

Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Katedra informačního inženýrství



Diplomová práce

Analýza vědeckých publikačních činností

Bc. Petr Klapka

© 2020 ČZU v Praze

ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Petr Klapka

Systémové inženýrství a informatika
Informatika

Název práce

Analýza vědeckých publikačních činností

Název anglicky

Analysis of scientific publications

Cíle práce

Hlavním cílem práce je analýza vědeckých publikačních činností na základě bibliografických citací a bibliografických referencí pomocí metody, která kvantifikuje vztahy mezi samotnými autory, dokumenty a vědeckými publikacemi.

Výstupem práce je interaktivní vizualizace bibliografické sítě, která může sloužit ke kontrole plagiátorství a citování mezi jednotlivými autory.

Metodika

Teoretická část práce bude obsahovat rešerši na základě analýzy důvěryhodných zdrojů. Nejprve poskytnete definice a příklady

využití základních metrických pojmů v oblasti informační vědy. V druhé části se bude zabývat bibliometrickými zákony.

Praktická část bude zpracována na základě sběru veřejně dostupných dat z databáze Scopus a Google Scholar, která budou dále zpracována a přenesena do grafické interaktivní vizualizace pomocí volně dostupného softwaru Gephi.

Doporučený rozsah práce

80 – 140 stran

Klíčová slova

Bibliografie, Citační analýza, Citační databáze, Grafická vizualizace, Scientometrie

Doporučené zdroje informací

KÖNIGOVÁ, Marie. Úvod do bibliometrie. 1. vyd. Praha: Česká informační společnost, 1993.

ŠVEJDA, Jan. Bibliometrie. In: KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online]. Praha : Národní knihovna ČR, 2003- [cit. 2014-12-12]. Dostupné z: http://aleph.nkp.cz/F/?func=direct&doc_number=000000344&local_base=KTD.

Předběžný termín obhajoby

2020/21 ZS – PEF (únor 2021)

Vedoucí práce

doc. Ing. Vojtěch Merunka, Ph.D.

Garantující pracoviště

Katedra informačního inženýrství

Elektronicky schváleno dne 19. 11. 2020

Ing. Martin Pelikán, Ph.D.

Vedoucí katedry

Elektronicky schváleno dne 19. 11. 2020

Ing. Martin Pelikán, Ph.D.

Děkan

V Praze dne 29. 11. 2020

Čestné prohlášení

Prohlašuji, že svou diplomovou práci "Analýza vědeckých publikačních činností" jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne 30.11.2020

Poděkování

Rád bych touto cestou poděkoval vedoucímu této práce panu doc. Ing. Vojtěchovi Meruňkovi, Ph.D. za podněty a konzultace při zpracování této práce.

Analýza vědeckých publikačních činností

Abstrakt

Diplomová práce se zabývá analýzou a vizualizací vědeckých publikačních činností. Vazby publikací jsou vytvářeny pomocí citací. Jsou tak vytvářeny grafy, kde uzly reprezentují jednotlivé publikace a hrany vazby mezi nimi. Vzniká tak bibliografická síť, která je graficky vizualizována.

V teoretické části jsou vymezeny jednotlivé metriky informační vědy. Jedná se o bibliometrické zákony, citační rejstříky a samotnou citační analýzu. Dále jsou v teoretické části rozvedena témata, vztahující se k publikacím v citačních databázích, formátům dat a nástrojům pro vizualizaci grafické reprezentace vazeb. V práci je dále podrobný popis získávání zdrojových dat z citační databáze a transformace těchto dat do formátu vhodného pro grafickou reprezentaci.

Vlastní práce se týká výběru reálného vzorku dat z databáze Scopus, zpracováním těchto dat do formy vhodné pro vizualizační nástroj Gephi a následného importu do tohoto nástroje. V rámci Gephi je provedena další úprava dat vnitřními procedurami tohoto nástroje. Výsledkem je vizualizace vazeb dokumentů ve formě síťového grafu. Tento graf je výchozím stavem pro bibliografické zkoumání množiny vybraných publikací. V další části práce je pak definováno několik analytických bibliografických úloh. Samotné úlohy jsou poté řešeny filtrováním dat ve vizualizačním nástroji. Výsledky jsou zobrazeny v grafické podobě. V poslední části vlastní práce je zobrazena možnost práce se síťovým grafem prostřednictvím externího prohlížeče JavaScript GEXF Viewer, který pracuje s daty exportovanými z vizualizačního nástroje Gephi ve specifickém formátu GEXF.

Klíčová slova: Bibliografie, Citační analýza, Citační databáze, Grafická vizualizace, Scientometrie, Scopus, Gephi, Citace, Citační rejstřík, GEXF,

Analysis of scientific publications

Abstract

The diploma thesis deals with the analysis and visualization of scientific publishing activities. Bindings of publications are created according to citations. Graphs are thus created, where the nodes represent individual publications and the edges the link between them. This creates a bibliographic network that is graphically visualized.

The theoretical part defines the individual metrics of informatics science. These are bibliometric laws, citation registers and the citation analysis itself. Furthermore, the theoretical part deals with topics related to publications in citation databases, data formats and tools for visualizing of the graphical representation of links. The work also contains a detailed description of obtaining data source from the citation database and the transformation of this data into a format suitable for graphical representation.

The actual work concerns with the selection of a real sample of data from the Scopus database, processing this data into a form suitable for the visualization tool Gephi and subsequent import into this tool. Within Gephi, further data modification is performed by the internal procedures of this tool. The result is the visualization of document links in the form of a network graph. This graph is the starting point for a bibliographic examination of a set of selected publications. In the next part of the work, several analytical bibliographic tasks are defined. The tasks themselves are then solved by filtering the data in a visualization tool. The results are displayed in graphical form. The last part of the work shows the possibility of working with a network graph through an external JavaScript viewer GEXF Viewer, which works with exported data from visualisation tool Gephi in specific format GEXF.

Keywords: Bibliography, Citation analysis, Citation database, Graphic visualization, Scientometry, Scopus, Gephi, Citation, Citation index, GEXF

Obsah

1 Úvod.....	13
2 Cíl práce a metodika	14
2.1 Cíl práce	14
2.2 Metodika	14
3 Teoretická východiska	15
3.1 Metriky informační vědy.....	15
3.1.1 Kvantifikace knihovní činnosti.....	16
3.1.2 Bibliometrie	16
3.1.2.1 Bibliometrické zákony.....	17
3.1.3 Scientometrie	17
3.1.4 Webometrie.....	19
3.1.5 Infometrie.....	19
3.2 Citační analýza	20
3.2.1 Citace	20
3.2.2 Nástroje citační analýzy	20
3.3 Citační rejstříky	21
3.3.1 Web of Science	23
3.3.2 Scopus.....	23
3.3.2.1 Hledání v databázi SCOPUS	24
3.3.2.2 Uspořádání výpisu vybraných dokumentů	25
3.3.2.3 Export dat z databáze SCOPUS	26
3.3.3 Google Scholar	28
3.4 Nástroje pro analýzu a vizualizaci	29
3.4.1 VOSviewer.....	30
3.4.1.1 Vizualizace	30
3.4.1.2 Používané techniky.....	30
3.4.2 Cytoscape.....	31
3.4.3 Gephi.....	32
3.4.3.1 Import dat do Gephi.....	32
3.4.3.2 Objekty s již přiděleným id.	32
3.4.3.3 Objekty bez předem přiděleného id.....	38
3.5 Formát souboru GEXF	46
3.5.1 Záhlaví	48

3.5.2	Topologie sítě.....	49
3.5.3	Deklarace uzlu.....	49
3.5.4	Deklarace hrany	49
3.5.5	Data sítě.....	50
3.5.6	Deklarace atributů	50
3.6	Externí prohlížeče grafů	52
4	Vlastní práce	53
4.1	Zdrojová data.....	53
4.1.1	Export dat z databáze	53
4.1.2	Import dat do Gephi	54
4.1.3	Úprava dat	54
4.2	Vizualizace	56
4.2.1	Importovaná data.....	56
4.2.2	Yifan Hu zobrazení	57
4.2.3	Force Atlas zobrazení.....	58
4.2.4	Fruchterman Reingoldovo zobrazení	59
4.3	Analýza.....	62
4.3.1	Nalezení nejkratší cesty mezi dvěma uzly.	62
4.3.2	Rozdělení dokumentů dle jejich typu a citací.	63
4.3.3	Hodnocení časopisů dle počtu citovaných článků	64
4.3.4	Dokumenty dle počtu citací.	65
4.4	Zobrazení grafu v externím prohlížeči	66
4.4.1	Interaktivní zobrazení	66
4.4.2	Zobrazení komplexní informace vybraného uzlu.	67
4.4.3	Vyhledávání	69
	Výsledky a diskuse	70
5	Závěr.....	71
6	Seznam použitých zdrojů.....	72
7	Přílohy	74

Seznam obrázků

Obrázek 1	- Průnik jednotlivých metrik.....	15
Obrázek 2	- Prostředí Gephi.....	33
Obrázek 3	- Import dat do prostředí Gephi	34
Obrázek 4	- Možnosti importu tabulky uzlů do prostředí Gephi	34
Obrázek 5	- Nastavení importu a volba datového typu.....	35
Obrázek 6	- Import tabulky hran do prostředí Gephi.....	36
Obrázek 7	- Tabulka uzlů v prostředí Gephi.....	36

Obrázek 8 – Náhled grafu příkladu.....	37
Obrázek 9 - Graf s popisky uzlů	38
Obrázek 10 - Import zdrojových dat.....	40
Obrázek 11 - Nastavení importu zdrojových dat	41
Obrázek 12 - Laboratoř dat v prostředí Gephi	41
Obrázek 13 - Graf izolovaných uzlů.....	42
Obrázek 14 - Vytvoření odkazů na uzle	42
Obrázek 15 - Tvorba hran v prostředí Gephi	43
Obrázek 16 - Graf propojených uzlů pomocí hran	43
Obrázek 17 – Vizualizace fiktivní sítě s kopií uzlu	44
Obrázek 18 - Odstranění kopií uzlů.....	45
Obrázek 19 - Vizualizace očištěné fiktivní sítě	46
Obrázek 20 - Jednoduchý graf v GEXF.....	46
Obrázek 21 - Orientovaný graf v GEXF.....	47
Obrázek 22 - Příklad grafu v GEXF	51
Obrázek 23 - Úprava dat.....	54
Obrázek 24 - Nahrazení dat	54
Obrázek 25 - Vytvoření sloupce a spojení hodnot s oddělovačem.....	55
Obrázek 26 - Spojení hodnot do jednoho sloupce	55
Obrázek 27 - Náhodná vizualizace grafu.....	56
Obrázek 28 - Yifan Hu zobrazení	57
Obrázek 29 - Force Atlas zobrazení	58
Obrázek 30 - Fruchterman Reingoldovo zobrazení	59
Obrázek 31 – Označení zdrojových dat (uzlů)	60
Obrázek 32 - Úprava zdrojových uzlů.....	60
Obrázek 33 - Zvýraznění zdrojových uzlů	61
Obrázek 34 - Zvýrazněné zdrojové uzle v grafu.....	61
Obrázek 35 - Nejkratší cesta mezi zvolenými uzly	62
Obrázek 36 - Graf rozdělení na základě typu publikace.....	63
Obrázek 37 - Graf zobrazující množství citování jednotlivých časopisů	64
Obrázek 38 - Graf velikosti uzlů na základě citovanosti publikace	65
Obrázek 39 - Graf citovanosti neznámých autorů	66
Obrázek 40 - Interaktivní zobrazení grafu GEXF	67
Obrázek 41 - Zobrazení informací o zvoleném uzlu	68
Obrázek 42 - Vyhledávání uzlů	69

Seznam tabulek

Tabulka 1 - Základní přehled citačních rejstříků	22
Tabulka 2 - Tabulka hran.....	32
Tabulka 3 - Tabulka uzlů	32
Tabulka 4 - Základní publikace	38
Tabulka 5 - Tabulka citovaných publikací	39
Tabulka 6 - Základní publikace	39
Tabulka 7 - Tabulka citovaných publikací	39
Tabulka 8 - Základní publikace	39
Tabulka 9 - Tabulka citovaných publikací	39

Seznam použitých zkratk

AHCI - Arts and Humanities Citation Index

ASEP - databáze výsledků vědecké práce v AV ČR

CSV - comma-separated values

CWTS - Centre for Science and Technology Studies, University Leiden

DOI - digital object identifier

GEXF - Graph Exchange XML Format

ISBN - International Standard Book Number

ISI - Institute for Scientific Information

JS - JavaScript

MIT - Massachusetts Institute of Technology

RIV . Registr informací o výsledcích

WWW - World Wide Web

SCI - Science Citation Index

SSCI - Social Science Citation Index

XML - Extensible Markup Language

1 Úvod

Každý rok je zveřejněna spousta publikací v rámci vědeckých činností. Tyto publikace obsahují výsledky prací jednotlivých autorů z různých vědeckých oborů. To přispívá k velkému rozvoji vědy jako takové.

Ze všech oborů lidské činnosti proudí v současné době nepřetržitý tok nových informací. Díky internetu a dalším možnostem informačních technologií se tento proces neustále zrychluje. Vycházejí nové specializované časopisy, vydávají se odborné knihy, spousta informačních zdrojů vzniká na akademické půdě, ve výzkumných institucích a rovněž vznikají i kvalitní díla v prostředí velikých světových firem. Pokud je třeba sledovat vývoj v určitém oboru, je téměř nemožné dosáhnout toho manuálně. Proto byly vyvinuty specializované databáze, které jsou plněny informacemi o nových bibliografických příspěvcích. Vývoje a implementace těchto nástrojů se zhostily některé univerzity, vědecká pracoviště a některá vydavatelství. Služby jsou obvykle poskytovány komerčně, avšak existuje řada možností přístupu zdarma např. pro studenty, školy apod. Ukládání do databází si vynutilo i jistý stupeň normalizace záznamů o jednotlivých příspěvcích.

Postupem vývoje se zjistilo, že důležitým informačním zdrojem je sledování citací jednotlivých materiálů. Vznikly proto specializované citační databáze, které provazují záznamy na základě citací. Hledání informace se pak skládá z prohlížení pouze základních charakteristik publikací, tj. autora, názvu, abstraktu, klíčových slov a referencí na další publikace. S ohledem na množství vydaných příspěvků v daném oboru, je počet citací jedním z kvalitativních ukazatelů. Přírozeným výstupem z databáze jsou tabulky. Pokud je však rozsah informace značně velký, je práce s tabulkami dosti nepřehledná, vyčerpávající a pomalu konvergující k požadovanému výsledku. Proto vznikly speciální grafické nástroje, které umožňují vizualizaci vazeb mezi publikacemi a jsou schopny i zobrazit v detailu i další důležité informace.

Tato diplomová práce je zaměřena na popis zpracování bibliografických dat exportovaných z databáze a dále zpracovávaných ve vybraném grafickém nástroji.

2 Cíl práce a metodika

2.1 Cíl práce

Hlavním cílem práce je analýza vědeckých publikačních činností na základě bibliografických citací a bibliografických referencí pomocí metody, která kvantifikuje vztahy mezi samotnými autory, dokumenty a vědními dokumenty.

Výstupem práce je interaktivní vizualizace bibliografické sítě.

2.2 Metodika

Teoretická část práce bude vypracována na základě studia důvěryhodných zdrojů z internetu a odborné literatury. Nejprve poskytneme definice a příklady využití základních metrických pojmů v oblasti informační vědy. V druhé části se bude zabývat bibliometrickými zákony.

Praktická část bude zpracována na základě sběru veřejně dostupných dat z databáze Scopus a Google Scholar, která budou dále zpracována a přenesena do grafické interaktivní vizualizace.

3 Teoretická východiska

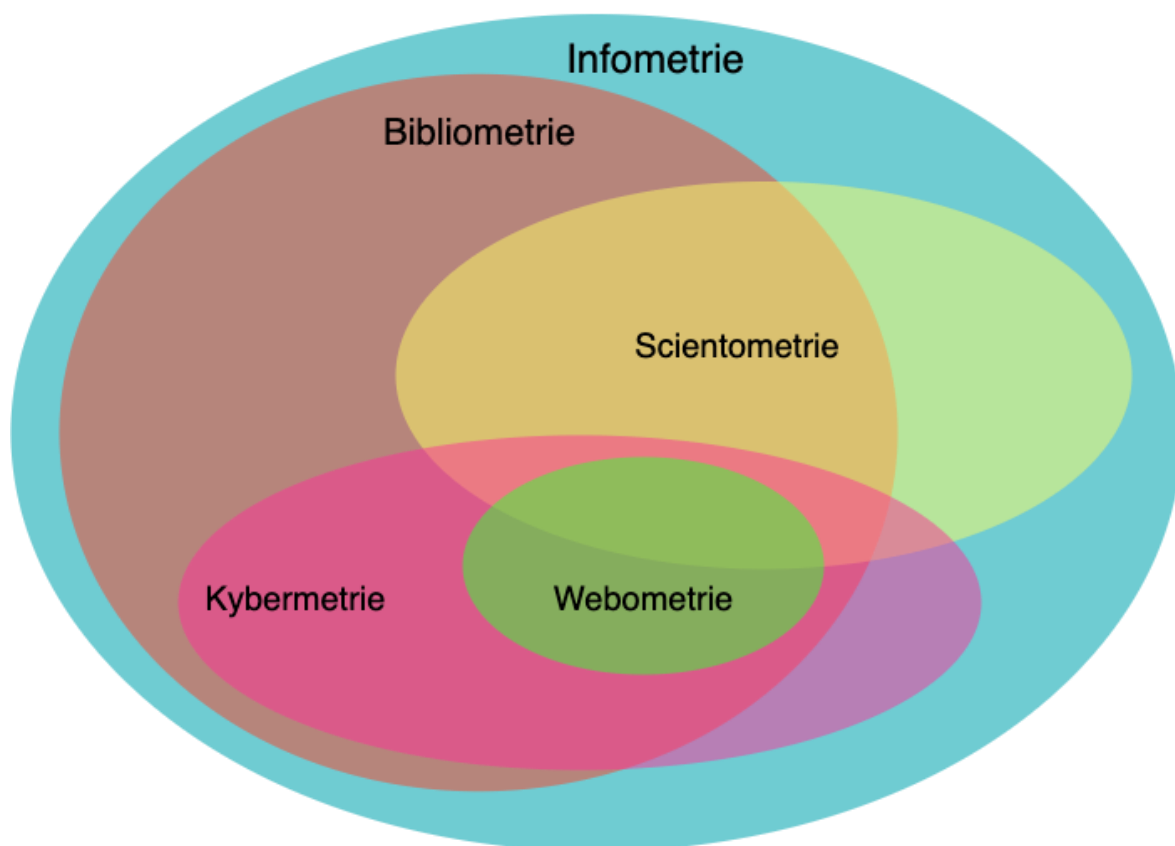
3.1 Metriky informační vědy

V posledních několika desetiletích v rámci oblasti knihovnictví a informační vědy bylo vyvinuto několik kvantitativních metod pro zkoumání informační vědy, které dnes známe pod slovem metrika (pochází z řeckého nebo latinského názvu „Metricus“).

V rámci tohoto rozvoje sehráli velkou roli akademici z různých odvětví. Mezi ně můžeme zařadit akademiky z oblasti chemie např. Tibora Brauna, dále pak z oblasti filozofie Vasilye Nalimova, kteří přišli s důležitými koncepty zkoumání.

V dnešní době se můžeme setkat s různými metrikami jako je např. knihometrie, webometrie a spoustou dalších, které však úzce souvisí zejména s bibliometrií a scientometrií. [1]

Obrázek 1 - Průnik jednotlivých metrik



Zdroj: Vlastní zpracování

3.1.1 Kvantifikace knihovní činnosti

Jedná se o nejstarší disciplínu, která vznikla v roce 1948 jako obecný popis pro aplikaci matematické a statistické techniky pro řešení problémů v knihovnách.

Ranganathan ji definoval jako kvantitativní analýzu různých aspektů v knihovních činnostech a knihovních dokumentech aplikováním matematických a statistických výpočtů k hledání řešení v knihovních problémech. Ranganathan použil tuto metodu například k uspořádání pořadí knih v archivech, tak aby každá další nejčastěji hledaná kniha byla co nejbližší u vchodu a méně hledané co nejdále, tím bylo zajištěno, že zaměstnanci knihovny nebudou muset chodit příliš daleko pro nejžádanější knihy.

Tuto metodu používal taktéž pro různé další úkoly jako například analýzu požadavků uživatele, oběh a uspořádání knih a periodické optimalizace pracovní zátěže jednotlivých zaměstnanců, dále také k alokaci rozpočtu a plánování rozvržení jednotlivých knihoven. [1]

3.1.2 Bibliometrie

Bibliometrie je založena na základě analýzy dokumentů vznikajících v rámci vědeckých publikačních činností. Jedná se zejména o dokumenty, které vychází z této vědní disciplíny, a tyto dokumenty odráží stav vědeckého zkoumání.

Poprvé se tento termín objevil v roce 1969. Do té doby byl používán termín statistická bibliografie, který však mohl být interpretován jako bibliografie o statistice.

Alan Prichard definoval bibliometrii následovně:

„Cílem bibliometrie je osvětlit proces psané komunikace a zaměření vědecké disciplíny (do té míry, jak se odráží v psané komunikaci) pomocí kvantitativní analýzy rozličných složek psané komunikace.“ [2]

Tuto definici můžeme charakterizovat jako průnik matematicko-statistických metod a informační vědy, který nám umožňuje vyhodnocovat údaje v oblasti vědeckých informací jako např. hodnocení vědeckého přínosu různých zemí, autorů, článků, oborů, jazyků apod.

Bibliometrie zkoumá různé oblasti jako je například analýza publikační činnosti vědců, zkoumání jednotlivých citačních vazeb, zkoumání mezinárodních publikačních trendů, vyhledávání informací, mapování vědy a mnohé další.

3.1.2.1 Bibliometrické zákony

Bibliometrické zákony nám slouží k popisu obecních zákonitostí jevů a jsou založeny na matematicko-statistických metodách.

Bradfordův zákon nám popisuje umístění jednotlivých publikací v určité vědní oblasti. Pomocí tohoto zákona lze stanovit závislost mezi rozložením publikací týkajících se konkrétního vědeckého oboru a celkovým počtem publikací. Z tohoto zákona vyplývá, že v minimálním množství časopisů tvořících základ oboru je umístěn maximální počet relevantních článků. [3]

Zipfův zákon nám popisuje distribuci jednotlivých slov v textu. Jednotlivá slova mohou být použita několikrát v textu, mnoho slov může být použito pouze jednou. Princip tohoto zákona byl realizován v rámci Zipfových výzkumů, kdy po prvé u knihy Odysseus Zipf použil frekvenci slov. Jednotlivá slova uspořádal podle jejich výskytu a ohodnotil je tzv. rankem, poté provedl součin frekvence slova a jeho ranku a zjistil, že tyto hodnoty jsou velmi blízké c , kdy $c = r \cdot f$. Této znalosti se využívá dodnes například při tvorbě frekvenčního slovníku. [4]

Lotkův zákon nám říká, že pouze malé množství produktivních autorů nám publikuje největší počet publikací v konkrétním oboru. Jako příklad, lze uvést, že pokud v konkrétním oboru publikuje celkem 100 autorů, tak 25 z nich má v tomto oboru publikovány 2 práce, 11 autorů 3 práce a 6 autorů práce čtyři. Pomocí tohoto zákona můžeme předem určit, kolika autory bylo publikováno více článků, pokud však známe celkový počet autorů, kteří publikovali pouze jeden článek a lze tedy určit produktivitu vědecké práce. [5]

3.1.3 Scientometrie

Scientometrie se zaměřuje na měření vědy, tedy růstem, strukturou, vzájemnými vztahy a produktivitou jednotlivých vědeckých disciplín. Na rozdíl od bibliometrie (3.1.1) se scientometrie neomezuje pouze na písemnou komunikaci, ale může zahrnovat i různé další zdroje jako je analýza množství doktorandů, financování institucí nebo další.

Scientometrie se dnes nejvíce využívá pro rozdělování financí jednotlivým institucím a hodnocením vědy. Existuje spousta další možností, jak v rámci vědy měřit množství a kvalitu. Scientometrie se tímto zabývá pomocí nalézání různých metrik a metod, kterými můžeme měřit výkonnost vědy. Tato měření jsou odvozena od interakcí v rámci komunikace ve vědě. Základní interakcí je citace.

Je však zapotřebí si uvědomit, že písemná komunikace analyzována pomocí scientometrie je omezena pouze na vědeckou literaturu, a proto mnoho studií provedených touto metodou by také mohlo být definováno jako bibliometrie. Většina těchto výstupů je zveřejněna v literární podobě. Lze tedy říct, že scientometrie je ve skutečnosti také bibliometrie. [6]

Scientometrie vznikla v 50. letech 20. století, kdy Eugene Garfield při realizaci expertních systémů založených na indexování vědeckých informací a citací zjistil, že indexovaná data mohou sloužit ke sledování růstu vědy. Následně pak byl u zrodu základů jedné z nejrozsáhlejších citačních databází současnosti, a to ISI Web of Science. Na principu této databáze vznikla spousta dalších jako je například databáze Scopus (3.4.1) , PubMed a u nás například ASEP, RIV apod. [7]

V rámci scientometrie byla vyvinuta spousta dalších statistických nástrojů pro výpočet kvantitativních aspektů jednotlivých publikací, jako jsou citační indexy, impakt faktor, H-index a další, kterými lze měřit některé kvantitativní aspekty práce jednotlivců, celých týmů, vědeckých oborů nebo celých zemí.

Například citační index autora nám určuje množství citací připadající na jeho autorské práce za určité období a čím častěji je autor citován jinými autory, tím je tento index vyšší. Naopak impakt faktor nám na základě časopisu jako celku umožňuje vyhodnotit a srovnávat vědecké časopisy. Výše faktoru časopisu, ve kterém je autorův článek zveřejněn nám určuje kvalitu autorova článku a zde platí, čím vyšší je tento faktor, tím lépe je jeho práce ohodnocena v rámci hodnocení vědecké publikační činnosti.

H-index neboli Hirschův index je založen na kvantitativním hodnocení kvality publikovaných článků jedním autorem, skupinou autorů nebo institucí. Umožňuje odstranit nevyváženou účast autora na různých člancích ať už z pohledu hlavního autora na méně citovaných člancích, nebo drobného přispěvovatele na velmi citovaném článku a ohodnotit průměrný podíl na citačním indexu jeho článků. [7]

3.1.4 Webometrie

Webometrie byla poprvé použita v roce 1997 T.C.Almindem a Peterem Ingwersenem. Tato vědecká disciplína za pomoci měření zkoumá World Wide Web. V počátcích byla tato disciplína taktéž nazývaná jako webová bibliometrie a internetometrie. Dnes je již stanoven oficiální termín webometrie, přestože hlavní magazín se stále jmenuje Kybermetrie. V roce 2004 Björneborn definoval webometrii jako:

„studium kvantitativních aspektů vytváření a užití informačních zdrojů, struktur a technologií celého internetu pomocí informetrických a bibliometrických metod“. [8]

Pomocí této metody můžeme provádět analýzu obsahu, analýzu struktury odkazů, analýzu využití (převážně pro vyhledávače) a analýzu webových technologií (včetně výkonu vyhledávače).

Kybermetrie úzce souvisí s informatikou a sdílením zájmů v oblastech kybernetické kartografie, webové archeologie, počítačové geografie, ekologie webu, těžení webu, analýzy webových grafů, dynamiky a inteligence webu. Na druhé straně webometrie je rozšířením klasických bibliometrických a informetických metod pro internet. Porovnávání odkazů s citacemi a referencemi, kdy jednotlivé citace jsou porovnávány s odkazy směřujícími na stránku a jednotlivé reference s odkazy směřující ze stránky. [9]

Webometrie se potkává s různými problémy, které souvisejí s webovou stránkou a nejsou obsaženi v rámci klasické bibliografie. Můžeme mezi ně zařadit například dezinformace, spolehlivost údajů, platnost a kvalitu různých tvrzení.

3.1.5 Infometrie

Tato vědní disciplína se zabývá obsaženým množstvím informací v dokumentu a změnami tohoto množství informací během informačního procesu, kde může docházet k různým transformacím. Jedná se o nejšířších ze všech zmíněných metrik, která zahrnuje všechny aspekty ze všech přechozích oblastí.

V rámci infometrie můžeme rozlišovat následující oblasti informatiky, jako je citační analýza, analýza souvisejících slov, studie růstu v čase apod. Jedním z hlavních problémů infometrie může být blokování prohledávání jednotlivých webů pomocí robotů, či webové stránky, které vyžadují autentizaci pro přístup na web. [10]

3.2 Citační analýza

Citační analýza nám hodnotí kvalitu a význam vědeckých publikací. Zakládá se na sledování míry četnosti jejich citování. Pomocí aplikace této analýzy můžeme mapovat a konstruovat citačních sít, optimalizovat informační toky a profilovat knihovní fondy. Umožňuje nám získat jméno autora konkrétní publikace a pomocí jednotlivých citačních odkazů mapovat ostatní autory, kteří na tento článek odkazují ve svých publikacích. Pomocí statistických údajů můžeme odvodit kvalitu článku pomocí množství citací a také jednotlivý časopis, ve kterém byl tento článek publikován.

Množství citací konkrétní publikace může být také označováno během určitého časového období pod pojmem citovanost. Tato míra nám ve většině případů určuje přínos jednotlivé publikace. Ve většině případů lze zjistit, jaké zdroje jsou v publikaci citovány a které se často nachází blízko sebe.

Pomocí výsledků citační analýzy můžeme modelovat vědecký rozvoj, zaměření výzkumů, nové technologie. [11]

3.2.1 Citace

Citaci můžeme definovat jako formalizovaný odkaz na konkrétní publikaci, která pochází z jiné publikace. Celkový počet těchto citací jedné publikace v určitém období označujeme za citovanost dokumentu. Dále se společně s tímto termínem můžeme setkat s pojmem autocitace, kdy autor cituje v rámci své publikace své vlastní publikace nebo publikace, u kterých je uveden jako spoluautor. Stejně tak se jedná o autora, který je citován některým z jeho kolegů, kteří jsou s ním uvedeni v díle jako spoluautoři.

3.2.2 Nástroje citační analýzy

Nástrojem citační analýzy je citační rejstřík nebo také citační databáze. Mezi největší a nejznámější můžeme zařadit Web of Science, Scopus nebo také Google Scholar. Každý z těchto rejstříků pokrývá různé zaměření a časový rozptyl.

3.3 Citační rejstříky

První citačním rejstřík, který byl spuštěn americkým institutem pro vědecké informace (ISI) v 50. letech 20. století se nazýval Current Contents. Jednalo se o rejstřík v papírové formě a ze začátku zahrnoval pouze biologii a medicínu. Později byly přidány i další obory. Byl publikován týdně a obsahoval pouze názvy článků z významných recenzovaných vědeckých časopisů. V rámci tohoto rejstříku byl uveden seznam autorů a klíčových slov, dále zde byly poskytnuty kontakty na jednotlivé autory, které sloužily pro čtenáře, aby mohli zasílat žádosti o dotisk kopií samotných článků.

Na začátku 60. let spustila ISI citační rejstřík s názvem Citační index (The Science Citation Index), který vytvořil Eugene Garfield. Tento rejstřík obsahuje kompletní bibliografickou informaci o každém článku, která zahrnuje jméno autora, adresu autora, název článku, název časopisu a počet stran, ale také použité citace a články citující další články. Pomocí tohoto rejstříku můžeme zjistit kdo daný článek publikoval, ale také jaké další články článek cituje. [12]

Základním principem tohoto rejstříku je přístup nejen k článku, který znám, ale také k dalším článkům, které mohou být také důležité. V rámci výzkumu tohoto rejstříku bylo zjištěno, že ve velmi malém množství časopisů se nachází většina významných informací, které byly každoročně publikovány a tyto časopisy tvoří nejkvalitnější jádro. Čím dále se vzdalujeme od tohoto jádra, tím nám kvalita publikací upadá. Tento rejstřík je seřazen abecedně podle jmen autorů a doplněn bibliografickými údaji jejich publikací, dále pod každou citovanou publikací je uveden autor a bibliografický popis citujícího pramene.

V následující tabulce můžeme vidět základní přehled výše zmíněných citačních rejstříků.

Tabulka 1 - Základní přehled citačních rejstříků

	Web of Science	Scopus	Google Scholar
Množství záznamů	Přes 12 000 časopisů	Více než 19 500 časopisů	Neuvedeno
Zaměření	<ul style="list-style-type: none"> • Vědecká činnost • Technologie • Sociální vědy • Umění • Humanitní vědy 	<ul style="list-style-type: none"> • Vědecká činnost • Technologie • Inženýrství • Medicína • Umění • Humanitní vědy 	<ul style="list-style-type: none"> • Vědecká činnost • Technika • Sociální vědy • Umění • Humanitní vědy • Obchod
Časový rozptyl	Záznamy od roku 1900	Záznamy od roku 1823, většina však až od roku 1996	Všechny online dostupné záznamy
Výhody	<ul style="list-style-type: none"> • Hlubší struktura záznamů jednotlivých časopisů • Využívá citační zprávy • Zaměřuje se převážně na výzkumy v U.S.A • Často obsahuje citační mapování pro vizualizaci 	<ul style="list-style-type: none"> • Přátelské uživatelské prostředí • Široká škála časopisů • Lze provádět export dat • Mezinárodní zaměření 	<ul style="list-style-type: none"> • Obsahuje kompletní obraz dopadu v indexech • Obsahuje i diplomové práce, knihy, abstrakty, články profesionálních sdružení, univerzit • Obsahuje nejvíce záznamů
Nevýhody	<ul style="list-style-type: none"> • Obsahuje chyby v citacích • Cena 	<ul style="list-style-type: none"> • Lze hledat citace pouze od roku 1996 • Slabší pokrytí humanitních oborů 	<ul style="list-style-type: none"> • Omezené vyhledávání • Nepřesné množství citací • Problémy s exportem • Obsahuje pouze online záznamy • Obsahuje kvantitu, nikoliv kvalitu citací

[11]

3.3.1 Web of Science

Web of Science je webovým rozhraním do tří databází s bibliografickými údaji z oblasti výzkumu a vývoje. Jedná se o databázi Science Citation Index, Social Science Citation Index a Arts and Humanities Index. Zprovozněn byl v roce 1997. Je součástí nadřazeného celku Web of Knowledge.

Poskytuje citační informace z jednotlivých článků a sleduje obsah více než 8000 recenzovaných časopisů. Jedná se o databázi, takže je možné jí prohledávat dle požadavků uživatele.

Současná tvorba citačních databází se opírá o jednoznačné identifikátory publikací. Mnohé z nich jsou již uloženy elektronicky. Je umožněn přístup k plným textům, které jsou rovněž uloženy v databázi, v depozitářích digitálních knihoven, archívu elektronických tisků případně přímý přístup na internet. Americký institut pro vědecké informace (ISI) je institucí, která se silně angažuje ve vybudování integrovaného informačního systému, který je zaměřen na citační vazby, které se v oblasti vědy jeví jako nejcennější. Strategie ISI směřuje k automatizovanému budování hypertextových vazeb na vybrané expertní zdroje plných znění textů.

Nadřazený Web of Knowledge, jehož je Web of Science součástí, umožňuje přístup do dalších databází jako je např. INSPEC, Current Contents nebo do rejstříků SCI, SSCI a AHCI. [13]

3.3.2 Scopus

SCOPUS je databáze odborné recenzované literatury. Obsahuje mj. abstrakty a citace jednotlivých literárních děl. Každý záznam obsahuje i link na plné znění textu. V současnosti jsou v databázi abstrakty z téměř 20 500 recenzovaných časopisů od více než 5000 vydavatelů z celého světa. Přístup do databáze je pro registrované uživatele. [14]

Databáze SCOPUS byla uvedena na trh nizozemským mezinárodním multimediálním vydavatelstvím Elsevier (od roku 1880), které se specializuje na oblasti:

- Vědecká
- Technologická
- Medicinská
- Sociální vědy

Obsah databáze SCOPUS je rozdělen do 2 sekcí a to Science & Technology a Health Sciences.

Cílové skupiny uživatelů jsou v sekci

1. Science & Technology:

- Akademické a vládní výzkumné organizace
- Podnikové výzkumné laboratoře
- Knihkupci
- Knihovníci
- Vědečtí výzkumníci
- Autoři
- Editoři

2. Health Sciences:

- Lékaři
- Zdravotní sestry
- Zdravotnické a pečovatelské školy a jejich studenti
- Výzkumní pracovníci
- Farmaceutické společnosti
- Nemocnice a výzkumné instituce

3.3.2.1 Hledání v databázi SCOPUS

Při hledání požadované bibliografické informace vyplní uživatel textové pole Search. Obsahem může být titul, abstrakt, klíčové slovo apod. Po prohledání databáze zobrazí výsledek hledání. Zpracovaný dotaz se zobrazí jako očíslovaná položka v poli dotazů na formuláři. Pokud je potřeba výběr zúžit, pokračuje se dalším dotazem Search, který je aplikován na množinu dokumentů z předchozích prohledávání. Mezi jednotlivým ploškami dotazu, resp. jednotlivými dotazy v poli dotazů je možno nastavit relaci pomocí logických operátorů AND a OR. [15]

Pokud je výběr dokumentů ukončen a zobrazen v okně aplikace, je možno v rámci těchto dokumentů nastavit ještě další filtry:

Year

Jedná se o roky, kdy byly jednotlivé dokumenty vydány. Jsou zobrazeny roky všech vybraných dokumentů ve formě check boxu. Zatržením aplikace umožňuje výběr položek z daného roku, resp. vyloučení položek.

Author name

Zobrazí se všichni autoři z vybrané skupiny dokumentů a zatržením jednotlivých check boxů je možno zúžit výběr na vybraného autora/autory.

Subject area

Zobrazí se všechny položky subjektů obsažených v dokumentech a je možno pomocí check boxů výběr zúžit.

Např: Medicine, Nursing, Social Sciences, ...

Document title

Pomocí check boxů je možno zúžit výběr typu z:

- Article
- Review
- Letter
- Editorial
- Note

3.3.2.2 Uspořádání výpisu vybraných dokumentů

Výpis dokumentů je možno seřadit dle různých kritérií. Databáze nabízí následující možnosti řazení:

- Date (newest)
- Date(oldest)
- Cited by (highest)
- Cited by (lowest)
- Relevance
- First Author (A-Z)
- First Author (Z-A)

- Source Title (A-Z)
- Source Title (Z-A)

Při výběru jedné položky seznamu pro zobrazení detailu je možno prohlížet kromě záhlaví dokumentu, též abstrakt a seznam klíčových slov.

V detailním zobrazení je možno provést:

- Export
- Download
- Print
- E-mail
- Save to PDF
- Add to List

3.3.2.3 Export dat z databáze SCOPUS

System umožňuje export vybraných dat ve zvoleném formátu pro další využití mimo prostředí SCOPUSu.

V zobrazeném seznamu je možno zatrhnout jen některé záznamy nebo označit celý seznam pro export.

Pro export je třeba dále vybrat požadovaná pole každé položky, která mají být do exportu zahrnuta.

Položky záznamu jsou rozděleny do 5 skupin. V každé této skupině je možno individuálně zatrhnout vybrané položky nebo zatrhnout celou skupinu.

Záznam obsahuje následující skupiny:

1. Citation information

- Author(s)
- Document title
- Year
- Source title
- Volume, Issue, Pages
- Citation count
- Source and Document Type

- DOI
2. Bibliographical information
 - Affiliations
 - Serial identifiers (e.g. ISSN)
 - PubMed ID
 - Publisher
 - Editor(s)
 - Language of Original Document
 - Correspondence Address
 - Abbreviated Source Title
 3. Abstract and Keywords
 - Abstract
 - Author Keywords
 - Index Keywords
 4. Funding Details
 - Number
 - Acronym
 - Sponsor
 - Funding text
 5. Other information (Check Box)
 - Tradenames and Manufactures
 - Accession numbers and Chemicals
 - Conference information
 - Include references

Nakonec je třeba vybrat formát požadovaného exportu. Výběr je z následujících možností:

- MENDELEY
- RefWorks
- SciVAL
- RIS Format (EndNote, Reference Manager)
- CSV (Excel)
- BibTeX
- Text (ASCII in HTML)

3.3.3 Google Scholar

Služba Google Scholar umožňuje vyhledávání akademických prací z jednoho místa. Je možno prohledávat materiály z mnoha vědních oborů.

Mohou to být:

- Diplomové práce
- Knihy
- Abstrakty
- Odborné studie
- Materiály vydané akademickými nakladatelstvími, odbornými společnostmi, univerzitami

Služba je určena pro vědecké pracovníky, pedagogy a studenty.

Funkce služby:

- Vyhledávání zdrojů.
- Vyhledávání článků, abstraktů, citací.
- Nalezení plného znění materiálu
- Informace o klíčových člancích v různých oborech

Vyhledávání může probíhat jednak spojením podmínek za pomoci logických operátorů a dále rozšířenými možnostmi, kdy je vyhledání omezeno na jednoho autora, nebo dobu publikování apod.

Služba je schopna třídit články podle kritérií, kdy je brán v potaz celý text včetně autora, zdroj publikace, počtu citací. Nejvýznamnější články jsou pak zobrazeny na prvních místech.

Je možno používat operátory:

- **Operátor “+”** ... je možno nastavit, aby výsledky hledání obsahovaly konkrétní slova nebo čísla
- **Operátor „-“** ... vylučuje všechny materiály obsahující zadaný výraz
- **Vyhledávání frází** ... ve výběru jsou pouze materiály obsahující přesně tuto frázi
- **Operátor “OR”** ... výsledky hledání obsahují alespoň jeden zadaný výraz (logický součet)
- **Operátor “intitle”** ... zobrazuje položky, které v titulu obsahují zadané slovo

3.4 Nástroje pro analýzu a vizualizaci

Množství informací v digitálním světě nám umožňuje provádět rozsáhlé analýzy těchto dat. Pomocí těchto analýz můžeme znázornit a vysvětlit řadu vědeckých, společenských a ekonomických jevů. Přestože máme dnes databáze, které obsahují rozsáhlé množství těchto dat, je velice náročné je zkoumat. Bez výpočetních nástrojů by to bylo téměř nemožné. Pro pochopení analýzy těchto dat nám však ani tyto nástroje nemusí dostatečně posloužit, proto je zapotřebí aplikovat metodu vizualizace dat, která analytikům umožňuje celkový pohled na síť těchto dat.

Tato vizualizace nám zprostředkovává samostatnou strukturu, pomocí které můžeme vidět hustotu sítě, topologii sítě a celkovou velikost zkoumané sítě dat. Pomáhá nám porozumět datům a provádět jejich analýzu co nejefektivněji. Je totiž obecně známo, že pomocí zraku jsme nejpřirozeněji schopni vnímat mnohem větší množství dat v daleko kratším čase. Dále je možné v takové vizualizaci operativně daleko více operací, než v samotných databázích dat.

Je však také nutné upozornit na to, že v případě, kdy síť obsahuje příliš mnoho dat, nám taková vizualizace nemusí vypovídat nic užitečného a to proto, že dochází k překrývání jednotlivých vrcholů a nečitelnosti popisků jednotlivých uzlů. Takováto struktura může být velmi přeplněná a nemůžeme provádět ani operace jako je filtrování dat, přibližování nebo posouvání obsahu. Je tedy zapotřebí před samotnou vizualizací upravit taková data pomocí jiných metod. Například se odstraní některá nepotřebná data nebo naopak zvýrazní data, která nás zajímají. Pomocí těchto metod se nám struktura sítě neporuší na rozdíl třeba od filtrování dat na základě různých atributů.

V následující části této kapitoly jsou stručně představeny tyto nástroje a práce s nimi.

3.4.1 VOSviewer

VOSviewer je programový nástroj pro konstrukci a vizualizaci bibliometrických sítí. Tento software byl vyvinut vývojáři Nees Jan van Eckem and Ludo Waltmanem na Leiden University v Centre for Science and Technology Studies (CWTS). Sítě mohou obsahovat časopisy, vědecké zprávy a knihy. Lze vytvářet síť na základě citací, bibliografických propojení, spolucitace nebo spoluautorských vztahů. Produkt nabízí možnost vytěžovat texty, které lze použít ke konstrukci a vizualizaci sítí společného výskytu důležitých výrazů získaných z vědecké literatury. [16]

3.4.1.1 Vizualizace

Vizualizace bibliometrických sítí lze podrobně prozkoumat pomocí funkcí přiblížení a posouvání podobných například Mapám Google. Inteligentní algoritmus značení zabraňuje tomu, aby se popisky uzlů navzájem překrývaly.

Hustota sítě a překryvné vizualizace. Hustota vizualizace poskytuje rychlý přehled hlavních oblastí v bibliometrické síti. Vizualizace překrytí lze například použít k zobrazení vývoje v čase.

Snímky obrazovky bibliometrických síťových vizualizací lze vytvářet ve vysokém rozlišení a lze je ukládat v mnoha grafických formátech souborů, bitmapových i vektorových.

3.4.1.2 Používané techniky

Pokročilé techniky rozložení a shlukování. K dispozici jsou nejmodernější techniky rozložení sítě a vytváření shluků uzlů. Výsledky rozložení a shlukování lze doladit pomocí různých parametrů.

Techniky zpracování přirozeného jazyka. K vytváření sítí pro souběžný výskyt termínů na základě textových dat v anglickém jazyce jsou k dispozici techniky zpracování přirozeného jazyka. Relevantní a nerelevantní pojmy lze rozlišit algoritmicky.

Vytváření bibliometrických sítí. K vytváření bibliometrických sítí je k dispozici řada pokročilých funkcí (např. Spoluautorství, bibliografická vazba a síť spolucitování). Například vliv publikací s mnoha autory, mnoha citacemi nebo mnoha odkazy lze snížit pomocí přístupu částečného počítání.

3.4.2 Cytoscape

Program Cytoscape je vizualizační program pro zobrazení a analýzu sítí všeho druhu. Knihovna byla vyvinuta v Donnelly Centre na University of Toronto. Původně sloužil k vizualizaci vztahů mezi makromolekulami. Funguje jak v komerčních projektech, tak i jako open source (vývoj jádra). K jeho vývoji přispívá velká komunita. Vzniklo tak množství užitečných pluginů. [17]

Hlavní charakteristiky produktu:

- Bohatá grafická knihovna napsaná v JS
- Open source licence (MIT) pro jádro Cytoscape.js knihovny
- Je orientován především k uživateli
- Důkladně optimalizován
- Podporuje renderování obrázků grafů
- Výstup umožňuje automatické i manuální pozicování uzlů
- Používá šablony stylů
- Podporuje operace teorie množin
- Zahrnuje algoritmy teorie grafů
- Animovatelné prvky grafu a výřezů
- Dobře udržovaný
- Týdenní frekvence vydání patche
- Kompatibilní se všemi moderními prohlížeči
- Modulární konstrukce
- Přesouvání celého grafu
- Měnit měřítko zobrazení
- Aplikovat algoritmus pro rozmístění vrcholů
- Analýza sítě a výpočet různých metrik
- Zobrazit síť s ohledem na hodnoty metrik (velikost a barva uzlů a hran)

Základní program neumožňuje zobrazení vývoje sítě v čase.

3.4.3 Gephi

Gephi je nástroj pro analýzu a vizualizaci sítě. V rámci této práce byl Gephi zvolen jako hlavní vizualizační nástroj pro vlastní část práce.

Sít se skládá ze dvou složek: seznamu objektů tvořících síť a seznamu vztahů (interakce mezi objekty). Objekty nazýváme uzly, vztahy jsou nazývány hrany.

Každý uzel *node* musí mít jednoznačný identifikátor *id* a další atribut *label*, který ho popisuje. Tento popisek slouží pak jako označení daného uzlu při grafické prezentaci. Uzel může mít řadu dalších atributů, které vyjadřují jeho vlastnosti.

Identifikátor *id* nemusí být vždy uměle vytvořen. S ohledem na charakter objektů lze využít přímo některého atributu, pokud bude zaručena jeho jednoznačnost. Pokud by se jednalo např. o tabulku oficiálně vydaných knih, pak lze využít ISBN.

Vztahy mezi uzly jsou vyjádřeny hranami *edge*. Hrana je dána párem *id1 – id2*, který vyjadřuje, že hrana propojuje uzel s identifikátorem *id1* s uzlem s identifikátorem *id2*.

3.4.3.1 Import dat do Gephi.

Pro import předpokládáme, že data jsou již předem předzpracována. Možnou reprezentací jsou tabulky MS Excelu ve formátu CSV.

3.4.3.2 Objekty s již přiděleným id.

V této variantě jsou předem připraveny dvě tabulky, kdy první obsahuje uzly (1 řádka=1 uzel) a druhá tabulka obsahuje dvojice *id*.

Příklad:

Tabulka 2 - Tabulka hran

id	Nazev	Atribut 1	Atribut 2	Atribut 3
11	Uzel 1	UA11	112	UA13
12	Uzel 2	UA21	222	UA23
13	Uzel 3	UA31	332	UA33
14	Uzel 4	UA41	442	UA43
15	Uzel 5	UA51	552	UA53
16	Uzel 6	UA61	662	UA63
17	Uzel 7	UA71	772	UA73
18	Uzel 8	UA81	882	UA83

Tabulka 3 - Tabulka uzlů

Source	Target
11	12
11	13
12	14
12	15
14	18
13	15
13	16
13	17

Zdroj: Vlastní zpracování

Ve formátu CSV to vypadá následovně

Uzly:

id; Navez; Atribut 1; Atribut 2; Atribut 3

11; Uzel 1; UA11;112; UA13

12; Uzel 2; UA21;222; UA23

13; Uzel 3; UA31;332; UA33

14; Uzel 4; UA41;442; UA43

15; Uzel 5; UA51;552; UA53

16; Uzel 6; UA61;662; UA63

17; Uzel 7; UA71;772; UA73

18; Uzel 8; UA81;882; UA83

Hrany:

Source; Target

11;12

11;13

12;14

12;15

14;18

13;15

13;16

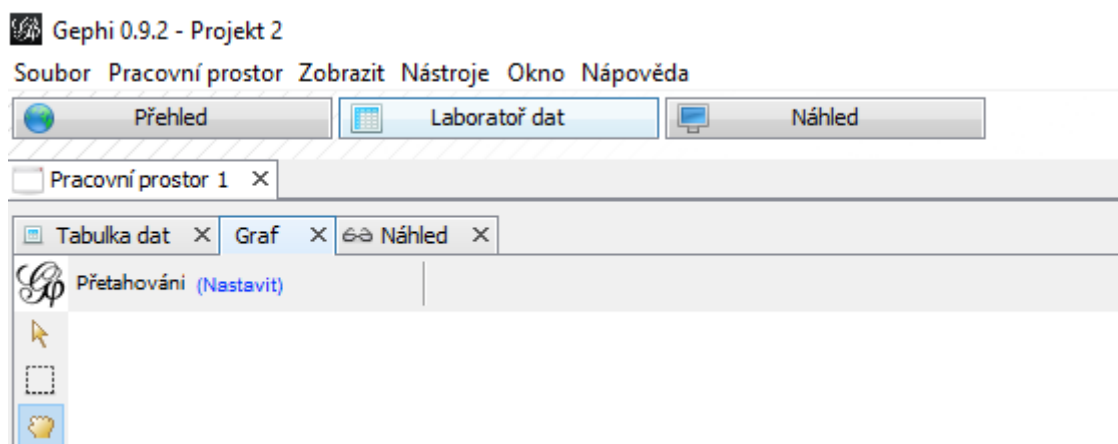
13;17

12;13

V rámci programu Gephi postupujeme následujícím způsobem.

Na úvodní obrazovce založíme nový projekt: **Soubor – Nový projekt**

Obrázek 2 - Prostředí Gephi

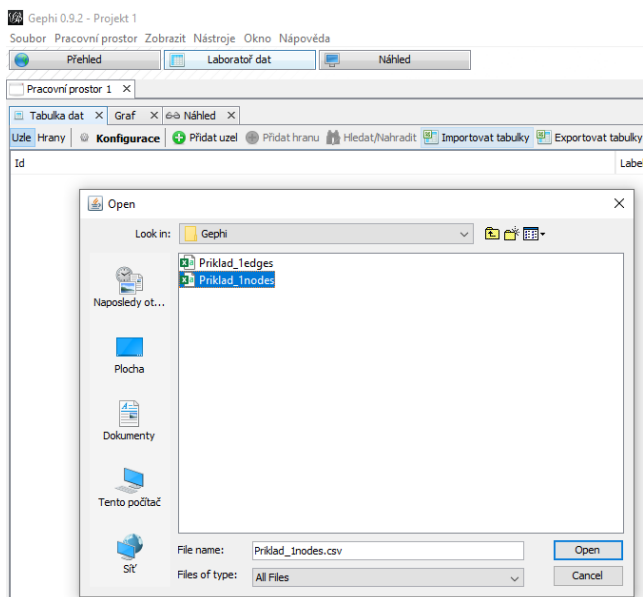


Zdroj: Vlastní zpracování

Přepneme na **Laboratoř dat** a zvolíme **Tabulka dat**. Nejprve budeme importovat tabulku uzlů. Přepínač **Uzle/Hrany** musí být nastaven na **Uzle**.

Následně zvolíme: **Importovat tabulky**.

Obrázek 3 - Import dat do prostředí Gephi

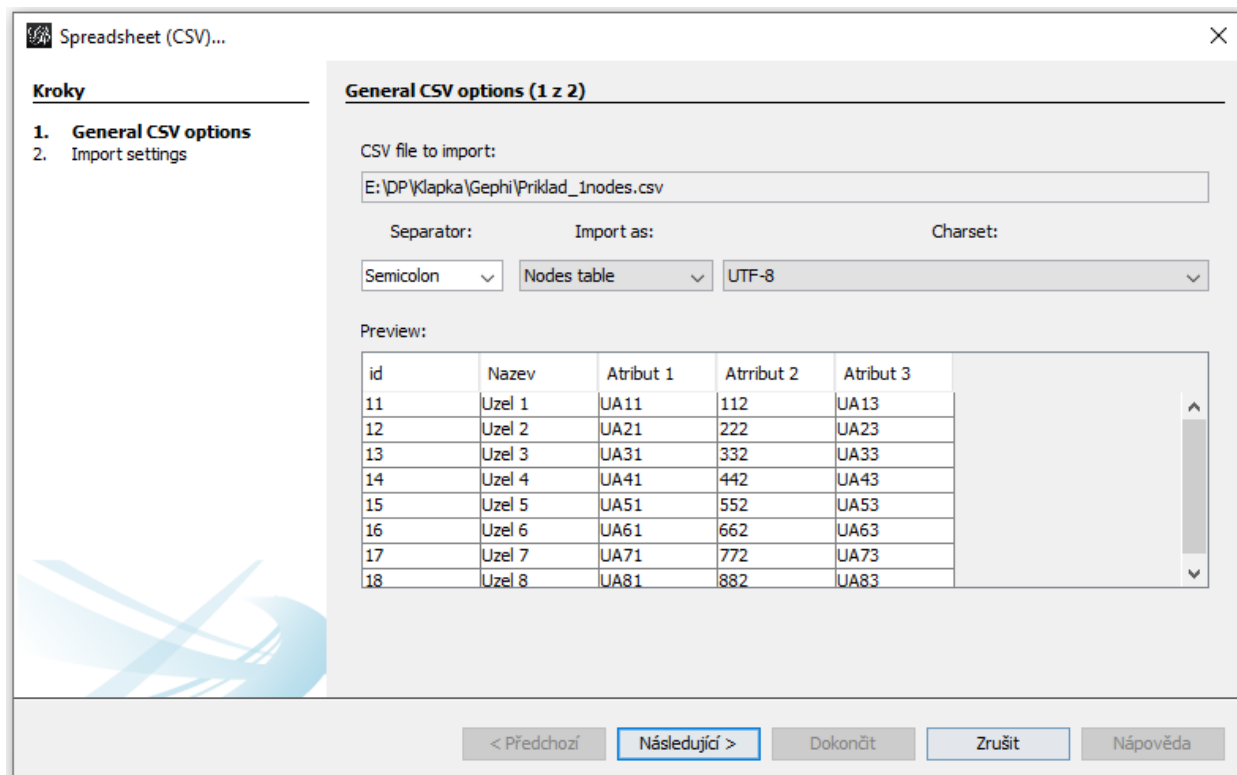


Zdroj: Vlastní zpracování

Nalistujeme soubor (CSV) s tabulkou uzlů a otevřeme ho.

Zobrazí se tabulkový náhled.

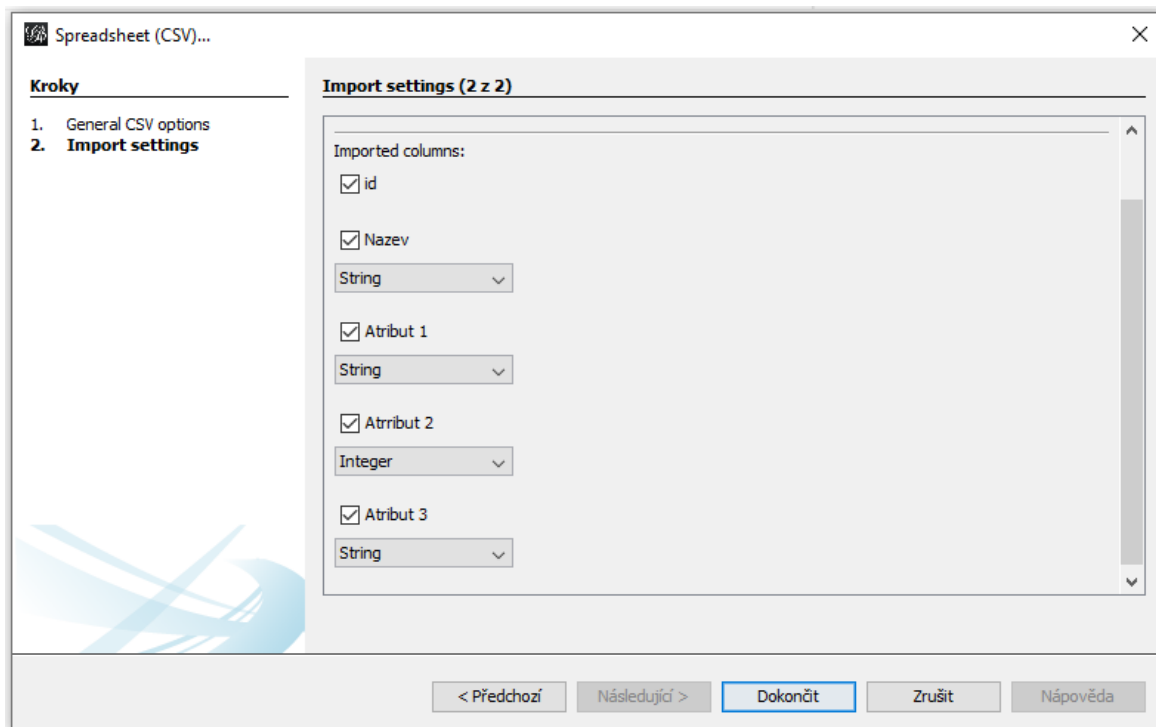
Obrázek 4 - Možnosti importu tabulky uzlů do prostředí Gephi



Zdroj: Vlastní zpracování

V dalším kroku se zobrazí první řádek tabulky, tj. záhlaví sloupců. Pomocí check boxů vybíráme sloupce, které chceme importovat. Současně Gephi „odhadl“ datový typ jednotlivých sloupců. Je nezbytně nutné projít toto přednastavení a provést případné korekce.

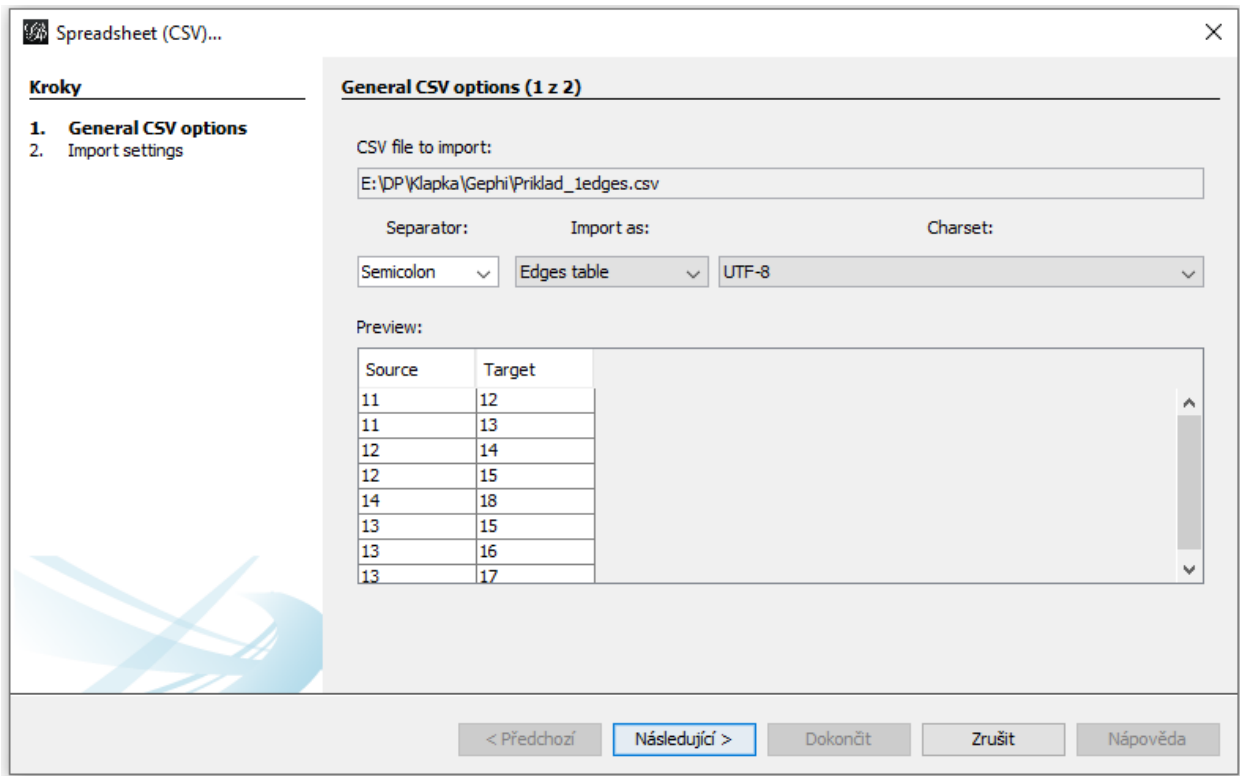
Obrázek 5 - Nastavení importu a volba datového typu



Zdroj: Vlastní zpracování

Po dokončení importu tabulky uzlů se obdobně provede import tabulky hran. Přepínač v horní liště se přepne do polohy **Hrany**, následně se provede import souboru s tabulkou hran.

Obrázek 6 - Import tabulky hran do prostředí Gephi

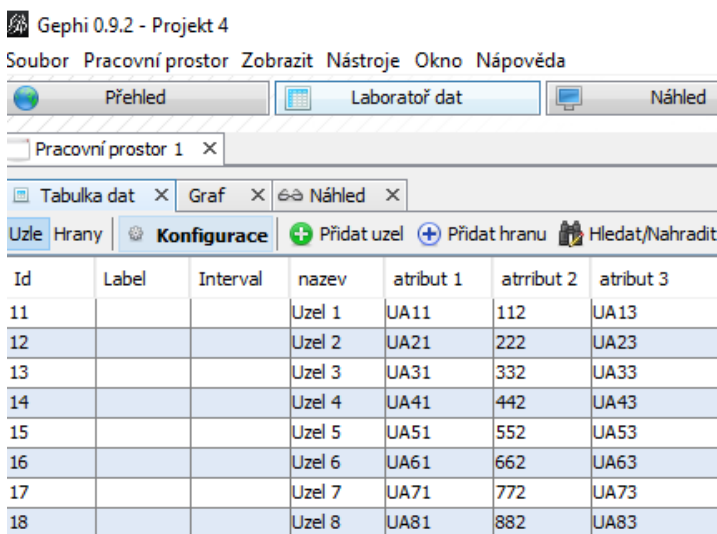


Zdroj: Vlastní zpracování

Zobrazí se náhled tabulky a následně se operace dokončí. Volbou **Tabulka dat** a **Uzle/Hrany** vidíme tabulky grafu uložené ve vnitřních strukturách Gephi.

V tabulce uzlů přibyl kromě jiného sloupec **Label**, který slouží jako popisek jednotlivých uzlů grafu. V tuto chvíli je zatím prázdný.

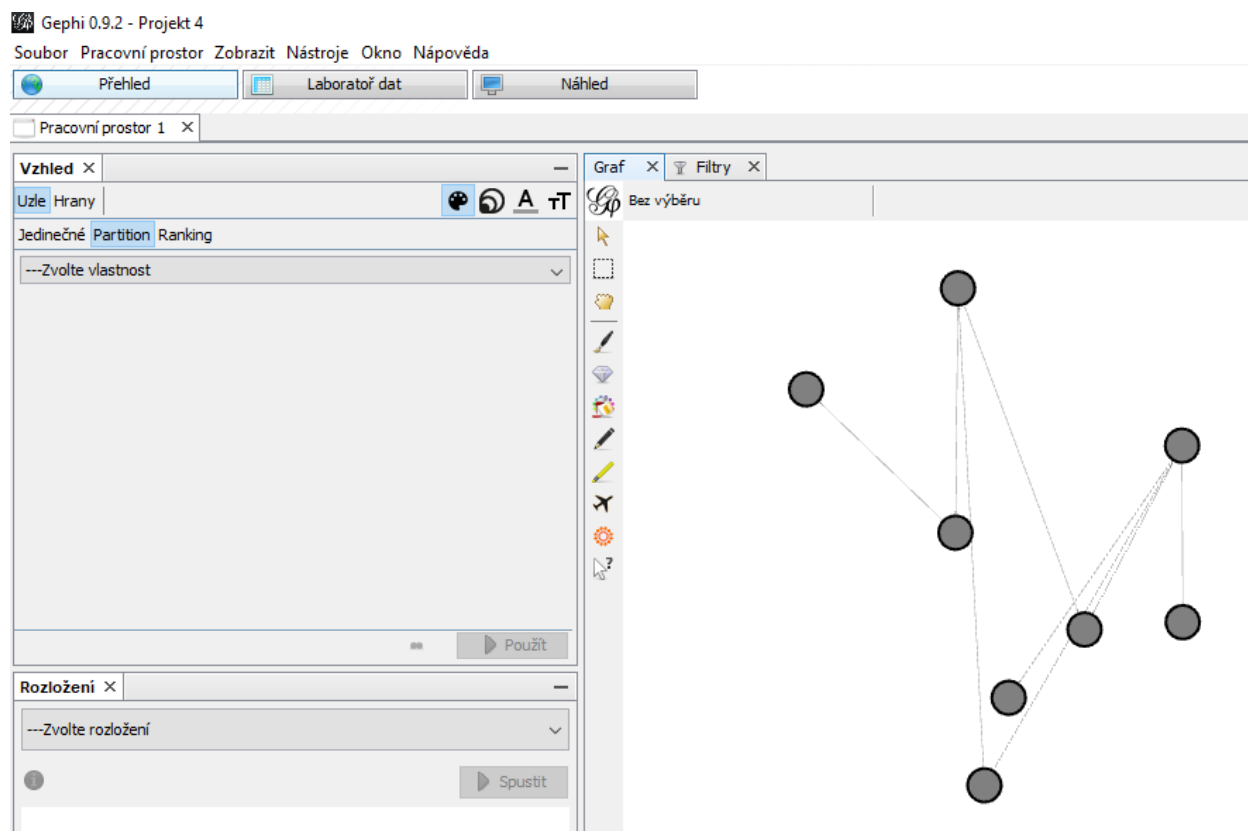
Obrázek 7 - Tabulka uzlů v prostředí Gephi



Zdroj: Vlastní zpracování

Náhled grafu můžeme zobrazit, pokud opustíme volbu **Laboratoř dat** a přepneme se na **Přehled**.

Obrázek 8 – Náhled grafu příkladu

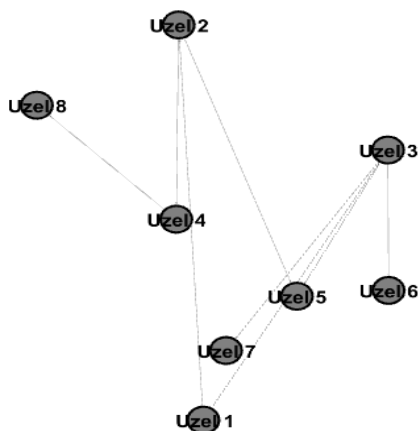


Zdroj: Vlastní zpracování

Tímto je import dat ukončen. V tuto chvíli je možno pro lepší orientaci ještě nastavit popisek uzlů. Nastavíme **Laboratoř dat** – Tabulka **dat** – **Uzle**. V dolní části obrazovky vybereme tlačítko **Kopírovat data na další sloupec** a v rozbalovacím menu vybereme položku, která bude charakterizovat každý uzel. V našem příkladu **nazev**. Vyskočí další okno, kde zvolíme cílový sloupec, v našem případě **Label**. [18]

Po zobrazení grafu jsou již součástí grafického vyjádření a popisky uzlů.

Obrázek 9 - Graf s popisky uzlů



Zdroj: Vlastní zpracování

3.4.3.3 Objekty bez předem přiděleného id.

Jako příklad lze uvést tabulku dokumentů (uzlů), z nichž každý referencuje jiné dokumenty. Obecně tyto dokumenty neobsahují nějaký jednoznačný klíč a je třeba je při importu provázat.

Pro každou položku (řádek) primární tabulky dokumentů je potřeba vytvořit sekundární tabulku s referencovanými N dokumenty (uzly). Při importu do Gephi je nejprve vložena položka primární tabulky a k ní se připojí všechny záznamy její sekundární tabulky. Gephi pak postupuje tak, že každé vstupující položce jak primární, tak i sekundární tabulky přidělí unikátní numerický klíč a automaticky vytvoří vnitřní tabulku hran, tj. relaci 1: N . Toto je třeba provést pro všechny položky primární tabulky. [19]

Příklad: Mějme 3 fiktivní publikace, z nichž každá cituje další publikace. Data jsou uspořádána v tabulkách.

Tabulka 4 - Základní publikace

Autor	Titul	Rok_vydani	Vydavatelstvi	Pocet_stran
Šulc Josef	Údržba střech	2017	ČVUT	230

Zdroj: Vlastní zpracování

cituje

Tabulka 5 - Tabulka citovaných publikací

Autor	Titul	Rok_vyda ni	Vydavatelst vi	Pocet_stra n
Karbaník Jiří	Klempířské nářadí	1992	Grada	180
Votýpka Zdeněk	Střešní materiály	2005	Grada	200
Kačl Jan	Bitumenové střešní krytiny	2012	Grada	220
Koran Jaroslav	Impregnační a dezinfekční přípravky	1995	Grada	99

Zdroj: Vlastní zpracování

Tabulka 6 - Základní publikace

Autor	Titul	Rok_vydani	Vydavatelstvi	Pocet_stran
Benešová Eva	Vaříme v restauraci	2012	Artia	350

Zdroj: Vlastní zpracování

cituje

Tabulka 7 - Tabulka citovaných publikací

Autor	Titul	Rok_vyda ni	Vydavatelst vi	Pocet_stra n
Pivota Jan	Kuchyňské stroje	1999	Grada	180
Koran Jaroslav	Impregnační a dezinfekční přípravky	1995	Grada	99
Stoček Petr	Bezmasá kuchařka	2000	Academia	180
Pospíchalová Iveta	Dietní recepty	2000	Academia	240

Zdroj: Vlastní zpracování

Tabulka 8 - Základní publikace

Autor	Titul	Rok_vydani	Vydavatelstvi	Pocet_stran
Vobuch Jiří	Strojní dílny	2017	Grada	200

Zdroj: Vlastní zpracování

cituje

Tabulka 9 - Tabulka citovaných publikací

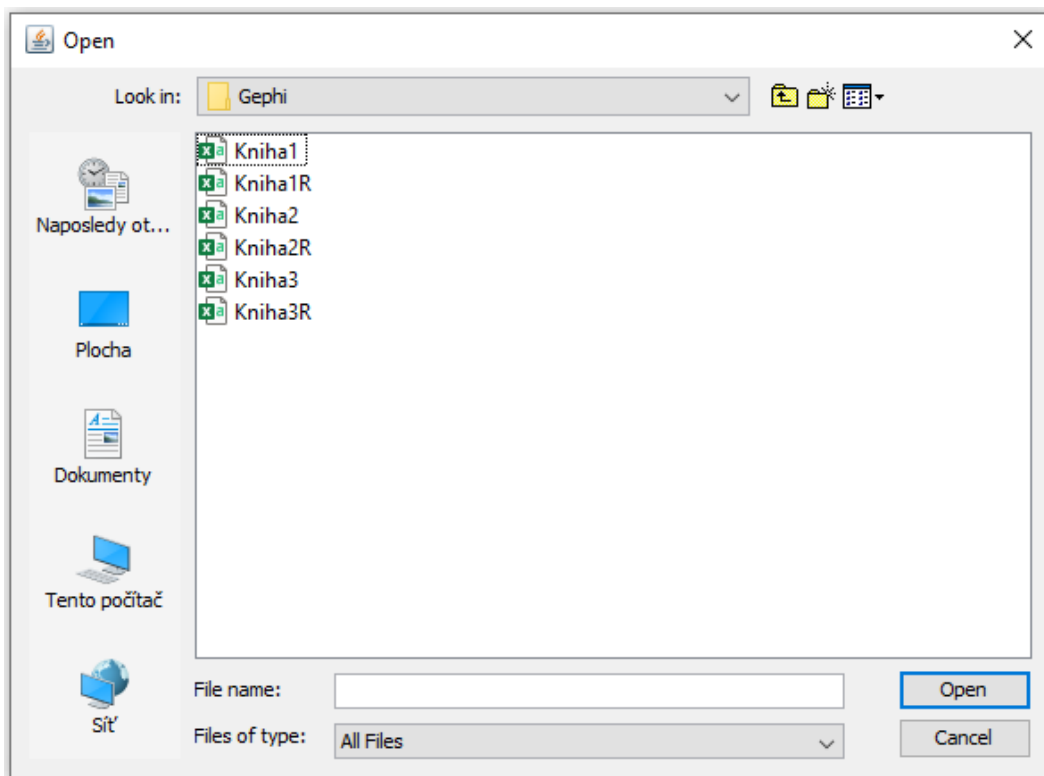
Autor	Titul	Rok_vydani	Vydavatelstvi	Pocet_stran
Vobuch Jiří	Točivé obráběcí stroje	2015	Grada	150
Nevara Petr	Průmyslové vysavače	2000	Kaercher	110
Javůrek Pavel	Bezpečnost práce při strojní výrobě	1995	ČVUT	150

Zdroj: Vlastní zpracování

Pro názornost je každá z těchto tabulek uložena na samostatném souboru CSV. Vždy publikace, která referencuje další publikace jako *Knihai* a jí referencované publikace na souboru *KnihaiR*.

V Gephi založíme nový projekt. V *Laboratoři dat* importujeme pouze tabulky uzlů. Vybereme soubor první publikace.

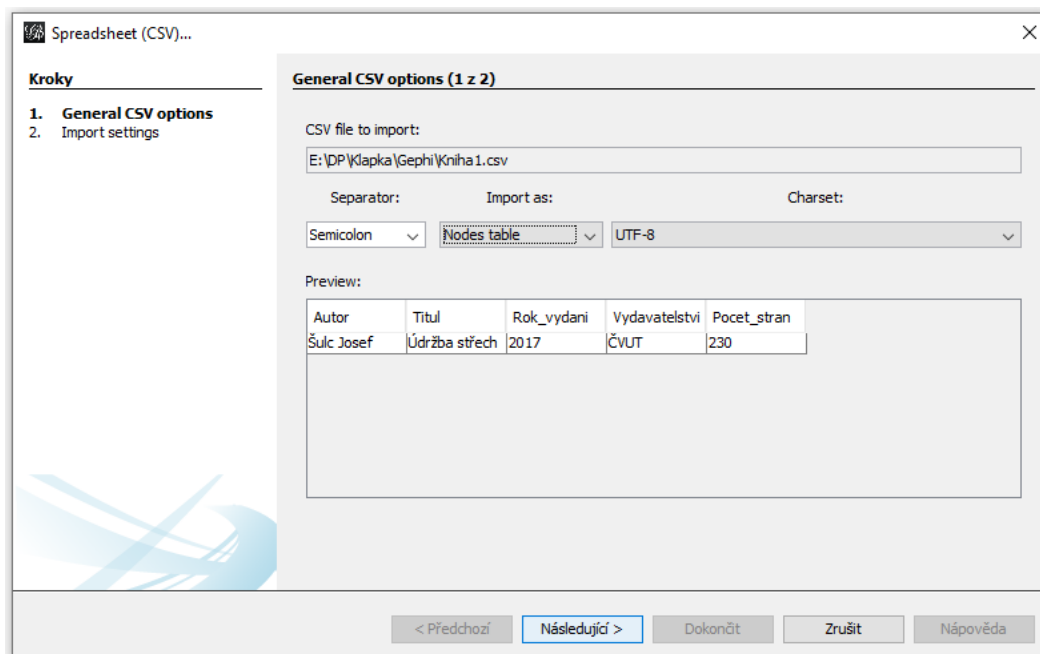
Obrázek 10 - Import zdrojových dat



Zdroj: Vlastní zpracování

Po otevření se zobrazí náhled na data. Je důležité zvolit u položky *Import as:* hodnotu *Nodes table*.

Obrázek 11 - Nastavení importu zdrojových dat

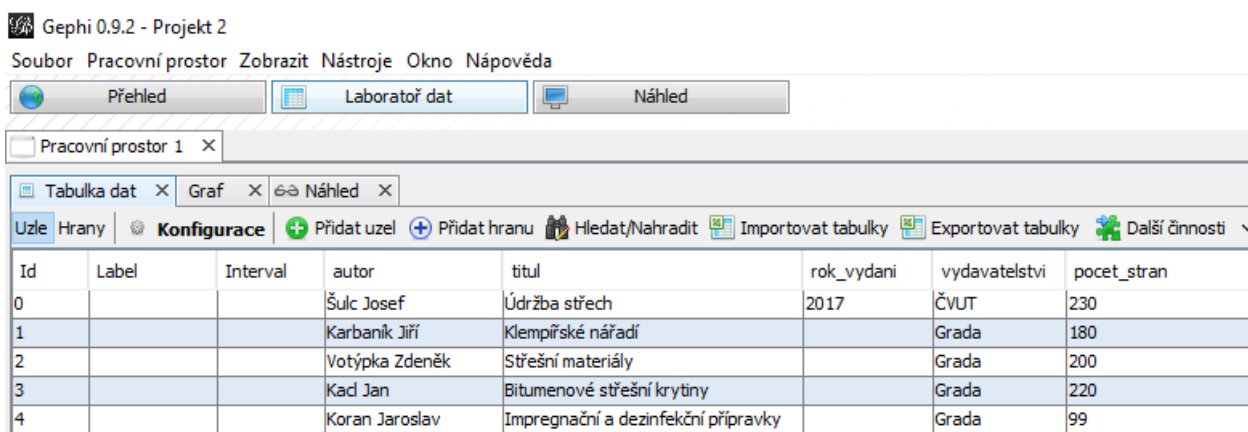


Zdroj: Vlastní zpracování

V dalším kroku je třeba zkontrolovat a případně korigovat datové typy. Viz výše. Nové položky vždy přidáváme ke stávajícím do stejného **Pracovního prostoru** Gephi.

Stejným způsobem importujeme i soubor referencí. Po importu obou souborů vypadá situace následovně.

Obrázek 12 - Laboratoř dat v prostředí Gephi



Zdroj: Vlastní zpracování

Každý řádek tvoří uzel grafu. Nyní je užitečné vyplnit sloupec **Label**, abychom se lépe orientovali při grafickém vyjádření. Jako popisec uzlů můžeme volit libovolný sloupec typu string, zde zvolíme atribut **titul**. Pokud si nyní zobrazíme **Přehled**, vidíme izolované uzly a jejich popisky. Vazby dosud nebyly definovány.

Obrázek 13 - Graf izolovaných uzlů



Zdroj: Vlastní zpracování

Nyní nastavíme vazbu mezi uzly. Je nutné znát základní publikaci a její citace. Označíme si levým tlačítkem myši všechny relevantní řádky.

Obrázek 14 - Vytvoření odkazů na uzle

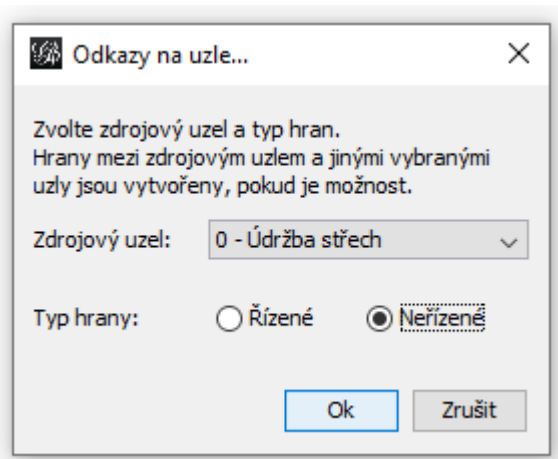
Gephi 0.9.2 - Projekt 2
Soubor Pracovní prostor Zobrazit Nástroje Okno Nápověda
Přehled Laboratoř dat Náhled
Pracovní prostor 1 X
Tabulka dat x Graf x Náhled x
Uzle Hrany Konfigurace Přidat uzel Přidat hranu Hledat/Nahrát Importovat tabulky Exportovat tabulky Další činnosti v

Id	Label	Interval	autor	titul	rok_vydani	vydavatelstvi	pocet_stran
0	Údržba střech		Šulc Josef	Údržba střech	2017	ČVUT	230
1	Klempířské nářadí		Karbaník Jiří	Klempířské nářadí		Grada	180
2	Střešní materiály		Votýpka Zdeněk	Střešní materiály		Grada	200
3	Bitumenové střešní krytiny		Kad Jan	Bitumenové střešní krytiny		Grada	220
4	Impregnační a dezinfekční přípravky		Koran Jaroslav	Impregnační a dezinfekční přípravky		Grada	99

Zdroj: Vlastní zpracování

Nyní se postavíme na základní publikaci a pravým klikem zahájíme proces spojování.

Obrázek 15 - Tvorba hran v prostředí Gephi

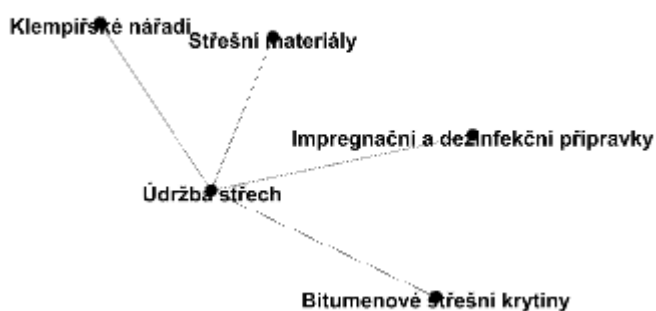


Zdroj: Vlastní zpracování

Gephi provede spojení zdrojového uzlu se všemi jeho referencemi.

V přehledu situace vypadá následovně.

Obrázek 16 - Graf propojených uzlů pomocí hran

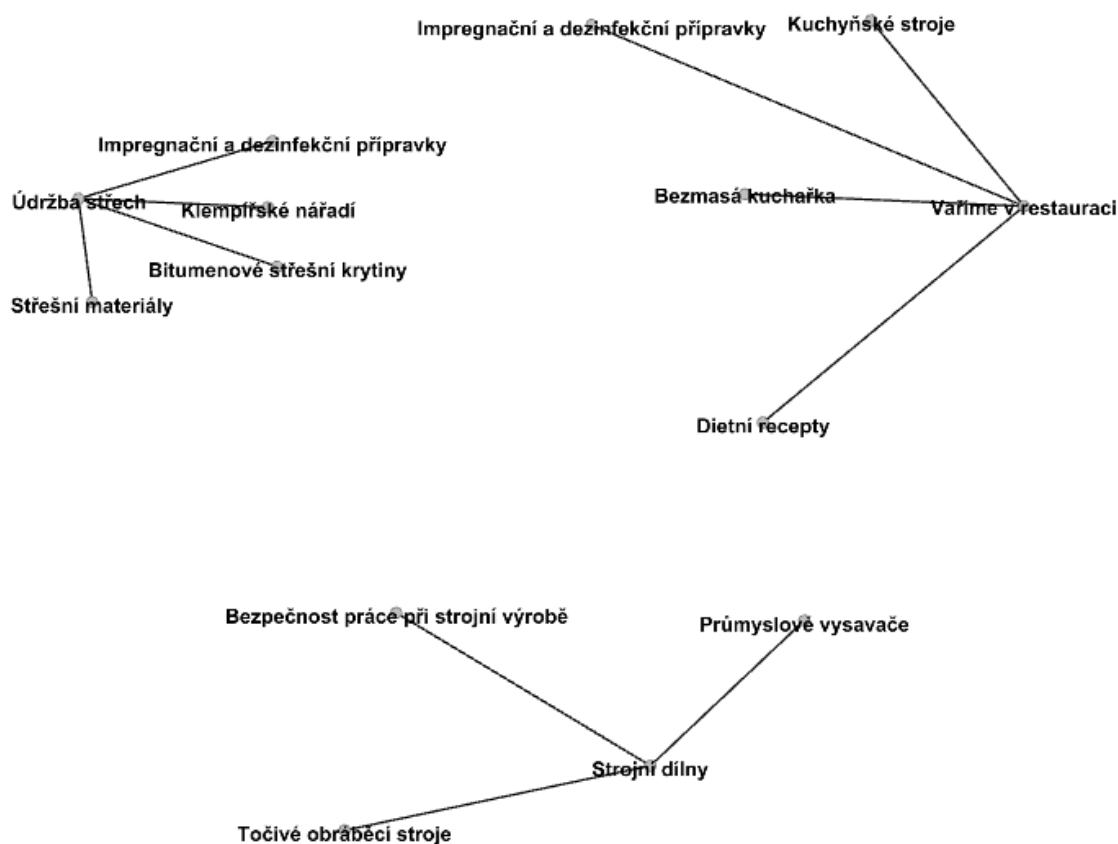


Zdroj: Vlastní zpracování

Stejným způsobem importujeme další CSV soubory, nastavíme popisky a propojíme zdrojové uzly s jejich odpovídajícími referenčními uzly.

Grafická reprezentace vypadá následovně.

Obrázek 17 – Vizualizace fiktivní sítě s kopií uzlu



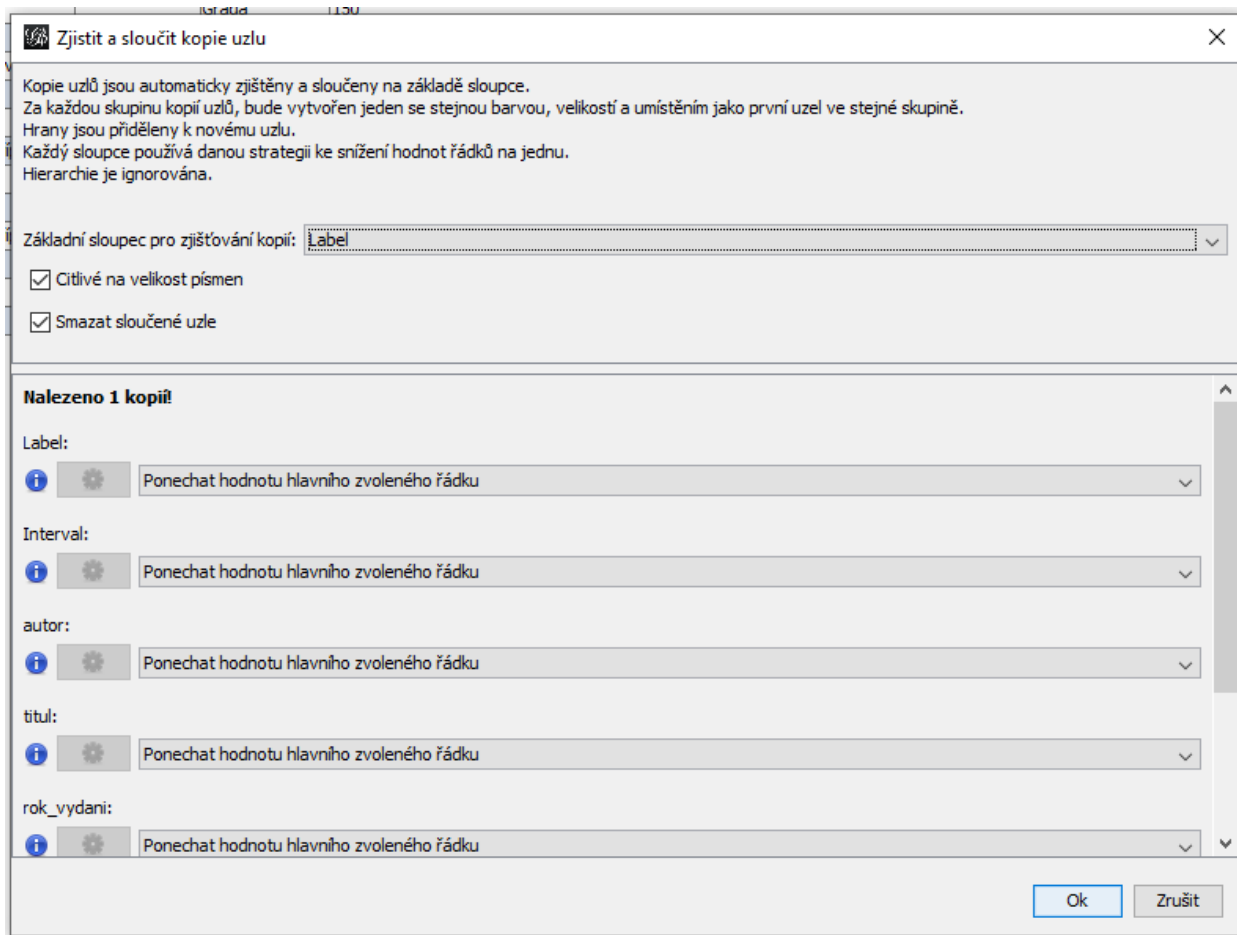
Zdroj: Vlastní zpracování

Uvedený příklad obsahuje duplicitní uzel. Je to reprezentace knihy Impregnační a dezinfekční přípravky.

Tuto duplicitu je žádoucí odstranit.

V *Laboratoři dat* vybereme tlačítko *Další činnosti* a po rozbalení volbu *Zjistit a sloučit kopie uzlu*.

Obrázek 18 - Odstranění kopií uzlů

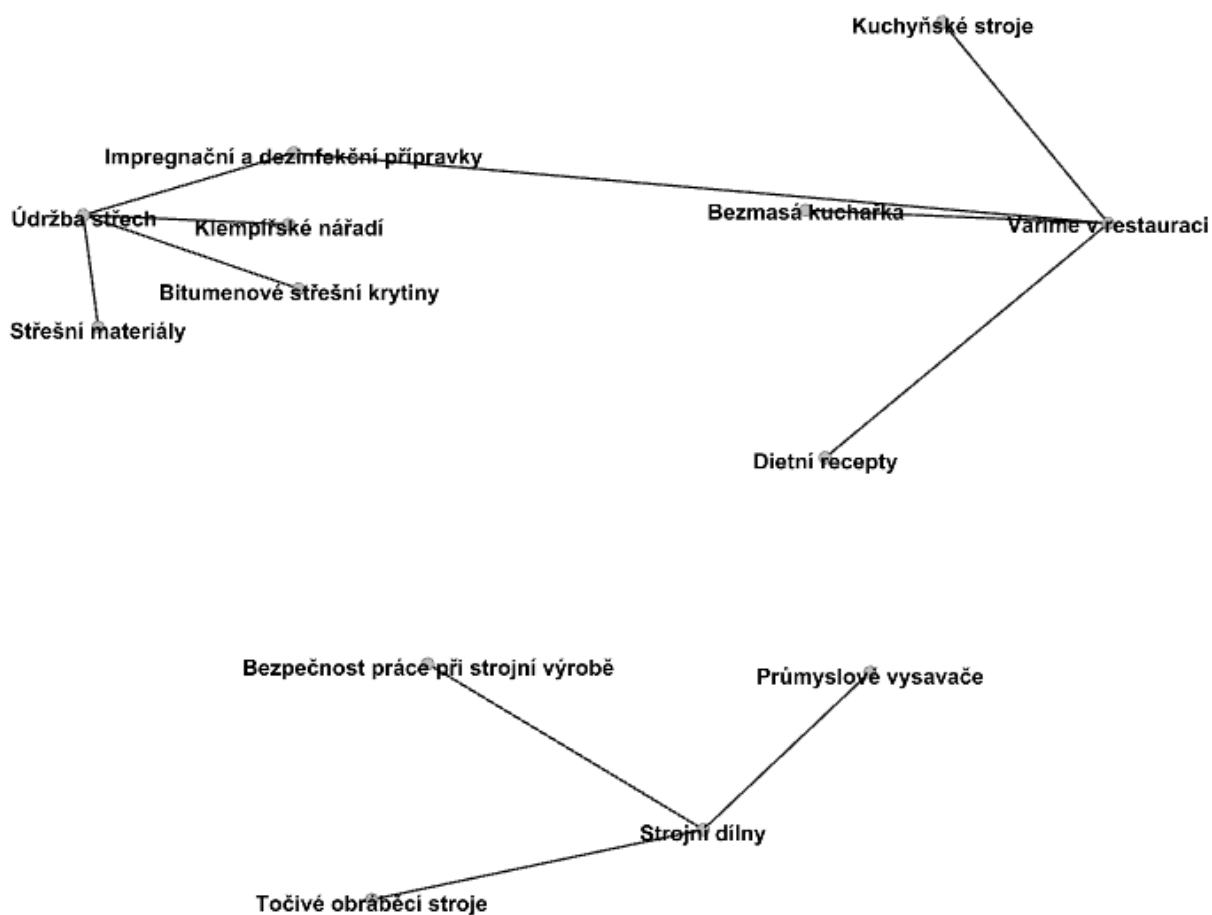


Zdroj: Vlastní zpracování

Po skončení operace Dephi vygeneruje nové unikátní *id*, překopíruje do této nové položky všechny původní atributy a původní duplicitní uzly smaže.

Graf nyní vypadá takto.

Obrázek 19 - Vizualizace očistěné fiktivní sítě



Zdroj: Vlastní zpracování

3.5 Formát souboru GEXF

GEXF (Graph Exchange XML Format) je jazyk, který popisuje komplexní strukturu sítě, s ní propojená data a dynamiku. [20]

Obrázek 20 - Jednoduchý graf v GEXF



Zdroj: Vlastní zpracování

Neorientovaný graf se 2 uzly a 1 hranou

Odpovídající popis v XML formátu je:

```
<graph defaultedgetype="undirected">
  <nodes>
    <node id="0" label=""/>
    <node id="1" label=""/>
  </nodes>
  <edges>
    <edge id="0" source="0" target="1"/>
  </edges>
</graph>
```

Nejdříve musí být definovány uzly pak teprve hrany. V tomto jednoduchém příkladu neřešíme barvy.

Mějme příklad grafu, který reprezentuje text „Toto je můj graf“. Každému uzlu přísluší jedno slovo. Vzhledem k tomu, že záleží na pořadí, bude tento graf orientovaný.

Obrázek 21 - Orientovaný graf v GEXF



Zdroj: Vlastní zpracování

Orientovaný graf se 4 uzly, 3 hranami a popisky uzlů.

Odpovídající popis v XML je:

```
<graph defaultedgetype="directed">
  <nodes>
    <node id="0" label="Toto"/>
    <node id="1" label="je"/>
    <node id="2" label="můj"/>
    <node id="3" label="graf"/>
  </nodes>
  <edges>
    <edge id="0" source="0" target="1"/>
    <edge id="1" source="1" target="2"/>
    <edge id="2" source="2" target="3"/>
  </edges>
</graph>
```

Je vidět, že každému uzlu zde přibyl jeho popis pomocí atributu *label*.

Úplný formát GEXF se skládá ze záhlaví, popisu uzlů a popisu hran.

Náš příklad by vypadal následně:

```
<?xml version="1.0" encoding="UTF-8"?>
<gexf xmlns="http://www.gexf.net/1.2draft"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.gexf.net/1.2draft
      http://www.gexf.net/1.2draft/gexf.xsd"
version="1.2">
<meta lastmodifieddate="2009-03-20">
  <creator>Gephi.org</creator>
  <description>Toto je můj graf file</description>
</meta>
<graph defaultedgetype="directed">
  <nodes>
    <node id="0" label="Toto"/>
    <node id="1" label="je"/>
    <node id="2" label="můj"/>
    <node id="3" label="graf"/>
  </nodes>
  <edges>
    <edge id="0" source="0" target="1"/>
    <edge id="1" source="1" target="2"/>
    <edge id="2" source="2" target="3"/>
  </edges>
</graph>
</gexf>
```

3.5.1 Záhlaví

Záhlaví je společné pro všechny GEXF dokumenty.

```
<?xml version="1.0" encoding="UTF-8"?>
<gexf xmlns="http://www.gexf.net/1.2draft"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.gexf.net/1.2draft
      http://www.gexf.net/1.2draft/gexf.xsd"
version="1.2">
<meta lastmodifieddate="2009-03-20">
  <creator>Gephi.org</creator>
  <description>Toto je můj graf file</description>
</meta>
```

První řádek je instrukce pro proces zpracování XML, kde je uvedeno, že odpovídá verzi 1.0 a znaky jsou kódovány v kódu UTF-8.

Další řádky záhlaví obsahují odkazy na jmenný prostor dokumentu a na definiční soubor XSD tohoto dokumentu. viz. Příloha č. 2.

Meta element informuje o datu poslední modifikace schématu a o tvůrci.

3.5.2 Topologie sítě

Sekce topologie sítě obsahující uzly a hrany se nazývá *graph*.

Element *graph* má vnořeny elementy uzlů *nodes* a elementy hran *edges*.

Element *nodes* obsahuje definice jednotlivých uzlů *node*, element *edges* obsahuje definice jednotlivých hran *edge*.

```
<graph defaultedgetype="directed">
  <nodes>
    <node id="0" label="Toto"/>
    <node id="1" label="je"/>
    ...
  </nodes>
  <edges>
    <edge id="0" source="0" target="1"/>
    ...
  </edges>
</graph>
```

Graf v GEXF může být orientovaný nebo neorientovaný. Přednastavená hodnota *defaultedgetype* je uvedena v záhlaví elementu *graph* a platí takto pro všechny hrany v tomto elementu. Je možno nastavit hodnoty *direct*, *undirect* a *mutual*.

Deklarace elementu *edges* musí následovat až po deklaraci elementu *nodes*.

3.5.3 Deklarace uzlu

Každý uzel *node* má identifikátor *id*, který musí být unikátní v celém souboru GEXF.

Typově je to string.

Uzel musí mít XML atribut *label*, typově string.

```
<node id="0" label="Toto"/>
```

3.5.4 Deklarace hrany

Každá hrana *edge* má identifikátor *id* a 2 koncové body *source* a *target*. Hodnoty koncových bodů ukazují na příslušné uzly, které musí být definovány v rámci souboru GEXF. Je možno definovat i smyčku, pak hodnota *source*=hodnota *target*. Uvedené položky jsou všechny typově string.

Každá hrana může mít volitelně další atributy:

- *label* ... popisek typově string
- *type* ... directed, undirected a mutual (směřovaný od zdroje k cíli a od cíle ke zdroji) typově string
- *weight* ... váha hrany, typově float

```
<edge id="0" source="0" target="1"/>
```

3.5.5 Data sítě

V praktických aplikacích využití grafů obvykle nestačí pracovat pouze se strohým grafickým vyjádřením sítě pomocí uzlů a hran. Informaci je třeba doplnit o další údaje ve formě připojených dat.

Data mohou být vnořena do atributů.

Každý atribut má pojmenování a hodnotu. Jméno atributu musí být deklarováno pro celý graf.

Hodnoty mohou být následujících typů: string, integer, double, float, boolean, date.

3.5.6 Deklarace atributů

Atributy jsou deklarovány uvnitř *attributes* elementu.

Třída XML-attribute je vnořena u uzlů jako *node* value, u hran jako *edge* value.

Příklad:

```
<graph defaultedgetype="undirected">
  <attributes class="node">                                ... deklarace atributů pro třídu
node
  <attribute id="0" title="url" type="string"/>            ... deklarace atributu
  0
  <attribute id="1" title="datum" type="date"/>            ...                               1
  <attribute id="2" title="je_v_archivu" type="boolean">    ...
  2
  <default>true</default>                                ... není-li atribut 2
  uveden, platí true
  </attribute>
</attributes>
<nodes>                                                  ... definice uzlů
  <node id="0" label="iDnes">                            ... uzel 0
  <attvalues>                                            ... hodnoty atributů pro uzel
0
  <attvalue for="0" value="http://www.idnes.cz"/>        ... hodnota atributu
0 pro uzel 0
```

```

    <attvalue for="1" value="2020-09-01"/> ... hodnota atributu
1 pro uzel 0
  </attvalues> ... hodnota atributu
2=default=true
  </node>
  <node id="1" label="Reporter"> ... uzel 1
    <attvalues>
      <attvalue for="0" value="https://reportermagazin.cz"/>
      <attvalue for="1" value="2020-09-01"/>
    </attvalues>
  </node>
  <node id="2" label="HospNoviny"> ... uzel 2
    <attvalues>
      <attvalue for="0" value="http://ihned.cz"/>
      <attvalue for="1" value="2020-09-01"/>
      <attvalue for="2" value="false"/>
    </attvalues>
  </node>
</nodes>
<edges> ... definice hran, atributy
pro hrany
  <edge id="0" source="0" target="1"/> ... nejsou
deklarovány
  <edge id="1" source="0" target="2"/>
  <edge id="2" source="1" target="2"/>
</edges>
</graph>

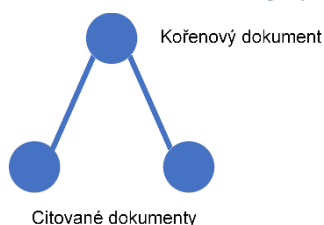
```

Praktický příklad:

Na tomto příkladu bude znázorněná vazba mezi dokumenty uloženými v citační databázi Scopus. Byl vybrán kořenový document a dva další dokumenty, které jsou v kořenovém dokumentu citovány.

Graficky lze vazbu dokumentů znázornit následovně:

Obrázek 22 - Příklad grafu v GEXF



Zdroj: Vlastní zpracování

Data všech tří dokumentů byla exportována z databáze Scopus do souborů *.CSV (Excel).

V databázi jsou všechna data dokumentů uložena strukturovaně a při exportu je tato struktura dodržena.

V souboru *.CSV se jedná o tabulku, kde první řádek je řádek názvů sloupců tabulky, tj. pojmenování jednotlivých položek dokumentu, další řádky pak nesou data jednotlivých dokumentů. Následně byla data importována do systému Gephi a uložena do jeho vnitřních struktur.

Poté byla data exportována do souboru *.GEXF. Vznikl soubor viz. příloha č. 1.

3.6 Externí prohlížeče grafů

Externí prohlížeče jsou aplikace třetích stran, které nám umožňují zobrazit grafy na základě dat z GEXF souborů. Některé obsahují funkcionality, které nám umožňují vyhledávat v grafech, zobrazují nám vlastnostu uzlů a jejich propojení pomocí hran. Většinou se jedná o prohlížeče, které jsou realizovány v rámci komunity uživatelů vizualizačních nástrojů.

Jedním z nich je Gephi JavaScript GEXF Viewer. Jedná se o aplikaci napsanou v JavaScriptu. Výhoda tohoto nástroje spočívá v tom, že při zobrazení grafu můžeme interaktivně sledovat plnou informaci, která se váže k zaměřenému uzlu a tento nástroj lze spustit v běžném internetovém prohlížeči. Na internetu se objevuje více těchto nástrojů, některé jsou placené a některé jsou volně dostupné například z GitHubu.

4 Vlastní práce

Vlastní práce se skládá ze dvou hlavních částí. První část této práce se zabývá výběrem vhodných dat, jejich exportem z databáze a následným zpracováním pro následující část diplomové práce.

Ve druhé části vlastní práce jsou data, která byla získána v první části práce importována do volně dostupného nástroje Gephi, ve kterém jsou upravena a následně je vyhotovena vizualizace sítě a je provedeno několik různých zobrazení sítě založených na algoritmech, které nástroj poskytuje.

Poté je proveden výběr nejvhodnější metody zobrazení pro naše data a na základě tohoto zobrazení je provedena analýza vědeckých publikačních činností s využitím několika metrik informační vědy.

4.1 Zdrojová data

Zdrojová data byla získána exportováním z databáze Scopus. Jedná se celkem o 150 publikací z oblasti počítačové vědy, které byly publikovány pod hlavičkou České zemědělské univerzity v Praze. Aktuálně se v rámci tohoto výběru celkem nachází 255 publikací, které nebylo možné v rámci této práce využít všechny z důvodu nedostatečného výkonu počítače, převážně operační paměti, pro následnou analýzu a vizualizaci.

Každá z těchto publikací obsahuje referenční publikace z různých vědeckých disciplín, celkem se jedná o 4639 citovaných publikací. V tomto vzorku dat jsou obsaženy citační a bibliografické informace, které nám tato databáze poskytuje.

4.1.1 Export dat z databáze

Pro export dat byl využit souborový tabulkový formát CSV, který lze otevřít v programu MS Excel a následně data importovat do prostředí Gephi. Příklad strukturůry tohoto formátu je uveden v teoretické části práce v kapitole 3.4.2.6.

V rámci exportu těchto dat bylo zapotřebí postupovat následovně, kdy nejprve byla exportována zdrojová publikace z výběru do separátního souboru a následovně byl proveden export referenčních publikací do dalšího souboru, které k této publikaci byly přidruženy v této databázi. Tento postup byl opakován pro všech 150 publikací, a to z důvodu přehlednosti pro další zpracování dat. Celkem tedy vzniklo 297 souborů,

neboť u 3 publikací, které při následném zpracování v programu Gephi byly označeny jako „seed“, nebyly nalezeny žádné reference.

4.1.2 Import dat do Gephi

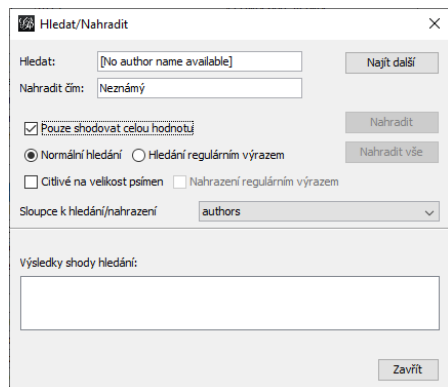
Vkládání dříve získaných dat z databáze Scopus bylo provedeno postupně po jednotlivých souborech. Nejprve byl importován soubor obsahující zdrojovou publikaci a následovně soubor referenčních publikací. Dále bylo provedeno propojení zdrojového a cílových úzlů pomocí funkce, kterou nám nástroj Gephi umožňuje pro automatické vytvoření tabulky hran. Tento postup importu dat je znázorněn na příkladu v kapitole (3.4.3.3).

Pro následnou přehlednost byl přidán sloupec „seed“ obsahující logický datový typ boolean a v prostředí Gephi je znázorněn jako checkbox, který nabývá hodnoty true/false. Pro každou zdrojovou publikaci, byl v tomto sloupci tento checkbox zaškrtnut, tedy sloupec „seed“ obsahuje hodnotu true, a to z důvodu označení zdrojové publikace. U referenčních publikací zůstává hodnota v tomto sloupci false, resp. je checkbox prázdný. Dále je tento sloupec využíván při vizualizaci.

4.1.3 Úprava dat

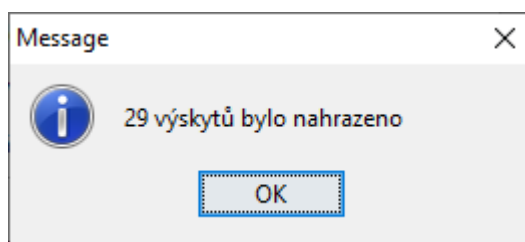
V části tohoto vzorku dat chybí některé záznamy, a to i včetně samotného autora. Tento výskyt se objevil u 29 publikací. Ve většině případů se ve sloupci „authors“ nacházela hodnota null nebo hodnota [No author name available]. Bylo tedy zapotřebí tyto dvě rozdílné hodnoty shrnout pod 1, aby bylo možné s těmito daty dále pracovat. Tyto hodnoty byly nahrazeny hodnotou „Neznámý“.

Obrázek 23 - Úprava dat



Zdroj: Vlastní zpracování

Obrázek 24 - Nahrazení dat

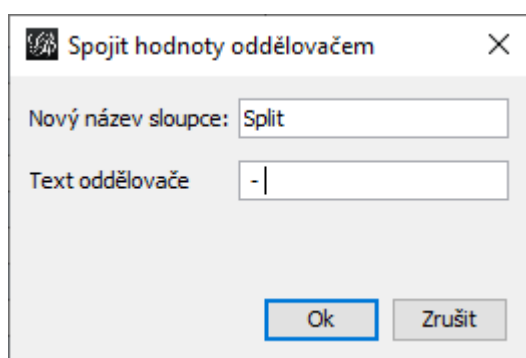


Zdroj: Vlastní zpracování

Dále se ve vzorku dat, celkem tedy 4789 publikací, vyskytlo několik duplicit. Tento jev vznikl během sběru dat, kdy zdrojové publikace odkazovaly na stejnou referenční publikaci, ale také v případě, kdy zdrojová publikace se stala publikací referenční. Tento jev se vyskytl v 205 případech. Odstranění tohoto jevu je popsáno na příkladu již v teoretické části v kapitole 3.4.3.3.

Nyní bylo zapotřebí vyplnit pole **Label** smysluplným textem, aby bylo možno rozpoznat, kterého dokumentu se uzel týká. Bylo zvoleno spojení polí **authors** a **title** s oddělovačem, který jsme si definovali. Do databáze byl přidán sloupec **Split**

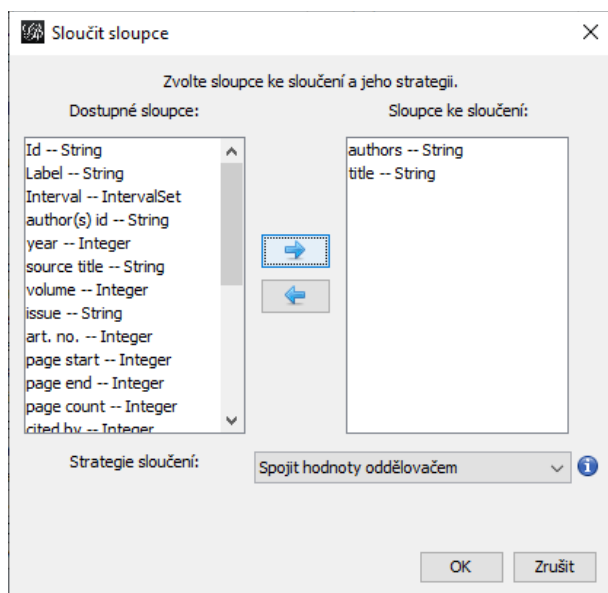
Obrázek 25 - Vytvoření sloupce a spojení hodnot s oddělovačem



Zdroj: Vlastní zpracování

a do tohoto sloupce byla kopírována tato složenina.

Obrázek 26 - Spojení hodnot do jednoho sloupce



Zdroj: Vlastní zpracování

Následně byl sloupec **Split** kopírován do sloupce **Label**.

4.2 Vizualizace

Vizualizace dat se v prostředí Gephi vytváří automaticky v náhodném rozvržení. V rámci této části práce je zobrazeno několik různých zobrazení na základě algoritmů a vybráno nejvhodnější pro použitá data. Toto zobrazení je dále využíváno v rámci analýzy.

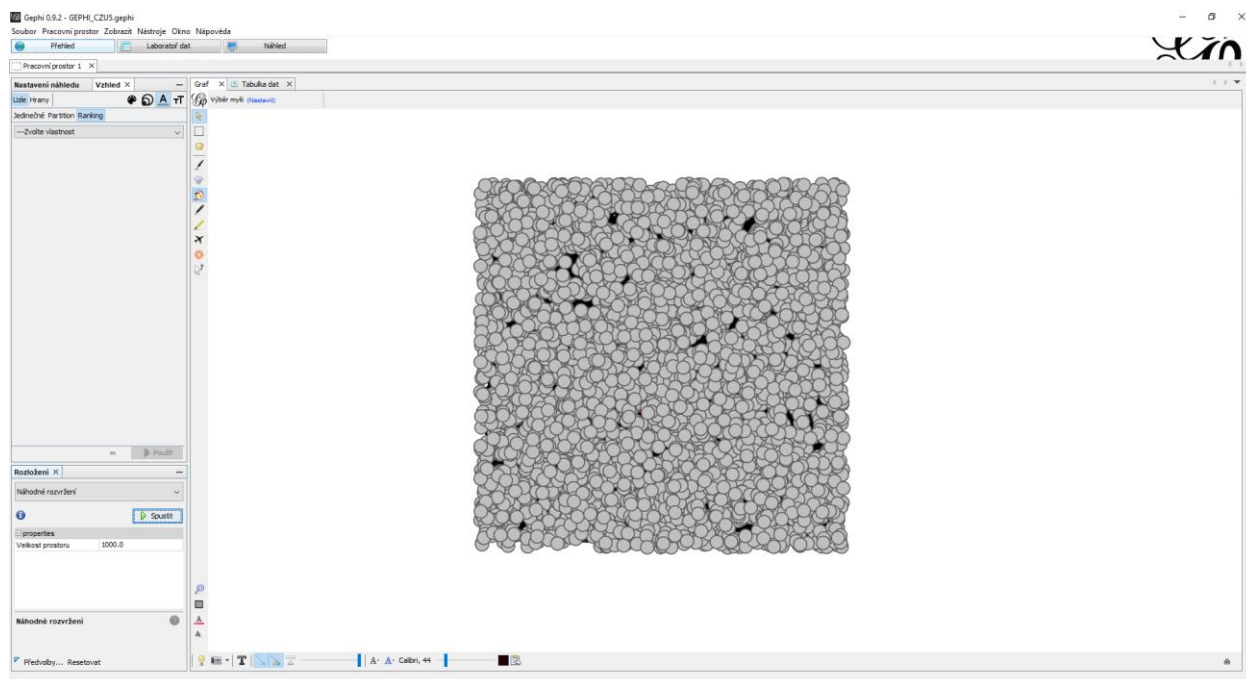
4.2.1 Importovaná data

Po importu dat je třeba udělat některé úpravy v samotných uzlech a dále zvolit vhodnou metodu zobrazení grafu, aby bylo možno dále s grafem pracovat a získávat požadované informace.

Data jsou natažena do vnitřní databáze Gephi a všechny referencované položky (cíle) jsou prolinkovány na zdrojovou publikaci (zdroj).

Pokud nyní provedeme základní zobrazení grafu, vypadá to následovně. Rozložení uzlů grafu je náhodné.

Obrázek 27 - Náhodná vizualizace grafu



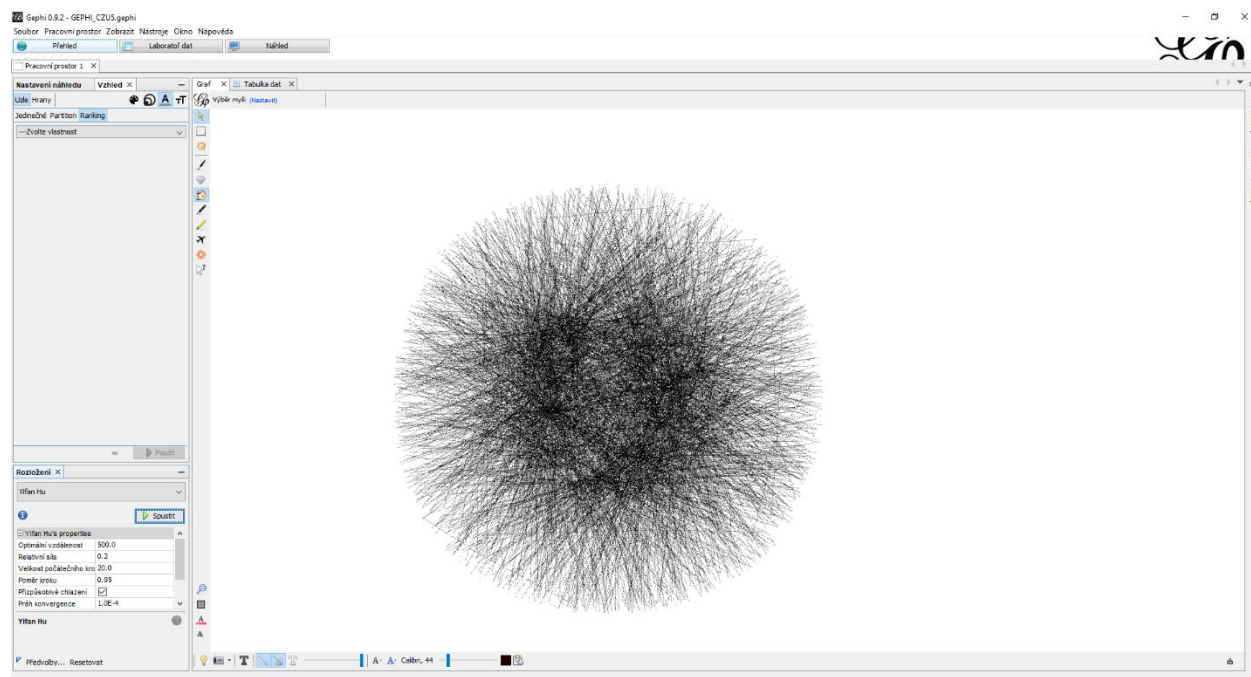
Zdroj: Vlastní zpracování

Dále je třeba zvolit metodu výpočtu rozložení grafu.

4.2.2 Yifan Hu zobrazení

Při volbě *Yifan Hu*, vypadá zobrazení následovně.

Obrázek 28 - Yifan Hu zobrazení



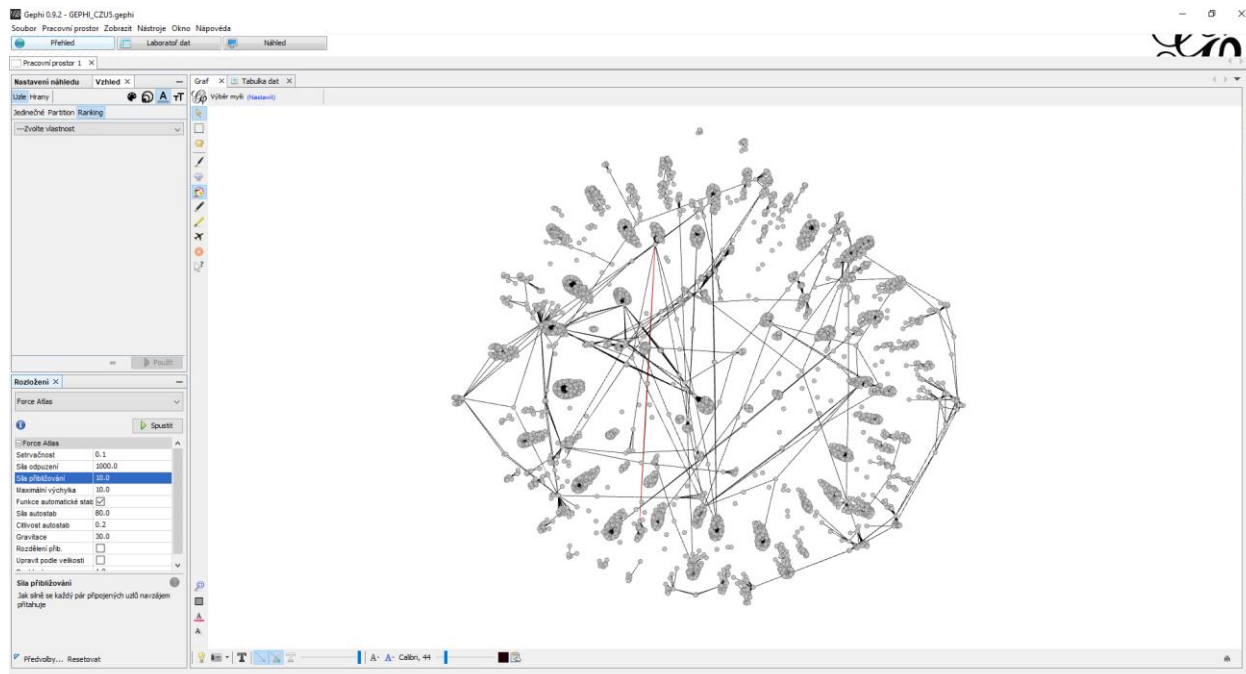
Zdroj: *Vlastní zpracování*

Je vidět, že pro naše data je toto zobrazení nevhodné.

4.2.3 Force Atlas zobrazení

Při nastavení metody *Force Atlas* vypadá nastavení následovně.

Obrázek 29 - *Force Atlas* zobrazení



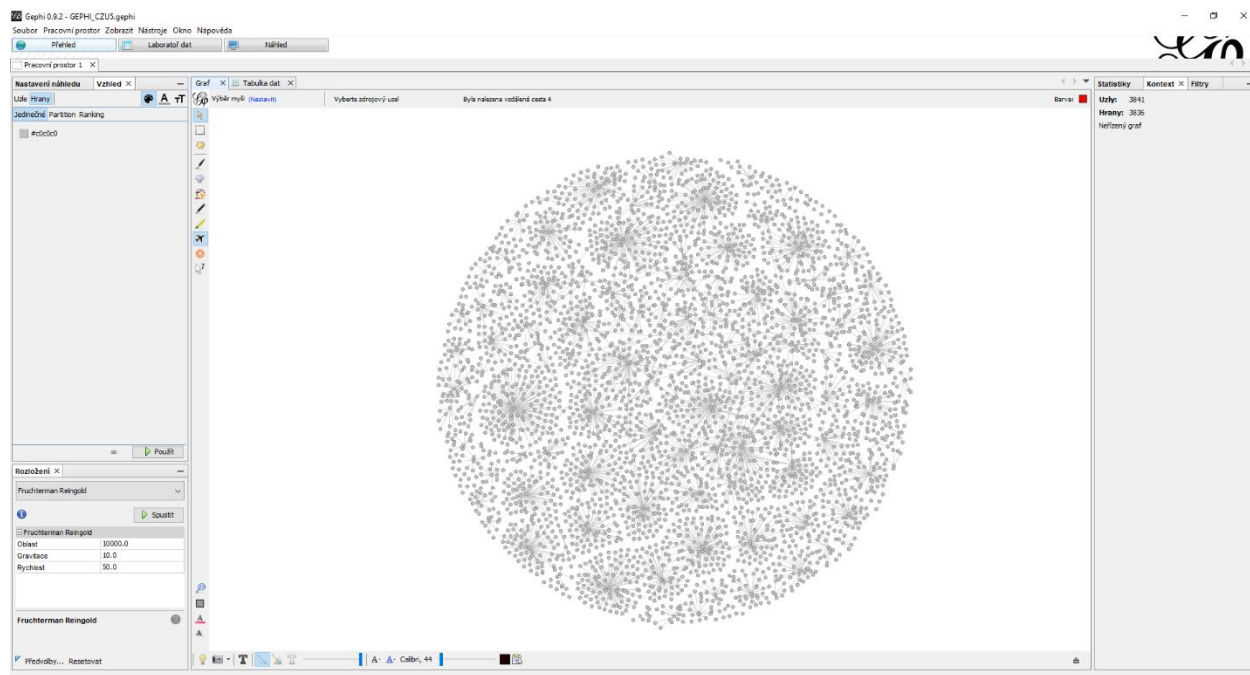
Zdroj: *Vlastní zpracování*

Ani toto zobrazení není pro naše účely ideální. Ukázalo se, že nejvíce vypovídající grafy jsou tvořeny pomocí algoritmu *Fruchtermana Reingolda*.

4.2.4 Fruchterman Reingoldovo zobrazení

Tento graf již má určitou vypovídací schopnost. Jsou zde vidět vazby mezi zdrojem a cílem

Obrázek 30 - Fruchterman Reingoldovo zobrazení



Zdroj: Vlastní zpracování

Aby bylo možné s daty dále pracovat, provedeme úpravu databáze tak, aby se nám ještě více opticky oddělily zdrojové publikace od referenčních cílových publikací.

Využíváme toho, že v rámci importu každé zdrojové publikace jsme tuto publikaci označili zaškrtnutím check boxu ve sloupci „seed“.

Obrázek 31 – Označení zdrojových dat (uzlů)

Id	Label	authors	author(s) id	title	year	source title	cited by	doi	link	ism	isbn	Seed	Split	Eccentricity	Closeness Ce...	Harmonic Closen...	Betweenness C...	Clustering Coeff...	Number of triangles
5029	Information sci...	Vesely A., ...	6602163241	Information sci...	2014	Neural Network World		10.14319/1NNW...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00004	0.0	0	
5038	The experient...	Kozak J., Bor...	861067700	The experient...	2014	GlobalSoft: Basis of...	1	10.1018/14500...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00004	0.0	0	
5050	Object orient...	Tóth J., Mar...	553253400	Object orient...	2014	Advances in Intelligen...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00044	0.0	0	
4424	Non-linearity	Papez V., Pap...	1208124600	Non-linearity	2015	2015 12th Internation...		10.1019/TELKS...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00001	0.0	0	
4432	Transformers	Papez V., Pap...	2408124600	Transformers	2015	2015 12th Internation...		10.1019/TELKS...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00001	0.0	0	
4437	Design and re...	Novák J., Šte...	2447229000	Design and re...	2015	2015 International Wo...		10.1019/IVCM...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00014	0.0	0	
4454	Smart Cities a...	Phelan T., Sm...	5620557000	Smart Cities a...	2015	2015 Smart Cities Sym...	3	10.12019/SCSP...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00007	0.0	0	
4472	Behavioral an...	Vorl J., Fran...	5618844700	Behavioral an...	2015	2015 Smart Cities Sym...	2	10.1019/SCSP...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00003	0.0	0	
4495	KORPI: An ef...	Kubelka K., M...	5629965700	KORPI: An ef...	2015	Computers and Electri...	2	10.1016/j.compe...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00041	0.0	0	
4522	Evaluation of...	Russo D., Val...	5533689800	Evaluation of...	2015	International Journal o...	54	10.3390/ijer50...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00025	0.0	0	
4586	The reliability	Kouba M., Bar...	5194471000	The reliability	2015	Ecological Informatic...		10.1016/j.ecoinf...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00031	0.0	0	
4616	Genetic algor...	Kozney V.	5646037200	Genetic algor...	2015	Expert Systems with A...	36	10.1016/j.eswa...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00081	0.0	0	
4655	GIS tool for...	Vopelka P., K...	5718962000	GIS tool for...	2015	Computers and Electri...	11	10.1016/j.compe...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00012	0.0	0	
4674	Open data in...	Ušanič M., Šte...	5531650900	Open data in...	2015	CEUR Workshop Proce...		https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00014	0.0	0	
4688	Markup lang...	Masner J., Va...	5195040700	Markup lang...	2015	CEUR Workshop Proce...	1	https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		0.117833	0.155382	0.0012	0.0	0	
4712	Cross-platfor...	Stoček M., Ma...	5527566500	Cross-platfor...	2015	Proceedings of the 11...		https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		0.150033	0.20159	0.0099	0.0	0	
4727	Information a...	Ušanič M., Tyr...	5531650900	Information a...	2015	Future Communication...	1	10.1018/18049...	https://www.sc...			<input checked="" type="checkbox"/>		0.229273	0.295621	0.01621	0.0	0	
4746	Spatial and ac...	Masner J., St...	5550590700	Spatial and ac...	2015	Future Communication...		10.1018/18049...	https://www.sc...			<input checked="" type="checkbox"/>		0.130186	0.172779	0.00179	0.0	0	
4759	Use of IT in...	Stoček M., Va...	5527566500	Use of IT in...	2015	Future Communication...		10.1018/18049...	https://www.sc...			<input checked="" type="checkbox"/>		0.121885	0.195537	0.00046	0.0	0	
4783	Mobile applic...	Stoček M., Va...	5527566500	Mobile applic...	2015	Future Communication...	2	10.1018/18049...	https://www.sc...			<input checked="" type="checkbox"/>		0.188511	0.235496	0.01148	0.0	0	
4799	Systems theo...	Habich C., Vo...	5056841525	Systems theo...	2015	Lecture Notes in Comp...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00001	0.0	0	
4807	Data modelin...	Tyrtyřák J., V...	5718422149	Data modelin...	2015	Advances in Intelligen...	1	10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.193604	0.28883	0.00029	0.0	0	
4830	Design of me...	Novák J., Šte...	8246646000	Design of me...	2015	International Journal o...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.17467	0.23035	0.00128	0.0	0	
4845	Riemann integ...	Molnář F., Mo...	5598095500	Riemann integ...	2015	APLIMAT 2015 - 14th...	2	https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.0	0.0	0	
4850	Study on proc...	Gojša O., Ha...	5658407500	Study on proc...	2015	International Journal o...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.216216	0.333344	0.00689	0.0	0	
4912	The role of s...	Manurka V., Th...	6598262926	The role of s...	2015	Communications in Co...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.17467	0.23035	0.00128	0.0	0	
4930	Critical eval...	Manurka V., Th...	6598262926	Critical eval...	2015	Advances in Intelligen...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.17467	0.23035	0.00128	0.0	0	
4949	Solving the ric...	Kuraz M.	3528070600	Solving the ric...	2015	Civil-Camp Proceedings		https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		0.502747	0.70455	0.00064	0.0	0	
4974	Information s...	Nedvedova K...	56349438700	Information s...	2015	International Archives...		10.5146/arsarc...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00001	0.0	0	
4981	Using legacy s...	Gomez C., Gh...	55485047400	Using legacy s...	2015	International Archives...		10.5146/arsarc...	https://www.sc...			<input checked="" type="checkbox"/>		0.159505	0.19038	0.00273	0.0	0	
5006	Advanced pra...	Bartolák J., D...	1291474800	Advanced pra...	2015	World		10.1016/j.worl...	https://www.sc...			<input checked="" type="checkbox"/>		0.127015	0.149494	0.00046	0.0	0	
5001	Internet addi...	Host'ovecký J...	57021815400	Internet addi...	2016	ICTEA 2016 - 14th IEEE...		10.1019/ICTEA...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00003	0.0	0	
5018	Advanced cha...	Žák J., Vach M...	56339050000	Advanced cha...	2016	2016 IEEE Radar Meth...	2	10.1109/RADW...	https://www.sc...			<input checked="" type="checkbox"/>		0.571429	0.708333	0.0	0.0	0	
5026	Use of bio-log...	Fantler M.S., ...	23978630000	Use of bio-log...	2016	Animal Biotechnology	8	10.1016/j.anbio...	https://www.sc...			<input checked="" type="checkbox"/>		0.706349	0.895951	0.00593	0.0	0	
5030	Productivity...	Bartolák J., D...	1291474800	Productivity...	2016	Applied Categorica SE...	3	10.10176/14M5...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00004	0.0	0	
5113	Use of differ...	Žák J., Vach M...	56339050000	Use of differ...	2016	2016 28th Internation...	1	10.1109/AICAD...	https://www.sc...			<input checked="" type="checkbox"/>		0.571429	0.708333	0.0	0.0	0	
5122	Using e-learning...	Svoboda R., J...	45162107600	Using e-learning...	2016	Smart Innovation, Syst...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00003	0.0	0	
5135	Graphs in eve...	Molnář F., Mo...	55980955000	Graphs in eve...	2016	APLIMAT 2016 - 15th...		https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		0.0	0.0	0.0	0.0	0	
5169	Feasibility of...	Nováková V., ...	56460372000	Feasibility of...	2016	Advances in Intelligen...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.149613	0.193788	0.00128	0.0	0	
5226	Forecasting S...	Maca P., Pech...	35339730500	Forecasting S...	2016	Computational Intellig...	17	10.1155/2016/3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.000153	0.0	0	
5278	Determinants...	Pechová M., L...	55974269000	Determinants...	2016	ISDMAT 2016 - Informa...	1	https://www.sc...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00004	0.0	0	

Zdroj: Vlastní zpracování

Nyní provedeme úpravu v **Laboratoři dat**, kdy označíme celou skupinu zdrojů, které máme označeny taktéž hodnotou „true“ ve sloupci „seed“ a dále můžeme hromadně s touto skupinou manipulovat. Je možné např. označit uzly barevně, nebo jinou velikostí apod.

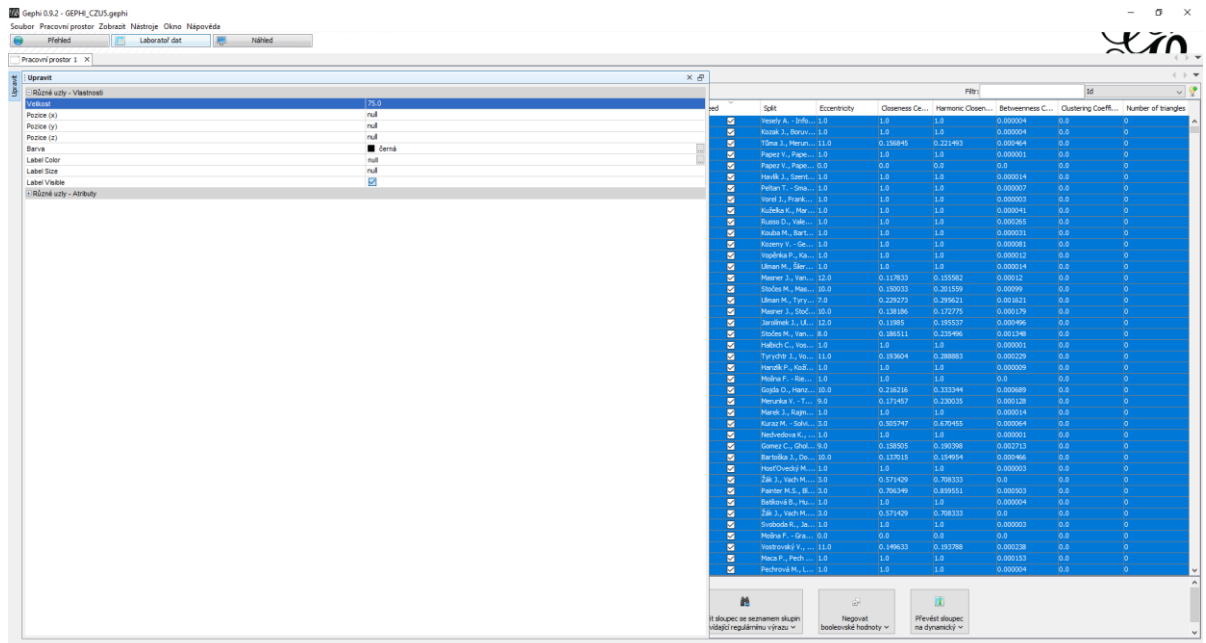
Obrázek 32 - Úprava zdrojových uzlů

Id	Label	authors	author(s) id	title	year	source title	cited by	doi	link	ism	isbn	Seed	Split	Eccentricity	Closeness Ce...	Harmonic Closen...	Betweenness C...	Clustering Coeff...	Number of triangles
2449	IT Ethics Rec...	Ušanič M., Ma...	5531650900	IT Ethics Rec...	2019	Journal of Computer In...		10.1008/108744...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00072	0.0	0	
2445	The intercon...	Manurka V., Th...	6598262926	The intercon...	2019	International Journal o...		10.1008/108744...	https://www.sc...			<input checked="" type="checkbox"/>		0.206005	0.26919	0.00870	0.00567	0	
2450	A versatile Va...	Mayer F., Heid...	570915209	A versatile Va...	2018	Lecture Notes in Comp...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		0.503407	0.769319	0.00169	0.0	0	
2566	Ecology and ...	Nedelcu V. N...	3275496300	Ecology and ...	2019	IEEE Transactions on L...	3	10.1109/ACCESS...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00058	0.0	0	
2717	Large exper...	Brown P., Zhu...	3700794300	Large exper...	2019	Databases	1	10.1016/j.dbs...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00035	0.0	0	
103	Quality perc...	Madry P., Dv...	5646037200	Quality perc...	2018	Advances in Intelligen...		10.1016/j.ecoinf...	https://www.sc...			<input checked="" type="checkbox"/>		0.129712	0.201832	0.01817	0.01367	0	
103	Eye-tracking	Andor V., Cse...	37211887900	Eye-tracking	2018	Advances in Intelligen...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00007	0.0	0	
1305	Open data a...	Vasíček V., ...	36588280100	Open data a...	2020	Advances in Intelligen...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00004	0.0	0	
1318	Dominance on...	Pešk M., Dem...	2398180000	Dominance on...	2020	Fuzzy Sets and System...		10.1016/j.fss...	https://www.sc...			<input checked="" type="checkbox"/>		0.80962	0.30273	0.00257	0.0	0	
1241	Loose meta-...	Fratiello V., ...	57219548300	Loose meta-...	2020	International Journal o...		10.1016/j.ijis...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00021	0.0	0	
1271	Generators of...	Navaia M., Pe...	3550275500	Generators of...	2020	Studies in Computatio...		10.1007/978-3...	https://www.sc...	18009491		<input checked="" type="checkbox"/>		0.146755	0.29097	0.00052	0.0	0	
1303	Influence of...	Šedl K., Kráží...	3017727800	Influence of...	2020	International Journal...		10.1016/j.ijis...	https://www.sc...			<input checked="" type="checkbox"/>		0.827461	0.796475	0.00038	0.0	0	
1315	Sensitivity of...	Madry P., Dv...	5646037200	Sensitivity of...	2020	International Journal o...		10.1016/j.ecoinf...	https://www.sc...			<input checked="" type="checkbox"/>		0.224972	0.263265	0.02515	0.0	0	
1395	Fighting Bas...	Wojcik M., S...	37212041400	Fighting Bas...	2020	Communications in Co...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00031	0.0	0	
1408	Microfit: Tool...	Žák J., Vach M...	4616161400	Microfit: Tool...	2020	Lecture Notes in Comp...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		3.84 M.	0.00011	0.0	0.0	0	
1411	The ubiquity...	Manurka V., Th...	6598262926	The ubiquity...	2020	Journal of Computatio...		10.1016/j.jco...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00072	0.0	0	
140	Virtual res...	Sappogno L., ...	37123195140	Virtual res...	2020	Ecological Informatic...		10.1016/j.ecoinf...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00039	0.0	0	
411	Wireless sen...	Taheri F., Dli...	5699958500	Wireless sen...	2020	Computers and Electri...		10.1016/j.compe...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.00018	0.0	0	
467	Surface qual...	Razavi P., Ga...	37212041400	Surface qual...	2020	International Journal...		10.1007/978-3...	https://www.sc...			<input checked="" type="checkbox"/>		1.0	1.0	0.0012			

Je třeba zdůraznit, že tyto úpravy slouží pouze pro optické rozlišení, na strukturu grafu nemají vliv.

Provedeme barevnou úpravu uzlů „otců“.

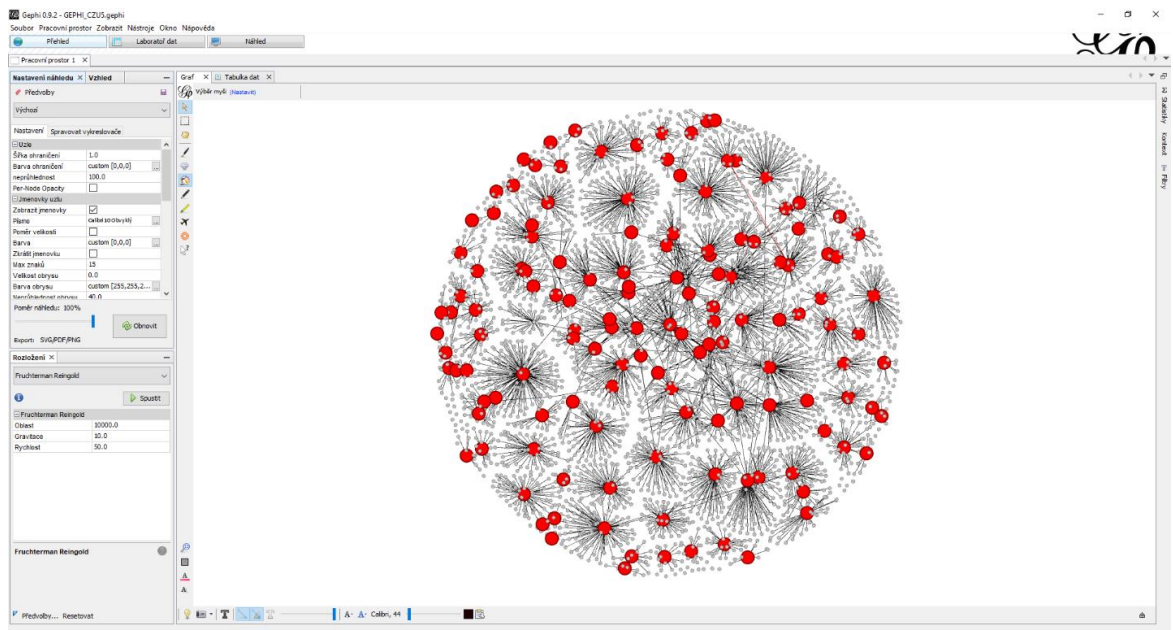
Obrázek 33 - Zvýraznění zdrojových uzlů



Zdroj: Vlastní zpracování

Zdrojové uzle jsme zvýraznili červenou barvou a jejich velikost nastavili na hodnotu 75.

Obrázek 34 - Zvýrazněné zdrojové uzle v grafu



Zdroj: Vlastní zpracování

4.3 Analýza

V této části je realizováno několik analytických příkladů. Postupně byly měněny hodnoty filtrů a poté zobrazeny výsledky, které bylo možno volně procházet a zkoumat jednotlivé položky. Analytické nástroje poskytují statistické informace v souhrnech.

4.3.1 Nalezení nejkratší cesty mezi dvěma uzly.

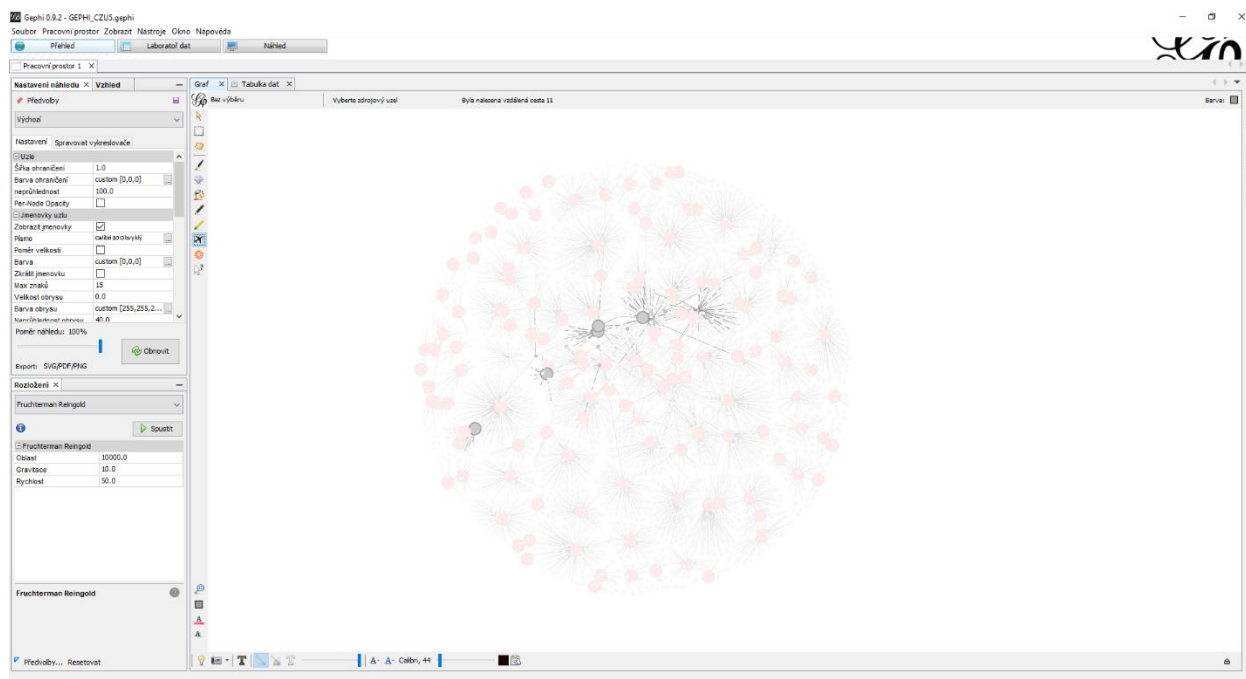
Jedná se o úlohu, kde máme vybraný uzel a hledáme jeho propojení na jiný uzel. Samozřejmě úloha nemusí mít vždy řešení, tj. cílový uzel je na “izolovaném ostrově”. Je tedy úlohou najít k vybranému uzlu jeho zdroj/cíl, k tomuto zdroji jeho zdroj/cíl, atd.

Jinak řečeno najít k vybranému dokumentu dokumentu, který primární document cituje, případně document, který je primárním dokumentem citován, atd.

Takovou úlohu umí Gephi řešit.

Příklad:

Obrázek 35 - Nejkratší cesta mezi zvolenými uzly

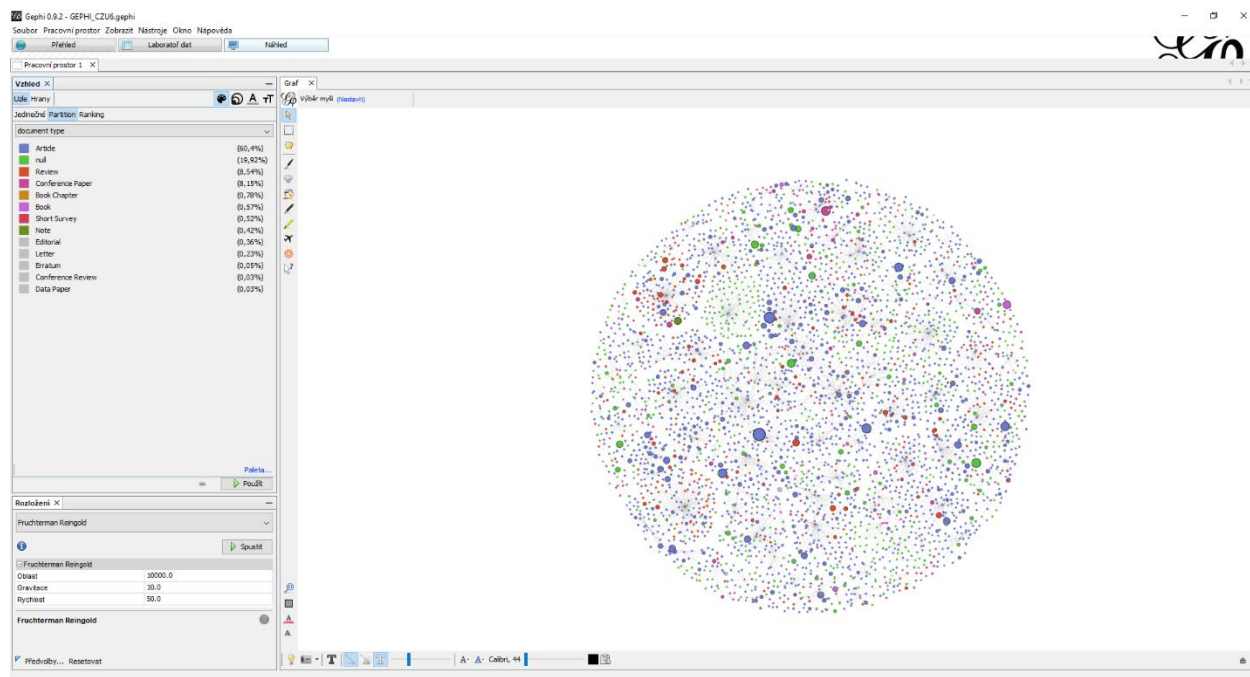


Zdroj: Vlastní zpracování

Uzly na nalezené cestě jsou graficky zvýrazněny. Nalezená cesta se skládá z 11 hran. Lze tedy říct, že zdrojová publikace je propojena s cílovou publikací přes 9 jiných publikací.

4.3.2 Rozdělení dokumentů dle jejich typu a citací.

Obrázek 36 - Graf rozdělení na základě typu publikace



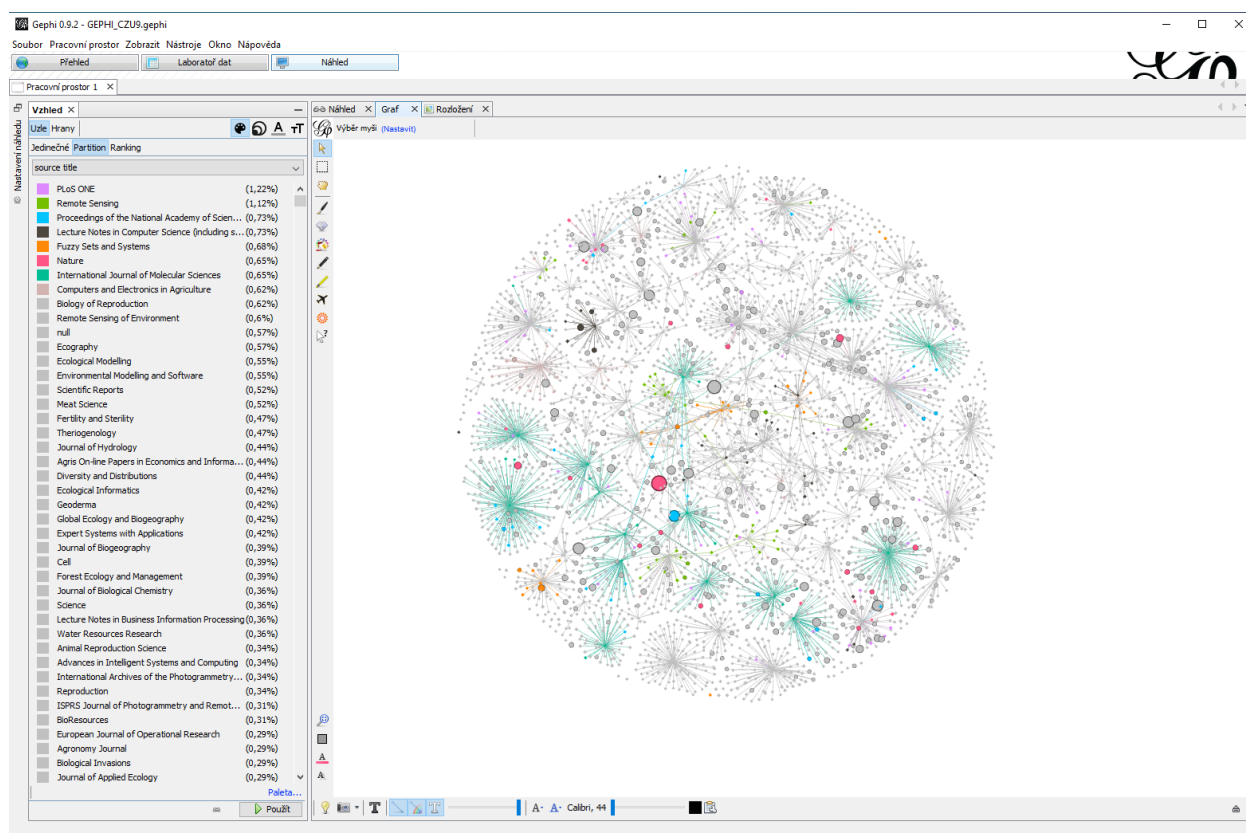
Zdroj: Vlastní zpracování

Zde velikosti uzlů určují počet citací a barvy odlišují typy dokumentů. V levé části obrazovky vidíme statistický souhrn.

Lze vidět, že nejvíce jsou zastoupeny články (60,4%). Současně vidíme, že 19,92% dokumentů nemá nastavený typ.

4.3.3 Hodnocení časopisů dle počtu citovaných článků

Obrázek 37 - Graf zobrazující množství citování jednotlivých časopisů



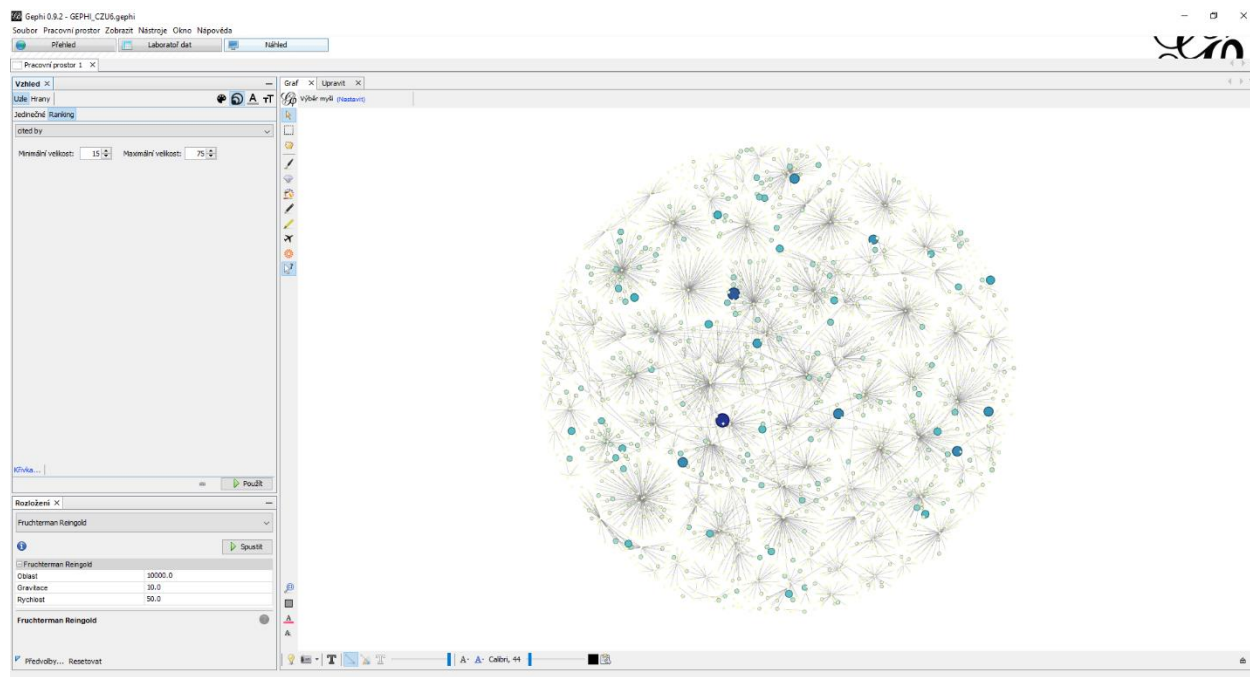
Zdroj: Vlastní zpracování

Vidíme, že nejvíce citovaných článků (1,22%) obsahuje časopis Plos ONE (<https://journals.plos.org/plosone/>) – fialová barva. Ale současně s ohledem na velikost příslušných uzlů jsou tyto články jsou málo citovány.

Lze tedy usuzovat, že v tomto časopise je publikována spíše kvantita než kvalita. Jedná se o online časopis, do kterého může veřejně publikovat své vědecké činnosti kdokoliv. Stačí si zde založit uživatelský účet.

4.3.4 Dokumenty dle počtu citací.

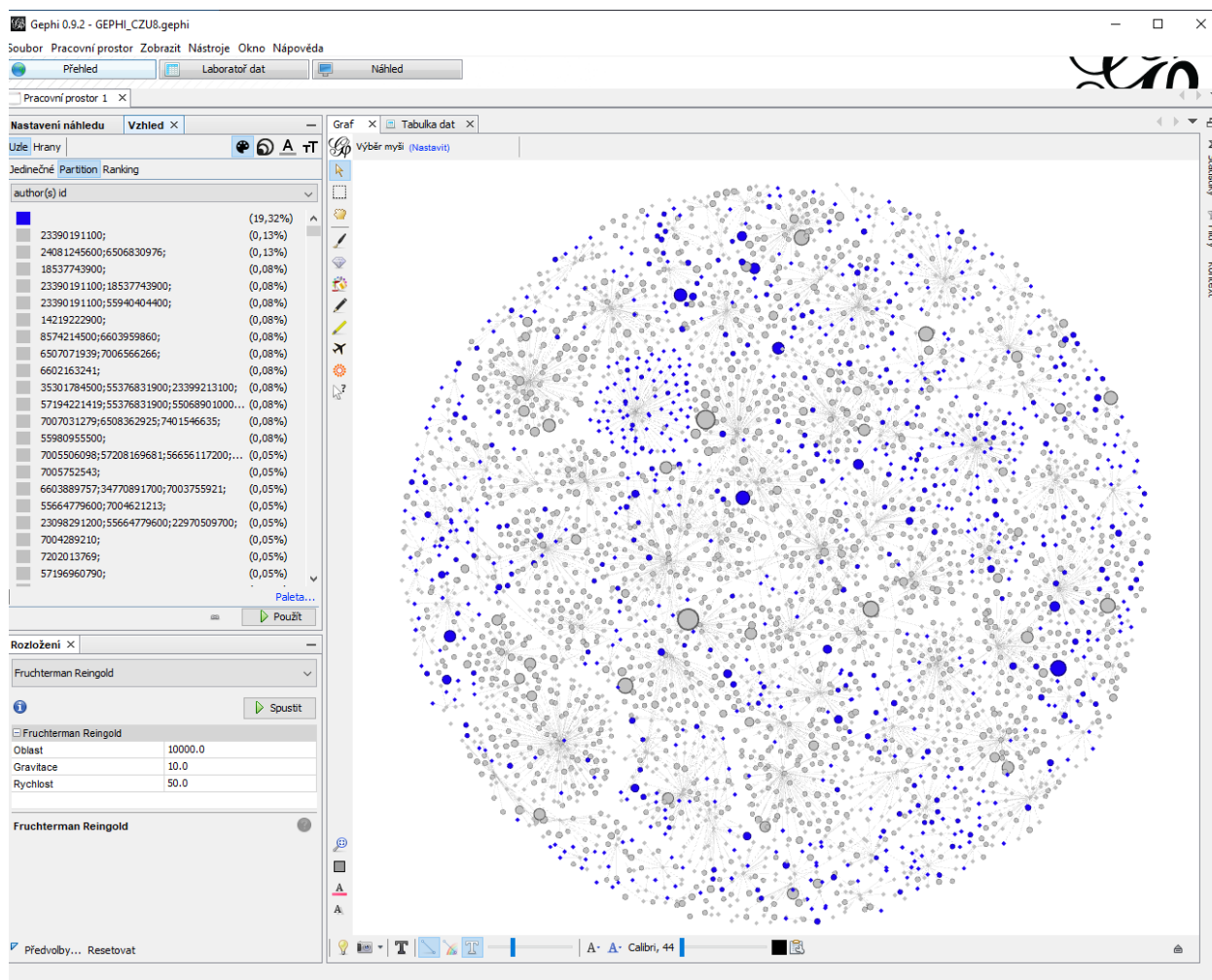
Obrázek 38 - Graf velikosti uzlů na základě citovanosti publikace



Zdroj: Vlastní zpracování

Velikost uzlů určuje, kolikrát byl příslušný dokument citován. Lze tedy vidět, že se v našem výběru dat nachází relativně dost kvalitních publikací, které jsou citovány.

Obrázek 39 - Graf citovanosti neznámých autorů



Zdroj: Vlastní zpracování

V tomto grafu lze vidět, že část z velmi citovaných publikací nemá uvedeno ID autora, může se tak jednat o chybu v citačním rejstříku, kdy byl autor špatně zadán nebo není autor v tomto rejstříku vůbec evidován. Jedná se celkem o 19,32% autorů z našeho výběru.

4.4 Zobrazení grafu v externím prohlížeči

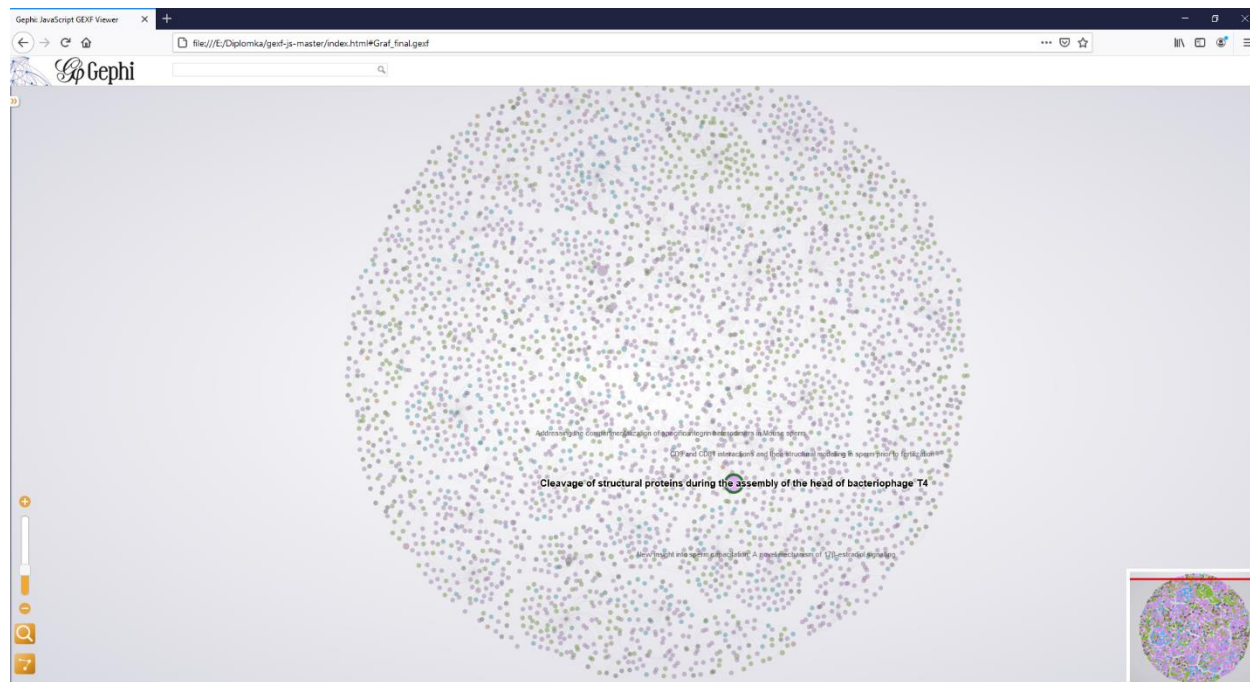
Zobrazení grafu v externím prohlížeči Gephi JavaScript GEXF Viewer, ve kterém lze interaktivně zobrazovat informace o zvoleném uzlu, vyhledávat a mít komplexní přehled o struktuře a velikosti celé sítě.

4.4.1 Interaktivní zobrazení

V přehledovém okně můžeme vyhledávat informace např. podle velikosti uzlů, nebo podle barev uzlů, které jsme si v rámci analýzy nastavili.

V pravém dolním rohu je miniatura celkové sítě, kde se ukazuje, ve které části sítě se pohybujeme. Při pohybu přes uzly se vždy zobrazuje popis uzlu. Pokud se zaměříme na konkrétní uzel, získáme detaily uzlu.

Obrázek 40 - Interaktivní zobrazení grafu GEXF

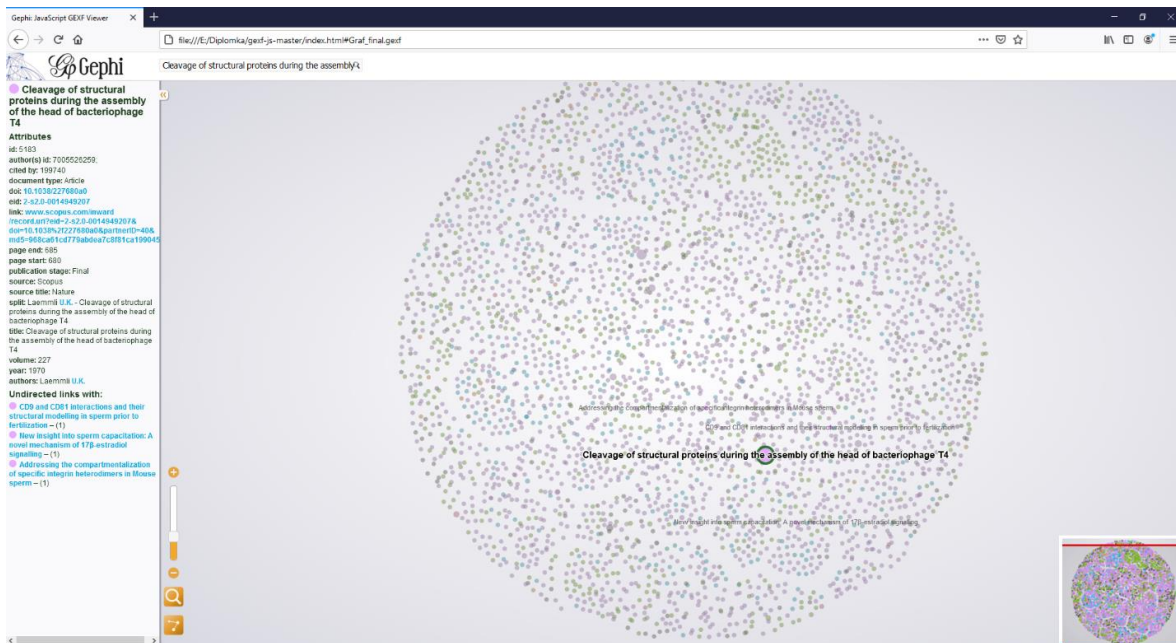


Zdroj: Vlastní zpracování

4.4.2 Zobrazení komplexní informace vybraného uzlu.

Po výběru konkrétního uzlu se v levé části obrazovky objeví komplexní informace o uzlu. Jsou zde zobrazeny všechny atributy uzlu, tj. veškeré informace, které byly uloženy v nástroji Gephi a následně exportovány prostřednictvím souboru GEXF a taktéž uzly na které zvolený uzel odkazuje.

Obrázek 41 - Zobrazení informací o zvoleném uzlu



Zdroj: Vlastní zpracování

4.4.3 Vyhledávání

V poli vyhledávání je možné do vstupního pole zadat název článku nebo autora. Tento text vyhledává adekvátní shodu ve sloupci ve kterém máme řetězec autor – title . V levém podokně se zobrazí všechny relevantní položky.

Při postupu přes tyto položky jsou jednak v grafické reprezentaci zobrazovány popisky příslušných uzlů a dále v okně detailu jsou zobrazovány komplexní informace o daném uzlu. Můžeme tedy vyhledat konkrétní článek a zobrazit články na které odkazuje.

Obrázek 42 - Vyhledávání uzlů

The screenshot displays the Gephi JavaScript GEXF Viewer interface. On the left, there is a sidebar with a search bar and a list of nodes. The selected node is titled "Cleavage of structural proteins during the assembly of the head of bacteriophage T4". The main area shows a network graph with numerous nodes and edges. A tooltip is visible over one of the nodes, displaying the article title: "Speaking their language - Development of a multilingual decision support tool for communicating invasive species risks to decision makers and stakeholders". The interface includes a browser window at the top and a search bar on the left.

Zdroj: Vlastní zpracování

Výsledky a diskuse

V práci je ověřeno praktické použití nástroje Gephi pro vizualizaci sítě vědeckých publikací. Pomocí sběru vzorku dat z oblasti informačních technologií lze dokázat, že jednotlivé vědecké disciplíny se navzájem propojují. Dále v tomto vzorku dat lze vidět jednotlivé vazby autorů a jejich publikací, kdy se navzájem citují. Tato vizualizace nám může posloužit například při rozvoji jednotlivých vědeckých disciplín, kdy je nutné si uvědomit, že je zapotřebí podporovat všechny vědní obory, a nikoliv se soustředit pouze na jeden.

Dále bylo v rámci této práce zjištěno, že citační rejstříky nejsou úplné a bylo by možné je ještě zdokonalit. Editoři by měli pracovat s maximální pečlivostí při vkládání dat do těchto rejstříků, aby nedocházelo ke ztrátě informací o jednotlivých publikacích či autorech. Tyto nedostatky nám můžou vizualizaci sítě značně zkreslovat.

5 Závěr

Hlavním cílem diplomové práce bylo provést analýzu vědeckých publikací na základě bibliografických citací a bibliografických referencí pomocí metody, která kvantifikuje vztahy mezi jednotlivými autory, dokumenty a vědeckými publikacemi. V práci byl použit datový vzorek z databáze Scopus a sice 150 publikací z oblasti počítačové vědy, které byly publikovány pod hlavičkou České zemědělské univerzity v Praze. Tyto zdrojové publikace se vlivem citací rozrostly na celkem 4789 zpracovávaných dokumentů. Data byla z databáze exportována do MS Excelu ve formátu CSV. Dalším krokem byla vizualizace těchto dat ve vhodném grafickém nástroji, který umožňuje provádět i analytické kroky.

Jako vizualizační nástroj byl vybrán nástroj Gephi, který je produktem velkého nizozemského vydavatelství odborné literatury. Tento univerzální nástroj zobrazuje jednotlivé entity jako uzly grafu a vazby mezi těmito entitami jako hrany grafu. Vzhledem k tomu, že data tvořila homogenní množinu, kde by v jednotlivých záznamech databáze existoval unikátní klíč, byla tato data do Gephi importována tak, aby vazby mezi zdrojovou a cílovou referencí dokumentů vytvořil nástroj Gephi sám. Následně bylo provedeno pomocí vnitřních nástrojů Gephi odstranění duplicit dokumentů. Dále byla v rámci grafické prezentace vazeb dokumentů zvolena nejvhodnější metoda zobrazení pro tento typ dat tak, aby výsledný obraz byl přehledný a měl co největší vypovídací hodnotu.

Během další práce s Gephi bylo vzorově provedeno několik analýz v množině předkládaných dat. Byly provedeny dílčí úlohy, jako jsou nalezení nejkratší cesty mezi dvěma uzly, rozdělení dokumentů dle jejich typu a citací, hodnocení časopisů dle počtu citovaných článků. V závěru práce byla zmíněna možnost exportu interní databáze Gephi do souboru GEXF, což je principiálně datový formát XML. Tento soubor může být importován do specializovaných prohlížečů grafů, které byly vytvořeny třetími stranami a které umožňují ještě větší prohlížeč a analytickou funkcionalitu než samotné Gephi.

6 Seznam použitých zdrojů

- [1] N.SENGUPTA, I. Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview. *Libri*. 1992, , 75-98. Dostupné z: doi:10.1515/libr.1992.42.2.75
- [2] KÖNIGOVÁ, Marie. *Úvod do bibliometrie: vysokoškolské skriptum pro 3.a 4.ročník studijního oboru 72-36-6 vědecké informace a knihovnictví*. 1.vyd. Praha: Česká informační společnost, 1993.
- [3] J. BENSMAN, Stephen. Bradford's Law and Fuzzy Sets: Statistical Implications for Library Analyses. *IFLA Journal*. Louisiana State University Libraries in Baton Rouge, Louisiana, USA, 2001, , 238-246. Dostupné z: doi:10.1177/034003520102700406
- [4] KINGSLEY ZIPF, George. *Human Behavior And The Principle Of Least Effort*. Cambridge, Massachusetts: Addison-Wesley Press, 1949. Dostupné z: doi:10.2307/2572028
- [5] COILE, Russell C. Lotka's frequency distribution of scientific productivity. *Journal of the American Society for Information Science*. 1977, **28**(6), 366-370. ISSN 00028231. Dostupné z: doi:10.1002/asi.4630280610
- [6] ABRIZAH, A., Mohammadamin ERFANMANESH, Vala Ali ROHANI, Mike THELWALL, Jonathan M. LEVITT a Fereshteh DIDEGAH. Sixty-four years of informetrics research: productivity, impact and collaboration. *Scientometrics*. 2014, **101**(1), 569-585. ISSN 0138-9130. Dostupné z: doi:10.1007/s11192-014-1390-8
- [7] Scientometrie. *Wikisofia.cz* [online]. [cit. 2020-11-28]. Dostupné z: <https://wikisofia.cz/wiki/Scientometrie>
- [8] Webometrie. *Wikipedia.org* [online]. [cit. 2020-11-28]. Dostupné z: <https://cs.wikipedia.org/wiki/Webometrie>
- [9] ABRAHAM, Ralph. *Webometry: Measuring the complexity of the World Wide Web*. World Futures, 1997. Dostupné z: doi:10.1080/02604027.1997.9972670
- [10] STOCK, W.G a S. WEBER. *Facets of informetrics. Information-Wissenschaft und Praxis* [online]. 2006 [cit. 2020-11-28]. Dostupné z: https://www.isi.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Informationswissenschaft/Dateien/Wolfgang_G._Stock/Daten_vor_2011/1166781846sonjaweber.pdf
- [11] PACÁKOVÁ, Jana. *CITAČNÍ ANALÝZA PŘEHLEDOVÁ STUDIE* [online]. In: . s. 1-34 [cit. 2020-11-28]. Dostupné z: https://is.muni.cz/th/gg8rr/Citacni_analyza.pdf
- [12] GARFIELD, Eugene. *Citation Analysis of Sports Medicine Research, 1981-1996 Productivity, Impact and Influence of Nations, Institutions and Researchers* [online]. In: . [cit. 2020-11-28]. Dostupné z: <http://www.garfield.library.upenn.edu/papers/sportsmed.html>
- [13] Web of Science: Direct Links: Home. *World's largest publisher-neutral citation index and research intelligence platform - Web of Science Group* [online]. [cit. 2020-11-28]. Dostupné z: <https://clarivate.libguides.com/directlinks>
- [14] Scopus - Quick Reference Guide. *Elsevier / An Information Analytics Business / Empowering Knowledges* [online]. [cit. 2020-11-28]. Dostupné z:

- https://www.elsevier.com/__data/assets/pdf_file/0008/207935/RD-Solutions-Scopus-Quick-Reference-Guide.pdf
- [15] Searching with Scopus. *YouTube.com* [online]. [cit. 2020-11-28]. Dostupné z: <https://www.youtube.com/watch?v=mMN7BxA0VAI>
- [16] Welcome to VOSviewer. *VOSviewer - Visualizing scientific landscapes* [online]. [cit. 2020-11-28]. Dostupné z: <https://www.vosviewer.com/>
- [17] What is Cytoscape?. *Cytoscape.org* [online]. [cit. 2020-11-28]. Dostupné z: https://cytoscape.org/what_is_cytoscape.html
- [18] Tutorial Visualization. *Gephi - The Open Viz Platform* [online]. [cit. 2020-11-28]. Dostupné z: www.gephi.org/users/tutorial-visualization
- [19] Vizualizace interního prolinkování webu pomocí nástroje Gephi. *Tyinternety.cz* [online]. 2009 [cit. 2020-11-28]. Dostupné z: <https://tyinternety.cz/prirucka-marketera/vizualizace-interniho-prolinkovani-webu-pomoci-nastroje-gephi/>
- [20] GEXF File Format. *Gephi - The Open Viz Platform* [online]. 2009 [cit. 2020-11-28]. Dostupné z: <https://gephi.org/gexf/format/schema.html>
- [21] Quick Start. *Gephi - The Open Viz Platform* [online]. [cit. 2020-11-28]. Dostupné z: www.gephi.org/users/quick-start

7 Přílohy

Příloha 1 - Příklad reálného GEXF souboru	75
Příloha 2 - XSD schéma pro soubor GEXF	79

```
<?xml version="1.0" encoding="UTF-8"?>
<gexf xmlns="http://www.gexf.net/1.3" version="1.3"
xmlns:viz="http://www.gexf.net/1.3/viz"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.gexf.net/1.3 http://www.gexf.net/1.3/gexf.xsd">
  <meta lastmodifieddate="2020-11-10">
    <creator>Gephi 0.9</creator>
    <description></description>
  </meta>
  <graph defaultedgetype="undirected" mode="static">
    <attributes class="node" mode="static">
      <attribute id="authors" title="authors" type="string"></attribute>
      <attribute id="author(s) id" title="author(s) id" type="string"></attribute>
      <attribute id="title" title="title" type="string"></attribute>
      <attribute id="year" title="year" type="integer"></attribute>
      <attribute id="source title" title="source title" type="string"></attribute>
      <attribute id="volume" title="volume" type="string"></attribute>
      <attribute id="issue" title="issue" type="string"></attribute>
      <attribute id="art. no." title="art. no." type="string"></attribute>
      <attribute id="page start" title="page start" type="integer"></attribute>
      <attribute id="page end" title="page end" type="integer"></attribute>
      <attribute id="page count" title="page count" type="string"></attribute>
      <attribute id="cited by" title="cited by" type="integer"></attribute>
      <attribute id="doi" title="doi" type="string"></attribute>
      <attribute id="link" title="link" type="string"></attribute>
      <attribute id="issn" title="issn" type="string"></attribute>
      <attribute id="isbn" title="isbn" type="long"></attribute>
      <attribute id="coden" title="coden" type="string"></attribute>
      <attribute id="document type" title="document type" type="string"></attribute>
      <attribute id="publication stage" title="publication stage" type="string"></attribute>
      <attribute id="access type" title="access type" type="string"></attribute>
      <attribute id="source" title="source" type="string"></attribute>
      <attribute id="eid" title="eid" type="string"></attribute>
    </attributes>
  </graph>
</gexf>
```

reprezentují

... atributy uzlů

názvy položek
dokumentu

```

<nodes>
  <node id="0" label="Stočes M., Vaněk J., Masner J., Jarolímek J."> ... uzel
0=kořenový
  <attvalues> ... dokument
  <attvalue for="authors" value="Stočes M., Vaněk J., Masner J., Jarolímek
J."></attvalue>
  <attvalue for="author(s) id"
value="55273566500;22986635600;55505508700;22985119700;"></attvalue>
  <attvalue for="title" value="Mobile application development options for news
and information portals"></attvalue>
  <attvalue for="year" value="2015"></attvalue>
  <attvalue for="source title" value="Future Communication, Information and
Computer Science - Proceedings of the International Conference on Future
Communication, Information and Computer Science, FCICS 2014"></attvalue>
  <attvalue for="page start" value="111"></attvalue>
  <attvalue for="page end" value="114"></attvalue>
  <attvalue for="cited by" value="2"></attvalue>
  <attvalue for="doi" value="10.1201/b18049-26"></attvalue>
  <attvalue for="link" value="https://www.scopus.com/inward/record.uri?eid=2-
s2.0-84928396106&doi=10.1201%2fb18049-
26&partnerID=40&md5=82a6d34fc0682c99cf23cbd08c818d23"></attvalue>
  <attvalue for="isbn" value="9781138026537"></attvalue>
  <attvalue for="document type" value="Conference Paper"></attvalue>
  <attvalue for="publication stage" value="Final"></attvalue>
  <attvalue for="source" value="Scopus"></attvalue>
  <attvalue for="eid" value="2-s2.0-84928396106"></attvalue>
  </attvalues>
  <viz:size value="10.0"></viz:size> ... data pro grafické
zobrazení
  </node>

```

<node id="1" label="Charland A., LeRoux B.">

... uzel 1=citovaný

<attvalues>

dokument

<attvalue for="authors" value="Charland A., LeRoux B."></attvalue>

<attvalue for="author(s) id" value="37260985600;7006434658;"></attvalue>

<attvalue for="title" value="Mobile application development: Web vs. native"></attvalue>

<attvalue for="year" value="2011"></attvalue>

<attvalue for="source title" value="Communications of the ACM"></attvalue>

<attvalue for="page start" value="49"></attvalue>

<attvalue for="page end" value="53"></attvalue>

<attvalue for="cited by" value="256"></attvalue>

<attvalue for="doi" value="10.1145/1941487.1941504"></attvalue>

<attvalue for="link" value="https://www.scopus.com/inward/record.uri?eid=2-

s2.0-

79955451802&doi=10.1145%2f1941487.1941504&partnerID=40&md5=2dc424

535a2435784638e90586a50931"></attvalue>

<attvalue for="issn" value="00010782"></attvalue>

<attvalue for="coden" value="CACMA"></attvalue>

<attvalue for="document type" value="Article"></attvalue>

<attvalue for="publication stage" value="Final"></attvalue>

<attvalue for="access type" value="Open Access"></attvalue>

<attvalue for="source" value="Scopus"></attvalue>

<attvalue for="eid" value="2-s2.0-79955451802"></attvalue>

</attvalues>

<viz:size value="10.0"></viz:size>

<viz:position x="107.80945" y="57.460793"></viz:position>

</node>

<node id="2" label="Xanthopoulos S., Xinogalos S.">

... uzel

2=citovaný

<attvalues>

dokument

<attvalue for="authors" value="Xanthopoulos S., Xinogalos S."></attvalue>

<attvalue for="author(s) id" value="55967780100;13006140900;"></attvalue>

<attvalue for="title" value="A comparative analysis of cross-platform development approaches for mobile applications"></attvalue>

<attvalue for="year" value="2013"></attvalue>

<attvalue for="source title" value="ACM International Conference Proceeding Series"></attvalue>

```

<attvalue for="page start" value="213"></attvalue>
  <attvalue for="page end" value="220"></attvalue>
  <attvalue for="cited by" value="77"></attvalue>
  <attvalue for="doi" value="10.1145/2490257.2490292"></attvalue>
  <attvalue for="link" value="https://www.scopus.com/inward/record.uri?eid=2-
s2.0-
84890514056&doi=10.1145%2f2490257.2490292&partnerID=40&md5=
0517dc
dd34fc9aeb6efff1aa67c9389b"></attvalue>
    <attvalue for="isbn" value="9781450318518"></attvalue>
    <attvalue for="document type" value="Conference Paper"></attvalue>
    <attvalue for="publication stage" value="Final"></attvalue>
    <attvalue for="source" value="Scopus"></attvalue>
    <attvalue for="eid" value="2-s2.0-84890514056"></attvalue>
  </attvalues>
  <viz:size value="10.0"></viz:size>
  <viz:position x="-176.793" y="428.27866"></viz:position>
</node>
</nodes>
<edges>
  <edge id="0" source="0" target="1"></edge>
  <edge id="1" source="0" target="2"></edge>
</edges>
</graph>
</gexf>

```

```

    <xs:restriction base="xs:string">
      <xs:enumeration value="1.2"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="variant" type="xs:string"/>
</xs:complexType>
<xs:element name="meta" type="ns1:meta-content"/>
<xs:element name="graph" type="ns1:graph-content"/>
<xs:complexType name="meta-content">
  <xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element ref="ns1:creator"/>
    <xs:element ref="ns1:keywords"/>
    <xs:element ref="ns1:description"/>
  </xs:choice>
  <xs:attribute name="lastmodifieddate" type="xs:date"/>
</xs:complexType>
<xs:element name="creator" type="xs:string"/>
<xs:element name="keywords" type="xs:string"/>
<xs:element name="description" type="xs:string"/>
<xs:complexType name="nodes-content">
  <xs:sequence>
    <xs:element minOccurs="0" maxOccurs="unbounded" ref="ns1:node"/>
  </xs:sequence>
  <xs:attribute name="count" type="xs:nonNegativeInteger"/>
</xs:complexType>
<xs:element name="node" type="ns1:node-content"/>
<xs:complexType name="edges-content">
  <xs:sequence>
    <xs:element minOccurs="0" maxOccurs="unbounded" ref="ns1:edge"/>
  </xs:sequence>
  <xs:attribute name="count" type="xs:nonNegativeInteger"/>
</xs:complexType>
<xs:element name="edge" type="ns1:edge-content"/>
<xs:simpleType name="defaultedgetype-type">
  <xs:annotation>
    <xs:documentation>Datatypes</xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:string">
    <xs:enumeration value="directed"/>
    <xs:enumeration value="undirected"/>
    <xs:enumeration value="mutual"/>
  </xs:restriction>
</xs:simpleType>

```

```
<xs:simpleType name="edgetype-type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="directed"/>
    <xs:enumeration value="undirected"/>
    <xs:enumeration value="mutual"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="id-type">
  <xs:union memberTypes="xs:string xs:integer"/>
</xs:simpleType>
<xs:simpleType name="idtype-type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="integer"/>
    <xs:enumeration value="string"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="mode-type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="static"/>
    <xs:enumeration value="dynamic"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="weight-type">
  <xs:restriction base="xs:float"/>
</xs:simpleType>
</xs:schema>
```