

**Univerzita Karlova**

**Filozofická fakulta**

**Ústav informačních studií a knihovnictví**

# **Autoreferát disertační práce**

Mgr. Josef Šlerka

**Normalized Social Distance**

Normalized Social Distance

Vedoucí práce doc. RNDr. Jiří Souček, DrSc.

2019

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 jednájícími. 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. komputačního 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 (Web) Distance můžeme chápat jako posun od de facto syntaxe k sémantice. V této práci navrhujeme 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 čtenář dostane 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. Klíčovým je pro nás chápání dat nikoli jako symbolicky zapsaného výsledku měření, ale jako symbolicky zapsané výsledky

nějakých akcí vůbec. Jednání uživatelů v prostředí online sociálních sítí pak představuje specifické sociální akce nadané intencionalitou a uchované a dostupné pro různé druhy analýz.

**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.

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 životě<sup>1</sup>, tak v sociálních sítích, jak ostatně potvrzuje řada výzkumů. Ať již jde o homofilii v oblasti věku či národnosti<sup>2</sup>, homofilii rasovou<sup>3</sup> či další.<sup>4</sup>

Vedle principu homofilie je druhým principem důležitým pro vznik vazeb v sociálních sítích tzv. princip propinquity. Jeho postulování je spojeno s výzkumem Festingera, Schachtea a Backa publikovaném v textu *The Spatial Ecology of Group Formation*.<sup>5</sup> Jádrem principu propinquity je teze, že lidé, kteří jsou si fyzicky blízko, mají častěji tendenci navazovat

---

<sup>1</sup> 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>.

<sup>2</sup> UGANDER, Johan. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503, 2011.

<sup>3</sup> 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>

<sup>4</sup> 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>.

<sup>5</sup> 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.

přátelství. Původní Festingerovy výzkumy se věnovaly studentskému spolubydlení, na ně pak navázala řada jiných.<sup>6</sup>

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 publikovaný pod názvem *Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook*.<sup>7</sup>

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 je 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.<sup>8</sup> A právě tato strategie je pro nás jednou z klíčových součástí přístupu k datům ze sociálních sítí.

---

<sup>6</sup> 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>.

<sup>7</sup> 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>

<sup>8</sup> 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>

Komplementárně k tomu přístupu stavíme analytický aparát vymezený Ervingem Goffmanem. Jeho klíčovou metaforou jsou pojmenování převzatá nebo odvozená z prostředí divadla. Svou teorii formuloval především v knize *The Presentation of Self in Everyday Life*<sup>9</sup>, kde vymezil několik základních pojmů, které slouží k uchopení lidské komunikace a interakce.

Pojmem **představení** označuje „souhrn všech činností, které jednotlivec provádí v době vyznačující se jeho trvalou přítomností ve společnosti konkrétního souboru pozorovatelů a která má pozorovatele nějaký vliv“.<sup>10</sup> 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.“<sup>11</sup>

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 rasa či pohlaví, jiné se rychle mění. Fasády si podle Goffmana spíše vybíráme, nikoli vytváříme.

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 zákazníci 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

---

<sup>9</sup> Dále citujeme dle českého překladu GOFFMAN, Erving. Všichni hrajeme divadlo: sebereprezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1.

<sup>10</sup> GOFFMAN, Erving. Všichni hrajeme divadlo: sebereprezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1. str. 29.

<sup>11</sup> GOFFMAN, Erving. Všichni hrajeme divadlo: sebereprezentace v každodenním životě. Praha: Nakladatelství Studia Ypsilon, 1999. ISBN 80-902-4824-1. str. 29.

zákulisí: místo, které se sice vztahuje k nějakému představení, ale dojem vyvolaný výkonem je tam vědomě popírán.

Oba teoretické přístupy pak tvoří základní východisko původního návrhu této práce: jak určit vzdálenosti mezi skupinami na sociálních sítích. Postup je odvozen od již zmíněné koncepce tzv. Normalized Information Distance, které se věnuje **část třetí**. Vycházíme přitom z tzv. Kolmogorovy teorie komplexity, která se snaží odpovědět na otázku „Co je nahodilý objekt?“. Michal Koucký uvádí následující příklad (volně dle jeho článku A Brief Introduction to Kolmogorov Complexity).<sup>12</sup> 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 Ludolfovo číslo a v případě C pak řetězcem 84354279521... a tak dále, 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 komplexnější. 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.

---

<sup>12</sup> 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>.

Teorie informační vzdálenosti představuje rozšíření Kolmogorovy komplexity o myšlenku vzdálenosti, resp. podobnosti mezi řetězci. 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 tohoto vzorce autoři Paul Vitányi a Rudi Cilibrasi odvozují obecnou normalizovanou informační vzdálenost (Normalized Information Distance), která by byla schopna produkovat i metrickou vzdálenost.

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.“<sup>13</sup> Obvykle se snaží algoritmicky nalézt opakující se sekvence znaků a vytvořit z nich slovník, který umožňuje odkaz na přesné místo, čímž se zmenší celková 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.

Protože NCD vyjadřuje blízkost (podobnost) založenou na syntaktické vrstvě, navrhli její autoři i její variantu založenou na myšlence podobnosti mezi kompresními algoritmy a internetovými vyhledávací. Podle Cilibrasiho každý kompresní algoritmus „only looks for

---

<sup>13</sup> 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](https://cs.wikipedia.org/wiki/Kompresi_dat).

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  $\pi$ .“<sup>14</sup> Symetricky je tak možné chápat výskyt slov na stránkách jako případ univerzální distribuce slov na stránkách a vyhledávač jako určitý druh pseudokompresního algoritmu, který zohledňuje všechny dimenze lidského mínění zachycené v dokumentech indexovaných vyhledávačem.

My tuto myšlenku rozšiřujeme dále o rozměr umožňující formalizovat vzdálenosti sociálních skupin v prostředí sociálních sítí, a to na základě jednání jejich uživatelů. A obracíme se tak k rovině pragmatické. Z perspektivy předchozích kapitol se lze totiž 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 sebe prezentace jedince. Právě to, že je součástí sebe prezentace 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 jedinec plně popsitelný 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.

**Čtvrtou část práce tvoří tři případové studie, které se snaží prozkoumat stabilitu a validitu postulované metriky. Prvním příkladem je rozsáhlá analýza publik tzv. antisystémových stránek na Facebooku v porovnání s publikem médií hlavního proudu. Na ní navazuje analýza porovnávací korelace výsledků voleb do Evropského parlamentu v různých zemích v porovnání s distančními maticemi reprezentujícími vzájemné vzdálenosti publik politických stran na Facebooku. Konečně pak třetí studie se věnuje porovnání distančních matic publik stejných celebrit mezi jednotlivými sociálními sítěmi pomocí tzv. Mantel testu. Všechny**

---

<sup>14</sup> CILIBRASI, Rudi Langston. Statistical Inference Through Data Compression: ILLC dissertation series. 1. Amsterdam: Lulu.com, 2007. ISBN 9061965403., str. 111.



případové studie naznačují, že navržený přístup je stabilní a minimálně prakticky použitelný pro specifická měření.

V **Závěru** se debata obrací k hranicím samotného přístupu a perspektivě, kterou tento výzkum otevírá. Pro potřeby našeho výzkumu je totiž člověk v zásadě popsateľný pomocí databáze, která v čase zaznamenává 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. Lidské postoje, hodnoty a podobně lze pak 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.

**Appendix** obsahuje krátký popis počítačových programů, které vznikly jako součást této disertační práce. Zdrojový kód ke všem programům je k dispozici na serveru Github. Jedná se o knihovny pro jazyk R, které zjednodušují přístup k API rozhraním některých sociálních sítí (Sugar, Geosugar) a nástroje pro prohlížení dat získaných při následném dataminingu sociálních sítí (Who are likers?, Page Affinity Analyzer, Normalized Social Distance Matrix Viewer a Normalized Facebook Distance Matrix Viewer.)

## Seznam použité literatury a dalších zdrojů (výběr)

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.

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>.

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.

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>.

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/>.

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/>.

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.

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.

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.

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>.

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](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.

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>.

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.

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.

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>.

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>.

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.

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

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

# Přehled odborné, akademické a publikační činnosti

## **Semináře vedené na Ústavu informačních studií a knihovnictví 2011–2019 (výběr)**

Dezinformace, propaganda a falešné zprávy

Digital Humanities

Informační věda a srovnávací mediální studia

Informační věda I

Open-source investigation and intelligence

Rko česky pro prváky stunome

Studia nových médií II

Digital Humanities as a Key For Research In Digital Era

## **Publikační činnost 2011–2019 (výběr)**

Odborné články

ŠLERKA, Josef a Jakub FIALA. 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.

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](http://link.springer.com/10.1007/978-3-319-65566-6_3).

ŠISLER, Vít., Švelch Jaroslav., ŠLERKA, Josef.: Video Games and the Asymmetry of Global Cultural Flows: The Game Industry and Game Culture in Iran and the Czech Republic. International Journal of Communication, 2017, č. • no. 11, s. • p. 3857-3879. ISSN 1932-8036.



ŠLERKA, Josef.: Patos distance jako cíl rétorické strategie. In Puc J., Fridmanová M.: Filosofie lidského, příliš lidského. Červený Kostelec, Pavel Mervart, 2011, s. • p. 147-158. ISBN 978-80-87378-86-1.

#### Popularizační články

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](http://www.dingir.cz/cislo/15/3/nova_a_socialni_media.pdf).

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>.

#### Blogposty

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. 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 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. 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>.

## Účast na konferencích (výběr)

Birds of a feather flock together: An Empirical Study of Czech journalist on Twitter

Cyberspace, Masarykova univerzita

30. 11. 2018, Brno

Stručná zpráva o jednom experimentu

IAC

5. 4. 2018, Praha

McLuhan, Fitbit a (ne)smrtebná zpětná vazba

New Media Inspiration

3. 3. 2018, Praha

Wikipedie ve službách zla?!

Wikikonference

2. 12. 2017, Praha

Facebook data mining – case study

Ludwig-Maximilians-Universität München

20. 6. 2017, Mnichov

Malý velký svět bublin na Facebooku

New Media Inspiration

18. 2. 2017, Praha

Jak se žije v Matrixu

Digital Rulleez

23. 11. 2016, Bratislava

New Media, Big Data, IoT aneb jak to je?!

Microsoft Czech Republic

27. 11. 2015 Praha

Svět viděný kompresním algoritmem

Vytváření vědění: big data, data mining a sociální sítě, Univerzita Karlova

3. 11. 2015, Praha