Kookurenční sítě mezi historickým vyprávěním a kulturně evolučním vysvětlením

(rukopis kapitoly pro knihu *Digitální obrat v českých humanitních vědách* – nešířit)
21. prosince 2019

Vojtěch Kaše,
SDAM, CAS, Aarhus Universitet, Dánsko & KFI FF, Západočeská univerzita v Plzni
vojtech.kase@gmail.com

Úvod

V této kapitole představím metody vytváření, vizualizace a analýzy kookurenčních sítí. Pomocí kookurenčních sítí můžeme text v přirozeném jazyce přetvořit do formálního síťového grafu, jehož uzly představují jednotlivá slova zdrojového textu a jehož hrany představují spoluvýskyty těchto slov, a následně vizualizovat a analyzovat pomocí standardizovaných metod síťové analýzy. Kookurenční sítě lze nahlížet jako spadající pod hlavičku metod výpočetní textové analýzy. Tyto metody jsou dnes zpravidla implementovány za využití počítačových algoritmů vyvíjených v informatice v oblastech zpracování přirozeného jazyka a dolování dat a dále rozvíjeny především v korpusové lingvistice. Ve většině případů tak nemáme co do činění s metodami a nástroji, které by byly samostatně vyvíjeny v komunitě digital humanities, ale spíše s metodami a nástroji, které jsou vyvíjeny na pomezí informatiky a lingvistiky a které jsou hojně využívány např. v marketingu.

Metody výpočetní textové analýzy se v kontextu digital humanities používají často víceméně zaměnitelně s termínem „distant reading“, který zpopularizoval literární vědec Franco Moretti.[1] Termín distant reading byl vytvořen v protikladu k termínu close reading: Zatímco většina badatelského úsilí v literárněvědném bádání je založena na detailním čtení vybraných literárních děl („close reading“), chceme-li v rozsáhlých literárních korpusech objevit obecnější vzorce, nezbývá nám podle Morettiho než opřít se o jiný typ metod, neboť detailní četba nám nutně nemůže stačit. Větší „vzdálenost“ pro Morettiho znamená „méně prvků, odtud ostřejší smysl jejich celkové propojenosti.“[2] Jak Moretti na jednom místě trefně

[1] Franco Moretti: *Distant Reading*, London – New York 2013. Genealogii tohoto termínu mapuje zejména Ted Underwood: A Genealogy of Distant Reading, *Digital Humanities Quarterly*, 11, č. 2 (2017), s. 1–12.
[2] F. Moretti: *Distant Reading*, s. 45.

poznamenává, „Číst ‘více’ je vždy dobrá věc, není to však řešení“.[3] Kde běžné čtení nestačí, je však možné použít metody výpočetní textové analýzy.

Jak již předchozí odstavce naznačují, v kontextu digital humanities se distanční čtení pojí primárně s literárněvědným a jazykovědným bádáním a výzkumnými otázkami, které si tyto disciplíny kladou (např. proměny žánru, témat, slovníku či syntaxe uvnitř vybraného korpusu).[4] Jak se však později pokusím ukázat, tyto metody mají širší aplikovatelnost, než by se mohlo na první pohled zdát, a to např. i při prozkoumávání hypotéz z oblasti kulturní evoluce náboženství.

K počítačové implementaci všech algoritmů rozebíraných v této kapitole využijeme programovací jazyk Python 3 s řadou rozšiřujících balíčků. Celý kód je přístupný ve formě jednoho uceleného notebooku dle standardu Jupyter[5] a je možné jej spustit v prohlížeči i bez lokální instalace Pythonu.[6] Je nepochybně pravda, že k aplikaci řady algoritmů, které si zde ukážeme, by bylo možné použít jednu po druhé různé webové aplikace (např. Voyant tools[7] či Infranodus[8]). Mám však za to, že může být zvláště názorné ukázat si, jak je možné tyto algoritmy použít bez použití těchto aplikací a implementovat je samostatně pomocí vlastního volně šiřitelného a upravovatelného skriptu uvnitř Jupyter notebooku.

Nový zákon v Českém ekumenickém překladu

První soubor dat, na němž si budeme demonstrovat možnosti kookurenčních sítí, představují texty Nového zákona v českém překladu. Ty představují objem textu, který je v silách jednotlivce přečíst od prvního do posledního slova během několika dnů. Celkově se jedná o 134 188 slov, což odpovídá cca. 550 normostranám. Zůstává tu proto otázka, co navíc nám v případě takto relativně krátkého korpusu, celkově sestávajícího z 27 novozákonních knih, můžou metody výpočetní textové analýzy nabídnout.

---

[3] Ibid., s. 46.
[4] Viz např. Richard Změlík: Close reading nebo Distant reading?, in: *Mezi Kritikou a Poezií. Ladislavu Soldánovi k Osmdesátinám*, ed. Jakub Sichálek, Opava 2018, s. 48–56. Viz také kap. xxx v tomto svazku.
[5] Jupyter notebooky představují velice užitečný způsob psaní a dokumentace kódu pro účely datového programování pro badatelské účely. Viz Adam Rule et al.: Ten Simple Rules for Writing and Sharing Computational Analyses in Jupyter Notebooks, *PLOS Computational Biology* 15, č. 7 (2019), s. e1007007.
[6] Příslušný Jupyter notebook je přístupný prostřednictvím GitHub repozitoře zde: https://github.com/kasev/distancni-cteni. Notebook je nakonfigurován tak, že ke spuštění kódu uvnitř něj je možné využít webovou platformu Google Colaboratory (https://colab.research.google.com/notebooks/welcome.ipynb#), a není tudíž třeba mít lokální instalaci Pythonu.
[7] https://voyant-tools.org
[8] https://infranodus.com

S ohledem na naše cíle je potřeba texty novozákonních knih podrobit dalším standardním procedurám umožňujícím aplikaci námi zvolených metod výpočetní textové analýzy. To znamená zejména odstranit z textů interpunkci, převést slova do jejich základních tvarů a zaměřit se pouze na slovní druhy, které jsou pro nás zajímavé. Pro tyto účely v našem skriptu využijeme především morfologického slovníku a značkovače MorphoDiTa,[9] který lze dotazovat pomocí webového API. To v praxi znamená, že sestavíme dotaz v podobě URL adresy obsahující řetězec českého textu, který chceme morfologicky analyzovat, a API na náš dotaz odpoví tím, že vrátí zpět výpis jednotlivých slov uvnitř řetězce spolu s jejich morfologickým určením a s jejich lemmatizovanými tvary, tj. infinitiv u sloves a první pád jednotného čísla u podstatných jmen, přídavných jmen, zájmen, číslovek a příslovcí. Vzneseme-li např. dotaz v podobě 3. verše 1. kapitoly Markova evangelia („Hlas volajícího na poušti: Připravte cestu Páně, vyrovnejte mu stezky!“),[10] dostaneme jako odpověď strukturovaný dokument, který si již snadno převedeme do následující tabulky:

prvek	tag	lemma
Hlas	NNIS1-----A----	hlas
volajícího	AGIS2-----A----	volající
na	RR--6----------	na
poušti	NNFS6-----A----	pouš'
:	Z:----------	:
Připravte	Vi-P---2--A----	připravit
cestu	NNFS4-----A----	cesta
Páně	NNMS2-----A---2	Pán
,	Z:----------	,
vyrovnejte	Vi-P---2--A----	vyrovnat
mu	PHZS3--3-------	on
stezky	NNFP4-----A----	stezka
!	Z:----------	!

Tabulka 1: Lematizace.

Řádky v této tabulce můžeme dále filtrovat za využití sloupce „tag“. Hodnoty v tomto sloupci obsahují standardizovaný řetězec o 15 znacích, přičemž každý znak na konkrétní pozici kóduje určitou morfologickou informaci (např. slovní druh, pád či osobu, číslo apod.). Pro naše potřeby je důležité písmeno na první pozici, které kóduje základní informaci o povaze

[9] Jana Straková – Milan Straka – Jan Hajič: Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition, in: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* Stroudsburg, PA, USA 2014, s. 13–18.

[10] „http://lindat.mff.cuni.cz/services/morphodita/api/tag?data=Hlas volajícího na poušti: Připravte cestu Páně, vyrovnejte mu stezky!“

analyzovaného prvku, tedy typicky slovní druh: „N“ značí podstatné jméno, „A“ přídavné jméno, „V“ sloveso, „D“ příslovce, „R“ předložku, „Z“ interpunkční znaménko apod. Na základě této informace jsme nyní schopni z morfologicky analyzovaného textu vybrat pouze podmnožinu slov určitých slovních druhů, respektive pouze jejich lematizovaných tvarů. Pro potřeby našeho příkladu se v dalším omezíme pouze na lemmatizované tvary podstatných jmen, přídavných jmen, číslovek, sloves a příslovcí. Z věty výše nám takto zůstane řada slov „hlas volající poušť připravit cesta Pán vyrovnat stezka“.

Nyní se vraťme k našemu datasetu čítajícímu 134 188 slov. Z tohoto datasetu pomocí právě popsané procedury získáme celkově 70709 lematizovaných podstatných jmen, přídavných jmen, číslovek, sloves a příslovcí. Textová data ve filtrované podobě určitě přehlížejí některé významové aspekty původního textu. Na druhou stranu však takto upravený text umožňuje aplikaci řady algoritmů z oblasti zpracování přirozeného jazyka, dolování dat a počítačové lingvistiky, z nichž nás zde zajímají především ty, které se používají v modelech automatické sumarizace textů.

Velice jednoduchý, přitom však velice efektivní model je model zvaný „bag-of-words“. U tohoto modelu pomocí série algoritmů získáme seznam párů slov a jejich frekvencí. Tento seznam si můžeme uspořádat od nejfrekventovanějších slov po ta nejméně častá a zohlednit relativní frekvenci daných slov uvnitř daného dokumentu. Dále stojí za povšimnutí, že v tomto modelu jsou zcela ignorovány vztahy mezi slovy. Navzdory této své reduktivnosti má model „bag-of-words“ velice širokou oblast použití a umožňuje např. velice důmyslně porovnat obsahovou podobnost vícero dokumentů, a to tak, že porovnáme určitý počet nejčastějších slov, které se v těchto dokumentech objevují. To si můžeme demonstrovat na příkladu čtyř novozákonních evangelií např. tak, že se u každého z nich podíváme na deset nejčastějších slov (Tabulka 2).

Mt	**Mk**	**L**	**J**
říci (0.01368)	říci (0.01518)	říci (0.01563)	říci (0.01935)
Ježíš (0.0094)	Ježíš (0.00996)	mít (0.00814)	Ježíš (0.0168)
mít (0.0069)	mít (0.00807)	Ježíš (0.00689)	otec (0.00911)
jít (0.0063)	přijít (0.00664)	člověk (0.0064)	mít (0.0084)
člověk (0.0063)	jít (0.00645)	přijít (0.00564)	přijít (0.00678)
syn (0.00595)	člověk (0.00636)	dát (0.00472)	odpovědět (0.00621)
přijít (0.00595)	učedník (0.00456)	jít (0.0044)	učedník (0.00565)
dát (0.00446)	duch (0.00351)	den (0.00429)	svět (0.00551)
učedník (0.00428)	moci (0.00351)	syn (0.00423)	jít (0.0053)
odpovědět (0.00422)	říkat (0.00351)	bůh (0.00374)	dát (0.00501)

Tabulka 2: Nejfrekventovanější slova v novozákonních evangeliích.

Asi nás v případě evangelií nepřekvapí, že jedním z nejčastějších slov u všech z nich je slovo „Ježíš“ a že stejně jako v českém jazyce obecně je u všech hojně zastoupeno sloveso „mít“ (sloveso „být“, které by bylo ve všech případech na prvním místě, jsme vypustili již dříve). Podíváme-li se však na tato data podrobněji, můžeme z nich vyčíst více. Vidíme zde např. určité hodnoty, na které jednotliví evangelijní autoři kladou důraz. Zatímco u všech tří synoptiků např. pozorujeme poměrně vysokou frekvenci slova „člověk“ (která se patrně pojí s titulem „syn člověka“), u Jana si můžeme povšimnout relativně vysoké frekvence slova „otec“. A mohli bychom pokračovat. Nedozvídáme se zde však příliš o tom, která slova jsou pro kterého autora typická a jak se k sobě jednotlivá slova uvnitř textu vztahují. Obojí postupně zohledníme v dalších příkladech.

Kookurenční sítě

Kookurenční sítě představují využití teorie sítí při analýze textových dat. Teorií sítí zde míním rozsáhlou oblast výzkumů vymezenou tzv. vědou o sítích na straně jedné[11] a analýzou sociálních sítí na straně druhé.[12] Základním východiskem je, že řadu jevů okolního světa lze formalizovat jako sestávající z uzlů a hran opatřených kvantitativními atributy. Takto formalizovaný jev lze následně analyzovat pomocí standardizovaných algoritmů.

Metody síťové analýzy lze použít i při práci s jazykovými daty, přičemž existuje vícero možností, jak textovou síť konceptualizovat.[13] Jednou z nich je analyzovat slova a vztahy mezi slovy uvnitř vybraného textu či korpusu s ohledem na to, jak se spolu uvnitř vybraného textu spoluvyskytují (odtud kookurenční). To opět může nabývat různých podob.

V případě novozákonního textu můžeme postupovat tak, že si pro jednotlivé knihy či autory vytvoříme síť založenou na vztazích mezi určitým počtem nejčastějších termínů (omezením počtu slov dosáhneme větší srovnatelnosti jednotlivých sítí mezi sebou a při vizualizacích dosáhneme větší přehlednosti). Námi použitá sestava algoritmů pro generování kookurenční sítě pracuje následovně.

Nejprve je pro příslušnou knihu vygenerována proměnná *lexicon*, která obsahuje množinu o námi zvoleném počtu nejčastějších slov uvnitř daného dokumentu, resp. lemat již dříve předvybraných slovních druhů – v našem příkladu nyní půjde o nejčastějších 500 slov. V

[11] Ulrik Brandes et al.: What Is Network Science?, *Network Science* 1, č. 1 (2013), s. 1–15.
[12] Lenka Buštíková: Analýza Sociálních Sítí, *Sociologický Časopis* 35, č. 2 (1999), s. 193–206.
[13] Rada Mihalcea – Dragomir Radev, *Graph-Based Natural Language Processing and Information Retrieval*, Cambridge 2011.

následujícím kroku program prochází od začátku do konce slovo po slově předzpracovanou (tj. lematizovanou a filtrovanou) verzi příslušné biblické knihy a extrahuje z ní tzv. *bigramy*, neboli páry bezprostředně sousedících slov. Tak například z prvního verše takto předzpracovaného Prvního listu Janova („počátek slyšet vlastní oko vidět hledět ruka dotýkat zvěstovat slovo život“)[14] získáváme bigramy „počátek slyšet“, „slyšet vlastní“, „vlastní oko“, „oko vidět“ atd. Z těchto párů slov jsou následně vybrány pouze ty, z nichž se obě slova nachází uvnitř proměnné lexicon. Bigramy obsahující tatéž dvě slova jsou následně sečteny. Získáváme tak seznam párů slov spolu s jejich četnostmi. V Tabulce 3 vidíme deset nejčastějších párů slov neboli bigramů ve čtyřech novozákonních evangeliích.

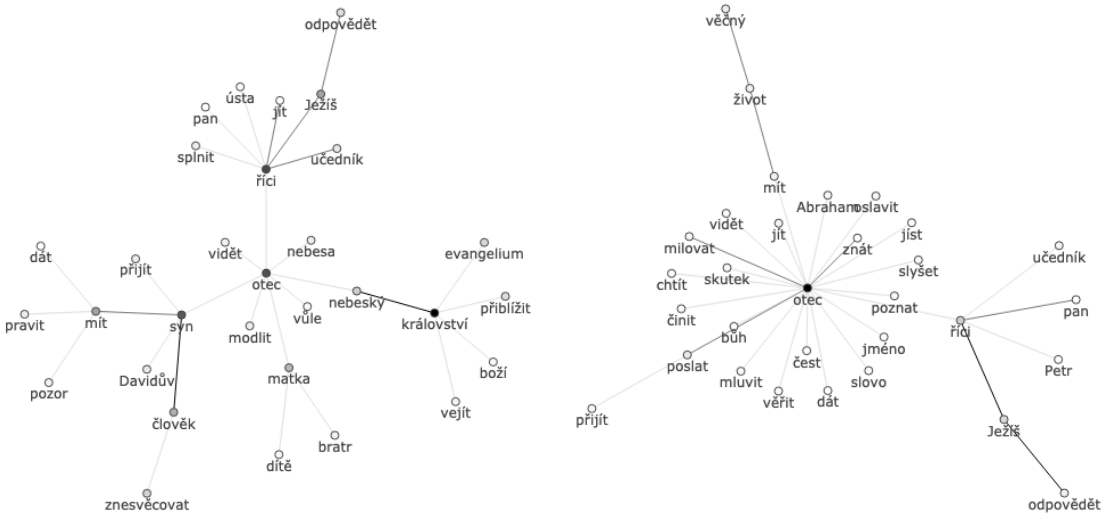
Mt	Mk	L	J
('království', 'nebeský', '31')	('Ježíš', 'říci', '26')	('boží', 'království', '32')	('Ježíš', 'říci', '66')
('syn', 'člověk', '31')	('syn', 'člověk', '15')	('syn', 'člověk', '27')	('Ježíš', 'odpovědět', '45')
('Ježíš', 'říci', '28')	('boží', 'království', '14')	('Ježíš', 'říci', '26')	('Petr', 'Šimon', '18')
('Ježíš', 'odpovědět', '20')	('říci', 'jít', '12')	('říci', 'učedník', '14')	('věčný', 'život', '18')
('říci', 'učedník', '17')	('duch', 'zlý', '12')	('duch', 'svatý', '14')	('život', 'mít', '17')
('syn', 'mít', '13')	('říci', 'učedník', '11')	('říci', 'pan', '13')	('říci', 'pan', '15')
('říci', 'jít', '12')	('duch', 'čistý', '10')	('říci', 'pán', '12')	('říci', 'jít', '13')
('nebeský', 'otec', '11')	('Ježíš', 'přijít', '7')	('Ježíš', 'odpovědět', '11')	('syn', 'člověk', '13')
('prorok', 'ústa', '11')	('Ježíš', 'odpovědět', '7')	('říci', 'jít', '11')	('otec', 'poslat', '12')
('farizeus', 'zákoník', '11')	('říci', 'pravit', '7')	('říci', 'mít', '11')	('Ježíš', 'přijít', '11')

Tab.3: 10 nejčastějších bigramů v novozákonních evangeliích.

Data v této podobě mohou být následně využita k vytvoření síťového grafu a ten dále analyzován. V našem případě z bigramových dat vytvoříme nesměrovou, váženou síť. Uzly sítě jsou jednotlivá slova, vazby mezi těmito uzly jsou slova, která se spolu objevují v bigramech. Hlavním kvantitativním atributem uzlů je jejich frekvence uvnitř příslušné novozákonní knihy, hlavním kvantitativním atributem hran, resp. jejich váhou, je pak četnost jejich spoluvýskytů uvnitř těchto knih.[15] Síťový graf v této podobě umožňuje aplikaci široké škály metrik z oblasti standardní síťové vědy: Pro jednotlivé uzly a hrany může být např. vypočítána jejich centralita, v rámci sítě mohou být identifikovány „komunity“ odpovídající určitým tématům textu apod.[16]

[14] „Co bylo od počátku, co jsme slyšeli, co jsme na vlastní oči viděli, na co jsme hleděli a čeho se naše ruce dotýkaly, to zvěstujeme: Slovo života“ (1J).
[15] Velice podobná metoda je popsána in Dmitry Paranyushkin: Identifying the Pathways for Meaning Circulation Using Text Network Analysis, *Nodus Labs*, 2011 (online), https://doi.org/citeulike-article-id:11684328.
[16] Mark E. J. Newman: Analysis of Weighted Networks, *Physical Review E* 70, č. 5, s. 056131.

Místo abychom k těmto sítím přistupovali jako celkům, zaměříme se nyní pouze na určité výseky těchto sítí. Velice často nás v rámci sítě např. zajímá pouze určitý počet nejbližších sousedů určitého uzlu v rámci dané sítě. Tato data získáme pomocí algoritmů pro výpočet nejkratší vzdálenosti uvnitř sítě. Můžeme předpokládat, že vzdálenost mezi uzly je tím kratší, čím větší váhu má daná hrana. Vzdálenost mezi dvěma sousedními uzly tak odpovídá inverzi jejich váhy. U uzlů, které spolu přímo nesousedí, je pak třeba vypočítat, přes které uzly k nim v rámci sítě vede nejkratší cesta a jaká je celková vzdálenost této cesty. Máme-li tato data, můžeme pro potřeby další analýzy a interpretace vybrat z původní sítě pouze určitý počet nejbližších uzlů, jako je tomu v příkladech na Obrázku 1.



Obrázek 1: Sousedství slova „otec" v Matoušově (vlevo) a Janově evangeliu (vpravo).

Na Obrázku 1 vidíme síť sestávající z 30 nejbližších uzlů k uzlu „otec" v Matoušově a Janově evangeliu. (Pozor, že vyobrazení obsahuje pouze nejkratší vazby ze zdrojového uzlu k ostatním uzlům a neobsahuje další vazby mezi uzly z původní sítě.) Tento typ sítě je vhodný k přiblížení toho, s jakými termíny je dané zdrojové slovo asociováno. V duchu distribuční sémantiky[17] můžeme dokonce uvažovat o tom, že takto vygenerovaná síť je určitou aproximací pro mentální asociační síť k danému slovu, a to buď autorů původního textu či v komunitách jejich recipientů.[18]

Srovnáme-li mezi sebou sítě sestávající ze sousedů slova „otec" v Matoušově a Janově evangeliu, vidíme již na první pohled řadu rozdílů. V případě Matoušova evangelia se zdá, že

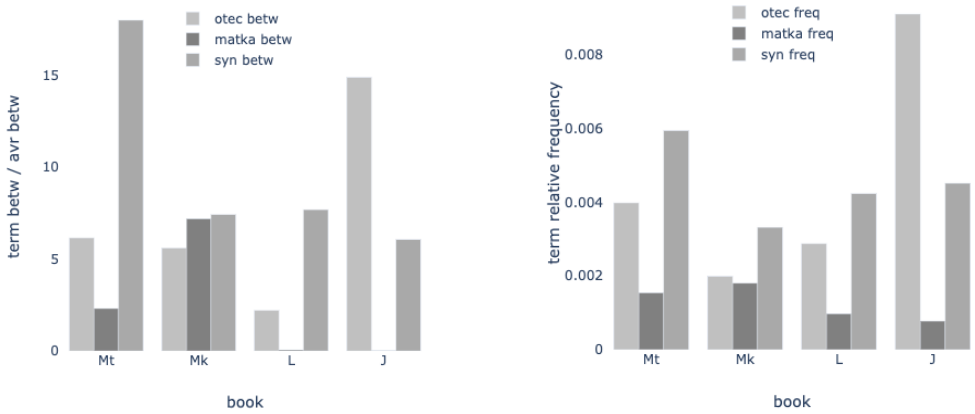
[17] Alessandro Lenci: Distributional Models of Word Meaning, *Annu. Rev. Linguist* 4 (2018), s. 151–171.
[18] David Galea – Peter Bruza: Deriving Word Association Networks from Text Corpora, *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science* (2015), s. 252–257.

slovo „otec“ je úzce provázáno s několika slovy, které zprostředkují vazbu na určité poměrně dobře vymezitelné a sémantické okruhy. Zaprvé zde můžeme vypozorovat poměrně zřetelně vymezený soteriologický okruh, zprostředkovaný vazbou na slovo „nebeský“. Odtud pak vede cesta ke slovům „království“, „evangelium“, „přiblížit“, „boží“ a „vejít“. Dále zde máme dva relativně dobře oddělitelné christologické okruhy. Na jedné straně je zde mesianistický okruh, zprostředkovaný vazbou na slovo „syn“, které se zde váže především na slova „člověk“ a „Davidův“[19], což odráží christologické tituly „syn člověka“ a „syn Davidův“. Na druhé straně zde máme učitelský okruh, zprostředkovaný vazbou na sloveso „říci“, a odtud dále ke slovům „učedník“, „Ježíš“, „jít“, „ústa“ a dalším. Nakonec můžeme mluvit ještě o rodinném okruhu, kdy přes slovo „matka“ máme zprostředkovanou vazbu na slova „dítě“ a „bratr“.

V případě sítě nejbližších sousedů vyvozené z Janova evangelia je situace odlišná. Zdá se, že zde slovo „otec“ představuje specifický teologický okruh či téma samo o sobě. Mezi celkovými 22 přímými vazbami zde pozorujeme zvláště silné vazby na slova „znát“, „milovat“, „poslat“. Vedle něj můžeme pouze vymezit christologický učitelský okruh zprostředkovaný opět slovesem „říci“.

To, jak ústřední polohu uvnitř sítě jako celku slovo „otec“ má, je možné kvantifikovaně vyjádřit prostřednictvím metrik centrality. Pro naše potřeby využijime zaužívanou metriku centrality zvanou „betweenness centrality“ neboli mezilehlost.[20] Aby však tato metrika byla přenositelná napříč našimi sítěmi, bude vhodné ji vždy vztáhnout k průměrné mezilehlosti všech uzlů uvnitř té které sítě. Na Obrázku 2 vidíme srovnání hodnot takto normalizované mezilehlosti a relativní frekvence termínů „otec“, „matka“ a „syn“. Můžeme si zde všimnout, že data relativní frekvence a mezilehlosti spolu do jisté míry korelují. Zatímco u Matouše pozorujeme, že nejvyšší hodnotu má v případě obou metrik slovo „syn“, v případě Jana je to pro změnu „otec“. V některých případech je však tato úměra výrazně slabší. To je např. situace slova „matka“ v Markově evangeliu, které má sice ze všech tří termínů nejnižší relativní četnost, ale hodnotu mezilehlosti má téměř stejně vysokou jako „syn“. To naznačuje, že slovo „matka“ může v případě Markova evangelia propojovat vícero tematických okruhů. Z určitého hlediska se tak jedná o důležitější termín, než by se mohlo zdát z frekvenčních dat.

[19] Slovo „Davidův“ nebylo algoritmem lematizováno. Odráží to nedostatečnost příslušného nástroje zvláště při nakládání s vlastními jmény.
[20] Viz Marc E. J. Newman: *Networks: An Introduction*, Oxford – New York 2010, s. 185–193.



Obrázek 2: Mezilehlosti a relativní frekvence vybraných termínů.

V odstavcích výše jsme záměrně pracovali s dobře známým korpusem v českém překladu. To mělo tu výhodu, že jsme mohli výsledky našich analýz bezprostředně srovnat s našimi znalostmi o biblických textech. Současně jsme si tak mohli ukázat, že Python plně postačuje k práci s jazykovými daty v češtině a že pomocí webového API můžeme textová data v češtině i snadno lematizovat. Avšak analyzovat texty Nového zákona v národním jazyce není pro humanitně-vědného badatele se specializací na antická náboženství dostačující. K tomu je potřeba pracovat s texty v řečtině, která byla originálním jazykem buď všech nebo alespoň většiny novozákonních knih. Učiníme tak spolu s obrácením pozornosti k řádově rozsáhlejšímu textovému korpusu.

Korpus LG

Díky iniciativě *Open Greek and Latin Project* při Lipské univerzitě vznikl veřejně dostupný lematizovaný a morfologicky označkovaný dataset antických řeckých textů z období antiky,[21] který obsahuje i řecký text Nového zákona. Ve zbývající části této kapitoly se budeme věnovat právě tomuto korpusu.

Korpus jako celek obsahuje dohromady 901 děl a 25 522 507 slov. V analýzách níže však budeme pracovat pouze s výsekem tohoto korpusu, a to s texty, které bylo možné pomocí veřejně dostupných zdrojů (1) jednoznačně klasifikovat buď jako křesťanské nebo pohanské[22], (2) datovat je na úrovni století a (3) na základě toho pojímat jako vzniklé v období mezi 8. stol.

[21] https://github.com/gcelano/LemmatizedAncientGreekXML
[22] Srovnání s texty židovské provenience si do budoucna zaslouží samostatné pojednání.

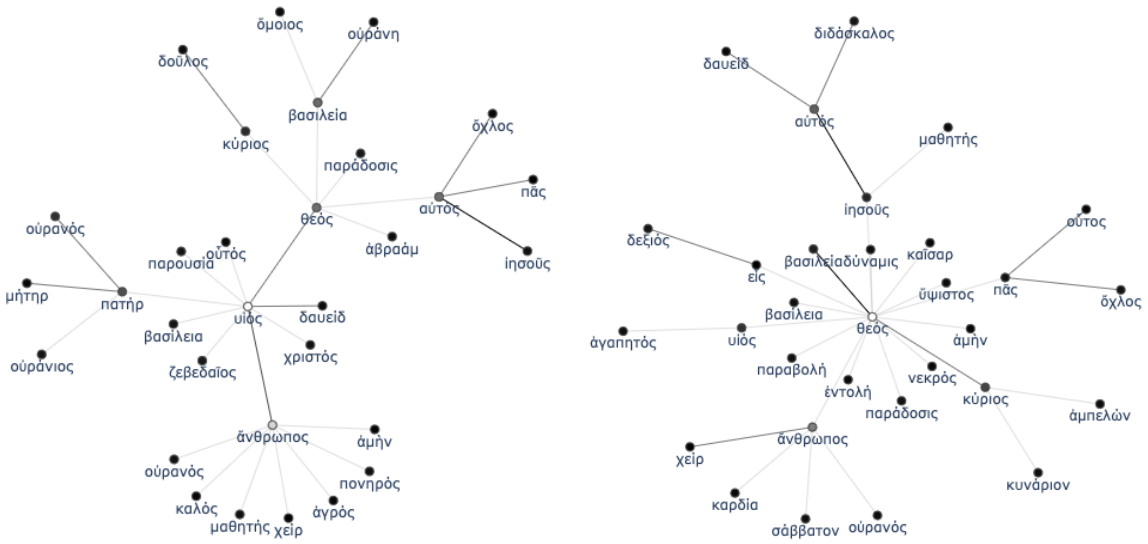
př.n.l. včetně a koncem 4. stol. n.l. včetně.[23] Tím se náš korpus zúžil na 687 dokumentů a 14 709 794 slov. V případě většiny analýz budeme korpus analyzovat jako sestávající ze tří, resp.čtyř, subkorpusů:

(1) archaický: texty vzniklé v 8.-6. stol.př.n.l.;

(2) klasický: texty vzniklé v 5.-4. stol.př.n.l.;

(3) římský/křesťanský: texty vzniklé v 1.-4. stol.n.l., z nichž část je klasifikována jako křesťanské provenience.

Obdobně jako v případě Nového zákona v češtině, i zde budeme postupoval tak, že využijeme morfologické značky k výběru lematizovaných tvarů pouze určitých slovních druhů.[24] V následujících analýzách tak budeme pracovat pouze s 3 826 927 slovy klasifikovanými jako podstatná jména a adjektiva. Můžeme hned pokročit k ukázání si, jak v tomto případě vypadá výsek sítě tvořený nejbližšími sousedy námi zvoleného slova. S ohledem na následující analýzy zde zaměříme pozornost na použití termínu *θεός*, neboli „bůh“. Stejně jako v případě české verze, i zde si vygenerujeme síť sestávající z 30 nejbližších sousedů tohoto termínu v případě řeckého textu novozákonních evangelii (viz Obrázek 3).[25]



Obrázek 3: Sousedství slova θεός vytvořené na základě řeckého textu Matoušova a Markova evangelia.

[23] Tabulka zahrnutých dokumentů s přehledovými metadaty je dostupná v příslušné repozitoři na githubu.

[24] Díky povaze korpusu však nebudeme lematizaci a morfologické značkování provádět sami. Je třeba mít na paměti, že tento korpus vznikl jako výsledek experimentu s automatickým nástrojem pro morfologické určování, a obsahuje proto mnoho chyb, které byly alespoň částečně odstraněny dodatečným čištěním.

[25] Velice podobná analýza řeckého textu je použitá také in István Czachesz, Network Analysis of Biblical Texts, *Journal of Cognitive Historiography* 3, č. 1–2 (2016), s. 43–67.

Díky lematizované podobě vstupních dat můžeme jednotlivá slova také poměrně snadno opatřit automatickým překladem do angličtiny pomocí dat z volně dostupných slovníků.[26] Vytvořenou metodu můžeme aplikovat na libovolný dokument v korpusu. V Tabulce 4 vidíme 5 nejbližších termínů slov k termínu *θεός* v Homérově Iliadě, v Platónově dialogu *Faidros* a v řeckém textu Matoušova evangelia, tedy v textech ze třech různých období.

Homer, *Iliad*	**dist**	**Plato, *Phaedrus***	**dist**	***Matthew***	**Dist**
πᾶς (all, the whole)	0.025	ἄνθρωπος (a man, one of the human race)	0.25	υἱός (a son, descendent)	0.111
ἄλλος (other, another)	0.03	ἕκαστος (each, every one)	0.25	ἄνθρωπος (a man, one of the human race)	0.142
ἀνήρ (a male human being, a man)	0.031	πᾶς (all, the whole)	0.333	δαυείδ	0.211
ἀθάνατος (undying, immortal)	0.045	σφέτερος (their own, their)	0.333	βασιλεία (kingship, sovereignty)	0.25
ἄναξ (a lord, master)	0.049	αὐτός (he, she)	0.476	οὐράνη (heaven)	0.309

Tabulka 4: 5 nejbližších sousedních slov k termínu θεός ve vybraných textech.

Kulturní evoluce moralizujících náboženství

Termín *θεός* jsme v analýzách výše nezvolili náhodně. Jde totiž o termín, jehož použití v antických řeckých textech může být zajímavé s ohledem na současnou debatu ohledně kulturní evoluce moralizujících náboženství. V této diskuzi se nezřídka odkazuje k dějinám náboženství antického Středomoří s cílem podpořit či oslabit tu či onu obecnou hypotézu o dlouhodobé dynamice náboženských představ v dějinách lidských společností.

Tato diskuze vychází z obecného pozorování, že náboženské dějiny nejkomplexnějších společností posledních několika tisíciletí jsou charakteristické tím, že jsou zde náboženské představy úzce provázány s oblastí morálky a že se zde často objevuje představa mocného božstva a posmrtných odměn a trestů. Díváme-li se však na společnosti méně komplexní, toto propojení mezi náboženskými představami a morálkou zpravidla chybí. Otázkou je, co za touto korelací stojí.

[26] Zde využíváme dva slovníky. V případě, že termín není identifikován a přeložen prvním, kratším slovníkem (https://raw.githubusercontent.com/biblicalhumanities/Dodson-Greek-Lexicon/master/dodson.xml), je proveden pokus o překlad ještě za využití druhého, rozsáhlejšího slovníku (https://raw.githubusercontent.com/gcelano/MorpheusGreekUnicode/master/).

Hypotéza velkých bohů tvrdí, že vztah mezi sociální komplexitou a morálním aspektem náboženství je vlastně sociálně-funkcionální: Tím, jak se společnost stává komplexnější, denním chlebem jejího fungování se stává spolupráce mezi jedinci, kteří se osobně neznají a nepojí je k sobě žádné příbuzenské vazby. Zastánci hypotézy velkých bohů tvrdí, že k motivaci jednotlivců k tomu, aby si za těchto podmínek navzájem důvěřovali, spolupracovali spolu a nepodváděli se, mohou právě velice dobře posloužit představy moralizujících (tj. odměňujících a trestajících) božstev, kterým se právě proto v těchto společnostech začíná dařit.[27]

Zastánci hypotézy blahobytu (*affluence*) se na toto funckionalistické vysvětlení dívají skepticky. Původ propojení mezi náboženstvím a morálkou nahlíží spíše jako jeden z mnoha projevů změněných preferencí u členů společnosti, která dosáhla určité materiální úrovně. Dosažení určité materiální úrovně vede jedince ke zpomalení tempa jejich „životních historií", tedy např. k pozdějšímu věku pro založení rodiny a snížení celkového počtu potomků. Na psychologické rovině se tato změna projevuje orientací na dlouhodobější cíle a rozšířením prostoru pro sociální realizaci. Zde je pak podle zastánců hypotézy blahobytu třeba hledat zdroj víry v posmrtný život s odměnami a tresty, nezřídka také spojovaný se zvláště mocnými božstvy.[28]

Další skupina badatelů obhajuje rituální hypotézu, která má blízko k hypotéze velkých bohů v tom, že se také soustředí na vývoj sociální komplexity. Klade však důraz na to, že tou klíčovou náboženskou inovací nebyla víra v moralizující božstva, ale nová forma kolektivní rituální praxe, která v určitých společnostech sehrála zásadní roli při posilování kolektivní identity a společenských institucí.[29]

Náboženské dějiny antického Středomoří jsou v diskuzi k těmto třem hypotézám nezřídka zmiňovány. Zastánci hypotézy velkých bohů tak např. odkazují ke křesťanství jako k vzorovému příkladu nového typu náboženského systému, který se nanejvýš hodil k podpoře spolupráce v tak komplexní společnosti, jako byla ta římská. Obhájci hypotézy blahobytu

[27] K hypotéze velkých bohů viz zejména Ara Norenzayan: *Big Gods: How Religion Transformed Cooperation and Conflict*, Princeton 2013; Ara Norenzayan et al.: The Cultural Evolution of Prosocial Religions, *Behavioral and Brain Sciences* 39, č. 1 (2016). Variantou této hypotézy je pak hypotéza nadpřirozeného trestu: Dominic Johnson: *God Is Watching You: How the Fear of God Makes Us Human*, New York 2016; Joseph Watts et al.: Broad Supernatural Punishment but Not Moralizing High Gods Precede the Evolution of Political Complexity in Austronesia, *Proceedings of the Royal Society of London B: Biological Sciences* 282, č. 1804 (2015): s. 20142556.

[28] Nicolas Baumard – Coralie Chevallier: The Nature and Dynamics of World Religions: A Life-History Approach, *Proceedings of the Royal Society B: Biological Sciences* 282, č. 1818 (2015), s. 20151593; Nicolas Baumard et al.: Increased Affluence Explains the Emergence of Ascetic Wisdoms and Moralizing Religions, *Current Biology* 25, č. 1 (2015), s. 10–15.

[29] Harvey Whitehouse et al.: Complex Societies Precede Moralizing Gods throughout World History, *Nature* 568, č. 7751 (2019), s. 226–229.

poukazují na klasické Řecko 5. a 4. stol.př.n.l. a takzvanou „osovou dobu“ jako na období, kdy dosažení určité úrovně blahobytu mohlo vést k celkově proměně preferenčního spektra, a tedy i k prosazení představ o posmrtných odměnách a trestech. Nakonec obhájci rituální teorie se velice kriticky vymezují vůči oběma předchozím s tím, že to podstatné se událo mnohem dříve, a to na samém prahu zemědělské revoluce, kdy první větší společnosti přistoupily k určitým rituálním inovacím.[30] S tímto přesvědčením v pozadí zastánci rituální hypotézy na adresu hypotézy blahobytu tvrdí, že řecké náboženství bylo v období cca. od 8. do 3. stol. př.n.l. relativně neměnné a že „[m]nohé klíčové znaky zde byly před tímto obdobím.“[31] Co se pak týče představ o bozích, tak tito autoři zdůrazňují, že nejenže „[ř]ečtí bohové určitě nebyli všemocní“ ale především „se příliš nezajímali o to, co lidé z morálního hlediska činí, dokud pokračovali v účasti na příslušných rituálech.“[32]

Ve vztahu k našemu korpusu a k tomu, jak reflektuje proměny v porozumění termínu *θεός*, můžeme z těchto tří hypotéz následně vyvodit tři odlišné predikce. Z hypotézy velkých bohů lze vyvodit predikci, že čím větší sociální komplexita, tím více lze očekávat propojení termínu *θεός* s termíny z oblasti morálky. Hypotéza blahobytu oproti tomu tvrdí, že propojení termínu *θεός* s termíny z oblasti morálky lze očekávat spíše v bohatších společnostech a že sociální komplexita jako taková je vedlejší. Nakonec zastánci rituální hypotézy explicitně tvrdí, že bez ohledu na sociální komplexitu či úroveň blahobytu, propojení termínu *θεός* s termíny z oblasti morálky je primárně záležitostí křesťanství.

Korpus byl výše podle období rozdělen na tři části: (1) archaickou, (2) klasickou a (3) římskou/křesťanskou. Lze předpokládat, že tato období lišila, jak co se týče sociální komplexity, tak co se týče blahobytu. Aniž bychom zacházeli do podrobností, již na základě obecného srovnání je možné konstatovat, že klasické Řecko bylo bohatší a sociálně komplexnější než archaické Řecko. Někteří renovovaní badatelé v posledních letech dokonce tvrdí, že klasické Řecko ve smyslu území spravovaného řeckými městskými státy bylo podle

[30] Harvey Whitehouse: Ritual and Social Evolution: Understanding Social Complexity Through Data, in *Computational History and Data-Driven Humanities*, ed. Bojan Bozic et al., Cham 2016, s. 3–14.
[31] Daniel Austin Mullins et al.: A Systematic Assessment of ‘Axial Age’ Proposals Using Global Comparative Historical Evidence, *American Sociological Review* 83, č. 3 (2018), s. 596–626.
[32] Ibid.

určitých kritérií v přepočtu na obyvatele dokonce bohatší než Římská říše.[33] Platí však, že Římská říše byla podle určitých měření sociálně komplexnější.[34]

Nás nyní zajímá, zda a případně jak by se tyto rozdíly mohly projevit v našich datech. V tomto smyslu se opět vrátíme k algoritmům pro tvorbu, analýzu a vizualizaci kookurenčních sítí. Budeme se zde však muset poprvé jasně a čelem popasovat s problémy, které s sebou použití těchto metod nese. Zatímco v případech výše jsem vždy pracovali pouze s jednotlivými, relativně krátkými dokumenty o podobném rozsahu,[35] tak nyní musíme na prvním místě rozhodnout, jak postupovat v případě řádově rozsáhlejšího korpusu.

Kookurenční sítě a korpus LG

Jedním z možných způsobů je vytvořit kookurenční sítě (a z nich vycházející sítě nejbližšího sousedství termínu *θεός*) pro každý jednotlivý dokument v našem korpusu. To nám následně umožní porovnat, jak se v průběhu času proměňuje kontext použití termínu *θεός*. Nás bude zajímat, zda a případně kdy se tento kontext stává více spjatý s oblastí morálky.

Za tímto účelem potřebujeme na prvním místě rozhodnout, která slova lze klasifikovat jako indikující morální kontext. Tento problém není v kontextu výpočetní textové analýzy nový. Řešila jej také skupina badatelů kolem Jonathana Haidta a Jesse Grahama, proponentů teorie o pěti či šesti základních pilířích morálky,[36] kteří pro tento účel vytvořili anglicko-jazyčný „Moral Foundation Dictionary".[37] Tento slovník tito badatelé původně použili k rozlišení morálních důrazů v kázáních proslovených v liberálních a konzervativních církvích ve Spojených státech.[38] Slovník obsahuje 295 anglických slov a slovních kořenů rozdělených do jedenácti kategorií. Zatímco prvních deset kategorií vychází z jednotlivých pilířů morálky podle Haidtovy teorie, kategorie 11 je věnována obecně morálním termínům, přičemž jedna část obsahuje pozitivní termíny (např. „right" či „good" apod.) a druhá termíny negativní (např. „wrong" či „bad"). V následujících analýzách použijeme pouze pozitivní termíny z této obecné

[33] Tak zejména Josiah Ober: *The Rise and Fall of Classical Greece*, Princeton – Oxford 2015, s. 147: „Comparisons can be misleading, but, by certain measures (aggregate and per capita economic growth, urbanization, and income distribution), the overall Greek economy of ca. 500–300 BCE appears to have outperformed the overall Roman economy of ca. 100 BCE–200 CE."

[34] Jako jedno z měřítek sociální komplexity se uvadí velikost hlavního města, kde Římská říše s milionovým Římem nemá v cca. čtvrtmilionových Aténách konkurenci. Srov. Ian Morris, *The Measure of Civilization: How Social Development Decides the Fate of Nations* (Princeton – Oxford 2013.

[35] Nejkratší Markovo evangelium má uvnitř GNT 11 277 slov, zatímco nejdelší Lukášovo evangelium má 19 456.

[36] Jonathan Haidt: *Morálka lidské mysli: Proč lidi rozděluje politika a náboženství*, Praha 2013.

[37] Viz https://moralfoundations.org/other-materials/.

[38] Jesse Graham – Jonathan Haidt – Brian A. Nosek: Liberals and Conservatives Rely on Different Sets of Moral Foundations, *Journal of Personality and Social Psychology* 96, č. 5 (2009), s. 1029–1046.

kategorie. Ty budeme následně považovat obecně považovat za morální indikátory. Jedná se celkově o 22 slov (resp. slovních kořenů).[39] Na tomto místě je potřeba zdůraznit, že se zde zcela záměrně ponecháváme tento seznam morálních indikátorů tak, jak je, a nepřidáváme do něj žádné další termíny, ani z něj žádné neubíráme. Činíme tak v zájmu minimalizace rozhodnutí, kterými bychom subjektivně zatížili naši další analýzu. S tímto se můžeme nyní vrátit k našemu korpusu lematizovaných antických řeckých textů.

Zde tedy pracujeme s korpusem o 687 dokumentech řeckých textů. V rámci tohoto korpusu 488 dokumentů obsahuje alespoň jeden výskyt termínu *θεός*. Když ze všech dokumentů v našem korpusu vytvoříme kookurenční síťě podle parametrů popsaných výše, vidíme, že termín *θεός* je součástí 456 z nich, což napovídá, že se v těchto dokumentech také nachází mezi 500 nejfrekventovanějšími lemmaty.[40] V případě těchto 456 dokumentů pak můžeme rovnou vytvořit síť 30 nejbližších sousedů obklopujících termín *θεός*.

V dalším kroku všechny termíny uvnitř těchto sítí podrobíme automatickému překladu. Tím získáváme alespoň jeden možný anglický význam v 91,23 % případů a v 67,71% případů více než jeden možný význam daného slova. Tyto automaticky získané významy mohou být následně porovnány s výše uvedenými morálními indikátory. Toho je dosaženo pomocí algoritmu, který pro každou síť nejbližších sousedů počítá termíny, u kterých alespoň jeden z možných významů začíná stejně jako některé slovo či slovní kořen mezi našimi 22 morálními indikátory. Pro každý dokument takto získáváme hodnotu mezi 0 (=žádný morální indikátor uvnitř příslušné sítě) a hypotetickými 30 (=všechny termíny uvnitř příslušné sítě klasifikovány jako morální indikátory). A tato hodnota představuje hlavní metriku, která nás nyní zajímá.

Indikátory morálky v sousedství termínu *θεός* v korpusu LG

Pro účely této studie se omezíme víceméně pouze na deskriptivní statistiku. Pokročilejší analýza by vyžadovala věnovat také nemalou pozornost otázkám měření sociální komplexity a blahobytu v historických společnostech obecně a v antickém Středomoří zvláště, což je samo

[39] Těmito termíny jsou: „righteous“, „moral“, „ethic“, „value“, „upstanding“, „good“, „goodness“, „principle“, „blameless“, „exemplary“, „lesson“, „canon“, „doctrine“, „noble“, „worth“, „ideal“, „praiseworthy“, „commendable“, „character“, „proper“, „laudable“, „correct“.

[40] Pro přípomenutí: Je tomu tak proto, že algoritmus vytvářející síť započítává hranu mezi slovy pouze tehdy, pokud se obě slova nachází v lexikonu určitého předvoleného množství nejčastějších slov. Ve všech případech zde jsme pracovali pouze s lexikonem o velikosti 500 slov. Dále: jelikož 287 z 456 dokumentů, ze kterých máme kookurenční sítě, obsahuje více než 500 unikátních lemmat, můžeme dovodit, že v naprosté většině případů 30 nejbližších sousedů představuje pouze zlomek z celé sítě, a tudíž nás skutečně informuje o kontextu použití zvoleného termínu. V 85,53 % dokumentů je kookurenční síť tvořena z textu čítajícího více než 300 unikátních lemat, kde tedy síť 30 nejbližších sousedů termínu *θεός* představuje nanejvýš 10 % všech uzlů.

o sobě velice rozsáhlé téma.[41] Pro naše potřeby postačí se v tomto smyslu podívat pouze na některé základní trendy v příslušných datech, které mohou být nějak vztaženy k hypotézám výše. Základní přehled si můžeme učinit pomocí Tabulky 5.

period	century	pagan docs	M	std	christian docs	M	std
archaic	8BC	2	0	0			
	7BC	3	0	0			
	6BC	1	0	0			
classical	5BC	51	0.82353	0.9941			
	4BC	83	1.50602	1.24326			
	3BC	1	0	0			
	2BC	1	1	0			
	1BC	0					
roman /christian	1CE	2	2.5	2.5	29	1.00	0.70711
	2CE	77	0.70130	0.90416	22	1.31818	0.99457
	3CE	32	0.87500	1.09985	27	1.14815	1.32153
	4CE	97	1.08247	1.02744	28	1.21429	1.03126

Tabulka 5. Přehled výsledků podle století; „docs" - počet dokumentů z daného období; „mean" - průměrný počet morálních indikátorů mezi 30 nejbližšími sousedy termínu θεός; „std" - standardní odchylka od průměru.

Alespoň jeden morální indikátor byl napočítán v 284 z celkových 456 dokumentů, což představuje přibližně 62,29 %. Zaprvé si můžeme povšimnout, že ani jeden morální indikátor nebyl napočítán v ego-sítích vytvořených z šestice dokumentů v našem korpusu, které spadají do tzv. *archaického* období, tj. 8. až 6. stol.př.nl.[42] Zcela jinak tomu je obrátíme-li pozornost k textům z *klasického* období, tj. z 5. a 4. stol.př.n.l., kde máme již výrazně vyšší počet dochovaných dokumentů. Jednu z nejvyšších průměrných hodnot počtu morálních indikátorů mezi nejbližšími sousedy termínu *θεός* nacházíme právě zde, v textech z 4. stol.př.n.l. (*N=83, M*=1.50602, *std*=1.24326).

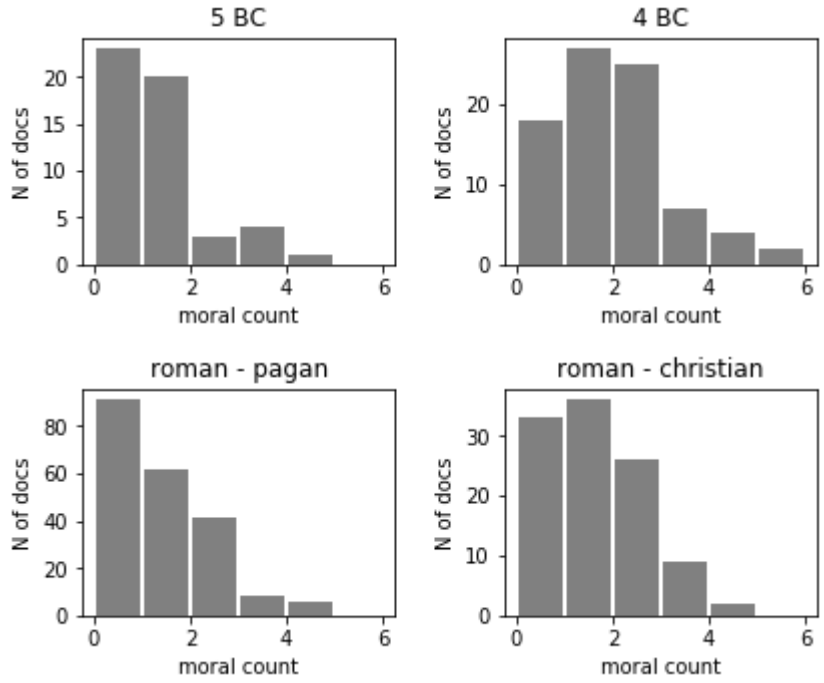
Srovnatelná data pak opět vidíme v dokumentech z římského období, kde jsou vedle sebe dokumenty křesťanské a pohanské.[43] V křesťanských dokumentech pozorujeme poměrně jasně vyšší průměrné hodnoty (*N= 106, M=1.16038, std=1.02480*) než v dokumentech pohanské provenience z téhož období (*N=208, M=0.92308, std=1.02333*). Neměli bychom

[41] Viz zejména Peter Turchin et al.: Quantitative Historical Analysis Uncovers a Single Dimension of Complexity That Structures Global Variation in Human Social Organization, *PNAS* 115, č. 2 (2018), s. E144–51; I. Morris: *The Measure of Civilization*.

[42] Konkrétně se jedná o homérovské eposy *Illias* a *Odysea*, Hésiodova díla *O původu bohů*, *Práce a dni* a *Štít* a ezopské bajky. Ve všech těchto dokumentech však termín *θεός* není nijak marginální.

[43] Do prvního století jsou v našem korpusu datovány pouze 2 pohanské texty, a totiž román *Příhody Chairea a Kallirhoy* od Charitóna z Afrodisias a *Řeči* Dióna z Prúsy. V tomto druhém dokumentu identifikujeme 4 morální indikátory, což se výrazně projevuje v průměrných hodnotách pro toto období.

však přehlédnout, že ani v případě křesťanských textů tyto hodnoty nejsou tak vysoké jako v případě pohanských textů z 4. stol.př.n.l. Tyto rozdíly jsou dobře patrné na histogramech na Obrázku 4.



Obrázek 4: Počty morálních indikátorů mezi nejbližšími sousedy termínu θεός.

Nyní bude vhodné si tato pozorování shrnout. Zaprvé, poměrně jasně se zde ukazuje, že hodnoty provázanosti mezi termínem *θεός* a obecně morálními termíny se v korpusu dokument od dokumentu výrazně liší. Zadruhé, v tomto ohledu si jistě všimneme určitého vzestupného trendu od nejstaršího období až po 4. stol.př.n.l. Zatřetí, v souladu s běžným očekáváním pozorujeme, že hodnoty pro křesťanské texty jsou vyšší než pro pohanské texty, ať už celkově, nebo zaměříme-li pozornost pouze na dokumenty ze srovnatelného období. Nakonec začtvrté, což je naopak poměrně zajímavá informace, křesťanské texty vykazují v průměru nižší hodnoty než texty ze 4. stol.př.nl.

Obtížnější je stanovit, co z těchto pozorování můžeme vyvodit s ohledem na dějiny náboženství a potažmo ve vztahu k výše představené debatě o kulturní evoluci moralizujících náboženství. Zde se skrývá celá řada úskalí. Co naše série algoritmů počítající obecně morální termíny v blízkosti termínu *θεός* vlastně měří? Analýza je postavena na celé řadě předpokladů, které jsme doposud pořádně nevyslovili. Zaprvé je zde předpoklad, že *kontext použití* termínu *θεός* uvnitř nějakého textu je vhodným výchozím bodem k tomu, chceme-li získat alespoň nějaké informace o tom, jak si autor vybraného textu představuje boha či bohy ve smyslu určité kategorie představ o nadpřirozených činitelích. Druhý nesamozřejmý avšak důležitý předpoklad je, že vyšší míra spoluvýskytů termínu *θεός* a slov, která jsme identifikovali jako

morální indikátory, znamená, že si autor příslušného textu představuje tyto nadpřirozené činitele jako více moralizující. Třetím předpokladem je, že tato míra provázanosti byla skutečně věrně zachycena námi implementovanými algoritmy. A nakonec je zde obsažený předpoklad, že na základě jednotlivých dokumentů a jejich autorů můžeme formulovat tvrzení o společenských trendech na úrovni populací s implikacemi pro debatu o obecných mechanismech v pozadí kulturní evoluce náboženství. Pokud na chvíli s vědomím všech možných nedostatků uznáme tyto předpoklady za odůvodněné, můžeme se zamyslet nad tím, co naše pozorování implikují pro výše nastíněnou debatu na tomto poli.

Zastánci hypotézy blahobytu ve svých publikacích předestřeli, že pokud chceme porozumět vzniku moralizujících náboženství, měli bychom se podívat na určité mezníky v ekonomickém vývoji nejvýznamnějších starověkých civilizací. V kontextu náboženských dějin antického Středomoří tak spatřují důležitý mezník v období klasického Řecka, se kterým se pojí řada kulturních inovací. Zastánci rituální hypotézy však v klasickém Řecku tento mezník nerozpoznávají. Zlomové okamžiky pro kulturní evoluci náboženství spatřují jinde, ať už v mnohem starší minulosti, na samotném úsvitu zemědělské revoluce, nebo v příchodu velkých doktrinálně orientovaných náboženských systémů, jako je křesťanství. Zcela v souladu s tímto stanoviskem tak proto Mullins se svými kolegy tvrdí, že řecké náboženství bylo v období před naším letopočtem víceméně neměnné, že nebylo nijak spjato s oblastí oblastí morálky a že morální element vstoupil zásadně do hry až s příchodem křesťanství.[44]

Výše představená pozorování však všechna tato tvrzení přinejmenším oslabují. Naše data naznačují, že v období před naším letopočtem se představy o bozích zásadně měnily. Zejména se zdá, že tyto představy byly směrem ke 4. stol.př.n.l. období více a více provázány s oblastí morálky. Nakonec nic v našich datech nenasvědčuje tomu, že by výrazné propojení mezi představami o bozích a morálkou bylo až záležitostí křesťanství. O to spíš platí, že naše data mohou být velice dobře interpretována v duchu hypotézy blahobytu a představovat pro ni určitou formu empirické podpory. Avšak jak již bylo výše předestřeno, detailní vyhodnocení této teze již překračuje možnosti této studie.

Závěr

V tomto textu jsme si představili metody vytváření, analýzy a vizualizace kookurenčních sítí jako příklad distančního čtení, které umožňuje získat cenné náhledy týkající se obsahu vybraných textů a způsobů, jakým jsou v nich použity vybrané termíny. Než jsme postoupili

[44] D. Mullins et al. A Systematic Assessment of 'Axial Age' Proposals, SI.

k samotným analýzám, ukázali jsme si nutné kroky v rámci předzpracování příslušných dat, zejména jejich lematizaci.

V prvním souboru příkladů jsme obrátili pozornost k českému překladu novozákonních evangelií, které nám díky tomu, že se jedná o relativně krátké a obecně dobře známý soubor textů, umožnili porovnat naše výsledky s našimi běžnými znalostmi. V následující části již jsme obrátili pozornost k řádově rozsáhlejšímu korpusu textů v řečtině, přičemž jsme se zaměřili na to, jak příslušné metody mohou být použity\ v zájmu prozkoumání ambiciózních makro-historických hypotéz. Na jednu stranu jsme si tak mohli ukázat potenciální síly představených metod, na stranu druhou tak však jasně vystoupila na povrch i značná rizika a problémy. Je to však právě uvědomění si těchto rizik a problémů, které nás posouvá dál k rozvíjení pokročilejších a pokročilejších nástrojů, pomocí nichž si můžeme „četbou na dálku“ povšimnout i věcí, kterých jsme si při „četbě zblízka“ nevšimli.