

esss
european summer school for scientometrics 2021

Zpráva z letní školy
Bc. Denis Škopan, ÚK UK

Datum akce: Zář 20-24 2021

Web pořadatele: <https://esss.info/>

Program: <http://web.archive.org/web/20210914223612/https://esss.info/programme/>



O letní škole¹

Cílem účasti na této letní škole je rozvoj znalostí v oblasti bibliometrie a scientometrie. Letní škola se zaměřuje nejen na teoretické poznatky v těchto oblastech, ale především na jejich aplikaci – součástí akce jsou praktické workshopy, které jsou rozvrhnuty 1:1 k teoretické výuce. Workshopy, krom aplikace teoretických znalostí, umožňují hlubší proniknutí do problematiky – jsou zaměřené jak na základní otázky bibliometrie a scientometrie (např. normalizace předmětových hesel pro citační analýzu), tak na komplexní (např. tvoření institucionálních reportů). Součástí praktických cvičení bude práce s daty přímo dané instituce, což usnadní vzhled do problematiky a také přenesení získaných poznatků do praxe.

Řeší se každoročně vybraná témata zaměření kombinující standardní znalosti s aktuálními trendy. Program letní školy je navíc zaměřen tak, aby přesahoval i do open science, tedy do tématu, který rezonuje světovou vědou i knihovnictvím.

Jedná se o intenzivní kurz, kde je kladen důraz na získání jak obecné znalosti o bibliometrii a scientometrii, tak na představení nejmodernějších metod. Jedná se o již 11. ročník a tato akce se vždy shledává s vysokým ohlasem, nejen vzhledem k tématu, ale i kvalitou.

Scientometrické postupy se stále častěji používají k analýze vývoje a trendů ve vědě a technice a také formují rozhodování na různých úrovních procesu hodnocení vědy. Rozhodnutí, která jsou v důsledku těchto postupů přijata, mají často vážné dopady. Optimální konstrukce indikátorů, zpracování a

¹ <https://esss.info/about/>

interpretace dat však vyžadují kompetentní odborné znalosti, které jsou v současné době k dispozici pouze v omezené míře kvůli nedostatku příležitostí ke školení.

Pomocí důkladného představení základních teoretických konceptů mohou účastníci esss očekávat, že se seznámí s nejčastěji používanými databázemi, naučí se získávat a zpracovávat relevantní data, jak konstruovat relevantní ukazatele a jak data adekvátně interpretovat.

Letní školu společně organizují University of Vienna (Rakousko), German Centre for Higher Education Research and Science Studies – DZHW (Německo), KU Leuven (Belgie) a EC3metrics Group of the University of Granada (Španělsko).

Hlavní témata

- Definice, historie a institucionalizace bibliometrie, scientometrie a informetrie
- Úvod do nejběžnějších zdrojů dat
- Přehled a použití nejdůležitějších a relevantních scientometrických ukazatelů
- Teorie, přístupy a omezení předmětových klasifikací
- Zpracování dat: získávání, čištění a zpracování dat
- Úvod do metrik časopisů
- Normalizace předmětu pro citační analýzu
- Síťová analýza: teoretická východiska a praktické příklady pomocí různých nástrojů
- Osvědčené postupy bibliometrických aplikací a služeb
- Úvod a praktické využití alternativních metrik

Vybrané přednášky

General introduction

Přednáška byla koncipována jako úvod do k bibliometrii.

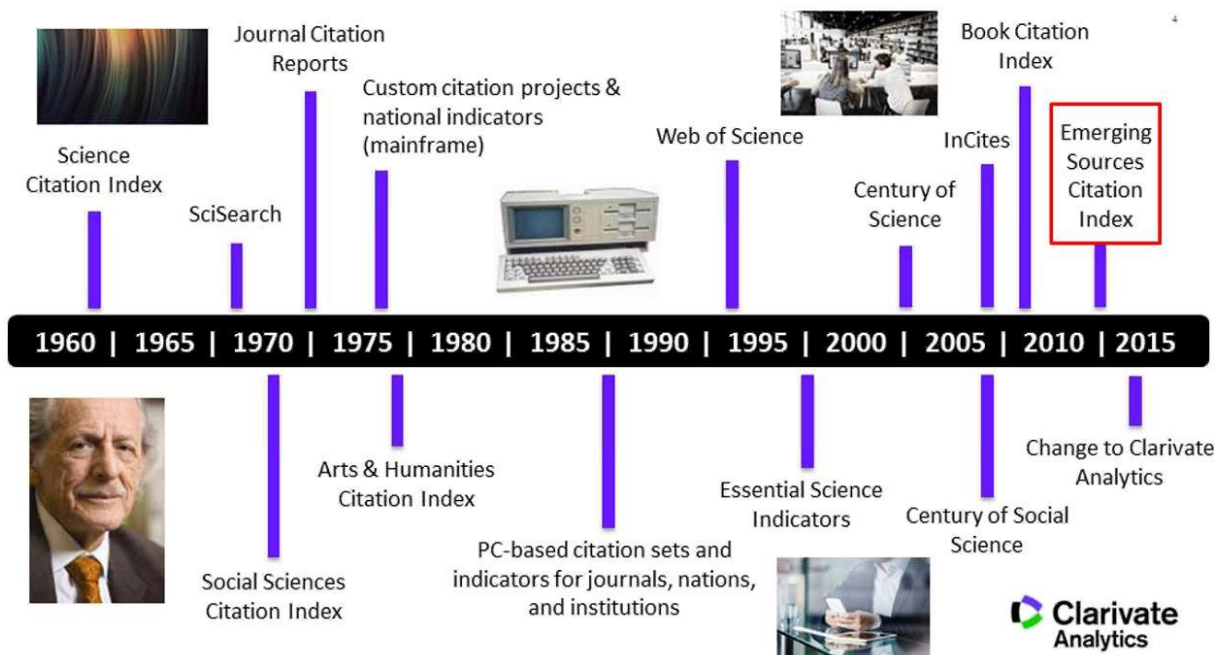
1. Vhodnost zdrojů dat (kombinovatelnost, přesnost, bias, trvalost, atd.)
2. Získávání a čištění dat (předmětová klasifikace)
3. Spolehlivost metodiky (co měřit, jak udělat z dat výstupy, kontext a benchmarking)
4. Síťová analýza (koautoství, kocitace)
5. Překračování hranic ve vědecké komunikaci (otevřená věda, altmetriky)
6. Bibliometrie v praxi (hodnocení vědy)

Introduction to bibliometric data sources

Blok přednášek, který představil zdroje dat pro bibliometrický výzkum, jejich silné stránky a slabá místa a to jak u tradičních velkých multidiscipinárních (WoS, Scopus) tak i oborových (MathScinet, SciFinder, astrophysics data system), nových alternativních (Dimensions, Google Scholar), datových (re3data.org, Databib, Data Citation Index, Mendeley Data) i otevřených (arXiv.org, CiteSeerX, Base, CRIS) a osvětlil problematiku standardizace bibliografických databází a jejich dat.



Knihovnický tip: Při analýze použít alespoň tři zdroje dat.



Datové zdroje se s postupem času mění, je potřeba na to myslet u retrospektivních analýz.

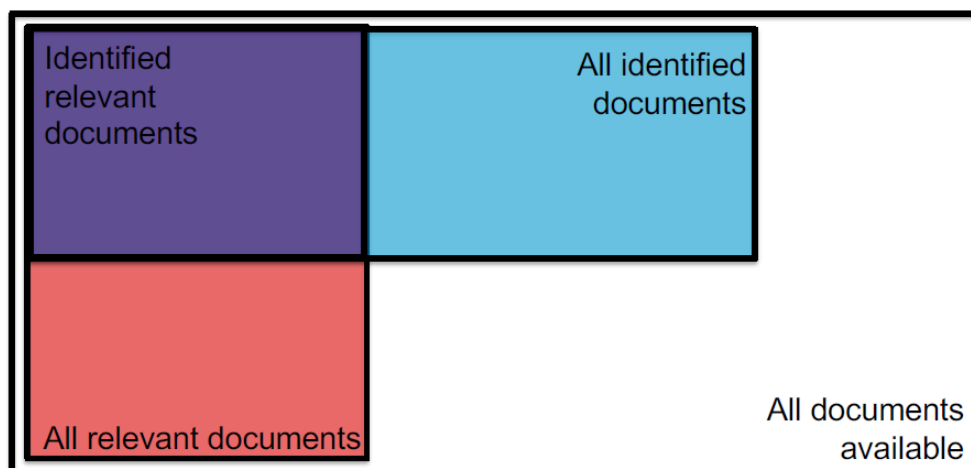
Accuracy and Beyond: Data Quality of Bibliometric Databases

Přednáška se zabývala kvalitou dat v bibliometrických databázích, konkrétně na dramatických rozdílech v ukazatelích ve WoS a Scopus na stejných příkladech. Dále byla také zmíněna problematika neúplných dat a rozmanitosti při zapisování afiliací k institucím.

Zajímavým bodem byla otázka vědecké mobility a sledování kariér subjektů z pohledu změn afiliací.

Designing effective queries

Precision (nebo také Positive and negative predictive values), aneb je vše, co jsem vyhledal relevantní vs **Recall** (nebo také Sensitivity and specificity), aneb našel jsem vše, co je relevantní?



True Positive = Relevant identified documents

False Positive = Identified but Irrelevant documents

False Negative = Relevant but unidentified documents

Precision = $TP / (TP + FP)$ = relevant \cap identified / identified

Recall = $TP / (TP + FN)$ = relevant \cap identified / relevant

💡 **Tip:** Je dobré v dotazech používat zástupné znaky jako *, ale musí se zvážit také možná ztráta významu (Electr* vs. Electron* vs. Electronic*).

💡 **Tip:** Pozor na přednost logických operátorů. Ve většině databází je posloupnost logických operátorů NOT > AND > OR. Závorky mají vždy přednost.

Scientometric Indicators in Use: an Overview

Pohled na indikátory jako na komplexní zástupné prvky reprezentující složitý koncept v realitě, reflektující skutečnosti, které nelze jednoduše měřit. Často se k nim vážou vágní termíny jako kvalita, výkonost, pokrok, užitečnost, význam atd. Vystává otázka nejen co měřit, ale proč to měřit.

Správný indikátor má být robustní (neovlivnitelný marginálními změnami), měřitelný (vhodný k měření toho, pro co byl zvolen) a reprodukovatelný.

- Bibliometrické indikátory – soustředí se hlavně na počty publikací
- Síťové indikátory – analyzují sítě pomocí koautorství a kocitací
- Citační indikátory – používají citaci k analýze dopadu výzkumu

Open Science & Bibliometrics

Krom tradičního úvodu do Open Science obecně přednášející zmínila, že přibližně 80% dat (pohled na 2000-2014) není vůbec citovaných a dále pouze 6% z těch citovaných má přiřazenou disciplínu.

Role bibliometrie v otevřené vědě bude v následujících letech budou následující:

- porozumět výstupům a rozpoznávat mechanismy, které poskytují smysluplné a zodpovědné metriky a indikátory
- zajistit spravedlivé a srovnatelné prostředky pro hodnocení výzkumu
- zabránit nepříznivému vlivu stimulů, metrik a indikátorů
- neustále přemýšlejte o své vlastní roli v hodnocení výzkumu

Altmetrics

Použití altmetrik při vyhledávání informací a hodnocení výzkumu vyvolává otázku: Jak spolu souvisí altmetriky a tradiční měření citací? Předpovídají sdílení na sociálních sítích s následnými citacemi pro daný článek? Pokud je nalezena korelace, mohlo by to naznačovat, že altmetrika a citace měří, alespoň do určité míry, stejný jev a že altmetrie jsou pouze ranými indikátory této základní kvality.

Výsledky raného výzkumu naznačují, že altmetriky souvisí s počty citací a mohou být schopny zachytit vliv vědeckých publikací na širší a odlišnou část jejich čtenářské základny než samotné počty citací, které odrážejí pouze chování autorů při publikování.

Subject Classification in Bibliometrics – Theory, Approaches & Limitations

Klasifikace v bibliometrii jsou silně orientovány na dominantní zdroje bibliometrických dat (WoS, SCOPUS)

- Klasifikace WoS (Clarivate)
 - 250 předmětových kategorií (SC)
 - Časopisy jsou řazeny do těchto kategorií. Některé časopisy mají více SC.
- Klasifikace podle SCOPUS (Elsevier)
 - 334 popisů (27 určují hlavní kategorie)

Harmonizace klasifikací naráží na problém rozsahu jednotlivých klasifikací a jejich slovníky (např. Article Abstracts vs. Patent Abstracts).

Subject Normalization for Citation Analysis

Publikační a citační chování vědců se liší napříč vědeckou komunitou, přičemž se také můžou postupem času měnit komunikační vzorce a pokrytí databáze, v důsledku čehož nejsou scientometrické ukazatele často srovnatelné napříč obory a po delší časové období. Toto má vliv na citační indikátory obecně, zejména na impakt faktor časopisů.

Nejdůležitější měřitelné faktory ovlivňující citační dopad:

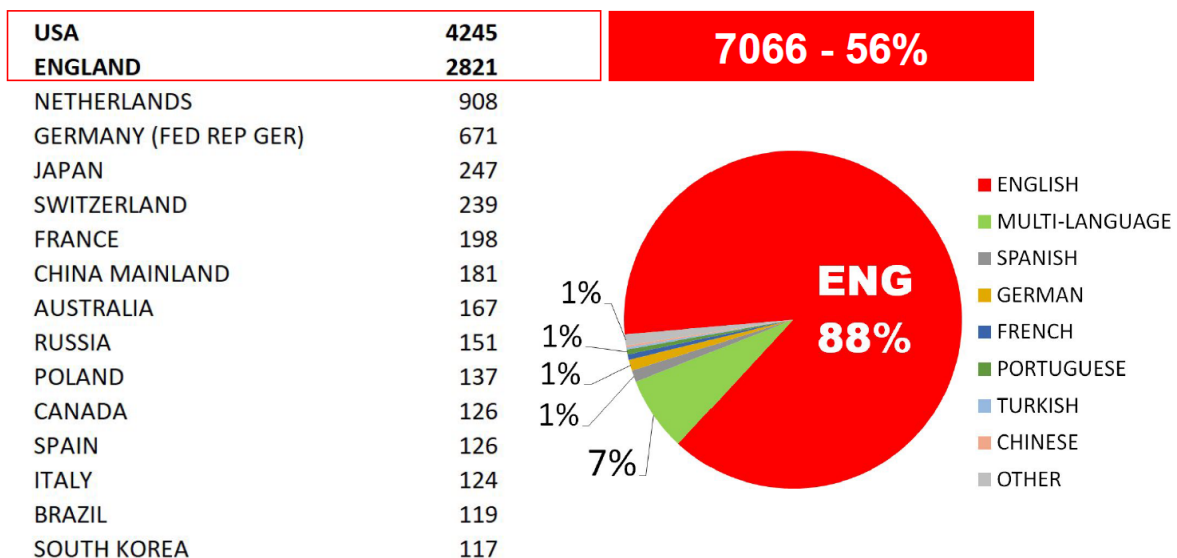
1. obor a v rámci oboru „úroveň abstrakce“
2. stáří článku
3. „sociální status“ status článku (prestiž spoluautorů a časopisu)
4. typ dokumentu
5. období pozorování

Normalizace může probíhat pomocí SNIP (poměr RIP -*Raw Impact per Paper* -/RDCP -*Relative Database Citation Potential*).

Journal Impact Measures

Hlavním problémem IF je jeho použití na individuální úrovni pro evaluaci jednotlivých autorů, což vede ke změnám lokálních cílů výzkumu ve snaze psát témata, která by mohla být lépe přijata mezinárodní vědeckou komunitou. 88% časopisů s IF je v angličtině.

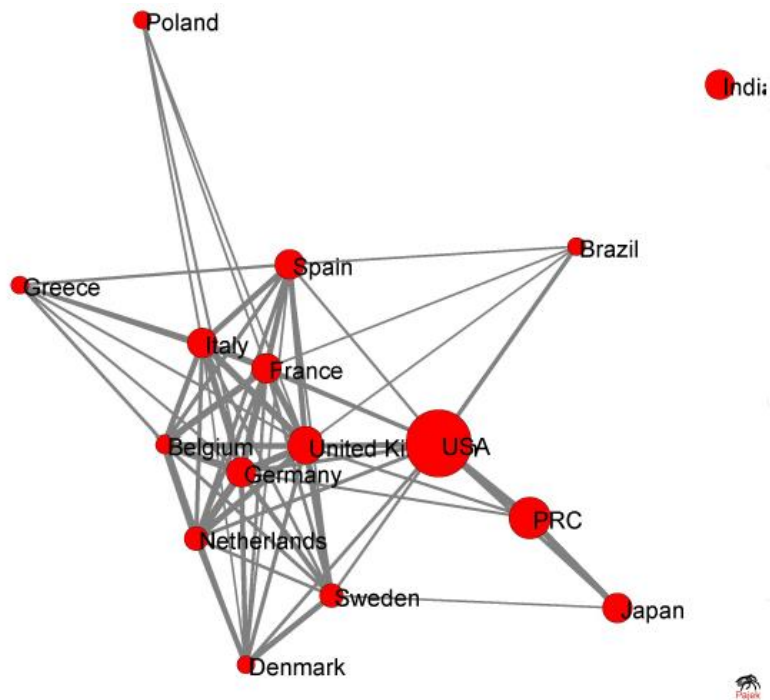
Countries and Languages coverage (IF Journals)



The Application of Network Analysis in Science Studies: Common Theoretical Background for Broad Applications

Síťová analýza je v kontextu studia vědy resp. bibliometrie použitím výkonných statistických nástrojů a analytických technik k odhalení struktury nebo vývoje vědy založené na vztazích mezi konkrétními entitami. Může být aplikována na všechny entity spojené s vědou, ať už publikace, obory, časopisy, instituce nebo výzkumné pracovníky, které jsou nějak propojeny prostřednictvím jakéhokoli druhu vztahu, jako je spolupráce, citace nebo společný výskyt.

S největší pravděpodobností jsou výsledky vykresleny v2D nebo 3D znázornění (mapě). Proto je postup často označován jako „mapování vědy“.



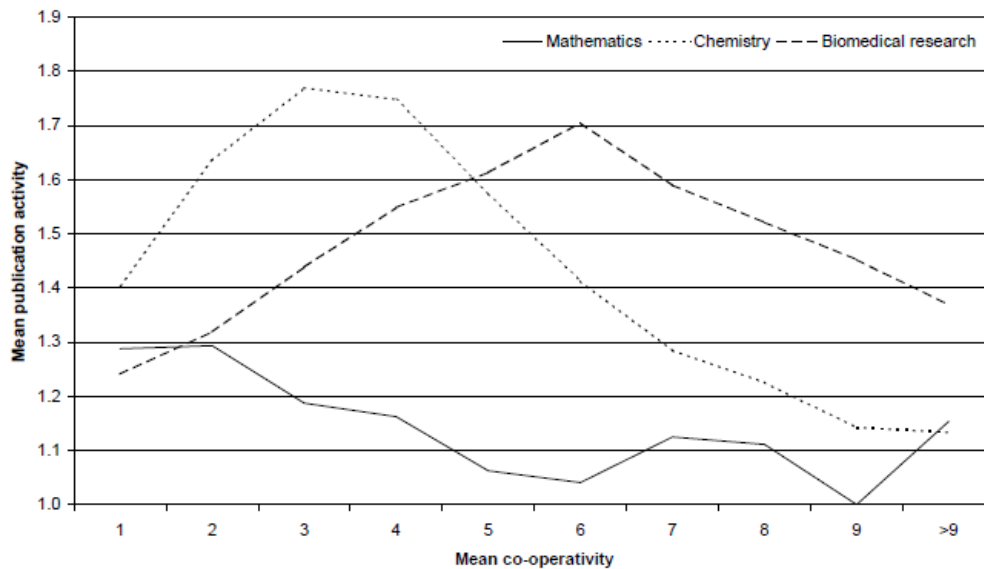
International collaboration in the field of environmental sciences
(Glänzel & Thijs, 2011, ISSI Newsletter)

Důležitost jednotlivých bodů mapy se hodnotí jejich vzdáleností od ostatních (closeness/farness), blízkostí ke středu (centrality) a počtem vazeb (degree) počtem vazeb, které zprostředkovávají ostatním bodům (betweenness).

Research Collaboration Measured by Co-Authorship

V novějších studiích o vědecké spolupráci napříč vědními obory bylo prokázáno, že „týmová práce“ vykazuje vyšší produktivitu než autorství jednotlivců, ale za prahem charakteristickým pro daný obor, produktivita s rostoucí spoluprací výrazně klesá.

Plot of average productivity vs. average co-operativity



GLÄNZEL, *Library Trends*, 2001

Dále vyvstává otázka jak spoluautorství ovlivňují mechanismy financování vědy, které mohou vést ke strategické volbě partnerů pro výzkum. Čestné autorství a hyperautorství nebo dokonce „povinné“ spoluautorství jsou projevem takovýchto strategií.

Praktická část

Praktická část letní školy spočívala v práci s daty (top publikace, potenciální partneři, OA/hybrid publikace, kopublikace) buď z University of Almeria (Španělsko) nebo University of Augsburg (Německo). První cvičení spočívalo v získávání a čištění dat z WoS (jak zvolit správné rejstříky, jak položit dotaz, jak extrahovat výsledky, atd.).

< BACK TO BASIC SEARCHES

Advanced Search Query Builder

Search in: **Web of Science Core Collection** ▾ Editions: **3 selected** ▾

Add terms to the query search preview

All Fields ▾ Example: liver disease india singh Add to query

More options ▲ Search Help

Query Preview

Enter or edit your query here. You can also combine previous searches e.g. #5 AND #2

+ Add date range Field Tags ▾ X Clear Search

Booleans : AND, OR, NOT

Field Tags :

TS=Topic	SO=[Publication Titles]	SA=Street Address	SU=Research Area	PUBL=Publisher
TI=Title	DO=DOI	CI=City	WC=Web of Science Categories	ALL=All Fields
AB=Abstract	PY=Year Published	PS=Province/State	IS= ISSN/ISBN	FPY=Final publication year
AU=Author	CF=Conference	CU=Country/Region	UT=Accession Number	
AI=Author Identifiers	AD=Address	ZP=Zip/Postal Code	PMID=PubMed ID	
AK=Author Keywords	OG=[Affiliation]	FO=Funding Agency	DOP=Publication Date	
GP=[Group Author]	OO=Organization	FG=Grant Number		
ED=Editor	SG=Suborganization	FD=Funding Details		
KP=Keyword Plus *		FT=Funding Text		

History

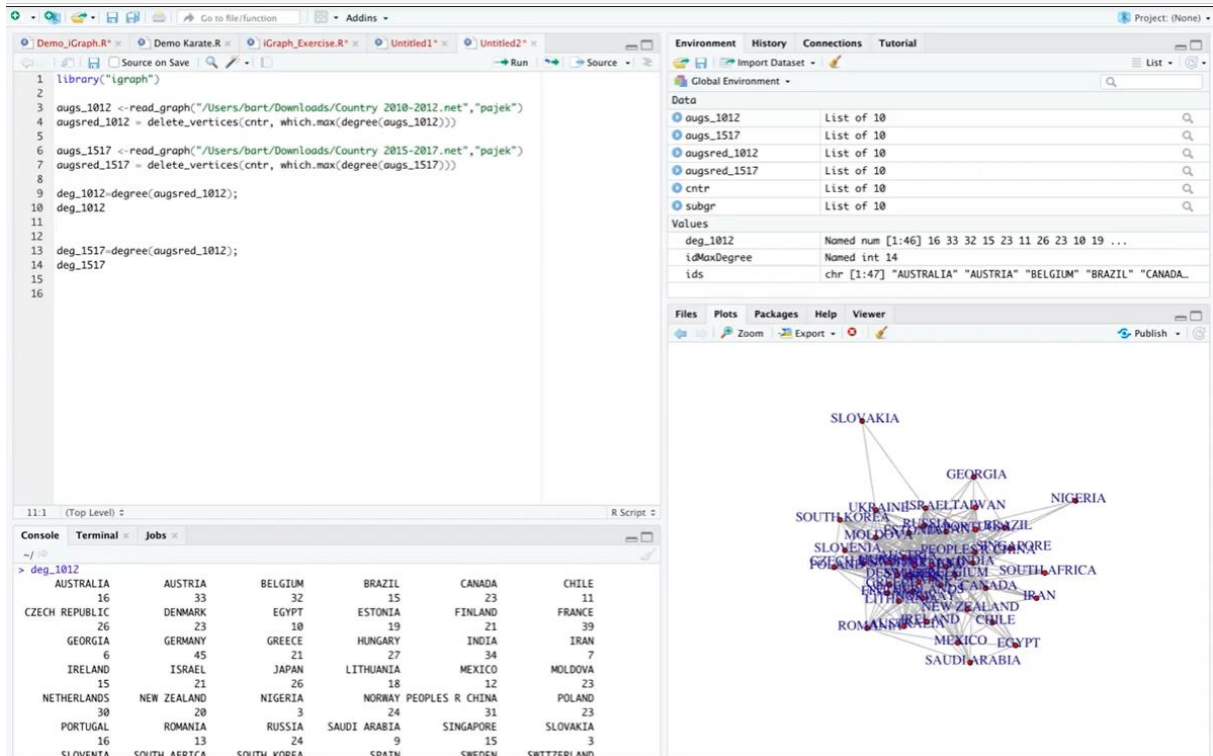
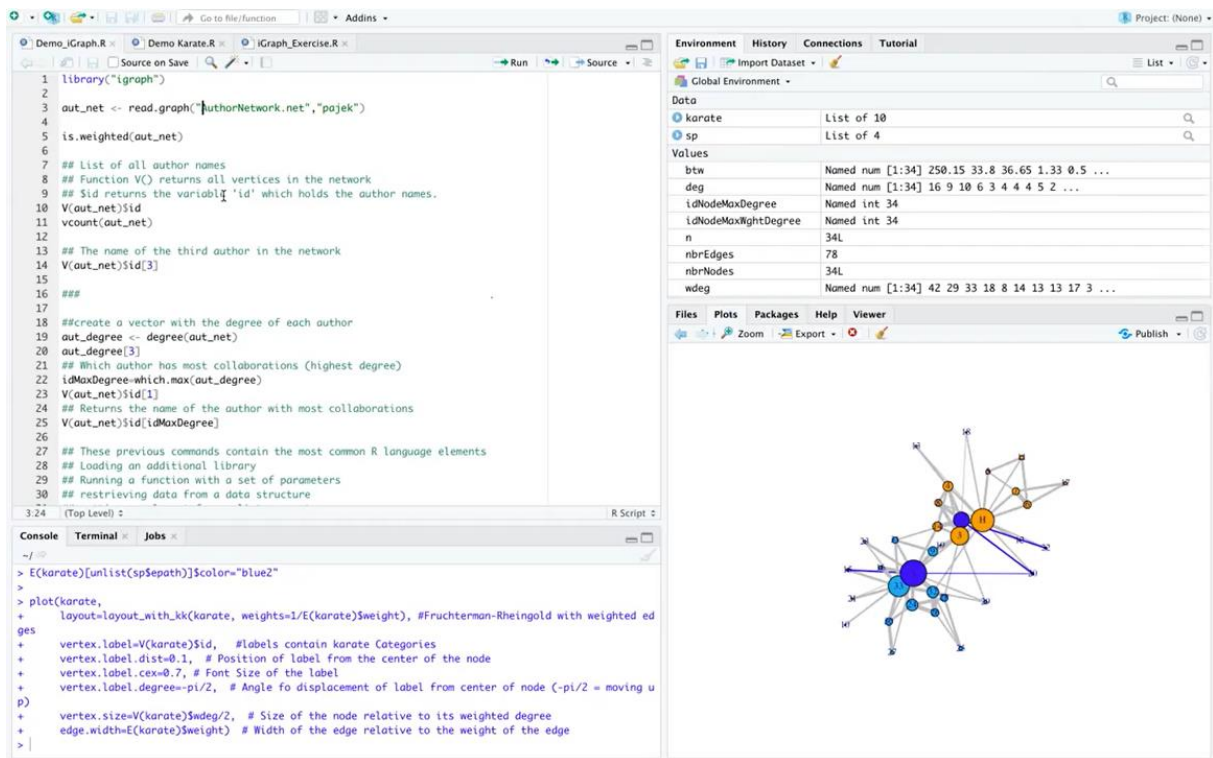
4 AD={ (UNIV AUGSBURG* SAME Germany) OR (WISSENSCH ZENTRUM UMWELT SAME AUGSBURG) OR ((AUGSBURG UNIV* NOT AUGSBURG UNIV APPL*) SAME Germany) OR ((LEHRSTUHL NOT MUNICH) SAME AUGSBURG)) AND Edit Add to Search 79

Nad daty jsme pracovali ve skupinách a následně prezentovali výsledky analýzy ostatním účastníkům letní školy.

	A	B	C	D	E	F	G
1	Subject Categories involved						
2							
3	Total Publication	Gold/ Gold Hybrid			Total Publication	Total	
4	Web of Science Categories	Record Count	% of 364		Web of Science Categories	Record Count	% of 2,741
5	Materials Science Multidisciplinary	70	19.231		Physics Applied	606	22.109
6	Physics Applied	61	16.758		Physics Condensed Matter	537	19.591
7	Multidisciplinary Sciences	58	15.934		Materials Science Multidisciplinary	534	19.482
8	Physics Multidisciplinary	53	14.56		Physics Multidisciplinary	251	9.157
9	Physics Condensed Matter	45	12.363		Chemistry Physical	178	6.494
10	Meteorology Atmospheric Sciences	39	10.714		Mathematics Applied	145	5.29
11	Environmental Sciences	30	8.242		Mathematics	123	4.487
12	Nanoscience Nanotechnology	23	6.319		Chemistry Multidisciplinary	112	4.086
13	Chemistry Multidisciplinary	18	4.945		Nanoscience Nanotechnology	107	3.904
14	Geosciences Multidisciplinary	18	4.945		Environmental Sciences	105	3.831
15	Chemistry Physical	17	4.67		Meteorology Atmospheric Sciences	103	3.758
16	Water Resources	17	4.67		Chemistry Inorganic Nuclear	100	3.648
17	Chemistry Inorganic Nuclear	13	3.571		Physics Mathematical	98	3.575
18	Public Environmental Occupational Health	10	2.747		Economics	89	3.247
19	Toxicology	10	2.747		Physics Atomic Molecular Chemical	89	3.247
20	Mathematics	8	2.198		Management	88	3.211
21	Optics	7	1.923		Multidisciplinary Sciences	77	2.809
22	Energy Fuels	6	1.648		Physics Fluids Plasmas	71	2.59
23	Psychology Multidisciplinary	6	1.648		Operations Research Management Science	63	2.298
24	Mathematics Applied	5	1.374		Computer Science Information Systems	60	2.189
25	Statistics Probability	5	1.374		Optics	58	2.116
26	Crystallography	4	1.099		Computer Science Theory Methods	55	2.007
27	Mathematics Interdisciplinary Applications	4	1.099		Computer Science Software Engineering	51	1.861
28	Metallurgy Metallurgical Engineering	4	1.099		Business	48	1.751
29	Chemistry Analytical	3	0.824		Geosciences Multidisciplinary	44	1.605
30							
31							
32							
33	PP Top 10 (Gold / Gold-hybrid)				PP Top 10 (Total publication)		
34	Web of Science Categories	Record Count	% of 67		Web of Science Categories	Record Count	% of 392
35	Multidisciplinary Sciences	14	20.896		Physics Applied	65	16.582
36	Environmental Sciences	11	16.418		Physics Multidisciplinary	64	16.327
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
74							
75							
76							
77							
78							
79							
80							
81							
82							
83							
84							
85							
86							
87							
88							
89							
90							
91							
92							
93							
94							
95							
96							
97							
98							
99							
100							

Dalším krokem bylo (po čištění extrahovaných dat) analyzovat výsledky prostřednictvím nástroje iGraph (<https://igraph.org/>), kde pomocí R nadefinujeme, co se má s daty dělat. iGraph je schopen efektivně zpracovávat velké sítě, lze jej produktivně použít s programovacím jazykem (kromě R také Python, Mathematica a C/C++) na vysoké úrovni a je podporováno interaktivní i neinteraktivní použitím.

K práci s nástrojem je potřeba stáhnout balíček korespondující ke zvolenému programovacímu jazyku. iGraph umí číst a zapisovat soubory Pajek a GraphML



Děkuji Asociaci knihoven vysokých škol České republiky, že jsem se díky její finanční podpoře mohl esss2021 zúčastnit.