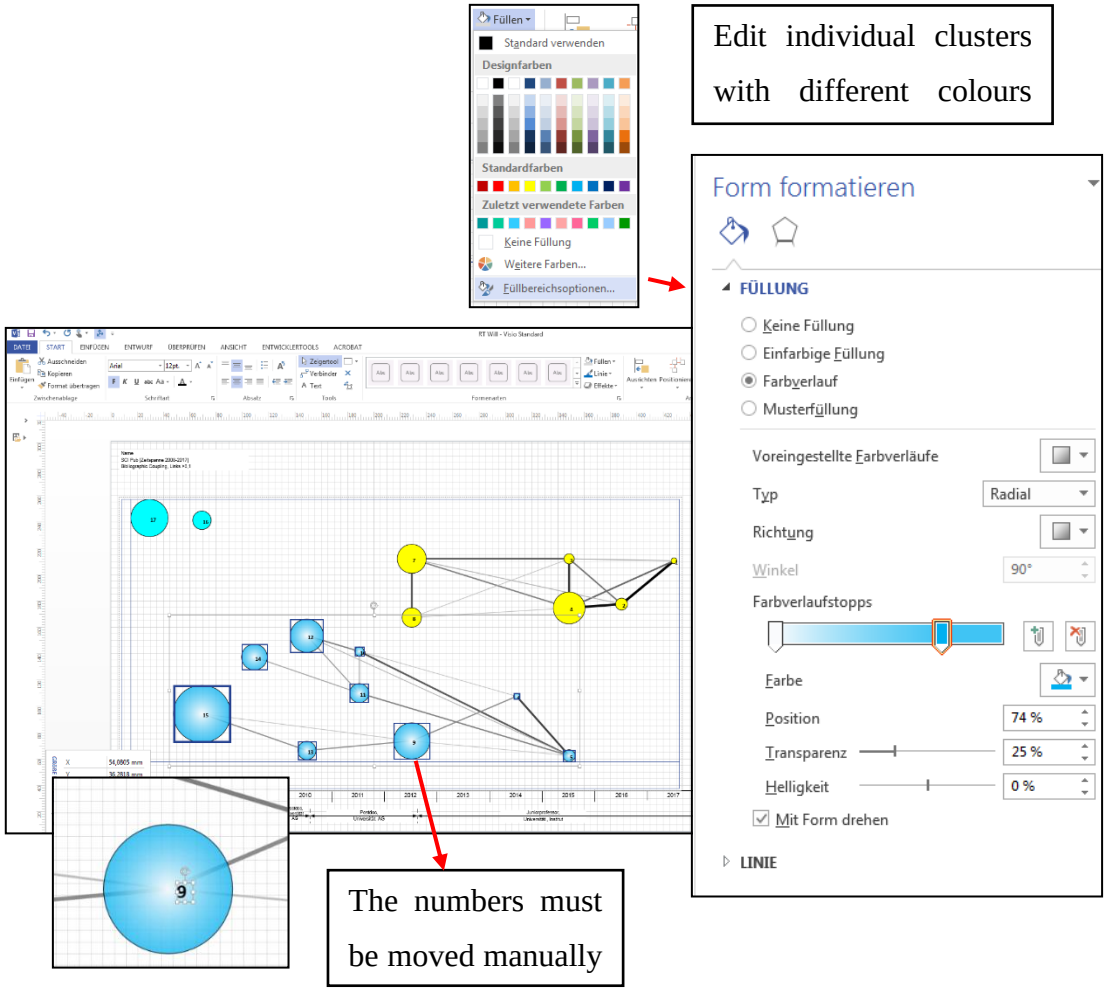
[Data origin] ([Year-Year])





If you **ungroup the Visio image**, you can also access individual vertices and edges. However, when you move the circles, the edges will not follow! They are not firmly connected to the circles in the traditional Visio way.

4.1.2. Add a radial gradient and move the numbers



4.2. Finishing the Research Trail using Inkscape and draw.io

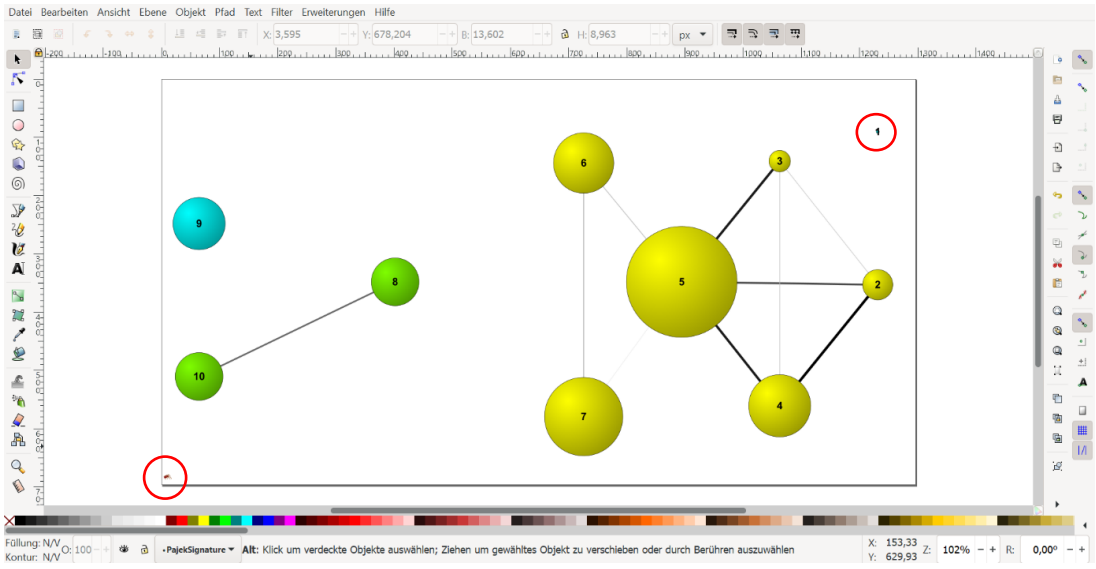
4.2.1. Basic edits with Inkscape

1. Navigate to the .svg file Pajek exported.

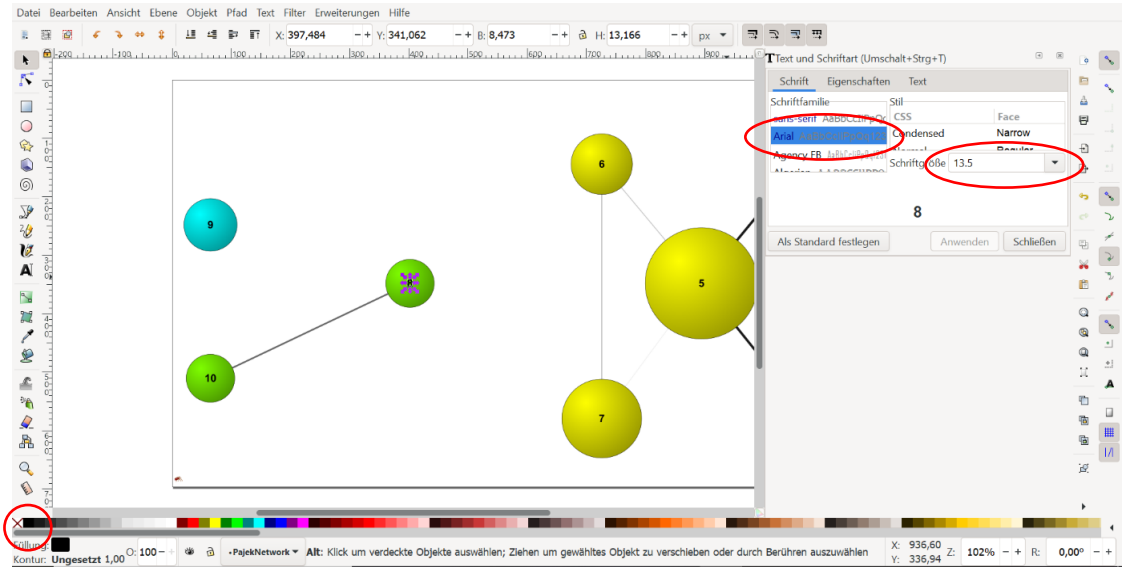
2. Right click->open with->Inkscape.

3. Delete the Pajek spider.

4. You may have to move some numbers to see smaller circles. Do so by clicking on them and drag them to you preferred location.



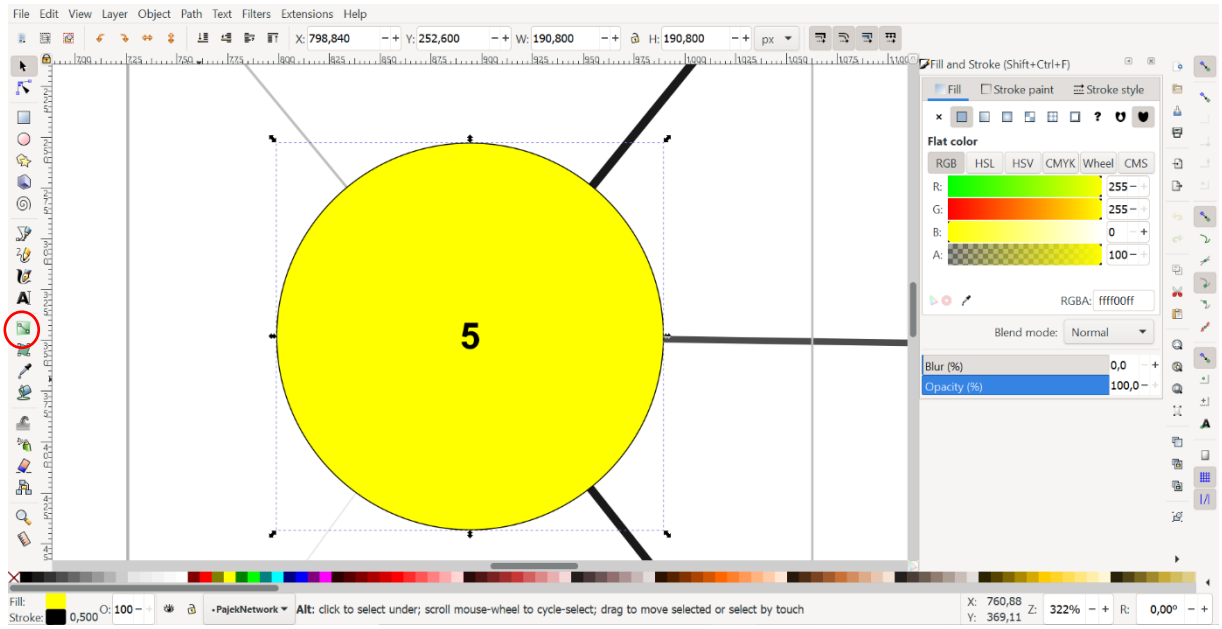
5. If you did not do this in Pajek, you can also move all the numbers. You also may colour them black and adjust size and font. Do this by selecting all of them using Ctrl+Shift and clicking on them. Click on “black” in the bottom left corner. Press Ctrl+Shift+t to open the text and font menu. Adjust font and size to a preferred one.



4.2.2. Adding a radial gradient with Inkscape

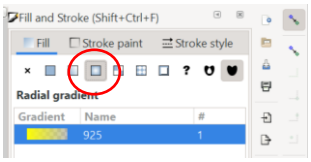
If you did not do this in Pajek, you can also add a radial gradient.

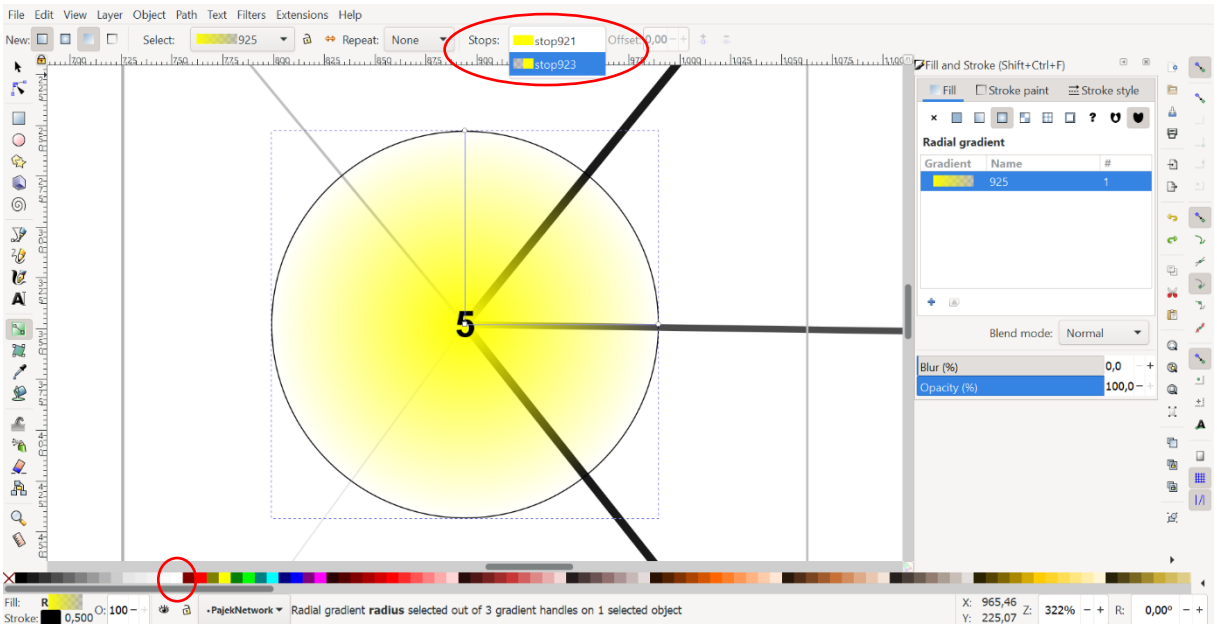
1. Press Ctrl+Shift+f. This will open the Fill and Stroke menu. Select the circle you want to edit by pressing Ctrl+Shift and clicking on them. Click on “Create and edit gradients” on the left side.



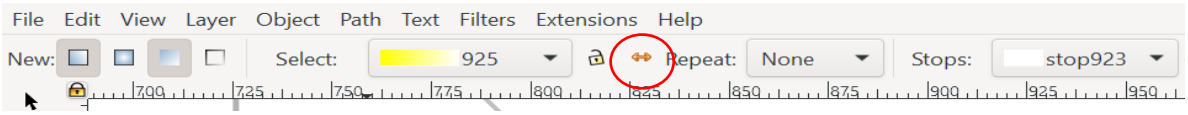
2. Click on “Radial Gradient”.

3. Change the colour of the gradient from transparent to white by clicking on “Stops”, selecting the second one and clicking on “white” in bottom left corner.

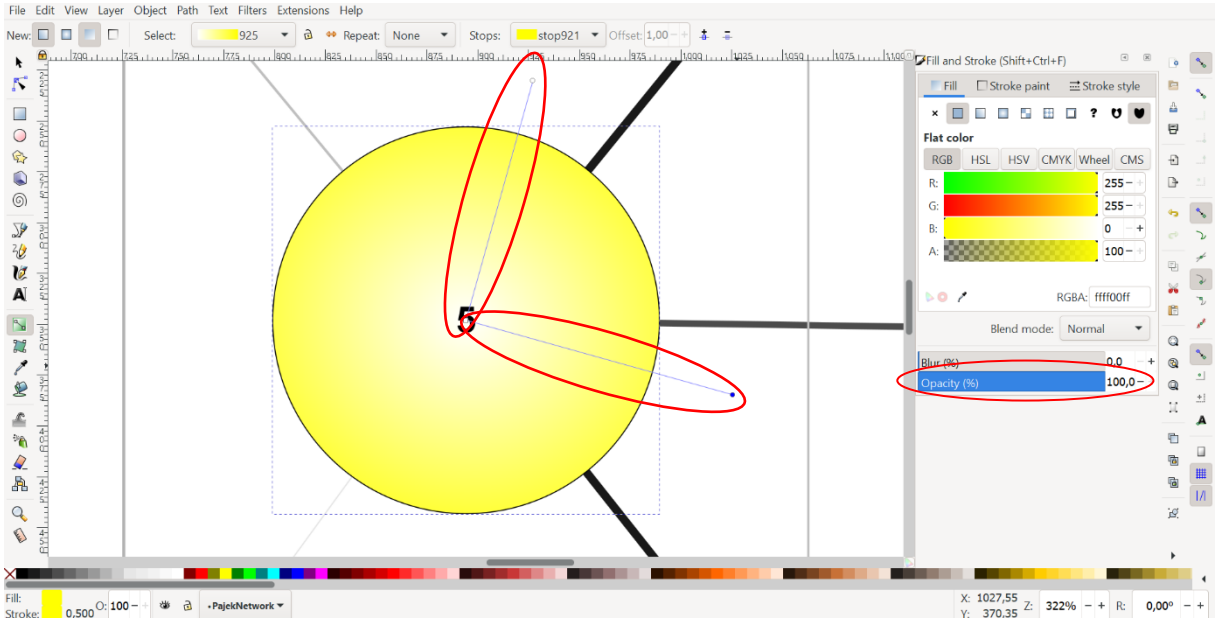




4. Reverse the gradient by clicking on the symbol on the top toolbar shown in the picture.



5. You may now play around with the gradient by dragging around the lines or the opacity by adjusting the slider shown in the picture.



6. During this process you may also change the colour of the circles, for example to colour all orphans in grey.

7. When done with your edits in Inkscape, just save the file.

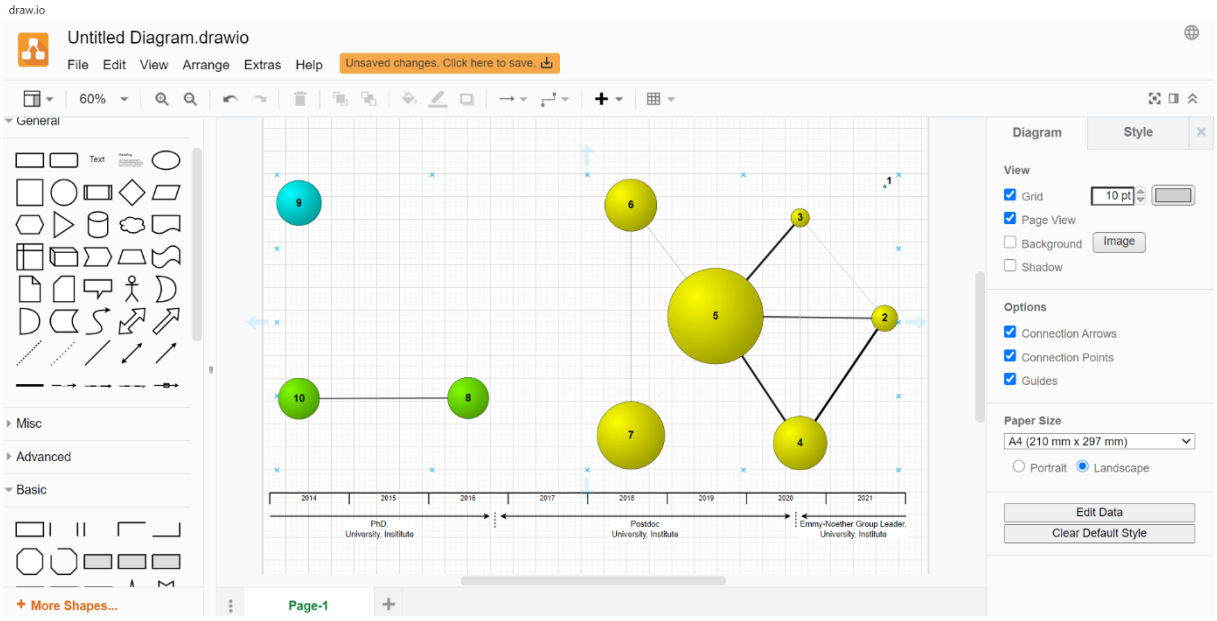
4.2.3. Basic edits with draw.io

1. Open draw.io. Set page size to A4 by clicking File->”Page Setup” and choosing “A4” and “Landscape”.

2. Navigate to the .svg file you just edited with Inkscape in the windows explorer and drag it into draw.io.

3. Open your timeline template and copy it into draw.io (or open a template draw.io-file to begin with).

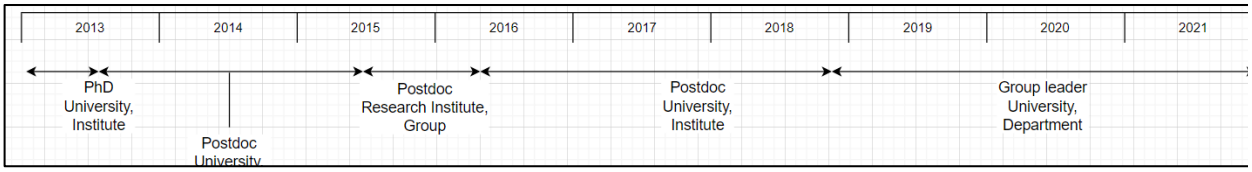
4. Align the timeline and the RT image by clicking and dragging the corners. You should doublecheck the alignment of the publication years and the circles.



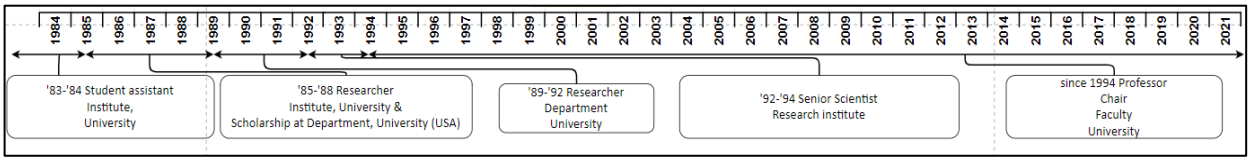
5. You may now name the clusters and add the titles of some publications if there is a reason. Also add the information about the affiliations of the researcher.

Just as with Visio, creating a **template for timelines** can save a good amount of time. At best you have two templates at hand depending on the length of your Research Trail. Two examples are shown below.

Spread out design for covering shorter periods of time:



Compact design for long timelines:



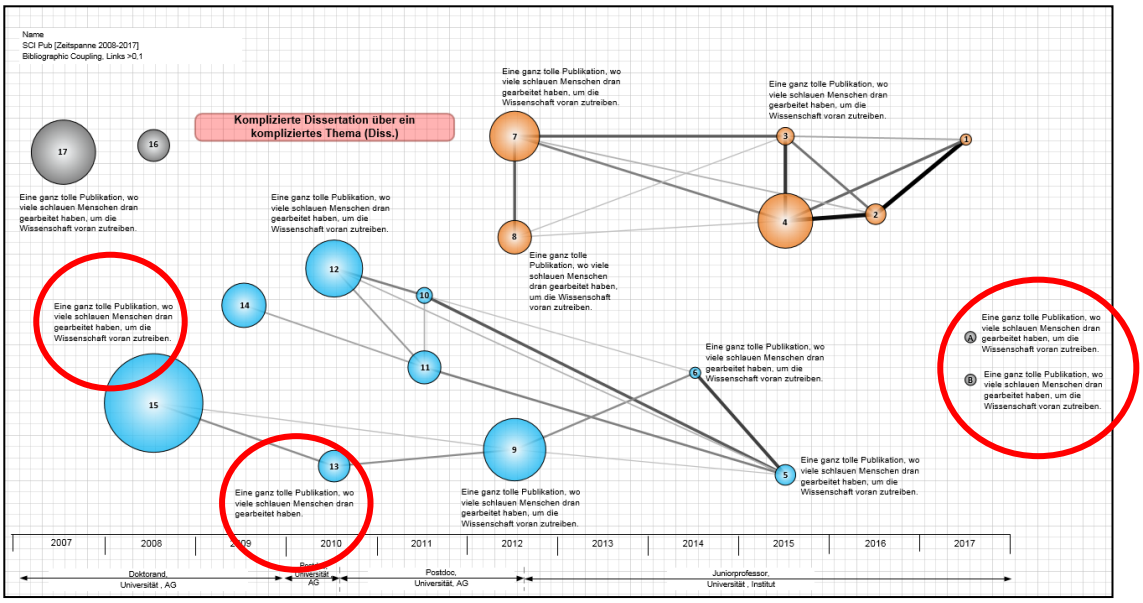
4.3. Creating Research Trails for internal and external use

In a last step you can add additional info into your Research Trail to make the contained information more accessible. For example, this can be done by highlighting and colour-coding important publications of the researcher and naming the different clusters according to their possible topic.

Furthermore, it is advised to create two versions of your Research Trail: one for internal use and another one for external use. This way you can keep the external version clean and condensed, while not losing track of additional info you may need for conducting an interview. Shown below is a Research Trail in two versions just as explained above.

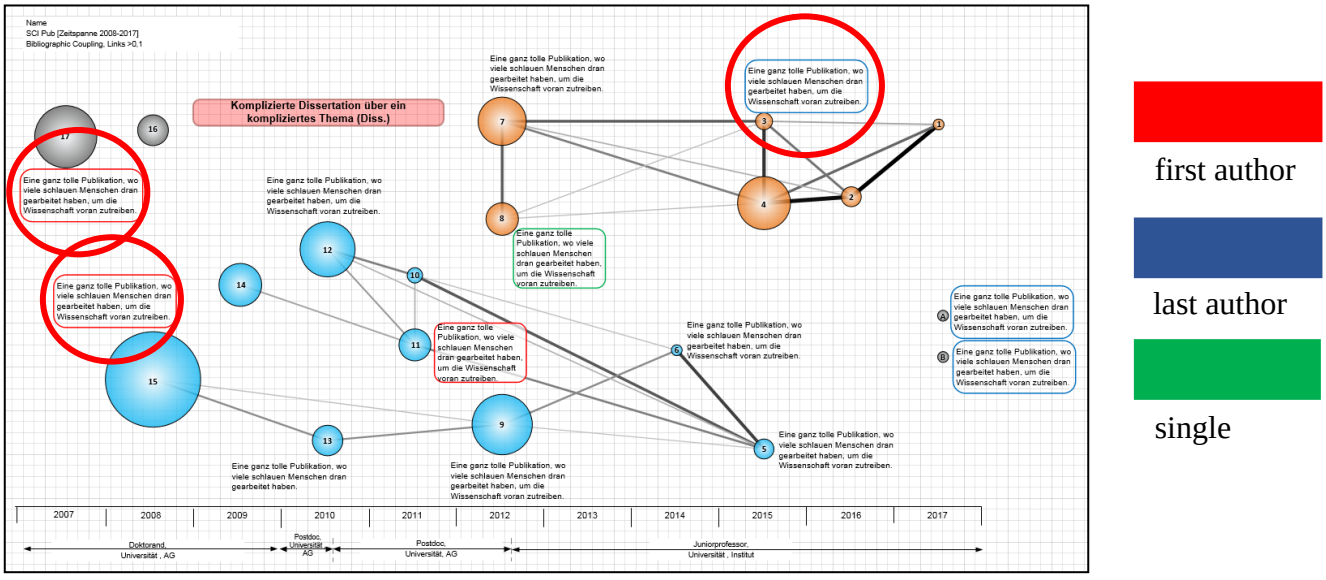
External Version

The titles of certain publications are added. This can include: first, last and single authorships and publications with a lot of citations:



Internal Version

For internal use the highlighted publications are getting colour-coded:



5. Keeping track of publications and clusters

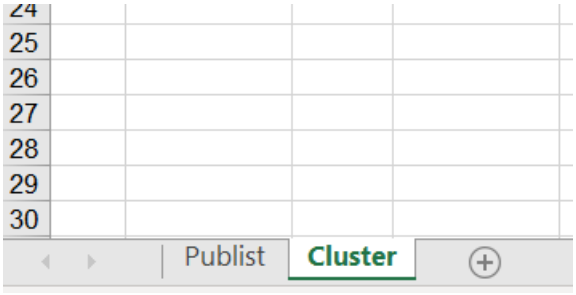
5.1. Publication list sorted by clusters

After creating a research trail with Pajek, the clusters generated should be examined again regarding their actual thematical overlap. For this purpose, a publication list sorted by clusters is useful. The steps could be as follows:

1. Create a separate Excel file with all publications used in the Research Trail. You can find all publications in the Excel file created by the macro itself. Your new Excel file should contain all relevant information about the publications:

Publication Number (this is important to identify each publication on the Research Trail), Author(s), Title, Keywords, Year, Journal Title, Type of Publication, DOI.

2. Create a new table named **Cluster** in your Excel file.



3. Use the Reseach Trail to identify the clusters with their respective publication numbers. Copy the publications of a cluster from your **Publist** table and paste them to your **Cluster** table. Colouring the publications according to the cluster from the Research Trail can help to better navigate the documents.

	A	B	C	D	E	F	G	H	I
1	Pub. Nr.	Authors	Title	Categories	Keywords	DOI	Year	Journal Title	Type of Pub.
2	WHISP Galaxies								
3	56	van Eymeren,	Lopsidedness in WHISP galaxies II. Morphological lopsidedn	surveys; galaxies: evolutic	SPIRAL GALAXIES	10.1051/0004	2011	ASTRONOMY	Article
4	57	van Eymeren,	Lopsidedness in WHISP galaxies I. Rotation curves and kiner	surveys; galaxies: evolutic	WESTERBORK HI	10.1051/0004	2011	ASTRONOMY	Article
5									
6	CALIFA								
7	23	Lopez-Coba, (	Star formation driven galactic winds in UGC 10043	ISM: jets and outflows; ga	INTEGRAL-FIELD	10.1093/mnra	2017	MONTHLY NC	Article
8	25	Mendez-Abre	Two-dimensional multi-component photometric decompos	galaxies: bulges; galaxies:	SURFACE-BRIGHT	10.1051/0004	2017	ASTRONOMY	Article
9	26	Sanchez, SF; C	CALIFA, the Calar Alto Legacy Integral Field Area survey IV. T	galaxies: evolution; galaxi	POTSDAM MULTI	10.1051/0004	2016	ASTRONOMY	Article
10	35	Garcia-Benito	CALIFA, the Calar Alto Legacy Integral Field Area survey III. S	surveys; techniques: spec	MASS-METALLICI	10.1051/0004	2015	ASTRONOMY	Article
11	50	Sanchez, SF; K	CALIFA, the Calar Alto Legacy Integral Field Area survey I. Su	techniques: spectroscopic	DIGITAL SKY SUR	10.1051/0004	2012	ASTRONOMY	Article
12	45	Husemann, B;	CALIFA, the Calar Alto Legacy Integral Field Area survey	techniques: spectroscopic	DIGITAL SKY SUR	10.1051/0004	2013	ASTRONOMY	Article
13									
14	GaBoDS								
15	85	Hildebrandt, H	GaBoDS: the Garching-Bonn deep survey - III. Lyman-break	galaxies : photometry; ga	STAR-FORMING (	10.1051/0004	2005	ASTRONOMY	Article
16	86	Erben, T; Schi	GaBoDS: The Garching-Bonn Deep Survey - IV. Methods for	methods : data analysis; t	CALIBRATION; G	10.1002/asna	2005	ASTRONOMIS	Article
17									
18	für mich auf die schnelle kein einheitliches Thema ersichtlich - im Großen und Ganzen wahrscheinlich nicht so weit weg vom Hauptcluster?								
19	63	van Eymeren,	A kinematic study of the irregular dwarf galaxy NGC 4861 us	galaxies: individual: NGC	HOMOGENEOUS	10.1051/0004	2009	ASTRONOMY	Article
20	65	Trachternach,	The baryonic Tully-Fisher relation and its implication for dar	dark matter; galaxies: kin	EXTRAGALACTIC	10.1051/0004	2009	ASTRONOMY	Article

4. Search for thematic overlaps via **Title** and **Keywords** of the publications and note them down if you find any. If a cluster does not have a common topic, write a comment about it in

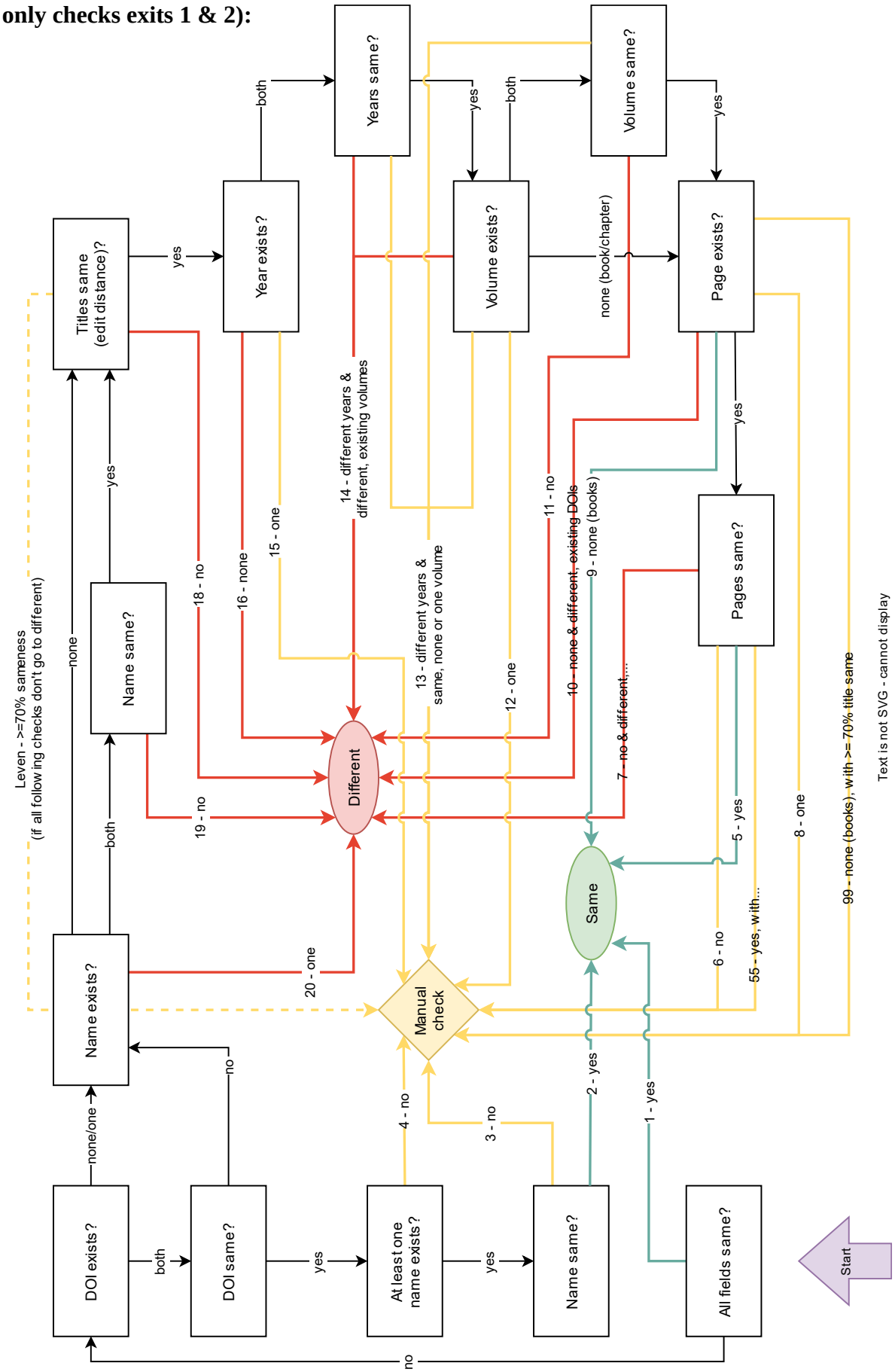
the Excel file. You can also use the titles and keywords of a cluster as a heuristic tool for finding a name of each cluster. Please note that this is only a helpful step in trying to find out what the clusters produced might mean. It will be necessary to read at least some publications and talk about the clusters in the interview to verify them. **If you realize that two publications, which should belong together are not in the same cluster, you may want to turn back to Step 2.2 and make a different research trail with another value for Salton's cosine. It is not possible to determine the 'correct' threshold before seeing and working with the outcome!**

5.2. Publication list for internal and external use

As stated above, you may want to use the publication list that the macro created for internal and external use. For the external use, e.g., to send to the researcher you are going to interview (or to bring with you at the interview), you can create a simple table listing publication number, author(s), title, journal, and maybe times cited. Without this publication list and the publication numbers, the research trail is not understandable. Here it is also advised to colour at least some parts of the rows in the cluster colours and to make the researcher's name **bold**.

No	Authors	Year	Title	Journal	TC
1	Fuenzalida, J; Hochrainer, A; Lemos, GB; Ortega, EA; Lapkiewicz, R; Lahiri, M; **Zeilinger**, A	2022	Resolution of Quantum Imaging with Undetected Photons	QUANTUM	2
2	Luo, YH; Chen, MC; Erhard, M; Zhong, HS; Wu, D; Tang, HY; Zhao, Q; Wang, XL; Fujii, K; Li, L; Liu, NL; Nemoto, K; Munro, WJ; Lu, CY; **Zeilinger**, A; Pan, JW	2021	Quantum teleportation of physical qubits into logical code spaces	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	5

6. End

We tried to gather everything we know about creating research trails. However, if you got any questions or want to make suggestions for improvement, do not hesitate to contact us (markus.hoffmann@tu-berlin.de). A lot of this document is based on what Grit Laudel did before, so a big thank you to her for the help.

Appendix – Decision tree for reference matching in "Extended_Trail" ("WoS_Trail" only checks exits 1 & 2):