

**Univerzita Karlova
Filozofická fakulta**

Ústav informačních studií a knihovnictví

Disertační práce

Mgr. Josef Šlerka

Normalized Social Distance

Normalized Social Distance

Vedoucí práce doc. RNDr. Jiří Souček, DrSc.

2019

Prohlašuji, že jsem disertační práci napsal/a samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 25. 3. 2019

Mgr. Josef Šlerka

Poděkování

Rád bych poděkoval vedoucímu své disertační práce a dalším akademikům, kteří mě inspirovali. Děkuji také svojí rodině a přátelům, kteří se mnou měli až nadpozemskou trpělivost.

Abstrakt

V této disertační práci se zabýváme aplikací koncepce informační vzdálenosti na analýzu dat ze sociálních sítí. Tato data chápeme jako zaznamenané akty sociálního jednání. Jako taková jsou výrazem určitých postojů, hodnot a záměrů. V práci zavádíme vzorec pro výpočet normalizované sociální vzdálenosti a na základě série případových studií dokazujeme užitečnost a validitu toho přístupu. Možnost aplikovat formální matematické a infromatické postupy na masivní datové záznamy jednání lidí v prostředí sociálních sítí je umožněna změnou, kterou přinášejí nová média a s nimi spojená technologická změna. Tato změna je provázena postupnou proměnou v metodách výzkumu v humanitních vědách označovanou jako nástup tzv. digital humanities. Pro tento přístup je charakteristické aplikování kvantitativních metod v oblasti humanitních věd a objevování nových datových oblastí využitelných pro analýzy. Specificky v případě dat ze sociálních médií pak dochází k setření kvantitativních a kvalitativních metod. Příkladem může být i výsledek této disertace, kdy informační teorie specificky spojuje metody klasické social network analysis a goffmanovského rámce analýzy konkrétního lidského jednání.

Klíčová slova: informační vzdálenost, normalized social distance, Kolmogorova komplexita, data mining, social media, digital humanities, Facebook, Twitter

Abstract

This dissertation thesis deals with the application of the concept of information distance to the social network data analysis. We consider this data as recorded acts of social action. As such, they express certain attitudes, values and intentions. We introduce a formula for calculating the Normalized Social Distance and, based on the series of case studies, we prove the usefulness and validity of this approach. The application of formal mathematical and computer science techniques to massive data records of human action in social network environments is enabled by the change brought by new media and the associated technological advancement. This change is accompanied by a gradual transition of research methods in the humanities, referred to as the onset of digital humanities. This approach is characterized by the application of quantitative methods in the field of humanities and the discovery of new data areas useful for analyses. In case of social media data, the differentiation between quantitative and qualitative methods is no longer valid. A good example is also this thesis, in which information theory specifically combines the methods of a traditional social network analysis and the Goffman's frame analysis of human action.

Keywords

Information distance, Normalized Social Distance, Kolmogorov complexity, social media, data mining, digital humanities, Facebook, Twitter

Úvod

1. Základní vymezení výzkumného prostoru a jeho kontextů
 - 1.1. Digital humanities
 - 1.2. Nová média
 - 1.3. Data, informace, znalosti a vědění
 - 1.4. Big data
 2. Výchozí výzkumný přístup pro analýzu jednání v sociálních sítích
 - 2.1. Sociální média
 - 2.2. Social network analysis
 - 2.2.1. Případová studie I. – Účty sledující účet @stunome na Twitteru
 - 2.2.2. Případová studie II. – Analýza geospatálních dat ze sociální sítě Foursquare
 - 2.3. Dramaturgická sociologie E. Goffmana a sociální média
 - 2.3.1. Případová studie I. - Jména uživatelů komentujících na facebookových stránkách
 - 2.3.2. Případová studie II. – Fanouškovství stránek na Facebooku jako klíč k pochopení povahy fanoušků
 - 2.3.3. Případová studie III – Úvodní fotografie fanoušků subkulturních stránek na Facebooku
 - 2.3.4. Případová studie IV. – Sebe prezentace IVČRN na českém Facebooku
 3. Normalized Social Distance
 - 3.1. Modely podobnosti
 - 3.2. Teorie Kolmogorovy komplexity
 - 3.3. Informační vzdálenost a Normalized Compression Distance (NCD)
 - 3.4. Normalized Web Distance
 - 3.5. Normalized Social Distance
 - 3.5.1. Případová studie I. – Analýza fanoušků DSSS
 - 3.5.2. Případová studie II. – Who is Shaping Your Agenda?
 4. Případové studie podporující validitu
 - 4.1. Případová studie I. – Tzv. prokremelské weby na českém Facebooku
 - Co sdílejí aktivní uživatelé
 - Kde ještě lajkují aktivní lajkeři?
 - 4.2. Případová studie II. – Politické mapy a volební výsledky
 - 4.3. Případová studie III. – Celebrity napříč sociálními sítěmi
 5. Závěr
- Appendix – Vytvořené aplikace a knihovny

Sugar

Geosugar

Who are likers?

Page Affinity Analyzer

Normalized Social Distance Matrix Viewer

Normalized Facebook Distance Matrix Viewer

Seznam použitých obrázků

Seznam tabulek

Seznam použité literatury

Úvod

Jádrem předkládané disertační práce a jejím výzkumným přínosem je návrh metriky pojmenované **Normalized Social Distance** (NSD), která slouží k měření vzdálenosti mezi různě vymezenými skupinami hodnot, zájmů či postojů v sociálních sítích na základě průniků mezi jednajícími. Průběžně v práci přinášíme řadu případových studií ověřujících tento přístup na základě dat z online sociálních sítí. Jedná se tak o příspěvek nejen k sociálním vědám z pohledu informatiky, ale také o ukázkou tzv. počítačného obratu v sociálních vědách či v širším smyslu obratu označovaného jako nástup tzv. digital humanities. Ty jsou jen důsledkem mnohem širšího fenoménu, který pracovně vymezujeme jako éru tzv. nových médií (byť s vědomím všech pojmových rizik, která se k tomuto pojmu váží).

Ačkoli práce spadá do věd humanitních a sociálních, její kořeny leží v informační vědě. Samotný vzorec NSD je totiž odvozen od myšlenky informační vzdálenosti (information distance) propagované zejména Paulem Vitanyim a Rudi Cilibrasim, která má svůj původ v Kolgomorově teorii komplexity a ústí do tzv. algoritmické teorie informací (algorithmic information theory). Ta představuje rozšíření klasické teorie informací, která tradičně svůj původ odvozuje od prací Clauda Shannona.

Vitanyi a Cilibrasi rozvinuli původně abstraktní myšlenku směrem k praktické aplikaci v oblasti měření informační vzdálenosti libovolného řetězce v podobě tzv. Normalized Compression Distance, která se de facto stará jen o syntax a od myšlenky kompresních algoritmů se dostali k vzniku tzv. Normalized Web Distance, která aplikuje základní myšlenku NCD na strojově měřitelnou sémantickou blízkost.

Posun od Normalized Compression Distance k Normalized Google Distance můžeme chápat jako posun od de facto syntaxe k sémantice. V této práci bychom rádi navrhli další posun, a to od sémantiky k pragmatice, konkrétně posun k měření blízkosti různých druhů lidských hodnot, jednání a postojů. Díky námi navrženému vzorci tak uceleně vidíme pohyb informatického přístupu od oblasti syntaxe přes sémantiku až po pragmatiku.

Než se dostaneme k jádru disertace, je třeba vysvětlit kontext, v němž se práce pohybuje. **První část** se věnuje celkovému rámci a kontextu, v němž se výzkum samotný odehrává.

Centrálními místy jsou pojmy digital humanities a pojem nových médií. Dále pak vymezení toho, co jsou vlastně data a některé pojmy s nimi spojené.

Část druhá se týká specifické oblasti nových médií, kterou jsou tzv. sociální média, respektive social network sites. Vycházíme přitom z vymezení dat produkovaných uživateli sociálních médií jako záznamů intencionálního jednání a věnujeme se propojení dvou klasických přístupů k této analýze. Prvním je tzv. social network analysis, která představuje strukturální pohled, kvantitativní pohled na konstrukci sociálních skupin a interakce v něm. Tento pohled je na straně druhé doplněn přístupem inspirovaným dramaturgickou sociologií E. Goffmana, která se naopak věnuje kvalitativním mikroanalýzám a podle našeho názoru zdařile doplňuje kvantitativní přístup SNA.

Oba teoretické přístupy pak tvoří základní východisko původního návrhu této práce: jak určit vzdálenosti mezi skupinami na sociálních sítích. Postup je odvozen od koncepce tzv. Normalized Social Distance, které se věnuje **část třetí**. Vycházíme přitom z tzv. Kolmogorovy teorie komplexity a z ní odvozené koncepce informační vzdálenosti. Tento koncept byl aplikován v oblasti informační vědy zejména Vitanyim a Cilibrasim a my jej doplňujeme o rozměr umožňující formalizovat vzdálenosti sociálních skupin v prostředí sociálních sítí, a to na základě jednání jejich uživatelů.

Čtvrtou část práce tvoří tři případové studie, které se snaží prozkoumat stabilitu a validitu postulované metriky, a to nejen uvnitř jedné sítě (Facebook), ale také mezi jednotlivými sítěmi (Facebook, Twitter, Instagram) a také vůči offline světu (Facebook vs. volby do Evropského parlamentu).

V **Závěru** se otevírá debata k hranicím samotného přístupu, ale také k řadě dalších možných aplikací. **Appendix** pak obsahuje krátký popis počítačových programů, které vznikly jako součást této disertační práce.

1. Základní vymezení výzkumného prostoru a jeho kontextů

V první části práce se nejdříve pokusíme zasadit vlastní výzkum do širšího oborového kontextu a soustředíme se na vymezení některých klíčových oblastí a pojmů, které hrají klíčovou roli pro současnost samotného výzkumu vycházejícího z dat ze sociálních médií. Východiskem pro nás bude základní uvedení metodologického přístupu pojmenovaného jako digital humanities, které tvoří celkový rámec našeho přístupu.

1.1. Digital humanities¹

Hnutí okolo tzv. digital humanities přichází jako reakce humanitních a sociálních věd na rozvoj výpočetní techniky a nových (digitálních) médií od poloviny minulého století. S nástupem počítačů se postupně objevují projekty využívající jejich výpočetní sílu a možnosti zpracovávat množství informací, kterou jsou jinak prakticky neuchopitelné jednotlivcem.

Za jednoho z prvních průkopníků v této oblasti se považuje jezuita Roberto Busa, který přišel již v roce 1949 s projektem kompletního zpracování index verborum díla Tomáše Akvinského. Projekt samotný byl nakonec realizován v letech 1974–1980 v tištěné verzi 56 svazků za podpory společnosti IBM. Při projektu bylo dílo Tomáše Akvinského kompletně přepsáno do děrných štítků, z nich byla poté vytvořena počítačem úplná konkordance. Dnes je celý *Index Thomisticus* přístupný na webu.² Díky tomu je možné během zlomků sekund vyhledat všechny výskyty daného slova v objemném díle.

Tento průkopnický projekt dobře ukazuje základy přístupu v digital humanities. Byl umožněn výpočetní silou a novými médii, mění samotný přístup k výzkumu, a konečně proměňuje i celkový přístup/interface k informacím, z nichž výzkum vychází. Právě tyto změny vymezují prostor digital humanities.

Postupný nástup počítačů v druhé polovině 20. století a zejména jejich neustále klesající cena přinášejí zásadní změnu i ve vědecké práci. Při původním využívání výpočetní síly ve vědě se

¹ Části této kapitoly jsou součástí doslovu k českému překladu knihy *Digital humanities*, která by měla vyjít v nakladatelství Academia.

² *Index Thomisticus* [online]. [cit. 2019-02-23]. Dostupné z: <http://www.corpusthomisticum.org/it/index.age>

výzkumník potýkal s relativně drahým sběrem dat ve větším množství a s jejich poměrně finančně náročným zpracováním. Proto byl velký důraz kladen především na formulaci hypotézy a její potvrzování (případně falzifikaci). Od 80. let se však situace mění, díky dostupnému výkonu a možnostem softwaru se postupně dostává do popředí tzv. explorační analýza dat, při níž může výzkumník procházet velké množství dat tak, aby našel větší počet hypotéz, které stojí za to dále testovat. K technikám explorační datové analýzy patří kupříkladu multidimensional scaling, PCA, ale také běžné vizualizace jako je box-plot či scatter-plot.

Původně čistě statistický přístup je doplněn o metody tzv. data miningu, tedy přístupu, který se snaží o získávání netriviálních informací z analýzy dat a který vychází především z tradic počítačové vědy. Mezi základní data-miningové techniky patří kupříkladu rozhodovací stromy, asociační pravidla, neuronové sítě, regresní analýza či shluková analýza. Náleží sem ovšem i nejrůznější formy sumarizace včetně vizualizací.³

Zatímco počítače představují motor nového, digitálního světa, tak digitalizovaná (či v digitálním prostředí vzniklá) data jsou jeho pohon. V procesu postupné digitalizace podstatné části lidského světa a jeho historie se otevírají zcela nové možnosti pro výzkum jak v oblasti humanitních, tak sociálních věd.

Můžeme sledovat kupříkladu ambiciózní projekt využívající poměrně komplikovaných postupů (jakým je kupříkladu aplikace algoritmu Latent Dirichlet Allocation) na modelování témat v rámci korpusu anglicky psané literatury 19. století jako ve studii *Significant themes in 19th-century literature* od Matthewa L. Jockerse a Davida Mimno.⁴ Autoři používají techniku modelování témat pro analýzu více než 3 400 děl americké, britské a irské literatury.

Jiné projekty se snaží využít možností, které nabízí vizualizace velkých obrazových korpusů podle definovaných vlastností. Můžeme tak třeba sledovat dynamiku vývoje světlosti obrazů moderny či zkoumat Mondrianovy obrazy s ohledem na jejich barevnost, jak to ukazují práce

³ Data mining – Wikipedia. Wikipedia [online]. [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Data_mining

⁴ JOCKERS, Matthew L. a David MIMNO. Significant themes in 19th-century literature. *Poetics*. 2013, 41(6), 750-769. DOI: 10.1016/j.poetic.2013.08.005. ISSN 0304422X. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0304422X13000673>].

Lva Manoviche v softwaru ImagePlot.⁵ Řada výsledků těchto analýz je dostupná v rámci projektu Cultural Analytics.⁶

Smyslem těchto přístupů je zajistit určité vzdálení se od zkoumaného korpusu. Franco Morretti v tomto ohledu hovoří o konceptu distant reading, jako opozici vůči klasickému anglosaskému modelu close reading. Cíle, které sledují jak Manovich, tak Morretti, jsou v jistém slova smyslu stejné. Mají zajistit možnost přehlédnout zkoumaný korpus v jeho větším celku, a objevit tak nějaké opakované vzory (patterns) či anomálie, které by jinak zůstaly skryty nebo se hůře vykazovaly.

Jako drobný příklad nám může posloužit analýza plakátů festivalu českého jazyka, řeči a literatury Šrámkova Sobotka, kterou autor této disertace připravil pro festivalový časopis *Splav!*. Jádrem byla vizualizace festivalových plakátů z jeho šedesátileté existence. Na ose X bylo pořadí plakátů podle ročníků, na ose Y pak světlost jednotlivých plakátů.

„Všimněte si, jak se Sobotka nebála dlouhou dobu experimentovat s formátem samotného plakátu. Jednou na šířku, jindy na výšku. Kromě klasických formátů volí i různé poměry stran. Obzvláště 19. a 21. ročník v tomto ohledu zřetelně vyčnívají. Proměňuje se také barevnost plakátů, ale i tak celkově převažuje modrá (a rudá), jak se na Šrámkův odkaz sluší. Témata i typy grafiky se různě střídají od ilustrací až k abstrakci a zase zpět při občasně tematické doslovnosti. Tato diverzita je přerušena téměř na 11 let období, které bych mohl nazvat „bílé“ a zasahuje léta 1986 až 1997. Během této doby se (více méně) zachovává jednotný princip. Vizualním jádrem je dívčí tvář a silueta Humprechtu ztvárněná malířem Zdeňkem Mlčochem, jenž oba motivy drobně variuje. Po smrti autora zůstávají ve své kanonické podobě. Toto období definitivně končí s koncem minulého století a plakáty se opět obracejí k větší diverzitě jak tematické, tak barevné, a vrací se tak zpět k původní rozmanitosti. S blížící se šedesátkou se ale objevuje další nový prvek: plakáty se formálně vrací k obyčejnosti, ale postmoderně si pohrávají se samotným žánrem plakátu,“ komentoval autor výsledky analýzy.⁷

⁵ Software Studies Initiative: ImagePlot visualization software: explore patterns in large image collections. Software Studies Initiative. Software Studies Initiative [online]. [cit. 2019-02-23]. Dostupné z: <http://lab.softwarestudies.com/p/imageplot.html>

⁶ Cultural Analytics [online]. [cit. 2019-02-23]. Dostupné z: <http://lab.culturalanalytics.info/>

⁷ ŠLERKA, Josef. Sobotka splaká tu [online]. 2016 [cit. 2019-02-23]. Dostupné z: <http://casopis.splav.cz/pdf/2016/splav-2016-07-03.pdf>



Obr. 1 – Vizualizace festivalových plakátů

Z historického pohledu ovšem hnutí digital humanities navazuje na historicky starší projekty spojené s využíváním kvantitativních metod v jazykovědě, literární vědě a podobných oborech. Tyto projekty drží krok s oblastmi sociálních věd, jako jsou sociologie a další. Řadu pokusů v tomto směru můžeme sledovat minimálně od počátku 20. století. Za všechny jmenujme třeba knihu německého teoretika Maxe Benseho *Teorie textu* z roku 1962 (dostupnou i v českém překladu) nebo knihu rumunského autora Solomona Marcuseho *Poetica matematică* z roku 1970. V těchto textech se rodí aplikace statistických a data-miningových metod zkoumání literárních textů. Mimochodem řadu inspirací z těchto oblastí najdeme i v některých vydáních *Poetiky* od Jozefa Mistríka, včetně originálního vzorce pro výpočet čitelnosti textu (tzv. readability formula), který se ve svém provedení výrazně liší od běžných vzorců.⁸

⁸ ŠLERKA, Josef a Filip SMOLÍK. Automatická měřítka čitelnosti pro česky psané texty. *Studie z aplikované lingvistiky*. 2010, 1(1), 33 - 44, 11 s. ISSN 18043240.

Specifickou kapitolu v tomto ohledu představuje práce s literárními korpusy a zrod moderní korpusové lingvistiky. Zde je dobré připomenut práci týmu *Korpusu českého verše*⁹, který ve svém projektu on-line nástrojů pro prohledávání a vizualizaci dat velmi dobře naplňuje představu o projektu, jaký patří do oblasti digital humanities.

Projekt už ale plně zapadá do oblasti výzkumů, jejichž využití v řadě oblastí humanitních a sociálních věd bylo umožněno až s nástupem počítačů a navazujícím rozvojem nových médií. Pro tyto nové způsoby výzkumu, které jsou ovšem často spojeny i se zcela novými druhy výsledků bádání, se vžilo označení digital humanities; byť, podobně jako v případě definování nových i sociálních médií, se vedou časté spory o přesnou definici.

Ve sborníku *Debates in the Digital Humanities*¹⁰ můžeme najít 21 odpovědí na otázku, co jsou digital humanities. Velká část těchto definic se týká především průniku mezi komputačními postupy a humanitními vědami, jako v případě Johna Unswortha z University of Illinois, který definuje digital humanities jako „využití komputačních nástrojů při práci humanitních věd“ (Gold 1998: 70).¹¹ Některá vymezení přidávají i fakt, že se jedná o komunitu výzkumníků, kteří se pohybují v oblastech zprostředkovaných digitálními médii. Tyto definice zjevně odkazují na skutečnost, že jedním z motorů výzkumu digital humanities je využívání nových médií jako prostředí pro samotné vytváření vědecké komunity.

Autoři učebnice *The Digital Humanities: A Primer for Students and Scholars*¹² vybírají dvě zásadně odlišné definice digital humanities, které podle nich dobře ilustrují napětí, jež s sebou koncept digital humanities přináší.

První definici, kterou autoři zmiňují, je definice z *Wikipedie*: „Digital humanities is an area of research and teaching at the intersection of computing and the disciplines of the humanities. Developing from the fields of humanities computing, humanistic computing and digital humanities praxis digital humanities embraces a variety of topics, from curating online collections to data mining large cultural data sets. Digital humanities (often abbreviated DH)

⁹ Korpus českého verše [online]. [cit. 2019-02-23]. Dostupné z: http://versologie.cz/v2/web_content/corpus.php

¹⁰ GOLD, Matthew K. *Debates in the digital humanities*. Minneapolis: Univ Of Minnesota Press, 2012. ISBN 978-0816677955.

¹¹ GOLD, Matthew K. *Debates in the digital humanities*. Minneapolis: Univ Of Minnesota Press, 2012. ISBN 978-0816677955. s. 70.

¹² GARDINER, Eileen a Ronald G. MUSTO. *The Digital Humanities: a Primer for Students and Scholars*. New York: Cambridge University Press, 2015. ISBN 978-110-7601-024.

currently incorporates both digitized and born-digital materials and combines the methodologies from traditional humanities disciplines (such as history, philosophy, linguistics, literature, art, archaeology, music, and cultural studies) and social sciences with tools provided by computing (such as data visualization, information retrieval, data mining, statistics, text mining, digital mapping), and digital publishing.“¹³

Tato definice je podle autorů učebnice v rozporu s inkluzivním vymezením Anne Burdick v publikaci MIT *Digital humanities*. „[Digital Humanities] asks what it means to be a human being in the networked information age and to participate in fluid communities of practice, asking and answering research questions that cannot be reduced to a single genre, medium, discipline, or institution.“¹⁴

Zatímco první z nich se orientuje hlavně na otázku metodologickou a vychází především z dialogu mezi počítačovými vědci a vědci z humanitních oblastí, v druhém případě se jedná o mnohem širší oblast, která de facto spadá do oblasti nazývané Manovichem „transcoding“, tedy vlivu digitální proměny lidské společnosti na člověka samotného. Autoři učebnice sami říkají, že by bylo dokonce lepší pojmenovat takto vymezené oblasti dvěma rozdílnými termíny. Metodologickou část pojmenovat jako digital humanities a část širší označit jako humanities computing. Sami se kloní k původnímu vymezení oblasti digital humanities jako primárně metodologické změny v samotné práci v humanitních vědách, která má svůj historický kořen především v nástupu počítačů. Navíc starší práce používají termín humanities computing jako označení samotného pole výzkumu, kde se protínají jednotlivé vědní oblasti a metodologie, jak to ilustruje obrázek z knihy Willarda McCartyho *Humanities Computing*¹⁵.

¹³ Digital Humanities – Wikipedia. Wikipedia [online]. [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Digital_humanities

¹⁴ BURDICK, Anne. *Digital humanities*. Cambridge, MA: MIT Press, c2012. ISBN 978-026-2018-470. str. vii.

¹⁵ MCCARTY, Willard. *Humanities Computing*. Paperback edition. Basingstoke, Hampshire: Palgrave Macmillan, 2014. ISBN 978-140-3935-045.

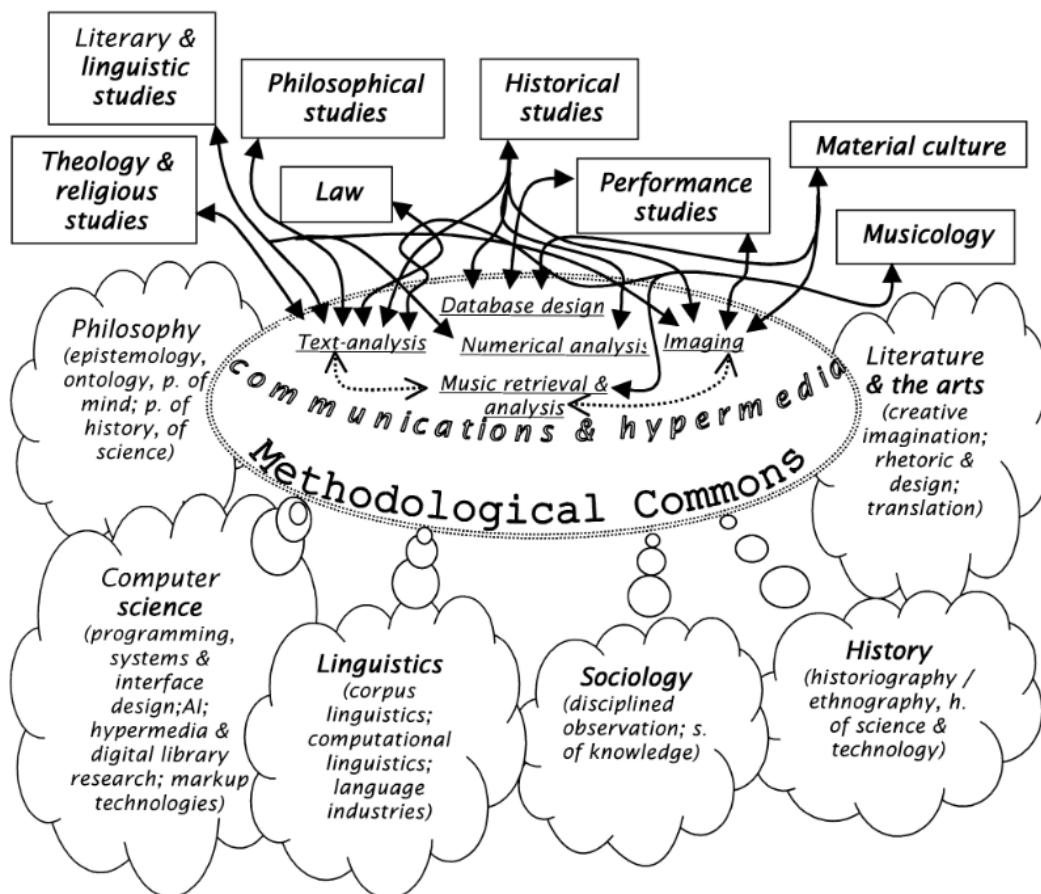


Fig. 1 A rough intellectual map for humanities computing. (Courtesy W. McCarty and H. Short.)

Obr. 2 – Mapa prostoru označovaného jako humanities computing

Nicméně obecná shoda v této oblasti nepanuje a snad ani není zcela nutná. Důležité je, že jsme vždy nejen na novém poli výzkumu, ale také přicházíme do styku s novým způsobem práce samotné výzkumnické komunity, která využívá nové technologie k efektivnější komunikaci, zrychlení cyklu průzkumu dat, zkoumání hypotéz, a konečně i k zpřístupňování velkých datových korpusů kolegům i širší veřejnosti.

Ačkoli je digital humanities relativně mladá disciplína (byť s celou řadou úctyhodných předchůdců), je podrobovaná řadě kritických pohledů nejen z pohledu kulturní kritiky, ale ze samotného pohledu metodologického. Tuto kritiku shrnují Bernhard Rieder a Theo Röhle v eseji *Digital Methods: Five Challenges*¹⁶ do pěti výzev, před kterými dle nich digital humanities dnes stojí. Jsou jimi: zdání objektivity (The lure of objectivity), síla obrazu (The

¹⁶ Theo Röhle a Bernhard Rieder: „Digital Methods: Five Challenges“ in BERRY, David M. Understanding digital humanities. New York: Palgrave Macmillan, 2012. ISBN 978-0-230-29265-9.

power of visual evidence), neznalost metody (Black-boxing), institucionální perturbace (Institutional perturbations) a problém universalismu (The quest for universalism).

Autor této disertace se kritiku pokusil shrnout v doslovu k českému vydání Morrettiho knihy *Graphs, maps, trees: Abstract models for a literary history*: „pokud věci počítáte, mohou se zdát objektivnější. Vizualizace dat s sebou nese často jen snadnější nahlédnutí pravdy, ale také rétorickou sílu, která může klamat. Moderní technologie přináší snadnou dostupnost matematických analýz pro všechny, aniž by bylo nutné rozumět jejich fungování. Často tak můžeme aplikovat datovou analýzu nevhodným způsobem, a i přesto dostat odpovědi, které dávají na první pohled smysl. Komputační metody s sebou nutně ponese větší zapojení vědců vzdělaných v jiných oblastech, a budou tak oslabovat původní kontexty, které humanitní vědy vymezují. Zapojení víceméně matematických modelů, ovšem také oživuje představu universální gramatiky světa, která svůj výraz nachází především v matematicky založených vědách.“¹⁷

Digital humanities představují dnes již relativně etablovaný přístup ve vědě. Jejich existence je ovšem mj. závislá na jiné změně ve světě, a to na vzniku zcela nové ontologické sféry vázané ve své existenci na oblast tzv. nových médií, o nichž budeme mluvit v následující kapitole.

¹⁷ J.Šlerka: Morettiho modely, Moretti jako model in Moretti: Grafy, modely, strom, str. 114 in MORETTI, Franco. Grafy, mapy, stromy: abstraktní modely literární historie. 1. Praha: Karolinum, 2014. Studia nových médií. ISBN 978-802-4626-093.

1.2. Nová média¹⁸

V běžném jazyce se můžeme setkat s označením nová média jako synonymem pro internet, mobilní telefony či herní konzole. Kupříkladu televizní společnosti mají oddělení nových médií, které se obvykle stará o prezentaci společnosti na internetu, případně sociálních sítích, či se věnuje výrobě aplikací pro počítače, tablety či chytré telefony. V obecnější rovině jsou tyto technologické artefakty spojeny s některými koncepty digitálnosti, interaktivity, hypertextuality, virtuality, síťovosti a simultánnosti.¹⁹

Digitálností je třeba rozumět číselnou, digitální reprezentaci informací, která přináší nejen dematerializaci existence artefaktů (zpravodajský web není spojený s papírem), ale také umožňuje komunikaci a práci prostřednictvím počítačů (například v internetové síti). Interaktivnost představuje možnost postoupit od pasivní recepce k zapojení diváka či čtenáře. Hypertextualita odkazuje k provázání textů mezi sebou, a především k nelineárnímu čtení. Virtualita pak znamená vytváření digitálních prostředí, která umožňují různé typy teleprezence. Síťovostí a simultánností se rozumí nejen nástup internetu, ale celkové uspořádání procesů ve společnosti.²⁰

Tyto změny existence a konzumace médií ale nespady z čistého nebe a jsou součástí širší změny, kterou obecně charakterizujeme jako nástup informační společnosti, v níž dochází k silnému prolínání techniky a kultury. Autoři monografie *New Media: a critical introduction* zasazují tuto současnou technokulturu do kontextu celospolečenských změn posledních 40 let. Jsou jim obrat od modernity k postmodernitě, intenzivní proces globalizace, přechod k postindustriální společnosti informačního věku na Západě a decentralizace geopolitických mocenských center.²¹ Tyto změny byly a jsou totiž spojeny právě s masivním rozvojem komunikačních a výpočetních technologií.

¹⁸ Některé části této kapitoly a kapitoly Sociální média jsou obsaženy v ŠLERKA, Josef. Nová a sociální média: Základní vymezení pojmů. DINGIR [online]. 2015(3) [cit. 2019-02-23]. Dostupné z: http://www.dingir.cz/cislo/15/3/nova_a_socialni_media.pdf

¹⁹ LISTER, Martin. *New media: a critical introduction*. 2nd ed. New York, N.Y.: Routledge, 2009. ISBN 978-041-5431-613, str. 13

²⁰ viz například CASTELLS, Manuel. *The rise of the network society*. 2nd ed., with a new pref. Malden, MA: Wiley-Blackwell, 2010. ISBN 978-140-5196-864.

²¹ LISTER, Martin. *New media: a critical introduction*. 2nd ed. New York, N.Y.: Routledge, 2009. ISBN 978-041-5431-613.str. 11

Je však nutné dodat, že kupříkladu samotná interaktivita či hypertextualita nejsou samy o sobě něčím kulturně novým a pro vymezení nových médií zcela dostačujícím. Autoři sborníku *The New Media Reader*²² představují čtenáři dlouhou kulturní historii řady těchto konceptů, kam patří mj. i tvorba J. L. Borgese, W. Burroughse či francouzské skupiny OuLiPo. Norský literární teoretik Espen Aarseth zase ukazuje, že texty určené k interakci nejsou novinkou ani z pohledu dějin masové kultury.²³ Pokud bychom chtěli hledat domácí předchůdce, pak příkladem interaktivního díla vzniklého dávno před nástupem nových médií je Radokův Kinoautomat.

Navíc z pohledu historického jsou vždy nějaká média (či obecně technologie) nová²⁴ a samotné označení nová média tak může mít povahu jen určité jazykové hry, jako označení pro média technologicky nová, nezastaralá.

Ostatně obtíže s vymezením pojmu nová média sleduje podrobně Jakub Macek ve své práci *Poznámky ke studiím nových médií*²⁵ a cílem této práce není kritická reflexe oboru. V Mackem citovaných Dahlbergových²⁶ třech základních přístupech k novým médiím se v této práci přikláníme k přístupu zaměřenému na technologickou determinaci spíše než na determinaci užitím nebo sociální determinaci. Tedy toho přístupu, který se projevuje na akcentaci technologické povahy mediální technologie.

Pro nás bude základním východiskem přístup amerického teoretika ruského původu Lva Manoviche. Ten ve své knize *The Language of New Media* z roku 2001 charakterizuje nová média pomocí pěti základních atributů.²⁷

²² WARDRIP-FRUIIN, Noah a Nick MONTFORT. *The NewMediaReader*. Cambridge, Mass.: MIT Press, c2003. ISBN 978-026-2232-272.

²³ viz AARSETH, Espen J. *Cybertext: perspectives on ergodic literature*. Baltimore, Md.: Johns Hopkins University Press, c1997. ISBN 978-080-1855-795.

²⁴ GITELMAN, Lisa. *Always already new: media, history and the data of culture*. Cambridge, Mass.: MIT Press, c2006. ISBN 978-026-2072-717.

²⁵ MACEK, Jakub. *Poznámky ke studiím nových médií*. Brno: Masarykova univerzita, 2013. ISBN 978-802-1064-768.

²⁶ DAHLBERG, Lincoln. *Internet Research Tracings: Towards Non-Reductionist Methodology*. *Journal of Computer-Mediated Communication*. 2004, 9(3), 00-00. DOI: 10.1111/j.1083-6101.2004.tb00289.x. ISSN 10836101. Dostupné také z: <https://academic.oup.com/jcmc/article/4614480>

²⁷ MANOVICH, Lev. *Jazyk nových médií*. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. *Studia nových médií*. ISBN 978-802-4629-612.

1. Princip číselné reprezentace, což znamená, že každý „[n]ovomediální objekt může být vymezen formálně, matematicky. Například obraz nebo tvar lze popsat matematickou funkcí“, z čehož vyplývá, že „[n]ovomediální objekt je předmětem algoritmické manipulace. Uplatněním vhodného algoritmu můžeme například automaticky odstranit zrnitost z fotografie, vylepšit její kontrast, rozpoznat tvary nebo změnit proporce. Řečeno ve zkratce, média se stávají programovatelnými“.²⁸ Jaký je přímý dopad? Kupříkladu ten, že mobilní aplikace, které nabízejí různé úpravy fotografií pomocí filtrů, jako jsou Hipstamatic nebo Instagram, v podstatě transformují estetiku starých fotoaparátů do matematických vzorců.

2. Princip modularity, kterou Manovich popisuje takto: „Jednotlivé prvky médií, obrazy, zvuky, tvary i jednání jsou reprezentovány jako soubory diskrétních vzorků, ať již jde o pixely, mnohoúhelníky, voxely, znaky, skripty. Na vyšší úrovni jsou tyto jednotky skládány do objektů, ale ponechávají si svojí oddělenou identitu.“²⁹ Právě modularita, daná předchozí možností číselné reprezentace, umožňuje existenci webových stránek, které se skládají z oddělených vrstev. Stejně tak je modularita základem pro další vlastnosti novomediálního díla. Kupříkladu libovolný text je po své digitalizaci připraven na různé transformace.

3. Princip automatizace jako další z atributů novomediálního díla vychází z číselného kódování a modulární struktury, které „umožňují automatizovat řadu operací při vytváření, manipulaci a přístupu k novým médiím“. Lidský záměr proto může být z tvůrčího procesu alespoň částečně odstraněn.³⁰ Dobrým příkladem může být kupříkladu služba IFTTT³¹ tedy If This Then That. V ní si uživatel může naklikat automatizovaný proces provozující několik služeb. Kupříkladu: kdykoli se objeví na Twitteru tweet od uživatele @realDonaldTrump, tak ho uložit do Google Documents.

4. Princip variability novomediálních artefaktů vychází z předchozích bodů. Novomediální objekt lze totiž různě verzovat a modifikovat, či dokonce přizpůsobovat individualitě konzumenta. „Stará média zahrnovala tvůrčí činnost člověka, který osobně sestavoval prvky textů, obrazů nebo zvuků do určité kompozice nebo sekvence. Tím, že byly uloženy do

²⁸ MANOVICH, Lev. Jazyk nových médií. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. Studia nových médií. ISBN 978-802-4629-612, str. 66

²⁹ MANOVICH, Lev. Jazyk nových médií. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. Studia nových médií. ISBN 978-802-4629-612, str. 68

³⁰ MANOVICH, Lev. Jazyk nových médií. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. Studia nových médií. ISBN 978-802-4629-612, str. 71

³¹ If This Then That [online]. [cit. 2019-02-23]. Dostupné z: <https://ifttt.com>

materiálu, je jejich souslednost pevně daná. Může být vytvořeno mnoho kopií původního originálu, které budou v souladu s logikou industriální společnosti zcela identické. Nová média jsou naopak charakteristická svou variabilitou. (Jiná označení tohoto principu nových médií jsou proměnlivost, nebo tekutost.) Namísto identických kopií novomediální objekt dává vzniknout mnoha různým verzím. Spíše než by je vytvářel lidský autor, jsou automaticky sestavovány počítačem.³²

5. Princip překódování je přirozeným důsledkem interakce kultury a digitálního světa. Platí zde totiž, že „Logika počítačů se vepisuje hluboko do kulturní úrovně médií již z toho důvodu, že nová média jsou vytvářena, rozšiřována, ukládána i archivována díky počítači. Způsoby, kterými počítače formují náš svět, reprezentují a zpřístupňují data, klíčové operace ovládající počítačové programy (vyhledávání, třídění, filtrování), rozhraní člověk-počítač (HCI), zkrátka vše, co můžeme označit za ontologii, epistemologii a pragmatiku počítače, to vše ovlivňuje kulturní úroveň nových médií, jejich organizaci, nové žánry, ale také obsah.“³³

Tyto základní principy stojí v samotném jádru novomediální změny, ať již se jedná o internet, mobilní telefon nebo digitální fotoaparát. Pokud si je osvojíme, umožňují nám pochopit nejen, v čem se liší od předchozích mediálních prostředí, ale také jaké nové možnosti lidského jednání otevírají.

Zároveň ale tyto změny jsou přesně tím, co umožňuje vznik digital humanities, protože dává vzniknout základnímu materiálu výzkumu, totiž digitálním objektům. Ty jsou dvojího druhu (jak je rozlišuje Richard Rogers³⁴): objekty přirozeně digitální a objekty digitalizované. Jejich rozdíl je ontologický: první vznikají již v prostředí samotného novomediálního prostoru, druhé jsou do něj transformovány: „An ontological distinction may be made between the natively digital and the digitized, that is, between the objects, content, devices, and environments that are ‚born‘ in the new medium and those that have ‚migrated‘ to it.“³⁵

³² MANOVICH, Lev. Jazyk nových médií. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. Studia nových médií. ISBN 978-802-4629-612, str. 75

³³ MANOVICH, Lev. Jazyk nových médií. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. Studia nových médií. ISBN 978-802-4629-612, str. 84

³⁴ ROGERS, Richard. Digital methods. Cambridge, Massachusetts: The MIT Press, [2013]. ISBN 978-026-2018-838.

³⁵ ROGERS, Richard. Digital methods. Cambridge, Massachusetts: The MIT Press, [2013]. ISBN 978-026-2018-838., str. 19

Uvedený ontologický rozdíl má své dopady v oblasti možnosti analýz novomediálních objektů, v každém případě ale novomediální artefakty existují jako digitální data. To přináší nejen otázky technické, spojené s jejich přenosem, ukládáním a dalšími aspekty, tak jak je řeší počítačová věda (computer science), ale i aspekty obecnější, a filozofičtější, které patří do oblasti informační vědy (information science).

Když se Lev Manovich snaží uchopit tuto zásadní změnu, hovoří o tom, že komputerizace dělá z média počítačová data. Tomu, jak ale vůbec v tomto kontextu rozumět označení data (a dalším klíčovým pojmům) se budeme věnovat v následující kapitole.

1.3. Data, informace, znalosti a vědění

Jak upozorňují Max Boisot a Agust Canals ve studii *Data, information and knowledge: have we got it right?*³⁶, Shannonova teorie informací, představená v textu *Mathematical Theory of Communication*³⁷ vůbec data nekonceptualizuje. Zaměřuje se na přenos informací a nezabývá se ostatními aspekty, jako je význam a podobně.

Faktickým cílem jeho zkoumání je měření efektivnosti přenosu a s ním spojené problémy. Klíčovým prvkem jsou tu možné varianty, které mohou přenášené symboly nabýt. Důležité ovšem je, že se „na tento repertoár stavu symbolů, které mohou být přenášeny, dívá jako na daný“.³⁸

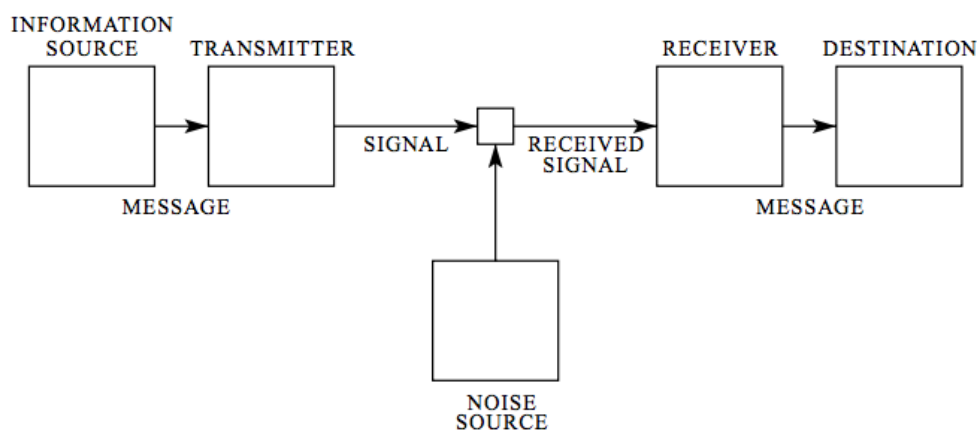


Fig. 1—Schematic diagram of a general communication system.

Obr. 3. – Základní Shannonovo komunikační schéma

Informace je tak zde definována v prostoru distribuce možných stavů,³⁹ který je předmětem zájmu v syntaktické rovině. Na tuto koncepci navazují i přístupy, které se snaží transponovat

³⁶ BOISOT, Max a Agust, CANALS. Data, information and knowledge: have we got it right?. Journal of Evolutionary Economics. 2004, 14(1), 43-67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9>

³⁷ SHANNON, C. E. A Mathematical Theory of Communication. Bell System Technical Journal. 1948, 27(3), 379-423. DOI: 10.1002/j.1538-7305.1948.tb01338.x. ISSN 00058580. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>

³⁸ BOISOT, Max a Agust, CANALS. Data, information and knowledge: have we got it right?. Journal of Evolutionary Economics. 2004, 14(1), 43-67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9>, str. 51

³⁹ FLORIDI, Luciano. The philosophy of information. New York: Oxford University Press, 2011. ISBN 978-019-9232-383.

tento přístup také do prostoru sémantického či pragmatického, jako je přístup Bar-Hillela a Carnapa⁴⁰ či Welse⁴¹.

Povaha a původ dat tak bývají řešeny spíše v širším kontextu informační vědy v rámci problému tzv. DIKW pyramidy neboli rozlišení mezi daty, informacemi, znalostmi a věděním.⁴² Tvůrce a propagátor této hierarchie Russell Ackoff ve svém textu *From Data to Wisdom*⁴³ začíná kritikou nepřiměřené pozornosti, která se věnuje technickým aspektům na úkor jiných: „Most of the time spent in school is devoted to the transmission of information and ways of obtaining it. Less time is devoted to the transmission of knowledge and ways of obtaining it (analytic thinking). Virtually no time is spent in transmitting understanding or ways of obtaining it (synthetic thinking). Furthermore, the distinction between data, information, and so on up to wisdom are seldom made in the educational process, leaving students unaware of their ignorance. They not only don't know, they don't know what they don't know.“

Ackoff pak navrhuje toto základní rozlišení: „Data are symbols that represent the properties of objects and events. Information consists of processed data, the processing directed at increasing its usefulness. For example, census takers collect data. The Bureau of the Census processes that data, converting it into information that is presented in the numerous tables published in the Statistical Abstracts. Like data, information also represents the properties of objects and events, but it does so more compactly and usefully than data. The difference between data and information is functional, not structural.“⁴⁴

Max Boisot a Agust Canals uvádějí pro rozdíl mezi funkcionálním a strukturálním pojetím následující příklad: „Effective cryptography protects information as it flows around the world. Encryption, by developing algorithms that bury information deeply in data, provides ‚the lock and keys‘ of the information age. Thus while the data itself can be made ‚public‘

⁴⁰ BAR-HILLEL, J. a R. CARNAP. Sémantická informace. In: DOLEŽAL, Lubomír. Teorie informace v jazykovědě. Praha: Nakladatelství Československé akademie věd, 1964, s. 165–175.

⁴¹ WELLS, R. Míra subjektivní informace. In: DOLEŽAL, Lubomír. Teorie informace a jazykověda. Praha: Nakladatelství Československé akademie věd, 1964, s. 187–195.

⁴² ROWLEY, Jennifer. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*. 2007, 33(2), 163-180. DOI: 10.1177/0165551506070706. ISSN 0165-5515. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/0165551506070706>

⁴³ ACKOFF, Russel. L. From data to wisdom. *Journal of Applied Systems Analysis*. 1989, (16), 3-9.

⁴⁴ ACKOFF, Russel. L. From data to wisdom. *Journal of Applied Systems Analysis*. 1989, (16), str.4

and hence freely available, only those in possession of the ‚key‘ are in a position to extract information from it. Cryptography, in effect, exploits the deep differences between data and information.“⁴⁵

Avšak samotná data definují trochu jinak než Ackoff, nikoli jen jako symboly reprezentující vlastnosti, ale „[d]ata can be treated as originating in discernible differences in physical states-of-the-world – that is, states describable in terms of space, time, and energy. Anything that can be discerned at all is discerned on the basis of such differences and is discerned by agents. Agents are bombarded by stimuli from the physical world, not all of which are discernable by them and hence not all of which register as data for them.“⁴⁶

Zavedením pojmu agenta (jednajícího), který v celkovém schématu je nejen rozdílem vystavován, ale také je produkuje, totiž z pouhé pyramidy vytvářejí dynamický systém, ve kterém zohledňují nejen postupnou transformaci dat ve vědění, ale také jeho zpětnovazebné působení do fyzického světa.

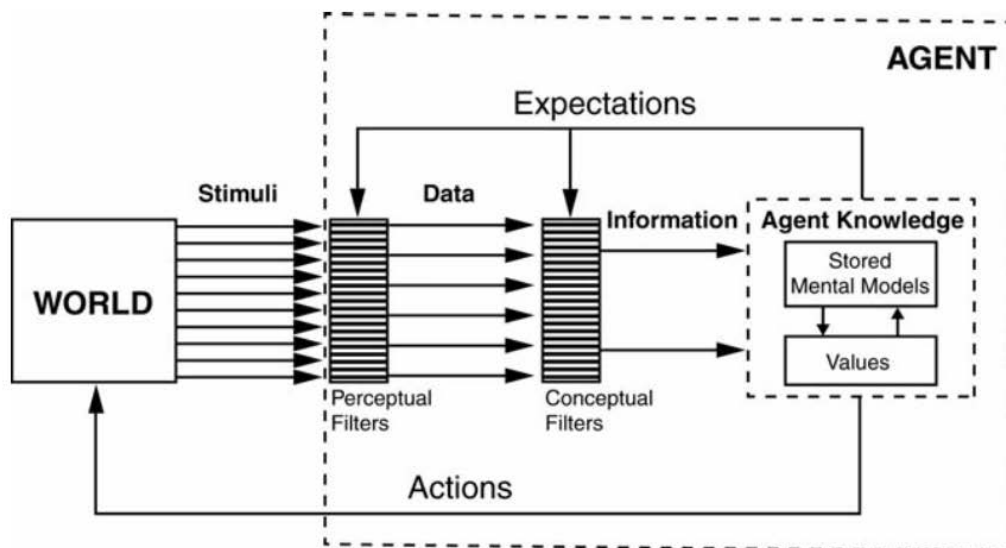


Fig. 1. The agent-in-the-world

Obr. 4. – Schéma interakce jednotlivých vrstev DIKW

⁴⁵ BOISOT, Max a Agust, CANALS. Data, information and knowledge: have we got it right?. Journal of Evolutionary Economics. 2004, 14(1), 43-67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9> str. 45

⁴⁶ BOISOT, Max a Agust, CANALS. Data, information and knowledge: have we got it right?. Journal of Evolutionary Economics. 2004, 14(1), 43-67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9> str. 46

Pro Maxe Boisota a Agusta Canalse jakožto ekonomy jsou jednající agenti především ekonomickým pojmem, pro naše účely ale můžeme hovořit o tom, že jakákoli data zachycují rozdíly, které vznikly působením různých sil a procesu jednajících aktérů, a to ať aktérů nadaných vědomím a intencionalitou, nebo aktérů neživých, kteří sice jednají, ale nelze jim připsat intencionalitu.

Pokud bychom chtěli tento pohled ještě trochu pootočit, můžeme využít peirceovskou klasifikaci znaků. Pro Peirce je znak cokoli, co pro někoho v nějakém aspektu něco jiného reprezentuje: „A sign, or representamen, is something which stands to somebody for something in some respect or capacity,“ říká Peirce doslova.⁴⁷ Autor této disertace se Peircově pojetí sémiozy věnoval ve své diplomové práci *Sémiotická interpretace Kantovy estetiky*⁴⁸, odkud volně přebíráme i následující charakteristiku základní sémiotické triády.

Peirce rozlišuje tři základní prvky procesu reprezentace. Znak (sign), jeho Objekt a jeho interpretant. Činností, která spojuje všechny tři dohromady, je tzv. Semiosis, která je potenciálně nekonečná, protože se do ní může zapojit celá triáda v nové pozici. Tak vzniká komplexní systém znaků. Nicméně nás zajímá jeho typologie ve vztahu mezi znakem a objektem, která je zároveň ta nejznámější. Jedná se o triádu ikon, index a symbol.

Ikon je znak, který se vztahuje k Objektu a denotuje ho jen díky svým vlastním rysům, které má bez ohledu na to, zda nějaký Objekt skutečně existuje, anebo ne. Tak je kresba kruhu ikonem slunečního kruhu či kupříkladu model parního stroje je ikonem skutečného parního stroje. V moderní terminologii mají modely vůči modelovanému ikonickou povahu.⁴⁹

Index je znak, který se vztahuje na Objekt a denotuje ho tím, že je jím skutečně ovlivněný. Příkladem indexu je klepání na dveře. Indexy ani ikony ovšem nic netvrdí, pouze ukazují. Tvrzení je až záležitostí symbolů.

Symbol je znak, který se vztahuje k Objektu a denotuje ho díky nějakému zákonu, většinou asociaci všeobecné ideje, a tento zákon způsobuje interpretaci. Jinými slovy znak tu nemá

⁴⁷ PEIRCE, Ch. S. *Collected Papers of Charles Sanders Peirce*. Vol. 2. Cambridge: Belknap, 1966., str. 228

⁴⁸ ŠLERKA, Josef. *Sémiotická interpretace Kantovy estetiky*.: Diplomová práce. 1. Praha, 2007.

⁴⁹ Viz OSOLSOBĚ, Ivo. *Ostenze, hra, jazyk: sémiotické studie*. Brno: Host, 2002. Teoretická knihovna. ISBN 80-729-4076-7.

vztah ani podobnosti a ani faktické souvislosti, ale jakési značky, pravidla, které nám umožňují spojení mezi jinak nespojitými věcmi.

Z tohoto pohledu jsou data symbolická (konvenční) vyjádřením indexů rozdílů a jako takové už obsahují nějaký pohled, z něhož se něco dává jako rozdílné a jako výsledek nějakého působení. Symboličnost a konvence je to, co umožňuje jejich uložení a opětovné přečtení v digitálním světě.

Řečeno ještě jinak: data nejsou jen symbolicky zapsané výsledky měření, ale symbolicky zapsané výsledky vůbec nějakých akcí.

Když se v tomto ohledu vrátíme zpět k vymezení nových médií a světu jeho artefaktů, které existují jako data a jsou jako data neustále počítači produkovány, bude třeba se ještě zastavit u pojmu tzv. big dat, která jsou přímým produktem této změny a dobře ukazují i změnu, ke které došlo v samotném chápání dat ve vědeckém výzkumu.

1.4. Big data

Původně se termín big data vztahuje na soubory dat, jejichž velikost je za schopností typických softwarových nástrojů je zachytit, ukládat, spravovat a analyzovat⁵⁰ a je spojen se zásadním nárůstem objemu dat uložených v počítačích po celém světě.

V oblasti vymezení velkých dat se nejběžněji setkáváme s přístupem, který mapuje jejich definici na klasickou koncepci 3V: volume, velocity a variety.⁵¹ Obvykle se uvádí, že je pro ně typický exponenciální růst jejich objemu (volume) a zvyšují se nároky na rychlost zpracování (velocity). Rovněž se objevují úlohy vyžadující okamžité zpracování velkého objemu průběžně vznikajících dat (z kamer či sociálních sítí).

⁵⁰ Big data: The next frontier for innovation, competition, and productivity. 1. New York, NY: McKinsey & Company, 2011. ISBN 9780983179696.

⁵¹ K bližšímu uvedení do problémů spojených s otázkou tzv. Big dat lze doporučit BERMAN, Jules J. Principles of big data: preparing, sharing, and analyzing complex information. Amsterdam: Elsevier, Morgan Kaufmann, [2013]. ISBN 978-0124045767. Dále pak IAFRATE, Fernando. From big data to smart data. Hoboken, NJ, 2015. Advances in information systems set, v. 1. ISBN 978-184-8217-553. a již zmíněný Big data: The next frontier for innovation, competition, and productivity. 1. New York, NY: McKinsey & Company, 2011. ISBN 9780983179696.

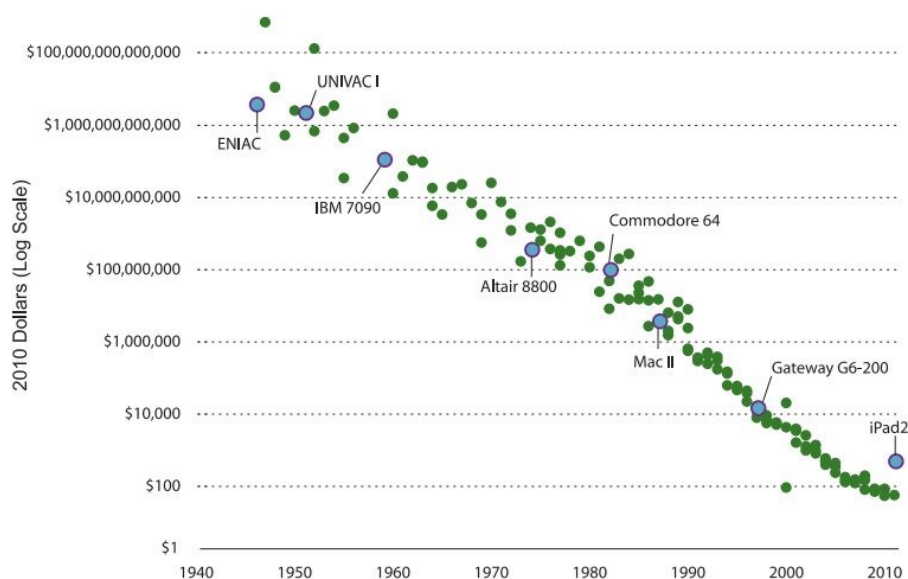
Oproti minulosti se také značně zvyšuje datová různorodost (variety), kdy se kromě potřeby zpracovat obvyklá strukturovaná data objevují požadavky na zpracování nestruturovaných dat jako například plain textů, ale i různých typů multimediálních dat.

Ke klasickému konceptu 3V se také přidává koncept věrohodnosti (veracity). Při omezení na tři klasické parametry totiž získáváme data se značně nejistou věrohodností v důsledku jejich nekonzistence, neúplnosti, nejasnosti a podobně.

Ve snaze lépe konceptualizovat, co vlastně velká data jsou, se také objevují přístupy, které se snaží pojem big data vymezit nikoli z pohledu technického, ale pragmatického. Tedy pohledu toho, k čemu jsou samotná uložená data určena. Jules Berman ve své knize *Principles of big data: preparing, sharing, and analyzing complex information* popisuje základní rozdíl pomocí pojmu small data.

Podle Bermana byla „small data“ designovaná a ukládaná pro odpověď na nějaké konkrétní otázky a základním rysem bylo jejich ukládání v agregacích. Pro Big Data však obvykle existuje jen rámcová představa o možnostech využití, důraz je proto kladen na co nejnížší granularitu dat při skladování.

Cost of Computing Power Equal to an iPad 2



Note: The iPad2 has computing power equal to 1600 million instructions per second (MIPS). Each data point represents the cost of 1600 MIPS of computing power based on the power and price of a specific computing device released that year.
Source: Moravec n.d..

Obr. 5. – Vývoj ceny výpočetní síly přepočtený na jeden iPad 2

Rovnoměrně s jejich nástupem se vyvíjejí technické možnosti zpracování. V posledních letech došlo k prudké demokratizaci v přístupu k velkým datům, který na jedné straně zapříčiňuje neustále klesající cena hardwaru (viz ilustrační obrázek z Floridiho knihy *Čtvrtá revoluce*). Zároveň dochází k prudkém rozvoji open source řešení pro zpracování těchto dat, například databází Hadoop či Elasticsearch a dalších. Souběžně s tím jsme svědky zpřístupnění samotných data miningových nástrojů a statistických programů (R, Rapidminer a další) nejširší veřejnosti.

Specifickým předmětem reflexe je pak dopad oblasti big data na samotné lidské poznání. Nejznámějším příspěvkem k této debatě je zřejmě provokativní esej *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* od Chrise Andersona, která vyšla v roce 2008 v časopise Wired.⁵² V něm Anderson shrnuje svoji pozici do jednoduchého konstatování: „Petabytes allow us to say: Correlation is enough. We can stop looking for models. We can analyze the data without hypotheses about what it might show.“ Jistěže lze

⁵² ANDERSON, Chris. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired [online]. 2008, 23.6.2008, 2008 [cit. 2019-02-23]. Dostupné z: <https://www.wired.com/2008/06/pb-theory/>

Andersonovu pozici smést velmi jednoduše pouhým konstatováním, že problematika vztahu mezi korelací a příčinností je složitější, jenže jádro Andersenova pohledu není ve vyzvednutí korelace. Je mnohem radikálnější: říká, že můžeme opustit vědu postavenou na modelech a nahradit ji vědou postavenou na statistice. Jinými slovy, že jsme svědky změny paradigmatu samotného pojetí vědy, která své oprávnění nezískává v modelování světa na základě hypotéz, ale je plně zapuštěna v explorační analýze obrovského množství dat.

Otázce toho, zda samotná big data přinášejí či nepřinášejí změnu paradigmatu, se věnuje celá řada dalších studií a knih. Asi nejznámější z nich je sborník *The Fourth Paradigm: Data-Intensive Scientific Discovery*,⁵³ z nějž Robin Kitch⁵⁴ vytvořil přehlednou tabulku:

Table 1. Four paradigms of science.

Paradigm	Nature	Form	When
First	Experimental science	Empiricism; describing natural phenomena	pre-Renaissance
Second	Theoretical science	Modelling and generalization	pre-computers
Third	Computational science	Simulation of complex phenomena	pre-Big Data
Fourth	Exploratory science	Data-intensive; statistical exploration and data mining	Now

Compiled from Hey et al. (2009).

Tab. 1 – Čtyři paradigmatu vědy

Svou esej končí Anderson provokativní otázkou: „The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways. **It's time to ask: What can science learn from Google?**“

Ačkoli je tato otázka myšlena řečnický, přesto na ni lze najít odpověď prostřednictvím Googlu. Je jí polemika mezi Peterem Norvigem a Noamem Chomským z roku 2011. Na začátku byla kritika známého lingvisty Noama Chomského v rámci jeho kritiky strojového

⁵³ HEY, Anthony J. G. *The fourth paradigm: data-intensive scientific discovery*. Redmond, Washington: Microsoft Research, [2009]. ISBN 978-098-2544-204.

⁵⁴ KITCHIN, Rob. *Big Data, new epistemologies and paradigm shifts*. 2014, 1(1). DOI: 10.1177/2053951714528481. ISSN 2053-9517. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/2053951714528481>

učení.⁵⁵ Ta byla založena na faktu, že se výzkumníci, pracující pouze statistickými metodami, snaží napodobit jazyk, aniž rozumějí tomu, jak funguje. Jde jim o nápodobu (a dodáváme: tím i predikci), nikoliv o porozumění. „It's true there's been a lot of work on trying to apply statistical models to various linguistic problems. I think there have been some successes, but a lot of failures. There is a notion of success ... which I think is novel in the history of science. It interprets success as approximating unanalyzed data,“ řekl Chomsky.

Na jeho kritiku obsáhle reagoval Peter Norvig, který v té době zastával pozici šéfa výzkumu ve společnosti Google a je nestorem strojového učení v americké vědě. Norvig obšírně odpovídá Chomskému a vyvrací jeho námitky⁵⁶. V poslední části své odpovědi ale Norvig přidává své polemice novou rovinu, když se odkáže na esej Leo Breimana *Statistical Modeling: The Two Cultures*,⁵⁷ v níž Breiman rozlišuje mezi dvěma přístupy ve vědě:

„First the **data modeling culture** (to which, Breiman estimates, 98 % of statisticians subscribe) holds that nature can be described as a black box that has a relatively simple underlying model which maps from input variables to output variables (with perhaps some random noise thrown in). It is the job of the statistician to wisely choose an underlying model that reflects the reality of nature, and then use statistical data to estimate the parameters of the model.

Second the **algorithmic modeling culture** (subscribed to by 2 % of statisticians and many researchers in biology, artificial intelligence, and other fields that deal with complex phenomena), which holds that nature's black box cannot necessarily be described by a simple model. Complex algorithmic approaches (such as support vector machines or boosted decision trees or deep belief networks) are used to estimate the function that maps from input to output variables, but we have no expectation that the form of the function that emerges from this complex algorithm reflects the true underlying nature,“ shrnuje Norvig Breimanovou pozici a dodává, že Chomského kritika se zakládá především na odmítání druhého modelu vědy.

⁵⁵ CASS, Stephen. Unthinking Machines. MIT Technology Review [online]. 4.5.2011 [cit. 2019-02-23]. Dostupné z: <https://www.technologyreview.com/s/423917/unthinking-machines/>

⁵⁶ NORVIG, Peter. On Chomsky and the Two Cultures of Statistical Learning. *Osobní stránka Petera Norviga* [online]. [cit. 2019-02-23]. Dostupné z: <http://norvig.com/chomsky.html>

⁵⁷ BREIMAN, Leo. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*. 2001, 16(3), 199-231. DOI: 10.1214/ss/1009213726. ISSN 0883-4237. Dostupné také z: <http://projecteuclid.org/euclid.ss/1009213726>

Cílem této kapitoly ani práce není rozsoudit spor ohledně dostatečnosti velkých dat pro vědecké poznání, ale upozornit na to, že spolu s nimi přichází změna v procesech vědeckého poznání a že tato změna postupně aspiruje na skutečnou změnu paradigmatu napříč vědou jako takovou.

V následující části této disertace se budeme věnovat specifickému dopadu této změny při výzkumu v oblasti analýzy lidského chování ve specifickém prostředí sociálních médií.

2. Výchozí výzkumný přístup pro analýzu jednání v sociálních sítích

2.1. Sociální média⁵⁸

Doposud jsme se v našich úvahách pohybovali v obecně novomediálním prostoru, za jehož nejlepší příklad nám slouží internet. Podle výzkumu Masarykovy univerzity⁵⁹ z roku 2015 jen 21,4 % procenta občanů České republiky nikdy nepoužilo internet. Neuživatelů počítače pak bylo 26,2 % a mobilní telefon nepoužívá jen 10,9 % obyvatel. Zdá se tedy, že komunikace a kultura zprostředkovaná pomocí nových médií je dnes i v naší současné společnosti natolik rozšířená, že představují specifické pole výzkumu pro všechny sociální a humanitní vědy.

Specifickou roli v počítačem zprostředkované komunikaci (jednání) jak dnes představují sociální média. Podle v úvodu citovaného výzkumu sociální média denně konzumuje 26,4 % obyvatelů České republiky. Ve věkové skupině 18–39 let to dokonce 64,3 % občanů, takže i u nás se jedná o nezanedbatelný fenomén.

Zdá se, že se situace v tomto ohledu prudce vyvíjí, i výzkumy komerčních firem totiž naznačují, že dochází k prudkému vývoji v této oblasti. Kupříkladu o tři roky mladší průzkum⁶⁰ AMI Digital Index agentury STEM/MARK provedený pro společnost AMI Digital už uvádí, že denně alespoň jednu sociální síť používalo 80 % českých uživatelů internetu. Podle výzkumu na internetu Češi strávili v roce 2018 skoro 150 minut denně?. Podle výzkumníků došlo meziročně k navýšení denních uživatelů o jedenáct procent. Oba výzkumy nelze metodologicky dobře porovnat. Pokud jde o demografii, mohli bychom zkusit přepočítat zmíněných 80 % z aktivních uživatelů na populaci obecně; vychází nám přibližně,

⁵⁸ Některé části této kapitoly a kapitoly Sociální média jsou obsaženy v ŠLERKA, Josef. Nová a sociální média: Základní vymezení pojmů. DINGIR [online]. 2015(3) [cit. 2019-02-23]. Dostupné z: http://www.dingir.cz/cislo/15/3/nova_a_socialni_media.pdf

⁵⁹ MACEK, Jakub a MACKOVÁ, Alena. *Stará a nová média v každodennosti českých publik (výzkumná zpráva)* [online]. 1. [cit. 2019-02-23]. ISBN 10.13140/RG.2.1.3589.9364. Dostupné z: https://www.researchgate.net/publication/280154724_Star_a_nova_media_v_kazdodennosti_ceskych_public_vyzkumna_zprava

⁶⁰ BREJČÁK, Peter. Infografika: Jak Češi využívají ty internety a sociální sítě v roce 2018. In: *Ty internety* [online]. 2018 [cit. 2019-02-23]. Dostupné z: <https://tyinternety.cz/socialni-site/infografika-jak-cesi-vyuzivaji-ty-internety-a-socialni-site-v-roce-2018/>

že na sociálních sítích bylo v roce 2018 denně nějak aktivních přibližně 60 procent obyvatelů České republiky. Jenže problematická zůstává hlavní otázka: jak definovat samotné sociální sítě.

Christian Fuchs⁶¹ upozorňuje na fakt, že nepanuje obecná shoda ohledně jejich přesného vymezení, když pojem sociální média zastřešuje jak blogy, tak tzv. social network sites (jako například Facebook), stejně jako mikroblogy (jako Twitter), různé wiki (kolektivně editované encyklopedie), ale také jiné stránky s uživatelsky generovaným obsahem nebo třeba stránky určené k sdílení obsahu.⁶²

Na množství různých teoretických pohledů upozorňuje i José van Dijck v *The Culture of Connectivity*⁶³ a navrhuje kriticky zkoumat sociální média jako techno-kulturní konstrukt a socioekonomickou strukturu⁶⁴ tvořenou šesti základními částmi: vlastnictvím, technologiemi, pravidly užívání (včetně přístupnosti a vlastnických licencí), uživateli, obchodním modelem a obsahem.⁶⁵

I přes různé názory na to, jak definovat sociální média jako specifický druh nových médií se všichni autoři shodují v tom, že se jedná o virtuální prostor, v kterém hraje zásadní roli možnost sociálních interakcí mezi uživateli a že tyto interakce mají specifický dopad na vytváření identity uživatelů a komunikačních situací.

Americká teoretička Danah M. Boyd se v textu *Social Network Sites: Definition, History, and Scholarship*⁶⁶ snaží sociální média, které označuje termínem social network sites, vymezit jako webové služby, které umožňují lidem:

⁶¹ FUCHS, Christian a Marisol SANDOVAL. *Critique, social media and the information society*. New York, 2014. ISBN 978-041-5841-856.

⁶² FUCHS, Christian a Marisol SANDOVAL. *Critique, social media and the information society*. New York, 2014. ISBN 978-041-5841-856., str. 4

⁶³ DIJCK, José van. *The culture of connectivity: a critical history of social media*. New York: Oxford University Press, c2013. ISBN 978-019-9970-780.

⁶⁴ DIJCK, José van. *The culture of connectivity: a critical history of social media*. New York: Oxford University Press, c2013. ISBN 978-019-9970-780., str. 28

⁶⁵ DIJCK, José van. *The culture of connectivity: a critical history of social media*. New York: Oxford University Press, c2013. ISBN 978-019-9970-780., str. 28

⁶⁶ BOYD, Danah a Nicole ELLISON. Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*. 2010, 38(3), 16-31. DOI: 10.1109/EMR.2010.5559139. ISSN 0360-8581. Dostupné také z: <http://ieeexplore.ieee.org/document/5559139/>

1. vytvářet veřejné nebo poloveřejné profily uvnitř ohraničeného systému
2. vytvářet seznamy uživatelů s nimiž jsem ve spojení prostřednictvím nějakých sociálních akcí
3. zobrazit a procházet seznamy těchto spojení a to nejen uživatelé samotnými, ale také ostatními uživateli sociální sítě.

Tyto základní charakteristiky podle ní odlišují sociální média od ostatních typů stránek, které umožňují běžné vytváření uživatelského obsahu.

Sociální média jako Facebook, Twitter či Instagram nabízejí celou řadu různých typů sociálních akcí. Na Facebooku je to kupříkladu přátelství, „to se mi líbí“, sdílení, komentář. Na Twitteru princip sledování či označení jako oblíbeného, a tak by se dalo pokračovat dále.

Všechna tato jednání můžeme chápat jako sociální jednání, jsou vybavena svou intencionalitou a mohou být podrobena tradičnímu druhu výzkumu, ať již kvalitativnímu nebo kvantitativnímu.

Oproti běžnému výzkumu ale sociální média přinášejí jednu novinku. Protože se jedná o média spadající do oblastí médií nových, jsou informace v nich archivovány v různých databázích. **Sociální akce jsou zveřejňovány, často archivovány a jsou v různých formách někdy k dispozici k dalšímu výzkumu.** Podle Lva Manoviche dnes můžeme prostřednictvím sociálních sítí a dalších zdrojů přistupovat k velmi rozsáhlým záznamům o lidském chování jak do hloubky, tak do šířky.⁶⁷ Nejsme již nuceni si vybrat mezi hloubkovým šetřením s malým počtem lidí, nebo dotazníkovým šetřením s velkým počtem respondentů.

Tak vznikají velká data ze sociálních sítí a zakládají dnes oblast výzkumu, o níž se také někdy hovoří jako o oblasti big social data. Přehled jednotlivých tendencí uvnitř výzkumu

⁶⁷ MANOVICH, Lev. Trending: The Promises and the Challenges of Big Social Data. Debates in the Digital Humanities. University of Minnesota Press, 2012, 2012-01-01, 460–475. DOI: 10.5749/minnesota/9780816677948.003.0047. ISBN 9780816677948. Dostupné také z: <http://minnesota.universitypressscholarship.com/view/10.5749/minnesota/9780816677948.001.0001/upso-9780816677948-chapter-47>

v této oblasti přináší studie *Conceptualizing Big Social Data*.⁶⁸ V ní navrhuje základní klasifikaci dat produkovaných uživateli v sociálních sítích: Digital self-representation data, Technology-mediated communication data a Digital relationships data, a klasifikaci formalizuje v následující tabulce:⁶⁹

Table 2 Classification of Big Social Data types

Category	Definition	Types of data
Digital self-representation data	Data related to <i>identity depiction</i> and <i>communicative body</i> in digital environment	<p><i>Profile data</i> (i) Login data (name/nickname/e-mail address and password); (ii) Mandatory data (services and application required data, for example, full name, citizenship, birthday); (iii) Extended data (profile pictures, education, tags of interests)</p> <p><i>Self-published content</i> (e.g., personal documents, pictures, videos, interests): (i) Disclosed data (to the public); (ii) Entrusted data (content sharing within trusted digital community)</p> <p><i>Data published by the community</i> (e.g., pictures, narrations, videos, posts): Relates to content shared by other users, which contribute to the digital identity creation</p>
Technology-mediated communication data	Data related to two-way communication, knowledge creation and distribution through technology	<p><i>Private communication data</i> instant 1-to-1 messaging and content sharing;</p> <p><i>Public communication data</i> 1-to-many messaging, commenting, information contribution and editing of existing entries;</p> <p><i>Collaborative communication data</i> many-to-many participatory content sharing, chats, video-conferences</p>
Digital relationships data	Data that reveal digital social relationships patterns	<p><i>Explicit data</i> Friendship data–Followee/Follower data;</p> <p><i>Implicit data</i> Data, which is revealed through technology-mediated communication data (e.g., tweets could be analyzed to infer connections between people)</p>

Tab. 2 – Klasifikace druhů big social data

Všimněme si, že takto klasifikovaná data již implicitně počítají s intencí uživatele sítě, která tvoří jejich základ. Data vyjadřující, vytvářející či ukazující identitu uživatele spadají do kategorie digital self-representation dat, data ukazující vztahy tvoří základ pro digital relationships data. Specifický typ dat představuje samotný obsah komunikace. (Otázkou

⁶⁸ OLSHANNIKOVA, Ekaterina, Thomas OLSSON, Jukka HUHTAMÄKI a Hannu KÄRKKÄINEN. *Conceptualizing Big Social Data*. *Journal of Big Data*. 2017, 4(1). DOI: 10.1186/s40537-017-0063-x. ISSN 2196-1115. Dostupné také z: <http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x>

⁶⁹ OLSHANNIKOVA, Ekaterina, Thomas OLSSON, Jukka HUHTAMÄKI a Hannu KÄRKKÄINEN. *Conceptualizing Big Social Data*. *Journal of Big Data*. 2017, 4(1). DOI: 10.1186/s40537-017-0063-x. ISSN 2196-1115. Dostupné také z: <http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x> str. 14

ovšem zůstává, zda navržená klasifikace je specifická právě pro velká data. Zdá se, že obecně platí pro data produkovaná sociálními sítěmi, bez ohledu na velikost samotných dat.)

Tento přístup je pro nás plně kompatibilní s pojetím, že data nejsou jen pasivním záznamem akcí, ale vždy výsledkem nějakého jednání a v případě jednání v sociálních sítích i výsledkem jednání s nějakým záměrem či vyjádřením určitého postoje. To se týká všech tří typů dat zmíněných výše. Nás budou dále zajímat data popisující uživatele a jejich vztahy, nikoli obsah komunikace. Pro zkoumání dat označovaných jako *Digital Relationships Data* bude naším východiskem tzv. social network analysis, pro data označovaná jako *Digital Self-Representation* data pak goffmanovský rámec analýzy označovaný jako dramaturgická sociologie.

2.2. Social network analysis

Vazby, které navazujeme v životě, nejsou nahodilé. Formuje je několik základních mechanismů sociální psychologie. Jako zásadní se ukazují především dva: princip homofilie a princip propinquity. Princip homofilie lze jednoduše popsat jako vysvětlení faktu, že za přátele máme především lidi, s nimiž sdílíme určité vlastnosti a hodnoty. Tento princip dominuje jak sítím ve skutečném životě⁷⁰, tak v sociálních sítích, jak ostatně potvrzuje řada výzkumů. Ať již jde o homofilii v oblasti věku či národnosti⁷¹, homofilii rasovou⁷² či další.⁷³

Vedle principu homofilie je druhým principem důležitým pro vznik vazeb v sociálních sítích tzv. princip propinquity. Jeho postulování je spojeno s výzkumem Festingera, Schachtera a Backa publikovaném v textu *The Spatial Ecology of Group Formation*.⁷⁴ Jádrem principu propinquity je teze, že lidé, kteří jsou si fyzicky blízko, mají častěji tendenci navazovat přátelství. Původní Festingerovy výzkumy se věnovaly studentskému spolubydlení, na ně pak navázala řada jiných.⁷⁵

Oba principy nelze od sebe oddělit a zkoumat je izolovaně, protože je to právě propinquity, co tvoří kontext, v němž se formují homofilní uspořádání. Vzájemný vztah mezi těmito principy v prostředí online sociálních sítí je stále předmětem výzkumů. K dispozici je kupříkladu výzkum role rasové homofilie a propinquity v prostředí sociální sítě Facebook

⁷⁰ MCPHERSON, Miller, Lynn SMITH-LOVIN a James M COOK. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*. 2001, **27**(1), 415–444. DOI: 10.1146/annurev.soc.27.1.415. ISSN 0360-0572. Dostupné také z: <http://www.annualreviews.org/doi/10.1146/annurev.soc.27.1.415>

⁷¹ UGANDER, Johan. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503, 2011.

⁷² WIMMER, Andreas a Kevin LEWIS. Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*. 2010, **116**(2), 583–642. DOI: 10.1086/653658. ISSN 0002-9602. Dostupné také z: <http://www.journals.uchicago.edu/doi/10.1086/653658>

⁷³ BAKSHY, E., S. MESSING a L. A. ADAMIC. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 2015, **348**(6239), 1130–1132. DOI: 10.1126/science.aaa1160. ISSN 0036-8075. Dostupné také z: <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa1160>

⁷⁴ FESTINGER, L., S. SCHACHTER a K. BACK. The Spatial Ecology of Group Formation. In: FESTINGER, L., S. SCHACHTER a K. BACK, ed. *Social Pressure in Informal Groups*, 1950. Oxford: Harper, 1950.

⁷⁵ MOUW, Ted a Barbara ENTWISLE. Residential Segregation and Interracial Friendship in Schools. *American Journal of Sociology*. 2006, **112**(2), 394–441. DOI: 10.1086/506415. ISSN 0002-9602. Dostupné také z: <http://www.journals.uchicago.edu/doi/10.1086/506415> nebo například PRECIADO, Paulina, Tom A.B.

SNIJDERS, William J. BURK a KERR. *Social Networks*. 2011. DOI: 10.1016/j.socnet.2011.01.002. ISSN 03788733. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0378873311000128>

publikovaný pod názvem *Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook*.⁷⁶

Právě to, že jsou vazby nenahodilé a že jsou výsledkem určitých principů, které odrážejí chování ve skupinách, zajišťuje možnost je systematicky zkoumat jak kvalitativně, tak kvantitativně. Náš zájem bude nyní primárně veden kvantitativním výzkumem sociálních sítí, který bývá zastřešován označením Social network analysis (dále jen SNA).

„Social network analysis (SNA) is not a formal theory in sociology but rather a strategy for investigating social structures. (...) Social network analysis, sometimes also referred to as ‚structural analysis‘, is not a formal theory, but rather a broad strategy for investigating social structures. The traditional individualistic social theory and data analysis considers individual actors making choices without taking the behaviour of others into consideration. This individualistic approach ignores the social context of the actor. One could say that properties of actors are the prime concern here. In SNA, however, the relationships between actors become the first priority, and individual properties are only secondary,“ interpretuje SNA Evelien Otte a Ronald Rousseau.⁷⁷

Pro naše potřeby budeme vycházet především z vymezení metodologie SNA podaného v monografii Stanley Wassermana a Katherine Faust *Social Network Analysis: Methods and Applications*⁷⁸ a ze souhrnné *Encyclopedia of Social Network Analysis and Mining*⁷⁹.

Zjednodušeně řečeno: sociální síť vzniká jako soubor aktérů (nikoli nutně lidských či jenom lidských), mezi kterými existují vztahy, které formalizujeme pomocí matice nebo grafu a dále analyzujeme. Soubor je množina prvků, které vytvářejí graf. Samotná množina může být zajímavá například pro měření průniků mezi různými množinami. Jednoduchý příklad

⁷⁶ WIMMER, Andreas a Kevin LEWIS. Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*. 2010, **116**(2), 583-642. DOI: 10.1086/653658. ISSN 0002-9602. Dostupné také z: <http://www.journals.uchicago.edu/doi/10.1086/653658>

⁷⁷ OTTE, Evelien a Ronald ROUSSEAU. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*. 2016, **28**(6), 441-453. DOI: 10.1177/016555150202800601. ISSN 0165-5515. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/016555150202800601>

⁷⁸ WASSERMAN, Stanley a Katherine FAUST. *Social network analysis: methods and applications*. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071.

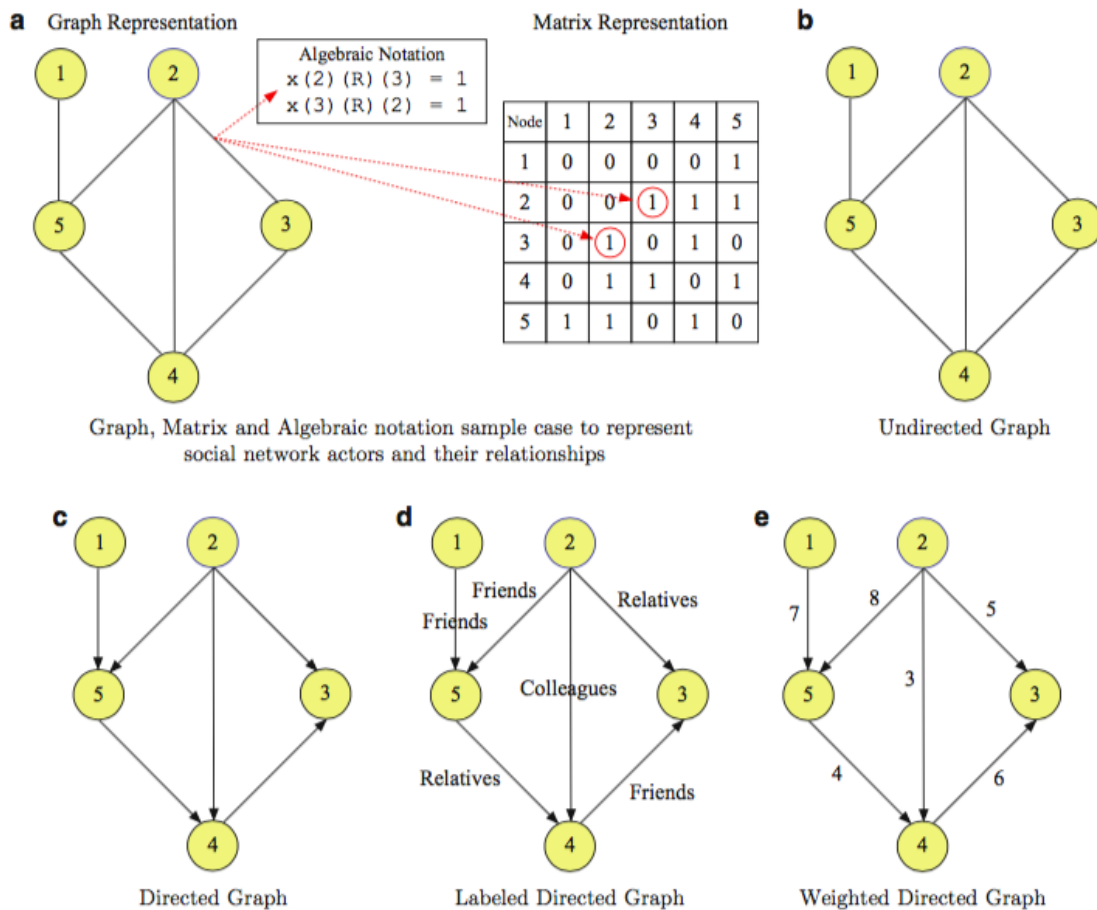
⁷⁹ ALHAJJ, Reda S. *Encyclopedia of social network analysis and mining*. New York: Springer, 2014. ISBN 978-146-1461-692.

sociální síť založené na vztazích může kupříkladu představovat punková píseň kapely *Zóna A*:

Igor má rád Evu, Eva jeho nie
a tak preto Igor stále smutný je.

Eva chce mít Vojta, Vojto Evu nie
a tak preto Eva stále smutná je.⁸⁰

Soubor je tvořen prvky Igor, Eva a Vojta. Vztahy mezi nimi se dají vyjádřit maticí nebo grafem. Jejich různé podoby popisující výrazně složitější případy než v písni *Zóny A* ilustruje následující obrázek.



Socio-Graph Representations, Concepts, Data, and Analysis, Fig. 1 A social network representation using a graph, its related matrix, and a sample algebraic notation (a), an undirected graph (b), a directed graph (c), a labeled graph (d), and a weighted graph (e) with $n = 5$ nodes and $m = 6$ links

⁸⁰ *Zóna A: Prečo Je Život Taký?* In: Karaoke Texty [online]. [cit. 2019-02-25]. Dostupné z: <https://www.karaoketexty.cz/texty-pisni/zona-a/preco-je-zivot-taky-38210>

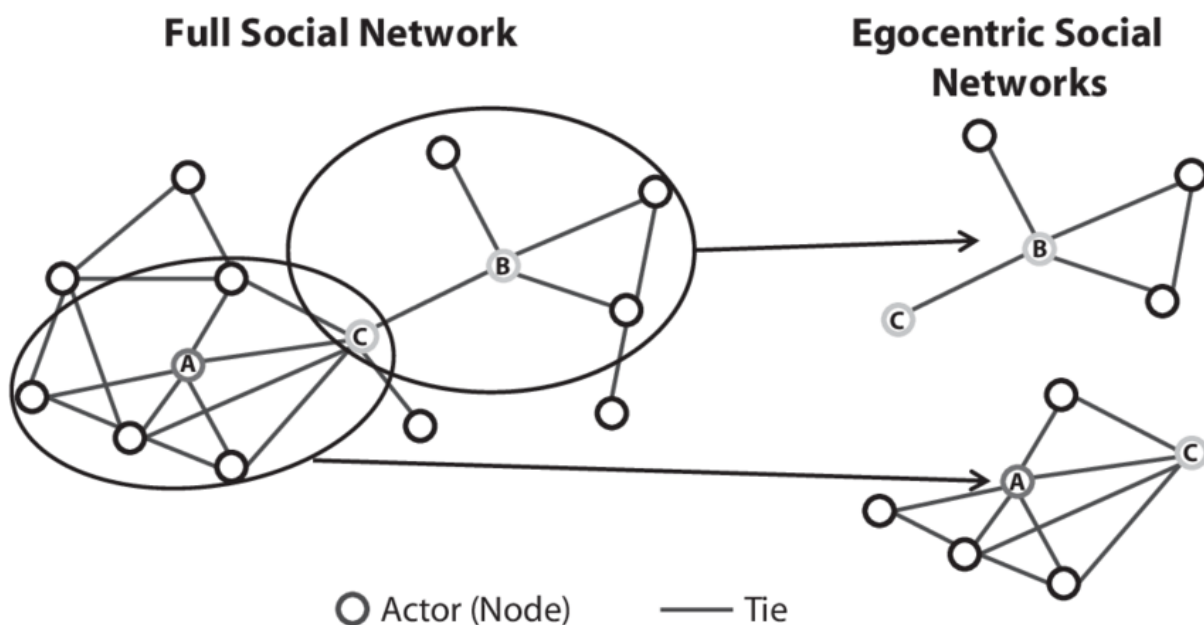
Mezi další příklady vztahů, které pokrývá zájem SNA, patří kupříkladu různé formy interakcí, jako jsou účasti na stejných událostech, členství v klubech, ale také různé typy vzájemného chování jako posílání zpráv, mluvení spolu, ale také autorství textů. Z nelidských akterů uvádí Faust kupříkladu řeky či cesty spojující dva body.⁸¹

Ze samotné definice SNA citované dle Boyd je na první pohled vidět, že v prostředí SNA můžeme najít celou řadu příkladů od jednodušších po komplikovanější vztahy. Na Twitteru je to například princip following. Na Facebooku nemusí vztah mezi lidmi zakládat pouze přátelství, ale také společné členství ve skupině či dávání „To se mi líbí“ stejným příspěvkům. Na sociální síti Flickr mohou být lidé propojeni tagy k fotkám a podobně. U sociální síti Foursquare/Swarm je to kupříkladu návštěva stejných míst.

Stavebními prvky social networks analysis jsou uzly, z nichž se síť skládají, a vazby mezi nimi. Pro oboje se používá celá řada termínů podle toho, k jaké oblasti vědění má autor blíže. Pro uzly se tak můžeme setkat s označením nodes, vertices, entities, items a podobně. Pro vazby zase ties, connections, relationships atd.

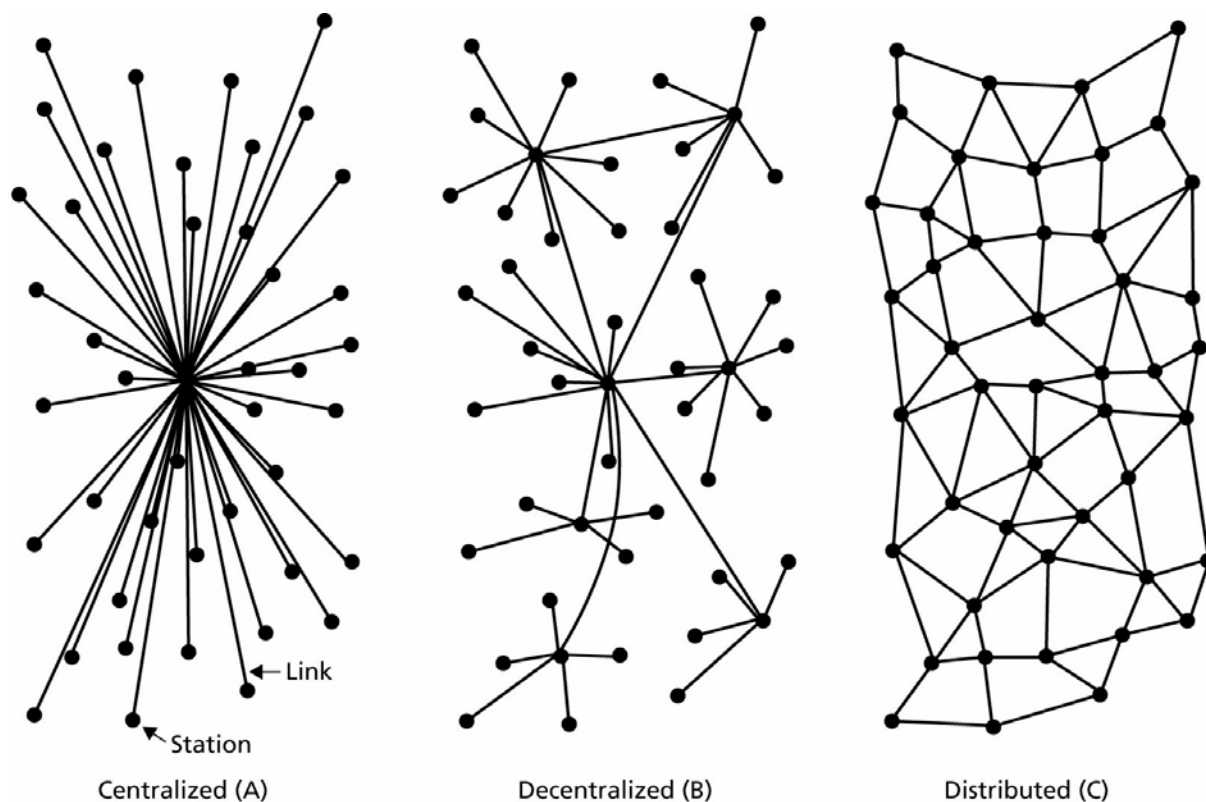
Vazby mohou mít z kvantitativního pohledu směr, případně váhu, či jich může být víc mezi stejnými body podle různých atributů. Z pohledu sociologie mohou mít i různé kvality – nejčastěji rozlišují sociologové vazby silné nebo slabé. Samotné grafy lze pak různě klasifikovat podle typů vazeb či podle míry a typu propojení.

⁸¹ WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071. str.18



Obr. 7. – Ukázka dekompozice grafu v sociální síti

V případě typů vazeb mluvíme o Unimodal networks (jedna přímá vazba), Multimodal networks (více přímých vazeb), Affiliation networks (vazba prostřednictvím např. akce), Multiplex Network networks (kvalitativně různé vazby). V případě míry propojení mluvíme o těchto základních typech grafů: Full (kdy je propojen každý s každým), Partial, Egocentric. Poslední klasická dělení se týkají tvaru propojení: Centralized, Decentralized, Distributed.



Obr. 8. – Příklad tří typů sítí: centralizované, decentralizované a distribuované

Při pohledu na SNA je zřejmé, že nabízejí řadu různých typů sítí. Jako uzly mohou vystupovat jednotliví uživatelé (ať již na Facebooku, Twitteru, Instagramu či dalších sítích), ale mohou to být i kupříkladu jednotlivé stránky či skupiny na Facebooku.

Mezi základní metody social network analysis patří měření postavení uzlů v síti a opakujících se vzorů, které vyjadřují postavení a pravděpodobnou roli uzlů v grafu, případně další informace o celkové podobě grafu.

Mezi základní metriky patří degree centrality, closeness centrality a betweenness centrality⁸², které měří postavení jednotlivých uzlů v rámci sítě a připisuje jim možné role v celé síti.

Základní metrikou je **degree centrality**. Ta vyjadřuje prostý počet vazeb daného uzlu, tedy počet přímých vazeb k dalším uzlům. V orientovaných grafech (directed graph) se případně měří i in-degree a out-degree centralita, tedy počet hran směřující k uzlu a z uzlu samotného. V praxi slouží jako pro základní měření aktivity uzlů v síti. Lidé s vysokou mírou konexí jsou

⁸² ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692., str. 1943

obvykle v sociálních sítích aktivnější, bývají starší (sociální síť se ve skutečném světě s věkem obvykle zvětšuje) apod. Uzly s vysokou hodnotou Degree Centrality jsou označovány jako „spojky“ nebo „střed“ v této síti.⁸³

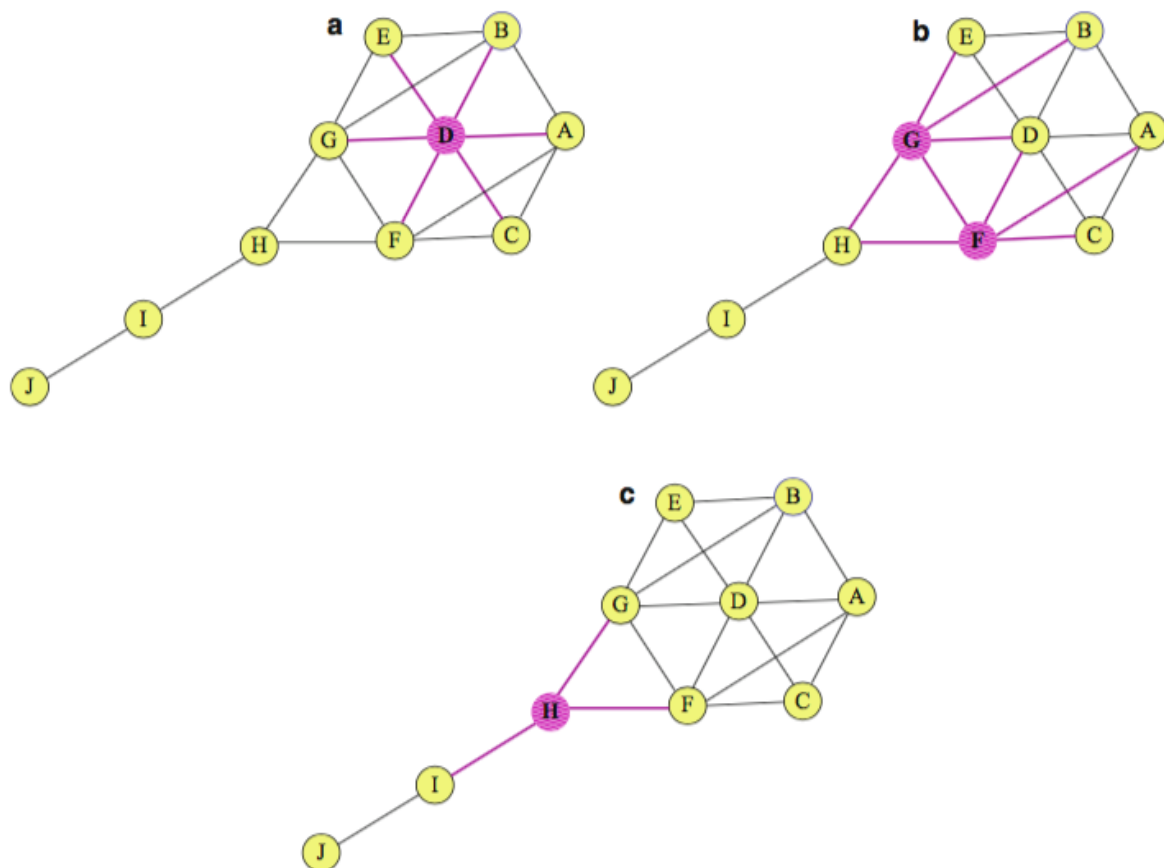
Druhou metrikou je **closeness centrality**. Ta vyjadřuje, jak blízko má uzel k jednotlivým dalším uzlům. Hodnota closeness centrality je nejvyšší, jestliže uzel má přímé spojení se všemi ostatními uzly sítě, takže všech ostatních uzlů je možné dosáhnout v jediném kroku. Vyjadřuje se jako nejmenší hodnota součtu vzdáleností k ostatním uzlům. Tyto uzly mohou snadno přijímat a přenášet inovace. Uzel s vysokou mírou blízkosti středu má velký vliv na to, co se v síti odehrává.⁸⁴

Poslední ze základních metrik je tzv. **betweenness centralita**. Metrika pro daný uzel je nejvyšší, pokud cesty mezi libovolnými dvěma uzly sítě vždy procházejí tímto uzlem. Betweenness centrality měří, kolik cest mezi dvojicí uzlů prochází daným uzlem. Někdy se o bodech s vysokou mírou betweenness centrality hovoří jako o závorách nebo mostech. Kontrolují totiž tok informací v síti a umožňují dobrou viditelnost všeho, co se děje v síti.⁸⁵

⁸³ ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692. str. 1943 nebo WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071. str. 178

⁸⁴ ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692. str. 1943 nebo WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071. str. 184

⁸⁵ ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692. str. 1943 nebo WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071. str. 189



Socio-Graph Representations, Concepts, Data, and Analysis, Fig. 2 A network shaped as a kite graph where each centrality measure yields a different central actor:

degree centrality (D), closeness centrality (F and G), and betweenness centrality (H). (a) Centrality degree. (b) Closeness degree. (c) Betweenness degree

Obr. 9. Ukázky bodů s vysokými hodnotami různých metrik

Kromě těchto základních metrik, které se orientují na roli uzlů, se pracuje i s řadou dalších, jako **eigenvector centrality**, která se zabývá mírou vlivu uzlu podle propojení s ostatními významnými uzly. Metriku **eigenvector centrality** lze počítat jen v neorientovaném grafu. Přidává k výsledku nejen počet vazeb uzlu, ale také počet vazeb uzlů, které jsou s ním spojené. V praxi tak odhaluje i nepřímý vliv. Metrika **Page Rank**, kterou používá Google pro stanovení hodnoty internetové stránky, je variantou **eigenvector centrality** pro orientované grafy. Dalším pak je **HITS** (Hyperlink-Induced Topic Search), který vyvinul Jon M. Kleinberg pro lepší hodnocení webových stránek věnovaných určitému tématu. Jeho výsledkem jsou hodnoty **authority** a **hub**. První vyjadřuje hodnotu obsahu stránky, druhá pak hodnotu linků, které ze stránky vedou. Poslední metrikou, **Reachability** je počet uzlů dosažitelných z jednoho uzlu na jeden, dva či tři kroky. Ukazují na potencialitu uzlů pro další šíření.

Tyto metriky sice nejsou jediné, pro výpočet postavení uzlů v síti jsou ale nejčastější. Jejich hodnota samozřejmě závisí i na samotných parametrech sítě. Za hlavní parametr je možné považovat samotnou hustotu sítě, tedy edge density, respektive její vyjádření pomocí **koeficientu shlukování**. Ten označuje, jak hustě jsou body v síti spolu propojeny. Vypočítá se jako poměr počtu hran v síti a jejich maximálního možného počtu. Čím vyšší koeficient, tím vyšší hustota sítě.⁸⁶

Ačkoli jsou jednotlivé metriky přísně formální, stávají se základem pro identifikaci sociálních rolí, které uzly v sítích sehrávají. Sociální role jsou z pohledu SNA definovány v zásadě formálně. Wasserman a Faust je definují jako „a role is identified as a position that has a distinct pattern of relations to other positions“.⁸⁷

Jako příklady těchto rolí můžeme zvolit role hubs nebo broker. Za huby označujeme osoby propojené s velkým množstvím uživatelů. Oproti tomu za brokery považujeme uživatele, kteří jsou jedinečnými propojovateli mezi komunitami. K těmto rolím, již z jejich definice, pak přináležejí vysoké hodnoty specifických metrik.⁸⁸

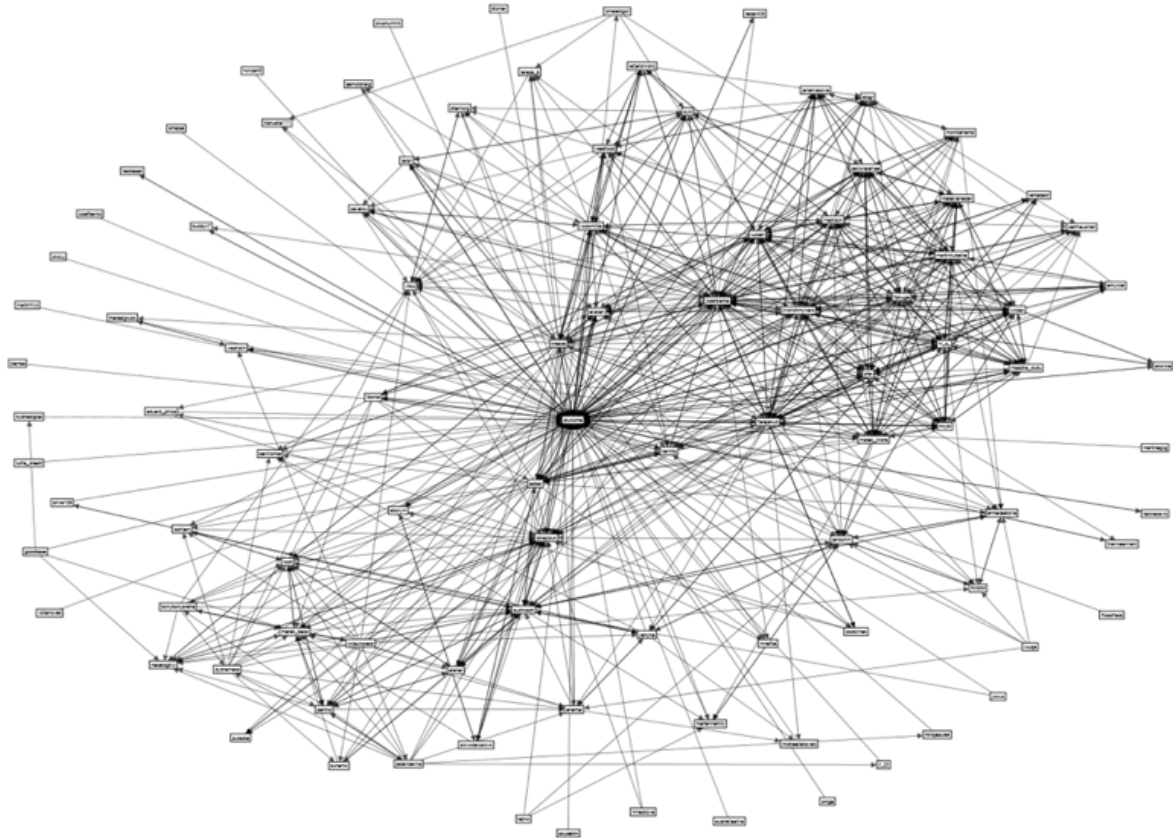
⁸⁶ ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692. str.512

⁸⁷ WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071, str. 175

⁸⁸ ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692., str. 786 a 1603

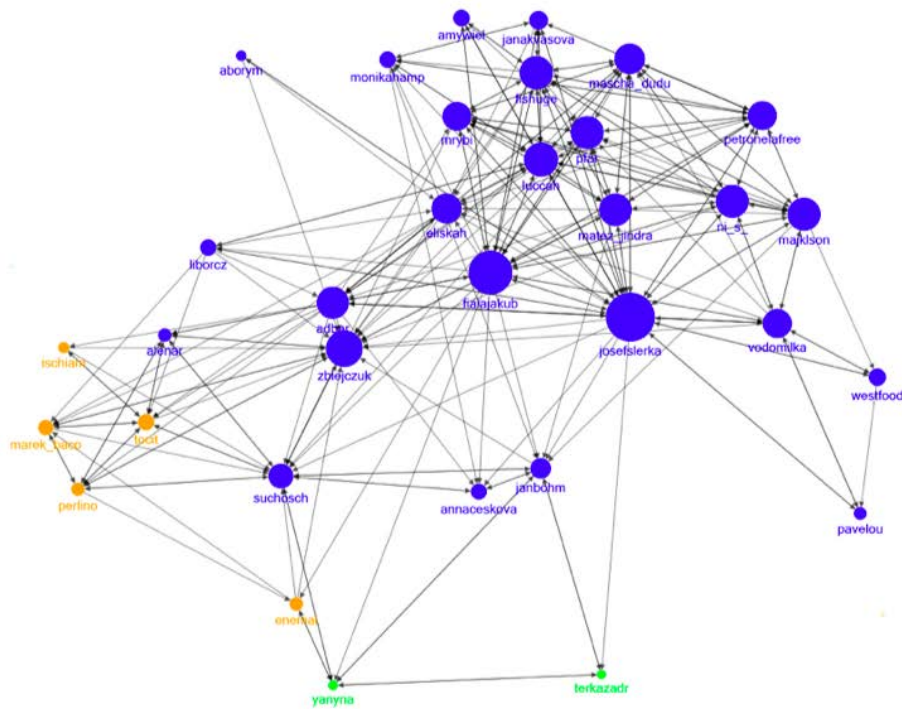
2.2.1. Případová studie I. – Účty sledující účet @stunome na Twitteru

V roce 2010 měl Twitter účet Studia nových médií několik desítek sledujících. Graf toho, jak se tito sledující sledovali navzájem, je možné vidět na následujícím obrázku:



Obr. 10. – Egocentrická síť účtu @stunome na Twitteru

Samotný graf je poměrně nepřehledný, pokud z něj ale odstraníme periferní uzly, největší centrální konektor (totiž účet @stunome), provedeme zklastrování účtu podle podobnosti a velikosti bodů a přidáme velikost uzlu, která odpovídá váze v roli brokera, získáme graf výrazně srozumitelnější.



Obr. 11. – Egocentrická síť účtu @stunome na Twitteru zbařená periferních bodů a zklastrovaná podle míry propojení

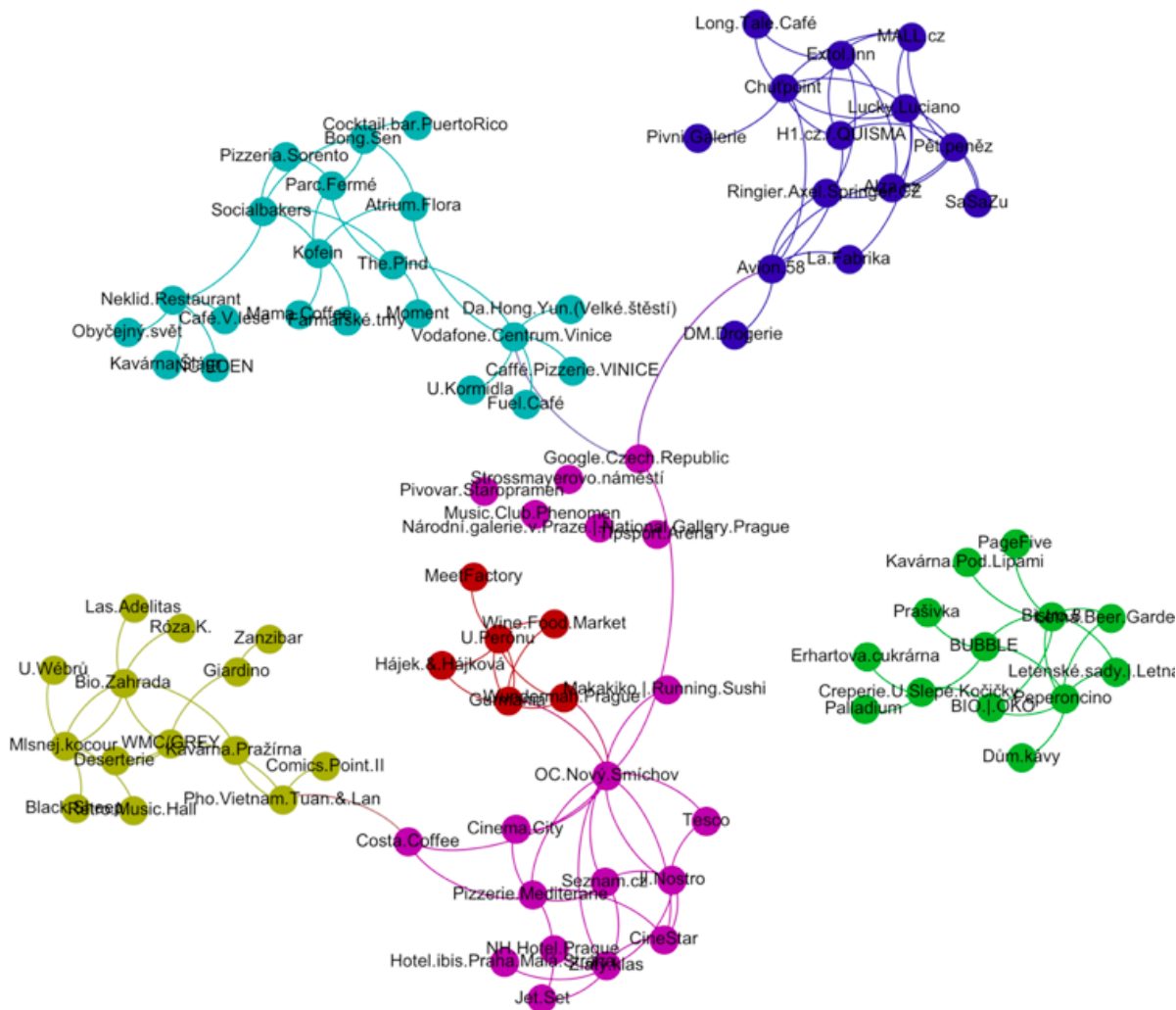
Vidíme v něm modrý klastr studentů, zelené budoucí studenty, kteří již chodili na akce SNM, a oranžový klastr lidí z internetového byznysu, kteří se zajímají o dění v oboru. Největší betweenness centrality má účet autora této práce a účet Adama Zbiejczuka, který patří mezi známé marketingové experty a který v té době studoval v PhD programu Informačních studií. Sociální postavení nás obou tak poměrně srozumitelně vysvětluje i naše postavení v síti. Na třetím místě můžeme najít účet tehdejšího studenta, dnes tajemníka oboru, Jakuba Fialy, který již tehdy patřil k těm nejvlivnějším účtům a dnes k nejvlivnějším lidem na oboru díky své funkci tajemníka SNM.

Všimněme si také, že podobně jako v případech dalších studií bylo nutné pro interpretaci sítě spojit jak znalosti kvantitativní, tak i znalost samotného prostředí. Bez ní by nebylo možné se posunout na pyramidě DIKW od dat a informací ke znalostem.

2.2.2. Případová studie II. – Analýza geospatiálních dat ze sociální sítě Foursquare

Analýza Twitter účtu @stunome patří ještě mezi analýzy klasické, takřka kanonické. Ovšem jinou perspektivu nabízí analýza dat ze sociální sítě Foursquare. Na síti Foursquare označují uživatelé místa, kde se nacházejí, tedy také restaurace nebo zaměstnání. Foursquare data od vás využívá mj. k službě “Kam chodí lidé po...”, která aktivně nabízí uživatelům možnosti pro další návštěvy. Data k jednotlivým místům jsou nabízena k automatickému zpracování, a tak je možné vytvořit následující graf: kam chodí lidé z vybraných internetových agentur po tom, co opustí kancelář (na oběd, po práci či z jiných důvodů). Pro stažení dat byla využita knihovna geosugar pro jazyk R, kterou vyvinul autor této disertace pro potřeby stahování a analýz geolokovaných dat⁸⁹.

⁸⁹ Knihovna Geosugar rozšiřuje možnosti knihovny RFacebooku [online]. [cit. 2019-02-23]. Dostupné z: <https://github.com/josefslerka/geosugar>



Obr. 12. – Síť míst na základě funkce „Kam chodí lidé po...“ na Foursquare

Pro graf byla zvolena maximální hloubka dva kroky, sleduje se tedy podnik po opuštění zaměstnání a podnik následující. Výsledný graf je pro přehlednost zpracován jednoduchým algoritmem pro shlukování. Graf samotný ukazuje nejen oblíbené podniky firem, ale také jednu zajímavost – holešovický podnik Avion 58. Ten propojuje Google a agenturu H1. Google samotný se dostal do grafu až jako další krok od ostatních agentur, avšak ukázal, kam chodí lidé z něj na obědy s lidmi z agentur a zřejmě i s lidmi z mediálního domu Ringier. Opět ale tento vhléd překračuje nad rámec samotných dat a je k němu nutná znalost prostředí. V tomto případě se jednalo o znalost agenturního provozu přístupnou autorovi práce díky tomu, že v agenturním prostředí pracoval.

2.3. Dramaturgická sociologie E. Goffmana a sociální média

Social network analysis je ve své základní podobě zaměřena primárně na formální modely fungování sociálních sítí. To, jak je motivováno a poháněno chování jednotlivců, tak není jejím primárním zájmem a pokud jej potřebuje výzkumník vysvětlit, musí sáhnout k jiným teoriím.

Jednu z možností, jak uchopit dynamiku chování uživatelů sociálních sítí (tzv. social network site, SNS) nabízí tzv. dramaturgická sociologie E. Goffmana, který ji zamýšlel jako určitý framework pro analýzu chování jednotlivců a jejich interakce na úrovni každodenní komunikace.

Klíčovou Goffmanovou metaforou jsou pojmenování převzatá nebo odvozená z prostředí divadla. Svou teorii formuloval především v knize *The Presentation of Self in Everyday Life*⁹⁰, kde vymezil několik základních pojmů, které slouží k uchopení lidské komunikace a interakce.

Pojmem **představení** označuje „souhrn všech činností, které jednatel provádí v době vyznačující se jeho trvalou přítomností ve společnosti konkrétního souboru pozorovatelů a která má pozorovatele nějaký vliv“.⁹¹ Představení se odehrává na **scéně**, která podle něj „zahrnuje nábytek, výzdobu, rozmístění objektů v prostoru apod., které vytvářejí kulisy a rekvizity pro množství lidských činností hraných před, uvnitř a na této scéně. Scéna se obvykle nehýbe, účinkující na ni vstupují.“⁹²

Základním rámcem představení a scény je pak **fasáda**, která určuje prvotní komunikační situaci. K osobní fasádě Goffman řadí především vzhled a vystupování. Podle něj můžeme do osobní fasády zahrnout odznaky úřadu, hodnosti, oblečení, pohlaví, věk a rasu, velikost a vzhled, držení těla, způsob mluvy apod. Některé tyto znaky jsou relativně neměnné jako třeba

⁹⁰ Dále citujeme dle českého překladu GOFFMAN, Erving. Všichni hrajeme divadlo: sebe prezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1.

⁹¹ GOFFMAN, Erving. Všichni hrajeme divadlo: sebe prezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1. str. 29

⁹² GOFFMAN, Erving. Všichni hrajeme divadlo: sebe prezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1. str. 29

rasa či pohlaví. Jiné se rychle mění. Fasády jsou dle Goffmana spíš vybírány, nikoli vytvářeny.

Goffman dále počítá s různými prostory komunikace, které jsou od sebe odděleny a ohraničeny bariérami vnímání. Ty nazývá **regiony**. Například ve větším obchodě můžeme vidět několik skupin – ohnisek komunikace mezi prodavačkou a zákazníkem či zákaznicí v rámci jedné scény.

Podle Goffmana jedinec kontroluje, jaké dojmy vyvolává. Goffman v tomto kontextu hovoří o **řízení dojmů** ve vztahu k představení, které hraje. Na základě této kontroly pak v obecné rovině rozděluje dva základní druhy tzv. regionů. **Přední region** je místo, které označuje, kde konkrétní představení probíhá, kde je scéna a podobně. Oproti němu stojí **zadní region** neboli zákulisí: místo, které se sice vztahuje k nějakému představení, ale dojem vyvolaný výkonem je vědomě popírán.

Příkladem takového chování je pro něj kupříkladu číšník v restauraci, který má přísně oddělenou část, kde obsluhuje hosty, a část v kuchyni, kde předává objednávku jídla. Zatímco v předním regionu obsluhy je v roli číšníka se specifickým chováním, držení těla a podobně, v zadním regionu se jeho chování dramaticky mění.

Pokud přijmeme Goffmanovu pozici za východisko porozumění chování uživatelů sociálních sítí a chápeme podobu konkrétního profilu uživatele jako fasádu, pak přijímáme i fakt, že volba jména, fotografie, míra nastavení soukromí, způsob vyplnění popisu a další jsou výrazem uživatelské identity,⁹³ respektive výraz toho, jak chce být uživatel přijímán. Identita sama ovšem vzniká z napětí mezi tím, jak se osoba vidí a jaké má postavení ve vztahu uvnitř skupiny a vztahů.⁹⁴

⁹³ Vztah mezi online identitou a sociálními akcemi v ní byl předmětem některých jiných výzkumů například WALLACE, Elaine, Isabel BUIL a Leslie DE CHERNATONY. Facebook 'friendship' and brand advocacy. *Journal of Brand Management*. 2012, 20(2), 128-146. DOI: 10.1057/bm.2012.45. ISSN 1350-231X. Dostupné také z: <http://link.springer.com/10.1057/bm.2012.45> či studie JENSEN SCHAU, Hope a Mary C. GILLY. We Are What We Post? Self-Presentation in Personal Web Space. *Journal of Consumer Research*. 2003, 30(3), 385–404. DOI: 10.1086/378616. ISSN 0093-5301. Dostupné také z: <https://academic.oup.com/jcr/article-lookup/doi/10.1086/378616>.

⁹⁴ JENSEN SCHAU, Hope a Mary C. GILLY. We Are What We Post? Self-Presentation in Personal Web Space. *Journal of Consumer Research*. 2003, 30(3), 385–404. DOI: 10.1086/378616. ISSN 0093-5301. Dostupné také z: <https://academic.oup.com/jcr/article-lookup/doi/10.1086/378616>

Výzkumů zabývajících se aplikací Goffmanova přístupu v oblasti nových médií a zejména médií sociálních je dnes celá řada. Jejich přehledné shrnutí přináší kupříkladu kniha *Facebooková (ne)závislost: Identita, interakce a uživatelská kariéra na Facebooku* od Marie Pospíšilové. Pro drtivou většinu z nich jsou ale základem různé podoby kvalitativního výzkumu. Pokud však platí naše teze, že sociální média přinášejí dostatek dat i ke kvantitativní analýze lidského jednání, pak bychom jimi měli být schopni podpořit platnost Goffmanova přístupu. Právě tomu se budou věnovat následující tři případové studie.

2.3.1. Případová studie I. - Jména uživatelů komentujících na facebookových stránkách

Prvním příkladem je volba uživatelského jména na sociální síti Facebook. Pravidla Facebooku zavazují uživatele k užití jména, pod kterým je uživatel znám ve svém každodenním životě, nepoužívání speciálních symbolů, přezdivek a podobně⁹⁵. Toto řešení je výsledkem kompromisu, který vyvolala změna pravidel Facebooku – pravidla původně vyžadovala jen a pouze skutečná jména.⁹⁶ Nicméně základním způsobem ověření identity uživatele při krádeži účtu pak bývá kopie občanského průkazu či jeho ekvivalentu.

Pokud uživatel pravidlo porušuje, dává tím jednoznačně najevo, že má nějaký problém s tímto pojetím. Důvodů může být celá řada od požadavků na silnou anonymitu (například z obav z politické represe) až po fakt, že jejich offline identita je spojena silně s nějakou přezdívkou jako v případě umělců.

Míra zastoupení těchto důvodů není dopředu známá, lze ale formulovat hypotézu, která předpokládá, že u některých typů stránek, obsahově spojených s velkou nedůvěrou ve stávající politický systém, bude podíl ne-skutečných jmen větší než u jiných. Důvodem je předpoklad, že uživatelé buď systému nedůvěřují zcela, nebo jsou si vědomi, že se v řadě svých názorů pohybují poblíž hranice tolerované právním řádem země.

Pro náš příklad jsme zvolili dataset 67 facebookových stránek: AC24, Aktuálně.cz, Alexander Ivanovič Možajev, Andrej Babiš, ANO, Antimajdan, Armáda České republiky (Czech Army), Armyweb, Bez politické korektnosti, Blesk.cz, Blok proti islámu, Blok proti islámu – Jižní čechy, Blok proti Islámu – Ústecko, Bohuslav Sobotka, Časopis dTest, Časopis Reflex, Časopis Respekt, Česká pirátská strana, Český kutil.cz, Český rozhlas 1 – Radiožurnál, ČSSD, ČT24, Deník Aha!, Deník Referendum, Domoobrana, Echo24.cz, EuroZprávy.CZ, Free Pub, G.cz, Generace Identity, Hnutí Úsvit, iDNES.cz, Imigranty v ČR nechceme, Ing. Radim Fiala, Islám v České republice nechceme, KDU-ČSL, Komunistická strana Čech a Moravy (KSČM), lidovky.cz, Literární noviny, Martin Konvička, Milan

⁹⁵ What names are allowed on Facebook? Facebook [online]. [cit. 2019-02-23]. Dostupné z: https://www.facebook.com/help/11214670538576?helpref=faq_content

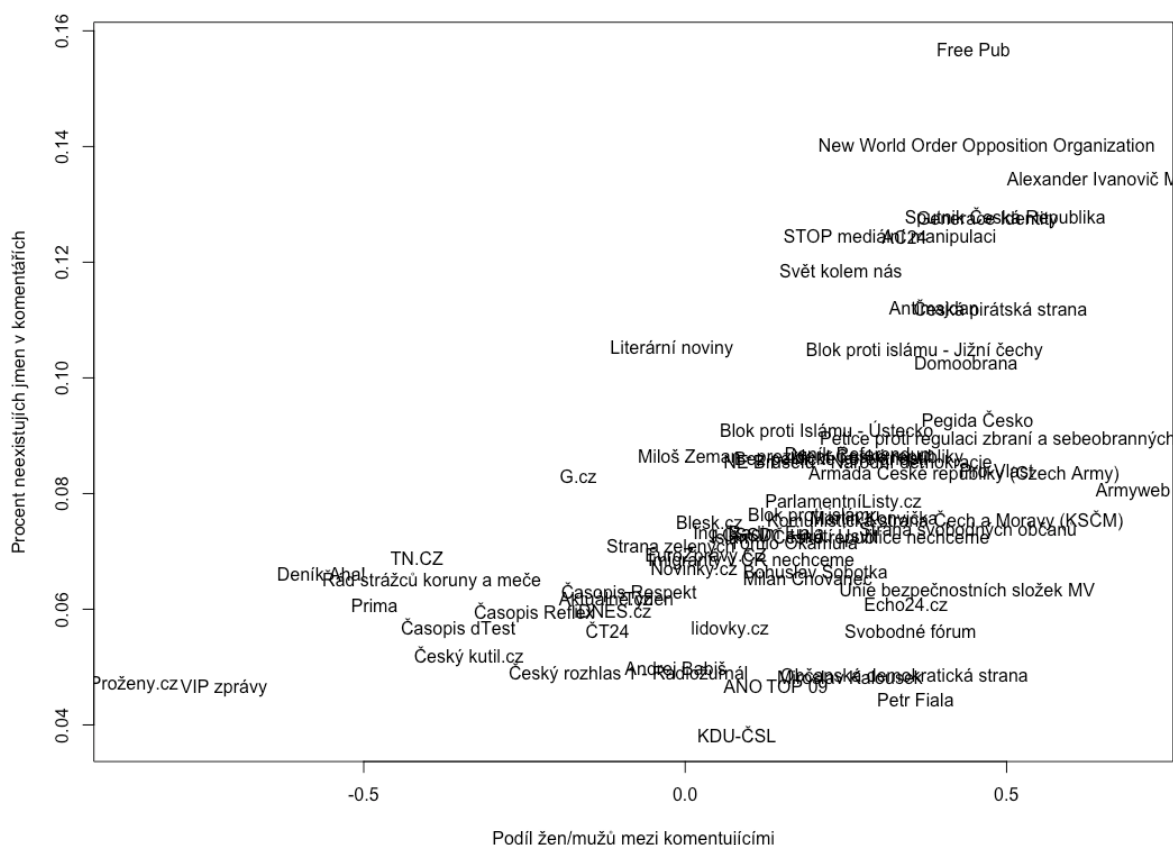
⁹⁶ Facebook real-name policy controversy. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Facebook_real-name_policy_controversy

Chovanec, Miloš Zeman – prezident České republiky, Miroslav Kalousek, NE Bruselu – Národní demokracie, New World Order Opposition Organization, Novinky.cz, Občanská demokratická strana, ParlamentníListy.cz, Pegida Česko, Petice proti regulaci zbraní a sebeobránných prostředků ze strany EU, Petr Fiala, Prima, Pro-Vlast, Proženy.cz, Řád strážců koruny a meče, Sputnik Česká Republika, STOP mediální manipulaci, Strana svobodných občanů, Strana zelených, Svět kolem nás, Svobodné fórum, TN.CZ, Tomio Okamura, TOP 09, Týden, Unie bezpečnostních složek MV, VIP zprávy. Tento seznam v sobě kombinuje stránky médií, politických stran a politiků a také různých alternativních skupin.

Ze všech těchto stránek jsme stáhli v červenci roku 2015 posledních 300 postů. Následně jsme k těmto postům stáhli všechny lajky na stránce, a získali tak facebooková jména lajkujících. Pro identifikaci toho, zda jméno existuje či neexistuje v českém prostředí, jsme zvolili seznam existujících jmen a příjmení v České republice ze stránek MV ČR⁹⁷. Pro samotnou identifikaci jsme vytvořili jednoduchý program, který identifikoval, zda facebookové jméno lze složit z jmen existujících v tomto seznamu. Abychom zohlednili fenomén, kdy si lidé vkládají mezi jméno a příjmení přezdívkou, považovali jsme za dostatečnou shodu dvou ze tří částí jména. Zároveň jsme seznam doplnili o časté zdrobněliny a specificky české formy (Josef vs Pepa apod.)

Na ose X vidíme podíl mužů/žen mezi jednotlivými lajkujícími, který nám ukazuje převahu ženského publika u čistého bulváru a specifických ženských serverů. Naopak úplně vpravo je vidět stránky kupříkladu věnované armádě.

⁹⁷ Četnost jmen a příjmení. Ministerstvo vnitra České republiky [online]. 2015 [cit. 2019-02-23]. Dostupné z: <http://www.mvcr.cz/clanek/cetnost-jmen-a-prijmeni-722752.aspx>



Obr. 13. – Graf podílu mužských a ženských jmen na stránkách (osa X) a podílu neexistujících jmen (osa Y)

Na ose Y vidíme pak podíl neexistujících jmen. Vidíme, že průměrně je 7,7 procent na stránku (medián byl 7,2 %). Outliery datasetu byly stránky s 12 % a patřily k nim následující: AC24, Alexander Ivanovič Možajev, Free Pub, Generace Identity, New World Order Opposition Organization, Sputnik Česká Republika, STOP mediální manipulaci a Svět kolem nás. Jedná se o stránky patřící do okruhu různých antisystémových webů, o nichž bude ještě řeč ve čtvrté části této práce. Zdá se tedy, že v souladu s naší hypotézou bylo uživatelů s neběžným jménem podstatně více na stránkách, které patří do skupiny antisystémových a extrémistických než na stránkách ostatních.

2.3.2. Případová studie II. – Fanouškovství stránek na Facebooku jako klíč k pochopení povahy fanoušků⁹⁸

Ve druhé případové studii budeme mluvit o akci, kterou Facebook označuje jako akci Stát se fanouškem. Dnes je k dispozici již celá řada výzkumů o motivaci uživatelů Facebooku stát se fanouškem nějaké stránky. Hlavním důvodem je (nepřekvapivě) zůstat v kontaktu s obsahem stránky,⁹⁹ přičemž již existují uspokojivé typologie fanoušků stránek podle vztahu k samotné značce.¹⁰⁰ Důležitou, byť ne absolutní roli hraje při stávání se fanouškem nějaké stránky potřeba sebevyjádření. Uživatel, který je fanouškem stránky, nejenže vyjadřuje svůj kladný postoj k obsahu postů stránky, ale také tímto krokem dává svůj postoj veřejně najevo.¹⁰¹ Lajkování stránek (ale i postů, jak ještě uvidíme) je tak svého druhu divadelním aktem (ve smyslu Goffmanovy dramaturgické sociologie). Výzkumy samozřejmě zohledňují i jiné motivace, jako kupříkladu fakt, že v postu stránky se objevuje oblíbená herečka uživatele.¹⁰² Všechny uvedené studie se věnují především facebookovým stránkám obchodních značek a typologie uživatelů je tak primárně zkoumána nad specifickým typem obsahu. Tomu je přizpůsobené i pojmenování jednotlivých skupin uživatelů. Kupříkladu Elaine Wallace hovoří o čtyřech typech: Fan-atics, Self-expressives, Utilitarians, Authentics. Distinktivní rysy v motivaci jednotlivých typů jsou dány především mírou loajality vůči brandu, mírou identifikace s brandem a potřebou být informován o dění kolem brandu. Jejich vlastnosti jsou přehledně shrnuty v následující tabulce z daného článku:

⁹⁸ Původní verze této případové studie vyšla na blogu autora této dizertace: ŠLERKA, Josef. Slušní lidé na Facebooku. Databoutique [online]. 1.7.2017 [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/161312186128/slu%C5%A1n%C3%AD-lid%C3%A9-na-facebooku>

⁹⁹ PELLETIER, M. a A. HORKY. The Anatomy of a Facebook Like: An Exploratory Study of Antecedents and Outcomes. *Annals of the Society for Marketing Advances*. 2013, (25).

¹⁰⁰ WALLACE, Elaine, Isabel BUIL, Leslie DE CHERNATONY a Michael HOGAN. Who “Likes” You ... and Why? A Typology of Facebook Fans. *Journal of Advertising Research*. 2014, 54(1), 92–109. DOI: 10.2501/JAR-54-1-092-109. ISSN 0021-8499. Dostupné také z:

<http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-54-1-092-109> nebo například BUNKER, Matthew P., K. N. RAJENDRAN a Steven B. CORBIN. The antecedents of satisfaction for Facebook “likers” and their effect on word-of-mouth. *Marketing Management Journal*. 2013, 23(2), 21–34.

¹⁰¹ WALLACE, Elaine, Isabel BUIL, Leslie DE CHERNATONY a Michael HOGAN. Who “Likes” You ... and Why? A Typology of Facebook Fans. *Journal of Advertising Research*. 2014, 54(1), 92–109. DOI: 10.2501/JAR-54-1-092-109. ISSN 0021-8499. Dostupné také z:

<http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-54-1-092-109>

¹⁰² LIPSMAN, Andrew, Graham MUDD, Mike RICH a Sean BRUICH. The Power of “Like”. *Journal of Advertising Research*. 2012, 52(1), 40–52. DOI: 10.2501/JAR-52-1-040-052. ISSN 0021-8499. Dostupné také z: <http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-52-1-040-052>

FAN-ATICS	SELF-EXPRESSIVES	UTILITARIANS	AUTHENTICS
<ul style="list-style-type: none"> • Brand is self-expressive • Highest brand loyalty • Highest brand love • Highest WOM • High number of Facebook friends • Highest self-monitoring • Highest materialism • Claims highest self-esteem • Opinion leader and opinion seeker • 'Likes' due to genuine interest and to create image • 'Like' for incentive although no the primary motivation for 'Liking' • Mainly females 	<ul style="list-style-type: none"> • Brand is self-expressive • Medium brand loyalty • Medium brand love • High WOM • Highest number of Facebook friends • High self-monitoring • Medium materialism • Medium self-esteem • Medium opinion leader and opinion seeker • 'Likes' mainly to create image • Mainly males 	<ul style="list-style-type: none"> • Brand is not self-expressive • Lowest brand loyalty • Lowest brand love • Lowest WOM • Average number of Facebook friends • Medium self-monitoring • Lowest materialism • Medium self-esteem • Least likely to be opinion leader and opinion seeker • 'Likes' only to gain incentives • Mainly males 	<ul style="list-style-type: none"> • Brand is not self-expressive • High brand loyalty • High brand love • Medium WOM • Lower number of Facebook friends • Lowest self-monitoring • Low materialism • High self-esteem • Low opinion leadership and medium opinion seeking • 'Likes' only because of genuine interest • Mainly females

Tab. 3 – Typologie fanoušků facebookových stránek

Zjednodušeně řečeno dominantní motivací pro to stát se fanouškem nějaké stránky je pozitivní vztah k samotné značce či obsahu postu. Existuje však skupina fanoušků, kteří jsou ve svých motivacích vedeni utilitární potřebou „zůstat v obraze“ či být prostě informován.

Pokud jde o výzkumy motivace udělení lajků pod samotné posty na Facebooku, nejsou zdaleka tak frekventované. Avšak výsledky studie *Facebook users' motivation for clicking the "Like" button*¹⁰³ naznačují značnou míru kompatibility s typologiemi uživatelů podle vztahu ke značce. Chin a Lu pracují s následujícími typy motivace pro lajk: hedonická, utilitární, udržovací, konformní a spřízněná. Jejich základní rysy shrnuli v následující tabulce:¹⁰⁴

¹⁰³ CHIN, Chih-Yu, Hsi-Peng LU a Chao-Ming WU. Facebook Users' Motivation for Clicking the "Like" Button. *Social Behavior and Personality: an international journal*. 2015, 43(4), 579–592. DOI: 10.2224/sbp.2015.43.4.579. ISSN 03012212.

¹⁰⁴ CHIN, Chih-Yu, Hsi-Peng LU a Chao-Ming WU. Facebook Users' Motivation for Clicking the "Like" Button. *Social Behavior and Personality: an international journal*. 2015, 43(4), 579–592. DOI: 10.2224/sbp.2015.43.4.579. ISSN 03012212.

Construct	Definitions	Operationalization
Hedonic motivation	Consumers expect to enjoy entertainment, amusement, and risks during consumption (Sit & Merrilees, 2005).	The content of the post is enjoyable to me. The content of the post seems fun. The content of the post interests me.
Utilitarian motivation	Consumers expect to achieve their expectations and get what they need, whether resources, information, or other things (Babin et al., 1994).	The content of the post is useful to me. The content of the post is helpful to me. The content of the post does not seem to benefit me.
Compliance motivation	An individual may change his/her own behaviors to meet the expectations, orders, or requests of others, so as to obey his/her reference groups (Etzioni, 1975).	I have always agreed with the poster of this post. I have always supported the poster of this post. I have always accepted the opinion of the poster of this post.
Conformity motivation	In interpersonal interactions, an individual may be influenced by others to change their behaviors, attitudes, or ideologies (Mowen & Minor, 1998).	I think this post has many "Like" clicks. I think very few clicked "Like" for this post. I think many people clicked "Like" for this post.
Affiliation motivation	Individuals long for positive affection from their interpersonal interactions (Hill, 1987).	The content of this post allows me to be informed about the poster's living status. The content of this post allows me to be informed about the poster's mood status. The content of this post allows me to be informed about the poster's recent status.

Tab. 4 – Typologie motivace pro lajkování postů

Výzkum tchajwanského týmu ukazuje, že všechny tyto motivace jsou přítomny v chování a podílejí se na chování uživatelů (svou roli hrají ještě subjektivní normy, ale ty nás v tuto chvíli nezajímají). Ze závěru studie pak vyplývá, že hedonická, utilitární a spřízněná motivace úzce souvisí s obsahem postu, zatímco udržovací především s tím, kdo postuje, a vztahem lajkujícího k němu. Poslední, konformní, je pak motivována celkovým vztahem k sociální skupině.

K tomu, aby uživatel mohl post lajkovat, jej musí neprve vidět. Nejčastějším místem, kde se s posty setká, aniž je musí sám aktivně vyhledávat, je tzv. news feed. Ten je Facebookem definován jako: „the constantly updating list of stories in the middle of your home page. News Feed includes status updates, photos, videos, links, app activity and likes from people, Pages and groups that you follow on Facebook.“¹⁰⁵

Samotný news feed ovšem neobsahuje všechny posty od přátel, stránek apod., nýbrž jejich výběr. Facebook o něm mluví takto: „The stories that show in your News Feed are influenced by your connections and activity on Facebook. This helps you to see more stories that interest you from friends you interact with the most. The number of comments and likes a post

¹⁰⁵ How News Feed Works. In: Facebook [online]. [cit. 2019-02-25]. Dostupné z: <https://www.facebook.com/help/www/1155510281178725>

receives and what kind of story it is (ex: photo, video, status update) can also make it more likely to appear in your News Feed.“¹⁰⁶

Postup, který Facebook užívá pro začlenění Stories (jak říká v tomto kontextu postům), není obecně znám; je předmětem obchodního tajemství Facebooku, a navíc se neustále vyvíjí. Obsah, který tak Facebook svým uživatelům nabízí, je sestaven co nejvíce na míru a přizpůsoben zájmům každého uživatele.

Facebookem zmiňovaný důraz na propojení mezi uživateli, které ovlivňuje, jaký obsah se v news feedu bude zobrazovat, zásadně akceleruje vliv jak principu homofilie, tak propinquity. To, že je obsah formován sítí přátel daného uživatele, potvrzuje i studie zaměstnanců Facebooku ve spolupráci s Michiganskou univerzitou publikovaná v časopise *Science* pod názvem *Exposure to ideologically diverse news and opinion on Facebook*¹⁰⁷. Ve studii se věnovali dopadu algoritmu tvořícího news feed na uživatele Facebooku. Výzkum potvrdil, že platí předpoklad názorové homofilie. Uživatelé, kteří se sami viděli jako liberálové, měli mezi svými přáteli signifikantně více liberálů, a naopak konzervativci více konzervativců. Výzkum také potvrdil, že liberálové byli více vystaveni liberálnímu obsahu a konzervativci konzervativnímu. Dále pak ukázal, že míra prokliku u postů, které přinášely opačný pohled na daný problém, byla nižší. Jinými slovy, hlavní roli podle výzkumníků nehrál algoritmus, ale skutečné sociální sítě lidí. Ti nejsou ochotni kliknout na obsah přinášející opačný pohled na problém.

Zdá se tedy, že interakce pod posty jsou výrazem zájmu uživatele, který vede k tomu, že se mu Facebook snaží nabízet obsah, u něhož lze předpokládat uživatelské zapojení. Lajky pod posty i lajky stránek samotných vyjadřují určitý postoj a zároveň jsou něčím motivované.

Pro ověření naší hypotézy jsme si vybrali stránku Slušní lidé, tedy stránku extrémistického brněnského uskupení, u jehož aktivních fanoušků předpokládáme právě afinitu k různým dalším extrémistickým uskupením a názorům. Stáhli jsme všech 154 postů publikovaných

¹⁰⁶ How News Feed Works. In: Facebook [online]. [cit. 2019-02-25]. Dostupné z: <https://www.facebook.com/help/www/327131014036297/>

¹⁰⁷ BAKSHY, E., S. MESSING a L. A. ADAMIC. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 2015, 348 (6239), 1130–1132. DOI: 10.1126/science.aaa1160. ISSN 0036-8075. Dostupné také z: <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa1160>

mezi 15. 1. 2017 až 15. 5. 2017. Ke každému postu jsme stáhli všechny ID uživatelů Facebooku, kteří měli s danými posty nějaký typ reakce (Like, Angry...). Celkem šlo o 22 874 reakcí od 8 344 uživatelů. Složení typu zapojení bylo následující:

	Var1	Freq	pct
3	LIKE	16210	0.708664860
1	ANGRY	4181	0.182783947
2	HAHA	1118	0.048876454
4	LOVE	899	0.039302265
5	SAD	277	0.012109819
6	WOW	189	0.008262656

Tab. 5 – Podíl druhů reakcí na facebookové stránce Slušní lidé

Z datasetu jsme vybrali pouze lajky a následně se zaměřili na uživatele, kteří lajkovali nejčastěji. Důvodem, proč jsme pro další analýzu zvolili jen ty, kdo dali lajk a nikoli ostatní typy reakcí, je relativní novost ostatních reakcí a absence relevantních studií o rozdílech, které přinášejí.

Z datasetu jsme tak vybrali 475 uživatelů, kteří dali 7 a více lajků pod posty v daném období. Celkem tvoří 41,5 % veškerých lajků na stránce.

Facebook neposkytuje přes API přístup k seznamu oblíbených stránek, a to ani u otevřených profilů. Použili jsme proto skript pro jednoduché stahování přímo z facebookových profilů. Pro stahování (tzv. scrapping) profilů na Facebooku jsme využil script, který napsal Jan Zikan, zdrojové kódy jsou dostupné na url: <https://github.com/janzikan/fb-data>. V našem případě jsme tak získali údaje k 311 profilům, které měly celkem v oblíbených 32 194 stránek.

Základní možností, kterou lze pro analýzu našeho datasetu využít, je spočítání míry afinity stránek oblíbených fanouškovským jádrem stránky. V zásadě se jedná o využití tzv. affinity indexu, který vyvinuli lidé v týmu Cahners Publications a Simmons Market Research Bureau pro potřeby vyhodnocování intenzity vztahu mezi médiem a cílovou skupinou reklamy. Jednoduchost výpočtu indexu přispívá k tomu, že je dnes běžným standardem v tzv. media planningu a svoje uplatnění nachází i jinde.

Server MediaGuru jej ve svém slovníčku definuje přehledně takto:

„Afinita (TAI, Target Affinity Index, vhodnost) je index, který popisuje vhodnost konkrétního reklamního nosiče (magazínu, TV pořadu) pro cílovou skupinu, kampaň. Charakterizuje, jak konkrétní cílová skupina sleduje dané médium ve srovnání s obecnou populací (obvykle dospělí 15+ nebo dospělí 12+). Počítá se jako sledovanost média v konkrétní cílové skupině děleno sledovaností v populaci. (...)

Čím vyšší afinita, tím je médium vhodnější pro oslovení konkrétní cílové skupiny. Obvykle afinita vyšší, než jedna znamená, že médium je vhodné pro použití v kampani (cílová skupina ho sleduje relativně více než populace).

Příklad:

Deník Sport čte v populaci 3,4 % lidí (mužů i žen) ve věku 12 a více let. Mezi muži ve věku 12 a více let najdeme 6,5 % mužů, kteří četli poslední vydání Sportu. Afinita deníku Sport v cílové skupině muži 12+ je rovna $6,5 \% / 3,4 \% = 1,9$. Muži čtou deník Sport relativně více než populace.“¹⁰⁸

U nás prvně aplikoval tuto metodu na facebookové stránky Jan Schmid, se kterým jsme společně publikovali několik postů k dané otázce na blogu Databoutique. Jeho východisko je následující:

- Víme, kolik procent fanoušků z našeho zkoumaného datasetu je zároveň fanouškem nějaké jiné stránky.
- Víme, kolik má tato stránka na Facebooku fanoušků.
- Můžeme spočítat kolik procent fanoušků domácí populace je fanouškem stránky, protože víme, kolik je přibližně aktivních uživatelů domácího Facebooku (pro nás to bude 4 000 000).

¹⁰⁸ Afinita. Mediaguru [online]. [cit. 2019-02-23]. Dostupné z: <https://www.mediaguru.cz/medialni-slovník/afinita/>

- Následně tak můžeme spočítat, kolikrát častěji se vyskytují fanoušci dané stránky uvnitř zkoumaného datasetu, než kdyby se jednalo o náhodně vybranou skupinu uživatelů Facebooku o stejné velikosti.

Ukažme si to na jednoduchém příkladu: V našem datasetu například víme, že 64 uživatelů je fanouškem stránky Hlasuji pro czexit. To je 20 % z celého datasetu. Přitom ale stránka Hlasuji pro czexit má na Facebooku 44 731 fanoušků, což znamená, že jejich fanouškem je 0,011 % uživatelů českého Facebooku. V praxi je tak v našem datasetu 18,5x více fanoušků a stránka Hlasuji pro czexit je tak silně afinitní vůči stránce Slušní lidé. Oproti tomu stránka Nametests.com český má v naší analyzované skupině uživatelů 16 % příznivců, v celém českého Facebooku ji pak má lajknutou 15 % všech uživatelů. Afinita se tu blíží 1, a tím pádem stránka není afinitní.

Kromě samotné afinity ale hraje roli i celkové procentuální zastoupení v datasetu.

V následující tabulce můžeme vidět stránky, které jsou nejvíce zastoupeny v našem zkoumaném datasetu, a zároveň mají afinitu větší než 10.

Name	About	FansOnFacebook	PctInFBPopulation	FansInDataset	PctInDataset	Affinita
Slušní lidé		7943	0.1986	278	89.3891	450.153
Imigranty v ČR nechceme	Účelem této stránky je sjednotit všechny slušné lidi a vlastence, kteří si uvědomují katastrofické následky probíhající arabsko-africké invaze do Evropy.	102199	2.5550	128	41.1576	16.109
Tomio Okamura - SPD	Oficiální stránka Tomia Okamury a jeho příznivců https://www.instagram.com/tomio.cz/ www.tomio.cz https://twitter.com/tomio_cz https://www.youtube.com/channel/UCJGNjN97BAk85Xb_c7shOQ	257413	6.4353	125	40.1929	6.246
Václav Klaus ml.	Veřejný facebook; články, názory; fotky; lehce osobní věci; pozvánky na akce ...	65966	1.6491	120	38.5852	23.397
We Are Here At Home .com	JESTLI MILUJES SVOU ZEM , KLIKNI A PODPOŘ SEBE A SVŮJ NÁROD., ČESKÁ REPUBLIKA ZŮSTANE NAŠE A PO NAŠEM. A TO HEZKY ČESKY!	83357	2.0839	90	28.9389	13.887
Miloš Zeman - prezident České republiky	Oficiální stránka Miloše Zemana - prezidenta České republiky www.zemanmilos.cz , @MZemanOficialni	93344	2.3336	87	27.9743	11.988
Blok proti islamizaci	..	21533	0.5383	79	25.4019	47.187
Moje země, moje pravidla	Hanba vlasti je i naší hanbou, rána vlasti je i naší ranou, smrt vlasti je i naší smrtí a její život je i naším životem a její sláva je i naší slávou	68283	1.7071	78	25.0804	14.692
Organizace proti multikulturnímu fanatismu	Naše iniciativa bojuje proti multikulturním fanatikům a upozorňuje na jejich bizarní chování	22538	0.5634	68	21.8650	38.805
Miloš Zeman - prezident ČR	NEOFICIÁLNÍ STRÁNKA FANOUŠKŮ. Oficiální stránky jsou zde: www.zemanmilos.cz www.facebook.com/prezidentor www.hrad.cz	34802	0.8700	67	21.5434	24.761
Hlasuji pro czexit	Usilujeme o vyhlášení referenda o vystoupení ČR z EU.	44731	1.1183	64	20.5788	18.402

Tab. 6 – Silně afinitní facebookové stránky vůči stránce Slušní lidé

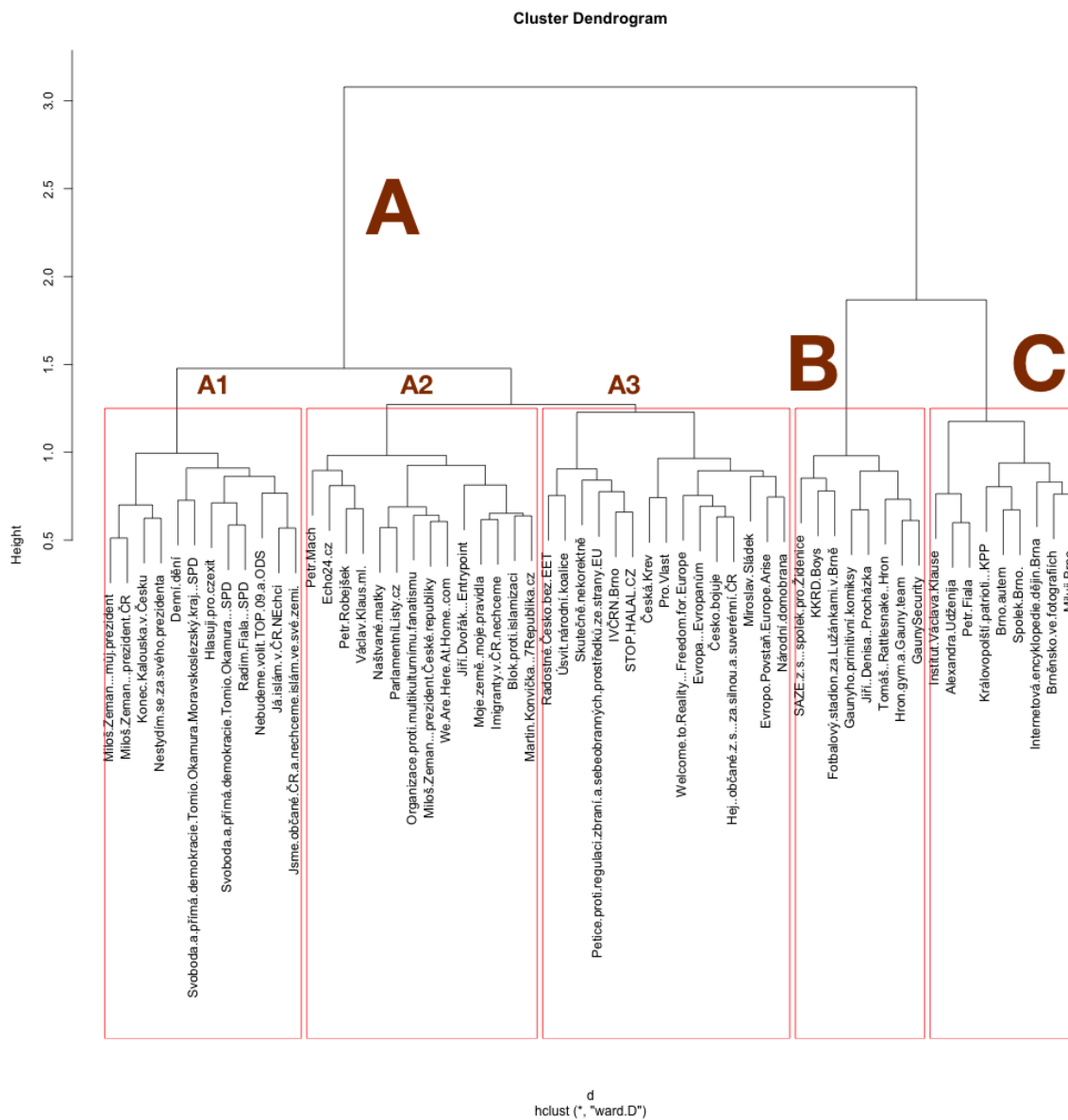
To, že samotná stránka Slušní lidé nemá 100 %, není nutně špatně, protože záleží na formě, kterou se posty dostávají k uživatelům. Můžete tak olajkovat post této stránky a zároveň nebyť jejím fanouškem. Facebook vám jej zobrazí například na základě interakcí vašich přátel. Podíváme-li se na seznam stránek, je vidět, že mezi nimi dominují stránky iniciativ a politiků spojených především s antiimigrační rétorikou. Což není nijak překvapivé vzhledem k rétorice stránky samotné.

Nevýhodou afinitního modelu je však to, že jeho výsledky jsou de facto jednorozměrné a neumožňují shluknout stránky podle jejich fanoušků, tedy říci, že některé stránky jsou navzájem více podobné než jiné na základě struktury oblíbených stránek mezi jejich fanoušky v datasetu.

Jako základ pro výpočet jsme vytvořili matici, v níž řádky tvořily facebookové stránky, které v našem datasetu měly alespoň 10 % společného průniku a afinitu rovnou nebo větší než 10. Sloupce byly ID fanoušků z našeho datasetu. V matici byla zapsána 1, pokud uživatel byl fanouškem dané stránky, nebo 0, pokud fanouškem nebyl.

Předmětem této práce však není úvod do shlukovacích postupů a případné zájemce raději odkážeme na internetové zdroje. Pro naše účely jsme zvolili metodu hierarchického shlukování¹⁰⁹, která se snaží spočítat míru podobnosti jednotlivých znaků (v našem případě facebookových stránek) a vizualizovat je pomocí tzv. dendrogramu. Výsledek naší jednoduché analýzy vidíte na výsledném obrázku:

¹⁰⁹ Hierarchical clustering. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Hierarchical_clustering



Obr. 14. – Dendrogram výsledků hierarchického klustrování pro stránku Slušní lidé

Základní čtení dendrogramů je relativně jednoduché, v našem případě vidíme na ose y vnesený koeficient podobnosti a na ose x jednotlivé objekty. Výška vertikální úsečky tak reprezentuje míru nepodobnosti. Čím je větší, tím jsou si facebookové stránky méně podobné (stále uvažujeme pohled preferencí jejich fanoušků v našem datasetu). Naopak pořadí objektů na ose x nese žádný jiný význam než vyjádření jejich rodové blízkosti. Mohlo by tak být i zrcadlově převrácené.

Vidíme, že se stránky poměrně rychle rozdělují do dvou, respektive tří větví. První větev, kterou jsem označil jako A, reprezentuje převážně stránky politických stran, médií a hnutí, které se vyznačují vysokou mírou nespokojenosti se stavem společnosti. Vidíme také, jak se

tato větev štěpí na další větve. Větev A1 reprezentuje především hnutí spojené s osobou Tomia Okamury a Miloše Zemana. Ve větvi A2 vidíme spíše pravicové subjekty ve spojení se silně protiimigračními médii a okruh stránek spojených s tzv. konvičkovci. Zde také najdeme politiky, jako jsou Petr Robejšek či Václav Klaus ml. Ve větvi A3 nacházíme hlavně stránky z okraje názorového spektra, převážně nacionalistické.

Větev B je silně spojená s částí brněnské scény. Je tu jak parodický KKRD Boys server, tak stránky spojené s Gaunym¹¹⁰ – jak s institucí GaunySecurity, s Gauneho primitivními komiksy a také se stránkami MMA bojovníků Tomáše „Rattlesnaka“ Hrona a Jiřího „Denise“ Procházky.

Poslední větev C by bylo možné rozdělit na dvě separované skupiny. ODS a její politiky (minulé či přítomné) a stránky vyjadřující silnou příslušnost k Brnu, což u skupiny působící v Brně nepřekvapí.

¹¹⁰ JKG 2017: Brněnská ODS si podává ruku s neonacisty a fotbalovými chuligány. In: *Antifa* [online]. 2017 [cit. 2019-02-23]. Dostupné z: <http://antifa.cz/content/jkg-2017-brnenska-ods-si-podava-ruku-s-neonacisty-fotbalovymi-chuligany>

2.3.3. Případová studie III – Úvodní fotografie fanoušků subkulturních stránek na Facebooku

Třetí případovou studii pro nás představují subkultury na Facebooku. Klíčovou součástí vymezení subkultur je styl, který je komunikován vůči okolí jako zásadní součást identity. „The communication of a significant difference, then (and the parallel communication of a group identity), is the ‚point‘ behind the style of all spectacular subcultures,“¹¹¹ říká Dick Hebdige. Z pohledu goffmanovsky orientované sociologie jsou stylové prvky kupříkladu fasády vědomě komunikované znaky, které nastavují rámec komunikace.

Pokud přijmeme toto východisko a pokud platí, že je profilová fotografie uživatele součástí jeho fasády, a pokud zároveň platí, že opakované „lajkování“ postů na stránkách mající deklarativně charakter blízký nějaké subkultuře (či kmeni) je výrazem blízkosti (či příslušnosti) k této subkultuře, tak by se to mělo v cover fotkách projevit výrazně větším zastoupením motivů spojených s danou subkulturou.

Překvapivě největší problém tvoří zajištění náhodného vzorku uživatelů na Facebooku. Protože není k dispozici žádný seznam facebookových stránek, kde by bylo možné klasicky vzorkovat (například každou 100. stránku apod.) museli jsme se uchýlit k jinému postupu. Na začátku jsme pracovali se seznamem cca 50 000 stránek, který nám poskytla společnost Socialbakers a který měla zavedený ve svém nástroji pro monitoring sociálních sítí Social Insider. Tento seznam vznikl kombinací požadavků klientů a ručním zařazením a jednalo se o seznam stránek, které publikují svůj obsah v češtině. V lednu 2017 jsme tento seznam, za použití Facebook API, rozšířili o stránky, které stránky z toho seznamu označily jako svoje oblíbené, a zároveň jsme u těchto stránek identifikovali, zda je dominantní počet fanoušků těchto stránek z České republiky. Vznikl tak seznam cca 300 000 českých facebookových stránek.

Z tohoto seznamu jsme náhodně vybrali 10 000 stránek, pro něž jsme stáhli posledních 300 postů. Z nich jsme opět náhodně vybrali jeden post a k němu stáhli všechny uživatele, kteří ho olajkovali. Takto vznikl dataset cca 300 000 lidí, z něž jsme náhodně vybrali 10 000 uživatelů jako náhodný vzorek.

¹¹¹ HEBDIGE, Dick. *Subculture: the meaning of style*. New York: Routledge, 1991. New accents (Routledge (Firm)). ISBN 04-150-3949-5., str. 102

Stahování profilových a úvodních fotografií přes API jsme provedli dvakrát.¹¹² Poprvé v únoru 2017 a podruhé v dubnu 2017. Tato data byla východiskem naší další analýzy a posloužila jako základní porovnávací datasety pro další analýzu.

Abychom nemuseli ručně klasifikovat obsah fotografií, využili jsme službu Google Cloud Vision¹¹³ pro automatické rozpoznání (otagování) objektů na fotce. Celkem se podařilo identifikovat objekty na 7 256 fotkách a přidělit jim 2 434 tagů.

Porovnání obou datasetů za únor 2017 a duben 2017 vykazuje poměrně velkou tematickou stabilitu. Výjimku tvoří pouze sezónní úvodní fotografie spojené se sněhem a obecně s počasím. V následující tabulce je vidět ukázka nejčastěji identifikovaných objektů v datasetu z února 2017 a v datasetu z dubna 2017. Sloupec afinita označuje, kolikrát častěji konkrétní objekty vyskytovaly se v dubnovém datasetu než v datasetu z února 2017.

¹¹² Pro stažení a analýzu dat z Facebooku jsme použili knihovnu sugaR pro jazyk R, kterou naprogramoval autor disertace <https://github.com/josefslerka/sugar>

¹¹³ Cloud Vision [online]. [cit. 2019-02-23]. Dostupné z: <https://cloud.google.com/vision/>

tag	bench_zastoupení	bench_procento	analyzed_zastoupení	analyzed_procento	affinita
person	1158	0.159592062	1184	0.164102564	1.0282627
people	983	0.135474090	976	0.135273735	0.9985211
photography	658	0.090683572	666	0.092307692	1.0179097
vehicle	627	0.086411246	607	0.084130284	0.9736034
black and white	540	0.074421169	504	0.069854470	0.9386371
child	531	0.073180816	535	0.074151074	1.0132584
photograph	491	0.067668137	454	0.062924463	0.9298980
mammal	481	0.066289967	482	0.066805267	1.0077734
image	479	0.066014333	495	0.068607069	1.0392753
font	459	0.063257993	421	0.058350658	0.9224235
sports	441	0.060777288	427	0.059182259	0.9737562
monochrome photography	438	0.060363837	417	0.057796258	0.9574649
winter	410	0.056504961	263	0.036451836	0.6451086
brand	403	0.055540243	388	0.053776854	0.9682503
snow	392	0.054024256	257	0.035620236	0.6593378
vertebrate	375	0.051681367	376	0.052113652	1.0083644
tree	371	0.051130099	343	0.047539848	0.9297820
flower	366	0.050441014	468	0.064864865	1.2859548
sea	360	0.049614112	373	0.051697852	1.0419989
monochrome	357	0.049200662	344	0.047678448	0.9690611
weather	344	0.047409041	223	0.030907831	0.6519396
black	331	0.045617420	298	0.041302841	0.9054182
screenshot	329	0.045341786	304	0.042134442	0.9292630
text	329	0.045341786	301	0.041718642	0.9200926

Tab. 7 - Objekty a témata identifikované v profilových fotkách uživatelů v datasetech z února a dubna 2017

Pro samotnou analýzu úvodních fotografií aktivních lajkujících jsme použili tři různé stránky. Dvě ryze subkulturní, a to stránku Otaku CZ/SK, tedy subkulturu spojenou s japonskými animovanými seriály anime, a stránku Hooligans.cz, tedy subkulturu spojenou s fotbalem. Pro porovnání jsme ještě vybrali stránku Miloš Zeman – prezident České republiky, která patří prezidentovi republiky a neměla by vykazovat rys subkultury.

Otaku Cz/Sk



Obr. 15. – Úvodní fotka stránky Otaku Cz/Sk na Facebooku v době naší analýzy

Stránka: Otaku Cz/Sk – 12.482 fanoušků (k 11. 4. 2017), přes API staženy všechny posty od 1. 1. 2017 do 31. 3. 2017, vybraný set cca 700 fanoušků, kteří olajkovali více než 20 postů, staženy jejich profilové fotografie a jejich úvodní fotografie. Pro kontrolu jsme se u 181 z nich podívali, jaké stránky mají označené ještě jako oblíbené. Výsledky, shrnuté v následující tabulce, odpovídají silnému zájmu o žánr anime.¹¹⁴

¹¹⁴ Viz následující případová studie věnovaná fanouškovským afinitám.

name	likes	normal	Freq	group	affinit
Otaku Cz/Sk	12482	0.004	166	0.91712707	257.166
Anime Cz/Sk	16860	0.005	100	0.55248619	114.692
Shirai.cz	20490	0.006	96	0.53038674	90.598
AnimeMania Cz/Sk	16200	0.005	94	0.51933702	112.202
PemiK	888663	0.254	75	0.41436464	1.632
Anime.Cz	10454	0.003	73	0.40331492	135.030
Gogomantv	449893	0.129	50	0.27624309	2.149
Naruto CZ/SK	39261	0.011	50	0.27624309	24.626
ANIME FAKTY	7284	0.002	46	0.25414365	122.117
EMEFKA	362094	0.103	44	0.24309392	2.350
Planeta Otaku	6473	0.002	42	0.23204420	125.468
VÍTE, ŽE?	285802	0.082	42	0.23204420	2.842
Fakty CZ&SK	244892	0.070	39	0.21546961	3.079
Pribeh Team	6325	0.002	39	0.21546961	119.232
Sarcasm	29100983	8.315	39	0.21546961	0.026
Anime Cz&Sk	5095	0.001	36	0.19889503	136.631
Hakuhodžó no Senši	4434	0.001	34	0.18784530	148.277
FattyPillow	238312	0.068	33	0.18232044	2.678
MenT	254660	0.073	32	0.17679558	2.430
Svět Otaku	4953	0.001	32	0.17679558	124.931
Zvrátený Humor	340595	0.097	32	0.17679558	1.817
Fantasy CZ/SK	37161	0.011	31	0.17127072	16.131
Trololol	366119	0.105	31	0.17127072	1.637
HERNÉ FAKTY	62852	0.018	30	0.16574586	9.230

Tab. 8 – Afinitní facebookové stránky vůči stránce Otaku Cz/Sk

Pro otagování úvodních fotografií jsme opět použili Google Cloud Vision API. Na výsledcích nejčastějších tagů je vidět, že motiv anime je v našem datasetu Otaku zastoupen 20x častěji než v běžné populaci a byl bezpečně rozpoznán v pětině fotografií.

tag	bench_zastoupení	bench_procento	analyzed_zastoupení	analyzed_procento	affinita
screenshot	329	0.0453417861	150	0.213068182	4.69915723
illustration	259	0.0356945976	150	0.213068182	5.96919972
anime	68	0.0093715546	135	0.191761364	20.46206551
font	459	0.0632579934	110	0.156250000	2.47004357
computer wallpaper	116	0.0159867696	96	0.136363636	8.52978056
cartoon	112	0.0154355017	83	0.117897727	7.63808847
brand	403	0.0555402426	79	0.112215909	2.02044327
text	329	0.0453417861	76	0.107954545	2.38090633
black and white	540	0.0744211687	75	0.106534091	1.43150253
darkness	212	0.0292171996	65	0.092329545	3.16010935
monochrome	357	0.0492006615	61	0.086647727	1.76110899
monochrome photography	438	0.0603638368	60	0.085227273	1.41189290
person	1158	0.1595920617	57	0.080965909	0.50733043
mangaka	13	0.0017916207	57	0.080965909	45.19143357
photography	658	0.0906835722	56	0.079545455	0.87717602
art	271	0.0373484013	51	0.072443182	1.93965951
sketch	116	0.0159867696	49	0.069602273	4.35374216
logo	146	0.0201212789	43	0.061079545	3.03556974
black	331	0.0456174201	41	0.058238636	1.27667536
drawing	150	0.0206725469	41	0.058238636	2.81719697
line	184	0.0253583241	40	0.056818182	2.24061265
photograph	491	0.0676681367	37	0.052556818	0.77668487
atmosphere	141	0.0194321940	37	0.052556818	2.70462605
album cover	187	0.0257717751	36	0.051136364	1.98420029

Tab. 9 – Zastoupení nejčastějších motivů v profilových fotkách neaktivnějších fanoušků stránky Otaku Cz/Sk v porovnání s průměrnou aktivní populací na Facebooku

Hooligans



Obr. 16. – Úvodní fotka stránky Hooligans.cz na Facebooku v době naší analýzy

Stránka: Hooligans.cz – 38 474 fanoušků (k 11. 4. 2017), opět přes API staženy všechny posty od 1. 2. 2017 do 31. 3. 2017. Vybraný set cca 700 fanoušků, kteří olajkovali více než 20 postů, přes API opět staženy jejich profilové fotografie a jejich úvodní fotografie a analyzovány pomocí Google Cloud Vision API.

tag	bench_zastoupení	bench_procento	analyzed_zastoupení	analyzed_procento	affinita
sport venue	89	0.0122657111	87	0.17718941	14.4459141
stadium	62	0.0085446527	78	0.15885947	18.5916825
structure	65	0.0089581036	73	0.14867617	16.5968353
person	1158	0.1595920617	65	0.13238289	0.8295080
sports	441	0.0607772878	60	0.12219959	2.0106128
arena	33	0.0045479603	56	0.11405295	25.0778251
font	459	0.0632579934	52	0.10590631	1.6741965
people	983	0.1354740904	47	0.09572301	0.7065780
brand	403	0.0555402426	39	0.07942974	1.4301294
player	101	0.0139195149	39	0.07942974	5.7063580
team	192	0.0264608600	34	0.06924644	2.6169382
black and white	540	0.0744211687	33	0.06720978	0.9031002
crowd	68	0.0093715546	33	0.06720978	7.1716784
screenshot	329	0.0453417861	31	0.06313646	1.3924563
image	479	0.0660143330	30	0.06109980	0.9255535
monochrome	357	0.0492006615	29	0.05906314	1.2004541
monochrome photography	438	0.0603638368	29	0.05906314	0.9784523
photography	658	0.0906835722	29	0.05906314	0.6513102
album cover	187	0.0257717751	28	0.05702648	2.2127493
soccer specific stadium	25	0.0034454245	28	0.05702648	16.5513646
text	329	0.0453417861	26	0.05295316	1.1678666
vehicle	627	0.0864112459	26	0.05295316	0.6128040
photograph	491	0.0676681367	23	0.04684318	0.6922487
soccer	62	0.0085446527	23	0.04684318	5.4821628

Tab. 10 – Zastoupení nejčastějších motivů v profilových fotkách neaktivnějších fanoušků stránky Hooligans.cz v porovnání s průměrnou aktivní populací na Facebooku

Výsledky opět ukazují na zastoupení tagů spojených se sportem.

Miloš Zeman – prezident České republiky



Obr. 17. – Úvodní fotka stránky Stránky Miloš Zeman na Facebooku v době naší analýzy

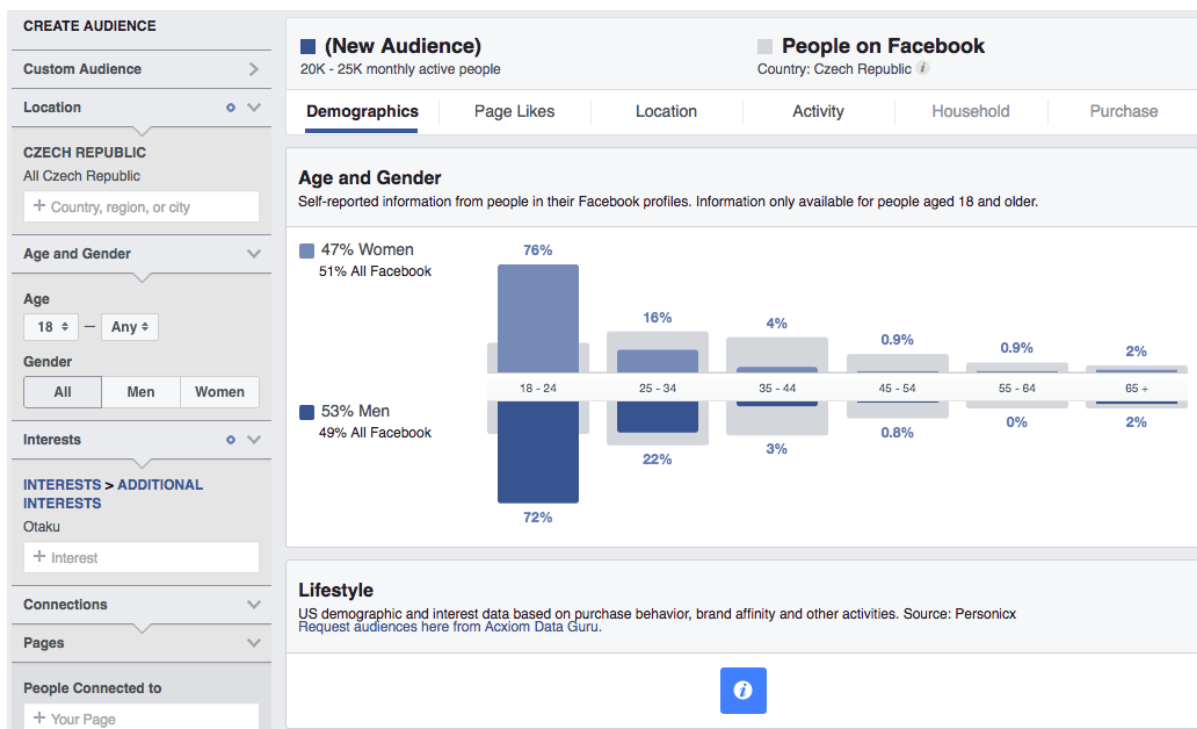
Stránky Miloš Zeman – prezident České republiky: 88 768 fanoušků (k 11. 4. 2017), přes API staženy všechny posty od 1. 2. 2017 do 31. 3. 2017, vybraný set cca 700 fanoušků, kteří olajkovali více než 20 postů, staženy jejich profilové fotografie a jejich cover fotografie. Přes API opět staženy jejich profilové fotografie a jejich úvodní fotografie a analyzovány pomocí Google Cloud Vision API.

tag	bench_zastoupení	bench_procento	analyzed_zastoupení	analyzed_procento	affinita
mammal	481	0.0662899669	39	0.134020619	2.0217331
vertebrate	375	0.0516813671	36	0.123711340	2.3937320
person	1158	0.1595920617	34	0.116838488	0.7321071
tree	371	0.0511300992	34	0.116838488	2.2851215
flower	366	0.0504410143	32	0.109965636	2.1800838
people	983	0.1354740904	31	0.106529210	0.7863438
dog	320	0.0441014333	28	0.096219931	2.1817869
dog like mammal	327	0.0450661521	28	0.096219931	2.1350820
plant	300	0.0413450937	25	0.085910653	2.0778923
vehicle	627	0.0864112459	24	0.082474227	0.9544386
sea	360	0.0496141125	16	0.054982818	1.1082092
land plant	153	0.0210859978	15	0.051546392	2.4445792
mountain	314	0.0432745314	15	0.051546392	1.1911485
vacation	280	0.0385887541	15	0.051546392	1.3357879
flora	136	0.0187431092	13	0.044673540	2.3834647
sky	255	0.0351433297	13	0.044673540	1.2711812
art	271	0.0373484013	12	0.041237113	1.1041199
cloud	208	0.0286659316	12	0.041237113	1.4385408
dog breed	141	0.0194321940	12	0.041237113	2.1221028
ecosystem	137	0.0188809261	12	0.041237113	2.1840620
fauna	88	0.0121278942	12	0.041237113	3.4001874
nature	152	0.0209481808	12	0.041237113	1.9685296
atmospheric phenomenon	227	0.0312844542	11	0.037800687	1.2082898
coast	200	0.0275633958	11	0.037800687	1.3714089
flowering plant	81	0.0111631753	11	0.037800687	3.3861949

Tab. 11 – Zastoupení nejčastějších motivů v profilových fotkách neaktivnějších fanoušků stránky Miloš Zeman v porovnání s průměrnou aktivní populací na Facebooku

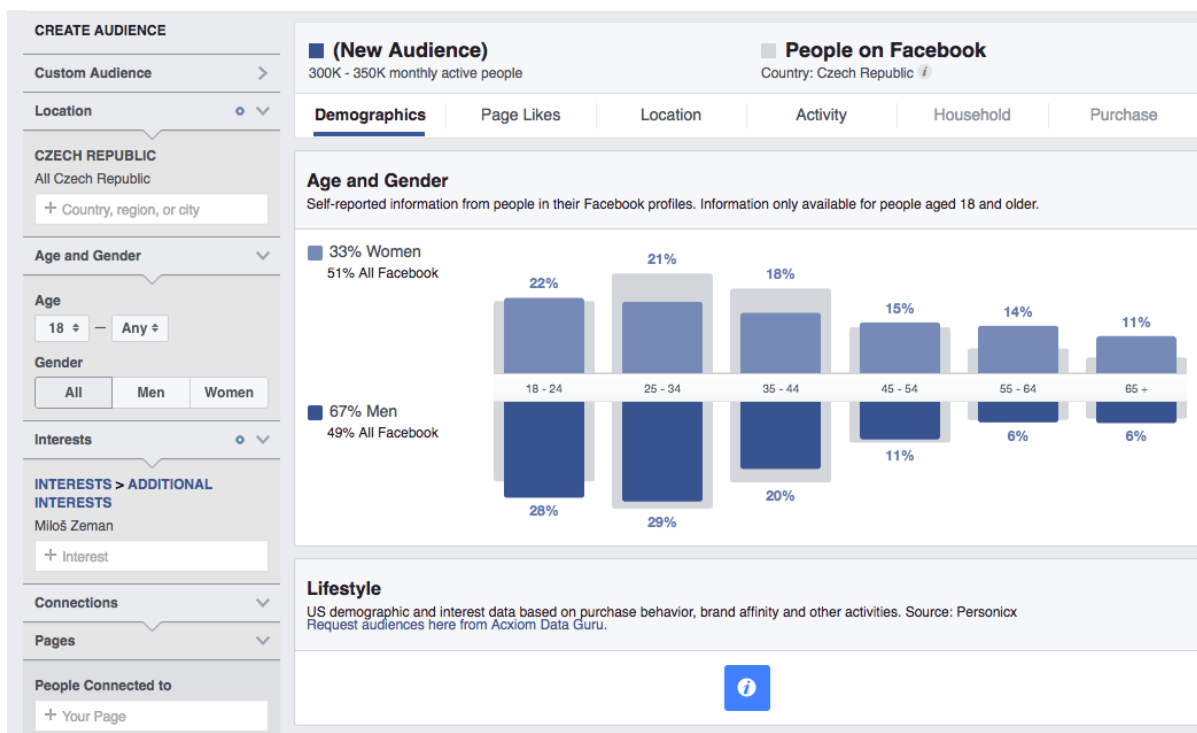
Podle očekávání se tu nevyskytují žádné silné afinity k nějakému specifickému tagu, byť některé tagy se tu objevují 2x až 3x častěji než v úplně běžném datasetu. Byť jsou to afinity slabé, dalo by se předpokládat, že mohou být zapříčiněny demografií uživatelů. Určitým vodítkem nám mohou být data z nástroje Facebook Audience Insights¹¹⁵, který slouží k lepšímu zacílení reklamy a obsahuje základní údaje o demografii uživatelů Facebooku. Tyto demografické údaje lze následně filtrovat pomocí řady parametrů včetně zemí, odkud uživatelé pocházejí, a poněkud vágní kategorie Interest, která pro Facebook vyjadřuje obecně zájem o nějaké téma. Téma anime a Miloš Zeman jsou tu jako možnost obsaženy a ukazují, že zatímco populace vyhodnocená Facebookem jako mající zájem o „otaku“ je výrazně zastoupena především mladšími uživateli:

¹¹⁵ Facebook Audience Insights [online]. In: . [cit. 2019-03-16]. Dostupné z: <https://www.facebook.com/ads/audience-insights/>



Obr. 18. – Věkový profil lidí zajímavých se o otaku na Facebooku podle Facebook Audience Insights

uživatelé, kteří mají podle Facebooku zájem o Miloše Zemana, jsou, oproti běžné populaci na Facebooku, o něco starší:

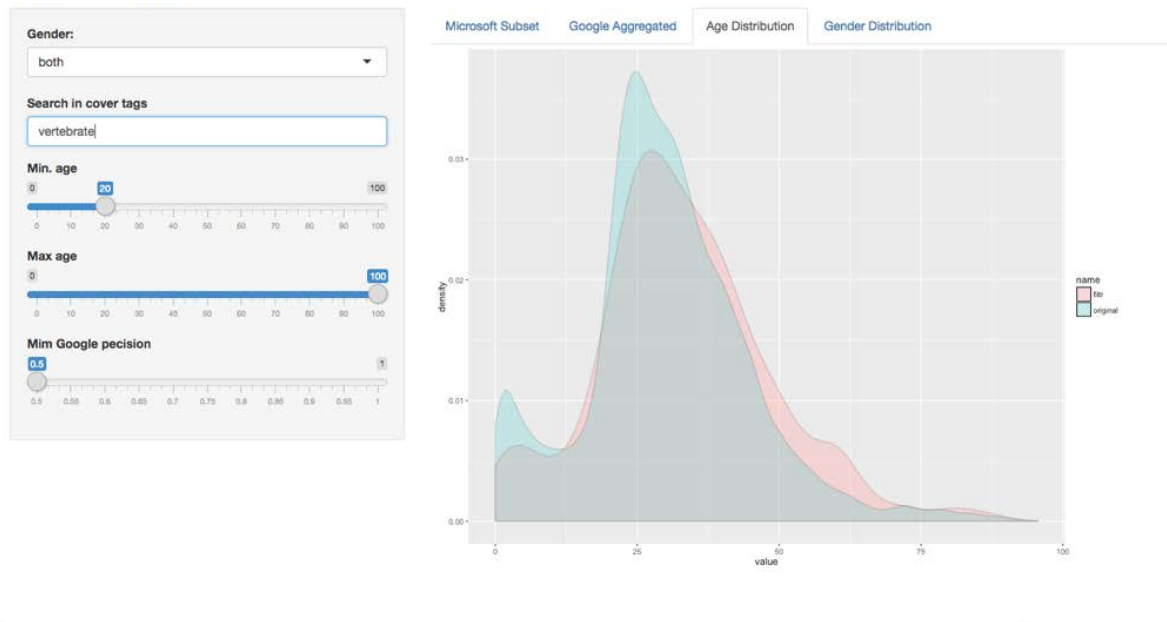


Obr. 19. – Věkový profil lidí zajímavých se o téma Miloš Zeman na Facebooku podle Facebook Audience Insights

Nabízí se tedy pracovní hypotéza, že některé typy předmětů v úvodních fotkách uživatelů budou zastoupeny i podle věku uživatelů. Bohužel však není jednoduchá možnost, jak přiřadit věk uživatelů k jejich profilům. Přesto se nám otevírá způsob, jak alespoň částečně ověřit platnost této hypotézy, a to opět pomocí strojového zpracování obrázkového obsahu.

Využili jsme přitom profilové fotografie uživatelů z datasetu únor 2017. V nich jsme pomocí služby Microsoft Face Recognition API nechali identifikovat osoby a věk na profilové fotografii. Celkem se identifikace zdařila u 5 261 profilů. Výsledky ukazují, že demografie užití specifických tagů slabě afinitních pro fanoušky stránky Miloše Zemana je skutečně věkově vychýlena směrem ke staršímu publiku.

Who are likers?



Who are likers?

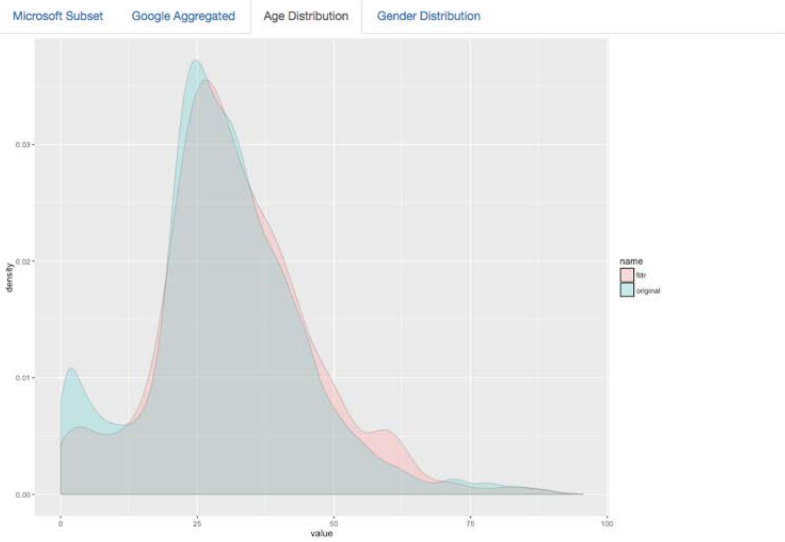
Gender:

Search in cover tags:

Min. age:

Max age:

Min Google precision:



Who are likers?

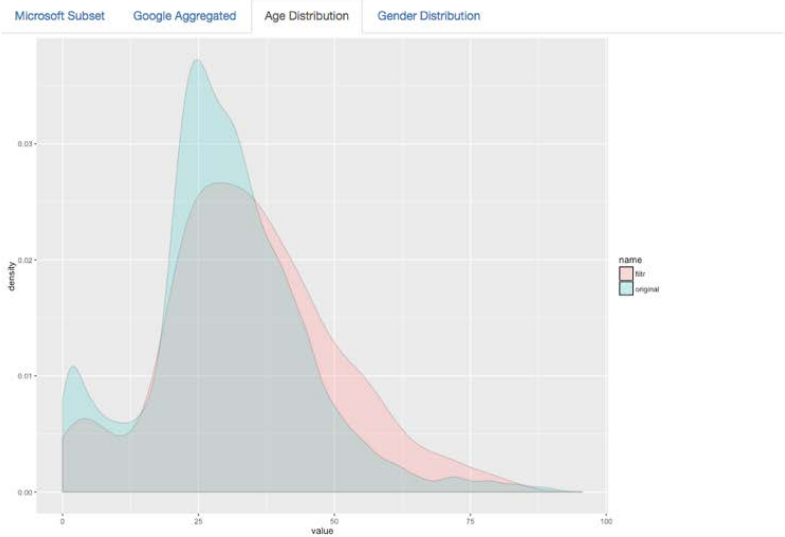
Gender:

Search in cover tags:

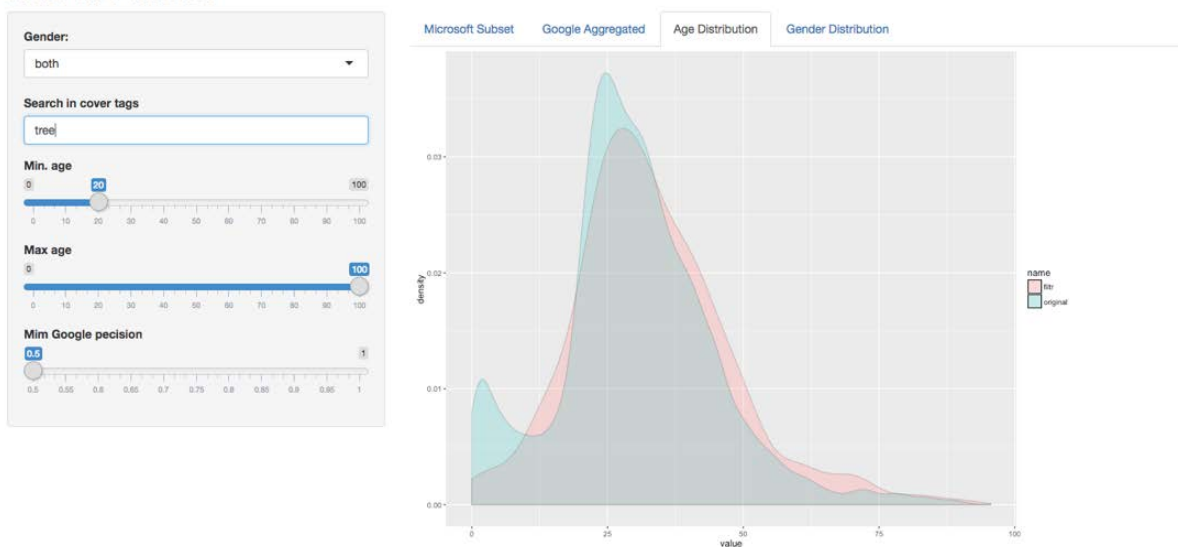
Min. age:

Max age:

Min Google precision:



Who are likers?



Obr. 20–23 rozložení věku uživatelů Facebooku podle objektu ve fotografii na úvodní fotografii osobního profilu

I tato případová studie potvrzuje, že základní goffmanovský rámec umožňuje dobře popsat specifické motivace jednání uživatelů. Z pohledu validity dat ovšem naráží na specifický problém, jak dobře vzorkovat data ze sociálních sítí. Zvolený postup sice zaručuje poměrně vysoký stupeň nahodilosti, přeci jen ale postihuje pouze uživatele aktivní na stránkách, nikoli všechny.

2.3.4. Případová studie IV. – Sebe prezentace IVČRN na českém Facebooku¹¹⁶

Výhodou samotného názvu stránky Islám v České republice nechceme je jednoznačná artikulace požadavku, který mají její příznivci a který je všechny spojuje. Takto vyslovený požadavek se však snaží zároveň jednoznačně odlišit samotné uskupení od jiných skupinových požadavků. Jde o princip minimální shody, pro jehož pochopení je třeba širšího kontextu.

Tím se pro nás stanou dvě různé, vzájemně se doplňující analýzy. První se věnuje analýze sítě stránek, které měla stránka IvČRn uvedeny jako svoje oblíbené stránky, tedy jako stránky, vůči kterým vyjadřuje vztah silného zájmu, který chce dát veřejně najevo i ostatním. Ty ukazují, kam se tvůrci stránky sami umisťují v prostoru českého Facebooku. Navazující analýza potom ukazuje, jaké stránky na Facebooku mají označené jako oblíbené fanoušci stránky IvČRn, a to zejména fanoušci patřící k tvrdému jádru.

Tyto dvě analýzy společně ukazují, že samotní aktivisté patří do širšího okruhu vzájemně se překrývajících populistických skupin.

Sít' oblíbených stránek stránky IvČRn

Sociální síť Facebook umožňuje administrátorům stránky označit jinou stránku jako svou oblíbenou. Administrátoři tak mohou veřejně ukázat svou náklonnost či vazbu k jiným stránkám. Protože nás zajímá primárně český prostor, vyřadili jsme pro další analýzu zahraniční stránky a stránky zpravodajských serverů. K 6. prosinci 2016 měla stránka IvČRn označeno 26 domácích stránek jako oblíbené. Jednalo se o tyto stránky:

Islám v ČR nechceme, Imigranty v ČR nechceme, Blok proti islamizaci, Blok proti islamizaci – Plzeňsko, Organizace proti multikulturnímu fanatismu, Naštvané matky, Moje země, moje pravidla, Martin Konvička, Svoboda a přímá demokracie Tomio Okamura – SPD, Pegida Česko, Václav Klaus ml., Blok proti islámu – Praha, Já islám v ČR Nechci, Jsme občané ČR a nechceme islám ve své zemi., Bubáci V Evropě, Prezident Zeman nerad islám. My také,

¹¹⁶ V upravené podobě vyšla tato studie jako samostatný příspěvek v ŠLERKA, Josef, FIALA, Jakub. O sebe prezentaci stránky Islám v ČR nechceme na českém Facebooku. In: OSTŘANSKÝ, B. Islamofobie po česku. Praha: Vyšehrad, 2017, s. 235-248. ISBN 978-80-7429-903-2.

Skutečně nekorektně, Konec Kalouska v Česku, Welcome to Reality, Radek John, Stop Islam a Bez politické korektnosti.

Tyto stránky lze podle témat rozdělit do několika skupin:

První z nich jsou stránky politiků, s nimiž je v různé míře spojena anti-islámská agenda. Jedná se o stránky Václav Klaus ml., Radek John a Martin Konvička.

Druhou skupinu stránek tvoří stránky politických hnutí či stran, které jsou opět explicitně spojeny s anti-islámskou tematikou. Jde o stránky: Blok proti islamizaci, Blok proti islamizaci – Plzeňsko, Svoboda a přímá demokracie Tomio Okamura – SPD, Pegida Česko a Blok proti islámu – Praha.

Třetí skupina je tvořena stránkami, které již v názvu vyjadřují určitý názorový postoj, ať už explicitně, nebo implicitně: Imigranty v ČR nechceme, Prezident Zeman nerad islám. My také, Stop Islam, Naštvané matky, Moje země, moje pravidla, Organizace proti multikulturnímu fanatismu, Já islám v ČR Nechci, Jsme občané ČR a nechceme islám ve své zemi a Bubáci v Evropě.

Čtvrtou skupinu pak tvoří stránky, které spojuje vnímání mainstreamového mediálního světa jako světa zkresleného a zároveň slibující nápravu tohoto zkreslení: Skutečně nekorektně, Welcome to Reality a Bez politické korektnosti.

Pátá skupina obsahuje jednu stránku, které se vymezují proti specifickým fenoménům. Jde o stránku Konec Kalouska v Česku.

Pokud bychom chtěli toto dělení ještě více zredukovat, můžeme skupiny 1–3 zařadit do skupiny spojené plně s anti-islámskou (případně anti-imigrační) agendou, zatímco skupiny 4 a 5 představují širší rámec spojený jak s obecnou nedůvěrou k médiím, tak k předchozím politickým reprezentacím.

Širší kontext sítě oblíbených stránek

V dalším kroku analýzy jsme se podívali na širší okruh těchto vybraných stránek. Cílem této části analýzy bylo vymezit širší kontext, v němž se pohybují tvůrci facebookových stránek.

Analyzovali jsme, jaké stránky mají tyto stránky označené jako oblíbené a jaké stránky mají za oblíbené označeny tyto stránky. Získali jsme tak síť, která má 1 598 uzlů a 2 863 hran. Jedná se o síť relativně řídkou, v rámci které má hodnotu vstupního spojení (in-degree) pět a více pouze horních 5 % stránek. Více než 10 příchozích spojení má pouze 28 stránek.

Jméno	Počet fanoušků	Hodnota indegree
Tomio Okamura – SPD	252931	33
Miloš Zeman – prezident České republiky	85029	13
Islám v ČR nechceme	72638	25
ParlamentníListy.cz	40501	14
Svoboda a přímá demokracie Tomio Okamura – SPD	26685	19
Stydím se za vládu ČR a chci její demisi	26189	10
Konec Kalouska v Česku	19334	12
STOP HALAL CZ	16191	13
Radim Fiala – SPD	10942	21
Svoboda a přímá demokracie Tomio Okamura Moravskoslezský kraj – SPD	4568	14
IVČRN Brno	4157	13
IVČRN Praha	2546	20
Svoboda a přímá demokracie Tomio Okamura Jihomoravský kraj – SPD	2420	14
Radek Rozvoral	2149	16
Svoboda a přímá demokracie Tomio Okamura Středočeský kraj – SPD	1937	14
Svoboda a přímá demokracie Tomio Okamura Plzeňský kraj – SPD	1801	14
Svoboda a přímá demokracie Tomio Okamura Liberecký kraj – SPD	1794	14
Svoboda a přímá demokracie Tomio Okamura Vysočina –	1769	11

SPD		
Svoboda a přímá demokracie Tomio Okamura Karlovarský kraj – SPD	1520	14
Svoboda a přímá demokracie Tomio Okamura Zlínský kraj – SPD	1398	15
Svoboda a přímá demokracie Tomio Okamura Praha – SPD	1351	14
Svoboda a přímá demokracie Tomio Okamura Královéhradecký kraj – SPD	1279	15
Svoboda a přímá demokracie Tomio Okamura Jihočeský kraj – SPD	1260	14
Svoboda a přímá demokracie Tomio Okamura Olomoucký kraj – SPD	1181	11
Svoboda a přímá demokracie Tomio Okamura Pardubický kraj – SPD	1058	14
Svoboda a přímá demokracie Tomio Okamura Ústecký kraj – SPD	1035	14
Jaroslav Holík	767	17

Tab. 12 – Přehled stránek s největším počtem indegree

Opět se nám tu vracejí stránky jednotlivých politiků, stran a médií.

Pokud se vrátíme k původnímu grafu, shlukovací algoritmus¹¹⁷ v něm identifikuje 16 shluků. Zde můžeme vidět opakování podobné struktury jako v původním vzorku.

Největší klastř je tvořen 182 stránkami, které v sobě zahrnují stránky lokálních organizací IVČRN, Bloku proti islamizaci, ale také stránky na podporu Miloše Zemana apod.

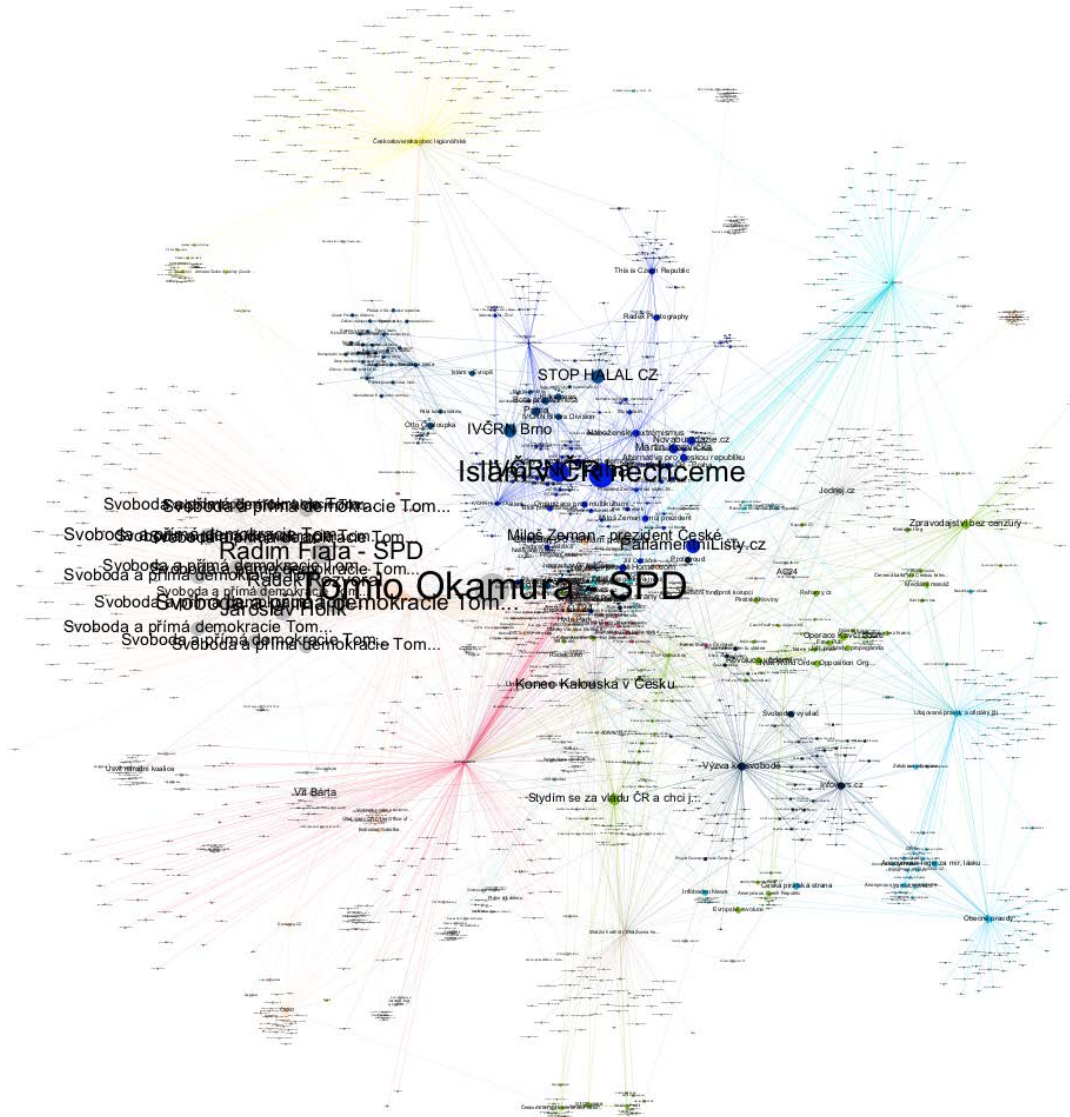
¹¹⁷ BLONDEL, Vincent D, Jean-Loup GUILLAUME, Renaud LAMBIOTTE a Etienne LEFEBVRE. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008, 2008(10). DOI: 10.1088/1742-5468/2008/10/P10008. ISSN 1742-5468. Dostupné také z: <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52>

Druhý největší klastr je tvořen stránkami mainstreamových médií a politiků (163 stránek). Ty jsou se zbytkem velkého uskupení spojeny stránkou Centra pro studium politického Islámu, která se prezentuje jako stránka náležící do oblasti mainstreamu.

Třetí největší klastr (153 stránek) je tvořen stránkami spojenými s alternativním zpravodajstvím: Zpravodajství bez cenzury, Revoluce vědomí, Free Pub, Operace Kavčí bouře, Disidenti.org, Pirátské Noviny a podobně. Dále pak jsou v něm zastoupeny anti-systémové stránky jako například Anonymous Czech Republic, Anonymous #czsk, anonymous media cz, New World Order Opposition Organization apod.

Podstatné je zastoupení stránek, které se explicitně nebo implicitně vymezují proti stávajícímu politickému systému: Stydím se za vládu ČR a chci její demisi či STOP vládě.

Specifickým případem je shluk stránek spojených s Okamurovým hnutím SPD, které se propojují vzájemně mezi sebou a vytvářejí tak kompaktní celek.



Obr. 24 – Síť stránek, které sledovala stránka Islám v ČR nechceme na Facebooku a stránky, které sledovaly tyto stránky

Z pohledu stránek, které stránka Islám v ČR nechceme měla v oblíbených, se sama řadila k širšímu proudu anti-islámského a anti-imigrantského hnutí včetně preferovaných politiků (Martin Konvička, Tomio Okamura, Václav Klaus ml. či Miloš Zeman). K tomuto základnímu vymezení se přidružuje určitá míra anti-systémového postoje a nedůvěry v klasická média. Toto základní vymezení potvrzuje i širší síť vztahů. Lze říci, že z pohledu vzájemného propojení se anti-islámské hnutí samo řadí do širšího anti-systémového hnutí.

Pokud se vrátíme do výchozího bodu našeho hledání kontextu a identity stránky IvČRN, předběžně vidíme, že i když je požadavek IvČRN formulován na principu minimální shody

vůči islámu, fakticky se řadí do prostoru protestních stránek a do kontextu stránek vyjadřujících nedůvěru jak vůči mainstreamovým médiím, tak vůči značné části politické scény. Jinými slovy vyvstává otázka, zda je požadavek odmítnutí Islámu takovým požadavkem, jehož uspokojení by vedlo k rozpuštění platformy, anebo zda je požadavek odmítnutí Islámu jen zástupným a nesplnitelným požadavkem, který zastřešuje obecnější nespokojenost. Tento předběžný závěr se potvrzuje i v další analýze.

Stránky oblíbené fanouškovským jádrem stránky Islám v ČR nechceme

Jak naznačuje předchozí výzkum, označení nějaké stránky jako oblíbené není jen výrazem našeho zájmu, ale také součástí naší sebe prezentace na Facebooku. Jinými slovy nejenže se prakticky zajímáme o danou tematiku, ale také chceme být vnímáni jako lidé, kteří se nějakým způsobem k tématu hlásí. Analýza toho, jaké stránky mají označené jako oblíbené aktivní fanoušci stránky, může tedy posloužit jako specifická sonda do postojů, které vzájemně sdílejí.

Pro naši analýzu jsme zvolili tzv. afinitní analýzu zájmů, kterou pro Facebook zavedl Jan Schmid a autor této disertace. Podstatou analýzy je informace o tom, které stránky má uživatel Facebooku označené jako oblíbené. Tato informace je v případě explicitního souhlasu uživatele dostupná na jeho profilu. V současnosti má takto otevřený profil přibližně polovina celé aktivní facebookové populace. V naší analýze jsme vyšli ze seznamu 1 000 nejaktivnějších uživatelů (těch, co na stránce zanechali nejvíce lajků pod jednotlivými posty v období od 1. 1. 2016 do 1. 7. 2016 a zároveň byli ještě alespoň jednou aktivní na znovuobnovené stránce IvČRn v zimě roku 2016). Celkem se podařilo získat údaje o oblíbených stránkách od 756 uživatelů; jednalo se o 71 228 stránek. Uživatel měl v průměru oblíbených 180 stránek, median byl 143 stránek.

Jak již bylo řečeno, pro afinitní analýzu stránek jsou důležité dva parametry. Seznam stránek, které vykazují signifikantně velký společný průnik zájmu, a výpočet afinity těchto stránek vůči analyzované populaci. Při ní se zohledňuje celková velikost stránky na Facebooku vůči zastoupení počtu fanoušků stránky uvnitř zkoumaného vzorku. Kupříkladu pokud máme ve zkoumaném vzorku o velikosti 1 000 uživatelů 100 fanoušků stránky X, je třeba ještě spočítat, kolik by jich tam bylo, kdyby se jednalo o naprosto náhodný vzorek. K tomuto údaji

se dostaneme, pokud vydělíme počet fanoušků stránky X velikostí celkové lokální populace na Facebooku a porovnáme ho s poměrem zastoupení ve zkoumaném vzorku.

V našem případě bylo mezi velmi aktivními 756 fanoušky stránky IvČRn přibližně 12 procent lidí, kteří byli zároveň fanoušky stránky obchodního řetězce Lidl Česká republika. Protože ale stránka Lidl Česká republika má celkově více než 54 000 fanoušků, můžeme říct, že je v našem vzorku zastoupena relativně stejně jako v populaci. Pokud počítáme aktivně populaci jako cca 3 500 000 fanoušků, tak by mělo být mezi sto náhodně vybranými fanoušky přibližně 15 procent fanoušků stránky Lidl, v našem případě jich je 12 procent. Lidl tedy není afinitní stránka pro jádro fanoušků IvČRn. Naopak stránka Martina Konvičky je oblíbená u skoro 40 procent zkoumaných fanoušků, což je 75x více, než by bylo v náhodném vzorku, a můžeme ji tedy považovat za velmi blízkou.

Mezi 71 228 stránkami je 11 stránek, u nichž můžeme najít společný průnik více než 20 procent fanoušků a zároveň silnou afinitu ke stránce Islám v ČR nechceme. Jedná se o následující stránky:

name	Počet fanoušků	Podíl ve vzorku	Afinita
Martin Konvička	17959	0,386	75x
STOP HALAL CZ	16312	0,313	67x
Blok proti islamizaci	12634	0,202	56x
Organizace proti multikulturnímu fanatismu	18820	0,269	50x
Naštvané matky	22773	0,253	39x
Václav Klaus ml.	42987	0,388	31x
Svoboda a přímá demokracie Tomio Okamura – SPD	27199	0,202	26x
Úsvit národní koalice	38668	0,223	20x
Imigranty v ČR nechceme	92611	0,427	16x
Miloš Zeman – prezident České republiky	85624	0,318	13x
We Are Here At Home.com	76953	0,271	12x

Tab. 13 – Přehled nejvíce afinitních facebookových stránek ke stránce Islám v ČR nechceme

Tyto stránky představují preference fanouškovského jádra. Vidíme, že seznam stránek je skoro úplnou podmnožinou stránek, které jsou zároveň označeny za oblíbené administrátory stránky Islám v ČR nechceme.

Podívejme se na širší seznam afinitních stránek. Budeme uvažovat takové, které mají překryv fanoušků alespoň deset procent a afinitu větší než 5.

name	Počet fanoušků	Podíl ve vzorku	Afinita
Alternativa pro Českou republiku	2498	0,116	163x
Martin Konvička	18070	0,386	74x
STOP HALAL CZ	16317	0,313	67x
Blok proti islámu – Praha	6090	0,105	60x
Petr Robejšek	9682	0,138	50x
Blok proti islamizaci	14378	0,202	49x
Organizace proti multikulturnímu fanatismu	19245	0,269	49x
Mešitu v Karlových Varech nechceme	7960	0,108	47x
Pegida Česko	10856	0,145	46x
Jsme občané ČR a nechceme islám ve své zemi.	13258	0,144	38x
Naštvané matky	23386	0,253	38x
Já jsem tady doma	17906	0,193	37x
Hej, občané z.s., za silnou a suverénní ČR	13334	0,132	34x
STOP Islámu	12264	0,119	33x
Národní domobrana	13178	0,126	33x
Náboženský extrémismus	12032	0,112	32x
Tomáš Ortel	10914	0,101	32x
Skutečně nekorektně	11361	0,103	31x
Islám v ČR nepreferujeme	13024	0,115	30x
Pravýprostor.cz	12061	0,104	30x
Václav Klaus ml.	46612	0,388	29x
Jiří Dvořák – Entrypoint	15774	0,120	26x
Miloš Zeman – můj prezident	17812	0,134	26x

Já islám v ČR NEchci	21079	0,158	26x
Svoboda a přímá demokracie Tomio Okamura – SPD	27222	0,202	26x
Miroslav Sládek	15474	0,109	24x
Nestydím se za svého prezidenta	16735	0,109	22x
Úsvit národní koalice	38624	0,223	20x
Pro-Vlast	23842	0,137	20x
Stop Islam	19133	0,104	19x
Bubáci V Evropě	29936	0,156	18x
Petice proti regulaci zbraní a sebeobránných prostředků ze strany EU	23072	0,112	17x
ParlamentníListy.cz	41887	0,194	16x
Imigranty v ČR nechceme	93950	0,427	15x
Evropo Povstaň/Europe Arise	33392	0,148	15x
Radostné Česko bez EET	27898	0,116	14x
Evropa – Evropanům	41726	0,154	12x
Miloš Zeman – prezident České republiky	86703	0,318	12x
We Are Here At Home.com	77809	0,271	12x
Welcome to Reality	41687	0,141	11x
Říkáme NE imigrantům v ČR	31744	0,107	11x
Moje země, moje pravidla	61335	0,199	11x
Strana svobodných občanů	71768	0,197	9x
Echo24.cz	41237	0,111	9x
Tomio Okamura – SPD	254114	0,509	7x
Armáda České republiky (Czech Army)	92341	0,138	5x

Tab. 14 – Přehled afinitních facebookových stránek ke stránce Islám v ČR nechceme

Rozšíření okruhu stránek afinitních k jádru fanoušků IvČRn ukazuje další pestrost antisystémového hnutí a potvrzuje tak validitu našeho předběžného závěru. Pozoruhodné je, jak se zde stále více prolínají v praxi zcela protikladné proudy. Vidíme zde podporovatele Miloše Zemana, kteří by měli patřit spíše ke konzervativní levici, vedle příznivců libertariánského hnutí u nás. Objevují se tu další drobnější politická hnutí a politici jako

Miroslav Sládek či Petr Robjšek a tyto polarity jsou vidět i na úrovni zastoupení médií, kdy Parlamentní listy jsou doplněny svým prakticky politickým protikladem, internetovým deníkem Echo 24.

3. Normalized Social Distance

3.1. Modely podobností

Viděli jsme, jakou zásadní roli hraje podobnost pro samotné sdružování do sítí a jak tvoří základní organizační princip sociálních sítí. Podobnost sama se zdá být pořadajícím principem pro řadu jevů nejen v lidské psychice a ve společnosti, ale také ve způsobu, jak světu rozumíme.^{118, 119} Jádrem je zřejmě empirická zkušenost, že se stoupajícím rozpoznáním podobností mezi A a B stoupá i pravděpodobnost, že B se bude chovat stejně vůči X, jako se chová A vůči X.¹²⁰ Díky tomuto očekávání jsme schopni na základě podobnosti dělat předpovědi v nových situacích. Podobní ptáci podobně zpívají.

I když má koncepce podobnosti svoje zásadní kritiky, jako například Nelsona Goodmana, který krizuje ontologický status podobnosti v duchu svého nominalismu¹²¹, přesto se dnes úloze podobnosti věnuje v kognitivních i dalších vědách plná pozornost. Jeden z problémů, které tato úloha přináší, je způsob, jak podobnost mezi objekty kvantifikovat. Dnes je převážně využíván jeden z pěti základních modelů:¹²² geometrický model, set-theoretic model, alignment-based model, transformační model a model vycházející z informační teorie.

První z těchto modelů, **tzv. model geometrický** je také nejznámější a nejlivnější. Tversky jej vymezuje takto: „These models represent objects as points in some coordinate space such that the observed dissimilarities between objects correspond to the metric distances between the respective points. Practically all analyses of proximity data have been metric in nature,

¹¹⁸ Ulrike Hahn ve svém článku HAHN, Ulrike, Nick CHATER a Lucy B RICHARDSON. Similarity as transformation. *Cognition*. 2003, **87**(1), 1–32. DOI: 10.1016/S0010-0277(02)00184-1. ISSN 00100277. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S0010027702001841>

odkazuje na celou řadu studií, které se tohoto principu týkají, ať již jde o mechanismus vzpomínání, kategorizace, vizuálního hledání, řešení problému, učení se, lingvistických znalostí a podobně.

¹¹⁹ GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press, 2005, s. 13–36. ISBN 9780521531016.

¹²⁰ GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press, 2005, ISBN 9780521531016., str. 11

¹²¹ GOODMAN, Nelson. Sedm výhrad proti podobnosti. In: Aluze [online]. 2008 [cit. 2019-02-23]. Dostupné z: http://aluze.cz/2008_02/07_studie_goodman.php

¹²² GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press, 2005, ISBN 9780521531016., str. 15–28

although some (e.g., hierarchical clustering) yield tree-like structures rather than dimensionally organized spaces.“¹²³

V praxi je obvykle vyžadováno, aby postup, které měří vzdálenost (podobnost – nepodobnost) produkoval metriky, které vyhovují třem základním požadavkům:

Minimality: $\delta(a, b) \geq \delta(a, a) = 0$.

Symmetry: $\delta(a, b) = \delta(b, a)$.

The triangle inequality: $\delta(a, b) + \delta(b, c) \geq \delta(a, c)$.

Splnění těchto základních požadavků totiž podmiňuje celou řadu dalších analýz jako kupříkladu velmi populární multidimensional scaling a další. Jak ale ukazují někteří kritici, kupříkladu Tversky, geometrický model občas naráží na problémy už v samotném očekávání symetrie podobnosti. Řada psychologických experimentů totiž ukázala, že není kupříkladu vždy funkční předpoklad symetrie podobnosti. V jednom z experimentů se kupříkladu předložilo 21 párů zemí dvěma skupinám studentů. V každém páru byla jedna země dominantní, kupříkladu pár Čína a Severní Korea. Studenti měli určit na dvacetimístné škále, jak je země A podobná zemi B. Byli rozděleni do dvou stejně velkých skupin, přičemž jediný rozdíl v zadání byl v pořadí zemí. Jedna měla kupříkladu určit, jak je Sovětský svaz podobný Polsku, zatímco druhá, jak je podobné Polsko Sovětskému svazu. Výsledky prokázaly zásadní asymetrii mezi oběma skupinami.¹²⁴ Prvním výsledkem bylo prokázání existence asymetrické podobnosti, druhým pak Tverského pokus o takový model, který by si s tímto problémem dokázal poradit.

Proto Tverský přinesl takzvaný **set-theoretic model**. „In light of the previous potential problems for geometric representations, Tversky proposed to characterize similarity in terms of a feature-matching process based on weighting common and distinctive features. In this model, entities are represented as a collection of features and similarity is computed by

¹²³ TVERSKY, Amos a Eldar SHAFIR. Preference, belief, and similarity: selected writings. Cambridge, Mass.: MIT Press, c2004. ISBN 978-0262700931.

¹²⁴ TVERSKY, Amos a Eldar SHAFIR. Preference, belief, and similarity: selected writings. Cambridge, Mass.: MIT Press, c2004. ISBN 978-0262700931., str. 17

$$\begin{aligned}
 S(A, B) \\
 = \theta f(A \cap B) - af(A - B) - bf(B - A) \\
 (2.2)
 \end{aligned}$$

The similarity of A to B is expressed as a linear combination of the measure of the common and distinctive features. The term $(A \cap B)$ represents the features that items A and B have in common. $(A - B)$ represents the features that A has but B does not. $(B - A)$ represents the features of B that are not in A. θ , a , and b are weights for the common and distinctive components.¹²⁵

Názornou ilustraci tohoto najdeme například v knize *Encyclopedia of the Mind*,¹²⁶ kde se porovnává, jak jsou si podobní roboti a zombie.

¹²⁵ GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. Cambridge Handbook of Thinking and Reasoning. Cambridge: Cambridge University Press, 2005, s. 13–36. ISBN 9780521531016., str. 21

¹²⁶ PASHLER, Harold E., ed. Encyclopedia of the mind. Los Angeles: SAGE, [2013]. SAGE reference. ISBN 978-1412950572.

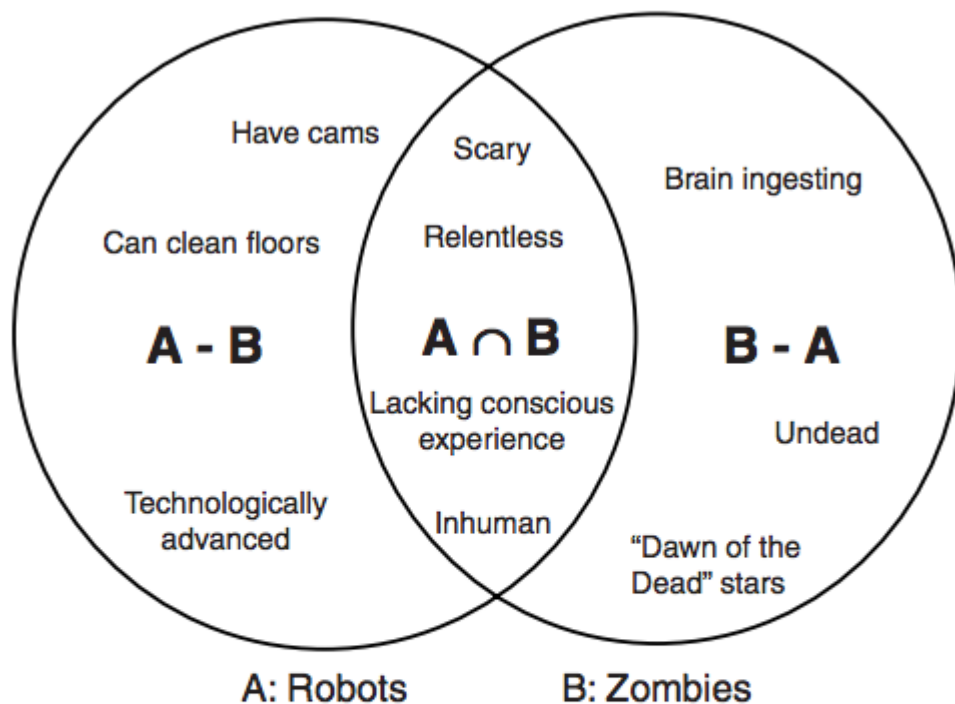


Figure 2 A Venn diagram showing some of the features of robots and zombies

Note: Featural models take the similarity of robots to zombies to be a positive function of the features shared by both and negative functions of the features possessed by one but not the other.

Obr. 25 – Model základních principů set-theoretic modelu

Kritici tohoto modelu namítali, že tento přístup nezohledňuje strukturální pozice jednotlivých vlastností. Jako reakce vznikl **alignment-based model of similarity**. Ten vychází nejen z průniku vlastností, ale také z jejich strukturální pozice. Jednoduchým příkladem může být auto se zelenými koly a nákladník se zelenou plachtou. Oba sdílejí vlastnost zelený, avšak to nezvyšuje jejich podobnost, protože kolo a plachta spolu nejsou na stejné strukturální pozici.¹²⁷

¹²⁷ GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. Cambridge Handbook of Thinking and Reasoning. Cambridge: Cambridge University Press, 2005., ISBN 9780521531016., str. 16

Vedle těchto přístupů, založených na myšlence porovnání mezi objekty, se můžeme setkat s přístupy dynamičtějšími, které opouštějí otázku podobnosti a tvrdí, že klíčovou roli pro podobnost hraje složitost transformace jednoho objektu na druhý. Sem patří model podobnosti nazývaný **transformační model podobnosti**. Vychází z myšlenky, že objekty jsou si tím podobnější, čím je jednodušší transformovat jeden objekt do druhého: „The simpler the transformation sequence, the more similar the entities are assumed to be. For example, the transformational complexity connecting 1 2 3 4 5 6 7 8 and 2 3 4 5 6 7 8 9 is small because the simple instruction ‚add 1 to each digit‘ suffices to transform one into the other. Experiments demonstrate that once reasonable vocabularies of transformation are postulated, transformational complexity does indeed predict subjective similarity ratings. Furthermore, when a new transformation is learned that turns Object A into Object B, A is seen as more similar to B.“¹²⁸

Existuje i celá řada pokusů klasifikovat podobnost podle míry náročnosti transformace a definovat, že podobnost klesá s rostoucí mírou náročnosti transformace. Shiro Imai uvádí čtyři základní transformace:¹²⁹

Type of Transformation		Pattern Pair	
Mirror-image	<i>M</i>	000●000●000●	●000●000●000
Phase	<i>P</i>	000●000●000●	00●000●000●0
Reversal	<i>R</i>	000●000●000●	●●●●●●●●●●
Wave length	<i>W</i>	00●●00●●00●●	00●●00●●00●●

Fig. 2. Illustrations of four basic transformations: Mirror-image transformation *M*, phase transformation *P*, reversal transformation *R*, and wave length transformation *W*.

Obr. 26 – Příklad základních principů transformačního modelu podobnosti

¹²⁸ PASHLER, Harold E., ed. Encyclopedia of the mind. Los Angeles: SAGE, [2013]. SAGE reference. ISBN 978-1412950572.

¹²⁹ IMAI, Shiro. Pattern similarity and cognitive transformations. Acta Psychologica. 1977, 41(6), 433-447. DOI: 10.1016/0001-6918(77)90002-6. ISSN 00016918. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/0001691877900026>, 433–447.

Pátým a zatím posledním silným modelem je přístup tzv. **informační vzdálenosti**, který navazuje na transformační model, ale opírá jej o algoritmickou informační teorii. Právě tento přístup, který odvozuje svůj původ od Kolmogorovy teorie komplexity, bude naším východiskem při návrhu metriky pro analýzu dat ze sociálních sítí. Věnovat se mu budeme v následujících kapitolách.

3.2. Teorie Kolmogorovy komplexity

Teorie Kolmogorovy komplexity se snaží odpovědět na otázku „Co je nahodilý objekt?“ Michal Koucký uvádí následující příklad (volně dle jeho článku A Brief Introduction to Kolmogorov Complexity).¹³⁰ Mějme k dispozici tři číselné řetězce v desítkové soustavě:

- A. 3333333333
- B. 3141596535
- C. 84354279521

Který z nich bychom považovali za náhodný? Podle Kouckého by většina lidí označila jako nahodilé řetězce B a C, protože řetězec A je prostě jen opakování čísla tři. Matematici by zřejmě souhlasili s tím, že ani řetězec B není náhodný, jedná se totiž o číslo pí. Jako nahodilý se tak jeví řetězec C. Pokud budeme o uvažovat o pravděpodobnosti, že někdo napíše některý z daných řetězců, pokud ho požádáme, aby napsal jedenáctimístné číslo, musíme uznat, že je to 10^{-11} . Náhodnost čísla (randomness) tak intuitivně nesouvisí s pravděpodobností.

Koucký pokračuje ve svém myšlenkovém experimentu a vyzývá nás, abychom si představili, že máme těmito řetězci čísel popsanou knihu. V případě A opakovaná čísla 3, v případě B Ludolfovým číslem a v případě C pak řetězcem 84354279521... a tak dále až dokud nedojde k vyčerpání počtu stran a teoreticky do nekonečna. Právě toto je intuitivní základ, který vede ke Kolmogorově komplexitě. Čím delší je popis postupu, který potřebujeme k popsání řetězce, tím je řetězec více komplexní. V Kolmogorově pojetí však nejde o popis v přirozeném jazyce, ale o existenci univerzálního počítačového stroje (Turingova stroje), který takový popis generuje, a délku tohoto programu.

¹³⁰ KOUCKÝ, Michal. A Brief Introduction to Kolmogorov Complexity [online]. In: 4. 5. 2006 [cit. 2019-02-23]. Dostupné z: <https://iuuk.mff.cuni.cz/~koucky/vyuka/ZS2013/kolmcomp.pdf>

3.3. Informační vzdálenost a Normalized Compression Distance (NCD)

Tato intuitivní pojetí podobnosti jsou i základem pro celou řadu způsobů, jak měřit podobnost mezi instancemi, a jsou i základem pro informační vzdálenost formalizovanou studií *Information Distance*¹³¹ a dále rozpracovanou v řadě dalších textů¹³². Teorie informační vzdálenosti představuje rozšíření Kolmogorovy komplexity o myšlenku vzdálenosti mezi řetězci, respektive jejich podobnosti. Podle ní je minimální informační vzdálenost mezi dvěma instancemi (řetězce x a y) vyjádřena délkou nejkratšího programu, který transformuje jeden řetězec na druhý a naopak. Univerzální informační vzdálenost je pak vyjádřena vzorcem $E(x,y) = \max\{K(x|y), K(y|x)\}$.

Od něj autoři odvozují obecnou normalizovanou informační vzdálenost (normalized information distance), která by byla schopna produkovat i metrickou vzdálenost. Výsledkem je následující vzorec¹³³:

$$e(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}}. \quad (3.3)$$

Teorie informační vzdálenosti je teoretická konstrukce, kterou není možné v praxi vytvořit, protože je závislá na nespočitatelné funkci K . Je však možné použít jiné funkce, které se v reálném světě o podobnou funkčnost snaží. Těmito programy jsou dle autorů kompresní algoritmy, které mají za úkol spočítat co největší bezztrátovou kompresi dat, tedy co největší redukci komplexit pomocí univerzálního programu. Kompresi dat definuje Wikipedie jako „zpracování počítačových dat s cílem zmenšit jejich objem (jednotka bajt) při současném zachování informací v datech obsažených. Úkolem komprese dat je zmenšit datový tok při jejich přenosu nebo zmenšit potřebu zdrojů při ukládání informací.“¹³⁴ Obvykle se snaží algoritmicky nalézt opakující se sekvence znaků a vytvořit z nich slovník, který umožňuje

¹³¹ BENNETT, Charles H. Information distance. IEEE Transactions on information theory. 1998, **44**(4), 1407–1423.

¹³² VITÁNYI, Paul M. B. Normalized information distance. In: Information theory and statistical learning. Boston: Springer, 2009, s. 45–82.

¹³³ VITÁNYI, Paul M. B. Normalized information distance. In: Information theory and statistical learning. Boston: Springer, 2009, str. 47

¹³⁴ Kompresi dat. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://cs.wikipedia.org/wiki/Kompresi_dat

odkaz na přesné místo, a tím zmenšit celkovou velikost souboru. Pokud se ve dvou souborech opakují stejné sekvence, bude výsledná komprese úspěšnější.

Odpovídá to i zkušenosti, kterou máme z jejich každodenního používání. Pokud kompresním programem zpracujeme dva soubory, které jsou si podobnější než jiné dva, je rozdíl mezi výslednou délkou nového souboru a délkou odpovídající součtu délek původních souborů menší.

Upravený vzorec vypadá následovně:¹³⁵

$$e_Z(x,y) = \frac{Z(xy) - \min\{Z(x), Z(y)\}}{\max\{Z(x), Z(y)\}}, \quad (3.5)$$

Přičemž Z je kompresní algoritmus a x a y zůstávají řetězce určené k porovnání.

Autoři NCD provedli sérii testů navrženého postupu na celé řadě druhů řetězců. Jejich výsledky se zdají potvrzovat univerzální charakter navrženého modelu, a to včetně klasifikace heterogenních řetězců. Vstupními daty pro experiment byla genetická data, ukázky z literárních textů, MIDI soubory, binární počítačové programy a zkompilevané programy ze zdrojových kódů programovacího jazyka Java. Pro kompresi sloužil kompresní algoritmus bzip a metoda quartet clustering¹³⁶. Výsledkem bylo korektní rozdělení jednotlivých souborů:

¹³⁵ VITÁNYI, Paul M. B. Normalized information distance. In: Information theory and statistical learning. Boston: Springer, 2009, str. 55

¹³⁶ CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403.

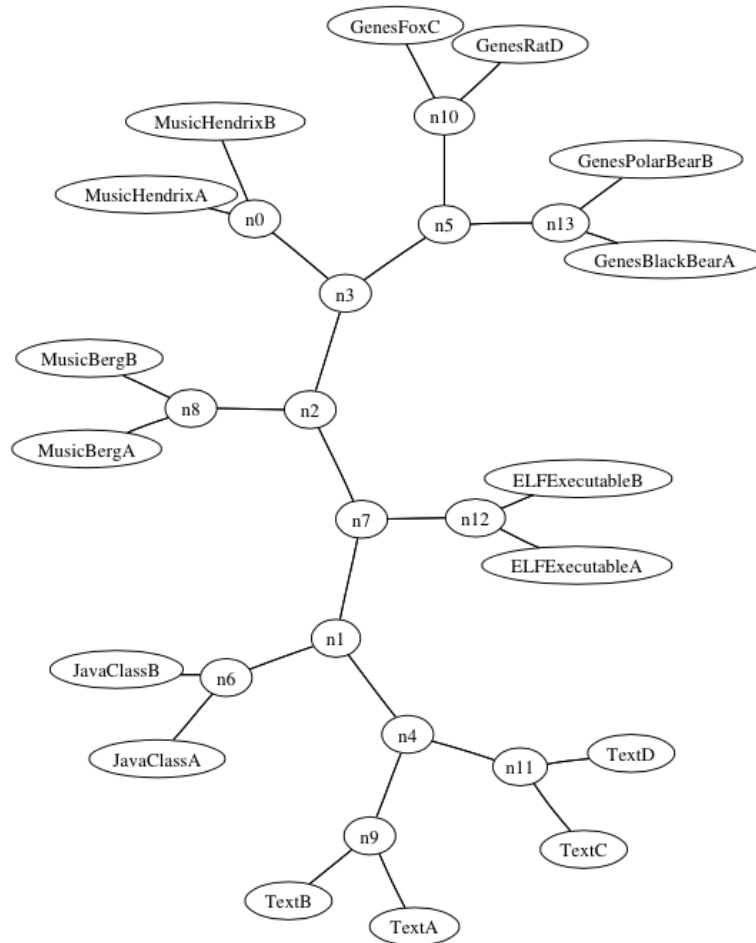


Fig. 3.4 Clustering of heterogeneous file types using the NCD, based on the **bzip2** compressor, and a quartet clustering method. The set of file contains four MIDI files, four genomes, four English texts, and two Java class files and Linux executables

Obr. 27 – Výsledky klastrování heterogenních řetězců pomocí NCD

Experimenty dalších autorů potvrzují výsledky původních experimentů, včetně předpokládané odolnosti NCD proti šumům v textu¹³⁷, dále se věnují jeho dalším aplikacím například pro automatickou evaluaci strojového překladu.¹³⁸ Další studie, za účasti autora původního týmu Paula Vitanyiho, pak sledují využití NCD při klastrování.¹³⁹

¹³⁷ CEBRIN, Manuel, Manuel ALFONSECA a Alfonso ORTEGA. The Normalized Compression Distance Is Resistant to Noise. *IEEE Transactions on Information Theory*. 2007, **53**(5), 1895–1900. DOI: 10.1109/TIT.2007.894669. ISSN 0018-9448. Dostupné také z: <http://ieeexplore.ieee.org/document/4167725/>

¹³⁸ VÄYRYNEN, Jaakko J. Normalized compression distance as an automatic MT evaluation metric. *Proceedings of MT*. 2010, (25), 21–22.

¹³⁹ CILIBRASI, R. a P. M. B. VITANYI. Clustering by Compression. *IEEE Transactions on Information Theory*. 2005, **51**(4), 1523–1545. DOI: 10.1109/TIT.2005.844059. ISSN 0018-9448. Dostupné také z: <http://ieeexplore.ieee.org/document/1412045/>

3.4. Normalized Web Distance

Aplikace teorie informační vzdálenosti v NCD se omezuje pouze na řetězce, nikoli na ideje nebo pojmy. Proto se její autoři rozhodli přijít s metrikou, která toto omezení překračuje, a to pomocí využití konceptu World Wide Webu.

Podle Cilibrasiho každý kompresní algoritmus „only looks for statistical biases, repetitions, and other biases in symmetrically defined local contexts, and cannot achieve compression even for very low-complexity meaningful strings like the digits of π .“¹⁴⁰ Symetricky je tak možné chápat výskyt slov na stránkách jako případ univerzální distribuce slov na stránkách a vyhledávač pak jako určitý druh pseudokompresního algoritmu, který zohledňuje všechny dimenze lidského mínění zachycené v dokumentech indexovaných vyhledávačem.

Vitányi pak formalizuje tyto úvahy takto:

„Let W be the set of pages of the World Wide Web, and let $x \subseteq W$ be the set of pages containing the search term x . By the Conditional Coding Theorem we have $\log 1/m(x|x \subseteq W) = K(x|x \subseteq W) + O(1)$, where m is the universal lower semicomputable discrete semimeasure. This equality relates the incompressibility of the set of pages on the Web, containing a given search term, to its universal probability. While we do not know how to compute m , a natural heuristic now is to use the distribution of x in the Web to approximate $m(x|x \subseteq W)$. (We give a simplified approach where we assume that every page contains at most one search term.) Let us define the probability mass function $g(x)$ to be the probability that the search term x appears in a page indexed by a given internet search engine G , that is, the number of pages returned divided by the overall number of pages indexed. Then the Shannon-Fano code associated with g can be set at

$$G(x) = \log \frac{1}{g(x)} .$$

Replacing $Z(x)$ by $G(x)$ in the formula in (3.5), we obtain the distance eG , often called the Normalized Web Distance (NWD):

¹⁴⁰ CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403., str. 111

$$\begin{aligned}
 e_G(x,y) &= \frac{G(xy) - \min\{G(x), G(y)\}}{\max\{G(x), G(y)\}} \\
 &= \frac{\max\{\log f(x), \log f(y)\} - \log f(x,y)}{\log N - \min\{\log f(x), \log f(y)\}}.
 \end{aligned}
 \tag{3.6}$$

where $f(x)$ is the number of pages containing x , the frequency $f(x, y)$ is the number of pages containing both x and y , and N is the total number of indexed pages. We can view the search engine G as a compressor using the Web, and $G(x)$ as the binary length of the compressed version of the set of all pages containing the search term x , given the indexed pages on the Web. The distance e_G is actually a family of distances parametrized with the search engine G . It was originally called Normalized Google Distance (NGD) and thus featured a particular search engine. The name Normalized Web Distance is more generic and more in line with the name NCD, which also does not mention a concrete compressor.¹⁴¹

Cilibrasi popisuje ve své disertaci *Statistical inference through data compression* základní kontrast mezi oběma přístupy takto:

„The first type is the NCD based on a literal interpretation of the data: the data is the object itself. The second type is the NGD masses of contexts expressing a large body of common-sense knowledge. It may be said that the first case ignores the meaning of the message, whereas the second focuses on it.“¹⁴²

Série experimentů provedné Cilibrasim a Vitányim pomocí výsledků vyhledávače Googlu přináší v tomto ohledu velmi uspokojivé výsledky.¹⁴³ Předmětem experimentů byly názvy děl holandských malířů 17. století, názvy anglických románů, čísla a barvy a názvy Shakespearových děl. Ve všech případech dokázal postup díla správně rozdělit. Předmětem experimentu byla i rekonstrukce vazeb a hierarchií vyjádřených experty ve WordNetu. Zde

¹⁴¹ VITÁNYI, Paul M. B. Normalized information distance. In: Information theory and statistical learning. Boston: Springer, 2009, str. 64

¹⁴² CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403. str. 113

¹⁴³ CILIBRASI, Rudi L. a Paul M.B. VITANYI. The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering. 2007, 19(3), 370–383. DOI: 10.1109/TKDE.2007.48. ISSN 1041-4347. Dostupné také z: <https://ieeexplore.ieee.org/document/4072748/>

byla přesnost mezi NGD a vazbou ve WordNetu 0.8725.¹⁴⁴ Zdá se, že idea, při které autoři „have replaced the obstinate objectivity of classical compressors with an anthropomorphic subjectivity derived from the efforts of millions of people worldwide“¹⁴⁵ přinesla své ovoce. Protože Google výrazně omezil využívání svého vyhledávače, začali autoři hovořit spíše o Normalized Web Distance, kde je Google možné prostě nahradit libovolnými vyhledávačem nad dostatečně velkým objemem textů.

¹⁴⁴ CILIBRASI, Rudi L. a Paul M.B. VITANYI. The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering. 2007, 19(3), 370–383. DOI: 10.1109/TKDE.2007.48. ISSN 1041-4347. Dostupné také z: <https://ieeexplore.ieee.org/document/4072748/>, str. 13

¹⁴⁵ CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403, str. 112

3.5. Normalized Social Distance

Posun od Normalized Compression Distance k Normalized Google Distance je posunem od de facto syntaxe k sémantice. **V tomto kontextu bychom rádi navrhli další posun, a to od sémantiky k pragmatice.** Princip homofilí, o kterém jsme hovořili v předcházejících částech jako o základním principu fungování sociálních skupin, lze propojit s přístupy, které tvrdí, že člověk je v podstatě sít' přesvědčení a hodnot, které zapříčiňují záměrné jednání. Lidé, kteří sdílejí toto či podobné pojetí, jakou jsou Richard Rorty, Donald Davidson či Daniel Dennet, budou v zásadě souhlasit s tím, že „přesvědčení a touhy se spolčují, aby zapříčinily, zracionalizovaly a objasnily záměrné jednání. Záměrně jednáme z nějakých důvodů a naše důvody vždy obsahují jak hodnoty, tak přesvědčení.“¹⁴⁶ Naše přesvědčení nikdy neexistují samostatně izolovaně. „Přesvědčení se nevyskytují nezávisle na sobě. Přesvědčení je určováno a tvořeno (mezi jinými faktory) vztahem k jiným přesvědčením. Věřím-li, že vidím kočku, musím zároveň věřit i mnoha jiným věcem: musím vědět, co je kočka, co je vidění, a především musím mít povědomí o možnosti svého omylu, jakkoli je nepravděpodobná.“¹⁴⁷

Cesta k získávání různých přesvědčení i hodnot je komplikovaná a individuální. Dalo by se dokonce říci, že právě existence jedinečné sítě hodnot a přesvědčení každého člověka zakládá lidskou individualitu. Nicméně s lidskou společností sdílíme řadu přesvědčení, která jsme získali působením celé řady vnějších faktorů. Princip homofilie říká, že některé obecné faktory, jako je rasa, etnicita, vzdělání, pohlaví, věk či náboženství, vysvětlují sdílení některých přesvědčení ve skupinách více než jiné.

Z této perspektivy se lze vrátit k sociálním akcím, jako je lajkování postů na Facebooku či sledování nějakého uživatele na Twitteru a podobně. Všechny tyto akce jsme označili za vědomé jednání, které je výrazem určitého přesvědčení a je součástí sebeprezentace jedince. Právě to, že je součástí sebeprezentace jedince, znamená, že je určena k interpretaci ostatními. V sociálních sítích, které nabízejí dostatečné množství možností vyjádřit svoje postoje a hodnoty snadno kvantifikovatelnými akcemi, lze tedy vystavět analogii, v níž je

¹⁴⁶ DAVIDSON, Donald. *Subjektivita, intersubjektivita, objektivita*. Praha: Filosofia, 2004. ISBN 80-700-7190-7., str. 147

¹⁴⁷ DAVIDSON, Donald. *Subjektivita, intersubjektivita, objektivita*. Praha: Filosofia, 2004. ISBN 80-700-7190-7., str. 145

jedinec plně popsatelný svými sociálními akcemi („jsi to, co lajkuješ“), stejně tak lze měřit blízkost přesvědčení dle souvýskytu v různých sadách jedinců.

Uživatel se tak stává analogií dokumentu jako v případě Normalized Google Distance, jen místo slov, která jsou v dokumentech obvykle spojována sémanticky, jsou tu záznamy jeho sociálních akcí, které jsou indexy jeho přesvědčení a hodnot. V praxi tak v dostatečně velké a uživatelsky pestré sociální síti můžeme získat představu o míře spolčení v jednotlivých přesvědčeních v daném vzorku populace.

Obecně tak lze navrhnout analogický způsob výpočtu vzdálenosti mezi dvěma sociálními postoji po vzoru Normalized Google Distance. **Budeme o ní mluvit jako o Normalized Social Distance.** Samotné označení Normalized Social Distance jsme odvodili od materiality toho, z čeho je počítána, nikoli od faktu, že se věnuje hodnotám nebo postojům a podobně. V tom se držíme linie předchozích vzorců.

Narozdíl od Normalized Google Distance ale x a y nejsou počty dokumentů, v nichž se vyskytují daná slova, ale počty unikátních uživatelů, kteří svou sociální akcí (například olajkováním postů daných stránek) vyjádřili svoje přesvědčení. Průnikem se pak myslí, kolik unikátních uživatelů udělalo obě aktivity. Místo celkového počtu indexovaných dokumentů pak N nahrazuje předpokládaný počet uživatelů dané sítě. Jelikož se jedná de facto o analog aplikace Normalized Google Distance, platí pro ni stejná matematická omezení včetně faktu, že za určitých okolností nespĺňuje požadavek euklidovské triangulace. V případě, že totiž není žádný průnik mezi sety, produkují nekonečnou vzdálenost. Výsledný vzorec pak vypadá takto:

$$NSD(x, y) = \frac{\max \{ \log f(x), \log f(y) \} - \log f(x \cap y)}{\log N - \min \{ \log f(x), \log f(y) \}}$$

Pro potřeby zkoumání jednotlivých sítí lze do vzorce dosadit konkrétní hodnoty pro jednotlivá sociální média. Lze například počítat Normalized Facebook Distance, kdy pro vyjádření abstraktní vzdálenosti mezi dvěma stránkami na Facebooku dosadíme do našeho vzorce následující parametry:

x je počet aktivních uživatelů na jedné stránce,

y vyjadřuje počet aktivních uživatelů na stránce druhé,
x, y počet aktivních uživatelů na obou stránkách
N je celkový počet uživatelů českého Facebooku.

Podobně lze počítat konkrétní aplikace pro další sítě, jako třeba Twitter nebo Instagram, v závislosti na zkoumaném problému. V kapitole věnované případovým studiím, které se snaží potvrdit smysluplnost užívání této metriky, ostatně uvidíme právě aplikace těchto metod na zmíněné sociální sítě.

Nyní však zůstaneme u dat ze sítě Facebook a v následujících dvou případových studiích si ukážeme, jaké možnosti přináší analýza dat z těchto sítí právě pomocí NSD. První studie *Analýza fanoušků DSSS* se věnuje postupu, při kterém jsme měli k dispozici dataset cca 300 000 facebookových stránek, k nimž jsme získali seznamy všech ID uživatelů, kteří na této stránce olajkovali nějaký post. Dataset vznikl v rámci komerčního projektu Social Insider ve společnosti Ataxo v roce 2013 a byl využíván primárně pro komerční analýzy. Tento způsob přístupu k datům umožňovalo tehdejší nastavení Facebook API a bohužel dnes již není reálně dostupné, přesto ale dobře ukazuje základní možnosti explorační analýzy pomocí NSD. Druhá případová studie *Who is Shaping Your Agenda?* pak aplikuje tuto metodologii na průzkum vzdálenosti předem definovaného seznamu stránek. Oba přístupy jsou vzájemně komplementární a dobře ilustrují základní možnosti analýzy.

3.5.1. Případová studie I. – Analýza fanoušků DSSS

Pokud chápeme popis uživatele jako síť rozdílů v preferovaných stránkách, nabízí se možnost nejen věnovat se celkové charakteristice fanoušků prostřednictvím distančního modelu, ale také jejich podrobnější charakteristice, přesněji nalezení zřetelně odlišených subskupin.

Východiskem pro takové hledání může být nikoli distanční model, který je fakticky jednorozměrný, nýbrž matice s údaji o dalších stránkách, na kterých uživatel něco olajkoval, případně se stránkami, jejichž fanoušky jsou aktivní lajkeři na analyzované stránce.

Takováto matice je v podstatě bimodální síť s relativně nízkou hustotou, může ale být podrobena některým klasickým exploračním technikám, jako je hierarchický klustering, multidimenzionální scaling či analýza základních komponent (PCA).

Pro průzkum takovýchto matic jsme vytvořili aplikaci Facebook profiling, která má na vstupu dva soubory. První je tabulka s distančním modelem fanoušků, druhým pak binární matice obsahující na řádcích ID uživatelů a ve sloupcích pak stránky, v nichž se fanoušci zkoumané stránky nejčastěji zapojují svým like.

Vyjděme opět z politického subjektu. Za příklad nám může sloužit stránka Dělnické strany sociální spravedlnosti, která je považována za tolerovanou formu neonacismu na české politické scéně. Minimální hranici pro průnik jsme v našem případě stanovili na 5 % a blízkost menší než 0,7.

Následující tabulka pak zahrnuje jak vypočítanou vzdálenost stránky od stránek DSSS, tak afinitu stránek pro lepší orientaci:

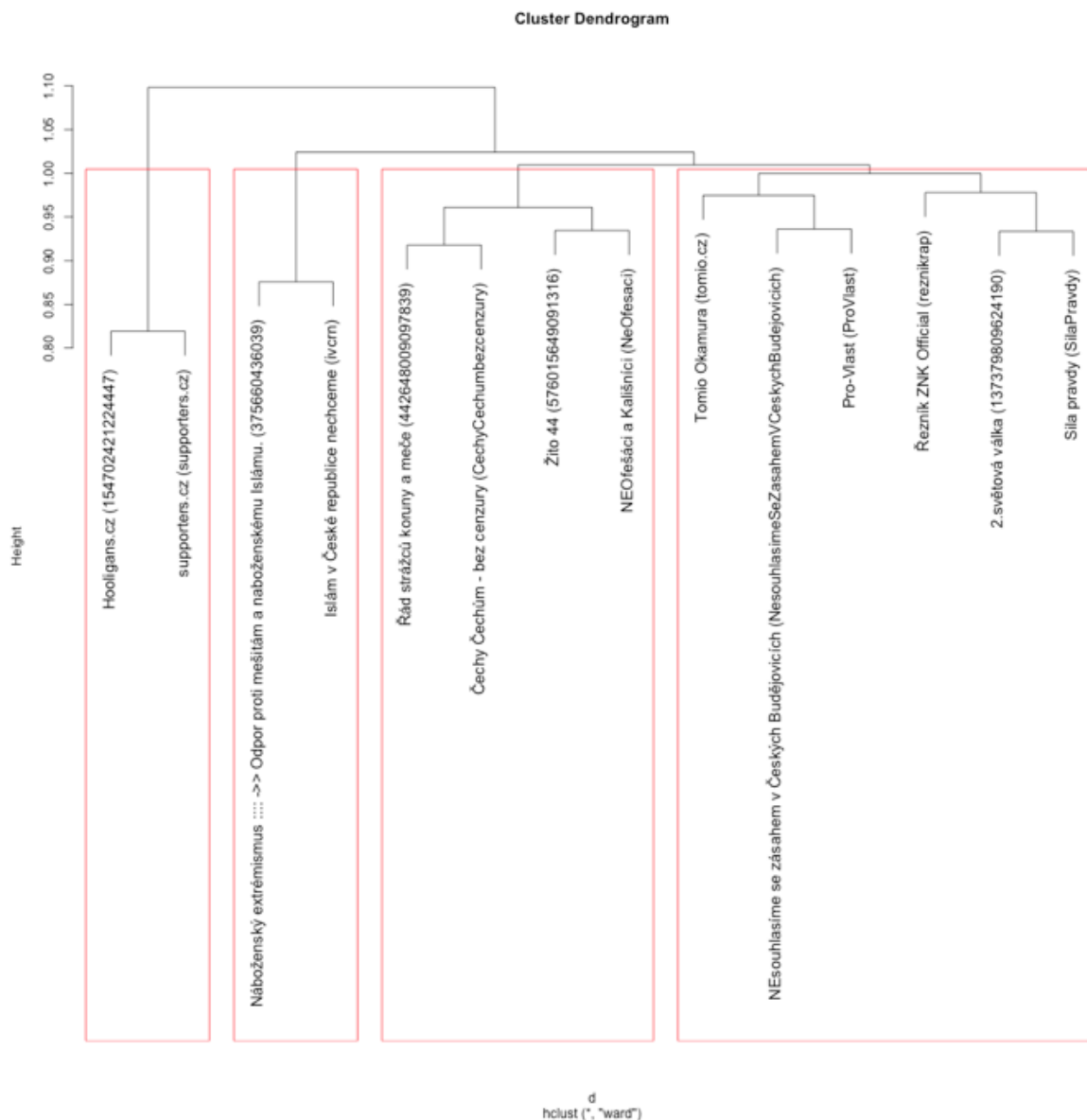
Jméno stránky	Počet unikátně lajkujících	Procentuální podíl	Afinita	NSD
Čechy Čechům – bez cenzury	217	10,91	167x	0,325
NEsouhlasíme se zásahem v Českých Budějovicích	109	5,48	126x	0,375
Pro-Vlast	100	5,028	115x	0,386
NEOfešáci a Kališníci	128	6,435	103x	0,388
Náboženský extrémismus::: ->> Odpor proti mešitám a náboženskému Islámu.	127	6,385	85x	0,414
Žito 44	103	5,178	68x	0,442

Islám v České republice nechceme	164	8,245	45x	0,497
Hooligans.cz	163	8,195	32x	0,539
2. světová válka	153	7,692	29x	0,554
supporters.cz	136	6,838	27x	0,565
Řezník ZNK Official	110	5,53	16x	0,635
Řád strážců koruny a meče	187	9,402	13x	0,654
Síla pravdy	110	5,53	12x	0,672
Tomio Okamura	176	8,849	11x	0,676

Tab. 15 – Přehled nejbližších facebookových stránek ke stránce DSSS na základě NSD

Seznam stránek nepůsobí překvapivě. Objevují se tu stránky vysloveně neonacistické, stránky se silně xenofobním naladěním, stejně jako stránky projektu Daniela Landy. Z celkového počtu 1 989 fanoušků ve zkoumaném datasetu jich na některé z těchto stránek alespoň jednou něco olajkovala přibližně polovina (973 fanoušků). Ze seznamu ale není vůbec jasné, zda se jedná o skupinu jednolitou, či zda mezi fanoušky existují nějaké rozdíly podle toho, na jakých dalších stránkách jsou ještě aktivní. Zde přichází čas na aplikaci různých postupů explorační analýzy dat.

V prvním kole nám může pomoci hierarchické shlukování stránek. Jeho grafický výsledek (včetně rozdělení do tří klastrů) vidíte na následujícím obrázku.

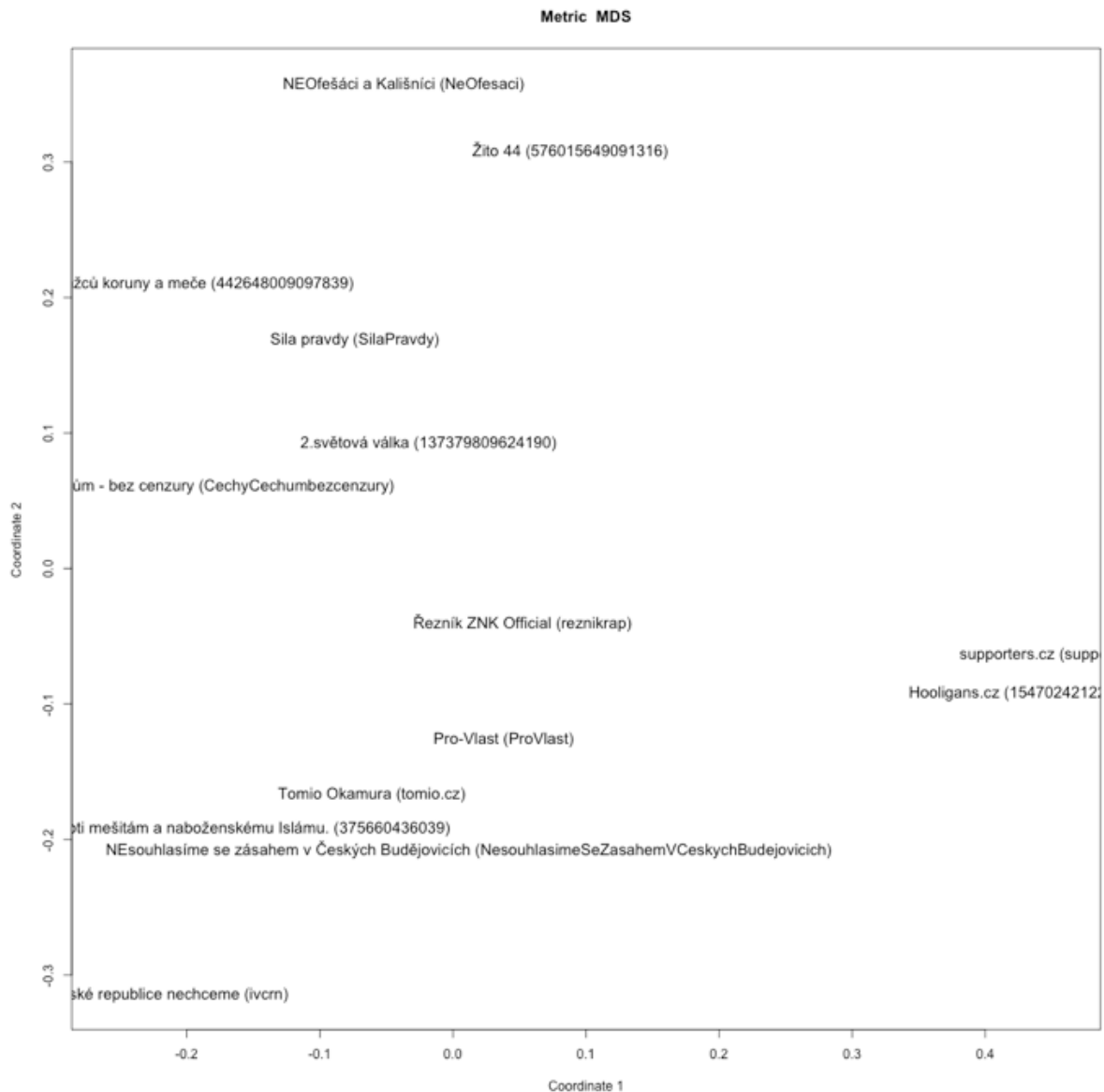


Obr. 28. – Dendrogram výsledků hierarchického klustrování pro DSSS

Předběžně vidíme čtyři základní klastry, které by šlo pojmenovat jako Hooligans, Islamofobové, Tradiční neonacisté a Okamurovci. Počet zvolených klastrů je dán hledáním rozumné míry mezi zřetelným odlišením, počtem klastrů a mírou granularity. Pro 14 stránek se zdá být počet čtyř klastrů celkem přijatelný.

Pro lepší vizualizaci vzájemných vztahů mezi stránkami samotnými můžeme využít techniku MDS, zmiňovanou již v předchozích částech, a pokusit se najít dimenze, které by šlo jednoznačně pojmenovat.

V našem případě vypadá mapa fanoušků následovně:

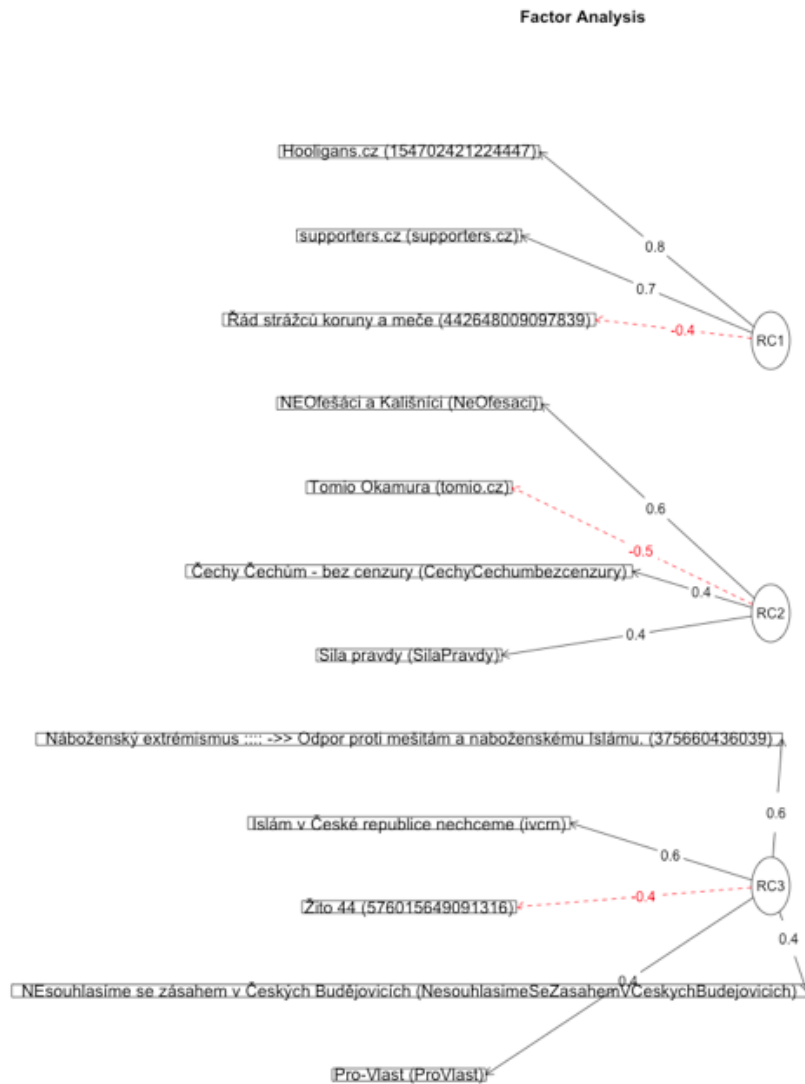


Obr. 29. – Graf MDS výsledků distanční matice na základě NSD pro DSSS

Osa X by mohla odpovídat míře apolitičnosti, osa Y pak přibližně odpovídá míře napětí mezi starým a novým politickým extrémismem. Starý byl spojen především s Danielem Landou jako symbolem, nový s Tomiem Okamurou.

Další pohled nám může poskytnout PCA analýza, při níž se pokusíme redukovat počet stránek na ty, které vnitřně od sebe nejlépe odlišují komponenty tvořené fanoušky DSSS na

Facebooku. Jelikož se jedná o exploratorní analýzu, je třeba takový poměr hledat v přiměřeném počtu komponent a jejich zatížení. V našem případě by se mohlo jednat například o tuto kombinaci:



Obr. 30. – Graf faktorové analýzy výsledků distanční matice na základě NSD pro DSSS

Ta nejen potvrzuje základní tendenci, kterou jsme viděli v shlukovací analýze a MDS, ale také ukazuje, jak je pro skupinu nových neonacistů landovský projekt Žito 44 nezajímavý, stejně jako je pro staré neonacisty nepřijatelný Tomio Okamura.

3.5.2. Případová studie II. – Who is Shaping Your Agenda?

V případové studii *Who is Shaping Your Agenda? Social network analysis of anti-Islam and anti-immigration movement audiences on Czech Facebook*¹⁴⁸ jsme aplikovali metodu Normalized Social Distance pro shlukování stránek českého anti-islámského hnutí v době vrcholící tzv. imigrační krize. Základním úkolem studie bylo prozkoumat, jak blízké jsou si facebookové stránky reprezentující stránky anti-islámského hnutí u nás, dále pak stránky politických stran a hnutí, politických think-tanků a politiků a také stránky vybraných politických stran. Celkem bylo na vstupu 56 ručně vybraných stránek politických stran a hnutí, osobností a domácích.¹⁴⁹ Pro potřeby anglické verze článku jsme doplnili kódování jednotlivých stránek dle typu:

(n) = news media – zpravodajské servery

(m) = political movement or party - politické strany a hnutí

(p) = individual politician – stránky individuálních politiků

přičemž jsme ve specifických případech uvedli v závorkách příslušné zařazení jako například Miloš Zeman [President].

Pro aplikování NSD na dataset jsme použili knihovnu sugaR pro programovací jazyk R, kterou autor této disertační práce napsal jako součást výzkumu v jejím rámci. Jako základní dataset posloužily všechny veřejné posty administrátorů těchto stránek mezi 1. zářím 2015 a

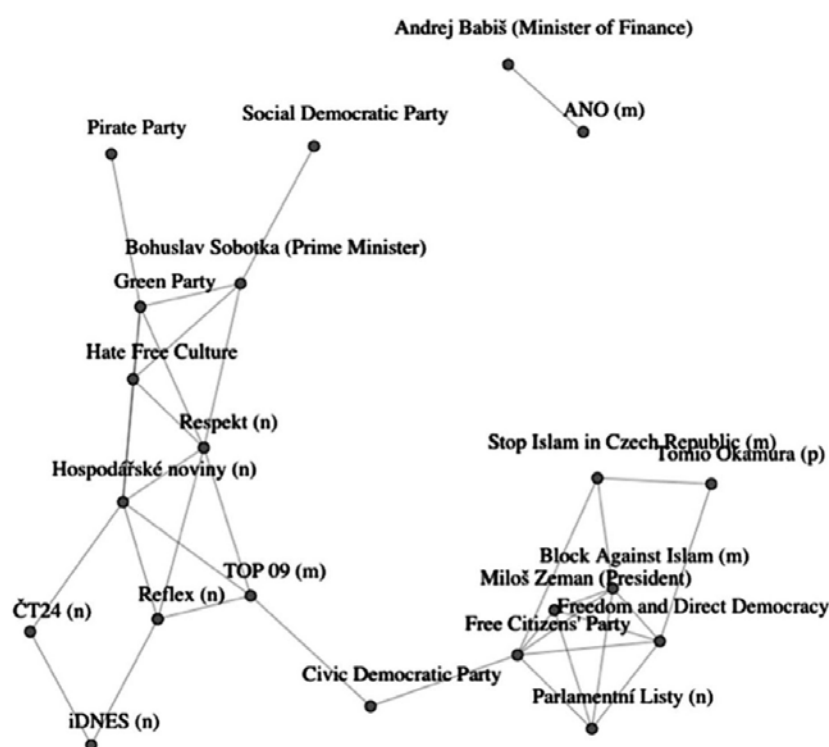
¹⁴⁸ Rozšířená verze této případové studie byla publikována v ŠLERKA, Josef a Vít ŠISLER. Who Is Shaping Your Agenda? Social Network Analysis of Anti-Islam and Anti-immigration Movement Audiences on Czech Facebook. In: Expressions of Radicalization. Cham: Springer International Publishing, 2018, 2018-12-20, s. 61–85. DOI: 10.1007/978-3-319-65566-6_3. ISBN 978-3-319-65565-9. Dostupné také z: http://link.springer.com/10.1007/978-3-319-65566-6_3

¹⁴⁹ Jednalo se o tyto stránky: Daniel Herman (Minister of Culture), Martin Stropnický (Minister of Defence), Jan Veleba (p), Svobodné fórum (n), Alexandra Udženija (p), Andrej Babiš (Minister of Finance), ANO (m), Pavel Bělobrádek (Minister of Science), Blesk (n), Block Against Islam (m), Pirate Party, Milan Chovanec (Minister of the Interior), Social Democratic Party, ČT24 (n), Echo24 (n), Jiří Dienstbier (Minister for Human rights), Referendum (n), European Commission CR, European Values, Generation Identity, Hate Free Culture, Freedom and Direct Democracy, iDNES (n), Hospodářské noviny (n), Miroslav Lidinský (p), Stop Islam in Czech Republic (m), Jana Černochová (p), Marian Jurečka (p), Miroslav Kalousek (p), Christian and Democratic Union (m), Martin Konvička (p), Communist Party, Lidové noviny (n), Lubomír Zaorálek (Minister of Foreign Affairs), Michaela Marksová-Tominová (Minister of Social Affairs), NO to Brussels – National Democracy (m), Svatopluk Němeček (Minister of Health), Novinky (n), Civic Democratic Party, Parlamentní Listy (n), Petr Fiala (p), Pravý břeh (n), Miloš Zeman (President), Czech Radio – Radiožurnál (n), Reflex (n), Karla Šlechtová (Minister of Regional Development), Bohuslav Sobotka (Prime Minister), Green Party, Pavel Svoboda (p), Free Citizens' Party, Tomio Okamura (p), Tomáš Zdechovský (p), TOP 09 (m), TV Noe (n), Respekt (n), Kateřina Valachová (Minister of Education)

28. prosincem 2015. Jednalo se celkem o 19 321 postů. Následně jsme pro tyto posty stáhli všechny uživatele, kteří jim dali lajk. Celkem se jednalo o seznam 540 775 unikátních ID. Tito uživatelé udělili celkem postům v daném období celkem 3 351 034 lajků. Jako základ pro počet aktivních uživatelů jsme vzali střední hodnotu počtu uživatelů Facebooku v České republice, tak jak ji uváděla společnost ve svém nástroji Facebook Audience Insights (2016), tedy 3 500 000 unikátních uživatelů.

Na základě těchto dat jsme spočítali matici obsahující vzájemnou vzdálenost jednotlivých stránek.

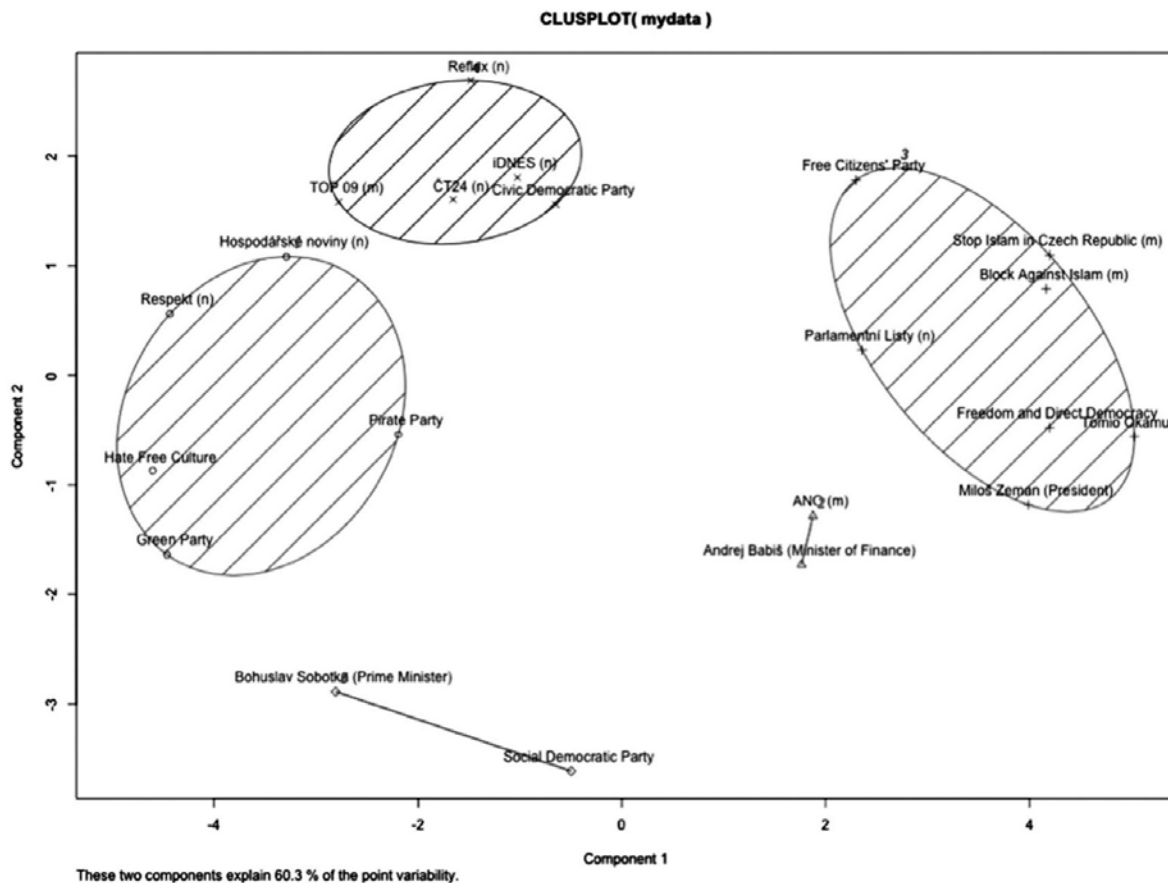
Data z výsledné matice lze vizualizovat a analyzovat celou řadou metod, z nichž jsme zvolili dvě. První je síť stránek.



Obr. 31. – Síťový graf blízkosti vybraných facebookových stránek na základě NSD lajkujících

V této vizualizaci tvoří stránky jednotlivé uzly a vazby mezi nimi jsou dány velmi silnou vazbou, tedy hodnotou rovnou nebo menší než 0,5.

Pro druhou vizualizaci jsme zvolili vizualizaci klastrování pomocí metody k-means clustering.



Obr. 32. – K-means clustering blízkosti vybraných facebookových stránek na základě NSD lajkujících

Metoda NSD nám umožňuje tzv. „distant reading“ sociálních sítí a uživatelů, kteří je lajkují. Toto čtení ukazuje strukturální aspekty, jež nemusejí být nutně viditelné na úrovni „close reading“ (tj. analýzy obsahu nebo rozhovorů).

V naší případové studii můžeme díky metodě NSD (jak ve vizualizaci grafů, tak v analýze clusterů) zjistit několik důležitých výsledků o anti-islámských a protiimigračních hnutích a o vzájemné blízkosti jejich lajkujících čtenářů, stejně jako o českých zpravodajských médiích a stránkách českých politických stran.

Na českém Facebooku výsledky ukazují několik úzce spojených clusterů stránek, jejichž lajkující čtenáři jsou si velmi blízko a sdílejí podobné postoje. Uživatelé v těchto clusterech lajkují a sdílejí podobný obsah a jen zřídka zasahují do jiných clusterů.

Mezi těmito seskupeními můžeme určit jeden cluster, který lze označit za anti-islámský, protiimigrační, nacionalistický a zaměřený proti Evropské unii. Tento cluster se skládá z uživatelů lajkujících stránky anti-islámských politických hnutí Blok proti islámu a Islám v České republice nechceme, protievropské a nacionální strany Svoboda a přímá demokracie a libertariánské/konzervativní Strany svobodných občanů. Všechny tyto strany používají silnou protiimigrační rétoriku. Důležité je, že tento cluster zahrnuje Miloše Zemana, prezidenta České republiky, který je znám silnými anti-islámskými a protiimigračními názory. Zpravodajský server Parlamentní listy, který rovněž v tomto clusteru hraje významnou roli, publikuje redakčně neupravené názory politiků a autorů z celého politického spektra, běžně se však pojí s konzervativními a nacionalistickými médii. Uživatelé lajkující výše uvedené stránky jsou dle NSD velmi těsně spojeni a vykazují výrazný překryv – spoléhají se na podobné nebo blízké zdroje zpráv a lajkují podobný obsah.

Další cluster zahrnuje uživatele lajkující na stránkách liberální/levicové Strany zelených, liberální Pirátské strany a antidiskriminační skupiny Hate Free Culture. Blízko k těmto uživatelům má i liberální týdeník Respekt a deník Hospodářské noviny. Stejně jako v předchozím clusteru se i tito uživatelé obvykle spoléhají na podobné nebo blízké zdroje zpráv a na Facebooku lajkují podobný obsah.

Na první pohled vypadá česká veřejná debata o migrační krizi velmi polarizovaně. Zpravodajské servery mají tendenci líčit českou společnost jako zcela rozdělenou do dvou táborů, což zhruba odpovídá výše uvedeným dvěma skupinám uživatelů českého Facebooku.¹⁵⁰ Stejně tak vnímá migrační krizi jako rozdělující téma mnoho politiků, kteří tuto otázku zařazují do svých programů a získávají tak důležité body.

Analýza NSD ukazuje, že ačkoli tyto dvě skupiny uživatelů jsou významné a velmi aktivní na jiných stránkách Facebooku, existují nejméně dva další podobně významné clustery, které

¹⁵⁰ ŠLERKA, Josef. Polarizovaná společnost? Nikoli, je to složitější. Magazín Reportér [online]. 1. 10. 2016 [cit. 2019-02-24]. Dostupné z: <https://reportermagazin.cz/a/i8sYJ/polarizovana-spolecnost--nikoli-je-to-slozitejsi>

primárně nesouvisejí s debatou o migrační krizi a jsou na zpravodajských serverech zřídka zmiňovány. Do těchto clusterů patří především uživatelé lajkující politické hnutí ANO a jeho vůdce Andreje Babiše (současného premiéra ČR). ANO je relativně nová politická strana, která byla často zobrazována jako protisystémová nebo populistická a jako alternativa ke starším „zkorumpovaným“ stranám. Strana má nejasně definovaný program a podporuje zejména „řádné“ technokratické řízení veřejných záležitostí.¹⁵¹ Další významnou skupinou, již se migrační debata netýká, jsou uživatelé lajkující Českou stranu sociálně demokratickou a bývalého premiéra Bohuslava Sobotku. Podstatný je fakt, že tyto dvě strany v současnosti tvoří vládu. Tyto clustery nevykazují žádnou významnou blízkost k jakýmkoli mediálním výstupům a nemají významné překrytí se dvěma aktivními clustery popsanými výše.

Z analýzy NSD tak vyplývá, že i když se veřejná debata o imigrační krizi zdá být výrazně polarizována do dvou kontrastních skupin, je přitom roztržštěná nejméně do čtyřech různých clusterů, jejichž lajkující uživatelé zřídka sdílejí stejný obsah a jejich postoje, vyjádřené lajky, se jen málokdy překrývají. Tato fragmentace negativně ovlivňuje veřejnou diskusi a v Habermasově terminologii¹⁵² je tak i možnost komunikativního jednání a kooperace značně omezená.

¹⁵¹ ŠLERKA, Josef. Polarizovaná společnost? Nikoli, je to složitější. *Magazín Reportér* [online]. 1. 10. 2016 [cit. 2019-02-24]. Dostupné z: <https://reportermagazin.cz/a/i8sYJ/polarizovana-spolecnost--nikoli-je-to-slozitejsi>

¹⁵² HABERMAS, Jürgen. *Strukturální přeměna veřejnosti: zkoumání jedné kategorie občanské společnosti*. Praha: Filosofia, 2000. Morální a politická filosofie. ISBN 80-7007-134-6.

4. Případové studie podporující validitu

V této kapitole budeme procházet tři různé případové studie zaměřené na podporu validity navržené metriky v rámci analýzy dat ze sociálních sítí.

První studie se věnuje konzistenci postojů uživatelů. Vycházíme z úvahy, že uživatelské postoje vyjádřené lajky by měly být konzistentní i s dalšími uživatelskými akcemi, jako jsou sdílení obsahu na své osobní zdi nebo stávání se fanoušky specifických stránek.

Východiskem pro nás bude analýza politicky silně polarizované mediální scény.

Druhá případová studie se zaměřuje na otázku, zda se politické postoje a preference uživatelů vyjádřené lajkováním postů jednotlivých stran odrážejí i v podobě výsledků samotných voleb, respektive v distribuci hlasů pro jednotlivé strany podle specifických oblastí.

Konečně pak **třetí případová studie** se věnuje otázce motivace sledování uživatelských účtů celebrit v prostředí Twitteru, Instagramu a Facebooku. Vychází primárně z myšlenky, že pokud je dominantní motivací pro sledování celebrit jejich osoba a nikoli jejich způsoby komunikace na jednotlivých kanálech, pak by jejich distanční matice měly vykazovat značnou podobnost.

4.1. Případová studie I. – Tzv. prokremelské weby na českém Facebooku¹⁵³

V únoru roku 2015 publikoval slovenský aktivista Juraj Smatana seznam webů, pro které se vžilo označení prokremelské dezinformační weby.¹⁵⁴ Tento seznam vzbudil mezi novináři rozsáhlou debatu a v průběhu let se stal víceméně východiskem pro další výzkumy.

¹⁵³ Tento výzkum proběhl v roce 2015 ve spolupráci s Lenkou Krsovou. Upravená podoba původní série blogpostů pak tvoří tuto případovou studii. Původní podoba je k dispozici ŠLERKA, Josef a Lenka KRISOVÁ. Tzv. proruské zpravodajské weby na Facebooku? In: Databoutique [online]. [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/113772071738/tzv-proruske-zpravodajske-weby-na-facebooku-1>

¹⁵⁴ SMATANA, Juraj. Neúplný, ale stále aktualizovaný zoznam webových stránok, ktorých linkovaním si koledujete o moju odbornú starostlivosť. In: Hadičky a udice [online]. [cit. 2019-02-23]. Dostupné z: <http://afinabul.blog.cz/1502/juraj-smatana-neuplny-ale-stale-aktualizovany-zoznam-webovych-stranok-ktorych-linkovanim-si-koledujete-o-moju-odbornu-starostlivosť>

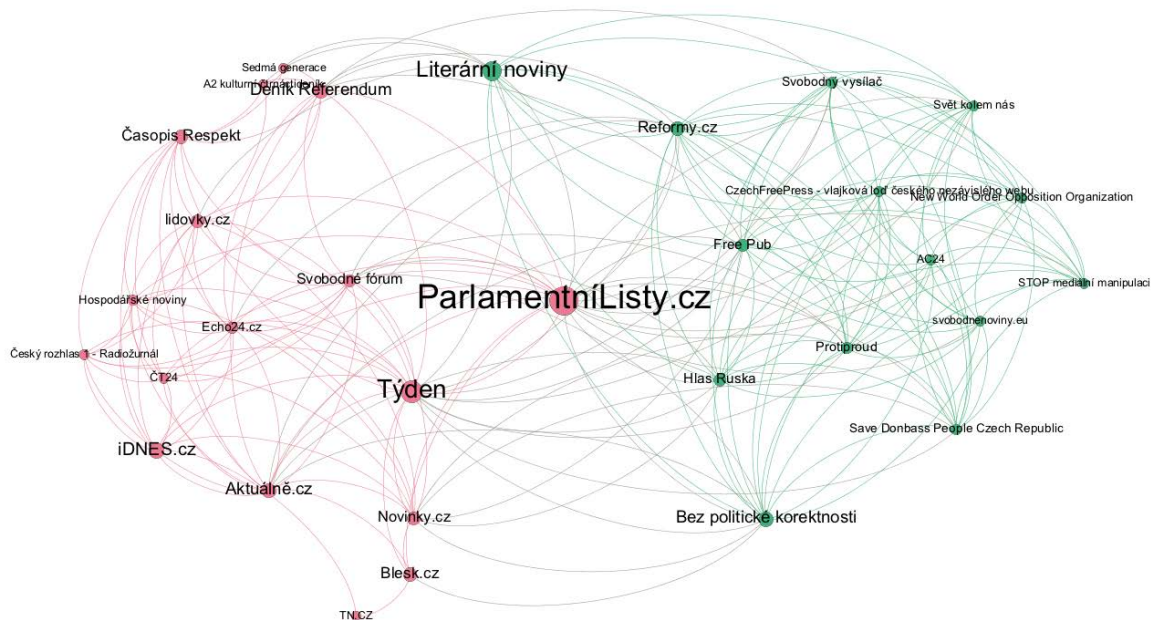
Protože velké procento z nich má na svých stránkách nějaký odkaz na svou nebo spřízněnou facebookovou stránku, rozhodli jsme se ověřit, zda lze nějak podpořit či naopak oslabit tvrzení, že dané servery skutečně patří do jednoho okruhu a že se jejich fanoušci aktivně podílejí na šíření informací z nich; jinými slovy validovat, zda expertní tvrzení má oporu v datech z naší analýzy.

Na základě předchozích kapitol předpokládáme, že tematicky podobné stránky přitahují podobné lidi. Tyto stránky by měly mít k sobě vzájemně blíže než ke stránkám mainstreamových zpravodajských serverů.

Pro ověření toho, jak moc jsou si blízké facebookové stránky webů uváděných v seznamech tzv. proruských webů, jsme zvolili seznam 30 facebookových stránek českých zpravodajských serverů doplněný o facebookové prezentace velkých zpravodajských serverů dle Netmonitoru a zohlednili také jejich celkovou velikost.

Z každé vybrané facebookové stránky jsme stáhli nejvýše 217 postů a k nim získali unikátní ID (identifikátory) jednotlivých lajkujících. Ty posloužily pro výpočet matice vzdáleností mezi stránkami pomocí metriky Normalized Social Distance.

Výslednou matici vzdáleností jsme se pro naše účely rozhodli vizualizovat a analyzovat pomocí několika exploračních metod. Nejdříve pomocí neorientovaného grafu, kdy vazbu mezi body vytváří normalizovaná vzdálenost menší než 0,55, vyjadřující velmi silnou afinitu (spřízněnost) mezi stránkami.



Obr. 33. – Síťový graf blízkosti vybraných facebookových stránek zpravodajských webů na základě NSD lajkujících

Pro samotnou vizualizaci jsme použili Force Atlas¹⁵⁵. Velikost uzlů (a textů popisků) pak vyjadřuje hodnotu tzv. betweenness uzlu. Zvýrazní se tak uzly (v našem případě reprezentují facebookové stránky), které tvoří mosty mezi částmi/klastry grafů. To jsou uzly, přes něž musíte projít, abyste se dostali z jedné strany na druhou, tak jak jsme o nich mluvili již v kapitole o social network analysis. Barvy uzlů a spojnic označují shluky bodů, mezi nimiž je více spojení než k ostatním. Zařazení uzlů do grafů zajistila strojová analýza pomocí algoritmu popsánoho Vincentem Blondem a dalším v článku *Fast unfolding of communities in large network*¹⁵⁶.

Zmíněné klastry lze ovšem vidět, i když stránky spojíme do shluků pomocí hierarchického klastrování.

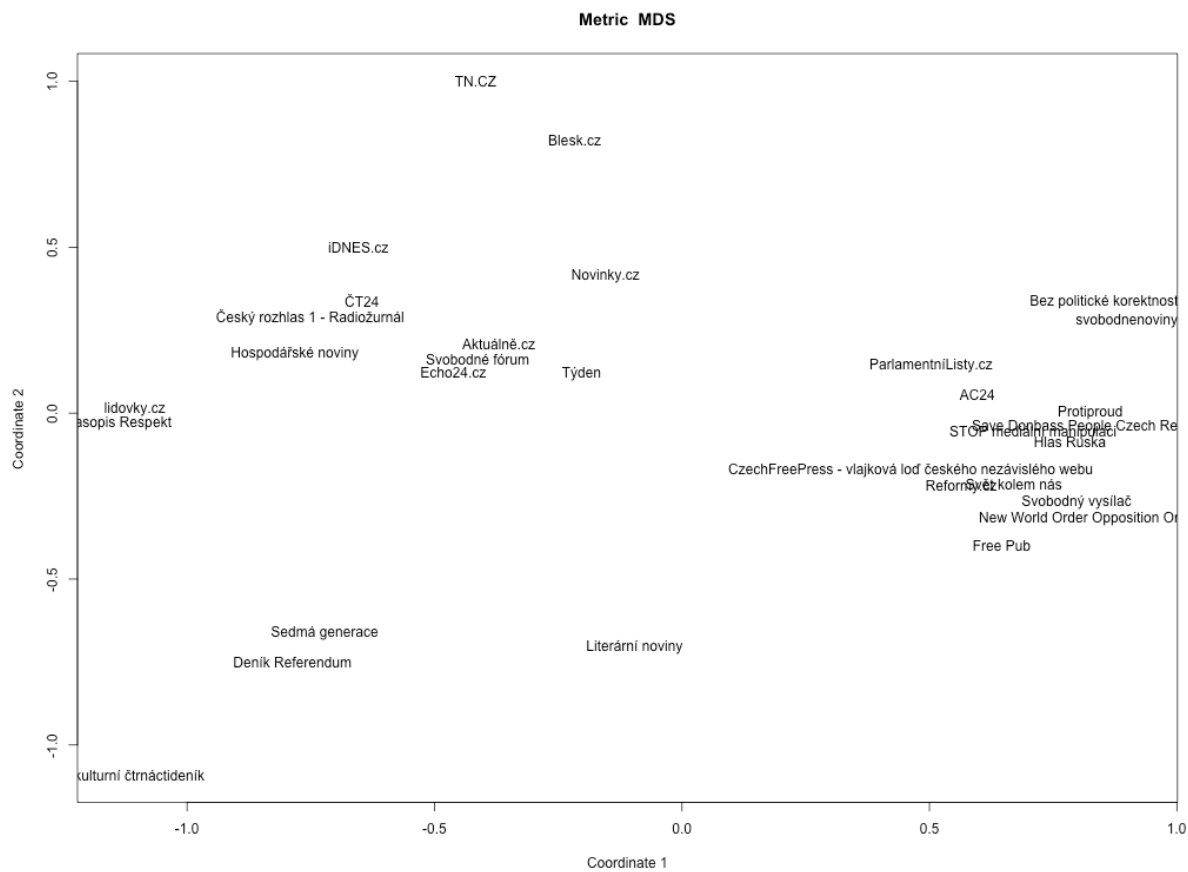
¹⁵⁵ Force-directed graph drawing. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-24]. Dostupné z: https://en.wikipedia.org/wiki/Force-directed_graph_drawing

¹⁵⁶ BLONDEL, Vincent D, Jean-Loup GUILLAUME, Renaud LAMBIOTTE a Etienne LEFEBVRE. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008, 2008(10). DOI: 10.1088/1742-5468/2008/10/P10008. ISSN 1742-5468. Dostupné také z: <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52>



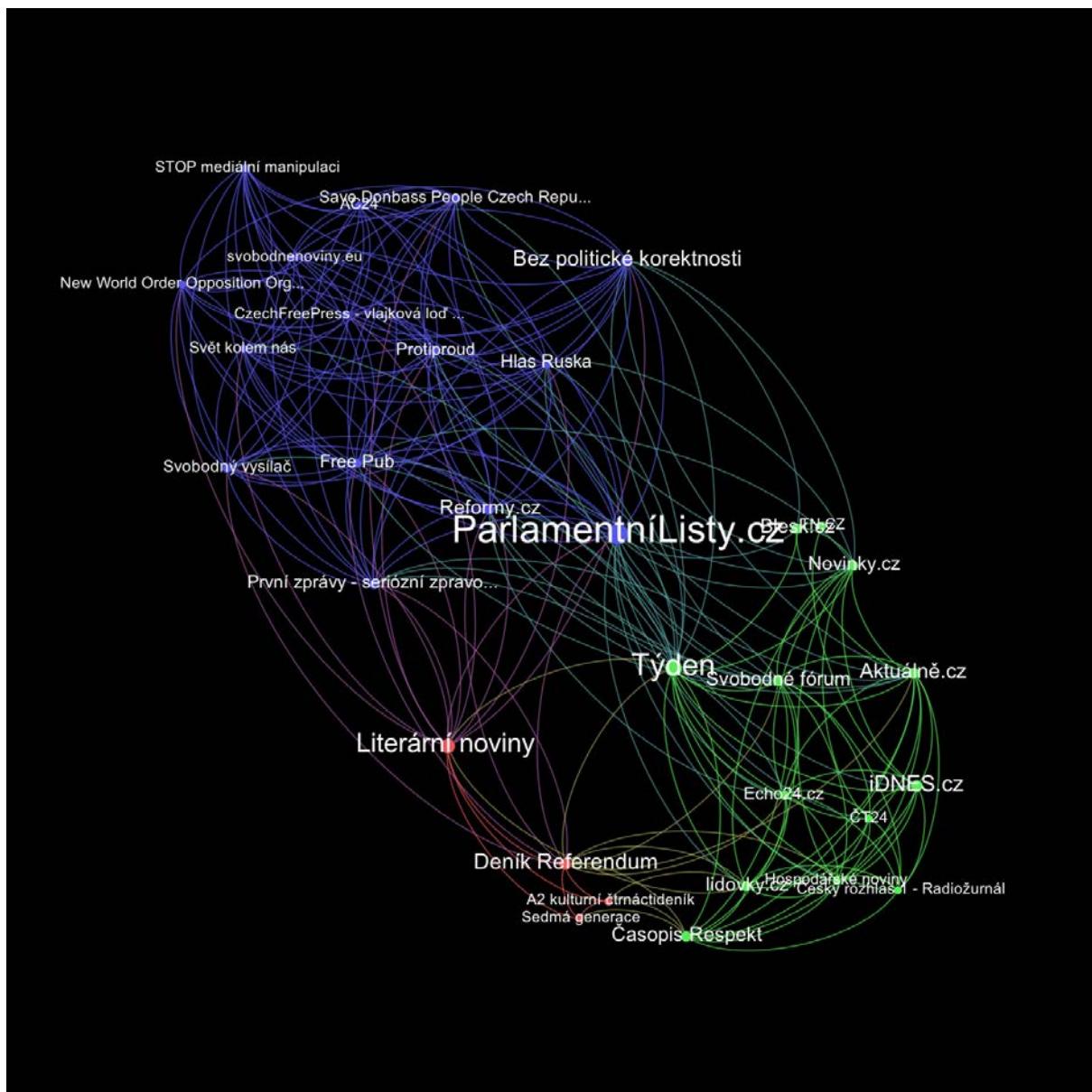
Obr. 34. – Dendrogram výsledků hierarchického klastrování na základě NSD lajkujících posty na zpravodajských stránkách

Stejně tak ilustrativní je vizualizace pomocí MDS, kde jsou vidět i dvě pomyslné osy. Osa X, kterou lze pojmenovat jako osu systémovosti x antisystémovosti (což se zdají potvrzovat další analýzy v této případové studii) a osu Y jako osu infotainmentu vs. politického zpravodajství.



Obr. 35 – Graf MDS výsledků distanční matice na NSD lajkujících posty na zpravodajských stránkách

Na základě zpětné vazby od expertů a novinářů jsme ještě tento seznam rozšířili o deset dalších zpravodajských i potenciálně prokremelských serverů a získali výslednou mapu, s níž jsme pracovali dál.



Obr. 36 – Výsledný síťový graf blízkosti vybraných facebookových stránek na základě NSD lajkujících zpravodajské stránky na Facebooku

Výsledná vizualizace pomocí sítě se stejnými parametry jako původní ukazuje existenci tří klastrů. Vnitřně nejsilněji provázaným je z nich shluk tvořený facebookovými stránkami převážně tzv. prokemelských serverů (v grafu modře/fialově) z původního seznamu. Zbylé dva odpovídají tzv. mainstreamovým (zeleně) a zhruba tomu, co lze označit jako alternativní média. Facebookové stránky Parlamentních listů, Prvních zpráv a Reforem se zdají pohybovat na hranici mezi dominantními klastry. Literární noviny, Parlamentní listy a Týden mají z celého grafu nejvyšší hodnotu betweenness uzlu. Specifické je zejména postavení Parlamentních listů, které se pohybují na pomezí dvou velkých světů a jejich zařazení mezi

nimi osciluje podle zvoleného typu klastrování. To znamená, že mají mezi ostatními uzly grafu zvláštní roli. Tyto tři servery figurují jako nejvytíženější uzly celé vizualizace. Pokud by se jednalo o dopravní mapu, sloužily by jako hlavní propojovací brány mezi třemi klastry, které se zobrazily.

Hypotéza o příslušnosti tzv. proruských serverů do jednoho klastru, zřetelně odlišného od ostatních, se zdá být pravděpodobná, a to alespoň z analýzy aktivit lidí, kteří se aktivně zapojují do interakce s jejich obsahem.

Co sdílejí aktivní uživatelé

Vraťme se ke goffmanovskému výchozímu bodu. Opakované lajkování postů na dané stránce mělo být příznakem silnějšího spojení s danými médii a jimi reprezentovaným okruhem zájmů. Pokud tom tak je, pak by se mělo toto spojení projevit i v tom, že aktivní fanoušci sdílejí obsah stránek, se kterými mají silnější spojení a který reprezentuje jim blízký postoj. Poměr mezi počtem unikátních uživatelů, kteří na stránce něco olajkovali, a počtem lajků na stránce se pohybuje mezi 1:2 až 1:4.

Facebook API nabízelo v roce 2015 možnosti strojově zpracovat odkazy, které lidé sdílejí na svém profilu, pokud nemají obsah svých „zdí“ nastavený jako soukromý. Rozhodli jsme se proto porovnat, z jakých serverů aktivně sdílejí fanoušci vybraných tzv. proruských stránek a vybraných stránek z mainstreamového klastru.

Analyzovali jsme 1 996 profilů lidí, kteří měli přes API Facebooku dostupný seznam odkazů sdílených na své osobní Zdi (FB wall) a alespoň pětkrát olajkovali nějaké posty následujících stránek patřících do mainstreamové oblasti:

- Aktuálně.cz (celkem analyzováno 217 postů)
- Blesk.cz (celkem analyzováno 217 postů)
- Časopis Respekt (celkem analyzováno 216 postů)
- Český rozhlas 1 – Radiožurnál (celkem analyzováno 218 postů)
- ČT24 (celkem analyzováno 217 postů)
- Hospodářské noviny (celkem analyzováno 217 postů)
- iDNES.cz (celkem analyzováno 217 postů)
- lidovky.cz (celkem analyzováno 218 postů)

- Novinky.cz (celkem analyzováno 217 postů)

U těchto lidí jsme vždy stáhli maximálně posledních 25 veřejně sdílených odkazů na jejich Zdi (jejich profily byly plně veřejné). Celkem šlo o 30 569 odkazů, které ukazovaly na 4 079 unikátních domén druhého a třetího řádu. Přehled top internetových domén, které měly alespoň cca 5 % unikátních sdílejících v našem datasetu, je následující:

doména	celkový počet odkazů z domény	unikátní počet sdílejících
www.facebook.com	13 178	1707
www.youtube.com	2627	903
www.novinky.cz	561	325
www.ceskatelevize.cz	512	335
www.rozhlas.cz	386	218
zpravy.idnes.cz	344	228
echo24.cz	275	159
cs.what-character-are-you.com	258	128
www.reflex.cz	227	170
respekt.ihned.cz	210	129
www.parlamentnilisty.cz	197	122
zpravy.aktualne.cz	189	137
tn.nova.cz	180	119
video.aktualne.cz	154	123
archiv.ihned.cz	142	123

www.lidovky.cz	115	92
----------------	-----	----

Tab. 16 – Přehled nejvíce zastoupených sdílených domén pro mainstreamové weby

Dále jsme připravili podobnou analýzu z okruhu tzv. prokremelských webů. Analyzovali jsme 1 488 profilů lidí, kteří měli přes API Facebooku dostupný seznam odkazů, které sdíleli na své osobní Zdi (i jejich profily, a tedy i analyzované odkazy byly plně veřejné) a zároveň alespoň pětkrát olajkovali nějaké posty následujících stránek:

- AC24 (celkem analyzováno 217 postů)
- Bez politické korektnosti (celkem analyzováno 217 postů)
- CzechFreePress – vlajková loď českého nezávislého webu (celkem analyzováno 217 postů)
- Free Pub (celkem analyzováno 217 postů)
- freeGlobe.cz – nový světový řád (celkem analyzováno 218 postů)
- Hlas Ruska (celkem analyzováno 217 postů)
- New World Order Opposition Organization (celkem analyzováno 216 postů)
- Občanský tribunál (celkem analyzováno 68 postů)
- Protiproud (celkem analyzováno 217 postů)
- První zprávy – seriózní zpravodajství (celkem analyzováno 217 postů)
- Reformy.cz (celkem analyzováno 217 postů)
- Save Donbass People Czech Republic (celkem analyzováno 217 postů)
- STOP mediální manipulaci (celkem analyzováno 217 postů)
- Svět kolem nás (celkem analyzováno 217 postů)
- svobodnenoviny.eu (celkem analyzováno 217 postů)
- Svobodný vysílač (celkem analyzováno 217 postů)
- Teplický KURÝR (celkem analyzováno 217 postů)
- Vlastní hlavou (celkem analyzováno 218 postů)
- Zvěděvec (celkem analyzováno 217 postů)

Také u těchto lidí jsme vždy stáhli nejvýše posledních 25 sdílených odkazů.

Celkem šlo o 23 132 postů, které ukazovaly na 2 655 unikátních domén druhého a třetího řádu. Přehled top internetových domén, které měly alespoň cca 5 % unikátních sdílejících v tomto datasetu, je následující:

Doména	celkový počet odkazů z domény	unikátní počet sdílejících
www.facebook.com	9.645	1301
www.youtube.com	1.979	697
www.parlamentnilisty.cz	521	287
www.novinky.cz	472	258
www.svetkolemnas.info	288	193
www.ac24.cz	244	164
www.hlavnespravy.sk	233	149
svobodnenoviny.eu	139	105
zpravy.idnes.cz	138	110
cs.what-character-are-you.com	128	81
www.ceskatelevize.cz	128	97
protiproud.parlamentnilisty.cz	126	83
www.bezpolitickekorektnosti.cz	125	87
www.nwoo.org	125	79
www.prvnizpravy.cz	125	90
aeronet.cz	124	88
czech.ruvr.ru	118	81

Tab. 17 – Přehled nejvíce sdílených zastoupených domén pro prokremelské weby

V obou skupinách je vidět průnik šesti top domén. Zcela bezpříznakově jsou to odkazy do samotného Facebooku a na server YouTube, které obecně na Facebooku patří k nejsdílenějším. Další sdílenou top doménou je server cs.what-character-are-you.com, který nabízí různé osobnostní testy na Facebooku. Ze zpravodajských kanálů jsou tu dva giganti: Novinky (nejnavštěvovanější zpravodajský server na českém internetu) a Česká televize (přibližně polovina přístupů byla na zpravodajství ČT24). Šestý společně sdílený server obou táborů jsou Parlamentní listy, které sice nedosahují takové návštěvnosti, ale vzhledem k výsledku předchozí analýzy se zdají být klíčovým mostem mezi oběma světy. Další sdílené odkazy z obou analyzovaných skupin uživatelů už míří většinou na servery patřící právě do shluků zpravodajských serverů analyzovaných v prvním blogpostu.

V případě tzv. proruských stránek tu dokonce zahlédneme v poslední době často zmiňovaný web Aeronet, který jinak vlastní facebookovou prezentaci nemá. Stejně tak narazíme na Hlas Ruska, který nebyl součástí výchozího seznamu. Zdá se tedy, že aktivně lajkující fanoušci zároveň aktivně sdílejí obsah serverů ze svých klastrů dle příslušnosti do svých klastrů, což potvrzuje představu konzistence postojů uživatelů

Kde ještě lajkují aktivní lajkeři?

Pro poslední analýzu jsme zvolili výpočet afinitních stránek, který jsme popsali v části věnované Goffmanově interpretaci sociálních interakcí. Předpokládali jsme, že politické orientace, zejména u politicky vyhraněného seskupení antisystémových webů, by se měly odrazit i v preferenci podobných stránek mimo zkoumaný dataset. Místo stažení stránek jsme ale využili databázi lajků popsanou v případové studii Analýza fanoušků DSSS.

Kde lajkují aktivní fanoušci mainstreamových zpravodajských serverů

Opět jsme vyšli ze vzorku aktivně lajkujících fanoušků (5 a více lajků) mainstreamových zpravodajských serverů. Celkem jsme analyzovali vzorek cca 2 000 uživatelů a do našeho seznamu zařadili jen stránky, u nichž něco olajkovalo alespoň 10 % uživatelů z našeho seznamu a vykazovaly afinitu 10x vyšší než v běžné populaci. Pro lepší přehled jsme stránky rozdělili do tří skupin: Zpravodajské stránky, Politické strany a jejich osobnosti a Ostatní.

Zpravodajské stránky a média

HNDialog, lidovky.cz, ParlamentníListy.cz, Týden, Český rozhlas 1 – Radiožurnál, Časopis Respekt, iHNed.cz, Aktuálně.cz, Echo24.cz, DVTV, Novinky.cz, Český rozhlas, ČT24, iDNES.cz.

Politické strany a jejich osobnosti

Občanská demokratická strana, Strana zelených, KDU-ČSL, ČSSD, TOP 09

Ostatní:

Znovu 89, Hyde Park, Žít Brno RIP, Klub přátel Člověka v tísní, Rekonstrukce státu, Tomáš Sedláček, Časopis dTest

Kde lajkují aktivní fanoušci tzv. proruských zpravodajských stránek

Opět jsme vyšli ze vzorku aktivně lajkujících fanoušků (5 a více lajků) ze vzorku tzv. proruských zpravodajských facebookových stránek z minulého blogpostu. Celkem jsme analyzovali vzorek 900 uživatelů a do našeho seznamu zařadili jen stránky, u nichž něco olajkovalo alespoň 10 % uživatelů z našeho seznamu a vykazovaly afinitu 10x vyšší než v běžné populaci. Pro lepší přehled jsme stránky opět rozdělili do tří skupin: Zpravodajské stránky, Politické strany a jejich osobnosti a Ostatní.

Zpravodajské stránky

CzechFreePress – vlajková loď českého nezávislého webu, Protiproud, Hlas Ruska, Free Pub, Svobodný vysílač, Informace bez hranic, Nazorobcana, Save Donbass People Czech Republic, Bez politické korektnosti, ParlamentníListy.cz, Reforms.cz, New World Order Opposition Organization, Zakázané informace, Hlavné Správy, Zpravodajská Alternativa, Zpravodajství bez cenzury, Svět kolem nás, ZEM A VEK, Pirátské Noviny, AC24, Slobodný vysílač

Politické strany a jejich osobnosti

NE Bruselu – Národní demokracie, O2H – Hnutie za Reálnu Demokraciu, Miloš Zeman – prezident České republiky, Hnutí Úsvit

Ostatní

Humanistická Iniciativa proti válce, Ruce pryč od Sýrie, Ukrajinské nepokoje, Přátelé Novoruska, Kalouskovo kníže – Knížetovo Kalousek /česká fraška, Illumináti v ČR a SR , Alexander Ivanovič Možajev, PRAVDA o TERORISMU, Utajované pravdy a oficiální lži, STOP mediální manipulaci, Chemtrails Czech Republic, Nechceme líbyjský scénár v Sýrii, Konec Kalouska v Čechách, Politici a politické strany ČR – Odstupte !!!, RADIO Jerevan, Holešovská výzva, TV BEZ CENZURY, Slované Славяне Słowianie, Svetová revolúcia, Stydím se za vládu ČR a chci její demisi, PRAVDA ŤA OSLOBODÍ! , Náboženský extrémismus ::: ->> Odpor proti mešitám a náboženskému Islámu., ProtiSystém, Celostnimedicina.cz, Odcházíme ze systému, Pro-Vlast, Za Slobodu Slova, Islám v České republice nechceme

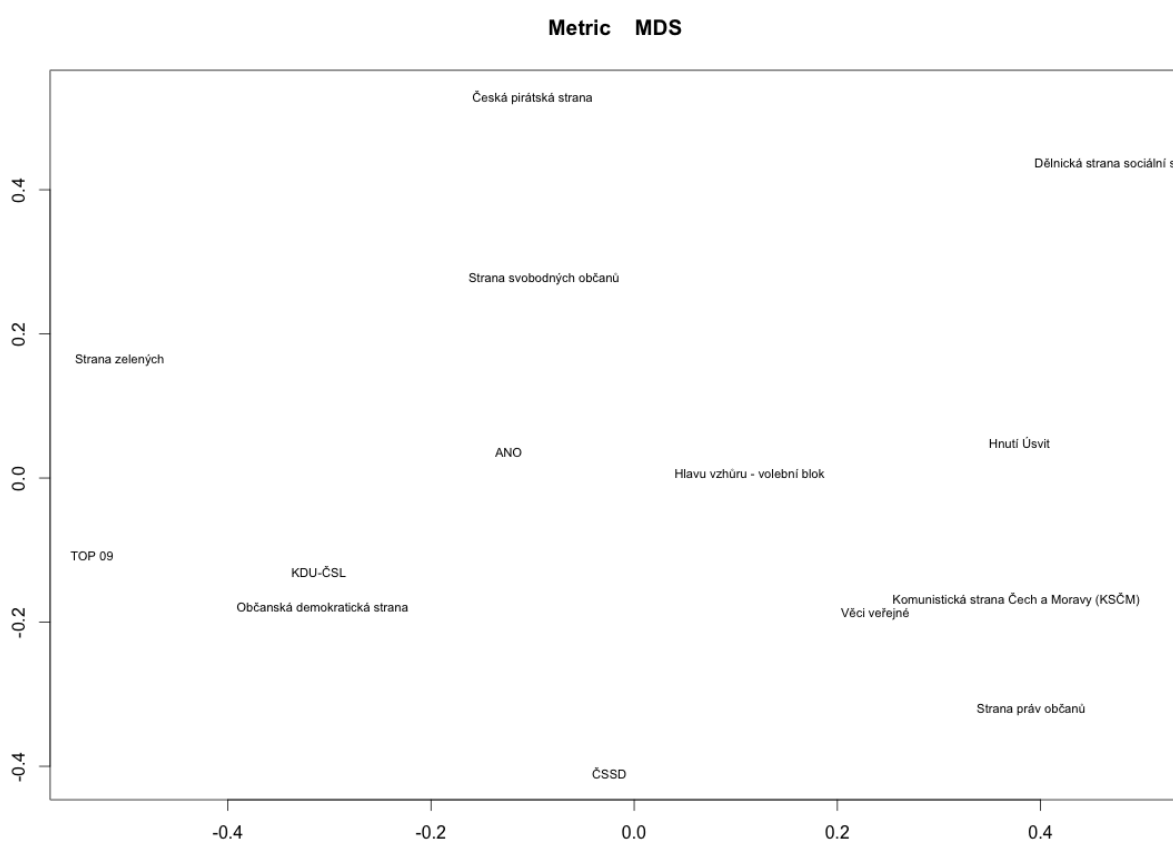
Výsledky u obou skupin opět ukazují silnou afinitu nejen ke zpravodajství, ale k politicky orientovaným tématům obecně. U fanoušků tzv. proruských serverů vidíme nejen vysoký podíl stránek věnovaných konfliktu na Ukrajině, ale také vysoký podíl stránek, které obecně vyjadřují radikální nespokojenost s politikou, a to jak domácí, tak zahraniční. Skladba stránek u obou skupin fanoušků naznačuje, že se nejedná o jeden homogenní celek, pro který by se hodilo tak úzké vymezení, kterým je označení pro- nebo protiruský. U fanoušků tzv. proruských serverů vidíme i silnou inklinaci k politickým subjektům a projektům, které je možné označit za antisystémové. Tedy k takovým, které neusilují o reformu v rámci stávajícího systému, ale o samotnou změnu podstaty tohoto systému. Výcházeli jsme přitom z analogie k antisystemovým stranám, které vymezuje politolog Sartori jako strany které „nesdílí hodnoty politického řádu, v němž operují.“¹⁵⁷

¹⁵⁷ SARTORI, Giovanni. Strany a stranické systémy: schéma pro analýzu. Brno: Centrum pro studium demokracie a kultury, 2005. Klasikové společenských věd. ISBN 80-732-5062-4. str. 138

4.2. Případová studie II. – Politické mapy a volební výsledky¹⁵⁸

Jak verifikovat, že existuje nějaký přesah mezi vhledy získanými z dat ze sociálních sítí a offline skutečnosti jinak, než prostřednictvím intuice? Kromě běžných metod matematického testování stability apod. se nabídla i nečekaná možnost v podobě výsledků voleb do Evropského parlamentu v roce 2014.

Při podzimních volbách 2013 vyšla Mapa fanoušků ze sledovaného období čtyř týdnů (12. srpna až 12. září 2013) takto:



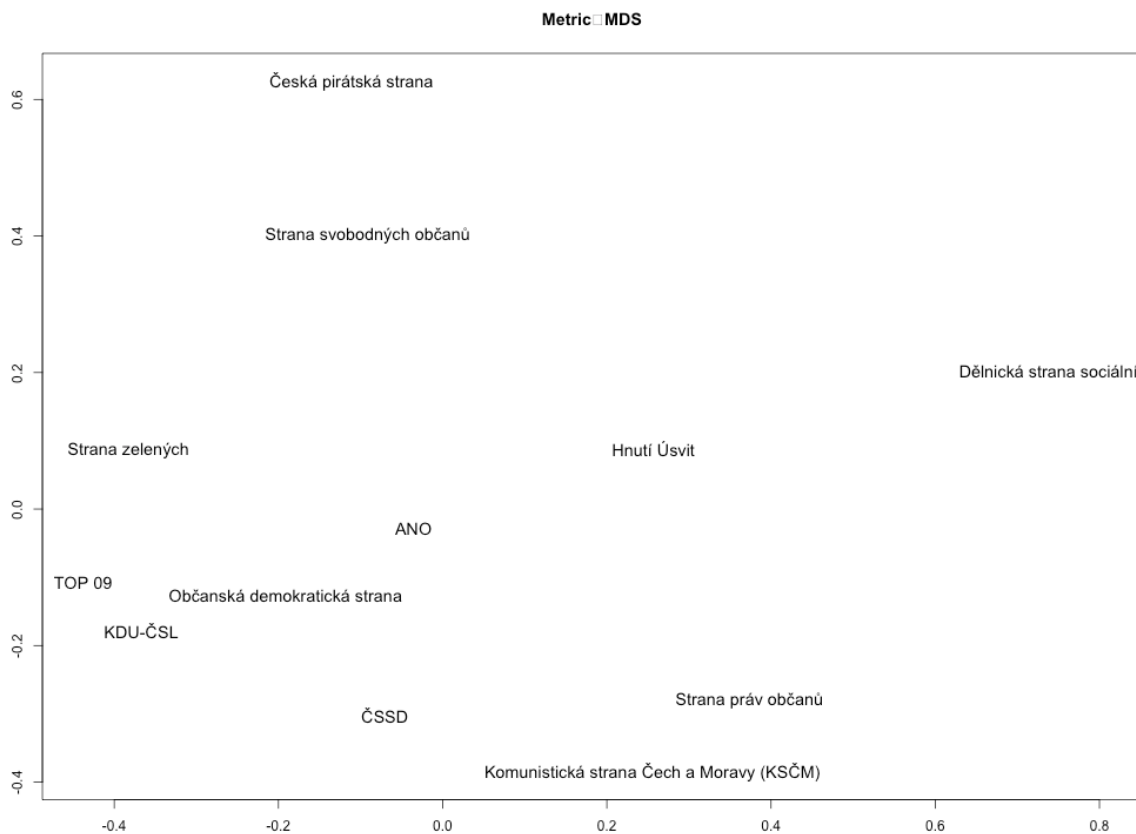
Obr. 37 – Graf MDS blízkosti českých politických stran na základě NSD lajkujících, podzim 2013

Tehdy autor této disertace spolu s Janem Schmidem navrhoval interpretovat její dvě dimenze jako dimenze věku (vertikální osa) a vzdělání (horizontální osa). Tedy kupříkladu mladší a

¹⁵⁸ Původní verze této případové studie byla publikována na blogu Databoutique. ŠLERKA, Josef. Korelace politických stran na Facebooku s výsledky voleb do Evropského parlamentu. In: *Databoutique* [online]. [cit. 2019-02-24]. Dostupné z: <http://databoutique.cz/post/87676505248/korelace-politick%C3%BDch-stran-na-facebooku-s-v%C3%BDsledky>

méně vzdělání voliči DSSS, spíše středního věku a vyššího vzdělání voliči TOP 09 atd. I když navrhovanou interpretaci ex-post nepřímo podpořily sociologické výzkumy voličů, nenapadlo nás, jak adekvátnost zachycených vztahů ověřit v offline světě. Zatímco na Facebooku mapa vizualizuje skutečně probíhající interakce, není a nebylo jisté, jak se má k světu mimo sociální síť. Ostatně tato naprosto fundamentální otázka vztahu sociálních sítí a reality visí prakticky nad každým výzkumem sociálních sítí a bývá většinou nezodpovězena.

Mapa vypočtená pouhých šest týdnů před volbami do Evropského parlamentu v roce 2014 vykazovala minimální změny v rozložení oproti mapě z předchozího roku:



Obr. 38 – Graf MDS blízkosti českých politických stran na základě NSD lajkujících, jaro 2014

Stabilita rozložení může znamenat, že MDS mapa skutečně zaznamenává stabilní rozdíly mezi stránkami na Facebooku. Důležité ale také je, kolik aktivních uživatelů zůstalo z podzimu 2013 do měření na jaře 2014. Neznamená totiž stabilita výsledků pouze to, že na Facebooku jsou aktivní totožní jedinci? Při analýze konkrétních ID lajkujících jsme zjistili, že překryv aktivních uživatelů téže strany z prvního a z druhého měření byl obvykle kolem 30

% . Jedná se tedy o indikátor toho, že model vyjadřuje spíše obecnější hodnotové rozdíly mezi stranami, než závislost na konkrétním seznamu uživatelů.

Pro Českou televizi jsem připravil celkem deset takovýchto map, z nichž některé měli diváci možnosti vidět v příjemnější grafice, než je surový výstup z nástroje, který pro analýzu používáme (statistický programovací jazyk R). Po jejich prezentaci se rozvinula krátká debata o relevantnosti takovýchto map mimo online prostředí. Mohou reprezentovat nějaké skutečné rozložení politických sil, zejména se zřetelem k tomu, že demografie facebookové populace je oproti populaci České republiky vychýlená? Tato otázka se vynořuje od prvních experimentů s touto metodou z času prezidentských voleb (viz Mapa blízkosti prezidentských kandidátů). Mapy blízkosti se sice osvědčují jako nástroj marketingového výzkumu v oblasti sociálních médií, ale jejich extenze mimo virtuální prostředí byla vždy špatně ověřitelná.

Po volbách do Evropského parlamentu se ale objevila možnost, jak ověřit, zda vzdálenosti měří nějakou reálnou veličinu. Stojí na relativně prosté, byť v některých aspektech kontrainuitivní úvaze a využívá korelaci volebních výsledků stran ve skutečných volbách.

Zjednodušený model vypadá takto: český volební systém poměrného zastoupení s relativně vysokým počtem stran způsobuje, že se strany ucházejí často o podobné voliče či voliče z podobných skupin.

Kupříkladu hypotetické strany Protestní radikální strana A (PRSA) a Méně radikální protestní strana B (MRPSB) cílí na silně nespokojené voliče, kteří preferují jednodušší řešení a mají nižší vzdělání. Takových voličů je ve společnosti určité procento. Ve volbách ho vyjádříme jedním číslem, procentuálním výsledkem pro tento typ protestních stran. Intuitivně vzato by měla být korelace procentních hlasů mezi oběma stranami negativní. Zisky pro PRSA jsou ztrátami pro MRPSB. Jenže strany obvykle nezískávají náhodné výsledky z celkového podílu voličů. Zpravidla bývá jedna úspěšnější při lovení nespokojených než druhá. PRSA získá celostátně třeba 14 % a MRPSB třeba 6 %. Důležité je, že jejich individuální výsledky jsou obvykle s nízkou mírou směrodatné odchylky a oslovené skupiny se ve volebních okrscích nevyskytují rovnoměrně. Díky tomu existuje vysoká pravděpodobnost pozitivní korelace výsledků obou stran v okrscích.

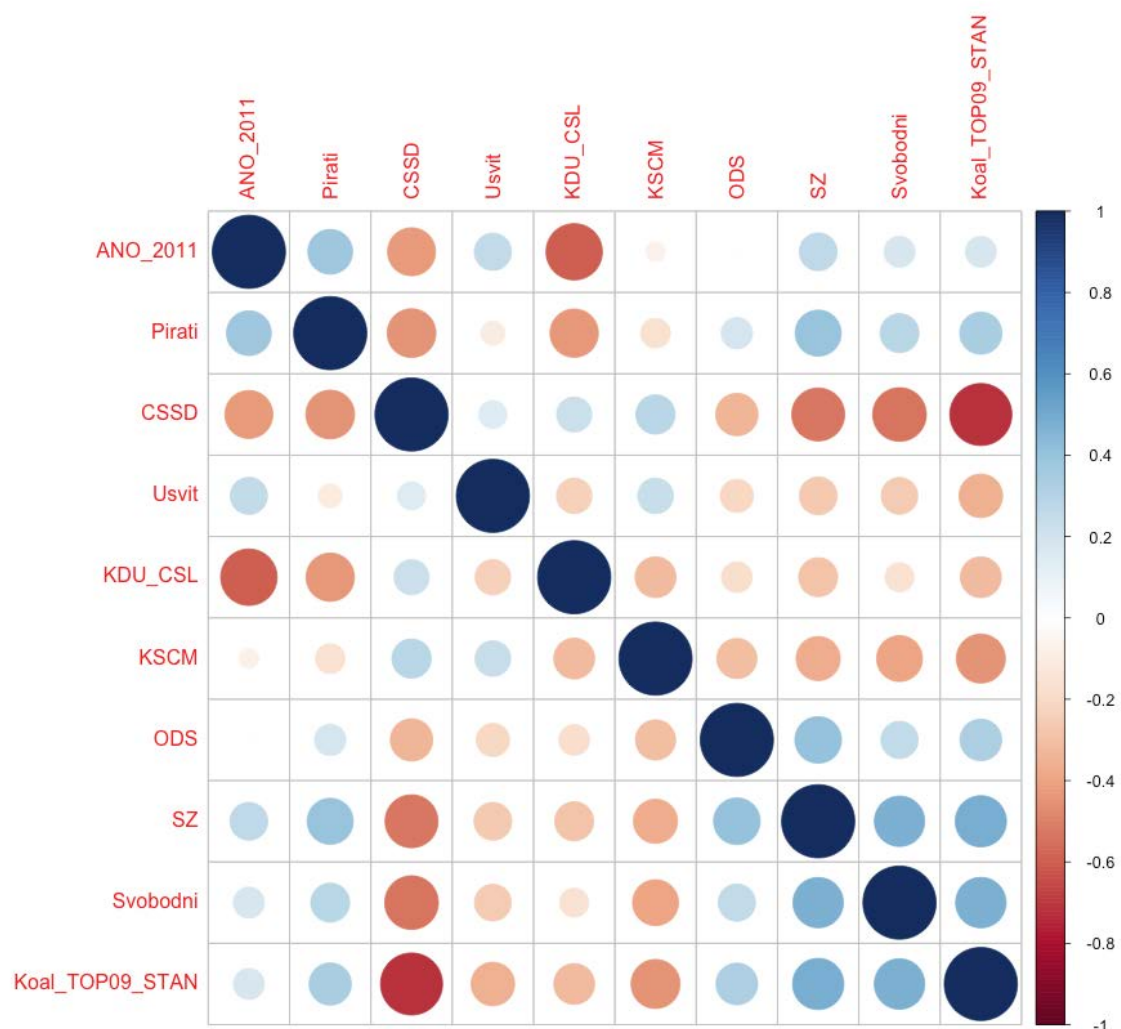
Voliči samotní jsou sice sloučeni do volebních okrsků, ale okrsky samotné většinou představují poměrně koherentní skupinu voličů. O tom svědčí už i pouhý pohled na volební mapu České republiky, která nemá jednolitou barvu nejúspěšnější strany, ale je naopak poměrně barevně pestrá.

Jaké korelace se objevují v praxi? Kupříkladu při podzimních volbách 2013 obvykle platilo, že výsledky KSČM a Úsvitu spolu silně pozitivně korelovaly napříč volebními obvody s alespoň 5 000 voliči. Pokud byly v daném volebním obvodu výsledky dobré pro komunisty, zpravidla byly dobré pro Okamuru a naopak¹⁵⁹. Vysvětlení v tomto případě zřejmě může být, že Okamura i komunisté oslovují především nespokojené voliče a získávají tak i podobné výsledky. Dalším příkladem je TOP 09 a Strana zelených, které cílí na podobné voliče: vzdělanější, bohatší, z měst.

Pokud facebookové mapy blízkosti skutečně odrážejí realitu podobnosti stran, měly by korelovat se vzájemnou korelací výsledků stran z voleb.

Na vstupu máme tedy dvě matice. Jednu se vzájemnou vzdáleností stránek politických stran na Facebooku a druhou s korelacemi stran na základě volebních výsledků v okrscích nad 5 000 obyvatel, která vypadá následovně:

¹⁵⁹ ŠLERKA, Josef. Housle a mřížky z voleb do parlamentu. In: Databoutique [online]. [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/66175359703/housle-a-mrizky-z-voleb-do-parlamentu>).

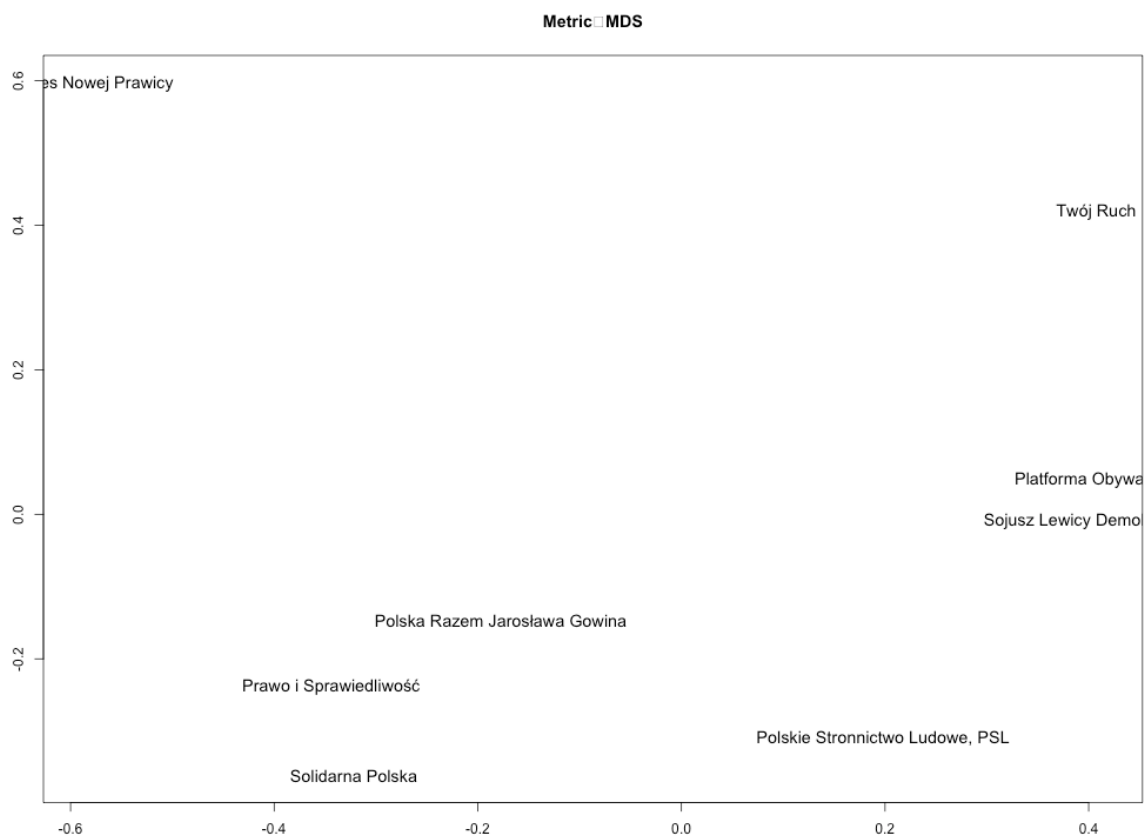


Obr. 39 – Korelační matice volebních výsledků českých politických stran z parlamentních voleb 2013

Čím lépe odráží mapa z Facebooku reálný svět, tím by měla být korelace s korelační maticí z reálných voleb nižší (maximálně provázaný vztah vyjadřuje hodnota -1). Hodnota 0 vyjadřuje naprostou nezávislost a data z Facebooku by pak neměla žádný vztah k offline světu. V sociálních vědách korelace menší než -0,2 obvykle poukazuje na existující vztah. A skutečně: pro Českou republiku vyšla korelace -0,71, což je na sociální vědy číslo nebývale dobré. (Korelace mají zápornou hodnotu, protože vztahy stran na Facebooku jsou vyjádřené jako blízkost – čím nižší, tím bližší, zatímco vztahy na základě volebních dat jsou vyjádřeny jako korelace – čím vyšší, tím bližší).

Zkusili jsme za podobných podmínek dopočítat korelace z podobných celků Polsko, Německo a Slovensko s rovněž překvapivě dobrými a podobnými výsledky.

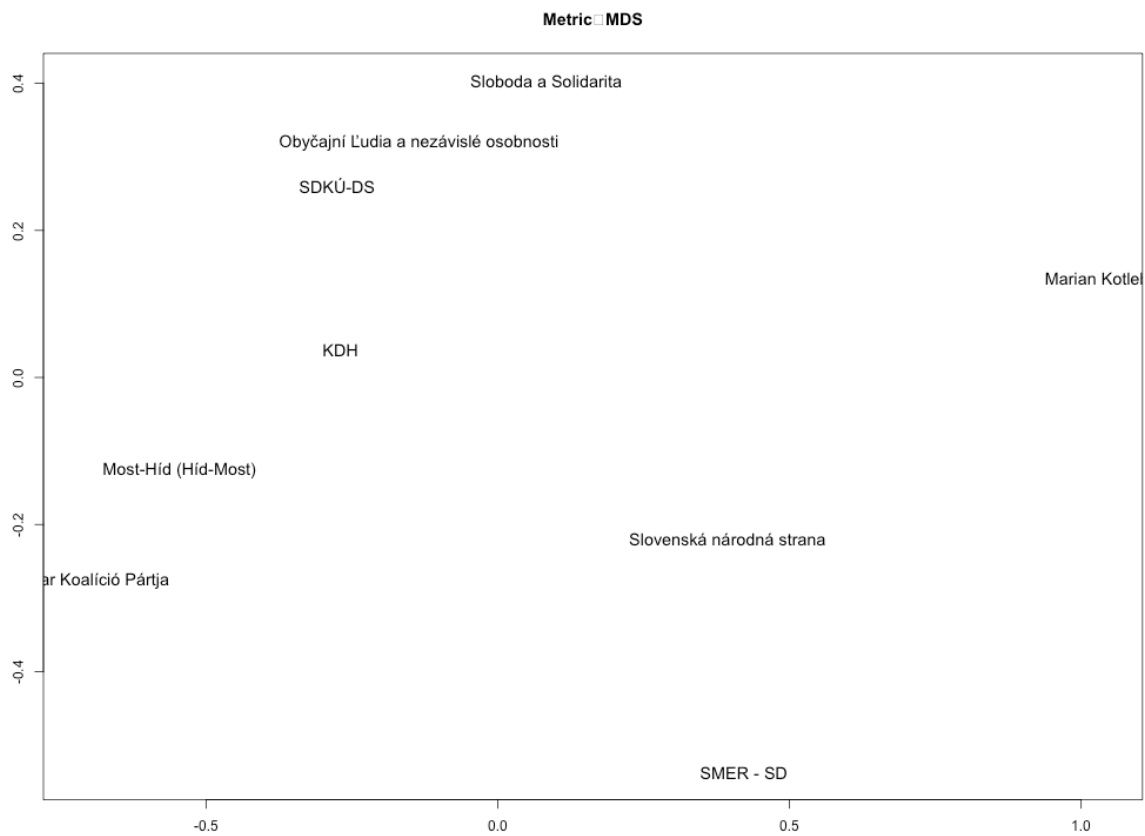
Pro mapu polské politické scény:



Obr. 40 – Graf MDS blízkosti polských politických stran na základě NSD lajkujících, jaro 2014

korelace -0,79 (počítáno na výsledky z krajů).

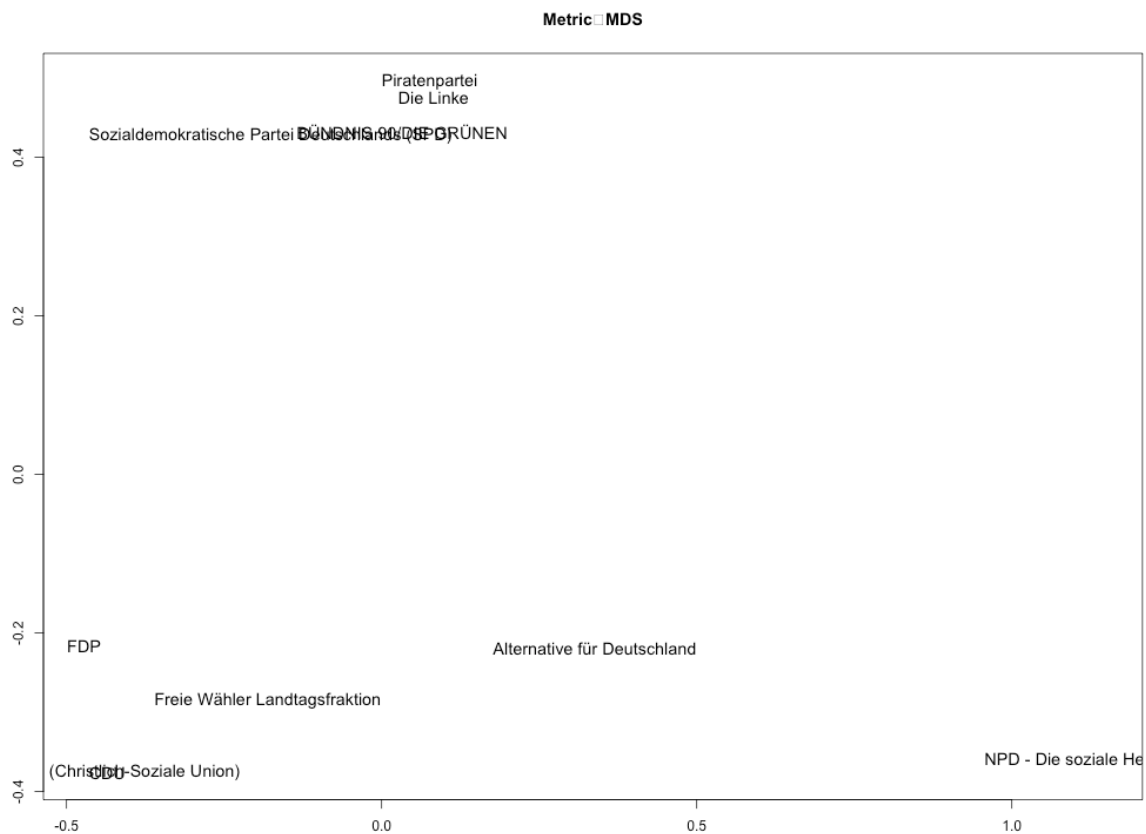
Pro mapu slovenské politické scény:



Obr. 41 – Graf MDS blízkosti slovenských politických stran na základě NSD lajkujících, jaro 2014

je to korelace -0,67.

A konečně pro Německo a jeho politickou scénu



Obr. 42 – Graf MDS blízkosti německých politických stran na základě NSD lajkujících jaro 2014

to byla korelace -0,71.

Spočetli jsme analýzu pro všechny země, k jejichž volebním datům se nám podařilo získat přístup, a ve všech případech jsme naměřili poměrně úzkou shodu. Problém ovšem je, že nevíme zcela přesně, jak silná tato korelace vlastně je, protože jsme museli kombinovat distanční a korelační matice. Tomuto problému se pokusíme vyhnout v příští případové studii.

4.3. Případová studie III. – Celebrity napříč sociálními sítěmi

Pokud jsou motivace pro to být fanouškem nějaké celebrity stejné, bez ohledu na to, zda jsme takovým fanouškem na Facebooku, Twitteru nebo Instagramu, pak by měly být matice vzdálenosti mezi jednotlivými celebritami, spočítané na základě podobnosti jejího publika, podobné na všech sociálních sítích.

Pro účely ověření této hypotézy jsme vytipovali skoro 30 účtů postav českého showbyznysu a youtuberů. Jednalo se o tyto účty: AtiShow, Ben Cristovao, Ektor, Gamekeepers_cz, GetTheLouk, Johnny Valda, Johny Machette, MadBros Channel, Makyna016, MenT, HouseBox, Šimon Vojta, Teri Blitzen, Theteengirllooks, VADAK, Vlada Videos, Vladimír 518, DJ Wich, Gabrielle Hecl, Hoggy, Jirka Král, ViralBrothers, Nejfake, Nici Koderová, Paulie Garand, PedrosGame, Sharlota, Stejk studio. Za období 1. 10. 2015 až 31. 1. 2016 jsme stáhli jejich posty z Facebooku, seznam všech followerů jejich twitterových účtů a seznam všech jejich followerů na Instagramu.

Následně jsme vypočetli pro každou ze sítí zvlášť matice vzájemných vzdáleností mezi účty a na výsledné tři matice pak aplikovali Mantelův test, což je metoda sloužící k výpočtu korelace mezi dvěma nebo více distančními maticemi mající svůj původ v biostatistice.¹⁶⁰ Pro naše účely představuje ideální metodu, jak ověřit podobnosti distančních matic mezi jednotlivými sítěmi.

Výsledky

Mantel test	Facebook	Instagram	Twitter
Facebook	1	0,83	0,8
Instagram	0,83	1	0,84
Twitter	0,8	0,84	1

Tab. 18 – Výsledky Mantel testu

¹⁶⁰ Mantel test. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-24]. Dostupné z: https://en.wikipedia.org/wiki/Mantel_test

V tabulce s výsledky můžeme vidět mimořádně vysokou korelaci mezi jednotlivými maticemi a zdá se tak, že motivace pro sledování jednotlivých celebrit a rozdílu mezi nimi jsou silnější než komunikace v jednotlivých médiích. Samozřejmě nevíme, a prakticky nemůžeme vědět, jak velký je skutečný průnik fanoušků mezi jednotlivými sítěmi. Do budoucna by pak bylo zajímavé soustředit se více na rozdíly mezi skupinami fanoušků mezi jednotlivými sítěmi a příčinu těchto rozdílů.

5. Závěr

Dostáváme se nyní na konec této práce. V ní je navržena transpozice metriky pro výpočet informační vzdálenosti z prostředí syntaktické a sémantické blízkosti do prostoru sociálních sítí a na sérii případových studií proběhl pokus ukázat oprávněnost takového přístupu. Samotná možnost tohoto posunu byla umožněna existencí širší změny, totiž nástupem nových médií na jedné straně a na druhé straně nástupem digital humanities, které se de facto staly výrazem měnícího se paradigmatu v humanitních a sociálních vědách. Během uplynulých let aplikoval autor této disertace metodologii uvedenou v práci v řadě komerčních projektů a publikoval o ní studie a blogposty. Dnes je již tedy možné setkat se s aplikací metriky Normalized Social Distance i v jiných pracích, jako kupříkladu v diplomové práci Dalibora Boboka pod názvem *Selective Exposure, Filter Bubbles and Echo Chambers on Facebook* z maďarské Central European University.¹⁶¹ Dokonce ani fakt, že některé sociální sítě omezují jednoduchý strojový přístup k potřebným datům, nepřinesl zásadní omezení pro aplikaci NSD. Jako příklad může posloužit blogpost Gabriela Totha¹⁶² z února 2018, kdy pro výpočet NSD kandidátů na slovenského prezidenta provedl celou analýzu na základě ručně vydolovaných dat přímo z facebookových stránek. Zdá se tedy, že navržený přístup je solidním základem pro další výzkumy.

V první řadě se nabízí jeho propojení se sociologickou koncepcí sociální vzdálenosti (anglicky social distance), která svůj původ odvozuje již od Simmelových analýz postupující urbanizace města a s ním i vzniku množství separovaných skupin v něm. Sociální vzdálenost není ovšem vzdáleností prostorovou (byť s ní může souviset). Snaží se vyjádřit míru pocitu známosti či blízkosti nějaké jiné skupiny.¹⁶³

Nejnámějším průkopníkem tohoto přístupu v kvantitativním výzkumu byl americký sociolog Emory S. Bogardus, který vyvinul tzv. Bogardusovu škálu sociální distance, zahrnující sedm základních otázek, které se týkaly různých etnických skupin v USA a

¹⁶¹ BOBOK, Dalibor. *Selective Exposure, Filter Bubbles and Echo Chambers on Facebook*: Diplomová práce [online]. Budapest, 2016 [cit. 2019-02-24]. Dostupné z: www.etd.ceu.hu/2016/bobok_dalibor.pdf

¹⁶² TÓTH, Gabriel. Velká analýza blízkosti podporovatelův kandidátů na prezidenta. In: Katedra komunikácie [online]. 4. 2. 2019 [cit. 2019-02-24]. Dostupné z: <https://katedrakomunikacie.sk/velka-analyza-blizkosti-kandidatov-na-prezidenta/>

¹⁶³ HODGETTS, Darrin a Otilie STOLTE. Social Distance. In: *Encyclopedia of Critical Psychology*. New York, NY: Springer New York, 2014, 2014, s. 1776–1778. DOI: 10.1007/978-1-4614-5583-7_559. ISBN 978-1-4614-5582-0. Dostupné také z: http://link.springer.com/10.1007/978-1-4614-5583-7_559

vyjadřovaly ochotu tazatele přijmout příslušníka jiného etnika do dané skupiny.¹⁶⁴ Jednalo se o tyto možnosti:

- 1) do blízkého příbuzenstva skrze manželství;
- 2) do svého klubu jako blízkého přítele;
- 3) za souseda v ulici;
- 4) do svého zaměstnání za spolupracovníka;
- 5) jako občana své země;
- 6) jen jako návštěvníka své země;
- 7) vyloučil by jej ze země¹⁶⁵.

Ve dvacátých letech Bogardus publikoval řadu výzkumů, které vycházely z této základní koncepce a snažily se fakticky vyjádřit sociální vzdálenost pomocí statistické analýzy odpovědí.¹⁶⁶

S celým konceptem Bogardusových výzkumů se pracuje dodnes a je dále rozvíjen. U nás můžeme odkázat kupříkladu na sborník editovaný Jiřím Šafrem pod názvem *Sociální distance: Interakce, relace a kategorizace*.¹⁶⁷ Nedim Karakayali¹⁶⁸ rozlišuje čtyři základní dimenze sociální distance. Původní Bogardusova je označena jako afektivní distance. Ta je doplněna konceptem normativní distance, tedy takové, která vychází z normativní příslušnosti do různých skupin. Dále pak sociální distance jako interaktivní distance, tedy distance postavená na míře interakcí mezi příslušníky skupin, a konečně pak kulturní a zvyková distance, které je odvozena od míry vzájemné imitace mezi skupinami. Poslední jmenovaná našla svou transformaci v teorii sociálního prostoru Pierra Bourdieua.

¹⁶⁴ Evoluci původního dotazníku s 60 otázkami do redukované podoby i se zdůvodněním popsal Bogardus v článku BOGARDUS, Emory S. A Social Distance Scale. *Sociology and Social Research*. 1933, (17), 265–271.

¹⁶⁵ Překlad znění otázek přejímám ze studie RYŠAVÝ, DAN. Sociální distance vůči Romům: Případ vysokoškolských studentů. *Sociologický časopis*. Praha, 2003, 39(1), 55–77.

¹⁶⁶ Viz kupříkladu Emory S. Bogardus. „Measuring Social Distances.“ *Journal of Applied Sociology* 9 (1925): 299–308.

¹⁶⁷ ŠAFR, Jiří, ed. *Sociální distance, interakce, relace a kategorizace: alternativní teoretické perspektivy studia sociální stratifikace*. Praha: Sociologický ústav Akademie věd České republiky, 2008. ISBN 978-80-7330-146-0.

¹⁶⁸ KARAKAYALI, Nedim. Social Distance and Affective Orientations. *Sociological Forum*. 2009, 24(3), 538–562. DOI: 10.1111/j.1573-7861.2009.01119.x. ISSN 08848971. Dostupné také z: <http://doi.wiley.com/10.1111/j.1573-7861.2009.01119.x>

Mnou navrhovaná koncepce Normalized Social Distance je s tímto přístupem v zásadě kompatibilní. Základní myšlenku sociální vzdálenosti ovšem aplikuje na model vzdálenosti odvozený od teorie informací, který umožňuje počítat vzdálenosti mezi skupinami nikoli jako výzkumníkem danými jednotkami, ale jako mezi různými druhy přijímaných či schvalovaných postojů. Důležité ovšem je, že při výpočtech vzdálenosti nevychází z expertně vytvořených okruhů, ale naopak postupuje odspoda nahoru. Na jedné straně využívá pestrost nabídky sociálních sítí a na druhé straně přistupuje k výzkumu neinvazivně, bez jakýchkoli dotazníků a podobně.

Přesto se nabízí otázka, zda je takový výzkum eticky správně? Neporušují výzkumy z této dizertace nějakým způsobem vědeckou etiku? Obvykle bývá etika výzkumu v prostředí sociálních sítí spojená s odpovědí na otázku, zda se jedná či nejedná o tzv. human subject research, tedy výzkum provozovaný na lidech. Ten je omezován a vymezován celou sérií pravidel, z nichž nejznámějším je tzv. Norimberský kodex, který vznikl jako reakce na neetickou (a často zřúdnou) praxi výzkumů provozovaných v době nacistické Třetí říše. Byť se jedná původně o deklaraci určenou pro výzkumy v oblasti medicíny, jeho základní vymezení i jeho modifikace platí pro všechny oblasti výzkumu, v nichž se experimentuje s lidmi a na lidech včetně sociologie nebo psychologie.

Jak je to v tomto kontextu s výzkumy prezentovanými v této práci? Jaká je jejich povaha? Jedná se vůbec o výzkum zařaditelný pod tzv. human subject research? Domníváme se, že nikoli. Nekoná se v něm totiž žádný experiment, v němž by svým jednáním výzkumník ovlivňoval jednání zkoumaných objektů. V oblasti klasických výzkumných metod v sociologii by se drtivá většina podobných výzkumů v době předinternetové dala v podstatě klasifikovat jako analýza dokumentů, tedy jako výzkum takových dokumentů, „které nebyly vytvořeny za účelem našeho výzkumu“.¹⁶⁹ Zejména v tomto ohledu platí, že „záznamem mohou být právě tak dobře psané dokumenty jako **jakékoli materiální stopy lidského chování**.“¹⁷⁰ (zvýraznil Josef Šlerka).

¹⁶⁹ DISMAN, Miroslav. Jak se vyrábí sociologická znalost: příručka pro uživatele. 3. vyd. Praha: Karolinum, 2000. ISBN 978-80-246-0139-7. str. 124

¹⁷⁰ DISMAN, Miroslav. Jak se vyrábí sociologická znalost: příručka pro uživatele. 3. vyd. Praha: Karolinum, 2000. ISBN 978-80-246-0139-7. str. 124

Tady se obloukem vracíme k námi navrženému chápání dat nikoli jako symbolicky zapsaných výsledků měření, ale jako symbolického záznamu výsledků obecných akcí v digitálním prostředí a jejich studiu v rámci digital humanities, které je umožněno technologickou povahou nových médií.

Spolu s ní ovšem přichází i změna toho, co pro tyto vědy znamená člověk. Pro potřeby našeho výzkumu je totiž člověk v zásadě popsateľný pomocí databáze, zaznamenávající v čase jeho jednotlivé sociální akce. Jedinec je tak popsateľný právě touto historií a lidstvo představuje různé formy toho, jak se k sobě postoje v čase váží, orientují a ovlivňují. To vede ke dvěma zásadním důsledkům. První je fakt, že firmy jako Google vlastní ve svých databázích dostatek informací k tomu, aby byly schopny zpracovat univerzální distribuci slov a slovních spojení. Jiné firmy, jako třeba Facebook, začínají mít k dispozici dostatek dat na to, aby pokryly univerzální distribuci hodnot a postojů lidstva a jejich proměnu v čase. Je jedno, že si tito technologičtí giganti často nechávají svá data pro sebe, změna pohledu na člověka už je přítomna.

Druhým důsledkem je fakt, že lidské postoje, hodnoty a podobně lze seskupovat do shluků podle blízkosti měřitelné na základě různých matematických transformací. Metaforicky řečeno: různé postoje jsou si tak vzdáleny, jak dlouhé by musely být výchovné programy, které by jedny dokázaly přeměnit na druhé.

Oba důsledky znějí poněkud strašidelně, ale vzato do důsledku tento přístup otevírá zcela nové možnosti výzkumu a s nimi i další dobrodružství v dějinách lidského poznání.

Appendix – Vytvořené aplikace a knihovny

Výzkumu dat ze sociálních sítí se autor této disertace věnoval posledních osm let jak v akademickém, tak byznysovém prostředí během svého působení ve společnosti Ataxo Interactive a Socialbakers. Během této doby průběžně vytvářel nástroje, které mu měly zjednodušit práci s daty ze sociálních sítí. Tyto nástroje a knihovny použil i pro analýzu dat v této práci a průběžně o nich informoval v práci samotné. Pro přehlednost je ale stručně shrňme v samostatné části.

Programy samotné lze rozdělit na dva druhy: nástroje pro stahování a transformaci dat ze sociálních sítí (kam patří knihovna Sugar a Geosugar) a nástroje pro vizualizaci dat (kam patří Who are likers?, Page Affinity Analyzer, Normalized Social Distance Matrix Viewer a Normalized Facebook Distance Matrix Viewer). Zdrojové kódy k nim jsou dostupné jako příloha k této disertaci na DVD; v případě, že jsou kódy veřejné, uvádíme i jejich repositář na serveru GitHub.

Sugar

Knihovna Sugar pro programovací jazyk R rozšiřuje možnosti knihovny RFacebooku pro stahování data z Facebooku a pro některé formy dataminingu.

Zdrojové kódy jsou dostupné zde: <https://github.com/josefslerka/sugar>

Geosugar

Knihovna Geosugar rozšiřuje možnosti knihovny RFacebooku a některých dalších, pro stahování dat z FourSquare a geolokovaných dat z Facebooku.

Zdrojové kódy jsou dostupné zde: <https://github.com/josefslerka/geosugar>

Who are likers?

Aplikace Who are likers? slouží k prohlížení datasetu cover a profilových fotografií několika tisíc náhodně vybraných uživatelů českého Facebooku. Aplikace je napsána v jazyce R a její zdrojové kódy jsou k dispozici na přiloženém DVD. Pro vytvoření webového rozhraní je využita knihovna Shiny.

Who are likers?

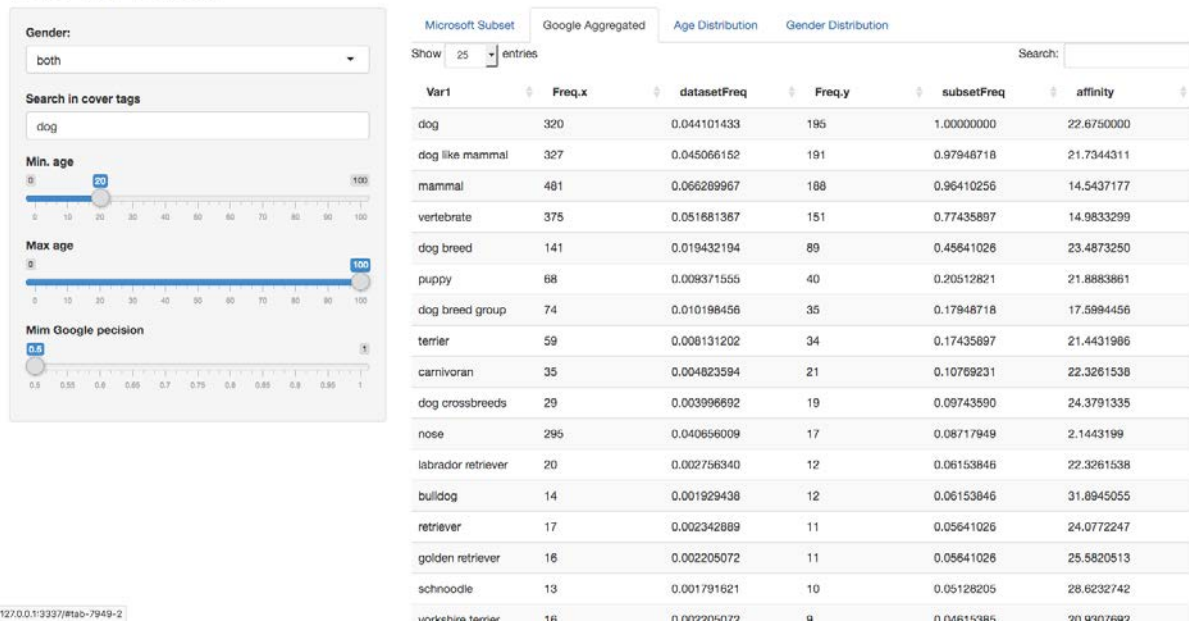
faceAttributes.gender	picture	faceAttributes.age	profile	cover
female	1000947069942280.jpg	28.5		
male	1003146029733558.jpg	32.7		
female	1003146029733558.jpg	24.3		

127.0.0.1:3337/#tab-7949-1

Obr. 43 – Ukázkový screenshot z aplikace Who are likers?

Aplikace umožňuje mj. porovnat, jak se mění distribuce věku a či pohlaví podle specifického obsahu cover fotografie.

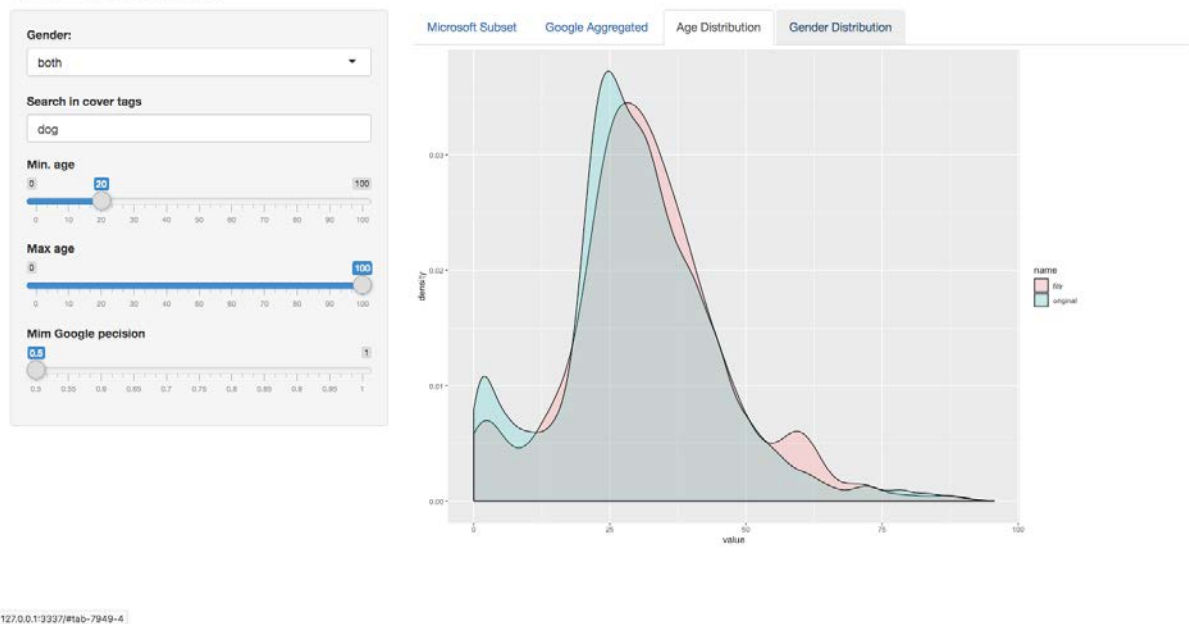
Who are likers?



Obr. 44 – Ukázkový screenshot z aplikace Who are likers?

Pro rozpoznání věku uživatelů z profilové fotografie slouží Microsoft Face Recognition API, pro otágování cover fotografií pak Google Vision API.

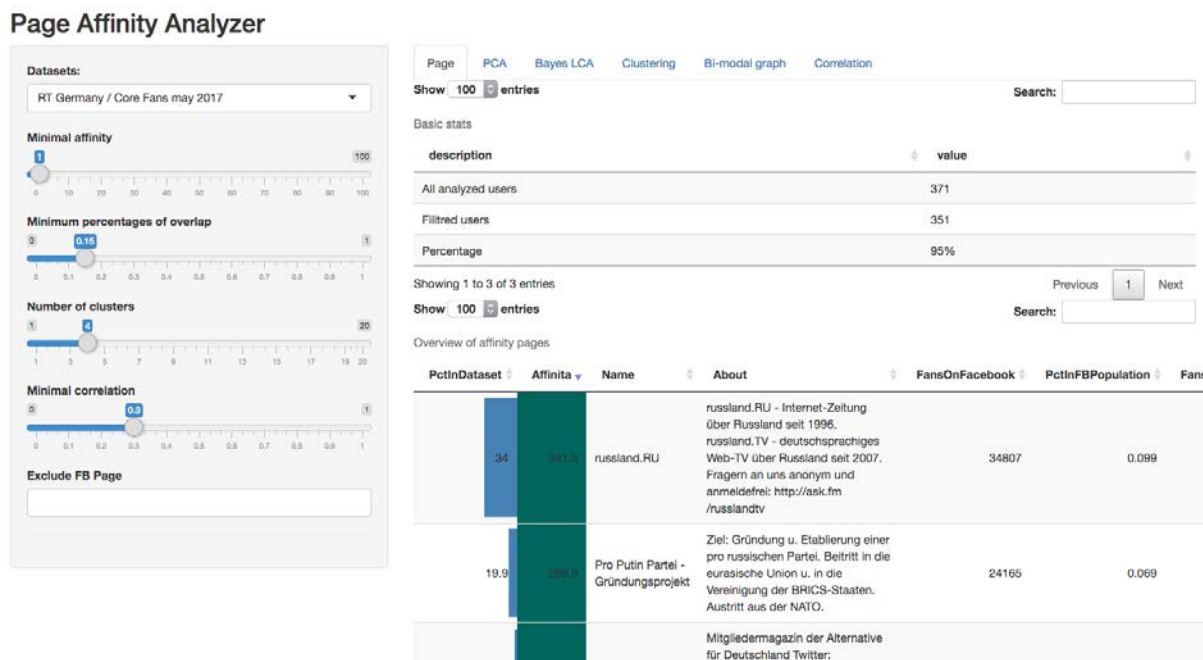
Who are likers?



Obr. 45 – Ukázkový screenshot z aplikace Who are likers?

Page Affinity Analyzer

Aplikace slouží k explorační analýze dat ze stažených oblíbených stránek použité kupříkladu v analýze fanoušků Slušných lidí. Aplikace je napsána v jazyce R a její zdrojové kódy jsou k dispozici na příloženém DVD. Pro vytvoření webového rozhraní je využita knihovna Shiny.



Obr. 46 – Ukázkový screenshot z aplikace Page Affinity Analyzer

Aplikace umožňuje filtrovat dataset podle velikosti a dalších parametrů a aplikovat na něj různé statistické postupy včetně PCA, Bayesian Latent Class Analysis, hierarchického klastrování a k-Means klastrování spolu s odpovídajícími vizualizacemi.

Aplikace je primárně určená k analýze dat z Facebooku, je však kompatibilní i s daty ze sítě Twitter.

Page Affinity Analyzer

Datasets:
RT Germany / Core Fans may 2017

Minimal affinity
0 100

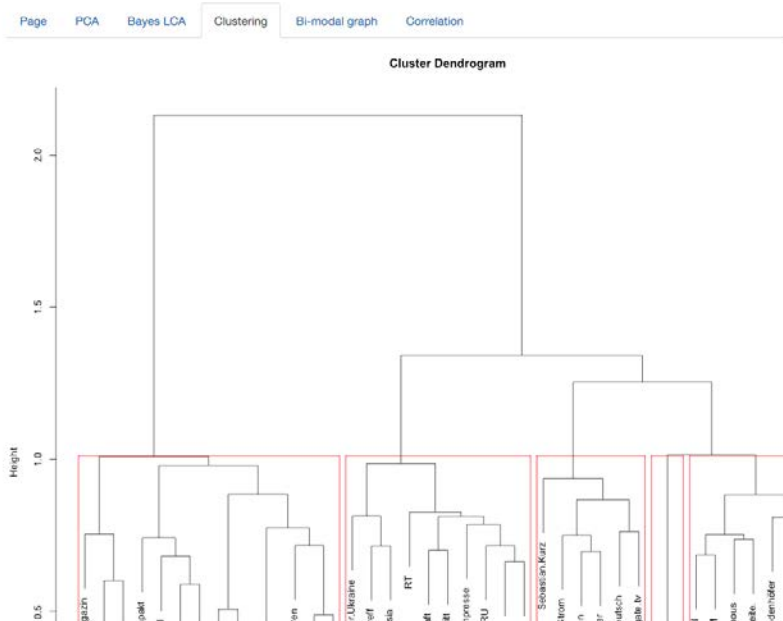
Minimum percentages of overlap
0 0.10 1

Number of clusters
1 4 20

Minimal correlation
0 0.3 1

Exclude FB Page

127.0.0.1:6855/#tab-2690-4



Obr. 47 – Ukázkový screenshot z aplikace Page Affinity Analyzer

Normalized Social Distance Matrix Viewer

Pro analýzu výsledných distančních matic bez ohledu na specifický původ dat jsme vytvořili aplikaci, která umožňuje snadnou explorační analýzu těchto dat. Aplikace je napsána v jazyce R a její zdrojové kódy jsou k dispozici na přiloženém DVD.

Normalized Social Distance Matrix Viewer

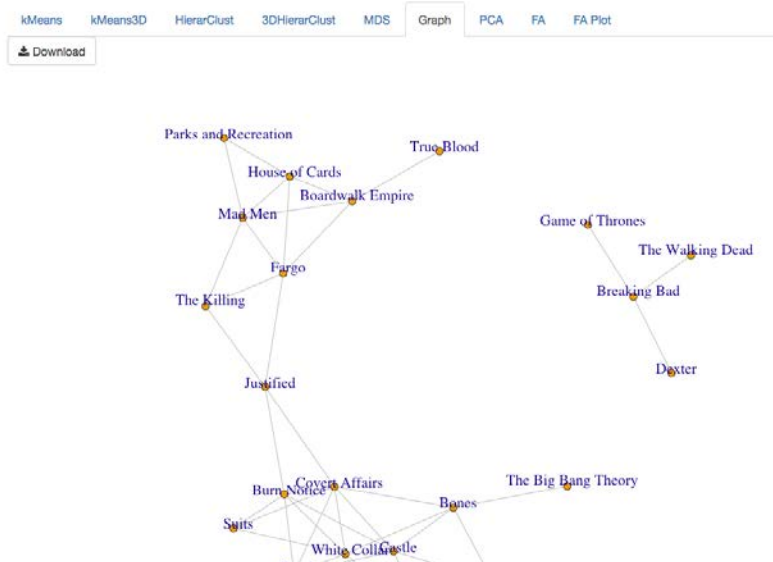
Choose MATRIX file to upload
Browse... usserialy_matrix.csv
Upload complete

Num. components/clusters
0 5 10

Min. distance
0 0.125 0.25 0.375 0.5 0.625 0.75 0.875 1 1.125 1.2

Significant threshold
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

Target:
 Anger Management
 Boardwalk Empire
 Bones
 Breaking Bad
 Burn Notice
 Californication
 Castle
 Cougar Town

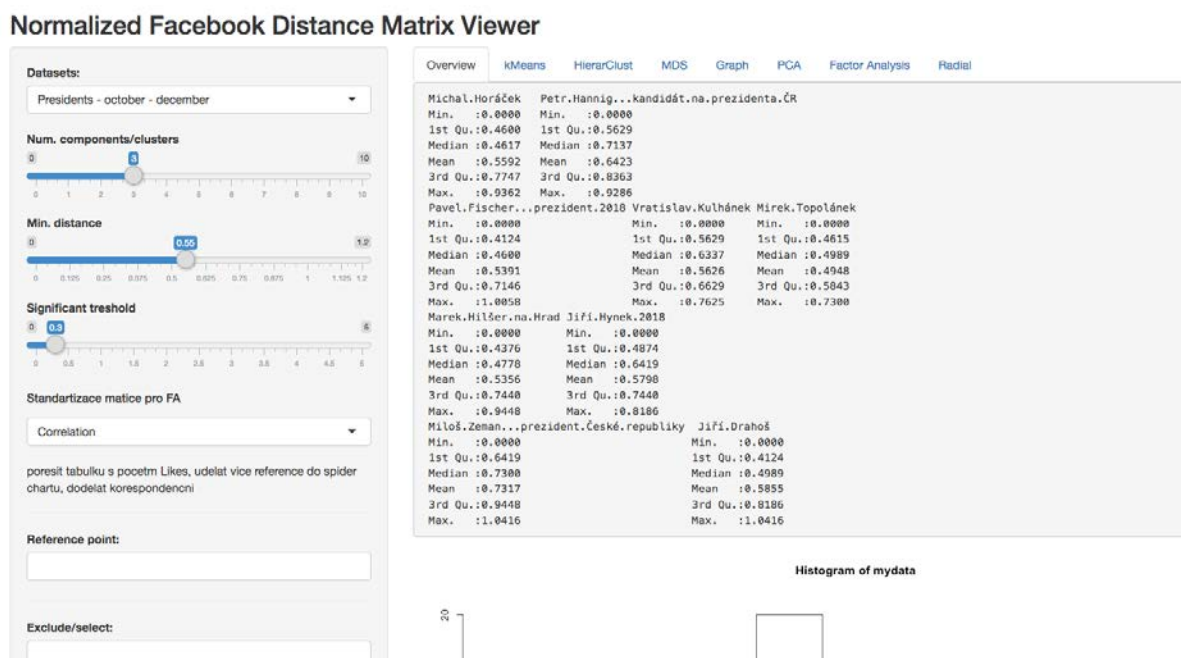


Obr. 48 – Ukázkový screenshot z aplikace Normalized Social Distance Matrix Viewer

Aplikace umožňuje filtrovat dataset podle prvků a dalších parametrů a aplikovat na dataset různé statistické postupy jako hierarchické klastrování a k-Means klastrování spolu s odpovídajícími vizualizacemi.

Normalized Facebook Distance Matrix Viewer

Protože jsem nejčastěji pracoval s daty ze sociální sítě Facebook, vytvořil jsem pro ni speciální aplikaci, která navazuje na obecný princip Normalized Social Distance Matrix Viewer a rozšiřuje ho o další specifické informace a možnosti vytvářet podrobnější vizualizace. Aplikace je napsána v jazyce R a její zdrojové kódy jsou k dispozici na přiloženém DVD.



Obr. 49 – Ukázkový screenshot z aplikace Normalized Facebook Distance Matrix Viewer

Seznam použitých obrázků

Obr. 1 – Vizualizace festivalových plakátů

Zdroj: ŠLERKA, Josef. Sobotka splaká tu [online]. 2016 [cit. 2019-02-23]. Dostupné z: <http://casopis.splav.cz/pdf/2016/splav-2016-07-03.pdf>

Obr. 2 – Mapa prostoru označovaného jako humanities computing

Zdroj: MCCARTY, Willard. Humanities Computing. Paperback edition. Basingstoke, Hampshire: Palgrave Macmillan, 2014. ISBN 978-140-3935-045.

Obr. 3 – Základní Shannonovo komunikační schéma

Zdroj: SHANNON, C. E. A Mathematical Theory of Communication. Bell System Technical Journal. 1948, 27(3), 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x. ISSN 00058580. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>

Obr. 4 – Schéma interakce jednotlivých vrstev DIKW

Zdroj: BOISOT, Max a Agust, CANALS. Data, information and knowledge: have we got it right?. Journal of Evolutionary Economics. 2004, 14(1), 43–67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9>

Obr. 5 – vývoj ceny výpočetní síly přepočtený na jeden iPad2

Zdroj: Cost of Computing Power Equal to an iPad2. In: Hamilton project [online]. 2011 [cit. 2019-02-25]. Dostupné z: http://www.hamiltonproject.org/charts/cost_of_computing_power_equal_to_an_ipad2

Obr. 6 – Základní typ grafu a jejich reprezentace

ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692.

Obr. 7 – Ukázka dekompozice grafu v sociální síti

Zdroj: SKIERA, Bernd. Social Media and Academic Performance: Does Facebook Activity Relate to Good Grades? [online]. In: . [cit. 2019-02-23]. Dostupné z:

https://www.researchgate.net/figure/Full-Social-Network-and-Two-Egocentric-Social-Networks_fig2_273450787

Obr. 8 – Příklad tří typů sítí: centralizované, decentralizované a distribuované

Zdroj: BARAN, P. (1964). On Distributed Communications, Memorandum RM-3420-PR. Santa Monica, Calif.: RAND Corporation. Available at:
http://www.rand.org/content/dam/rand/pubs/research_memoranda/2006/RM3420.pdf
(Accessed May 29, 2017).

Obr. 9 Ukázky bodů s vysokými hodnotami různých metrik

Zdroj: ALHAJJ, Reda S. Encyclopedia of social network analysis and mining. New York: Springer, 2014. ISBN 978-146-1461-692.

Obr. 10 – Egocentrická síť účtu @stunome na Twitteru

Obr. 11 – Egocentrická síť účtu @stunome na Twitteru zbavená periferních bodů a zklastrovaná podle míry propojení

Obr. 12 – Síť míst na základě funkce „Kam chodí lidé po...“ na Foursquare

Obr. 13 – Graf podílu mužských a ženských jmen na stránkách (osa X) a podílu neexistujících jmen (osa Y)

Obr. 14 – Dendrogram výsledků hierarchického klustrování pro stránku Slušní lidé

Obr. 15 – Úvodní fotka stránky Otaku Cz/Sk na Facebooku v době naší analýzy

Obr. 16 – Úvodní fotka stránky Hooligans.cz na Facebooku v době naší analýzy

Obr. 17 – Úvodní fotka stránky Miloš Zeman na Facebooku v době naší analýzy

Obr. 18 – Věkový profil lidí zajímajících se o Otaku na Facebooku podle Facebook Audience Insights

Obr. 19 – Věkový profil lidí zajímajících se o téma Miloš Zeman na Facebooku podle Facebook Audience Insights

Obr. 20–23 – rozložení věku uživatelů Facebooku podle objektu ve fotografii na úvodní fotografii osobního profilu

Obr. 24 – Síť stránek, které sledovala stránka Islám v ČR nechceme na Facebooku a stránky, které sledovaly tyto stránky

Obr. 25 – Model základních principů set-theoretic modelu

Zdroj: PASHLER, Harold E., ed. Encyclopedia of the mind. Los Angeles: SAGE, [2013]. SAGE reference. ISBN 978-1412950572.

Obr. 26 – Příklad základních principů transformačního modelu podobnosti

Zdroj: IMAI, Shiro. Pattern similarity and cognitive transformations. Acta Psychologica. 1977, 41(6), 433-447. DOI: 10.1016/0001-6918(77)90002-6. ISSN 00016918.

Obr. 27 – Výsledky klastrování heterogenních řetězců pomocí NCD

Zdroj: CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403.

Obr. 28. – Dendrogram výsledků hierarchického klastrování pro DSSS

Obr. 29 – Graf MDS výsledků distanční matice na základě NSD pro DSSS

Obr. 30 – Graf faktorové analýzy výsledků distanční matice na základě NSD pro DSSS

Obr. 31 – Síťový graf blízkosti vybraných facebookových stránek na základě NSD lajkujících

Obr. 32 – K-means klastering blízkosti vybraných facebookových stránek na základě NSD lajkujících

Obr. 33 – Síťový graf blízkosti vybraných facebookových stránek zpravodajských webů na základě NSD lajkujících

Obr. 34 – Dendrogram výsledků hierarchického klastrování na základě NSD lajkujících posty na zpravodajských stránkách

Obr. 35 – Graf MDS výsledků distanční matice na NSD lajkujících posty na zpravodajských stránkách

Obr. 36 – Výsledný síťový graf blízkosti vybraných facebookových stránek na základě NSD lajkujících zpravodajské stránky na Facebooku

Obr. 37 – Graf MDS blízkosti českých politických stran na základě NSD lajkujících, podzim 2013

Obr. 38 – Graf MDS blízkosti českých politických stran na základě NSD lajkujících, jaro 2014

Obr. 39 – Korelační matice volebních výsledků českých politických stran z voleb do Evropského parlamentu a parlamentních voleb 2013

Obr. 40 – Graf MDS blízkosti polských politických stran na základě NSD lajkujících, jaro 2014

Obr. 41 – Graf MDS blízkosti slovenských politických stran na základě NSD lajkujících, jaro 2014

Obr. 42 – Graf MDS blízkosti německých politických stran na základě NSD lajkujících, jaro 2014

Obr. 43 – Ukázkový screenshot z aplikace Who are likers?

Obr. 44 – Ukázkový screenshot z aplikace Who are likers?

Obr. 45 – Ukázkový screenshot z aplikace Who are likers?

Obr. 46 – Ukázkový screenshot z aplikace Page Affinity Analyzer

Obr. 47 – Ukázkový screenshot z aplikace Page Affinity Analyzer

Obr. 48 – Ukázkový screenshot z aplikace Normalized Social Distance Matrix Viewer

Obr. 48 – Ukázkový screenshot z aplikace Normalized Facebook Distance Matrix Viewer

Seznam tabulek

Tab. 1 – Čtyři paradigmatata vědy

Zdroj: KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. 2014, 1(1). DOI: 10.1177/2053951714528481. ISSN 2053-9517. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/2053951714528481>

Tab. 2 – Klasifikace druhů big social data

OLSHANNIKOVA, Ekaterina, Thomas OLSSON, Jukka HUHTAMÄKI a Hannu KÄRKKÄINEN. Conceptualizing Big Social Data. Journal of Big Data. 2017, 4(1). DOI: 10.1186/s40537-017-0063-x. ISSN 2196-1115. Dostupné také z: <http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x> str. 14

Tab. 3 – Typologie fanoušků facebookových stránek

Zdroj: WALLACE, Elaine, Isabel BUIL, Leslie DE CHERNATONY a Michael HOGAN. Who “Likes” You ... and Why? A Typology of Facebook Fans. Journal of Advertising Research. 2014, 54(1), 92–109. ISSN 0021-8499. Dostupné také z: <http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-54-1-092-109>

Tab. 4 – Typologie motivace pro lajkování postů

CHIN, Chih-Yu, Hsi-Peng LU a Chao-Ming WU. Facebook Users' Motivation for Clicking the “Like” Button. Social Behavior and Personality: an international journal. 2015, 43(4), 579–592. DOI: 10.2224/sbp.2015.43.4.579. ISSN 03012212.

Tab. 5 – Podíl druhů reakcí na facebookové stránce Slušní lidé

Tab. 6 – Silně afinitní facebookové stránky vůči stránce Slušní lidé

Tab. 7 – Objekty a témata identifikované v profilových fotkách uživatelů v datasetech z února a dubna 2017

Tab. 8 – Afinitní facebookové stránky vůči stránce Otaku Cz/Sk

Tab. 9 – Zastoupení nejčastějších motivů v profilových fotkách nejaktivnějších fanoušků stránky Otaku Cz/SK v porovnání s průměrnou aktivní populací na Facebooku

Tab. 10 – Zastoupení nejčastějších motivů v profilových fotkách nejaktivnějších fanoušků stránky Hooligans.cz v porovnání s průměrnou aktivní populací na Facebooku

Tab. 11 – Zastoupení nejčastějších motivů v profilových fotkách nejaktivnějších fanoušků stránky Miloš Zeman v porovnání s průměrnou aktivní populací na Facebooku

Tab. 12 – Přehled stránek s největším počtem indegree

Tab. 13 – Přehled nejvíce afinitních facebookových stránek ke stránce Islám v ČR nechceme

Tab. 14 – Přehled afinitních facebookových stránek ke stránce Islám v ČR nechceme

Tab. 15 – Přehled nejbližších facebookových stránek ke stránce DSSS na základě NSD

Tab. 16 – Přehled nejvíce zastoupených sdílených domén pro mainstreamové weby

Tab. 17 – Přehled nejvíce zastoupených sdílených domén pro prokremelské weby

Tab. 18 – Výsledky Mantel testu

Seznam použité literatury

AARSETH, Espen J. *Cybertext: perspectives on ergodic literature*. Baltimore, Md.: Johns Hopkins University Press, c1997. ISBN 978-080-1855-795.

ACKOFF, R. L. From data to wisdom. *Journal of Applied Systems Analysis*. 1989, (16), 3–9.

ALHAJJ, Reda S. *Encyclopedia of social network analysis and mining*. New York: Springer, 2014. ISBN 978-146-1461-692.

ANDERSON, Chris. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired* [online]. 2008, 23.6.2008, 2008 [cit. 2019-02-23]. Dostupné z: <https://www.wired.com/2008/06/pb-theory/>

BAKSHY, E., S. MESSING a L. A. ADAMIC. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 2015, 348(6239), 1130–1132. DOI: 10.1126/science.aaa1160. ISSN 0036-8075. Dostupné také z: <http://www.sciencemag.org/cgi/doi/10.1126/science.aaa1160>

BAR-HILLEL, J. a R. CARNAP. Sémantická informace. In: DOLEŽAL, Lubomír. *Teorie informace v jazykovědě*. Praha: Nakladatelství Československé akademie věd, 1964, s. 165–175.

BENNETT, Charles H. Information distance. *IEEE Transactions on information theory*. 1998, 44(4), 1407–1423.

BERMAN, Jules J. *Principles of big data: preparing, sharing, and analyzing complex information*. Amsterdam: Elsevier, Morgan Kaufmann, [2013]. ISBN 978-0124045767.

BERRY, David M. *Understanding digital humanities*. New York: Palgrave Macmillan, 2012. ISBN 978-0-230-29265-9.

BLONDEL, Vincent D, Jean-Loup GUILLAUME, Renaud LAMBIOTTE a Etienne LEFEBVRE. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008, 2008(10). DOI: 10.1088/1742-5468/2008/10/P10008. ISSN 1742-5468. Dostupné také z: <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008?key=crossref.46968f6ec61eb8f907a760be1c5ace52>

BOBOK, Dalibor. Selective Exposure, Filter Bubbles and Echo Chambers on Facebook: Diplomová práce [online]. Budapest, 2016 [cit. 2019-02-24]. Dostupné z: www.etd.ceu.hu/2016/bobok_dalibor.pdf

BOGARDUS, Emory S. A Social Distance Scale. *Sociology and Social Research*. 1933, (17), 265–271.

BOISOT, Max a CANALS, Agust. Data, information and knowledge: have we got it right?. *Journal of Evolutionary Economics*. 2004, 14(1), 43–67. DOI: 10.1007/s00191-003-0181-9. ISSN 0936-9937. Dostupné také z: <http://link.springer.com/10.1007/s00191-003-0181-9>

BOYD, Danah a Nicole ELLISON. Social network sites: definition, history, and scholarship. *IEEE Engineering Management Review*. 2010, 38(3), 16–31. DOI: 10.1109/EMR.2010.5559139. ISSN 0360-8581. Dostupné také z: <http://ieeexplore.ieee.org/document/5559139/>

BREIMAN, Leo. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*. 2001, 16(3), 199–231. DOI: 10.1214/ss/1009213726. ISSN 0883-4237. Dostupné také z: <http://projecteuclid.org/euclid.ss/1009213726>

BREJČÁK, Peter. Infografika: Jak Češi využívají ty internety a sociální sítě v roce 2018. In: *Ty internety* [online]. 2018 [cit. 2019-02-23]. Dostupné z: <https://tyinternety.cz/socialni-site/infografika-jak-cesi-vyuzivaji-ty-internety-a-socialni-site-v-roce-2018/>

BUNKER, Matthew P., K. N. RAJENDRAN a Steven B. CORBIN. The antecedents of satisfaction for Facebook “likers“ and their effect on word of-mouth. *Marketing Management Journal*. 2013, 23(2), 21–34.

BURDICK, Anne. Digital humanities. Cambridge, MA: MIT Press, c2012. ISBN 978-026-2018-470.

CASS, Stephen. Unthinking Machines. MIT Technology Review [online]. 4. 5. 2011 [cit. 2019-02-23]. Dostupné z: <https://www.technologyreview.com/s/423917/unthinking-machines/>

CASTELLS, Manuel. The rise of the network society. 2nd ed., with a new pref. Malden, MA: Wiley-Blackwell, 2010. ISBN 978-140-5196-864.

CEBRIN, Manuel, Manuel ALFONSECA a Alfonso ORTEGA. The Normalized Compression Distance Is Resistant to Noise. IEEE Transactions on Information Theory. 2007, 53(5), 1895-1900. DOI: 10.1109/TIT.2007.894669. ISSN 0018-9448. Dostupné také z: <http://ieeexplore.ieee.org/document/4167725/>

CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403.

CILIBRASI, Rudi L. a Paul M.B. VITANYI. The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering. 2007, 19(3), 370–383. DOI: 10.1109/TKDE.2007.48. ISSN 1041-4347. Dostupné také z: <https://ieeexplore.ieee.org/document/4072748/>

CILIBRASI, R. a P. M. B. VITANYI. Clustering by Compression. IEEE Transactions on Information Theory. 2005, 51(4), 1523–1545. DOI: 10.1109/TIT.2005.844059. ISSN 0018-9448. Dostupné také z: <http://ieeexplore.ieee.org/document/1412045/>

DAHLBERG, Lincoln. Internet Research Tracings: Towards Non-Reductionist Methodology. Journal of Computer-Mediated Communication. 2004, 9(3), 00-00. DOI: 10.1111/j.1083-6101.2004.tb00289.x. ISSN 10836101. Dostupné také z: <https://academic.oup.com/jcmc/article/4614480>

DAVIDSON, Donald. Subjektivita, intersubjektivita, objektivita. Praha: Filosofia, 2004. ISBN 80-700-7190-7.

DIJCK, José van. The culture of connectivity: a critical history of social media. New York: Oxford University Press, c2013. ISBN 978-019-9970-780.

DISMAN, Miroslav. Jak se vyrábí sociologická znalost: příručka pro uživatele. 3. vyd. Praha: Karolinum, 2000. ISBN 978-80-246-0139-7.

FESTINGER, L., S. SCHACHTER a K. BACK. The Spatial Ecology of Group Formation. In: FESTINGER, L., S. SCHACHTER a K. BACK, ed. Social Pressure in Informal Groups, 1950. Oxford: Harper, 1950.

FLORIDI, Luciano. The philosophy of information. New York: Oxford University Press, 2011. ISBN 978-019-9232-383.

FUCHS, Christian a Marisol SANDOVAL. Critique, social media and the information society. New York, 2014. ISBN 978-041-5841-856.

GARDINER, Eileen a Ronald G. MUSTO. The digital humanities: a primer for students and scholars. New York: Cambridge University Press, 2015. ISBN 978-110-7601-024.

GITELMAN, Lisa. Always already new: media, history and the data of culture. Cambridge, Mass.: MIT Press, c2006. ISBN 978-026-2072-717.

GOFFMAN, Erving. Všichni hrajeme divadlo: sebezprezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1.

GOLD, Matthew K. Debates in the digital humanities. Minneapolis: Univ. Of Minnesota Press, 2012. ISBN 978-0816677955.

GOLDSTONE, Robert L. a Ji Yun SON. Similarity. In: HOLYOAK, Keith J. a Robert G. MORRISON. Cambridge Handbook of Thinking and Reasoning. Cambridge: Cambridge University Press, 2005, s. 13–36. ISBN 9780521531016.

GOODMAN, Nelson. Sedm výhrad proti podobnosti. In: Aluze [online]. 2008 [cit. 2019-02-23]. Dostupné z: http://aluze.cz/2008_02/07_studie_goodman.php

HABERMAS, Jürgen. Strukturální přeměna veřejnosti: zkoumání jedné kategorie občanské společnosti. Praha: Filosofia, 2000. Morální a politická filosofie. ISBN 80-7007-134-6.

HAHN, Ulrike, Nick CHATER a Lucy B RICHARDSON. Similarity as transformation. *Cognition*. 2003, 87(1), 1–32. DOI: 10.1016/S0010-0277(02)00184-1. ISSN 00100277. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S0010027702001841>

HEBDIGE, Dick. *Subculture: the meaning of style*. New York: Routledge, 1991. New accents (Routledge (Firm)). ISBN 04-150-3949-5.

HEY, Anthony J. G. *The fourth paradigm: data-intensive scientific discovery*. Redmond, Washington: Microsoft Research, [2009]. ISBN 978-098-2544-204.

HODGETTS, Darrin a Otilie STOLTE. Social Distance. In: *Encyclopedia of Critical Psychology*. New York, NY: Springer New York, 2014, 2014, s. 1776–1778. DOI: 10.1007/978-1-4614-5583-7_559. ISBN 978-1-4614-5582-0. Dostupné také z: http://link.springer.com/10.1007/978-1-4614-5583-7_559

CHIN, Chih-Yu, Hsi-Peng LU a Chao-Ming WU. Facebook Users' Motivation for Clicking the “Like” Button. *Social Behavior and Personality: an international journal*. 2015, 43(4), 579–592. DOI: 10.2224/sbp.2015.43.4.579. ISSN 03012212.

IAFRATE, Fernando. *From big data to smart data*. Hoboken, NJ, 2015. *Advances in information systems set*, v. 1. ISBN 978-184-8217-553.

IMAI, Shiro. Pattern similarity and cognitive transformations. *Acta Psychologica*. 1977, 41(6), 433-447. DOI: 10.1016/0001-6918(77)90002-6. ISSN 00016918. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/0001691877900026>

JENSEN SCHAU, Hope a Mary C. GILLY. We Are What We Post? Self-Presentation in Personal Web Space. *Journal of Consumer Research*. 2003, 30(3), 385–404. DOI:

10.1086/378616. ISSN 0093-5301. Dostupné také z: <https://academic.oup.com/jcr/article-lookup/doi/10.1086/378616>

JOCKERS, Matthew L. a David MIMNO. Significant themes in 19th-century literature. *Poetics*. 2013, 41(6), 750–769. DOI: 10.1016/j.poetic.2013.08.005. ISSN 0304422X. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0304422X13000673>

KARAKAYALI, Nedim. Social Distance and Affective Orientations. *Sociological Forum*. 2009, 24(3), 538–562. DOI: 10.1111/j.1573-7861.2009.01119.x. ISSN 08848971. Dostupné také z: <http://doi.wiley.com/10.1111/j.1573-7861.2009.01119.x>

KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. 2014, 1(1). DOI: 10.1177/2053951714528481. ISSN 2053-9517. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/2053951714528481>

KOUCKÝ, Michal. A Brief Introduction to Kolmogorov Complexity [online]. In: . 4. 5. 2006 [cit. 2019-02-23]. Dostupné z: <https://iuuk.mff.cuni.cz/~koucky/vyuka/ZS2013/kolmcomp.pdf>

LIPSMAN, Andrew, Graham MUDD, Mike RICH a Sean BRUICH. The Power of “Like”. *Journal of Advertising Research*. 2012, 52(1), 40–52. DOI: 10.2501/JAR-52-1-040-052. ISSN 0021-8499. Dostupné také z: <http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-52-1-040-052>

LISTER, Martin. *New media: a critical introduction*. 2nd ed. New York, N.Y.: Routledge, 2009. ISBN 978-041-5431-613.

MACEK, Jakub. *Poznámky ke studiím nových médií*. Brno: Masarykova univerzita, 2013. ISBN 978-802-1064-768.

MACEK, Jakub a Alena MACKOVÁ. *Stará a nová média v každodennosti českých publik (výzkumná zpráva)* [online]. 1. [cit. 2019-02-23]. ISBN 10.13140/RG.2.1.3589.9364. Dostupné z:

https://www.researchgate.net/publication/280154724_Stara_a_nova_media_v_kazdodennosti_ceskych_publik_vyzkumna_zprava

MANOVICH, Lev. Trending: The Promises and the Challenges of Big Social Data. *Debates in the Digital Humanities*. University of Minnesota Press, 2012, 2012-01-01, 460–475. DOI: 10.5749/minnesota/9780816677948.003.0047. ISBN 9780816677948. Dostupné také z: <http://minnesota.universitypressscholarship.com/view/10.5749/minnesota/9780816677948.001.0001/upso-9780816677948-chapter-47>

MANOVICH, Lev. *Jazyk nových médií*. Praha: Univerzita Karlova, nakladatelství Karolinum, 2018. *Studia nových médií*. ISBN 978-802-4629-612.

MCCARTY, Willard. *Humanities Computing*. Paperback edition. Basingstoke, Hampshire: Palgrave Macmillan, 2014. ISBN 978-140-3935-045.

MCPHERSON, Miller, Lynn SMITH-LOVIN a James M COOK. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*. 2001, 27(1), 415–444. DOI: 10.1146/annurev.soc.27.1.415. ISSN 0360-0572. Dostupné také z: <http://www.annualreviews.org/doi/10.1146/annurev.soc.27.1.415>

MORETTI, Franco. *Grafy, mapy, stromy: abstraktní modely literární historie*. 1. Praha: Karolinum, 2014. *Studia nových médií*. ISBN 978-802-4626-093.

MOUW, Ted a Barbara ENTWISLE. Residential Segregation and Interracial Friendship in Schools. *American Journal of Sociology*. 2006, 112(2), 394–441. DOI: 10.1086/506415. ISSN 0002-9602. Dostupné také z: <http://www.journals.uchicago.edu/doi/10.1086/506415>

NORVIG, Peter. On Chomsky and the Two Cultures of Statistical Learning. *Osobní stránka Petera Norviga* [online]. [cit. 2019-02-23]. Dostupné z: <http://norvig.com/chomsky.html>

OLSHANNIKOVA, Ekaterina, Thomas OLSSON, Jukka HUHTAMÄKI a Hannu KÄRKKÄINEN. Conceptualizing Big Social Data. *Journal of Big Data*. 2017, 4(1). DOI: 10.1186/s40537-017-0063-x. ISSN 2196-1115. Dostupné také z: <http://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0063-x>

OSOLSOBĚ, Ivo. *Ostenze, hra, jazyk: sémiotické studie*. Brno: Host, 2002. Teoretická knihovna. ISBN 80-729-4076-7.

OTTE, Evelien a Ronald ROUSSEAU. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*. 2016, 28(6), 441–453. DOI: 10.1177/016555150202800601. ISSN 0165-5515. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/016555150202800601>

PASHLER, Harold E., ed. *Encyclopedia of the mind*. Los Angeles: SAGE, [2013]. SAGE reference. ISBN 978-1412950572.

PEIRCE, Ch. S. *Collected Papers of Charles Sanders Peirce*. Vol. 2. Cambridge: Belknap, 1966.

PELLETIER, Mark J. a Alisha BLAKENEY HORKY. Exploring the Facebook Like: a product and service perspective. *Journal of Research in Interactive Marketing*. 2015, 9(4), 337–354. DOI: 10.1108/JRIM-09-2014-0059. ISSN 2040-7122. Dostupné také z: <http://www.emeraldinsight.com/doi/10.1108/JRIM-09-2014-0059>

PELLETIER, M. a A. HORKY. The Anatomy of a Facebook Like: An Exploratory Study of Antecedents and Outcomes. *Annals of the Society for Marketing Advances*. 2013, (25).

PRECIADO, Paulina, Tom A.B. SNIJDERS, William J. BURK, Håkan STATTIN a KERR. *Social Networks*. 2011. DOI: 10.1016/j.socnet.2011.01.002. ISSN 03788733. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0378873311000128>

ROGERS, Richard. *Digital methods*. Cambridge, Massachusetts: The MIT Press, [2013]. ISBN 978-026-2018-838.

ROWLEY, Jennifer. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*. 2007, 33(2), 163–180. DOI: 10.1177/0165551506070706. ISSN 0165-5515. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/0165551506070706>

RYŠAVÝ, DAN. Sociální distance vůči Romům: Případ vysokoškolských studentů. Sociologický časopis. Praha, 2003, 39(1), 55–77.

SARTORI, Giovanni. Strany a stranické systémy: schéma pro analýzu. Brno: Centrum pro studium demokracie a kultury, 2005. Klasikové společenských věd. ISBN 80-732-5062-4.

SHANNON, C. E. A Mathematical Theory of Communication. Bell System Technical Journal. 1948, 27(3), 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x. ISSN 00058580. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>

SKIERA, Bernd. Social Media and Academic Performance: Does Facebook Activity Relate to Good Grades? [online]. [cit. 2019-02-23]. Dostupné z: https://www.researchgate.net/figure/Full-Social-Network-and-Two-Egocentric-Social-Networks_fig2_273450787

SMATANA, Juraj. Neúplný, ale stále aktualizovaný zoznam webových stránok, ktorých linkovaním si koledujete o moju odbornú starostlivosť. In: Hadičky a udice [online]. [cit. 2019-02-23]. Dostupné z: <http://afinabul.blog.cz/1502/juraj-smatana-neuplny-ale-stale-aktualizovany-zoznam-webovych-stranok-ktorych-linkovanim-si-koledujete-o-moju-odbornu-starostlivost>

ŠAFR, Jiří, ed. Sociální distance, interakce, relace a kategorizace: alternativní teoretické perspektivy studia sociální stratifikace. Praha: Sociologický ústav Akademie věd České republiky, 2008. ISBN 978-80-7330-146-0.

ŠLERKA, Josef. Nová a sociální média: Základní vymezení pojmů. DINGIR [online]. 2015(3) [cit. 2019-02-23]. Dostupné z: http://www.dingir.cz/cislo/15/3/nova_a_socialni_media.pdf

ŠLERKA, Josef. Housle a mřížky z voleb do parlamentu. In: Databoutique [online]. [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/66175359703/housle-a-mrizky-z-voleb-do-parlamentu>).

ŠLERKA, Josef. Sémiotická interpretace Kantovy estetiky.: Diplomová práce. 1. Praha, 2007.

ŠLERKA, Josef. Polarizovaná společnost? Nikoli, je to složitější. Magazín Reportér [online]. 1.10.2016 [cit. 2019-02-24]. Dostupné z: <https://reportermagazin.cz/a/i8sYJ/polarizovana-spolecnost--nikoli-je-to-slozitejsi>

ŠLERKA, Josef. Slušní lidé na Facebooku. Databoutique [online]. 1. 7. 2017 [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/161312186128/slu%C5%A1n%C3%AD-lid%C3%A9-na-facebooku>

ŠLERKA, Josef. Sobotka splaká tu [online]. 2016 [cit. 2019-02-23]. Dostupné z: <http://casopis.splav.cz/pdf/2016/splav-2016-07-03.pdf>

ŠLERKA, Josef a Jakub FIALA. O sebezprezentaci stránky Islám v ČR nechceme na českém Facebooku. In: OSTRÁNSKÝ, B. Islamofobie po česku. Praha: Vyšehrad, 2017, s. 235–248. ISBN 978-80-7429-903-2.

ŠLERKA, Josef a Lenka KRISOVÁ. Tzv. proruské zpravodajské weby na Facebooku?. In: Databoutique [online]. [cit. 2019-02-23]. Dostupné z: <http://databoutique.cz/post/113772071738/tzv-prorusk%C3%A9-zpravodajsk%C3%A9-weby-na-facebooku-1>

ŠLERKA, Josef a Filip SMOLÍK. Automatická měřítka čitelnosti pro česky psané texty. Studie z aplikované lingvistiky. 2010, 1(1), 33–44, 11 s. ISSN 18043240.

ŠLERKA, Josef a Vít ŠISLER. Who Is Shaping Your Agenda? Social Network Analysis of Anti-Islam and Anti-immigration Movement Audiences on Czech Facebook. In: Expressions of Radicalization. Cham: Springer International Publishing, 2018, 2018-12-20, s. 61–85. DOI: 10.1007/978-3-319-65566-6_3. ISBN 978-3-319-65565-9. Dostupné také z: http://link.springer.com/10.1007/978-3-319-65566-6_3

ŠLERKA, Josef. Korelace politických stran na Facebooku s výsledky voleb do Evropského parlamentu. In: Databoutique [online]. [cit. 2019-02-24]. Dostupné z:

<http://databoutique.cz/post/87676505248/korelace-politick%C3%BDch-stran-na-facebooku-s-v%C3%BDsledky>

TÓTH, Gabriel. Veľká analýza blízkosti podporovateľov kandidátov na prezidenta. In: Katedra komunikácie [online]. 4. 2. 2019 [cit. 2019-02-24]. Dostupné z: <https://katedrakomunikacie.sk/velka-analyza-blizkosti-kandidatov-na-prezidenta/>

TVERSKY, Amos a Eldar SHAFIR. Preference, belief, and similarity: selected writings. Cambridge, Mass.: MIT Press, c2004. ISBN 978-0262700931.

UGANDER, Johan. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503, 2011.

VÄYRYNEN, Jaakko J. Normalized compression distance as an automatic MT evaluation metric. Proceedings of MT. 2010, (25), 21–22.

VITÁNYI, Paul M. B. Normalized information distance. In: Information theory and statistical learning. Boston: Springer, 2009, s. 45–82.

WALLACE, Elaine, Isabel BUIL, Leslie DE CHERNATONY a Michael HOGAN. Who “Likes” You ... and Why? A Typology of Facebook Fans. Journal of Advertising Research. 2014, 54(1), 92–109. DOI: 10.2501/JAR-54-1-092-109. ISSN 0021-8499. Dostupné také z: <http://www.journalofadvertisingresearch.com/lookup/doi/10.2501/JAR-54-1-092-109>

WALLACE, Elaine, Isabel BUIL a Leslie DE CHERNATONY. Facebook ‘friendship’ and brand advocacy. Journal of Brand Management. 2012, 20(2), 128–146. DOI: 10.1057/bm.2012.45. ISSN 1350-231X. Dostupné také z: <http://link.springer.com/10.1057/bm.2012.45>

WARDROP-FRUIIN, Noah a Nick MONTFORT. The NewMediaReader. Cambridge, Mass.: MIT Press, c2003. ISBN 978-026-2232-272.

WASSERMAN, Stanley a Katherine FAUST. Social network analysis: methods and applications. New York: Cambridge University Press, 1994. ISBN 978-052-1387-071.

WELS, R. Míra subjektivní informace. In: DOLEŽAL, Lubomír. Teorie informace a jazykověda. Praha: Nakladatelství Československé akademie věd, 1964, s. 187–195.

WIMMER, Andreas a Kevin LEWIS. Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*. 2010, 116(2), 583–642. DOI: 10.1086/653658. ISSN 0002-9602. Dostupné také z: <http://www.journals.uchicago.edu/doi/10.1086/653658>

Četnost jmen a příjmení. Ministerstvo vnitra České republiky [online]. 2015 [cit. 2019-02-23]. Dostupné z: <http://www.mvcr.cz/clanek/cetnost-jmen-a-prijmeni-722752.aspx>

Kompres dat. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://cs.wikipedia.org/wiki/Kompres_dat

Facebook real-name policy controversy. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Facebook_real-name_policy_controversy

What names are allowed on Facebook?. Facebook [online]. [cit. 2019-02-23]. Dostupné z: https://www.facebook.com/help/112146705538576?helpref=faq_content

Digital Humanities – Wikipedia. Wikipedia [online]. [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Digital_humanities

Mantel test. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-24]. Dostupné z: https://en.wikipedia.org/wiki/Mantel_test

Korpus českého verše [online]. [cit. 2019-02-23]. Dostupné z: http://versologie.cz/v2/web_content/corpus.php

Force-directed graph drawing. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-24]. Dostupné z: https://en.wikipedia.org/wiki/Force-directed_graph_drawing

Cultural Analytics [online]. [cit. 2019-02-23]. Dostupné z: <http://lab.culturalanalytics.info/>
Software Studies Initiative: ImagePlot visualization software: explore patterns in large image collections. Software Studies Initiative. Software Studies Initiative [online]. [cit. 2019-02-23]. Dostupné z: <http://lab.softwarestudies.com/p/imageplot.html>

Data mining – Wikipedia. Wikipedia [online]. [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Data_mining

Index Thomisticus [online]. [cit. 2019-02-23]. Dostupné z: <http://www.corpusthomisticum.org/it/index.age>

Big data: The next frontier for innovation, competition, and productivity. 1. New York, NY: McKinsey & Company, 2011. ISBN 9780983179696.

Afinita. Mediaguru [online]. [cit. 2019-02-23]. Dostupné z: <https://www.mediaguru.cz/medialni-slovník/afinita/>

Knihovna Geosugar rozšiřuje možnosti knihovny RFacebooku [online]. [cit. 2019-02-23]. Dostupné z: <https://github.com/josefslerka/geosugar>

Hierarchical clustering. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-02-23]. Dostupné z: https://en.wikipedia.org/wiki/Hierarchical_clustering

If This Then That [online]. [cit. 2019-02-23]. Dostupné z: <https://ifttt.com>

JKG 2017: Brněnská ODS si podává ruku s neonacisty a fotbalovými chuligány. In: Antifa [online]. 2017 [cit. 2019-02-23]. Dostupné z: <http://antifa.cz/content/jkg-2017-brnenska-ods-si-podava-ruku-s-neonacisty-fotbalovymi-chuligany>

Cloud Vision [online]. [cit. 2019-02-23]. Dostupné z: <https://cloud.google.com/vision/>

How News Feed Works. In: Facebook [online]. [cit. 2019-02-25]. Dostupné z: <https://www.facebook.com/help/www/327131014036297/>

Zona A: Prečo Je Život Taký?. In: *Karaoke Texty* [online]. [cit. 2019-02-25]. Dostupné z: <https://www.karaoketexty.cz/texty-pisni/zona-a/preco-je-zivot-taky-38210>