

naše společnost

our society



09/ **Tři generace metod
měření korupce
v Evropě**

KRISTÝNA BAŠNÁ

35/ **Homogamie
a internet:
Pomáhají online
seznamky rozumět
výběrovému
párování?**

MARKÉTA ŠETINOVÁ

57/ **Velká data jako
alternativa
výběrových šetření
v kvantitativním
sociálněvědním
výzkumu**

JOHANA CHYLÍKOVÁ



Sociologický ústav
Akademie věd ČR



ČSDA

Časopis Naše společnost je odborným periodikem Centra pro výzkum veřejného mínění vydávaným od roku 2003 Sociologickým ústavem AV ČR, v.v.i., od roku 2017 ve spolupráci s Českým sociálněvědním datovým archívem. Počínaje rokem 2007 je časopis recenzovaný.

R E D A K C E

Martin Ďurďovič, šéfredaktor, Martin Vávra, redaktor

R E D A K Č N Í R A D A

Ilona Bažantová (PrF UK, Praha), Adéla Gjuričová (ÚSD AV ČR, Praha), František Kalvas (FF ZČU, Plzeň), Jindřich Krejčí (SOÚ AV ČR a FSV UK, Praha), Jiří Šubrt (FF UK, Praha), Tomáš Trampota (FSV UK, Praha), Jiří Vinopal (SOÚ AV ČR a FF UK, Praha), Milan Zeman (SÚ SAV, Bratislava), Lukáš Novotný (FF UJEP)

Z A M Ě Ř E N Í Č A S O P I S U

Stati vycházejí primárně z výzkumů veřejného mínění a jeho výsledků, a to s důrazem na širší mezioborové souvislosti a aktuálnost témat. Příspěvky čerpají z dat kontinuálních šetření CVVM i jiných výzkumů. Časopis přijímá rovněž náměty, které se opírají o paradigmatu nabízená sociologií, demografií, politikou, ekonomikou, mediálními studii, soudobými dějiny a jinými příbuznými disciplínami, včetně metodologie sociálního výzkumu. Vydávání rubrik Metodologie sociálního výzkumu a Data a datové zdroje je podpořeno z projektu MŠMT Český sociálněvědní datový archiv (kód LM2015060).

Časopis vychází dvakrát ročně v nákladu 400 výtisků.

Naše společnost online: <http://www.cvvm.soc.cas.cz>

K O R E S P O N D E N C I Z A S Í L E J T E N A A D R E S U :

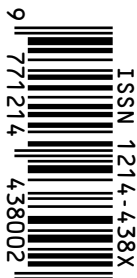
CVVM, SOÚ AV ČR, v.v.i., Jiřská 1, 110 00 Praha 1 nebo na e-mail redakce.ns@soc.cas.cz

U K Á Z K A C I T O V Á N Í Č L Á N K Ů Z T O H O T O Č Í S L A :

Příjmení, Jméno. 2018. „Název článku.“ *Naše společnost* 16 (1): pp-pp, <http://dx.doi.org/10.13060/1214438X.2018.116.xxx>

Praha 2018

ISSN 1214-438X (Print), ISSN 2336-1646 (Online)



naše společnost
our society

editorial

Vážené čtenářky, vážení čtenáři,

nenechte se ničím zmást – ano, opravdu máte před sebou další číslo časopisu *Naše společnost*. Jak je patrné na první pohled, v letošním roce jsme se odhodlali k další inovaci.

Navazujeme tím na kroky započaté v letech předchozích. Vzpomeňme, že předloni v časopisu došlo k rozšíření podílu anglicky psaného textu v odborných statích (rozšířený abstrakt, překlady grafů a tabulek), což bylo vedeno snahou zpřístupnit obsah těchto statí alespoň zčásti zahraničnímu čtenáři. Loni byl na základě spolupráce s Českým sociálněvědním datovým archivem časopis rozšířen o pravidelnou rubriku zaměřenou na odborné statě k metodologii sociálního výzkumu a od stejné doby jsou součástí časopisu také rubriky s texty neprocházejícími recenzním řízením. Letošní rok přináší změny ve vzhledu časopisu.

První výrazná změna se týká zmenšení formátu. To je nepřehlédnutelné především u tištěné verze. Zmenšením formátu přibližujeme *Naší společnost* mezinárodně běžným formátům odborných sociálněvědních časopisů. Další změny úzce souvisí s první. Zmenšení formátu jednak podnítilo vznik nového motivu obálky, jednak si vyžádalo změnu grafické úpravy textu, tabulek, grafů a dalších položek. Snažili jsme se *Naší společnosti* vdechnout graficky současnější vzezření, které by vyhovovalo čtenářům a hospodárně využívalo prostor jednotlivých stránek. Doufáme, že ve svém souhrnu tyto změny přispějí k tomu, že četba časopisu a práce s ním budou ještě pohodlnější a příjemnější, než tomu bylo doposud.

Pokud jde o obsah tohoto čísla, autorky odborných statí v něm prezentují témata, která jsou neotřelá, aktuální a chytlavá: měření korupce, seznamování přes internet a využití tzv. velkých dat (big data). Převážně přehledový charakter jejich statí uvádí čtenáře do problematiky. Všem třem je ale zároveň společné, že reflektují nejnovější trendy, analyzují je a zaujímají k nim stanovisko. V českém kontextu tak načrtávají některé nové obzory.

První stať Kristýny Bašné shrnuje dosavadní vývoj přístupů k měření korupce, popisuje tři „generace“ těchto přístupů a na základě jejich korelační analýzy zkoumá, nakolik měří totožný jev. Dotýká se také otázky posunu k objektivnějšímu měření, které kombinuje vnímání korupce, zkušenost s ní a „tvrdá data“ dovolující usuzovat na existenci korupce. Druhá stať Lenky Šetinové se zabývá seznamováním přes internet. Lze se v ní dočíst, jaký dopad mají hledání partnera přes internet a jeho výzkum pro teorie výběrového párování. Pozornost je věnována tomu, zda seznamování přes internet homogamii spíše tlumí, nebo spíše posiluje. Metodologická stať Johany Chylíkové na jednu stranu objasňuje koncept velkých dat a možnosti jejich využití v sociálním výzkumu, na druhou stranu poukazuje na limity tohoto typu dat a na výzvy spojené s jejich zpracováním a analýzou. Čtenář si tak může udělat ucelený obrázek o tomto

interdisciplinárním tématu, které je v posledních letech skloňováno nejen sociálními a technickými vědci, nýbrž také politiky (viz např. dokumenty související s národní iniciativou Průmysl 4.0).

Děkuji všem, kdo se na přípravě tohoto čísla podíleli, a za redakci časopisu přeji zajímavé čtení!

Martin Ďurđovič,
šéfredaktor

obsah

ODBORNÉ STATI

09/ **Tři generace metod měření korupce v Evropě**

KRISTÝNA BAŠŇÁ

35/ **Homogamie a internet: Pomáhají online seznamky rozumět výběrovému párování?**

MARKÉTA ŠETINOVÁ

ODBORNÉ STATI K METODOLOGII SOCIÁLNÍHO VÝZKUMU

57/ **Velká data jako alternativa výběrových šetření v kvantitativním sociálněvědním výzkumu**

JOHANA CHYLÍKOVÁ

RECENZE

81/ **Recenze knihy: Niels Brügger, Ralph Schroeder (eds.). 2017. *The Web as History: Using Web Archives to Understand the Past and the Present*. London: UCL Press, 296 s.**

MARTIN VÁVRA

DATA A DATOVÉ ZDROJE

85/ **Data z výzkumů ČT dostupná v Českém sociálněvědním datovém archivu**

MARIE POSPÍŠILOVÁ

Tři generace metod měření korupce v Evropě¹

Kristýna Bašná, Sociologický ústav AV ČR, v.v.i.

Three Generations of Methods for Measuring Corruption in Europe

EXTENDED ABSTRACT:

This article focuses on methods for measuring corruption, first describing three generations of corruption indicators and then comparing them qualitatively and quantitatively. Corruption is a clandestine activity that is extremely difficult to measure; there are no official statistics on the number of corruption cases. For this reason, corruption can only be measured indirectly, by various proxies, and it is extremely hard to state whether these indicators are reliable and indeed measure the corruption phenomena in a given country. A large number of different indicators have been developed over the years that try to capture and quantify corruption. Some authors measure perceptions of corruption, others try to use “hard data” to explore the level of corruption in a country, and even others combine different measurements, weight them, and then publish composite indicators to capture the overall level of corruption in a country.

This article aims to evaluate the quality of the different indicators using quantitative and qualitative methods. Possible uses and value of each individual indicator are discussed in terms of quality and practical considerations. First-generation indicators combine expert evaluations, surveys, and other data on corruption. The article focuses on the two best-known and most frequently used indicators – the Control of Corruption measurement by the World Bank and the Corruption Perception Index by Transparency International. Second-generation indicators are based on opinion surveys that ask respondents about their perceptions of and own experience with corrup-

KEYWORDS:

corruption
surveys
methods, quantitative

DOI LINK:

<http://doi.org/10.13060/1214438X.2018.116.418>

¹ Tento článek byl podpořen grantem Ministerstva školství, mládeže a tělovýchovy, číslo LM2015066.

tion. There are two types of such surveys, one focusing on the public and the other one on businesses. The second-generation indicators include surveys such as Eurobarometer, World Values Survey (WVS), European Social Survey (ESS), Global Corruption Barometer (GCB), WB BEEPS, or International Social Survey Programme (ISSP). Finally, the third generation of corruption measures is based on “hard data” and usually focuses on a specific sector in a specific country.

After a descriptive presentation of all three generations of indicators, the quality of these indicators is discussed from the quantitative and qualitative perspectives. The author describes the time and country coverage of each indicator, the different wordings and definitions used in the questionnaires, the different target groups sampled, and also the possible implications of different methods for the results. Finally, the author presents a correlation analysis of different indicators over time and across countries, indicating whether and how strongly these three generations correlate with one another. The article concludes with a discussion on whether the three generations of indicators measure the same type of corruption and whether they can be used interchangeably. As for results, it seems that there are statistically significant associations between different types of indicators; however, this significance is, in most cases, not high enough to allow for interchangeability. Therefore, it is of utmost importance to carefully select the type of indicator used for scientific research, as results and conclusions might differ significantly depending on the indicator used.

Korupce je jedním z nejaktuálnějších a nejpalcivějších problémů České republiky. Česká republika se bohužel každoročně umísťuje na spodních příčkách mezinárodních průzkumů zjišťujících percepci korupce (Transparency International – TI), ale také v průzkumech tázajících se na reálné setkání s korupčním jednáním (World Bank či Eurobarometer). Například podle TI se v roce 2014 Česká republika umístila na 22. místě z 28 států Evropské unie a na 53. místě ze 175 států světa. Podobnou situaci jako v České republice můžeme najít ve více evropských státech, hlavně postkomunistických nebo ve státech jižní Evropy. Není proto překvapivé, že boj proti korupci si mnoho politických stran klade do volebních programů a že snaha o snížení korupce je často mezi prioritami nových vlád. Nutnost boje proti korupci dokládá i to, že podle mnoha autorů korupce snižuje kvalitu státní správy v řadě oblastí, důvěru občanů v politické instituce i důvěru lidí mezi sebou, zpomaluje vývoj občanské společnosti nebo může dokonce vyvolávat občanské nepokoje [Brown et al. 2011; Griesshaber, Geys 2011; Pellegata 2012]. Korupce však není nebezpečná pouze z pohledu společnosti a její nefunkčnosti, ale podle mnoha autorů také například brání ekonomickému růstu [Kaufmann et al. 1999]. K úspěšnému boji proti korupci je prvním krokem určení rozsahu, výše a závažnosti korupce ve státě, je tedy velmi důležité korupci správně změřit.

Existuje rozsáhlá debata o tom, jak měřit korupci a zda je vůbec možné ji změřit. Korupce je ilegální činnost a nejsou k dispozici žádné robustní oficiální statistiky o počtu případů korupce². Na rozdíl od většiny ostatní trestné činnosti není v případě korupce motivace hlásit případy na policii. Obě strany, které se na korupci podílejí, mají motivaci tuto činnost skrývat a třetí strana, která ztratila peníze nebo moc kvůli korupci, většinou ani neví, že došlo k zločinu. Tato skrytá povaha korupce zásadně ztěžuje výzkum o korupci, díky ní neexistují žádné přímé způsoby, jak ji měřit. Nicméně existuje několik nepřímých způsobů, jak dostat informace o úrovni korupce v zemi [Tanzi 1998]. Avšak z důvodu, že jsou tyto informace nepřímé, zůstává nejasné, zda tyto metody měří skutečnou úroveň korupce, nebo zda měří rozličné formy vnímání korupce ovlivněné vnějšími faktory. Bohužel tato otázka nejspíše nebude nikdy zodpovězena, protože neexistuje a s velkou pravděpodobností nikdy nebude existovat způsob měření reálné míry korupce.

Tento článek předkládá přehled nejpoužívanějších existujících indikátorů porovnávacích evropské země a představuje dělení indikátorů na první, druhou a třetí generaci. Mezi indikátory první generace je zařazena Control of Corruption (CC) od World Bank (WB) a Corruption Perception Index (CPI) of Transparency International (TI). Do druhé generace jsou zařazeny průzkumy veřejného mínění a průzkumy firem, v tomto článku jsou to průzkumy European Social Survey (ESS), World Values Survey (WVS), Eurobarometer, Global Corruption Barometer (GCB) a International Social Survey Programme (ISSP). Firemní průzkumy zahrnuté do tohoto článku jsou The Business Environment and Enterprise Performance Survey (BEEPS) a Eurobarometer. Konečně, do třetí generace je zahrnut Index of Public Integrity (IPI). Je důležité si uvědomit, že se můžeme pouze přibližovat k reálnému číslu korupce, ale že žádný z indikátorů neměří reálnou korupci. Je také nemožné činit závěry, který indikátor je blíže reálné míře korupce, můžeme pouze posuzovat kvalitu indikátorů mezi sebou. Článek tudíž diskutuje výhody a nevýhody každého indikátoru a představuje jejich porovnání. Článek uzavírá diskuzi nad tím, zda indikátory tří generací měří stejný fenomén a zda by popřípadě byla možná jejich záměna.

Definice korupce

Abychom mohli korupci měřit, je velmi důležité ji definovat. Přestože existuje mnoho různých definic korupce, není žádná, která by byla obecně přijímaná vědeckou komunitou či mezinárodními institucemi [OECD 2008; United Nations 2004]. Korupci lze pozorovat na vysoké (politické) úrovni (tzv. grand corruption), kde se na korupci podílejí politici a mezinárodní hráči, na

² Samozřejmě na úrovni každého státu existují statistiky o počtu odsouzení za korupci, velmi kvalitní a detailní statistiky existují i v České republice. Mohlo by se zdát, že tyto statistiky měří reálnou míru korupce, ale bohužel reálný počet případů korupce a počet odsouzení či stíhání za korupci spolu vůbec nemusí korelovat. V zemích, kde policie a soudy fungují kvalitně, může být stíhána a souzena většina případů korupce, počet odsouzení tedy může být relativně vysoký. Naopak v zemi, kde policie a soudy fungují špatně, může být počet odsouzení velmi nízký, i když korupce v zemi bují. Oficiální statistiky počtu lidí stíhaných a odsouzených za korupci tedy mohou být velmi zavádějící.

střední úrovni (firmy) nebo konečně na úrovni jednotlivců (tzv. petty corruption), kde se na korupci podílí veřejnost a běžní občané. Existuje také několik různých forem zneužívání moci ze strany veřejných činitelů, jako je uplácení, podvádění, střet zájmů, manipulace, nepotismus, klientelismus nebo patronáž [Heidenheimer, Johnson 2002: 27]. Korupce může být definována zákonem, firemním kodexem nebo veřejností. Každá definice nebo vnímání korupce se může lišit v závislosti na expozici, zapojení a zkušenosti s korupcí. Samotné chápání korupce může být, a velmi pravděpodobně je, kulturně podmíněné, a tudíž nesmírně různorodé v různých zemích, což ještě více znesnadňuje definici korupce [Charron et al. 2013: 22]. To, co se chápe jako korupční chování v jedné zemi, může být považováno za standardní chování v jiné zemi³, a z tohoto důvodu se článek zaměřuje pouze na korupci v rámci Evropy, kde se díky společné historii a relativně homogenní kultuře dá předpokládat podobné vnímání a porozumění fenoménu korupce.

Podobný problém v nejednoznačnosti definice můžeme najít i v indikátorech měřících korupci. V literatuře se nejčastěji používá definice, že korupce je „zneužití svého postavení pro soukromý zisk“ [Treisman 2000; World Bank 1997]. Tato definice je totožná s tou, kterou používá TI pro své CPI [Transparency International]. WB definuje účely pro CC pomocí definice, která je také velmi podobná již zmíněné: „CC zachycuje vnímání toho, do jaké míry je postavení zneužito pro soukromý zisk, toto zahrnuje tzv. grand korupci i petty korupci, stejně tak jako ‚uchvácení‘ státu elitami a soukromými zájmy“ [World Bank].⁴ Průzkumy veřejného mínění většinou korupci nijak nedefinují, nechávají na respondentech, co si pod pojmem „korupce“ představí. Jediná výjimka je Eurobarometer, který dává respondentům instrukce: „Od teď, pokud zmíníme korupci, myslíme tím korupci v širokém slova smyslu, jako nabídnout, dát, požadovat či obdržet úplatek, provize, hodnotný dar či důležité laskavosti“⁵ [European Commission 2014]. Je patrné, že nejen vědecká komunita, ale ani indikátory se neshodnou na jedné definici korupce, srovnání různých indikátorů je proto problematické už z tohoto důvodu.

Možnosti výzkumu korupce

Jak již bylo řečeno v úvodu, neexistuje možnost, jak měřit korupci přímo, korupce je skrývaná činnost. Měření korupce prošlo velmi dlouhým vývojem, první výzkumy snažící se konzistentně zachytit korupci jsou datovány již do poloviny 90. let 20. století. Dnes je indikátorů měřících korupci celá řada, ale objevuje se zde více metodologických proudů, které rozdělují indikátory

³ Pro různé teorie a výzkumy týkající se této problematiky viz [Bukovansky 2006; Heidenheimer, Johnson 2002; Rothstein, Torsello 2013].

⁴ “CC Captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as “capture” of the state by elites and private interests.”

⁵ “From now on, when we mention corruption, we mean it in a broad sense, including offering, giving, requesting and accepting bribes or kickbacks, valuable gifts and important favours, as well as any abuse of power for private gain.”

na různé typy neboli generace. Dělení na určité generace indikátorů měřících korupci je možné podle více kritérií, pro účely tohoto článku budu používat dělení podle Graycara a Smithe [2011]. Graycar a Smith dělí indikátory do tří generací; první generace zahrnuje kompozitní indikátory, které jsou založené především na názorech expertů. Kompozitní indikátory jsou sice dodnes nejpoužívanější metodou studia korupce, ale zároveň jsou silně kritizovány z důvodu netransparentní a nepřesné metodologie.

V odpovědi na tuto kritiku začaly některé instituce publikovat průzkumy veřejného mínění a průzkumy firem, kde se respondentů dotazují na přímou zkušenost s korupcí a na jejich názory na rozšířenost korupce, tyto průzkumy se řadí mezi indikátory druhé generace. Přímá zkušenost s úplatky sice může být objektivní metodou zjišťování prevalence korupce, jsou zde ale také zásadní problémy. Zkušenost s korupcí u firem a občanů ukazuje pouze malý výsek korupce a navíc je korupce citlivé téma, u kterého mohou respondenti lhát. Percepce korupce zase může být u respondentů ovlivněna médii či společnostmi. Vzhledem k tomu, že ani jedna metoda není schopna zachytit korupci reliabilně, objevují se v posledních letech indikátory třetí generace.

Tyto indikátory nemají za cíl popsat korupci v celém jejím spektru, ale spíše se snaží o nalezení tvrdých dat v určité oblasti korupce. Jde tedy o to, pomocí již existujících dat zachytit buď existující korupci v této oblasti, nebo riziko korupce. Indikátory třetí generace představují novou metodu měření korupce, tyto indikátory jsou tudíž zatím omezeny na určitou oblast v určité zemi a možnosti mezinárodního srovnávání jsou zatím velmi omezené (ale první pokusy se již objevují, jak je blíže popsáno níže). Tabulka 1 ukazuje tyto tři generace měření korupce a nejpoužívanější zástupce těchto generací. Dále vyznačuje výhody a nevýhody každého jmenovaného postupu. V následujících kapitolách si každou možnost zkoumání a měření korupce rozebereme.

Tabulka 1: Možnosti měření korupce

Typ výzkumu	Příklady	Výhody	Nevýhody
První generace	kompozitní indikátory (CC, CPI)	nejstarší, zahrnují nejvíce zemí	nejasné, netransparentní metody, nízká validita
Druhá generace	průzkumy veřejného mínění (Eurobarometer, ESS, ISSP, GCB, WVS)	zjišťují reálnou zkušenost lidí s úplatky	nebezpečí lhaní, vyšší odmítnutí odpovídat, měří pouze úplatky
	firemní výzkumy (BEEPS, Eurobarometer)	relevantní, objektivní, měří reálnou zkušenost	pouze jeden výsek korupce, nebezpečí lhaní
Třetí generace	IPI	„tvrdá“ data, měří reálný risk korupce	složité mezinárodní srovnávání, často jeden výsek korupce

Table 1: Corruption measurement options

Type of study	Examples	Advantages	Disadvantages
First generation	Composite indicators (CC, CPI)	Oldest, most countries covered	Unclear and non-transparent methods, low validity
Second generation	Public opinion polls (Eurobarometer, ESS, ISSP, GCB, WVS)	People's real-life experiences with bribes measured	Risk of lying, higher rejection rate, only bribes measured
	Business surveys (BEEPS, Eurobarometer)	Relevant, objective, measuring real-life experiences	A single segment of corruption, risk of lying
Third generation	IPI	"Hard" data, measuring real risk of corruption	Complicated international comparison, often a single segment of corruption

První generace

Indikátory první generace se spíše než o měření reálné míry korupce pokoušely vůbec otevřít diskuzi na téma korupce a zvýšit povědomí o tomto problému [Heinrich, Hodess 2011: 29]. Tyto indikátory se objevily v 90. letech 20. století a patří mezi ně hlavně kompozitní indikátory. Kompozitní indikátory se skládají z více zdrojů, které hodnotí míru korupce, jedná se většinou o průzkumy veřejného mínění, expertní hodnocení a podobně. Kompozitní indikátory jsou nejpoužívanějšími indikátory mezi vědci, kteří zkoumají empiricky korupci [Kapoor and Ravi 2012; Treisman 2000; Treisman 2007; Uslaner 2009; Zakaria 2013], a použití této metody jako prostředku hodnocení zemí stále rychle roste [Bandura 2008]. Autoři kompozitních indikátorů se snaží zachytit co nejvíce pohledů a názoru na výši korupce v dané zemi s vírou, že po agregování těchto percepce vznikne indikátor, který se reálné míře korupce v zemi alespoň přiblíží [Heller 2009]. Nejznámější kompozitní ukazatele jsou Corruption Perception Index (CPI) od Transparency International (TI) a Control of Corruption (CC), která je součástí Good Governance indikátorů, zveřejňovaných World Bank (WB). CPI Transparency International je starší, datuje se do roku 1995, zatímco indikátor CC byl poprvé publikován v roce 1996.

Kompozitní indikátory pokrývají široký výběr zemí a jsou prováděny každoročně, což je mnohem častěji než jakékoliv indikátory druhé či třetí generace. Někteří autoři také poukazují na výhodu, že díky používání více zdrojů je chyba měření jednoho zdroje minimalizována [Charron et al. 2013]. Nicméně i když jsou kompozitní indikátory široce používány v oblasti výzkumu korupce, čelí také velké kritice.

Je zde například obava, že kompozitní indikátory měří spíše předsudky a stereotypy o korupci než reálnou míru korupce [Donchev, Ujhelyi 2014; Treisman 2007; Treisman 2013]

nebo že CC měří spíše názor podnikatelských elit [Rohwer 2009] a CPI zase privilejuje názory expertů pracujících v mezinárodních firmách či institucích [Heywood 1997]. Je také velmi nepravděpodobné, že kompozitní indexy měří tzv. grand korupci, protože odborníci většinou nemají přímou zkušenost s korupcí na této úrovni [Fazekas et al. 2013]. Tento typ dat je publikován například institucí Freedom House⁶ nebo s kupinou PRS, která publikuje International Country Risk Guide (ICRG)⁷. Experimentální posouzení se může zdát jako validní metoda měření úrovně korupce, nicméně i zde existují zásadní problémy. Například tato metoda je méně transparentní, protože přesné metodiky pro posouzení míry korupce nejsou veřejně k dispozici [Mungiu-Pippidi 2015]. Navíc odborníci obvykle mají určitou představu o tom, jak různé další ukazatele hodnotí korupci v dané zemi, je zde tudíž nebezpečí, že se odborníci těmito ukazateli inspiroují (free-riding) [Charron et al. 2013: 43; Treisman 2000]. Free-riding může tím pádem zásadně snižovat interní validitu kompozitního indikátoru, jenž by se v tomto případě spoléhal sám na sebe [Malito 2014]. Dále kompozitní ukazatele používají primárně jako zdroje nereprezentativní průzkumy, a proto jsou náchylné k chybám, které mohou být umocněné malou velikostí vzorku [Golden and Picci 2005]. A konečně – další argument proti používání kompozitních ukazatelů je ten, že se téměř každý rok spoléhají nejen na rozdílné zdroje, ale i na rozdílný počet zdrojů, srovnání napříč roky tedy může být velmi problematické.

Jak bylo řečeno, nejznámějšími a nejpoužívanějšími indikátory první generace jsou Corruption Perception Index (CPI) od Transparency International (TI) a Control of Corruption (CC) od World Bank (WB). TI používá škálu 0–100 (0–10 před rokem 2012), 0 představuje velmi vysokou mírou korupce a 100 (10 před rokem 2012) velmi nízkou mírou vnímané korupce. V roce 2012, po procesu vnitřní i vnější kontroly [Transparency International 2012b], TI změnila metodiku výpočtu i výslednou škálu, tato metoda je nyní replikována každý rok. Indikátor CC od WB se pohybuje od –3 do 3, kdy země s hodnotou 3 má nejvyšší kontrolu nad korupcí. S výjimkou metodických rozdílů⁸ existují také rozdíly ve zdrojích, které tyto kompozitní indikátory používají. WB využívá více zdrojů než TI, například v roce 2008 TI použilo 11 zdrojů a WB použila nejen stejných 11 zdrojů, ale i dalších 14, které nebyly používány TI [Rohwer 2009]. Navíc Rohwer poukazuje na to, že WB používá jako zdroje názory odborníků jak ve veřejném, tak i soukromém sektoru, TI na druhou stranu zahrnuje pouze názory expertů ve veřejném sektoru [2009].

Další specifický rozdíl mezi indikátory CC a CPI je důležitý především pro vědce, kteří analyzují vývoj korupce v čase. Jak bylo zmíněno, TI změnila svou metodiku v roce 2012, a proto údaje před tímto rokem nejsou srovnatelné v čase a TI sama odrazuje vědce od porovnání jejich indikátoru v čase [Transparency International 2012a]. Na druhou stranu WB používá stejnou metodiku od počátku ukazatelů Good Governance, a indikátor CC proto může být používán s opatrností při analýze vývoje korupce v čase [Kaufmann et al. 2010]. Nicméně jak

⁶ <https://freedomhouse.org/report/nations-transit/nations-transit-2015#.VtBQp33hBiw>

⁷ <https://www.prsgroup.com/about-us/our-two-methodologies/icrg>

⁸ Rozdíly v metodice těchto indikátorů nemají dramatický vliv na rozdíly ve výsledcích indikátorů a v hodnocení států, tento článek se proto těmito rozdíly blíže nezabývá. Pro podrobnější informace k metodice CC viz např. [Kaufmann et al. 2004] a CPI viz [Lambsdorff 2001]. Pro podrobnější analýzu rozdílů mezi CPI a CC viz Malito [2014] nebo Rohwer [2009].

Tabulka 2: Počet zemí v CC a CPI a vzájemné korelace indikátorů

Rok	CPI	CC	Korelace
1995	18	–	–
1996	20	39	0,9518
1998	29	39	0,9502
1999	35		
2000	33	39	0,9829
2001	32		
2002	34	39	0,9643
2003	38	40	0,9742
2004	39	40	0,9804
2005	39	40	0,9852
2006	39	40	0,9862
2007	40	40	0,9876
2008	40	40	0,9892
2009	40	40	0,9916
2010	40	40	0,9946
2011	40	40	0,9913
2012	40	40	0,9921
2013	40	40	0,989
2014	40	40	0,988

Zdroj: WB, TI. V této tabulce jsou zahrnuty pouze evropské země. Všechny korelace jsou signifikantní na $p < 0,001$.

Tabulka 3: Zástupci druhé generace měření korupce

Jméno	Druh evaluace	Roky	Země
Eurobarometer	firemní průzkum	2014, 2015	EU + NO, LI, CH
WB BEEPS	firemní průzkum	1999, 2002, 2005, 2009, 2011–2014	východní Evropa a střední Asie (29 zemí v 2011–2014)
ESS	veřejné mínění	2004, 2010	25 zemí Evropy + Turecko, Izrael
ISSP	veřejné mínění	2006, 2016	30 zemí světa (19 evropských v roce 2006)
Eurobarometer	veřejné mínění	2005, 2007, 2009, 2011, 2013	EU + NO, LI, CH
GCB	veřejné mínění	2004, 2005, 2006, 2007, 2009, 2010/11, 2013, 2016	země celého světa (107 v roce 2013, 30 evropských)
WVS	veřejné mínění	1996–1998	světový průzkum (53 zemí, 27 evropských)

Zdroj: Eurobarometer, WB, ESS, ISSP, GCB, WVS

Table 2: Number of countries covered by the CC and the CPI; correlation between the indicators

Year	CPI	CC	Correlation
1995	18		
1996	20	39	0.9518
1998	29	39	0.9502
1999	35		
2000	33	39	0.9829
2001	32		
2002	34	39	0.9643
2003	38	40	0.9742
2004	39	40	0.9804
2005	39	40	0.9852
2006	39	40	0.9862
2007	40	40	0.9876
2008	40	40	0.9892
2009	40	40	0.9916
2010	40	40	0.9946
2011	40	40	0.9913
2012	40	40	0.9921
2013	40	40	0.989
2014	40	40	0.988

Source: WB, TI. Only European countries are accounted for in the table. All correlations are significant at $p < 0.001$.

Table 3: Examples of second-generation corruption measurements

Name	Type of study	Years	Countries
Eurobarometer	Business survey	2014, 2015	EU + NO, LI, CH
WB BEEPS	Business survey	1999, 2002, 2005, 2009, 2011–2014	Eastern Europe and Central Asia (29 countries in 2011–2014)
ESS	Public opinion	2004, 2010	25 countries of Europe + Turkey, Israel
ISSP	Public opinion	2006, 2016	30 countries of the world (19 European in 2006)
Eurobarometer	Public opinion	2005, 2007, 2009, 2011, 2013	EU + NO, LI, CH
GCB	Public opinion	2004, 2005, 2006, 2007, 2009, 2010/11, 2013, 2016	Global coverage (107 countries in 2013, 30 European)
WVS	Public opinion	1996–1998	Global coverage (53 countries, 27 European)

Source: Eurobarometer, WB, ESS, ISSP, GCB, WVS

bylo uvedeno výše, oba kompozitní ukazatele občas mění zdroje, takže změny ve výši korupce mohou být spíše v důsledku zahrnutí údajů z nových organizací než ve změnách v reálném vnímání korupce. Kromě toho TI někdy používá stejné průzkumy ve více letech [Treisman 2007], a jak Kaufmann et al. ukazují, asi polovina z rozptylu změny v CC v průběhu času je důsledkem změn ve vážení agregovaných dat a ze změn zdrojů [2002].

Tabulka 2 ukazuje, kolik zemí zahrnuly do svého výzkumu TI a WB, a také korelace mezi indikátory CC a CPI. Jak již bylo řečeno výše, CPI byl proveden častěji v dřívějších letech. Na druhou stranu WB zahrnuje větší počet evropských zemí v dřívějších letech. Nicméně od roku 2007 oba indikátory zahrnují všech 40 evropských zemí analyzovaných v tomto článku a jsou zveřejněny ročně, rozdíly v počtu zemí zde tudíž nejsou. Tabulka 2 také ukazuje korelace mezi CPI a CC. Tyto korelace jsou extrémně vysoké, v některých případech dosahují až 0,99, je zde tudíž vysoká pravděpodobnost, že tyto indikátory jsou vzájemně zaměnitelné.

Druhá generace

Druhá generace indikátorů se snaží o jiné uchopení fenoménu korupce. Oproti sdružování mnoha informací o korupci a publikování jednoho kompozitního indikátoru se snaží o jasnější a přímější měření korupce. Toto měření nabývá dvou podob, a to měření percepce korupce a měření přímé zkušenosti s korupcí, a většinou se provádí ve dvou oblastech, a to dotazováním firem a běžných občanů. Tabulka 3 ukazuje firemní průzkumy a průzkumy veřejného mínění, jimiž se zabývá tento článek.

Průzkumy ve firmách

Průzkumy ve firmách se mohou zdát jako velmi validní metoda pro zkoumání výše korupce, protože firmy se mohou setkávat s úplatky i ve vyšších sférách, je zde tedy potenciál pro zachycení širších projevů úplatkářství. Na druhou stranu průzkumy ve firmách mají i své slabé stránky. Velkým problémem je to, že zástupci firem nemusí odpovídat pravdivě, není jisté, že se firmy přiznají k tomu, že je v jejich odvětví běžné uplácet. Dále je zde problém toho, že tyto průzkumy zachycují pouze jedinou formu korupce, a to úplatkářství. U Eurobarometru je velmi jasné také vidět problém toho, že jen málo firem mělo možnost uplácet v oblastech, na které se průzkum ptal, velikost vzorku je tudíž velmi malá a reprezentativnost nejspíše tristní. Avšak přesto, že není jisté, jak validní jsou firemní průzkumy, mohou nám pomoci vyzpozorovat alespoň určité trendy (jako například u BEEPS).

The Business Environment and Enterprise Performance Survey (BEEPS)

The Business Environment and Enterprise Performance Survey (BEEPS) je průzkum Světové banky a Evropské banky pro obnovu a rozvoj (EBRD). BEEPS se zaměřuje pouze na země východní Evropy a regionu střední Asie⁹ a její m cílem je analyzovat podnikatelské prostředí z více úhlů pohledu. BEEPS bylo proveden v roce 1999, 2002, 2005, 2009 a v letech 2011–2014 a zahrnuje firmy vybrané statistickými metodami výběru, které mají minimálně 5 zaměstnanců na plný úvazek. Počet dotazovaných firem roste v průběhu let; v roce 1999 to bylo 4 100 podniků ve 25 zemích, v roce 2009 BEEPS provedl rozhovory již v téměř 12 000 podnicích ve 29 zemích [EBRD 2010] a konečně v roce 2011–2014 bylo zahrnuto 16 000 podniků ve 30 zemích východní a střední Evropy a Asie [EBRD 2015]. Průzkum BEEPS je organizován jako panelová studie, většina firem je tedy zahrnuta vícekrát a je možné sledovat vývoj u jedné firmy v čase. Tento panel má bohužel některá omezení; například v roce 1999 EBRD neuchovávala informace o dotazovaných firmách, možnost panelového srovnání je tedy až od roku 2002. Také, jak bylo uvedeno výše, data existují (s malými výjimkami) pouze pro východní Evropu a střední Asii, je tedy nemožné srovnávat výsledky ze západní Evropou v čase. Zajímavostí je, že data BEEPS ukazují značné rozdíly v čase, korelace mezi 1999 a 2012–2014 je jen 0,28 (nesignifikantní). Srovnáme-li korelace mezi CPI (TI) nebo CC (WB), korelace na stejných letech a zemích je 0,78 (CPI) a 0,86 (CC).¹⁰

Eurobarometer

Eurobarometer je mezinárodní průzkum veřejného mínění, který je pravidelně prováděn Evropskou komisí už od roku 1973. Eurobarometer zahrnuje různá témata a zaměřuje se na členské země Evropské unie nebo kandidátské země (jako například v případě Candidate Country Eurobarometeru provedeného v roce 2001). Eurobarometer se kromě průzkumu veřejného mínění v oblasti korupce také dvakrát zaměřil na průzkumy firem v rámci tzv. Flash Eurobarometeru¹¹, konkrétně se Eurobarometer dotázal zástupců podniků na přímé zkušenosti s úplatky v roce 2013 a pak znovu v roce 2015. Bohužel průzkum Eurobarometeru má velmi zásadní problémy, které znemožňují použití tohoto průzkumu pro další analýzy. Průzkum Eurobarometeru totiž zahrnul firmy, které byly v kontaktu s veřejným činitelem pouze v několika málo oblastech. Tento úzký výběr odfiltroval většinu podniků, například v roce 2015 pouze 43 % všech společností mělo tento druh kontaktu, v analýze přímé zkušenosti s korupcí tedy zůstalo pouze 3 365 firem na úrovni Evropy. To samozřejmě vedlo k velmi nízkému počtu firem v každé jednotlivé zemi,

⁹ BEEPS zahrnul v roce 2005 i Irsko, Španělsko, Německo, Portugalsko a Řecko jako referenční země.

¹⁰ CC nebyl publikován v roce 1999, je proto použit průměr 1998–2000. Pro 2012–2014 je použit rok 2013.

¹¹ Flash Eurobarometer je rychlý telefonický sběr dat na téma, o které požádá jakákoliv součást Evropské komise. Pro více informací viz http://ec.europa.eu/public_opinion/archives/flash_arch_en.htm.

což extrémně snižuje reprezentativitu a činí tento průzkum mimořádně nevhodným pro analýzu na úrovni jednotlivých zemí. Například v Irsku 13 % (v reálných číslech to je 7 respondentů) manažerů připustilo, že někdo od nich očekával úplatek za posledních 12 měsíců, je však zajímavé, že toto číslo je nula v roce 2015 (respondentů bylo 27). Z důvodu těchto metodologických problémů nezahrnu výzkumy zjišťující postoje a zkušenosti s korupcí mezi představiteli firem ze série Eurobarometer do další analýzy.

Průzkumy veřejného mínění

Průzkumů veřejného mínění, které se týkají korupce, je celá řada. Mezi nejznámější, do kterých je zapojena i Česká republika patří například ESS, což je mezinárodní komparativní výzkum hodnot, postojů a chování. Probíhá v dvouletých intervalech od roku 2002. Dalším zástupcem je GCB, který vydává Transparency International od roku 2004. ISSP je další mezinárodní komparativní výzkum, který byl zahájen již v roce 1985 a pokračuje dodnes. Nové údaje jsou zveřejněny v ročním cyklu a skládají se z různých témat. Dalším průzkumem je WVVS, což je mezinárodní průzkum, který sahá až do roku 1981. Do dnešního dne bylo provedeno šest vln zahrnujících až sto zemí. Poslední průzkum veřejného mínění je Eurobarometer, který již byl popsán výše. Do průzkumů veřejného mínění jsou většinou zařazeny otázky týkající se vnímání korupce, což znamená, že respondenti odpovídají na otázky ptající se na soukromý názor respondenta na rozšířenost korupce. Dalším typem otázek, které se velmi často v průzkumech veřejného mínění objevují, jsou otázky na zkušenost s úplatky. Každý z těchto typů korupce má své výhody a nevýhody a žádný není ideální, i když může posloužit jako vodítko pro další zkoumání korupce.

Percepce korupce

Vnímání korupce je sledováno mnoha průzkumy veřejného mínění. Největším problémem tohoto typu otázky je to, že není jisté, zda percepce korupce měří reálnou míru korupce. Je zde možnost, že lidé mohou mít určitou, relativně správnou, představu o tom, jak je korupce v jejich zemi rozšířená, ale je zde také reálná možnost, že vnímání korupce je daleko od reality [Rose, Peiffer 2012] a názory lidí na toto téma mohou být ovlivněny mnoha dalšími faktory, například ze strany médií [Golden, Picci 2005]. Pokud média hodně často o korupci píšou, protože vychází najevo mnoho korupčních případů, může to být tím, že policie pracuje velmi efektivně a je schopná korupční případy stíhat. Navíc svobodný tisk podává zprávy o korupčních kauzách bez omezení, zatímco v autokratických režimech se může veřejnost o případech korupce dozvědět pouze zřídka z důvodu cenzury tisku, přičemž to nemusí znamenat, že v demokratických režimech je vyšší korupce. Jak Karklins dodává, tento nepoměr může uvést lidi do mylného přesvědčení, že autoritářské režimy jsou méně zkorumpované než demokratické, a lidé poté mohou podporovat autoritářštější formu vlády či pociťovat nostalgii po předchozí diktatuře, jak je tomu často v některých segmentech postkomunistické společnosti [Karklins 2005: 7]. Vnímání korupce může být ovlivněno také

aktuální nespokojeností s politickou nebo hospodářskou situací [Kurtz, Schrank 2007; Mungiu-Pippidi 2013]. Nicméně i když existují zjevné nedostatky při měření vnímání korupce, existují zde i určité výhody. Na rozdíl od expertního hodnocení občané obvykle nemají přesné znalosti o tom, jaké skóre v korupci jejich země získala v minulých letech, takže je zde velmi malé nebezpečí takzvaného „free-ridingu“. Dále jsou průzkumy veřejného mínění vysoce korelované v čase napříč zeměmi, což odráží určitou míru spolehlivosti [Treisman 2000], a konečně, zdá se, že subjektivní hodnocení korupce má vliv na rozhodování vlády i politické chování občanů [Mauro 1995].

ESS zahrnul otázku na percepci korupce v roce 2010, kde byla zařazena pod modulem *Spravedlnosti*. Respondenti byli dotázáni, jak často se domnívají, že policie a soudci berou úplatky ve své zemi. Eurobarometer zařadil téma korupce do průzkumu Special Eurobarometer celkem pětkrát, v letech 2005, 2007, 2009, 2011 a 2013. Dotazník je velmi podrobný a důkladný, proto nabízí základy pro velmi podrobnou analýzu korupce v Evropské unii. Výhoda Eurobarometeru je ta, že byl prováděn relativně často a zahrnuje stejné země, je zde tedy velmi dobrý prostor pro srovnání jednotlivých zemí v čase. Bohužel srovnání v čase není možné v posledním průzkumu v roce 2013, protože Eurobarometer významně pozměnil dotazník. Na druhou stranu změna v dotazníku měla své opodstatnění pro studium přímé zkušenosti s korupcí, protože Eurobarometer začal filtrovat respondenty, kteří neměli kontakt s žádnou státní institucí, to znamená, že ani nemohli přijít do styku s nabídkou úplatku. Eurobarometer zahrnul otázky jak na percepci korupce, tak na přímou zkušenost s korupcí.

GCB byl proveden nejčastěji ve srovnání s jinými průzkumy veřejného mínění (7x) a má také největší pokrytí zemí. Bohužel na rozdíl od Eurobarometeru, GCB nezahrnul vždy ty samé země ve všech vlnách. Kromě toho změnil GCB formulace otázek, velmi často nejsou proto výsledky v čase přímo srovnatelné. V roce 2006 začal GCB filtrovat respondenty, kteří nebyli v kontaktu s žádnou institucí, data tudíž i z tohoto důvodu nemohou být srovnána v čase. A konečně, GCB má velmi vysokou míru non-response [Mungiu-Pippidi 2015: 41], což dále snižuje možnosti použití pro další analýzy.

Otázky týkající se korupce byly v ISSP položeny v modulu *Role vlády* v roce 2006. K dispozici jsou tři otázky týkající se korupce a úplatkářství, z nichž jedna se týká zkušenosti s korupcí a zbývající dvě se týkají percepcie zkorumpovaných úředníků a politiků. Tyto otázky jsou opět zahrnuty do modulu *Role vlády* v roce 2016.

WVS zahrnul otázky na korupci do svých dotazníků pouze jednou, do třetí vlny WVS (1995–1998), kam zařadil pouze otázku na percepci korupce, což z WVS činí nejstarší existující mezinárodní šetření o vnímání korupce v Evropě.

Přímá zkušenost s korupcí

Měření přímé zkušenosti respondenta s korupcí by mohlo být věrohodnou metodou měření korupce, měří totiž zkušenosti těch lidí, kteří se s korupcí setkali, jsou to tedy objektivnější data než u percepcie korupce. Nicméně tato metoda měří korupci pouze na nejnižší úrovni, tzv. „petty“ korupci, získat údaje o zkušenosti s korupcí na všech úrovních je téměř nemožné. Je zde také

reálná hrozba toho, že respondent bude lhát nebo odmítne odpovědět, speciálně v případě, že se bude výzkumník ptát na aktivní účast při úplatkářství. Navíc v Evropě (hlavně západní) je velmi málo respondentů, kteří se s korupcí setkali, což velmi snižuje reprezentativitu pro účely další analýzy, hlavně mikroanalýzy. Některé průzkumy dále zkoumají individuální zkušenost s korupcí a další zkušenosti jak respondenta, tak lidí z jeho rodiny či domácnosti, což může mít teoreticky za následek zdánlivě vyšší prevalenci korupce. Některé průzkumy se ptají na zkušenosti v posledních 12 měsících a jiné v posledních pěti letech, což také může snižovat validitu komparací těchto průzkumů. A konečně, některé průzkumy neberou v úvahu možnost, že respondent nebyl v kontaktu s žádnou institucí, tyto lidi neodfiltrují, čímž mohou způsobit zkreslené výsledky.

Zkušenost s korupcí v ESS byla dotazována v roce 2004 dvěma otázkami, jedna se týkala pasivní participace na úplatcích a druhá aktivní. Eurobarometer, ISSP a GCB zařadily otázku na přímou zkušenost s korupcí do dotazníku zároveň s percepcí korupce.

Třetí generace

První a druhá generace indikátorů korupce se zaměřila hlavně na mezinárodní srovnání, avšak hlubší analýza uvnitř státu nebo dokonce uvnitř jednoho sektoru je mnohem obtížnější a u indikátorů první generace dokonce nemožná. Existuje zde i problém, že indikátory první i druhé generace se opírají o názory buď expertů, nebo občanů či manažerů firem, nejsou založeny na tvrdých datech, tudíž nikdo nemůže posoudit, zda měří reálnou míru korupce, či názory, které nemusí být zdaleka spolehlivé.

V rámci jednotlivých států se tudíž začaly utvářet různé jiné indikátory pro měření korupce nebo rizik korupce pomocí tvrdých dat. Většinou se pro účely indikátorů třetí generace používají big data a měří se korupce v určitých odvětvích, ne na úrovni státu. Tyto indikátory velmi často vznikly za pomoci výsledků indikátorů druhé generace, které poukázaly na problémy v určitém odvětví státní správy, a tyto problémy začaly být hlouběji analyzovány za pomoci dalších dat [Graycar, Smith 2011]. Různé snahy o uchopení korupce objektivněji za pomoci „tvrdých dat“ můžeme zahrnout do rodiny indikátorů třetí generace. Z důvodu, že existuje mnoho přístupů a mnoho metod k měření korupce uvnitř určitých sektorů, tato třetí generace indikátorů není jednoduchá na kvalitativní popis. Jeden pokus v popisu těchto indikátorů učinili Heinrich a Hodess [2011], kteří argumentují, že indikátory třetí generace mají určité společné znaky, jako například vlastnictví dat státem (pro tvorbu indikátorů se zde, na rozdíl od nezávislých institucí či výzkumných agentur, které sbíraly data pro indikátory první a druhé generace, používají data sbíraná státními institucemi), triangulace dat (kromě „tvrdých dat“ se indikátory třetí generace soustředí i na provádění rozhovorů s aktéry státní správy či na analýzu fungování institucí), nebo svůj důraz na konkrétní použití výsledků ke zlepšování státních institucí oproti pouhému prezentování výsledku jako v případě indikátorů první či druhé generace.

Avšak i přes to, že se indikátory třetí generace snaží používat „tvrdá“ a spolehlivější data než indikátory první a druhé generace, které jsou založeny spíše na analýzách názorů respondentů či

vědců, stále není možné souhlasit s tím, že indikátory třetí generace měří reálnou výši korupce. Hlavní důvod je ten, že není možné získat tvrdá data ohledně výše korupce, speciálně na mezinárodní komparativní úrovni, indikátory třetí generace tudíž spíše naznačují možný budoucí vývoj měření korupce.

Mezi indikátory třetí generace můžeme zařadit například výzkum Golden a Picciho z roku 2005, kteří sledovali rozdíl mezi kvalitou existující infrastruktury a penězi alokovanými na vybudování této infrastruktury v Itálii [Golden, Picci 2005], podobný přístup zvolil Olken v Indonésii [2009]. Tito autoři získali nezávislé odborné posudky na reálnou cenu infrastruktury v závislosti na její kvalitě, použitých materiálech, době stavby a podobně a srovnali tuto cenu s cenou, již město vydalo na zaplacení této infrastruktury. Podle autorů může být rozdíl mezi těmito cenami přisuzován korupci. Dalším příkladem je výzkum Amoreho a Bennedsena [2013], kteří zjišťovali míru nepotismu v Dánsku na základě příbuzenských konexí mezi politiky a majiteli firem, nebo Goldman et al. sledující politické konexe firem, které získaly veřejné zakázky [2013]. Výzkum korupce za pomoci dat z veřejných zakázek se objevil u více autorů [Coviello, Gagliarducci 2010; Fazekas et al. 2013]. Některé z těchto indikátorů je potenciálně možné srovnávat i napříč státy a existují vědecké týmy, které na mezinárodním srovnání pracují, bohužel je však tento výzkum na počátku.

Jedním z příkladů mezinárodně srovnatelného indikátoru třetí generace nám může poskytnout Alina Mungiu-Pippidi [Mungiu-Pippidi 2015; Fazekas et al. 2013], která se svým týmem vytvořila v roce 2016 indikátor s názvem Index of Public Integrity (IPI)¹², kterým hodnotí státy za pomoci dat, která nejsou závislá ani na názorech odborníků, ani běžných občanů, ale opírají se o reálné ukazatele. Tento indikátor neměří korupci per se, ale spíše riziko nerušeného bujení korupce [Mungiu-Pippidi and Dadasov 2015]. Z důvodu, že tento indikátor je jediným existujícím indikátorem třetí generace, který je možné srovnávat napříč evropskými státy, je také jediným, který tento článek zařadil do srovnávací korelační analýzy, i když, jak je diskutováno níže, má tento indikátor zásadní problémy.

IPI se skládá z šesti proměnných: jsou jimi nezávislost justice, otevřenost obchodu, eobčanství, administrativní zátěž, transparentnost rozpočtu a svoboda tisku. Přestože je tento indikátor zařazen mezi indikátory třetí generace, není použit pouze tvrdých dat pro vývoj indikátoru stoprocentní. Hned první součást, nezávislost justice, není postavena na objektivních datech, ale na průzkumu vnímání nezávislosti justice mezi řídicími pracovníky (Executive opinion survey), který každoročně provádí World Economic Forum.¹³ Druhá proměnná, otevřenost obchodu, se skládá z počtu dokumentů, které jsou potřeba pro export a import, a z času, který je potřeba na kompletaci exportu či importu. Administrativní zátěž se skládá z počtu procedur, které jsou nutné k založení podnikání, z času, který je na toto vynaložen, kolikrát ročně je třeba platit daně a kolik času je na zaplacení daní. E-občanství je založeno na datech o používání internetu, konkrétně se jedná o procento z populace, které má přípojku k internetu, z podílu uživatelů internetu a z podílu uživatelů Facebooku. Transparentnost rozpočtu je založena na

¹² Data jsou dostupná na <http://integrity-index.org/>.

¹³ Více informací na <http://reports.weforum.org/global-competitiveness-report-2015-2016/>.

Tabulka 4: Korelační analýza mezi indikátory první a druhé generace¹

Typ	Průzkum	Rok	N	CPI	CC
Percepce: veř. čin.	WVS	1996	27	-0,94***	-0,84***
Zkušenost: firmy	BEEPS	1999	17 CPI 18 CC	-0,53*	-0,47*
Zkušenost: firmy	BEEPS	2002	16 CPI 18 CC	-0,76**	-0,76**
Percepce: soudnic.	GCB	2004	30	-0,88***	-0,87***
Zkušenost	GCB	2004	29	-0,74***	-0,77***
Zkušenost	ESS	2004	25	-0,81***	-0,82***
Percepce: soudnic.	Eurobar	2005	25	-0,85***	-0,86***
Zkušenost	Eurobar	2005	25	-0,74***	-0,74***
Percepce: soudnic.	GCB	2005	28	-0,93***	-0,92***
Zkušenost	GCB	2005	27	-0,69**	-0,73***
Zkušenost: firmy	BEEPS	2005	24	-0,74***	-0,8***
Percepce: politici	ISSP	2006	19	-0,87***	-0,91***
Zkušenost	ISSP	2006	19	-0,86***	-0,87***
Percepce: soudnic.	GCB	2006	27	-0,93***	-0,93***
Zkušenost	GCB	2006	26	-0,61**	-0,65**
Percepce: soudnic.	Eurobar	2007	27	-0,81***	-0,8***
Zkušenost	Eurobar	2007	27	-0,72***	-0,75***
Percepce: soudnic.	GCB	2007	30	-0,91***	-0,91***
Zkušenost	GCB	2007	26	-0,7**	-0,7**
Percepce	Eurobar	2009	27	-0,84***	-0,84***
Zkušenost	Eurobar	2009	27	-0,82***	-0,85***
Percepce	GCB	2009	27	-0,87***	-0,89***
Zkušenost	GCB	2009	26	-0,69**	-0,73***
Zkušenost: firmy	BEEPS	2009	19	-0,83***	-0,86***
Percepce: soudci	ESS	2010	27	-0,91***	-0,91***
Percepce	Eurobar	2011	27	-0,86***	-0,85***
Zkušenost	Eurobar	2011	27	-0,79***	-0,81***
Percepce	GCB	2011	31	-0,89***	-0,88***
Zkušenost	GCB	2011	32	-0,77***	-0,79***
Percepce: stát. zást.	Eurobar	2013	28	-0,65**	-0,64**
Zkušenost	Eurobar	2013	28	-0,52**	-0,61**
Percepce	GCB	2013	30	-0,91***	-0,91***
Zkušenost	GCB	2013	26	-0,73***	-0,73***
Zkušenost: firmy	BEEPS	2012–14	19	-0,66**	-0,7**

Zdroj: ESS, BEEPS, GCB, Eurobarometer, ISSP, WVS, CC, CPI. V roce 1999 nejsou data pro CC, je to tudíž průměr CC 1998–2000. Spearmanův korelační koeficient. *p < 0,05; **p < 0,01; ***p < 0,001.

¹ Do analýzy není zahrnut Eurobarometer z důvodu velmi nízkého počtu respondentů, což zásadně snižuje reprezentativitu průzkumu, jak je uvedeno výše.

Table 4: Correlation analysis of first- and second-generation indicators¹

Type	Survey	Year	N	CPI	CC
Perceptions: Public officials	WVS	1996	27	-0.94***	-0.84***
Experience: Businesses	BEEPS	1999	17 CPI 18 CC	-0.53*	-0.47*
Experience: Businesses	BEEPS	2002	16 CPI 18 CC	-0.76**	-0.76**
Perceptions: Judiciary	GCB	2004	30	-0.88***	-0.87***
Experience	GCB	2004	29	-0.74***	-0.77***
Experience	ESS	2004	25	-0.81***	-0.82***
Perceptions: Judiciary	Eurobar	2005	25	-0.85***	-0.86***
Experience	Eurobar	2005	25	-0.74***	-0.74***
Perceptions: Judiciary	GCB	2005	28	-0.93***	-0.92***
Experience	GCB	2005	27	-0.69**	-0.73***
Experience: Businesses	BEEPS	2005	24	-0.74***	-0.8***
Perceptions: Politicians	ISSP	2006	19	-0.87***	-0.91***
Experience	ISSP	2006	19	-0.86***	-0.87***
Perceptions: Judiciary	GCB	2006	27	-0.93***	-0.93***
Experience	GCB	2006	26	-0.61**	-0.65**
Perceptions: Judiciary	Eurobar	2007	27	-0.81***	-0.8***
Experience	Eurobar	2007	27	-0.72***	-0.75***
Perceptions: Judiciary	GCB	2007	30	-0.91***	-0.91***
Experience	GCB	2007	26	-0.7**	-0.7**
Perceptions	Eurobar	2009	27	-0.84***	-0.84***
Experience	Eurobar	2009	27	-0.82***	-0.85***
Perceptions	GCB	2009	27	-0.87***	-0.89***
Experience	GCB	2009	26	-0.69**	-0.73***
Experience: Businesses	BEEPS	2009	19	-0.83***	-0.86***
Perceptions: Judges	ESS	2010	27	-0.91***	-0.91***
Perceptions	Eurobar	2011	27	-0.86***	-0.85***
Experience	Eurobar	2011	27	-0.79***	-0.81***
Perceptions	GCB	2011	31	-0.89***	-0.88***
Experience	GCB	2011	32	-0.77***	-0.79***
Perceptions: Public prosecutors	Eurobar	2013	28	-0.65**	-0.64**
Experience	Eurobar	2013	28	-0.52**	-0.61**
Perceptions	GCB	2013	30	-0.91***	-0.91***
Experience	GCB	2013	26	-0.73***	-0.73***
Experience: Businesses	BEEPS	2012–14	19	-0.66**	-0.7**

Source: ESS, BEEPS, GCB, Eurobarometer, ISSP, WVS, CC, CPI. Due to absence of 1999 data for CC, the mean value of CC for 1998–2000 is input. Spearman's correlation coefficient. * p < 0.05; ** p < 0.01; *** p < 0.001.

¹ As mentioned above, Eurobarometer is not included due to its extremely low sample and the ensuing low representativeness.

specifických datech o transparentnosti jednotlivých položek rozpočtu. Svoboda tisku je získána od Freedom House a kromě objektivních dat je založena i na názorech odborníků v této oblasti.

Z popisu indexu IPI je vidět, že indikátory třetí generace, které by srovnávaly korupci na mezinárodní úrovni, jsou teprve na začátku vývoje. IPI sice může být zařazen mezi indikátory třetí generace, ale podrobný popis jednotlivých částí indikátorů ukázal, že IPI je zatím na pomezí třetí generace, jelikož zdroji nejsou pouze objektivní data, ale částečně i názory odborníků. Tento popis ukazuje, že indikátory třetí generace nejsou zatím dostatečným nahrazením indikátorů první a druhé generace, v této chvíli neexistují objektivní „tvrdá“ data, která by umožnila měřit reálnou míru korupce či srovnávat míru korupce napříč státy.

Kvantitativní srovnání indikátorů první, druhé a třetí generace

Následující část analyzuje korelace první, druhé a třetí generace indikátorů. Taková analýza je schopná ukázat, zda percepce expertů (v případě první generace), percepce běžných občanů a reálné zkušenosti občanů či firem (v případě druhé generace) a relativně objektivní data složená z více zdrojů (v případě třetí generace) poukazují na stejné či podobné trendy v korupci. Tato analýza může vznést světlo do debaty, zda kompozitní indikátory měří stejný jev jako průzkumy měřící reálné zkušenosti s korupcí a zda jsou výsledky podobné výsledkům z „objektivních“ dat. Jak bylo diskutováno výše, je otázkou, zda indikátory měří reálnou korupci, či zda měří jiné jevy. Statisticky signifikantní výsledky korelační analýzy by nám mohly ukázat, zda zde existuje určitá reliabilita těchto indikátorů. Je však nutné podotknout, že i když nám tato korelační analýza může přiblížit určité kvalitativní znaky jednotlivých průzkumů, není možné učinit závěry o validitě a reliabilitě těchto průzkumů veřejného mínění a firem pouze na základě těchto výsledků z toho důvodu, že většina průzkumů byla provedena v různých letech a také v různých zemích, v některých případech chybějí celé oblasti Evropy (například BEEPS provádí průzkumy pouze ve východní Evropě, jak bylo uvedeno výše).

Tabulka 4 ukazuje korelace mezi první a druhou generací indikátorů měřících korupci, to znamená korelace mezi kompozitními ukazateli a průzkumy veřejného mínění a firem. Tyto korelace jsou rozděleny podle roku, kdy byl průzkum či kompozitní ukazatel publikován. Tato metoda je volena z toho důvodu, že korupce má určitý vývoj a srovnávání indikátorů ve stejných letech zajistí alespoň základní validitu.

Z tabulky je na první pohled vidět, že indikátory první a druhé generace jsou navzájem korelované. Všechny korelace jsou relativně silné, což může poukazovat na určitou reliabilitu těchto průzkumů, podle výsledků analýzy to vypadá, že indikátory první i druhé generace měří relativně podobný jev.¹⁴ Na druhou stranu je jasně vidět, že indikátory první a druhé generace nejsou

¹⁴ Samozřejmě otázka z úvodu přetrvává: Měří indikátory první a druhé generace korupci, nebo něco jiného? Tato analýza ukázala, že obě generace měří podobný jev, může se tedy zdát, že se blíží reálné míře korupce, není to ale dostatečný důkaz.

Tabulka 5: Korelační analýzy IPI s indikátory první a druhé generace

Typ	Průzkum	Rok	N	IPI
Kompozitní indikátor	CPI	2014	34	0,95***
Kompozitní indikátor	CC	2014	34	0,95***
Percepce: stát. zást.	Eurobar	2013	27	-0,66**
Zkušenost	Eurobar	2013	27	-0,58**
Percepce	GCB	2013	28	-0,86***
Zkušenost	GCB	2013	23	-0,76***
Zkušenost: firmy	BEEPS	2012–14	17	-0,55*

Zdroj: BEEPS, GCB, Eurobarometer, CC, CPI. Spearmanův korelační koeficient. *p < 0,05; **p < 0,01; ***p < 0,001.

Table 5: Correlation analysis of IPI with first- and second-generation indicators

Type	Survey	Year	N	IPI
Composite indicator	CPI	2014	34	0.95***
Composite indicator	CC	2014	34	0.95***
Perceptions: Public prosecutors	Eurobar	2013	27	-0.66**
Experience	Eurobar	2013	27	-0.58**
Perceptions	GCB	2013	28	-0.86***
Experience	GCB	2013	23	-0.76***
Experience: Businesses	BEEPS	2012–14	17	-0.55*

Source: BEEPS, GCB, Eurobarometer, CC, CPI. Spearman's correlation coefficient. * p < 0.05; ** p < 0.01; *** p < 0.001.

vzájemně zaměnitelné, i když jsou korelace statisticky signifikantní, indikátory nejsou naprosto totožné (speciálně u přímé zkušenosti s úplatky).

Nejsilnější korelace můžeme nalézt mezi kompozitními indikátory a mezi průzkumy veřejného mínění zjišťujícími percepce korupce. Tento výsledek se jeví jako očekávatelný, protože obě tyto míry měří korupci relativně podobně. Kompozitní indikátory měří především názory expertů na percepce korupce a průzkumy veřejného mínění zase měří názory veřejnosti na rozšířenost korupce, tyto dva indikátory tedy měří totožnou věc a vysoká korelace znamená, že veřejnost má na rozšíření korupce podobné názory jako experti. Nejvyšší korelaci můžeme nalézt v percepce korupce mezi WVS v roce 1996 a CPI od Transparency International (0,94). Naopak nejnižší korelace je mezi Eurobarometerem z roku 2013 a CC (0,64).

Zajímavější výsledky nám ukazují korelace mezi indikátory první generace a zkušeností s korupcí jak v průzkumech veřejného mínění, tak ve firmách. Zde je vidět, že korelace jsou na první pohled nižší. Přestože zde vždy existuje statisticky signifikantní vztah, v indikátorech měřících stejný jev by se očekávala korelace vyšší. Nejméně korelovaný je opět Eurobarometer z roku 2013, kde je korelace s CPI pouze 0,52, naopak nejvyšší korelaci u průzkumů veřejného mínění můžeme nalézt mezi ISSP z roku 2006 a CPI (0,87). U firemního průzkumu BEEPS je

nejvyšší korelace v roce 2009 – 0,86 s CC a nejnižší v roce 1999 – 0,47. Rozdíl mezi průzkumy tázající se na zkušenosti s korupcí může být v důsledku více fenoménů, například výsledky může ovlivňovat způsob, jakým je otázka položena. Eurobarometer a GCB se ptají na zkušenosti s korupcí v posledních 12 měsících na rozdíl od pěti let jako v případě ESS nebo ISSP. Dále zde mohou být rozdíly v tom, že některé průzkumy se ptají na respondenta a některé na domácnost. A některé průzkumy nefiltrují respondenty, kteří nepřišli do styku s žádnou státní institucí. Jak je vidět z této analýzy, přinesla nám sice odpověď na otázku, zda první a druhá generace indikátorů měří podobný jev, stále však není jasné, zda je tento jev reálnou mírou korupce. Částečnou odpověď na tuto otázku nám může nabídnout tabulka 5, která ukazuje výsledky korelační analýzy IPI s indikátory první a druhé generace.

Možnost korelace je v případě analýzy indikátoru třetí generace velmi omezena, protože IPI vznikl v roce 2016, nejsou zde starší data. Pro účely této srovnávací analýzy proto použijeme nejnovější indikátory, to znamená CC a CPI z roku 2014, GCB a Eurobarometer z roku 2013 a BEEPS z roku 2012–2014. Výhodou je, že máme srovnání se všemi typy indikátorů. Výsledky analýzy jsou v mnohém překvapující, IPI má korelaci s indikátory první generace na úrovni 0,95, což je extrémně vysoká korelace. Zde je možné, že tyto dva indikátory měří totožný jev, a je tedy možné je zaměnit. Je velmi zajímavé, že indikátor, který je z větší části založen na objektivních datech, má totožné výsledky s indikátorem, který měří pouze názory expertů. Takto silná korelace je až podezřelá, nabízí se proto otázka, jak přesně autoři jednotlivé části indikátoru vážili. Na základě výsledků korelační analýzy první a třetí generace indikátorů již nejsou překvapující výsledky korelace druhé a třetí generace. Nemůžeme přímo srovnávat sílu korelací, jelikož jsou zde zastoupeny jiné země, ale trend je zde totožný. Percepce korupce je korelována relativně silněji, zkušenost s korupcí jak firem, tak občanů je korelována slaběji.

Závěr

Tento článek představil tři generace indikátorů měření korupce v Evropě. První generace, která je založena na kompozitních indikátorech, představuje světoznámé a nejpoužívanější indikátory, jako je Control of Corruption od World Bank nebo Corruption Perception Index od Transparency International. I když jsou tyto indikátory dodnes nejpoužívanější ve vědeckém výzkumu korupce, existují zde zásadní problémy, které vznášejí pochyby o jejich validitě. Druhá generace indikátorů je založena na průzkumech veřejného mínění nebo na průzkumech ve firmách. Do těchto indikátorů patří jak sledování percepce korupce, tak dotazování na přímou zkušenost s korupcí. Speciálně dotazování na zkušenost s korupcí se snaží o objektivnější měření korupce, ale i to má svoje metodologické nevýhody. Z popisu vyplývá, že ani indikátory první generace, ani indikátory druhé generace neměří korupci na základě tvrdých dat; kompozitní indikátory jsou založené především na názorech expertů, indikátory druhé generace na výpovědích firem či občanů, kde se jedná výlučně o úplatkářství, a další formy korupce jsou zanedbané. Tento rozpor se snaží překonat indikátory třetí generace, které se snaží stavět svoje výsledky na tvrdých

datech. Indikátory třetí generace se zaměřují spíše vždy na jednu zemi a jeden sektor, možnost srovnání jako u indikátorů první či druhé generace je zatím velmi omezená. Tento článek představuje pionýra ve třetí generaci měření korupce, indikátor IPI, který však nestaví pouze na tvrdých datech, ale používá i „měkčí“ zdroje, není tudíž ještě plnohodnotným zástupcem indikátorů třetí generace.

Rozdíly mezi indikátory byly analyzovány pomocí korelační analýzy. Korelační analýza mezi indikátory první a druhé generace korupce ukázala, že zde existují velmi silné korelace mezi percepcí korupce v průzkumech veřejného mínění a mezi indikátory první generace. Korelace mezi přímou zkušeností s korupcí jak v průzkumech veřejného mínění, tak v průzkumech firem byly signifikantní, ale slabší. Podobné výsledky jsme získali korelováním třetí generace s indikátory první a druhé generace. IPI je extrémně silně korelovaný s indikátory první generace a slaběji s indikátory druhé generace, z nichž je silněji korelovan s percepcí korupce než se zkušeností s úplatky. Vysoké korelace IPI s CC a CPI však mohou také poukazovat na problém s indikátorem IPI, doporučila bych tedy počkat na další ročník IPI nebo na jiný indikátor třetí generace srovnávající země před použitím IPI pro výzkum korupce.

Analýza korelací tří generací indikátorů nám sice ukázala, že zde existuje statisticky signifikantní vztah mezi těmito generacemi, což může naznačovat určitou reliabilitu, ale přesto tyto korelace nabývají někdy i velmi nízkých hodnot, speciálně u zkušeností s korupcí a kompozitních indikátorů. Tyto výsledky mohou poukazovat na to, že percepcie korupce není to samé co reálná korupce a tyto dvě proměnné není možné bez problémů zaměňovat, což je se bohužel stává u mnoha autorů zkoumajících korupci [Blake, Martin 2006; Treisman 2000; Pellegata 2012; Mauro 1995]. Pro výzkum korupce je tedy zásadní mít na paměti, jaký typ indikátoru výzkumník používá, jaké má tento indikátor výhody a nevýhody a v jakých případech se může používat. V neposlední řadě je nutné klást důraz na to, že je chyba zaměňovat percepcii korupce s reálnou korupcí. Indikátory třetí generace, které se snaží dále přiblížit nepřímému měření korupce, jsou sice zatím pouze na počátku praktického využití, je ale velmi pravděpodobné, že jejich rozmach způsobí naprosto rozdílné výsledky výzkumů o důvodech a důsledcích korupce, pro něž byly doposud používány pouze indikátory první či druhé generace.

LITERATURA

- Amore, Mario D., Morten Bennesen. 2013. „The value of local political connections in a low-corruption environment.“ *Journal of Financial Economics* 110 (2): 387–402.
<https://doi.org/10.1016/j.jfineco.2013.06.002>.
- Bandura, Romina. 2008. *A Survey of Composite Indices Measuring Country Performance*. New York: United Nations Development Programme – Office of Development Studies.
- Blake, Charles H., Christopher G. Martin. 2006. „The dynamics of political corruption: Re-examining the influence of democracy.“ *Democratization* 13 (1): 1–14.
<https://doi.org/10.1080/13510340500378191>

- Brown, David.S., Michael Touchton, Andrew Whitford. 2011. „Political Polarization as a Constraint on Corruption: A Cross-national Comparison“. *World Development* 39 (9): 1516–1529.
- Bukovansky, Mlada. 2006. „The hollowness of anti-corruption discourse“. *Review of International Political Economy* 13 (2): 181–209.
- Charron, Nicholas, Victor Lapuente, Bo Rothstein. 2013. *Quality of Government and Corruption from a European Perspective: A Comparative Study of Good Government in EU Regions*. Northampton: Edward Elgar Publishing.
- Coviello, Decio, Stefano Gagliarducci. 2010. *Building political collusion: Evidence from procurement auctions*. IZA Discussion Paper No. 4939.
- Donchev, Dilyan., Gergely Ujhelyi. 2014. „What Do Corruption Indices Measure?“ *Economics & Politics* 26 (2): 309–331. <https://doi.org/10.1111/ecpo.12037>.
- EBRD. 2010. *The Business Environment and Enterprise Performance Survey (BEEPS)*. European Bank for Reconstruction and Development.
- EBRD. 2015. *The business environment in the transition region*. European Bank for Reconstruction and Development.
- European Commission. 2014. *Special Eurobarometer 39 – Corruption – Report*. European Commission.
- Fazekas, Mihaly, István J. Toth, Lawrence P. King. 2013. *Anatomy of grand corruption: A composite corruption risk index based on objective data*. Budapest: Corruption Research Center Budapest.
- Golden, Miriam A., Lucio Picci. 2005. „Proposal for a New Measure of Corruption, illustrated with Italian data“. *Economics & Politics* 17 (1): 37–75.
- Goldman, Eitan, Jörg Rocholl, Jongil So. 2013. „Politically Connected Boards of Directors and The Allocation of Procurement Contracts“. *Review of Finance* 17 (5): 1617–1648.
- Graycar, Adam, Russell G. Smith. 2011. *Handbook of Global Research and Practice in Corruption*. Northampton: Edward Elgar Publishing.
- Griesshaber, Nicolas, Benny Geys. 2011. „Civic Engagement and Corruption in 20 European Democracies“. *European Societies* 14 (1): 57–81.
- Heidenheimer, Arnold J., Michael Johnston. 2002. *Political Corruption – Concepts and Contexts*. New Brunswick: Transaction Publishers.
- Heinrich, Finn., Robin Hodess. 2011. „Measuring corruption.“ Pp. 18-33 In: Adam Graycar, Russel G. Smith (eds.), *Handbook of Global Research and Practice in Corruption*. Northampton: Edward Elgar Publishing.
- Heller, Nathaniel. 2009. „Defining and measuring corruption.“ Pp. 47-65 In: Rotberg, Robert I. (Ed.), *Corruption, Global Security, and World Order*. Washington: Brookings Institution Press.
- Heywood, Paul. 1997. „Political Corruption: Problems and Perspectives“. *Political Studies* 45 (3): 417–435. <https://doi.org/10.1111/1467-9248.00089>.

- Kapoor, Mudit, Shamika Ravi. 2012. „Determinants of Corruption: Government Effectiveness vs. Cultural Norms“. *The B.E. Journal of Economic Analysis & Policy* 12 (1): 1935-1682. <https://doi.org/10.1515/1935-1682.3049>.
- Karklins, Rasma. 2005. *The system made me do it : corruption in post-communist societies*. Armonk, NY; London: M.E. Sharpe.
- Kaufmann, Daniel, Aart Kraay, Massimo Mastruzzi. 2004. *Governance Matters III: Governance Indicators for 1996-2002*. Washington: World Bank.
- Kaufmann, Daniel, Aart Kraay, Pablo Zoido-Lobaton. 1999. *Governance Matters*. Washington: World Bank.
- Kaufmann, Daniel, Aart Kraay, Eduardo Lora, Lant Pritchett. 2002. „Growth without Governance“. *Economía* 3 (1): 169–229.
- Kaufmann, Daniel, Massimo Mastruzzi, Aart Kraay. 2010. *The Worldwide Governance Indicators: Methodology and Analytical Issues*. The World Bank.
- Kurtz, Marcus J., Andrew Schrank. 2007. „Growth and Governance: Models, Measures, and Mechanisms“. *The Journal of Politics* 69 (2): 538–554. <https://doi.org/10.1111/j.1468-2508.2007.00549.x>.
- Lambsdorff, Johann.G. 2001. *Background Paper to the 2001 Corruption Perception Index*. Transparency International.
- Malito, Debora.V. 2014. *Measuring Corruption Indicators and Indices*. EUI working papers.
- Mauro, Paolo. 1995. „Corruption and growth“. *The Quarterly Journal Of Economics* 110 (3): 681–712. <https://doi.org/10.2307/2946696>.
- Mungiu-Pippidi, Alina. 2013. *The Good, the Bad and the Ugly: Controlling Corruption in the European Union. Policy paper*. European Parliament.
- Mungiu-Pippidi, Alina. 2015. *The Quest for Good Governance: How Societies Develop Control of Corruption*. Cambridge: Cambridge University Press.
- Mungiu-Pippidi, Alina, Ramin Dadasov. 2015. *A note on What causes change in governance and How to measure public integrity*. Berlin: Hertie School of Governance.
- OECD. 2008. *Corruption*. OECD.
- Olken, Benjamin. A. 2009. „Corruption perceptions vs. corruption reality“. *Journal of Public Economics* 93(7-8): 950–964.
- Pellegata, AI. 2012. „Constraining political corruption: an empirical analysis of the impact of democracy“. *Democratization* 20 (7): 1195–1218. <https://doi.org/10.1080/13510347.2012.688031>.
- Rohwer, Anja. 2009. *Measuring Corruption: A Comparison Between The Transparency International's Corruption Perceptions Index And The World Bank's Worldwide Governance Indicators*. Munich: Center for Economic Studies Munich.
- Rose, Richard, Caryn Peiffer. 2012. *Paying Bribes to Get Public Services: A Global Guide to Concepts and Survey Measures*. Glasgow: Centre for the Study of Public Policy.
- Rothstein, Bo., Davide Torsello. 2013. *Is Corruption Understood Differently in Different Cultures?* Gothenburg: The Quality of Government Institute.

- Tanzi, Vito. 1998. „Corruption Around the World. International Monetary Fund“. *IMF Staff Papers* 45 (4): 559-594.
- Transparency International. 2012a. Corruption Perceptions Index 2012. Transparency International.
- Transparency International. 2012b. Corruption Perceptions Index 2012: An Updated Methodology. Transparency International.
- Transparency International. FAQs on Corruption. Dostupné na: https://www.transparency.org/whoweare/organisation/faqs_on_corruption.
- Treisman, Daniel. 2000. „The Causes of Corruption: A Cross-National Study“. *Journal of Public Economics* 76: 399–457.
- Treisman, Daniel. 2007. „What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?“ *Annual Review of Political Science* 10, 1: 211–244.
- Treisman, Daniel. 2013. „Corruption in the Post-communist Transition.“ Pp 209–216. In: Paul Hare and Gerard Turley (eds.), *Handbook of the Economics and Political Economy of Transition*. London: Routledge.
- United Nations. 2004. *United Nations Convention against Corruption*. United Nations.
- Uslaner, Eric. M. 2009. „Corruption“ Pp. 127–142. In: Geert T. Svendsen, Gunnar L.H. Svendsen (eds.), *The Handbook of Social Capital: The Troika of Sociology, Political Science and Economics*. Northampton: Edward Elgar Publishing.
- World Bank. 1997. *Helping Countries Combat Corruption*. Washington: World Bank.
- World Bank. *Control of Corruption*. Washington: World Bank.
- Zakaria, Patty. 2013. „Is corruption an enemy of civil society? The case of Central and Eastern Europe“. *International Political Science Review* 34 (4): 351–371. <https://doi.org/10.1177/0192512112466880>.

Z D R O J E D A T :

- ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3. 4. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3. 2. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- European Commission (2012): Eurobarometer 64.3 (Nov–Dec 2005). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA4415 Data file Version 1. 0. 1, doi:10.4232/1.10971

- European Commission (2012): Eurobarometer 68.2 (Nov–Dec 2007). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA4742 Data file Version 4. 0. 1, doi:10.4232/1.10986
- European Commission (2012): Eurobarometer 72.2 (Sep–Oct 2009). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA4976 Data file Version 3. 0. 0, doi:10.4232/1.11137
- European Commission and European Parliament, Brussels (2014): Eurobarometer 76.1 (2011). TNS OPINION & SOCIAL, Brussels [Producer]. GESIS Data Archive, Cologne. ZA5565 Data file Version 4. 0. 0, doi:10.4232/1.11847
- European Commission, Brussels (2015): Eurobarometer 79.1 (2013). TNS Opinion, Brussels [producer]. GESIS Data Archive, Cologne. ZA5687 Data file Version 2. 0. 0, doi:10.4232/1.12417
- ISSP Research Group (2008): International Social Survey Programme: Role of Government IV – ISSP 2006. GESIS Data Archive, Cologne. ZA4700 Data file Version 1. 0. 0, doi:10.4232/1.4700
- Transparency International (2004–2011). Global corruption Barometer. Data z: http://archive.transparency.org/policy_research/surveys_indices/gcb
- Transparency International (1995–2014). Corruption perception index. Data z: <http://www.transparency.org/research/cpi/>
- World Values Survey. Wave 3 1995–1998 official aggregate v.20140921. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: Asep/JDS, Madrid SPAIN.
- World Bank. (1996–2014). Control of corruption. Worldwide Governance Indicators (www.govindicators.org).
- World Bank. (1999–2014) EBRD-World Bank Business Environment and Enterprise Performance Survey (BEEPS). Panel Database. 1999, 2002–2009, 2009–2014.

O AUTORECH



Kristýna Bašná získala Ph.D. v oboru sociologie na Karlově univerzitě a titul MSc. ze sociologie na University of Oxford. Pracuje jako výzkumná pracovníce Sociologického ústavu AV ČR v oddělení Hodnotové orientace ve společnosti. Její výzkum se zaměřuje na témata korupce a postojů k sociálnímu státu v Evropě, zejména v postkomunistických zemích.

Kontaktovat ji lze na adrese kristyna.basna@soc.cas.cz.

ZKRATKY:

BEEPS	The Business Environment and Enterprise Performance Survey
CC	Control of Corruption
CPI	Corruption Perception Index
ESS	European Social Survey
GCB	Global Corruption Barometer
ISSP	International Social Survey Programme
TI	Transparency International
WB	World Bank

Homogamie a internet: Pomáhají online seznamky rozumět výběrovému párování?¹

Markéta Šetinová, Institut sociologických studií, Fakulta sociálních věd, Univerzita Karlova

Homogamy and the Internet: Can Online Dating Sites Help Us Understand Assortative Mating?

ABSTRACT

Non-random selection of reproductive partners in the human population, i.e. assortative mating, has been a stable occurrence for decades and across societies, including the Czech Republic. Social sciences have paid primary attention to homogamy, marriage between similar partners, also due to its potential impact on society. High levels of homogamy in a society may imply high closedness of the different groups, prevent social mobility, suggest racial tensions, or lead to higher inequality.

Three factors are deemed as responsible for homogamy. Structural factors are represented by the partners' physical closeness as they meet, for example, at school or other institutions. Second, there are individual factors, namely one's personal preferences for a relationship with someone similar to one. Finally, social factors represent the influence of social groups as they have their own stake in preserving group cohesion through marriage between their members. However, evidence of assortative mating and its mechanisms mostly comes from attitudinal data or questionnaire surveys, which make it difficult to differ between individual and structural factors, among others. In this review article, I focus on the potential of data from online dating sites to help us better understand assortative mating. Is the Internet changing the ways people meet? How does assortative mating work in the online environment and what can it tell us about the mating process in general?

KEYWORDS:

homogamy
assortative mating
online dating
partner selection

DOI LINK:

<http://doi.org/10.13060/1214438X.2018.116.417>

¹ Tato publikace vznikla v rámci projektu GA UK (829417) „Seznamování přes internet: Dopad online seznamek na hledání partnera“ a GA ČR (17-12099S) „Změna poměru mužů a žen ve vyšším vzdělání a transformace sňatkového trhu“

The paper starts by presenting the main theories of assortative mating and adding information on the specifics of the online environment. There are basically two contradictory expectations of the effect of the Internet on the mating process. The first group assumes that the easy access to thousands of potential partners across all sociodemographic groups an online dating service provides, plus the absence of direct influence of intermediaries such as community members, will lead to higher diversity of the ensuing marriages. In short, the level of assortative mating will decrease. The opposite expectation is based on the assumption that people generally prefer contact with similar partners and online dating services help them more effectively filter mates by preferred criteria, thus avoiding contact with people who do not fit their expectations. As a result, assortative mating will grow.

The theory section is followed by a review of contemporary international research of online dating services with regard to assortative mating. I conclude the detailed review by arguing that the human preference for similar partners occurs in the online environment as well, and therefore, assortative mating likely cannot be explained by structural factors only. Moreover, all points to the fact that data from real-life user interaction in online dating services can importantly widen our knowledge of assortative mating.

Láska není tak náhodná, jak nám občas může připadat. Vše nasvědčuje tomu, že zvířata i lidé preferují reprodukční partnery, kteří jsou jim podobní. V přírodě si zvířecí páry odpovídají především fyzicky, třeba velikostí nebo barvou, v lidské populaci nacházíme rozsáhlou evidenci pro podobnost párů především na základě sociodemografických charakteristik [Kalmijn 1998]. Prakticky to ve společnosti znamená, že vysokoškolačky mají tendenci provdat se za vysokoškolačky, křesťané si častěji za životní partnery volí křesťanky a běloši preferují bělošky. Tento nenáhodný výběr sexuálních partnerů v rámci reprodukce nazýváme výběrovým párováním. Ve společenských vědách používáme jako ukazatel výběrového párování homogamii, tedy manželství mezi lidmi s podobnými charakteristikami, například vzděláním, náboženstvím nebo etnicitou. Dekády výzkumu potvrzují, že tendenci k homogamii v základních charakteristikách pozorujeme jako víceméně stabilní jev napříč společnostmi (shrnutí výzkumů ve společenských vědách se věnují například [Kalmijn 1998] a [Schwartz 2013]).

Jak homogamie vlastně vzniká? Obecně máme pro vzorce homogamního chování k dispozici tři potenciální vysvětlení neboli mechanismy [Kalmijn 1998]. Prvním je fyzická blízkost podobných partnerů, například díky seznámení se ve škole nebo skrz další instituce, tedy tak zvané strukturální faktory. Seznámit se lze jen s tím, koho reálně potkáváte. Druhým jsou individuální faktory, tedy preference jednotlivce pro vztah s partnerem, který mu je podobný. S takovým partnerem se lépe navazuje konverzace i vychovávají děti. Třetí mechanismus je sociální, tedy vliv sociálních skupin na jednotlivce. Skupiny mají vlastní zájem na udržení koheze komunity skrz manželství jejich příslušníků. Předmětem akademické diskuze nadále je, zda má jeden z těchto mechanismů primární roli, případně jaké je jejich vzájemné působení. Zde ale tradiční výzkum často naráží na limitaci dat.

Většina studií homogamie se opírá o pozorování již vzniklých svazků, například využívá data ze sčítání lidu nebo dotazníkových šetření. Sociodemografické údaje sezdaných párů nám ale neřeknou nic o tom, kdy a jak podobnost vznikla. Byli si jednotlivci podobní už ve chvíli, kdy se seznámili, nebo se stali podobnými až v průběhu vztahu? Zcela konkrétně – našla si vysokoškolačka partnera vysokoškoláka, nebo si našla partnera středoškoláka, kterého namotivovala ke studiu? Výsledek je v dotaznících stejný, mechanismus vzniku homogamie ale zcela odlišný. Další způsob měření homogamie je pomocí postojových dotazníků, kde respondenti udávají své preference ohledně potenciálních partnerů [Katrňák 2008]. Postoje jsou ale náchylné ke zkreslení z důvodu stylizace respondenta (tzv. sociální desirabilita odpovědí). Toto zkreslení pozorujeme například u otevřenosti vůči partnerovi z jiné etnické skupiny: v postojových dotaznících udávají lidé vysokou toleranci k mezietnickým sňatkům, reálně k nim ale dochází velmi zřídka [Wang 2012].

S ohledem na limitace tradičního výzkumu jsou velmi zajímavá právě data ze seznamek, která umožňují sledovat reálné uživatelské chování v první fázi hledání partnera. Nemusíme se tedy zabývat otázkou, kdy podobnost páru nastala, ani brát v potaz sociální desirabilitu. Data ze seznamky navíc nabízí velký detail analýzy, od bohatých sociodemografických údajů uživatelů (některé seznamky jsou například propojené s Facebookem nebo Instagramem), po informace o uživatelské interakci. Jaká kritéria volí uživatelé, když potenciální partnery na seznamce vyhledávají? Koho kontaktují? Kdo jim odepisuje? Jak dlouho si píší? Dokážeme tak například velmi jednoduše vyhodnotit, jaké pohlaví má k homogamním vztahům větší tendenci, což je informace, kterou z tradičních dat získáváme jen těžko.²

Online seznamování nenabízí jen zcela nová data, ale také zásadně mění způsob poznávání potenciálních partnerů: jednotlivci se pohybují v prostředí, kde mají neomezený přístup k celému spektru nezadaných. Proměňuje tato skutečnost typ partnera, kterého si hledají? Není výběrové párování jen důsledkem seznamování v offline prostředí, kde na rozdíl od internetu nemáme jednoduchý přístup k lidem z jiných sociálních skupin? Nebo naopak bude online seznamování – i díky jednoduchému filtrování v databázi nezadaných – výběrové párování podporovat a zvyšovat? Tyto otázky jsou namísto především v posledních letech, kdy internet pomalu nahrazuje tradiční formy hledání partnera, jako je seznámení ve škole, přes společné přátele nebo v práci [Rosenfeld, Thomas 2012]. Mezi manželstvími uzavřenými v letech 2005–2012 v USA bylo 35 % započato online [Cacioppo, Cacioppo, Gonzaga, Ogburn, VanderWeele 2013] a 11 % všech amerických dospělých přiznává osobní zkušenost s online seznamkou [Smith, Duggan 2013]. Ač pro Česko nemáme příliš reprezentativních dat, vše napovídá tomu, že zájem o seznamky také narůstá [Syslová 2011]. Zatímco ještě pár let zpátky se k seznamování přes internet uchýlovali především vysokoškolsky vzdělaní lidé z větších měst, které lákala efektivita online seznamování, dnes se minimálně v USA sociodemografie seznamkových uživatelů do velké míry podobá běžné populaci lidí bez partnera a s přístupem k internetu [Sautter, Tippett, Morgan 2010]. Z okrajové aktivity se tedy stává norma – promění naši volbu partnera?

² Pro malý příklad možných analýz uživatelského chování na seznamce doporučuji blog Christiana Ruddera, matematika a zakladatele jedné z největších světových seznamek OkCupid (theblog.okcupid.com).

Současný výzkum výběrového párování v online prostředí se soustřeďuje především na dva aspekty. Zaprvé, zda již dříve identifikované šablony homogamního chování v offline světě odpovídají i párování v online prostředí. Zadruhé, pokud tendenci k výběrovému párování nacházíme, jak mohou zcela nová data ze seznamky rozšířit naše porozumění tomuto fenoménu. Na tyto dvě hlavní otázky se zaměřuje i má přehledová studie, která se skládá ze dvou sekcí. V první se soustřeďuji na existující vysvětlení výběrového párování, která dále rozšiřuji o aspekt online prostoru. Druhá sekce shrnuje dosavadní empirické poznání v problematice výběrového párování s ohledem na náboženství, etnicitu a vzdělání. Nejprve představuji tradiční výzkum „offline“ prostředí, kde cituji studie vycházející z výběrových šetření i z demografických dat. Navazuji prvními výzkumy online prostředí a názorně vysvětluji, jakým způsobem mohou data z online seznamek rozšířit naše porozumění výběrovému párování.

Vysvětlení výběrového párování: Teorie

Lidské preference hledat si partnery „sobě podobné“ si společnost všímá už od Aristotela a společenské vědy se tímto fenoménem zabývají nejméně od začátku 20. století [McPherson, Smith-Lovin, Cook 2001]. Obecně nazýváme vyšší frekvenci kontaktu s lidmi, kteří jsou si v určitých charakteristikách podobní, pojmem homofilie. Tu nepozorujeme jen v rámci partnerských vztahů, ale silně také v přátelství [Verbrugge 1983], v komunitách nebo v pracovních kolektivech [Ibarra 1997]. Největší akademický zájem vzbuzuje právě homogamie při výběru partnera, a to díky svému potenciálnímu dopadu na společnost. Vysoká homogamie ve společnosti může znamenat velkou uzavřenost jednotlivých skupin, zabraňovat sociální mobilitě a například u etnické homogamie i poukazovat na rasové napětí ve společnosti [Allport 1954]. Nejvíce akademické, ale i mediální pozornosti přitahuje možný dopad výběrového párování na stoupající nerovnost [např. The Economist 2014]. Ve chvíli, kdy bohatý jedinec spojí své vysoké příjmy s bohatým partnerem, rozdíl mezi takto vzniklým párem a chudým párem bude výrazně vyšší, než kdyby se bohatí a chudí partneři párovali náhodně. Homogamní manželství tak teoreticky mají potenciál zvyšovat nerovnost celé společnosti a nejen proto má smysl pochopit, co za nimi stojí [Schwartz 2013].

Pro tendence k výběrovému párování existuje rozsáhlá evidence. Vysvětlit a empiricky doložit mechanismy, které jsou za pozorovanými vztahy, je ale pro akademický výzkum stále výzvou. [Kalmijn 1998] shrnuje tři faktory, kterými společenské vědy výběrové párování vysvětlují: individuální, strukturální a sociální. [Schwartz 2013] tyto faktory dále rozvíjí o trendy, které vysvětlují zesílení nebo snížení výběrového párování v čase a s ohledem na proměny společnosti. V této části jejich práci shrnuji a rozšiřuji o hypotetické dopady online prostředí na výběrové párování.

Strukturální faktory

Tradiční sňatkový trh je přirozeně limitován příležitostmi fyzického kontaktu jednotlivců: vztahy se nejjednodušeji navazují s lidmi, které potkávejte pravidelně. Lidé málokdy hledají partnera účelově, seznámení je spíše vedlejší produkt jejich aktivit v rámci institucí a sociálních skupin, do kterých přirozeně patří [Kalmijn 1998]. Mezi nejčastější místa seznámení již dlouhou dobu patří škola, práce a přátelské skupiny, tedy místa, která se vyznačují velkou mírou homogenity a kde lidé tráví většinu svého bdělého času [Rosenfeld, Thomas 2012]. Kalmijn [1998] teoretizuje, že čím vyšší homogenita jednotlivců v rámci populace, tím nižší je pravděpodobnost kontaktu s jinou skupinou. Toto platí na úrovni státu, ale například i na úrovni školy nebo sousedství. Kromě velikosti sociální skupiny na sňatkovém trhu je důležité i její geografické rozložení: početně malá skupina s vysokou koncentrací na jednom místě bude jevit vyšší úroveň homogenie [Blau, Blum, Schwartz 1982]. Strukturální faktory jako příležitosti a omezení v hledání partnera tedy podporují výběrové párování bez ohledu na jednotlivcovy preference. Lidé musí svůj protějšek vybírat z jednotlivců, kteří jsou k dispozici.

Díky seznamování přes internet by strukturální faktory mohly hypoteticky zeslábnout [Rosenfeld, Thomas 2012]. Na online seznamkách má jednotlivec možnost vystoupit ze svých běžných sociálních kruhů a institucí a potkat potenciální partnery, které by v offline světě fyzicky neměl příležitost poznat. Právě tato skutečnost je důležitým důvodem, proč se lidé rozhodnou s online seznamováním začít: cítí, že v offline světě jsou jejich možnosti seznámení vyčerpány, a začnou hledat partnera účelově. Seznamka jim na pár kliků zprostředkuje kontakt s lidmi z jiných sociálních a třeba i geografických skupin. Pokud by výběrové párování bylo ovlivněno především strukturálními faktory, a nikoliv preferencemi, v rámci seznamování přes internet by se nemuselo projevovat nebo by mělo být slabší.

Individuální preference

Lidé hledající partnera se pohybují na tak zvaném sňatkovém trhu, kde své potenciální partnery hodnotí na základě socioekonomických a kulturních zdrojů [Becker 1973; Oppenheimer 1988]. Jakmile se dvojice sezdá, své zdroje zkombinují a odvíjí od nich například svůj společenský status. Ani jedna strana nechce manželstvím klesnout statusem a hledá si tak partnera s co nejvyšším příjmovým potenciálem. Kulturní zdroje jako hodnoty, společný jazyk nebo postoje nejsou na rozdíl od těch ekonomických při výběru partnera založené na maximalizaci, ale na podobnosti. Lidé preferují kulturně podobné partnery, protože s lidmi ze stejného kulturního prostředí je snadnější komunikace, trávení volného času i výchova dětí [Kalmijn 1994]. Maximalizace ekonomických zdrojů a preference pro podobnost v kulturních aspektech však v rámci sňatkového trhu v konečném důsledku obě vedou k homogamii. V prvním případě se bohatí lidé párují spolu, zatímco ti chudší na sebe zbydou. A přestože kulturní zdroje nemusí nutně znamenat podobnost v dalších charakteristikách (například v úrovni vzdělání nebo etnicity), často s nimi silně korelují a podporují tak výběrové párování v celé škále sociodemografických charakteristik [Kalmijn 1998].

Pokud jsou tedy dominantním mechanismem výběrového párování individuální preference, tak s ohledem na ekonomické i kulturní zdroje lze v online prostředí očekávat, že výběr partnera na základě těchto charakteristik by se mohl stát ještě efektivnějším [Potarca 2017]. Většina uživatelů online seznamek již na svém veřejném profilu uvádí nejvyšší dosažené vzdělání a u některých seznamek je možné zadat i majetek a úroveň příjmu. Stejná situace platí pro kulturní zdroje: součástí běžných profilů jsou i hodnotové ukazatele jako například náboženství. Seznamky, které své uživatele propojují pomocí algoritmů, tak většinou činí na základě odpovědí na postojové otázky. Uživatelé tedy vidí, kdo z potenciálních partnerů odpovídal na otázky podobně, a seznamka ty s nejvyšší shodou pak podporuje v individuálním kontaktu. Z pohledu fungování sňatkového trhu online seznamky tedy přináší lepší informovanost o potenciálních partnerech, a tak i snadnější možnost zvýšit nebo zachovat svůj ekonomický i kulturní status. Hypoteticky by tedy mělo výběrové párování v online prostředí přetrvávat a díky menším bariérám v hledání možná i sílit.

Se zvyšující se rovností mužů a žen se předpokládají další proměny sňatkového trhu, které ale budou homogamii spíše posilovat. Ve chvíli, kdy i od žen očekáváme příspěvek do rodinného rozpočtu, muži budou na svatebním trhu preferovat partnerky s vysokým příjmem [Sweeney, Cancian 2001]. V rovnější společnosti je také více příležitostí ženu se stejným socioekonomickým statutem potkat, jak fyzicky, tak v online prostředí. Počet žen na vysokých školách ve většině rozvinutých zemí předčil počty mužů a ženy čím dál častěji i během pracovního života vstupují do „mužských“ odvětví [Esteve, García-Román, Permanyer 2012]. Tyto dva trendy tedy mluví pro nárůst homogamie v rámci online i offline seznamování. Na druhou stranu, [Sweeney 2002] teoretizuje, že ve chvíli, kdy ženy nebudou závislé na příjmu svého partnera, mohou hledání partnera věnovat více času a jejich preference se mohou přeorientovat na jiné charakteristiky, než je příjem. To by homogamii alespoň v některých aspektech mohlo snižovat. Není ale vyloučené, že tyto všechny trendy působí současně: některé aspekty vyšší rovnosti mezi pohlavími výběrové párování zvyšují a další ho ve stejný moment snižují [Schwartz 2013].

Sociální faktory

Dalším faktorem, který podporuje výběrové párování, je působení sociálních skupin na jednotlivcův výběr partnera. Sociální skupiny mají tradičně zájem na tom, aby se jejich členové párovali mezi sebou – zajistí tak soudržnost skupiny i její kontinuitu do budoucích generací. U svých členů tedy podporují identifikaci se skupinou, která se může odvíjet od společné historie a hodnot nebo od jinakosti v porovnání s okolím, což nejčastěji pozorujeme u etnických skupin. Jednotlivcova síla identifikace se skupinou a tedy i sklon k homogamii se často odvíjí od uzavřenosti skupiny. Vedle pozitivní podpory k výběru partnera ze stejné skupiny jsou skupinové sankce druhým mechanismem, kterým sociální skupiny přispívají k výběrovému párování. Může se jednat o neformální sankce, například otevřený nesouhlas rodičů manželského páru se sňatkem, až o formální zákaz smíšeného manželství. To například historicky vidíme u církve,

kteřá nepovolovala manželství lidí rozdílného náboženství, nebo v USA, kde v období segregace nebylo dovoleno manželství mezi bělochy a černochoy [Kalmijn 1998].

Trendy v souvislosti s vlivem sociální skupiny na výběrové párování se často odvíjí od modernizační teorie. Vzhledem k tomu, že lidé v západoevropské společnosti svou identifikaci a úspěch čím dál častěji odvozují od získaných charakteristik (například vzdělání) oproti těm vrozeným nebo jinak připsaným (například rasa nebo náboženství), očekáváme pokles vlivu tradičních sociálních skupin jako církve nebo rodiny na výběr partnera. Naopak výběrové párování v rámci získaných charakteristik, především vzdělání, by mělo přetrvávat [Schwartz 2013; Rosenfeld, Thomas 2012]. Toto se samozřejmě velmi jednoduše může projevit i v rámci online seznamek, kde se navíc ve větší míře seznamují lidé nad 35 let, s již rozběhnutou kariérou, a vliv získaných charakteristik by tímto samovýběrem mohl být ještě posílen. Lze tedy očekávat, že na seznamkách najdeme silnější výběrové párování na základě vzdělání, ale vliv náboženství nebo etnicity jako vrozených charakteristik by měl být slabší.

Výběrové párování v datech

Jak měřit míru výběrového párování ve společnosti a mapovat jeho trendy? Jednou možností je absolutní míra homogamie, kdy se pomocí popisných nástrojů díváme na procento homogamních párů z celkového množství uzavřených manželství. Toto číslo je ale v čase náchylné ke zkreslení kvůli měnící se kompozici sociálních skupin v populaci [Schoen, Weinick 1993]. Pokud se kompozice začne měnit, například přibude více vysokoškoláků nebo se společnost stane více multikulturní, ve srovnáních absolutních čísel budeme pozorovat trendy reflektující změny v kompozici a ne ve sklonu k výběrovému párování. Většina výzkumů proto používá relativní míru výběrového párování, která kompozici společnosti zohledňuje a ukazuje šance jednotlivce na uzavření homogamního sňatku. Poměry šancí se pak využívají ke srovnání šance na homogamní vztah jednotlivce s určitou charakteristikou (např. žena s VŠ) a šance jednotlivce s odlišnou charakteristikou (např. žena se ZŠ). V praxi to znamená využití analytických nástrojů, které dokáží měřit šance a jejich poměry, například log-lineární modely a multinomické logistické regrese [Katrňák 2008].

Výběrové párování pozorujeme v mnoha charakteristikách – nejen sociodemografických, ale i osobnostních nebo fyzických. V empirické části se věnuji jen etnicitě, vzdělání a náboženství. Tímto výběrem zprvé navazuji na shrnutí [Kalmijn 1998] a [Schwartz 2013]. Zadruhé tak vybírám charakteristiky, pro které existuje dostatek evidence v offline výzkumech a zároveň se jim aspoň v nějaké míře věnují výzkumy online prostředí. Nejprve představuji závěry tradičního výzkumu výběrového párování, které povětšinou využívají data z výběrových šetření nebo demografická data a často prezentují vývoj výběrového párování v čase. Pokud uvádím procenta, odkazuji na absolutní míru výběrového párování. Pokud uvádím šance, jedná se o relativní míru. Následně navazuji prvními výzkumy online prostředí, které analyzují reálné uživatelské

chování. Tato data jsou vždy plně anonymizovaná a jejich analýza probíhá se souhlasem seznamovací služby³.

Náboženství

Výběrové párování s ohledem na náboženství je po dekády stabilně pozorovaným jevem napříč společnostmi. Blackwell a Lichter [2004] ukazují, že náboženská homogamie je v USA stále normou pro všechna náboženství a všechny druhy svazků od randění přes kohabitaci až po manželství. Silnou preferenci pro svazek mezi lidmi stejného náboženství pozorujeme mimo jiné i v Německu [Klein, Wunder 1996], v Nizozemsku [Kalmijn, Graaf, Janssen 2005], Irsku [O'Leary 2001] a Libanonu [Klat, Khudr 1986]. Nejvyšší míru homogamie nacházíme u uzavřených náboženských skupin, například u židů v USA (kolem 80 %) nebo u katolíků a protestantů v severním Irsku, což odpovídá teoretickým předpokladům plynoucím z jednotlivcovy identifikace se sociální skupinou [Kalmijn 1998]. Náboženská homogamie je samozřejmě nejrelevantnější v zemích s vysokou religiozitou obyvatelstva [Katrňák 2008]. V Česku se tématu náboženské homogamie výzkum příliš nevěnoval, i vzhledem k tomu, že Česko patří k nejsekulárnějším zemím v Evropě [Hamplová 2000].

Míra nábožensky smíšených manželství se v posledních dekadách zvyšuje, a to především u liberálnějších náboženských skupin a u lidí s vyšším vzděláním [Sherkat 2004]. Nacházíme tu oporu pro modernizační teorii, i když tento faktor bude do určité míry souviset i s klesajícím významem náboženství v některých západních společnostech [Katrňák 2008]. Velká výzkumná pozornost se věnuje také souvislosti mezi náboženskou homogamií a vyšší spokojeností manželů se svazkem, kterou především na amerických datech pozorujeme již několik generací [z novějších např. Curtis, Ellison 2002]. Myers [2006] si ale na longitudinálních datech všímá, že souvislost mezi náboženskou homogamií a manželskou spokojeností v posledních letech klesá. U mladých manželských párů zaprvé častěji pozorujeme náboženskou heterogamii a zadruhé u nich má tento druh svazku menší vliv na manželskou spokojenost. To by hypoteticky mohlo souviset se sekularizací společnosti i nižším významem sociálních skupin: manželé v nábožensky heterogamních svazcích nemusí čelit nepochopení nebo sankcím.

Výzkumů zaměřených na výběrové párování s ohledem na náboženství v online prostředí je velmi málo. Zčásti proto, že lidé na mainstreamových seznamkách často své náboženství neuvádějí. Ti jednotlivci, pro které je náboženství u partnera velmi důležité, se pravděpodobně budou rovnou seznamovat přes specializované náboženské seznamky jako je například JDate nebo Christian Connection, v českém prostředí pak třeba přes seznamku serveru Katolík.cz. Fiore a Donath [2005] analyzují profily a soukromé konverzace více jak 200 tisíc uživatelů anonymní

³ Ve většině případů zůstává jméno konkrétní seznamky známo jen konkrétním výzkumníkům – pokud se ve výzkumném textu jméno seznamky vyskytuje, vždy uvádím. Naprostá většina seznamek analyzuje data i pro své vlastní potřeby, interně nebo externě, a v obchodních podmínkách tuto skutečnost sdělují svým zákazníkům.

americké seznamky s ohledem na výběrové párování, od fyzického vzhledu po sociodemografické ukazatele. Potvrzují, že ve srovnání s dalšími sociodemografickými ukazateli je právě náboženství charakteristikou, kde uživatelé často volí možnost „preferuji neuvádět“.

Přesto nějakou oporu pro výběrové párování na základě náboženství v online prostoru nacházíme. Fiore a Donath [2005] vypočítávají o 17 % vyšší šanci na kontakt mezi lidmi se stejnou náboženskou afiliací, než jaká by vznikla náhodně. Podobné číslo ve svých analýzách pozorují i [Hitsch, Hortacsu, Ariely 2006], kteří mají k dispozici informace a uživatelské chování z více než 22 000 profilů online seznamky z amerického Bostonu a San Diega. V jejich datovém souboru mají vyšší preferenci pro partnery se stejným vyznáním ženy. Tyto výsledky reflektují i vlastní analýzy seznamky OKCupid – lidé si zpravidla vymění více zpráv s uživateli se stejným náboženstvím, ale tento trend neodpovídá všem náboženstvím a u většiny uživatelů není nijak silný [Rudder 2009].

Lze konstatovat, že míra náboženské homofilie v online prostředí je v porovnání s offline prostředím nízká. Vystávají ovšem otázky, nakolik to znamená větší otevřenost lidí na online seznamkách kontaktovat potenciální partnery z jiné náboženské skupiny a jak velkou roli hraje absence dat o náboženství u většiny profilů nebo existence specializovaných seznamek. Je pravděpodobné, že lidé na svých profilech náboženství hned neuvádí, ale v následných konverzacích nebo prvních schůzkách tato informace zazní a uživatelé se na jejím základě rozhodují, zda v kontaktu pokračovat. Srovnání sociodemografických ukazatelů u již vzniklých svazků totiž naznačuje, že jednotlivci preferují partnera stejného vyznání podobnou měrou, ať se seznámili online nebo offline. Potarca [2017] na amerických panelových datech ukazuje, že náboženská homogamie je stejná u párů, které se poznaly online a v heterogenních offline prostředích, například v práci. Je ale silnější u dvojic, které se potkaly v homogennějším kontextu, například skrz školu nebo rodinu. Toto zjištění tedy potvrzuje vliv strukturálních faktorů, na druhou stranu rozhodně nevylučuje osobní preferenci pro partnera stejného náboženství.

Vzdělání

Nejspíš nejvíce společenskovedních prací o výběrovém párování se zaměřilo na homogamii ve vzdělání, tedy preferenci pro partnera se stejným vzděláním. Nejvyšší dosažené vzdělání jednotlivce silně koreluje s jeho nebo jejím socioekonomickým statusem a tento typ výběrového párování tak ovlivňuje úroveň sociální mobility ve společnosti a proměňuje sociální struktury [Ultee, Luijckx 1990]. Evidenci pro homogamii s ohledem na vzdělání nacházíme napříč fází vztahu [Blackwell, Lichter 2004] i napříč společnostmi, od dlouhé tradice výzkumu v anglosaských zemích [například v USA: Rosenfeld 2008; v Británii: Halpin, Chan 2003] po západní Evropu [pro Španělsko: Esteve Palós, Cortina 2006; pro Francii: Goux, Maurin 2003], střední Evropu [Katrnak, Kreidl, Fonadova 2005], Latinskou Ameriku [Esteve, López, McCaa 2013] i jihovýchodní Asii [Smits, Park 2009].

Šablona vzdělanostní homogamie se ale napříč zeměmi liší, což ve svém výzkumu srovnávajícím 65 zemí potvrdili i [Smits, Ultee, Lammers 1998]. Ukazuje se, že demokratické společnosti mají nižší míru vzdělanostní homogamie, stejně jako země s protestantskými kořeny. Zajímavý je závěr ohledně ekonomického rozvoje, který má na výběrové párování s ohledem na vzdělání zřejmě non-lineární vztah (ve tvaru obráceného U). Nejdříve s růstem ekonomické úrovně roste i výběrové párování, v určitý bod se ale zlomí a klesá. Smits s kolegy [1998] se domnívají, že s rostoucí industrializací nahrazuje úroveň vzdělání u partnera připsané charakteristiky a vliv rodičů, v souladu s modernizační teorií. Jak ale společnost bohatne, význam statusu se snižuje, do popředí vstupují individuální faktory a výběrové párování upadá.

S mezinárodním srovnáním na [Smits et al. 1998] navazují [Domański, Przybysz 2007], kteří porovnávají vzdělanostní homogamii ve 22 evropských zemích. Nejvyšší úroveň výběrového párování nachází v postkomunistických společnostech jako je Slovensko, Česká republika nebo Polsko, kde autoři dále rozvíjí modernizační teorii tím, že odklon od komunismu ke kapitalismu zvýšil provázanost vzdělání se socioekonomickým statusem, a tedy preferenci pro vzdělanostní homogamii. Dle [Katrňák 2008] tento závěr není zcela přesný. Na českých datech z let 1974 až 2004 ukazuje, že „pravidlo vzdělanostní homogamie proměna společenského uspořádání od socialismu ke kapitalismu nepoznamenává. Pouze posiluje jeho intenzitu. Vysokoškolsky vzdělaní muži a ženy měli největší šance na vzdělanostně homogamní sňatek v socialismu a mají je i dnes“ [Katrňák 2008: 88].

Mnohem náročnější otázkou je, zda se výběrové párování s ohledem na vzdělání s časem zvyšuje nebo snižuje, a to i kvůli rozdílným přístupům k metodologii. Otázka trendů zaměstnává především americké výzkumníky. Absolutní míra vzdělanostní homogamie v USA roste vzhledem k lepšímu přístupu žen ke vzdělání, zda roste i relativní míra homogamie je nejasné: například [Hou, Myles 2008] pozorují nárůst relativní vzdělanostní homogamie, další spíše stabilní nebo klesající trend [Fu, Heaton 2008]. Česká data ale ukazují, že trend homogamie je za poslední tři dekády spíše stabilní a homogamních je necelých 60 % svazků [Katrňák 2008].

I v online prostředí nacházíme evidenci pro vzdělanostní homofilii. Vše ale nasvědčuje tomu, že neplatí vždy a pro všechny sociodemografické skupiny. Dle již zmíněného výzkumu americké seznamky od [Hitsch et al. 2006] jsou to především ženy, které hledají muže se stejným vzděláním. Pokud se žena rozhodne muže na seznamce oslovit jako první, s největší pravděpodobností kontaktuje muže s podobným vzděláním, zejména jedná-li se o vysokoškolačku. Homogamní šablonu hledání partnera nachází [Hitsch et al. 2006] i u mužů se středoškolským vzděláním, ale ne u mužů-vysokoškoláků, kteří stejně často kontaktují i ženy s nižším vzděláním.

Skopek, Schulz a Blossfeld [2011] se zaměřují na prvotní kontakt uživatelů německé seznamovací služby: jejich data zahrnují informace z uživatelských profilů a konverzace mezi nimi. Nacházejí potvrzení pro homofilní výběr partnera na základě vzdělání, tedy že podobnost ve vzdělanostní úrovni signifikantně zvyšuje šanci prvního kontaktu mezi dvěma uživateli. Stejně jako u [Hitsch et al. 2006] pozoruje větší ochotu mužů kontaktovat ženy s nižším vzděláním, ale ženy jen výjimečně odepisují mužům, kteří mají nižší vzdělání než ony.

Hitsch, Hortacsu a Ariely [2008] dále porovnávají míru výběrového párování mezi vzorkem uživatelů již zmíněné seznamky v Bostonu a San Diegu s daty reálných sezdaných párů v těchto dvou metropolích, které mají ze sčítání lidu. Za pomoci Gale-Shapley modelu zkoumají, zda by dynamika na online seznamce vytvořila stejnou šablonu výběrového párování, jako pozorují v offline prostoru. Uzavírají, že s ohledem na věk a etnicitu by se tak stalo, ale u vzdělání by výsledná míra výběrového párování byla zřejmě nižší. Ač i na seznamce ukazují lidé preferenci pro partnera se stejným vzděláním, vysoká míra (60 %) výběrového párování s ohledem na vzdělání ve vybraných dvou metropolích je zřejmě ovlivněna strukturálními faktory, konkrétně seznamováním v rámci institucí, kde lidé potkávají především protějšky stejné vzdělanostní úrovně.

Z uvedených zjištění vyplývá, že v online prostoru sklon k výběrovému párování s ohledem na vzdělání nacházíme, avšak tíhnou k němu spíše ženy než muži. Roli mohou hrát sociální faktory, například větší stigma pro ženy, které naváží vztah s méně vzdělaným, a tedy statusově nižším, mužem. Nebo pozorujeme dopad individuálních faktorů, kdy žena bere nízké vzdělání jako signál nižší finanční bonity potenciálního partnera. Na druhou stranu, výběrové párování s ohledem na vzdělání je na internetu v celkové míře zřejmě nižší, než pokud se lidé seznamují offline, což poukazuje na vliv strukturálních faktorů: střední a vysoká škola jsou jedny z nejčastějších míst k seznámení [Rosenfeld, Thomas 2012].

Etnicita

Dalším silným faktorem výběrového párování je etnicita. Bez ohledu na to, zda jednotlivec hledá životního partnera nebo nezávazné rande, jeho preference pro partnera ze stejné etnické skupiny je velmi silná [Blackwell, Lichter 2004]. Míra mezietnického manželství bývá často považována za dobrý ukazatel vztahu etnických skupin na celospolečenské úrovni a tudíž se téma etnické homogamie stalo výzkumně atraktivním tématem [Kalmijn 1998].

I vzhledem k historii země proběhlo nejvíce výzkumu zaměřeného na homogamii rasy a etnicity v USA. Rosenfeld [2008] dochází k závěru, že rasa je v USA „pořád nejsilnější sociální bariérou na sňatkovém trhu“. Přestože v průzkumech veřejného mínění se proti smíšeným manželstvím vyjadřuje jen méně než 10 % Američanů, reálně k nim dochází jen velmi málo [Wang 2012]. Americký běloch má přibližně čtyřikrát větší pravděpodobnost homogamie než heterogamie [Kopf 2016]. Rasovou homogamii nacházíme i v dalších zemích s vyšší etnickou heterogenitou obyvatelstva, například v Německu [González-Ferrer 2005], Švédsku [Dribe, Lundh 2008] a Nizozemí [Kalmijn, van Tubergen 2006]. V Česku výzkum zaměřený na etnickou homogamii nevznikl, ale na Slovensku se tématu věnovali [Fónadová, Katrňák 2016]. U slovenských manželství uzavřených mezi lety 1992–2012 byla šance na etnický homogamní sňatek stabilně výrazně vyšší než na heterogamní, bez větších výkyvů a nejsilněji mezi Romy. Vzhledem k podobným reáliím v Česku se dá usuzovat, že i zde bude homogamie s ohledem na rasu a etnicitu přítomná a pravděpodobně vysoká.

I přestože výběrové párování v rámci rasy a etnicity je ve srovnání s dalšími charakteristikami výběrového párování velmi silné, analýzy posledních dekád ukazují lehce se zvyšující tendenci lidí tvořit mezietnické páry, a to jak v USA [Qian, Lichter 2007], tak v západní Evropě [Lucassen, Laarman 2009]. Vyšší ochotu navazovat mezietnické vztahy pozorujeme především u mužů a u mladších a vzdělanějších jednotlivců [Yancey 2007; Feliciano, Robnett, Komaie 2009]. Tento trend sice odpovídá předpokladům modernizační teorie, ale zatím se jedná o příliš malý efekt, který se navíc neprojevuje na všech datových souborech.

Do výběru partnera mimo svou etnickou skupinu se promítá i faktor společenské hierarchie etnických skupin, který různá etnika seřazuje do pomyslného žebříčku dle společenské vzdálenosti jednotlivých skupin [Gordon 1964]. Na základě této hierarchie by nejvíc preferovanými partnery v USA měli být běloši, následně Hispánci a Asiaté a na samém konci černoši. Právě vůči černochům jsou stále nejvíce rozšířené negativní postoje, především kvůli dlouhé historii rasismu mezi bělošskou většinou a černošskou menšinou [Feliciano et al. 2009]. V Evropě jsou Hispánci (Španělé) a Asiaté preferováni jako partneři spíše než Arabové nebo Afričané, vzhledem ke kulturním souvislostem a nově i negativním sentimentům plynoucím z imigrantských vln [Potárcá, Mills 2015]. Hierarchie etnických skupin je patrná i ze slovenských dat. Nejmenší pravděpodobnost heterogenních sňatků je mezi Slováky a Romy, jejichž vztahy jsou dlouhodobě vyhrocené, a sociální faktory, jako neformální sankce za sňatky s druhou skupinou, mohou hrát důležitou roli. U Maďarů, Čechů nebo Poláků žijících na Slovensku je sňatková heterogenita se Slováky výrazně vyšší [Fónadová, Katrňák 2016]. Je ale důležité mít na paměti, že definice etnicity je flexibilní a může se projevit v odpovědích respondentů: etnicky smíšená rodina se může považovat za Slováky, i když bychom ji mohli definovat i třeba jako romsko-maďarskou.

Velmi podobné trendy partnerských preferencí nacházíme i v prvních studiích romantického chování v online prostředí. Hitsch s kolegy [2006] v analýzách vychází z dat anonymní americké seznamky, která se na začátku seznamovacího procesu uživatelů ptá na preferenci etnicity partnera. Skoro 40 % žen uvádí, že hledá partnera stejné etnicity. To samé tvrdí jen 18 % mužů. Toto číslo se navíc liší v závislosti na etnicitě uživatele, nejvyšší preferenci pro stejnou etnicitu mají běloši, nejnižší černoši. Zadaná preference odpovídá i uživatelskému chování na samotné seznamce: i po kontrole dalších proměnných se ukazuje, že uživatelé mají silnou preferenci pro kontakt s lidmi stejné etnicity. A to i ti, kteří v úvodním dotazníku deklarují, že jim na etnicitě nezáleží.

Rozdíly v postojích k mezietnickému romantickému kontaktu různých sociodemografických skupin zkoumá také [Yancey 2007] pomocí analýz seznamky amerického serveru Yahoo, kde uživatelé zadávají seznamovací inzeráty. Nejvýraznějším faktorem se stejně jako u [Hitsch et al. 2006] ukazuje být pohlaví: muži jsou k navazování mezietnických vztahů přes online seznamku otevřenější než ženy. Na rozdíl od výzkumů offline svazků se v inzerátech Yahoo neukázaly rozdíly v postojích k mezietnickým vztahům mezi vzdělanostními nebo věkovými skupinami. Především vliv věku je ale pravděpodobně zkreslen demografií uživatelů Yahoo, kde jsou v převažující míře mladší ročníky.

Data ze seznamky serveru Yahoo používá také [Feliciano et al. 2009], který potvrzuje, že v online prostředí jsou k romantickému kontaktu s jinou etnickou skupinou otevřenější muži

než ženy. V datech se ukazuje i určitá rasová hierarchie: bílí Američané jsou nejotevřenější směrem k Hispáncům, na druhém konci spektra jsou lidé z Blízkého východu a jižní Indie. Toto zjištění odpovídá klasické verzi teorie asimilace, protože právě lidé z těchto zemí patří v USA mezi nové etnické skupiny.

Do analýz rasových preferencí svých uživatelů se pustila i jedna z největších amerických seznamek OkCupid. Její uživatelé vyplňují dotazník, na základě kterého jim pak seznamka nabízí protějšky, kteří odpovídali podobně. Zatímco v roce 2008 udávalo v dotazníku 42 % procent jejích uživatelů silnou preferenci pro vztah s partnerem stejné etnicity, v roce 2014 to už bylo jen 31 %. Pokud se ale podíváme na hodnocení atraktivity uživatelů, které seznamka také nabízí, rasové preference jednotlivců se mezi lety 2008 a 2014 příliš neproměnily: ženy všech ras nejlépe hodnotí muže stejné rasy, muži jsou k jiným etnicitám obecně otevřenější, ale penalizují černošky [Rudder 2014]. Ukazuje se tedy podobný trend jako v analýzách [Yanceyho 2007], [Feliciana et al. 2009] i [Hitsch et al. 2006].

Potárcá a Mills [2015] kombinují data seznamky eDarling z několika evropských zemí s údaji o velikosti menšin napříč Evropou a nabízí tak unikátní mezinárodní srovnání partnerských preferencí v online prostředí. Podobně jako v USA, i Evropané bez ohledu na státní příslušnost preferují vztah s partnery ze stejné etnické skupiny. Čím větší velikost má etnická skupina v dané zemi, tím větší je preference jejich členů párovat se v rámci ní. Potvrzuje se i hierarchie etnických skupin napříč Evropou a lehce překvapivě se ukazuje silná preference pro navázání vztahu s Evropany u lidí všech ras a etnik. Zde je však důležité omezení kategorie „Evropan“, která nedovoluje analyzovat národnostní rozdíly v preferenci uživatelů, které se jistě vyskytují. V souladu s teorií asimilace má přítomnost cizinců v zemi pozitivní dopad na otevřenost lidí k romantickému vztahu s menšinou i v online prostředí. Zatímco v homogenním Polsku jsou preference pro partnera ze stejné etnické skupiny silné, v heterogenním Švýcarsku jsou lidé mezi-etnickým vztahům více otevřeni. Toto zhruba odpovídá i procentu mezi-etnických sňatků v daných zemích.

Zajímavou nástavbou na výzkumy jednotlivých sociodemografických ukazatelů v rámci výběrového párování je jejich vzájemná interakce, kterou seznamky umožňují analyzovat. Lin a Lundquist [2013] mají k dispozici více jak 9 milionů profilů a 200 milionů zpráv z jedné z největších amerických seznamek, kterou ale nejmenují. Ve svých analýzách potvrzují existenci rasové hierarchie, která odpovídá teoretickým předpokladům: nejvíce odpovědí od všech etnických skupin získávají běloši, uprostřed jsou Asiaté a Hispánci, černoši pak bývají nejčastěji ignorováni. Stejně jako v předchozích studiích, i v tomto případě jsou muži mezi-etnickému kontaktu otevřenější než ženy – je ale otázkou, zda je za tím větší rasová tolerance nebo naopak předpoklad, že ženy jiných etnik mají tradičnější přístup k manželství. Důležitým zjištěním je, že vzdělání neovlivňuje uživatelské etnické preference: vysokoškolská běloši obou pohlaví spíše reagují na bělocha bez vysokoškolského vzdělání než na černocha s vysokou školou.

Výzkumy výběrového párování s ohledem na etnicitu v rámci online seznamek zatím do velké míry potvrzují závěry, které pozorujeme v „tradičních“ výzkumech tohoto fenoménu. Rozšiřují je ale o zajímavá poznání: lidé mají pro partnera ze své etnické skupiny silnou preferenci

i ve chvíli, kdy mají díky online prostoru k seznámení s různými etnicitami jednoduchý přístup. Individuální preference je v tomto případě zřejmě silnější, než jsou strukturální faktory. Ženy upřednostňují partnery stejné etnicity častěji než muži a vše nasvědčuje tomu, že do výběru partnera v online prostředí se promítá nejen homofilie, ale i rasová hierarchie. Výběrové párování s ohledem na etnicitu je natolik silné, že ho příliš neredukuje homofilie ve vzdělávání, ale pravděpodobně ho snižuje celková společenská otevřenost k jiným národnostem. Podobně jako tradiční výzkumy, i data z online seznamek ukazují rozdíl mezi deklarovanou postojovou otevřeností lidí a jejich reálným chováním na seznamce. Online data navíc ukazují, že se jedná o tytéž uživatele: v jednu chvíli zadají, že jsou mezietnickému kontaktu otevřeni, za pár minut už ale komunikují jen s lidmi stejné etnicity.

Závěr

Predikce ohledně dopadu internetu na proces hledání partnera se dají shrnout do dvou v zásadě protichůdných očekávání. První skupina teoretizuje, že jednoduchý přístup k tisícům potenciálních partnerů napříč všemi sociodemografickými skupinami a hledání partnera zbavené přímého vlivu prostředníků, jako je vlastní komunita, povede k větší diverzitě vznikajících partnerství a výběrové párování tak bude klesat [Rosenfeld 2008]. Druhý tábor očekává opak. Vychází z předpokladu, že lidé obecně preferují kontakt s podobnými partnery a na online seznamkách mohou efektivněji filtrovat protějšky podle preferovaných charakteristik a zamezit tak kontaktu s lidmi, kteří neodpovídají jejich představám [Potarca 2017].

Empirický výzkum výběrového párování na internetu, především na online seznamkách, je stále v počátcích. Zatím vcelku konzistentně nachází podporu pro výběrové párování v online prostoru na základě hlavních sociodemografických charakteristik (etnicita, náboženství, vzdělání), a potvrzuje tak závěry tradičních výzkumů homogamie. Ač mají uživatelé internetu ničím neomezený přístup k potenciálním partnerům různorodých charakteristik, vybírají mezi nimi ty sobě podobné. Homogamní šablony, které pozorujeme online i offline, se tedy zřejmě nedají vysvětlit jen strukturálními faktory a příležitostmi k seznámení. Osobní preference jednotlivců pro podobnost je důležitá.

Stejně jako offline, obzvláště silnou tendenci k výběrovému párování online pozorujeme u etnicity [Hitsch et al. 2006]. Do internetového seznamování se promítají známé šablony rasové hierarchie a v zemích s vyšší tolerancí vůči různým etnikům vidíme i vyšší ochotu uživatelů online seznamek komunikovat s lidmi jiných etnicit [Potárcá, Mills 2015]. Zajímavé je, že ženy jsou k mezietnickému kontaktu méně otevřené než muži [Yancey 2007]. Podobně je tomu i u vzdělání, kde podporu pro výběrové párování nacházíme také disproporčně u žen. Jsou to především vysokoškolačky, které se neseznamují s muži s nižším statusem [Hitsch et al. 2006]. Přestože výběrové párování s ohledem na vzdělání online pozorujeme, vše nasvědčuje tomu, že v menší míře než u offline seznamování. Strukturální faktory tedy budou vysvětlovat aspoň část trendu, který se vyskytuje v offline datových souborech.

Trochu odlišný příběh o seznamování v online prostoru vypráví výběrové párování s ohledem na náboženství. Zatímco při offline seznamování je náboženská homogamie silná, při prvních kontaktech na online seznamkách tomu tak není. Výzkumy již vzniklých párů ale potvrzují, že i vztahy navázané online jsou nábožensky homogamní [Potarca 2017]. Za slabým trendem tak spíše může stát malá ochota uživatelů vyplňovat náboženské vyznání do svého profilu nebo seznamování nábožensky založených lidí na specializovaných seznamkách [Fiore, Donath 2005]. Stejně jako u předchozích charakteristik, i podobnost v náboženství preferují na seznamkách více ženy [Hitsch et al. 2006]. Vyšší tendence žen k výběrovému párování je sama o sobě zajímavým pozorováním. V souvislosti s interpretací [Kalmijn 1998] se nabízí dát tuto tendenci do souvislosti s nižší mírou hypogamie u žen a vidět etnicitu, socioekonomický status a do určité míry i náboženství jako ukazatel statusu, na který by mohly být ženy citlivější. Tyto závěry jsou ale na úrovni hypotéz a jejich potvrzení by vyžadovalo spíše kvalitativní výzkum.

Podle dostupných výzkumů uživatelského chování na seznamkách internet žádnou revoluci ve výběrovém párování nepřináší. Toto shrnutí dále potvrzuje [Potarca 2017], která srovnává míru endogamie u německých a amerických párů, které se poznaly online, s těmi, které se potkaly v tradičnějším prostředí. Šablony výběrového párování nachází u všech párů, bez ohledu na místo seznámení. Nejsilnější sklon k endogamii pozoruje u párů, které se seznámily ve škole nebo skrz rodiny, slabší je naopak u těch, které se potkaly online nebo skrz přátele. Výsledky však může zkreslovat relativně malá velikost vzorku párů, které se poznaly online – u obou zemí je to méně než 10 % z celkového počtu analyzovaných párů. I přesto nic nenasvědčuje tomu, že by výběrové párování v online prostředí bylo výrazně jiné než offline.

Samotné seznamky navíc samy cítí, že lidé preferují kontakt s podobnými protějšky. Některé seznamky se specializují pouze na seznamování na základě jednotlivých charakteristik. Můžou to být náboženské seznamky (JDate.com pro židy nebo Muslima.com pro muslimy), elitní seznamky (třeba The League zaměřená jen na vyšší socioekonomické třídy) nebo seznamky zaměřené na určitý životní styl (VeggieDate.org pro vegetariány, SmokingPassions.com pro kuřáky, Fitness-Singles.com pro sportovce). I velké mainstreamové seznamky typu OKCupid nebo eDarling, které využívají párovací algoritmus, propojují své uživatele na základě podobnosti.

Samozřejmě je ale příliš brzy na jednoznačné závěry. Je nutné zdůraznit, že tento článek vychází jen z pár prvních výzkumů, které se opírají o data několika seznamovacích služeb. Závěry jsou tak platné primárně jen pro určitou část seznamkového trhu. Co se děje na seznamkách, jejichž majitelé s výzkumníky spolupracovat nechtějí nebo nemají tu možnost, s jistotou nevíme. Také se z velké části opírám o výzkumy zaměřené na USA, kde je seznamování přes internet zřejmě rozšířenější než v jiných rozvinutých zemích, a závěry tak nemusí být přenositelné do dalších kontextů. V neposlední řadě je důležité připomenout, že porovnávat trendy výběrového párování sezdaných dvojic s výběrovým párováním v samotných začátcích seznamovacího procesu na seznamce znamená srovnávat dvě odlišné fáze vztahu. Rozdíl mezi nimi tak nemusí být nutně zapříčiněn online prostorem.

I přes všechny limitace dokáže výzkum online seznamek významně rozšířit naše poznání o šablonách výběrového párování. Nejen kvůli porovnání již zaznamenaných trendů, ale především jako hloubková sonda do mechanismu výběrového párování. Strukturální a sociální faktory rozhodně nestačí k vysvětlení výběrového párování a individuální faktory budou mít důležitý vliv. Z prvních analýz také vyplývá, že jsou to spíše ženy, které upřednostňují sobě podobného partnera [Hitsch et al. 2008]. Data ze seznamkových profilů dále nabízí analýzy homofilie na základě charakteristik, které nejsou v klasických výzkumných datech běžné, a jsou tedy do velké míry neprozkoumané, i přestože pro uživatele mohou být klíčové během filtrování potenciálních partnerů. Jedním z důležitých ukazatelů se ukazuje být například přístup k rodině (chuť mít děti a počet dětí z předchozích vztahů), typ postavy (například „atletická“) a zda je uživatel kuřák. V analýzách uživatelského chování na seznamce [Fiore, Donath 2005] měly tyto charakteristiky v hledání partnera větší váhu než ukazatele typu vzdělání nebo náboženství.

Proměny, které internet vnáší do sňatkového trhu, se určitě vyplatí dále zkoumat i s ohledem na fakt, že internet čím dál silněji vstupuje do našich životů. Velký smysl by měl kvalitativní výzkum, který by zkoumal racionalizaci uživatelského chování na seznamkách, ale třeba i motivační nebo naopak bariéry pro využití seznamovacích služeb. Závěry z prvních výzkumů je třeba validovat na dalších datových souborech. Pro tento účel je důležitá spolupráce se seznamkami, které mají bohatá data reálného uživatelského chování. Důležité rozšíření by přinesl i výzkum českého kontextu, kde je online seznamování méně rozšířené než v USA a rodinný život se opírá o jiné historické reálie. Dalším zajímavým výzkumným směrem je trajektorie vztahu, který vznikl online. Potvrdí se první výzkumy, že co se délky vztahu a spokojenosti v něm týče, seznámení na seznamce je spojeno s mírně lepšími výsledky než seznámení offline [Cacioppo et al. 2013]?

O AUTORECH



Markéta Šetinová je socioložka a psychoterapeutka. Absolvovala magisterské studium sociologie na univerzitě v Oxfordu, nyní pokračuje s doktorským výzkumem při Institutu sociologických studií na Fakultě sociálních věd UK. Ve svém výzkumu i psychoterapeutické praxi se věnuje především seznamování, singles a partnerským vztahům. Markétu Šetinovou můžete kontaktovat na marketa@modernilaska.cz.

L I T E R A T U R A

- Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, Mass.: Addison-Wesley Pub. Co.
- Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46. <https://doi.org/10.1086/260084>.
- Blackwell, Debra L., Daniel T. Lichter. 2004. "Homogamy Among Dating, Cohabiting, and Married Couples." *Sociological Quarterly* 45 (4): 719–37. <https://doi.org/10.1111/j.1533-8525.2004.tb02311.x>.
- Blau, Peter M., Terry C. Blum, Joseph E. Schwartz. 1982. "Heterogeneity and Inter-marriage." *American Sociological Review* 47 (1): 45–62. <https://doi.org/10.2307/2095041>.
- Cacioppo, John T., Stephanie Cacioppo, Gian C. Gonzaga, Elizabeth L. Ogburn, Tyler J. VanderWeele. 2013. "Marital Satisfaction and Break-Ups Differ across on-Line and off-Line Meeting Venues." *Proceedings of the National Academy of Sciences* 110 (25): 10135–40. <https://doi.org/10.1073/pnas.1222447110>.
- Curtis, Kristen Taylor, Christopher G. Ellison. 2002. "Religious Heterogamy and Marital Conflict: Findings From the National Survey of Families and Households." *Journal of Family Issues* 23 (4): 551–76. <https://doi.org/10.1177/0192513X02023004005>.
- Domański, Henryk, Dariusz Przybysz. 2007. "Educational Homogamy in 22 European Countries." *European Societies* 9 (4): 495–526. <https://doi.org/10.1080/14616690701314119>.
- Dribe, Martin, Paul Nystedt. 2013. "Educational Homogamy and Gender-Specific Earnings: Sweden, 1990–2009." *Demography* 50 (4): 1197–1216. <https://doi.org/10.1007/s13524-012-0188-7>.
- Esteve, Albert, Joan García-Román, Iñaki Permanyer. 2012. "The Gender-Gap Reversal in Education and Its Effect on Union Formation: The End of Hypergamy?" *Population and Development Review* 38 (3): 535–46. <https://doi.org/10.1111/j.1728-4457.2012.00515.x>.
- Esteve, Albert, Luis Ángel López, Robert McCaa. 2013. "The Educational Homogamy Gap between Married and Cohabiting Couples in Latin America." *Population Research and Policy Review* 32 (1): 81–102. <https://doi.org/10.1007/s11113-012-9263-4>.
- Esteve Palós, Albert, Clara Cortina. 2006. "Changes in Educational Assortative Mating in Contemporary Spain." *Demographic Research* 14 (17): 405–28. <https://doi.org/10.4054/DemRes.2006.14.17>.
- Feliciano, Cynthia, Belinda Robnett, Golnaz Komaie. 2009. "Gendered Racial Exclusion among White Internet Daters." *Social Science Research* 38 (1): 39–54. <https://doi.org/10.1016/j.ssresearch.2008.09.004>.
- Fiore, Andrew T., Judith S. Donath. 2005. "Homophily in Online Dating: When Do You Like Someone Like Yourself?" In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, 1371–1374. CHI EA '05. New York, NY, USA: ACM. <https://doi.org/10.1145/1056808.1056919>.

- Fónadová, Laura, Tomáš Katrňák. 2016. "Etnická homogamie na Slovensku v letech 1992 až 2012." *Sociologický časopis / Czech Sociological Review* 52 (2): 157–78. <https://doi.org/10.13060/00380288.2016.52.2.245>
- Fu, Xuanning, Tim B. Heaton. 2008. "Racial and Educational Homogamy: 1980 to 2000." *Sociological Perspectives* 51 (4): 735–58. <https://doi.org/10.1525/sop.2008.51.4.735>.
- Gordon, Milton Myron. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. Oxford University Press.
- Goux, Dominique, Eric Maurin. 2003. "Who Marries Whom In France?" In *Who Marries Whom?*, edited by Hans-Peter Blossfeld and Andreas Timm, 57–78. European Studies of Population 12. Springer Netherlands. https://doi.org/10.1007/978-94-007-1065-8_4.
- Halpin, Brendan, Tak Wing Chan. 2003. "Educational Homogamy in Ireland and Britain: Trends and Patterns*." *The British Journal of Sociology* 54 (4): 473–95. <https://doi.org/10.1111/j.1468-4446.2003.00473.x>.
- Hamplová, Dana. 2000. "Šetření ISSP 1998 - Náboženství / International Social Survey Program 1998 - Religion." *Sociologický Časopis / Czech Sociological Review* 36 (4): 431–40.
- Hitsch, Günter J., Ali Hortacsu, Dan Ariely. 2006. "What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating" [online]. Rochester, NY: Social Science Research Network [cit. 1. 6. 2018]. Dostupné z: <https://papers.ssrn.com/abstract=895442>.
- Hitsch, Günter J. 2008. "Matching and Sorting in Online Dating." [online]. Rochester, NY: Social Science Research Network [cit. 1. 6. 2018]. Dostupné z: <https://papers.ssrn.com/abstract=1113243>.
- Hou, Feng, John Myles. 2008. "The Changing Role of Education in the Marriage Market: Assortative Marriage in Canada and the United States since the 1970s" [online]. *Canadian Journal of Sociology* 33 (2) [cit. 1. 6. 2018]. Dostupné z: <https://journals.library.ualberta.ca/cjs/index.php/CJS/article/view/551>.
- Ibarra, Herminia. 1997. "Paving an Alternative Route: Gender Differences in Managerial Networks." *Social Psychology Quarterly* 60 (1): 91–102. <https://doi.org/10.2307/2787014>.
- Kalmijn, Matthijs. 1994. "Assortative Mating by Cultural and Economic Occupational Status." *American Journal of Sociology* 100 (2): 422–52. <https://doi.org/10.1086/230542>
- Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology*, 1998. <https://doi.org/10.1146/annurev.soc.24.1.395>
- Kalmijn, Matthijs, Paul M. de Graaf, Jacques P. G. Janssen. 2005. "Intermarriage and the Risk of Divorce in the Netherlands: The Effects of Differences in Religion and in Nationality, 1974–94." *Population Studies* 59 (1): 71–85. <https://doi.org/10.1080/0032472052000332719>.
- Katrňák, Tomáš. 2008. *Spříznění Volbou? Homogamie a Heterogamie Manželských Párů v České Republice*. Praha: Slon.

- Katrnak, Tomas, M. Kreidl, Laura Fonadova. 2005. "Trends in Educational Assortative Mating in Post-Socialist Central Europe: Czech Republic, Slovakia, Poland, and Hungary Between 1988 and 2000." [online]. USA: California Center for Population Research [cit. 1. 6. 2018]. Dostupné z: <http://escholarship.org/uc/item/3dz4r17q>.
- Klat, Myriam, Adèle Khudr. 1986. "Religious Endogamy and Consanguinity in Marriage Patterns in Beirut, Lebanon." *Social Biology* 33 (1–2): 138–45. <https://doi.org/10.1080/19485565.1986.9988631>.
- Klein, Thomas, Edgar Wunder. 1996. "Regionale Disparitäten Und Konfessionswechsel Als Ursache Konfessioneller Homogamie." *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie* 48: 96–125.
- Kopf, Dan. 2016. "Why Is Interracial Marriage on the Rise?" [online]. USA: Priceonomics [cit 1. 6. 2018]. Dostupné z: <https://priceonomics.com/why-is-interracial-marriage-on-the-rise/>.
- Lin, Ken-Hou, Jennifer Lundquist. 2013. "Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education." *American Journal of Sociology* 119 (1): 183–215. <https://doi.org/10.1086/673129>.
- Lucassen, L., C. Laarman. 2009. "Immigration, Inter marriage and the Changing Face of Europe in the Post War Period." *History of the Family* 14 (1): 52–68. <https://doi.org/10.1016/j.hisfam.2008.12.001>.
- McPherson, Miller, Lynn Smith-Lovin, James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (1): 415–44. <https://doi.org/10.1146/annurev.soc.27.1.415>.
- Myers, Scott M. 2006. "Religious Homogamy and Marital Quality: Historical and Generational Patterns, 1980 – 1997." *Journal of Marriage and Family* 68 (2): 292–304. <https://doi.org/10.1111/j.1741-3737.2006.00253.x>.
- O'Leary, Richard. 2001. "Modernization and Religious Inter marriage in the Republic of Ireland." *The British Journal of Sociology* 52 (4): 647–65. <https://doi.org/10.1080/00071310120084517>.
- Oppenheimer, Valerie Kincade. 1988. "A Theory of Marriage Timing." *American Journal of Sociology* 94 (3): 563–91. <https://doi.org/10.1086/229030>.
- Potârca, Gina. 2017. "Does the Internet Affect Assortative Mating? Evidence from the U.S. and Germany." *Social Science Research* 61 (January): 278–97. <https://doi.org/10.1016/j.ssresearch.2016.06.019>.
- Potârca, Gina, Melinda Mills. 2015. "Racial Preferences in Online Dating across European Countries." *European Sociological Review* 31 (3): 326–41. <https://doi.org/10.1093/esr/jcu093>.
- Rosenfeld, Michael J. 2008. "Racial, Educational and Religious Endogamy in the United States: A Comparative Historical Perspective." *Social Forces* 87 (1): 1–31. <https://doi.org/10.1353/sof.0.0077>.

- Rosenfeld, Michael J., Reuben J. Thomas. 2012. "Searching for a Mate: The Rise of the Internet as a Social Intermediary." *American Sociological Review* 77 (4): 523–47. <https://doi.org/10.1177/0003122412448050>.
- Rudder, Christian. 2009. "How Races and Religions Match in Online Dating" [online]. USA: The OkCupid Blog [cit. 1. 6. 2018]. Dostupné z: <https://theblog.okcupid.com/how-races-and-religions-match-in-online-dating-e74ccdc5f6f6>.
- Rudder, Christian. 2014. "Race and Attraction, 2009–2014" [online]. USA: The OkCupid Blog [cit. 1. 6. 2018]. Dostupné z: <https://theblog.okcupid.com/race-and-attraction-2009-2014-107dcbb4f060>.
- Sautter, Jessica M., Rebecca M. Tippett, S. Philip Morgan. 2010. "The Social Demography of Internet Dating in the United States*." *Social Science Quarterly* 91 (2): 554–75. <https://doi.org/10.1111/j.1540-6237.2010.00707.x>.
- Schoen, Robert, Robin Weinick. 1993. "Partner Choice in Marriages and Cohabitations." *Journal of Marriage and Family*, no. 2: 408–414. <https://doi.org/10.2307/352811>.
- Schwartz, Christine. 2013. "Trends and Variation in Assortative Mating: Causes and Consequences." *Annual Review of Sociology* 39 (1): 451–70. <https://doi.org/10.1146/annurev-soc-071312-145544>.
- Sherkat, Darren E. 2004. "Religious Inter-marriage in the United States: Trends, Patterns, and Predictors." *Social Science Research* 33 (4): 606–25. <https://doi.org/10.1016/j.ssresearch.2003.11.001>.
- Skopek, Jan, Florian Schulz, Hans-Peter Blossfeld. 2011. "Who Contacts Whom? Educational Homophily in Online Mate Selection." *European Sociological Review* 27 (2): 180–95. <https://doi.org/10.1093/esr/jcp068>.
- Smith, Aaron, Maeve Duggan. 2013. "Online Dating & Relationships" [online]. USA: Pew Research Center: Internet, Science & Tech [cit. 1. 6. 2018]. Dostupné z: <http://www.pewinternet.org/2013/10/21/online-dating-relationships/>.
- Smits, Jeroen, Hyunjoon Park. 2009. "Five Decades of Educational Assortative Mating in 10 East Asian Societies." *Social Forces* 88 (1): 227–55. <https://doi.org/10.1353/sof.0.0241>.
- Smits, Jeroen, Wout Ultee, Jan Lammers. 1998. "Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables." *American Sociological Review* 63 (2): 264–85. <https://doi.org/10.2307/2657327>.
- Sweeney, Megan M. 2002. "Two Decades of Family Change: The Shifting Economic Foundations of Marriage." *American Sociological Review* 67 (1): 132–47. <https://doi.org/10.2307/3088937>.
- Sweeney, Megan M., Maria Cancian. 2001. "The Changing Importance of Economic Prospects for Assortative Mating" [online.] USA: California Center for Population Research [cit. 1. 6. 2018]. Dostupné z: <http://escholarship.org/uc/item/0zb3420q>.

- Syslová, Jana. 2011. "Třetina Čechů hledá partnera po síti, vyjde to každému pátému" [online]. Praha: iDNES.cz [cit. 1. 6. 2018]. Dostupné z: http://ona.idnes.cz/tretina-cechu-hleda-partnera-po-siti-vyjde-to-kazdemu-patemu-prd-/vztahy-sex.aspx?c=A110615_173819_vztahy-sex_abr.
- The Economist. 2014. "Sex, Brains and Inequality" [online]. London: The Economist [cit. 1. 6. 2018]. Dostupné z: <https://www.economist.com/news/united-states/21595972-how-sexual-equality-increases-gap-between-rich-and-poor-households-sex-brains-and>.
- Ultee, Wout C., Ruud Luijkx. 1990. "Educational Heterogamy and Father-to-son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction?" *European Sociological Review* 6 (2): 125–49. <https://doi.org/10.1093/oxfordjournals.esr.a036554>.
- Valkenburg, Patti M., Jochen Peter. 2007. "Who Visits Online Dating Sites? Exploring Some Characteristics of Online Daters." *CyberPsychology & Behavior* 10 (6): 849–52. <https://doi.org/10.1089/cpb.2007.9941>.
- Verbrugge, Lois M. 1983. "A Research Note on Adult Friendship Contact: A Dyadic Perspective." *Social Forces* 62 (1): 78–83. <https://doi.org/10.2307/2578348>.
- Wang, Wendy. 2012. "Chapter 4: Public Attitudes on Inter marriage" [online]. USA: Pew Research Center's Social & Demographic Trends Project [cit 1. 6. 2018]. Dostupné z: <http://www.pewsocialtrends.org/2012/02/16/chapter-4-public-attitudes-on-intermarriage>.
- Yancey, George. 2007. "Homogamy over the Net: Using Internet Advertisements to Discover Who Interracially Dates." *Journal of Social and Personal Relationships* 24 (6): 913–30. <https://doi.org/10.1177/0265407507084190>.

Velká data jako alternativa výběrových šetření v kvantitativním sociálněvědním výzkumu¹

Johana Chylíková, Sociologický ústav AV ČR, v.v.i.

Big Data as an Alternative to Surveys in Quantitative Social Research

EXTENDED ABSTRACT :

The goal of this article is to inform social scientists, especially those of a quantitative orientation, about the basic characteristics of Big Data and to present the opportunities and limitations of using such data in social research. The paper informs about three basic types of Big Data as they are distinguished in contemporary methodological literature, namely administrative data, transaction data and social network data, and exemplifies how they can be utilized by quantitative social research. According to many, questionnaire-based sample survey as the dominant method of quantitative social research has found itself in a crisis, especially as response rates have decreased in most developed countries and public confidence in opinion polling has declined.

The author presents the characteristics and specifics of Big Data compared to survey research – a method whose primary distinguishing characteristic is the capacity to quantify individual behaviour, social action and attitudes at the level of populations. In this context, the article draws attention to the differences between Big Data and survey data typically presented in scholarly literature, namely that datasets are not representative of known populations, the values of observed variables are systematically biased, there is a limited number of variables in Big Data sets, there is uncertainty about the meaning of observed values, and social environment has direct influence on

KEYWORDS :

big data

survey

data quality

epistemology

methodology

DOI LINK :

<http://doi.org/10.13060/1214438X.2018.116.416>

¹ Tento článek byl podpořen grantem „CSDA Český sociálněvědní datový archiv“ (MŠMT, kód: LM2015060) a také grantem „SERISS“ (EU Horizon 2020, kód 654221).

the behaviours captured by Big Data. Attention is also paid to such characteristics of Big Data that pose an obstacle to smooth integration of this type of data in the social scientific mainstream. First, the collection, processing and analysis of Big Data is extremely demanding in terms of programming skills, something social scientists typically do not have. Second, the availability of Big Data is limited as they are normally possessed by private corporations, some of which (Facebook, Google) have undoubtedly come to form data oligopolies – and their management is mostly unwilling to share their data with traditional academics.

Based on the above-mentioned specifics, differences and limitations, it is argued that Big Data currently do not have the potential of becoming a full-fledged source of social science data and replacing sample surveys as the dominant research method. Finally, the article draws attention to the specifics of different types of Big Data as they are primarily generated for purposes other than social research and result from specific situations framed by existing social relations – and it is from this perspective that Big Data should be viewed by social researchers.

V současném sociálněvědním akademickém výzkumu jsou primární kvantitativní metodou výběrová šetření, a to nejen v České republice, ale i v dalších zemích Evropy či v USA. Šetření generují velké množství dat pro sociálněvědní analýzy a široké využití v akademickém i komerčním výzkumu nacházejí zhruba od konce 2. světové války. V Česku tato metoda spíše vzkvétá, než aby upadala, o čemž svědčí zapojení českých akademiků a akademiček do většiny velkých evropských a světových komparativních projektů výběrových šetření (European Social Survey, International Social Survey Programme, Survey of Health, Aging and Retirement in Europe, EU SILC, European Values Study) či nedávný vznik velkého panelu domácností v rámci sociologického projektu Proměny české společnosti. Pokud mají být výběrová šetření užitečnou a spolehlivou výzkumnou metodou, musejí výzkumníci průběžně vyhodnocovat jejich účinnost – šetření by měla přinášet validní informace o chování lidí, jejich postojích, názorech a hodnotách a neměla by být jen záminkou pro akademickou hru s čísly odtrženou od empirické reality.

V posledních desetiletích se ovšem začínají objevovat problémy, v jejichž důsledku výběrová šetření ztrácejí prestiž [Gobo, Giamietto 2011]. Odborníci se shodují na tom, že největším problémem výběrových šetření je klesající návratnost, která se snižuje i navzdory úsilí investovanému do opakovaných oslovení a penězům na odměny pro respondenty [Savage, Burrows 2007; Couper 2013, Massey, Tourangeau 2013]. Populace je přezkoumaná a lidé se už šetření zúčastňovat nechtějí; jsou často zaneprázdněni, nejsou k dostižení a častými žádostmi o rozhovor jsou unaveni. Za snižující se návratnost nemůže jen akademický výzkum, ale především soukromé firmy a agentury, které ve velké míře zjišťují veřejné mínění či spotřebitelské chování, případně obchodníci, kteří prostřednictvím telefonických hovorů prodávají své zboží. V současnosti se v USA návratnost z akademických šetření, která využívají pravděpodobnostní výběr, pohybuje kolem 40 až 60 procent, přičemž od devadesátých let klesá výrazně rychleji než dříve [Massey,

Tourangeau 2013], v případě komerčních šetření je situace ještě výrazně horší; CATI šetření prestižní americké agentury Pew Research Center mají v posledních pěti letech návratnost pouhých 9 procent [Pew Research Center 2017]. Situaci v evropské vědě může dobře ilustrovat návratnost z šetření European Social Survey (ESS), které využívá pravděpodobnostní výběr z populace. V Německu je návratnost jedna z nejnižších; od roku 2010, tedy v posledních třech zveřejněných vlnách výzkumu, se pohybuje jen kolem 30 procent, přitom v předchozích vlnách, tedy mezi lety 2002 až 2008, činila návratnost kolem 50 procent (podobně nízkou návratnost mají německé pravděpodobnostní výběry i v šetření International Social Survey Programme (ISSP), v roce 2014 to bylo 35 procent [Jutz, Scholz 2016]). Ve Velké Británii mají šetření ESS dlouhodobě návratnost přes 50 procent, ale v poslední zveřejněné vlně z roku 2014 činila návratnost jen 44 procent. Česká republika si v ESS v porovnání s jinými zeměmi vede docela dobře: v posledních čtyřech vlnách šetření měla návratnost 70, resp. 68 procent [ESS 2017].

V pořadí druhým velkým problémem výběrových šetření je snižující se prestiž metody, jak mezi výzkumníky, tak u veřejnosti [Gobo, Giamietto 2011], přičemž v druhém případě je již možné mluvit o částečné ztrátě důvěry v průzkumy veřejného mínění. Velkou ránu důvěře veřejnosti utrhly události roku 2016, a to hlasování o brexitu a volba amerického prezidenta. Oběma těmito lidovým hlasováními předcházelo velké množství průzkumů, které v situaci rovnoměrného rozložení podpory oběma soupeřícím stranám přinášely informace o těsném vítězství Hillary Clinton a odpůrců brexitu. V obou případech dopadly volby opačně. V obou případech však velký podíl na zklamání veřejnosti nesla i média, která o každém procentním bodu preferencí referovala jako o významném a reálném rozdílu a spoluvytvářela očekávání, která se nakonec nevyplnila. Nízká důvěra laické veřejnosti ve výběrová šetření může negativně ovlivnit budoucnost sociálněvědního výzkumu; nedůvěra se může projevit v dalším poklesu návratnosti nebo třeba v menší ochotě politiků financovat výzkum využívající výběrová šetření. V České republice zjišťovalo důvěru veřejnosti ve výběrová šetření Centrum pro výzkum veřejného mínění (CVVM), a to dosud pouze jednou, v březnu 2018. Výzkumům veřejného mínění podle CVVM rozhodně či spíše důvěřuje 49 procent respondentů, naopak 37 procent výběrového souboru spíše nebo rozhodně nedůvěřuje [Hanzlová 2018]. Z tohoto údaje se neodvažují soudit, zda je v současnosti zjištěná důvěra české populace ve výzkumy veřejného mínění vysoká, či nízká, neboť nemám časové srovnání a informaci o tom, zda důvěra v posledních letech klesá, či nikoliv.

S přihlédnutím k výše uvedeným problémům výběrových šetření se nabízí otázka, zda nehledat lepší zdroj výzkumných dat. Nebylo by lepší využít technologického pokroku a poohlédnout se po nějaké jiné metodě? V akademickém prostředí se v uplynulých deseti letech začaly zvedat kritické hlasy, které právě k takovému poohlédnutí vybízejí. Savage a Burrows [2007] vyzývají k tomu, aby byla výběrová šetření ve výzkumu upozaděna a do středu zájmu nastoupila takzvaná velká data a sběr dat prostřednictvím internetu. Naznačují, že vědci a vědkyně, které se drží kritizované metody výběrových šetření, by se brzo mohli ocitnout v pozici toho, který „chvíli stál a již stojí opodál“. Článek Savage a Burrowse vzbudil velkou pozornost a následovalo ho několik přímých reakcí, ať už pozitivních (např. [Webber 2009]), či kritických (např. [Crompton 2008;

Couper 2013]). Vyšly i další texty zabývající se velkými daty, ať už z epistemologického hlediska, které předpovídá změny, jež velká data přinesou do procesu vědeckého poznávání [Kitchin 2014; McFarland, Lewis, Goldberg 2016; Chandler 2015], či z metodologického hlediska, které velká data kriticky komentují jako omezený výzkumný nástroj sociálních věd [Grimmer 2018; Couper 2013; Boyd, Crawford 2012; Murphy et al 2011; Hargittai 2015].

Ačkoliv si velká data ještě neprorazila cestu do středního proudu sociologického výzkumu, lze očekávat, že se tak brzy může stát. Sociální vědci a vědkyně se opatrně seznamují s daty ze sociálních sítí a uvažují o možnostech, jak využívat i jiná velká data. Ovšem ještě před tím, než velká data vstoupí do sociologického výzkumu v plné síle, je nutné reflektovat jejich vlastnosti a identifikovat výhody a nevýhody, které využívání tohoto typu dat v sociálních vědách má. V další části tohoto článku přináším informace o tom, co jsou velká data, jaká je jejich specifická povaha, jak je lze využít pro výzkum a které jejich problematické charakteristiky ve výzkumu zohledňovat.

Co jsou velká data?

Veřejnost i odborníci používají název velká data (z angl. Big data) běžně, v sociálních vědách se částečně ujal i alternativní název *organická data*, jehož autorem je významný sociálněvědní metodolog Robert Groves, tento termín se ale používá výjimečně [Couper 2013; Habermann, Kennedy, Lahiri 2016]. Velká data mohou být definována jako soubory dat, které jsou tak velké, že je nedokáže zpracovávat běžný statistický software [Snijders, Matzat, Reips 2012]. O jak velké soubory se tedy jedná? Může to být např. objem dat o velikosti 2,5 petabytů (peta = 10^{15}), která obsahují informace o jednom milionu nákupních transakcí amerického řetězce WalMart, jež byly zaznamenány během jedné hodiny. Nebo to mohou být dvě a půl miliardy komentářů, 2,7 miliardy lajků a 300 milionů fotek, které v roce 2012 nahráli uživatelé Facebooku za jeden den [Kitchin 2014]. Pro velká data není ale typická jen kvantita, ale i kvalita. Jsou vysoce granulovaná a detailní a v některých případech zachycují realitu v každém okamžiku („moment-to-moment“ data), např. data z GPS, která obsahují údaje o poloze objektu v průběhu pohybu. Někteří dokonce tvrdí, že velká data jsou tak detailní, že v podstatě zachycují realitu jako takovou [Chandler 2015].

Pro popis toho, co velká data vlastně jsou, existuje několik definic. Nejpoužívanější z nich je definice třemi V, ve které každé V odkazuje k jedné typické vlastnosti velkých dat [Couper 2013; Chandler 2015; Japec et al. 2015; Daas, Roos, van de Ven, Neroni 2012]. Volume (objem) velkých dat překračuje kapacitu tradičních metod pro uchování a zpracování dat; velocity (rychlost) vyjadřuje, že velká data jsou získávána v reálném čase, tj. v okamžiku, kdy se událost skutečně děje, a variety (různorodost či variabilita) znamená, že velká data jsou syrová, neuspořádaná, nestrukturovaná a nepřipravená k analýze.

Další definice [Boyd, Crawford 2012] popisuje velká data jako „kulturní, technologický a vědecký fenomén“, jehož existence je dána kombinací několika činitelů. Tím prvním je

technologie, konkrétně maximalizace výpočetní síly a algoritmické přesnosti získávat, analyzovat, spojovat a porovnávat velké soubory dat. Tím druhým je analýza, která umožňuje identifikovat ve velkých datových souborech struktury, což lze využít pro ekonomické, sociální a právní účely. Tím třetím je mytologie, konkrétně rozšíření přesvědčení, že velké datové soubory nabízejí vyšší formu poznání, jež přináší dosud neobjevené informace a jež „má auru pravdy, objektivitu a přesnosti“.

Velká data lze rozdělit na tři typy podle toho, z jakého zdroje vznikají:

1. Administrativní data vznikají pro účely veřejné či státní správy a disponují jimi zpravidla státní či veřejné instituce. Administrativními daty jsou záznamy ministerstev, státních úřadů nebo třeba data z elektronické evidence tržeb [Couper 2013].
2. Transakční data jsou generována při elektronických transakcích a mohou jimi být údaje o platbách z debetních/kreditních karet, informace z věrnostních karet obchodů, záznamy telefonních operátorů o hovorech a přenosech dat, údaje z internetového vyhledávače či historie internetového prohlížeče. Transakční data jsou „automatický vedlejší produkt transakcí a aktivit“ [Couper 2013] a jejich primárním účelem je vyřízení transakce nebo uživatelského požadavku na službu. Transakční data bývají zpravidla využívána k marketingovým účelům a příjemci služeb firem si často vůbec nejsou vědomi toho, že data o jejich chování jsou ukládána a dále využívána.
3. Data ze sociálních médií či sítí, jako je Facebook, Instagram, Twitter, Tumblr a další, vznikají aktivitou uživatelů na síti, která představuje sdílení obsahu, jako jsou krátké texty, fotografie nebo internetové odkazy, a komunikaci s ostatními uživateli sítě [Murphy, Hill, Dean 2013].

Anglický název Big data odkazuje k fenoménu Velkého bratra (Big Brother) z Orwellova románu 1984 [Boyd, Crawford 2012]. Velká data podobně jako Velký bratr obsahují informace o všem, co člověk (uživatel v rámci jedné služby) dělá (placení kartou, prohlížení internetu, lajkování na Facebooku). Lidé, kteří disponují těmito daty, získávají velké množství leckdy intimních informací o velkém množství lidí. Podobnost s Velkým bratrem nutí k otázkám týkajícím se ochrany osobních údajů; tento problém je ale velice komplexní, nemá jednoznačné řešení a představuje jeden z největších praktických, právních a etických problémů práce s velkými daty [Habermann, Kennedy, Lahiri 2016; Couper 2013; Boyd, Crawford 2012; Savage, Burrows 2009; Kalyvas, Overly 2015; Murphy et al 2014].

Případné využití velkých dat v sociálněvědním výzkumu je specifické tím, že výzkumníci přímo nekontrolují jejich sběr [Japac et al. 2015]. Velká data jsou „rutinně získávána jako vedlejší produkt institucionálních transakcí“ [Savage, Burrows 2007] a jejich sběr není nijak designován. Tinati a kolegové [Tinati, Halford, Carr, Pope 2014] tuto specifickou vlastnost popisují jako „sbírání dat v divočině“ (in the wild), někdy se o velkých datech mluví jako o „digitálních stopách“ (digital footprints), které za sebou zanechávají jednotlivci v každodenním offline i online životě [Chandler 2015; Golder, Macy 2014; McFarland, Lewis, Goldberg 2016].

Využití velkých dat ve výzkumu je různorodé a lze jej ilustrovat na reálných příkladech. Např. administrativní data mohou v metodologii posloužit ke zjištění validity dat

z výběrových šetření. Ansolabehere a Hersh [2012] srovnávali výpovědi respondentů v internetovém dotazníkovém šetření o účasti ve volbách s údaji z voličských registrů. Informace z registrů spojili s neanonymními daty ze šetření a porovnáním zjistili, že respondenti svou účast ve volbách nadhodnocovali. Data ze sociálních sítí lze využít např. ke studiu změn nálad uživatelů. Bollen a kolegové [2011] analyzovali vztah mezi náladou, kterou lidé vyjadřovali na sociální síti Twitter, a vývojem na burze. S pomocí tzv. mood tracking nástrojů měřili průměrnou náladu na Twitteru a zjistili, že zahrnutí této proměnné do modelu vývoje burzovního indexu Dow Jones dokáže odhady vývoje zpřesnit. Mediálně známým se stal experiment, který z vlastní iniciativy provedl Facebook v roce 2012 [Griffin 2014] a jehož cílem bylo zjistit, zda má tato sociální síť schopnost ovlivňovat náladu uživatelů. Do experimentu vstoupilo celkem 700 tisíc uživatelů, kteří byli rozděleni do dvou skupin; první skupině Facebook zobrazoval pozitivní mediální obsah, druhé negativní. Výsledkem bylo zjištění, že Facebook dokáže prostřednictvím zobrazovaného obsahu uživatele činit buď smutnějšími, nebo šťastnějšími. Rovněž Bond a kolegové [2012] manipulovali zobrazování obsahu několika skupinám uživatelů sociální sítě, aby zkoumali vliv zobrazovaného obsahu na volební chování. Podobné experimenty mohou provádět pouze výzkumníci a výzkumnice, kteří jsou přímo navázáni na provozovatele sociálních médií. Ostatním zbývá pouze sekundární využití dostupných dat, jako to např. udělali Kosinski a kolegové [Kosinski, Stillwell, Graepel 2013], kteří ukázali, že z facebookových lajků lze predikovat charakteristiky jednotlivých uživatelů Facebooku, jako je sexuální orientace, etnicita či politická orientace. Další výzkumníci využili transakční data z vyhledávače Google, aby se z nich pokusili odhadnout, zda se blíží chřipková epidemie [Lazer et al. 2014].

Velká data, především transakční data, kterými disponují velké firmy, nalézají obrovské využití v marketingu. Využívání transakčních dat je poměrně levné, protože jsou vedlejším produktem prodeje služeb či produktů, a tak nevznikají další náklady na sběr dat o zákaznících. Např. americký maloobchodní řetězec analyzoval data o nákupech za období dlouhé dvanáct let a našel v nich vztahy, které odhalily vzorce v nákupním chování zákazníků; některé výrobky zákazníci nakupovali společně s konkrétními jinými výrobky. Řetězec na základě vzorců upravil umístění zboží v regálech, aby zboží, které zákazníci kupují společně, bylo umístěno blízko sebe, a dosáhl tím 16procentního nárůstu tržeb [Kitchin 2014]. V podstatě stejný princip dlouhodobě využívá internetový obchod Amazon. Tento gigant mezi e-shopy uchovává informace o tom, co si jednotliví zákazníci v minulosti koupili, a při každé další návštěvě webu jim navrhuje, aby si koupili další zboží. Nabízí jim přesně to zboží, které si v minulosti koupili jiní lidé, kteří si mimo jiné už koupili to samé zboží, co zákazník, kterému je zobrazována nabídka. Marketingová strategie vyjadřovaná nedokončenou větou „Lidé, kteří si koupili stejnou věc jako vy, si rovněž koupili...“ se dlouhodobě vyplácí a převzala ji další webové obchody a stránky (např. IMDB, International Movie Database). Velká data k zajištění obchodního úspěchu použila i internetová televize Netflix, která produkuje vlastní televizní seriály. Při plánování natáčení seriálu House of Cards Netflix nejprve analyzoval ve své databázi preference svých předplatitelů, aby zjistil, které herce a režiséry mají nejvíc

v oblíbě. Na základě tohoto průzkumu obsadil do hlavní role Kevina Spaceyho a režii svěřil Davidovi Fincherovi, neboť právě tyto dva se ukázali jako neoblíbenější umělci předplatitelů Netflixu [Barnes 2013; Carr 2013].

Velká data jako data svého druhu

V roce 2008 vzbudil velkou pozornost článek internetového magazínu Wired, ve kterém jeho autor zvěstoval příchod nového vědeckého věku, kdy velká data dokážou přinést naprosto věrný, detailní a objektivní obraz reality [Anderson 2008]. Podle Andersona může věda konečně objektivně zachytit skutečnost v celé její šíři a hloubce, kterou výzkumníkům odhalí velká data. Na Andersonův komentář přímo reagovalo několik akademiček a akademiků [Kitchin 2014; McFarland, Lewis, Goldberg 2016; Boyd, Crawford 2012; Chandler 2015] a vytýkali mu především jeho představu o povaze velkých dat. Velká data podle kritiků nejsou obrazem reality, nemluví „sama za sebe“ a nerodí objektivní, čisté vědění. I velká data vznikají v určitých situacích a za určitých podmínek, které jsou plně dány lidskými aktéry. Prostředí, ve kterém velká data vznikají, není sociální a kulturní vakuum a není to ani nově vytvořené „hřiště“, jehož zákonitosti a pravidla nebyla ovlivněna člověkem. Reagující akademici proto volají po tom, aby velká data byla reflektována jako fenomén svého druhu, nikoliv jako objektivní obraz reality [Kitchin 2014; McFarland, Lewis, Goldberg 2016; Boyd, Crawford 2012].

Ve svém článku dále Anderson představuje radikální tvrzení, že díky velkým datům věda již nebude potřebovat teorii. V situaci, kdy „data mluví sama za sebe“, jsou podle Andersona vědecké aproximace a modely zbytečné, protože poznání se bude samo vynořovat z velkých objemů dat. Deduktivnímu způsobu poznávání tak podle něj odzvonilo. Chandler [2015] ho v tomto kontextu doplňuje informací, že analytici, kteří pracují s velkými daty, nepřicházejí k datům s přesnými představami o tom, co v datech hledat, ale prvotní informace o zkoumaném fenoménu získávají nejprve z dat. Anderson a další, kteří vidí opuštění deduktivního způsobu poznávání jako velkou šanci pro rozvoj vědy, doufají, že se bude zvyšovat schopnost technologie zachycovat jevy, takže „svět začne mluvit sám za sebe bez zprostředkování omylným lidským interpretem“ [Chandler 2015].

V kvantitativní sociologii je dedukce základním principem poznávání. Při práci s daty z výběrových šetření si výzkumníci přirozeně kladou otázky, na které hledají odpovědi deduktivním přístupem, a s tímto účelem přistupují k výzkumu od začátku až do konce. Induktivní přístupy neumožňují testovat hypotézy a vytvářet modely vztahů, což pro současnou praxi znamená obrovské omezení. Někteří autoři [Boyd, Crawford 2012; Barnes 2013] kritizují, že bez deduktivního přístupu nebudeme schopni studovat příčiny a následky sociálních jevů. Ovšem podle Savage a Burrowse [2007] to není problém; podle nich jsou sociální vědy ve studiu kauzality dlouhodobě neúspěšné a měly by se této ambice úplně vzdát. Naopak by se měly přeorientovat jen na popis sociálních jevů, k čemuž jsou podle nich velká data vhodnější než data z výběrových šetření.

Sociologie využívající velká data samozřejmě může používat induktivní analytický přístup, který opustil klasický princip. Z velkých dat lze např. získávat popisy společnosti či lidského chování. Příkladem sociálněvědní studie využívající indukci může být studie Webbera [2009], který z obrovského množství údajů o křestních jménech a příjmení odhadl informace o etnicitě a příslušnosti ke společenské třídě velkého množství jednotlivců žijících na daném území. Velká data tak využil k získání informace, kterou by jinak mohlo přinést jen velmi rozsáhlé výběrové šetření. Avšak navzdory možnému použití indukce není pravděpodobné, že se sociální vědy deduktivního přístupu v nejbližší budoucnosti vzdají. Dedukce bude nadále využívána pro analýzu dat z výběrových šetření a bude se snažit přežít i v oblasti analýzy velkých dat. Již v současné době využívají analytici a analytičky klasický statistickoanalytický přístup k analýze dat ze sociálních sítí, jako je např. obsahová analýza a popisná statistika k testování hypotéz o výskytu různých slov, klíčových slov a hashtagů, frekvencích sdílení různých obsahů či podobnosti jednotek v různých skupinách. Příklady takových prací nacházíme i v České republice (viz např. [Karašćáková 2013; Hrdina 2016; Vochocová, Mazák, Štětka 2016]). Deduktivní přístup se také jistě udrží v analýze dat, která vznikají propojením velkých dat s klasickým dotazníkovým šetřením, např. spojením dat z facebookového účtu s individuálními daty z neanonymního dotazníkového šetření. Tento druh analýzy může přinášet velmi zajímavé výsledky i navzdory omezením neanonymního dotazování [Kosinski, Stillwell Graepel 2013; Ansolabehere, Hersh 2012].

Reflexe povahy velkých dat pro využití ve výzkumné praxi

Ať už budeme velká data používat pro deduktivní, nebo induktivní poznávání, musíme je podrobit důkladnému rozboru z hlediska jejich povahy. Je potřeba se zabývat širšími souvislostmi používání velkých dat ve výzkumu, a to zejména ontologickou podstatou dat ze sociálních sítí a transakčních dat. V jakých podmínkách vznikají? Jak samotné sociální sítě a prostředky transakčních dat mění chování lidí? Jaké sociální skupiny přednostně vytvářejí velká data? V neposlední řadě je nutné reflektovat, jak ovlivňuje využití velkých dat samotný proces výzkumu; velká data vznikají bez dohledu výzkumníků a jejich samotný charakter svádí výzkumníky k určité perspektivě. Povaha těchto dat formuje design výzkumu a ovlivňuje výzkumníky v tom, jaké si kladou otázky a výzkumné cíle [Boyd, Crawford 2012]. Jako výzkumníci a výzkumnice využívající kvantitativní paradigmaty musíme věnovat pozornost všem charakteristikám, ve kterých se velká data liší od dat z výběrových šetření, a zohlednit jejich netradiční vlastnosti ve statistické analýze a při interpretaci výsledků. Musíme se věnovat odchylkám od pravých hodnot zjišťovaných proměnných a dalším nepřesnostem, které snižují validitu závěrů výzkumu. Systematická analýza odchylek ve velkých datech však nebude jednoduchá, neboť tato data jsou výrazně komplexnější a komplikovanější než data z šetření. V následujícím textu se na využití velkých dat dívám optikou metodologie výběrových šetření a implicitně porovnávám jejich charakter s vlastnostmi, které mají data z výběrových šetření, jež nacházejí využití v klasické

statistické analýze. Postupně se zastavuji u několika zásadních problémů, které využívání velkých dat ve výzkumu doprovází. Jsou jimi náročnost analýzy velkých dat, jejich dostupnost, reprezentativita, vychýlení výběru jednotek, vliv sociálního okolí na hodnoty proměnných a další problematické jevy.

Náročnost analýzy velkých dat

V současnosti představuje překážkou uvedení velkých dat do běžné výzkumné praxe skutečnost, že pro analytickou práci s nimi jsou potřeba programátorské schopnosti [Manovich 2011; Golder, Macy 2014; McFarland, Lewis, Goldberg 2016; Kitchin 2014]. Velká data je nutné umět sbírat, skladovat a analyzovat, spojovat je, případně je validizovat údaji z externích zdrojů. Dále je potřeba je uspořádat a připravit pro analýzu, neboť data sama o sobě nejsou strukturována tak, aby mohla být analyzována bez úpravy [Golder, Macy 2014]. Všechny uvedené operace jsou náročné na provedení a bez programátorského tréninku je nelze zvládnout. Sociální vědci a vědkyně s klasickým vzděláním v analýze dat v současnosti nemají potřebné znalosti k tomu, aby mohli velká data analyzovat, a proto se analýzou velkých dat zabývají především programátoři. Tomu napovídá i to, že většina dosud publikovaných studií využívajících velká data byla zpracována programátory a datovými vědci [Golder, Macy 2014; McFarland, Lewis, Goldberg 2016].

Angažmá IT expertů v hájemství sociálních věd je podle některých autorů problematické, neboť většina z nich nemá dostatečné sociálněvědní vzdělání. Jejich výzkum je proto často „redukcionistický, funkcionalistický a ignoruje vlivy kultury, politiky, veřejné politiky a kapitálu“ [Kitchin 2014]. Další autorky zase upozorňují na významnou genderovou disproporci v komunitě počítačových vědců, mezi kterými jsou nedostatečně zastoupeny ženy [Boyd, Crawford 2012]. Je známo, že gender výzkumníka nepřímo ovlivňuje tematické zaměření plánovaných studií i konkrétní výzkumné otázky a může mít vliv i na interpretace zjištění z výzkumu [Harding 2010; de Madariaga 2012.]. Silná převaha mužů mezi počítačovými experty, kteří analyzují velká data pro účely sociálněvědních analýz, tak může deformovat výzkumnou agendu směrem k tématům a problémům, které jako prioritní vidí především mužská část populace.

Na absenci sociálních vědců a vědkyň v oblasti velkých dat ukazuje i fakt, že v současné době nejsou k dispozici učebnice, příručky či manuály představující analytické metody práce s velkými daty. Významní výzkumníci a výzkumnice v oboru kvantitativní sociologie v uplynulých desítkách let publikovali množství nejrůznějších učebnic a instruktážních textů, které vysvětlují zásady práce s daty z výběrových šetření a principy mnoha typů statistické analýzy (např. [Babbie 2015; Bollen 2009; Saris, Gallhofer 2014; Tabachnick, Fidell, Osterlind 2001; Hox, Roberts 2011]), k analýze velkých dat však učebnice nejsou. Chybí také obecnější pojednání o tom, jaký typ úloh je možné s velkými daty realizovat či jaký druh výzkumných otázek lze s jejich využitím řešit. Dosud vycházejí především texty, které shrnují velká data jako fenomén, který může mít na sociálněvědní výzkum v budoucnosti velký vliv (např. [Boyd, Crawford 2012; Couper 2013; Murphy et al 2014; McFarland, Lewis, Goldberg 2016; Savage, Burrows 2007, Burrows, Savage 2014]).

Dostupnost velkých dat pro sociálněvědní výzkum

Omezená dostupnost velkých dat je považována za největší překážku, která stojí velkým datům v cestě do sociálněvědního mainstreamu [Herman, Kennedy, Lahiri 2016; Couper 2013; Savage, Burrows 2007; Boyd, Crawford 2012; Manovich 2011; McFarland, Lewis, Goldberg 2016; Driscoll 2014]. Data ze sociálních sítí jsou omezena méně než transakční data, což se odráží i v jejich častějším používání v sociálněvědním výzkumu [Murphy et al. 2014]. Facebook či Twitter umožňují veřejnosti přístup do datového úložiště, ze kterého lze při znalosti programátorských postupů získat aktuálně dostupná data. I zde se však postupně utahují kohouty: informační giganti jako Facebook se v posledních letech stali datovými oligopoly, disponujícími obrovským množstvím informací, a uvědomují si výsadní postavení, které díky informacím získávají. Aby si toto postavení udrželi, poskytují veřejnosti, tedy i výzkumníkům, stále menší objemy dat [Boyd, Crawford 2012; Driscoll 2014; McFarland, Lewis, Goldberg 2016]. Vědcům je údajně k dispozici pouze 1 až 10 procent veřejných tweetů [Boyd, Crawford 2012; Driscoll 2014], množství a typ dat, které je možné získat z Facebooku, je rovněž regulováno. V oblasti analýzy dat ze sociálních sítí tak získávají rozhodující konkurenční výhodu vědci, kteří jsou u firem, jako je Facebook, *přímo zaměstnání*. Nejenže jim jsou dostupná všechna existující data, ale získávají i možnost realizovat na sociální síti vlastní experimenty. Jejich vědecká práce tedy může nad prací akademiků v mnohém vyniknout [Manovich 2011; Savage, Burrows 2007].

Administrativní a transakční data jsou sociálním vědkyním a vědcům dostupná velice málo nebo vůbec; administrativními daty disponuje stát a jeho úřady a podléhají pravidlům ochrany soukromí, jejichž přísnost se liší v jednotlivých zemích. Transakční data mají plně ve své moci soukromé firmy, jako jsou banky, telefonní operátoři či provozovatelé internetových vyhledávačů a prohlížečů. Mnozí upozorňují na to, že se mocenská pozice soukromých firem oproti minulosti posiluje právě díky tomu, že velké firmy díky digitalizaci disponují obrovským množstvím informací o jednotlivcích i společnosti jako celku. S tím se na druhou stranu oslabuje pozice sociálních věd, které ztrácejí monopol na využívání zdrojů informací o společnosti, a v důsledku i monopol na popis a vysvětlování společenských jevů. Savage a Burrows [2007] pro tuto společenskou situaci používají termín převzatý od Thriфта [2005] *knowing capitalism*, který označuje kapitalistickou společnost, kde soukromé firmy mají prostředky k získávání vědění a disponují mocí toto vědění rekonstruovat a přetvářet ve svůj prospěch. V tomto novém společenském uspořádání ztrácejí svoji moc instituce, jako je stát a na něj navázaná akademická sféra.

Neschopnost získávat velká data pociťují sociální vědci jako obtíž, která může mít zásadně negativní vliv na budoucnost vědy a výzkumu. Metodolog Robert Groves vyzývá, aby věda v součinnosti se státem začala tento problém neprodleně řešit. Stát by měl akademické sféře nabídnout pomoc při vyjednávání s velkými firmami disponujícími velkými daty a iniciovat vznik institucí na pomezí akademické, resp. státní a soukromé sféry, které budou spolupracovat s vlastníky velkých dat a přesvědčovat je o nutnosti používání těchto dat ve výzkumu [Herman, Kennedy, Lahiri 2016]. Čím konkrétně by mohli akademici velké firmy přesvědčit, aby jim data poskytovaly, však Groves neuvádí.

Reprezentativita za populaci

Výzkumníci, kteří běžně pracují s daty z výběrových šetření, kladou velký důraz na reprezentativitu, případně kvazireprezentativitu dat za populaci národního státu či za jinou, specifickou populaci (studenti, ženy na rodičovské dovolené atp.). Velká data, která nejsou reprezentativní za běžně zkoumané populace, jsou jim podezřelá [Savage, Burrows 2007]. Často vyjadřují námitku, že „velikost není všechno“ a že obrovské objemy velkých dat nejsou schopny vyvážit to, že zahrnutí jedinci proporcčně neodpovídají obecné (nebo jiné známé) populaci [Boyd, Crawford 2012; Couper 2013; Hargittai 2015; McFarland, Lewis, Goldberg 2016; Golder, Macy 2014].

Problém (ne)reprezentativity se týká dat ze sociálních médií i transakčních dat. Transakční data firem obsahují jen údaje o lidech, kterým daná firma dodává služby; jsou to tedy např. uživatelé platebních karet jedné banky či klienti konkrétního telefonního operátora. V marketingovém výzkumu pro účely firem není nereprezentativita problém, pokud jsou ale taková data využívána pro účely sociálněvědního výzkumu, musí být analýza doprovázena důrazným upozorněním, že výsledky nejsou zobecnitelné na populaci národního státu. Obdobně populace lidí využívajících sociální sítě není v současné době stejná jako obecná populace. Z analýz existujících profilů na Twitteru a na Facebooku vyplývá, že např. v USA jsou uživatelé sociálních sítí mladší, vzdělanější a bohatší [Golder, Macy 2014] a jednotlivci na síti se od obecné populace liší i v pohlaví a rase [Hargittai 2015]. Ani v Česku nezahrnují sociální sítě celou populaci; o demografickém složení profilů na Facebooku informuje např. Dočekal [2015].

Z hlediska sociologické metodologie lze porušení zásady reprezentativity u velkých dat považovat za problém. Existují však specifické případy, kdy je možné podmínku reprezentativity uvolnit a získat druh informace, kterou lze jinak zjistit jen obtížně. Ve výběrových šetřeních máme problém získávat data o běžných denních činnostech lidí; když používáme deníky, respondenti do nich často zapisují nepřesně, leccos opomenou, něco si třeba naopak vymyslí. Navíc tím respondenty velice zatěžujeme. Velká data nám mohou pomoci právě při získávání deníkového typu informací; jejich přesnost by mohla vyvážit nereprezentativitu. Data z platebních a věrnostních karet umožňují zjistit, v kterou denní dobu dělají lidé běžné nákupy nebo v jakém období nakupují speciální zboží, jako je třeba nábytek; data z vyhledávačů, prohlížečů nebo navigačních aplikací dodávají např. informace o tom, v kterou dobu lidé zhruba začínají pracovat či jakým způsobem a za jakým účelem se pohybují ve veřejném prostoru. Data z různých mobilních aplikací, které lidé využívají ke sportu, dietě nebo kontrole spánku a mnoha dalším činnostem, jsou rovněž zdrojem velice detailních informací o běžných činnostech a návycích.

Vychýlené soubory jednotek

Soubory dat ze sociálních sítí, ale i transakčních dat obsahují jednotky různého typu; mohou to být lidé, nákupy, platby kartou, tweety, fotografie, lajky, statusy a další jednotky informací

[Murphy et al. 2014]. Tyto soubory jsou vždy do jisté míry odchýlené od populace, na kterou by z dat šlo usuzovat (uživatelé Facebooku, zákazníci obchodu, klienti banky atd.), neboť ne všechny jednotky populace jsou stejně aktivní. Pro transakční data a data ze sociálních sítí platí, že obsahují jen informace pocházející od jednotlivců, kteří učinili nějakou akci. Do výběru se dostávají jen lidé, kteří mají věrnostní kartu obchodu a používají ji, lidé, kteří použili hashtag, lidé, kteří zaplatili platební kartou, lidé, kteří sdílejí na Facebooku fotografie, případně hashtagy, které byly použity v daném období, fotografie, které byly zveřejněny v daném období na sociální síti, atp. Do výběru se naopak nedostanou ti, kteří sice nakupují, ale bez využívání věrnostních a platebních karet, nebo např. ti, kteří tweetují, ale nepoužívají hashtagy. Couper [2013] tuto vlastnost velkých dat vyjadřuje výstižnou větou „Big data tends to focus more on the ‘haves’ and less on the ‘have-nots’“. V důsledku tohoto fenoménu vypadává z analyzovaných souborů velkých dat podstatná část lidí, kterých se případný zkoumaný problém týká, a výzkumníci tak získávají hodnoty, které jsou odchýlené od pravé hodnoty platné pro zkoumanou populaci, tedy např. uživatele Facebooku, klientů spořitelny atp.

V případě sociálních sítí existuje více možností uživatelské aktivity, než je jen existence profilu na síti. Jak velká je proporce lidí, kteří sociální médium aktivně používají? Sociální sítě jsou plné „posluchačů“, tedy lidí, kteří především konzumují obsah ostatních uživatelů, ale sami vytvářejí velmi málo obsahu, případně se neprojevují vůbec [Boyd, Crawford 2012]. O těchto posluchačích tak výzkumníci nezískávají žádná data. Skutečností je i to, že lidé mají různou pravděpodobnost stát se aktivním uživatelem sítě [Murphy et al. 2014]; např. manuálně pracující mohou být v obsahu sociálních médií reprezentováni méně než lidé, kteří tráví naprostou většinu dne v kanceláři či doma u počítače. A nejde jen o uživatele, kteří se do souboru nedostanou, protože nejsou aktivní, ale i o uživatele, kteří sice aktivní jsou, ale de facto neexistují. Část profilů na sociálních sítích tvoří duplicitní účty a není ani známé množství falešných účtů, které jednotlivci zakládají z různých důvodů. Samostatný problém tvoří účty lidí, kteří jsou placeni za šíření propagandy. Tito lidé na sítích diskutují jako normální uživatelé, ovšem s účelem šířit propagandu [King, Pan, Roberts 2016; Murphy et al. 2014]. Příbuzným problémem je aktivita robotů, kteří vytvářejí obsah, jenž je na první pohled nerozeznatelný od obsahu, který vytváří skutečný uživatel [Boyd, Crawford 2012; McFarland, Lewis, Goldberg 2016; Bessi, Ferrara 2016; Murphy et al. 2014]. Roboty mohou být např. falešné twitterové účty, využívané pro účely politického boje s cílem ovlivnit diskurz na sociální síti. Existují sice metody, jak v datech identifikovat takové robotické účty, sociálněvědní výzkumníci je však běžně neovládají [Bessi, Ferrara 2016].

K problematice vychýlených souborů jednotek ze sociálních sítí se váže i fakt, že lidé používají konkrétní sociální síť ke sdílení konkrétního typu obsahu [Hargittai 2015]. Např. Twitter lidé primárně používají k verbálně psané komunikaci nebo k odkazování na jiný internetový obsah, naopak na Instagram dávají jenom fotky. Jednotliví uživatelé často neduplikují obsah na svých twitterových, facebookových a instagramových účtech, takže obsahy těchto médií se nepřekrývají, ale doplňují. Pokud analyzujeme obsah na sociálních sítích a nemáme spojená data ze všech profilů uživatelů na různých sociálních sítích, nemůžeme mít celkový přehled o obsahu, který

uživatelé sdílejí přes všechny své profily na sociálních sítích. Zkoumá-li např. analytik vztah mezi lajkováním na Facebooku a fotografiemi, které uživatel nahrává na svůj facebookový profil, mohou mu unikat podstatné informace, neboť významná část uživatelových fotek se možná nachází na Instagramu. V analýzách je proto nutné explicitně uvádět, že usuzování z dat z jednoho sociálního média se týká pouze a výhradně jen účtů na konkrétním zkoumaném médiu.

Množství proměnných ve velkých datech

Jeden z rozdílů mezi velkými daty a daty z výběrového šetření není jen v množství případů v jednom souboru, ale v i množství proměnných. „Velká data obsahují hodně případů a málo proměnných, zatímco data z šetření obsahují hodně proměnných, ale málo případů“, píše ve svém článku zjednodušeně Couper [2013]. Nedostatkem je podle něj malé množství demografických proměnných či jejich úplná absence. Uvádí, že internetové prohlížeče, které generují transakční data, často nemají údaje o demografii uživatele, a tak musí základní demografické údaje odhadovat z vlastních dat o vyhledávání. Tyto odhady ovšem obsahují chyby; Google odhaduje pro účely svých výzkumů gender uživatelů, je jej však schopen identifikovat správně pouze v 75 procentech případů [Couper 2013]. Ani na sociálních sítích není identifikace genderu a dalších demografických proměnných zcela přesná; uživatelé o sobě často demografické informace neudávají, případně poskytují nepravdivé údaje. Savage a Burrows [2007] navrhuje, aby sociální vědci tento problém řešili inspirací v marketingovém výzkumu, kde se prosadilo používání proměnné *místo bydliště*. Tato proměnná je údajně velmi silným prediktorem, schopným nahradit základní demografické proměnné, jako je třída, etnicita či vzdělání [Webber 2009]. Místo bydliště by prý mohlo suplovat množství demografických proměnných, se kterými sociální věda pracuje dnes; proti tomuto nápadu se však rychle zvedly kritické hlasy [Crompton 2008]. Zajímavou ukázkou odvozování demografických proměnných přinesl Webber [2009], který ve velkých datech z údajů o křestním jménu a příjmení dokázal s vysokou pravděpodobností odhadnout etnicitu a společenskou třídu jedinice. K problematice množství a charakteru proměnných ve velkých datech je třeba dodat, že vývoj v oblasti digitálních technologií jde kupředu velmi rychle a jednotlivci a firmy generující a zpracovávající velká data dosahují podstatného pokroku v tom, jak různorodé informace jsou schopni získávat. Zlepšují se v odhadování demografických proměnných a zvyšují množství proměnných, které dokážou v datech identifikovat.

Nejistota ohledně motivace k chování

Transakční data poskytují výhradně údaje o chování (např. platba kartou) a provedených akcích (např. vyhledávané heslo v Googlu), neřikají nám ale nic o motivacích k tomuto chování či o postojích a hodnotách, které člověka k chování či akci vedou. Data ze sociálních sítí dodávají o něco méně informací o chování, mohou ale lépe sloužit jako zdroj informací o postojích a hodnotách [Couper 2013]. I dat ze sociálních sítí se však často týkají pochybnosti o tom, co pozorované chování vlastně znamená. Na Twitteru např. nevíme, zda uživatel/ka retweetováním

vyjadřuje danému obsahu podporu, či zda se jedná o subversivní tweet, který může být sdílen z jiných důvodů, než je vyjádření souhlasu nebo podpory; typickým příkladem takového subverzivního tweetu je retweetování obsahu twitterového účtu kontroverzního politika či jiné veřejně známé osoby, které má upozornit na pozoruhodný či bizarní obsah. To samé platí o lajkování na Facebooku, ani u něj neznáme motivaci uživatele k lajkování.

Příkladem neznámé motivace k akci v transakčních datech je vyhledávání v internetovém vyhledávači. O tom, jak velkou roli hrají motivace k vyhledávání, se přesvědčili výzkumníci, kteří studovali rozšířenost užívání drogy zvané *Salvia Divinorum* [Murphy, Dean, Hill, Richards 2011]. Domnívali se, že ke zjištění rozšířenosti drogy je možné použít data z vyhledavače Google a z Twitteru, indikátorem rozšířenosti měla být podle jejich předpokladu četnost vyhledávání a tweetování o této látce. Z dat Googlu a Twitteru zjistili frekvence výskytu klíčového slova *Salvia Divinorum* v určitém období a porovnali je s daty projektu National Survey of Drug Use and Health, ve kterém respondenti vypovídali, zda tuto konkrétní látku užíli, či nikoliv. Porovnání obou datových zdrojů ukázalo, že ve velkých datech byla četnost výskytu mnohem větší než v datech z šetření, což donutilo výzkumníky pátrat po příčinách. Zjistili, že bylinu ve zkoumaném období kouřila zpěvačka Miley Cyrus na oslavě svých narozenin a video zachycující tuto událost se dostalo na YouTube. Příčinou velkého výskytu drogy ve vyhledávání Googlu a na Twitteru bylo, že se lidé touto událostí bavili, nikoliv to, že by si drogu chtěli opatřit.

Vliv sociálního okolí

Stejně jako v offline životě jsou lidé, kteří za sebou nechávají digitální stopy, pod vlivem sociálního okolí. Na sociálních sítích se tento vliv může projevat jako cenzura, kdy je administrátorem či majitelem účtu vymazáván obsah a blokování uživatelé, nebo jako autocenzura, kdy sám uživatel omezuje svůj projev z důvodu obav z reakce sociálního okolí, případně za účelem budování vlastní image. Následkem cenzury i autocenzury je to, že některé jednotky pozorování, jako jsou komentáře, lajky, tweety či jiný typ obsahu, jsou buď vymazány cenzorem (cenzura), nebo vůbec nevzniknou (autocenzura).

Cenzura na sociálních sítích někdy probíhá ze strany provozovatele sítě z důvodu toho, že cenzurovaný obsah odporuje zásadám komunity (např. pornografie), jindy ji uplatňují majitelé profilů, jako jsou soukromé firmy, značky, celebrity nebo politici. Obsah na sociální síti je tak často filtrován z hlediska zájmů firem, politiků a dalších významných lidí, kteří nechtějí, aby jim uživatelé nevhodnými komentáři kazili budovanou image.

Autocenzura na sociálních sítích je důsledkem existence společenských norem a přímého kontaktu s ostatními uživateli sítě. Uvnitř sociálních sítí existuje velké množství menších sítí uživatelů, pro které jsou typické sdílené postoje a hodnoty. V těchto menších sítích panují mocenská pravidla, která jsou zčásti dána reálnými mocenskými silami ve společnosti, vtělenými v charakteristiky, jako je pohlaví, vzdělání či etnicita, zčásti nově vytvářená v síti aktivitou uživatelů. Uživatelé těmto mocenským pravidlům přizpůsobují svoje chování, postupem času se učí, co si mohou a nemohou dovolit, a sdílejí obsah, který je více či méně v souladu s pravidly

subpopulace. Jejich komentáře, lajky, sdílené fotografie a statusy tedy vznikají pod přímým vlivem ostatních uživatelů sítě. Cílem přizpůsobení se pravidlům v síti je minimalizace sankcí za chování, postoje či názory, se kterými většina v subpopulaci nesouhlasí. Sankcemi na síti zpravidla bývají negativní, či dokonce agresivní komentáře ostatních uživatelů či zablokování přístupu na profil, případně zablokování účtu. Autocenzura v důsledku může vést i k tomu, že některé skupiny lidí se na některých místech sítě vůbec nevyjadřují, např. proto, že patří k diskriminované skupině, která je pro diskutující většinou snadným cílem útoku. Představovat si sociální síť jako svobodné virtuální fórum, které přináší skutečné a objektivní informace o tom, co si lidé myslí, je liché.

Vliv sociálního okolí může aktivitu uživatelů nejen potlačovat, ale také ji podporovat či moderovat. Takový tlak může vytvořit např. vysoká popularita příspěvku na sociální síti vyjádřená množstvím lajků. Sama popularita příspěvku může ovlivnit postoj uživatele k zobrazovanému obsahu; nejenže se např. na Facebooku vysoce populární obsah zobrazuje více lidem, čímž může získat více lajků, ale i samotný fakt, že je příspěvek populární, může v uživatelích vzbudit sympatii k zobrazovanému obsahu, kterou uživatel vyjádří lajkem. Aktivita ostatních uživatelů v síti hraje velkou roli v diskuzích, kde uživatelé přímo reagují jeden na druhého a většina komentářů je přímo závislá na ostatních komentářích.

Právě přímý vliv sociálního okolí je činitelem, který je zodpovědný za jeden z hlavních rozdílů mezi daty z výběrových šetření a daty ze sociálních sítí. Data z výběrových šetření obsahují jednotky, které jsou na sobě vzájemně nezávislé v tom smyslu, že nejsou přímo ovlivněny anticipovanými reakcemi ostatních lidí. V momentě dotazování je přítomno mnohem méně činitelů, které ovlivňují prezentovaný postoj či názor, a panuje v něm často výrazně příjemnější klima než např. v diskuzi na Facebooku. Ve výzkumném rozhovoru se respondent do značné míry vyjadřuje nezávisle na mínění ostatních lidí, neboť neví, jak odpovídají ostatní respondenti ve výběru, a navíc se neobává, že prezentovaný názor vzbudí hlasitou a leckdy nevybíravou kritiku. Na sociálních sítích je situace zcela jiná; vliv sociálního okolí je zde přímý a reakce uživatelů jsou závislé na reakcích ostatních uživatelů, neboť obsahují inklinaci k chování skupiny, ovlivnění ostatními uživateli či obavy z projevení vlastního názoru. Nezávislost pozorování v datech z výběrových šetření je kvalita, které nemohou data ze sociálních sítí konkurovat.

Vliv sociálního okolí hraje svoji roli i v transakčních datech, zde má však spíše charakter odchylky vzniklé v důsledku sociální desirability než přímého ovlivnění ostatními lidmi. Z transakčních dat mohou vypadat údaje reflektující chování, jež lidé považují za společensky nežádoucí, jako je např. nákup alkoholu či erotických pomůcek. Nákupy tohoto zboží mohou být výrazně častěji placeny v hotovosti, nikoliv platební kartou či bankovním převodem [Couper 2013]. Pokročilejší uživatelé internetu mohou rovněž maskovat svůj pohyb v online prostoru, pokud mají pocit, že se dopouštějí sociálně nevhodného chování. Dalším příkladem vlivu sociální desirability v transakčních datech může být selektivní používání mobilních aplikací zaměřených na fitness a zdravou výživu, které svým uživatelům poskytují informace o kalorických hodnotách jídel, množství spotřebované energie při určitých typech pohybu atd. (např. www.kaloricketabulky.cz). Uživatelé těchto aplikací si zde vedou dietní deníky a provozovatelé

aplikace získávají informace, které lze využít pro marketingové účely. Na data z těchto aplikací však nelze plně spoléhat, neboť uživatelé nemusí být v reportování o své dietě zcela upřímní.

Online vs. offline život

Někteří autoři upozorňují, že uživatelé se na sociálních sítích chovají jinak než v offline životě [Boyd, Crawford 2012; Golder, Macy 2014]. V prostředí, kde se většinou komunikuje psaným textem a kde existuje jistá míra anonymity, se u lidí mohou projevit vlastnosti, které se v offline prostředí neuplatňují, např. agresivita. Podle Goldera a Macyho [2014] si však uživatelé do virtuálního prostředí přenašejí především své skutečné charakterové vlastnosti a z chování na síti se dá o psychologických a sociálních vlastnostech jednotlivce usuzovat leccos. Např. člověk, který vyhledává ve svém životě úspěch a obdiv ostatních, se tak bude chovat i na sociální síti.

Sociální sítě poskytují relativně velkou kontrolu nad sebe prezentací, čímž lidem umožňují budovat si image způsobem, který byl před vznikem internetu nemožný [Pospíšilová 2016]. Sociální média mají oproti reálnému světu tu výhodu, že na pomyslné goffmanovské jeviště sociálního jednání se většinou může dostávat jen to, co o sobě sám uživatel zveřejní (pomíne-li extrémní případy, kdy uživatelé zveřejňují citlivý obsah týkající se jiných uživatelů). Kontroly nad zveřejňovaným obsahem využívá v menší či větší míře každý uživatel sociálních sítí a buduje tak vlastní obraz, který o sobě chce šířit na veřejnosti [Couper 2013; Manovich 2011]. V tom uživatelům pomáhají i funkce sociálních sítí, které umožňují skrýt před veřejností obsah, který vznikl v emotivní chvíli či v důsledku snížené pozornosti. V jazyce výběrových šetření se jedná o jakýsi haló efekt, kdy si jsou jednotlivci vědomi toho, že jsou sledováni, a přizpůsobují tomu své chování; nezveřejňují fotografie, kde jim to nesluší, a vědomě sdílejí jen informace, o kterých se domnívají, že je v očích těch, na kterých jim záleží, nepoškodí.

Chování lidí na síti dále ovlivňují uživatelé, kteří jsou na síti a často i v offline světě silnými hráči. Těmito silnými hráči jsou facebookové či instagramové účty firem a značek, které ovlivňují uživatele přímo prostřednictvím svých účtů nebo zprostředkovaně přes celebrity a blogery, kteří dané značky propagují. Uživatelé jsou prostřednictvím účtů vyzýváni, aby ke svým fotografiím přidávali konkrétní hashtagy či se fotili s daným produktem, např. za účelem účasti v soutěži. Často jsou také nabádáni, aby komentovali určitý obsah, a je jim sděleno, o čem konkrétně mají v komentáři psát; mohou být rovněž požádáni, aby lajkovali určitý obsah či ho sdíleli na svých profilech. Některé velké značky jako Apple nebo Starbucks se již dostaly do tak silné pozice, že uživatelé používají jejich produkty jako statusový symbol a sdílejí fotografie těchto produktů na sociálních sítích sami od sebe, aniž by je k tomu musel kdokoliv vyzývat.

Kromě výše uvedených specialit online světa, které je nutné mít na paměti, když používáme data ze sociálních sítí k výzkumu, je potřeba vzít v úvahu i fakt, že analýza dat ze sociálních sítí může přinášet zkreslený obraz sociálních vztahů mezi lidmi. V prostředí sociálních sítí může docházet ke vzniku vzájemných, leckdy silných vazeb mezi lidmi, kteří se ve skutečném světě nikdy nepotkali. Ze silné vazby identifikované na sociální síti tak nelze se stoprocentní jistotou usuzovat na silnou vazbu v reálném světě, ze které vyplývají sociálním vědám známé benefity,

jako je např. pomoc při získání zaměstnání. Nevalidní informace o sociálních vazbách se však netýkají jen sociálních médií, matoucí mohou být i údaje z transakčních dat. Např. GPS data z mobilních telefonů lze použít k rekonstrukci vzájemné prostorové blízkosti s ostatními lidmi, nemělo by z nich ale být automaticky usuzováno na kvalitu a sílu vztahu mezi lidmi, kteří spolu tráví hodně času. Chybu, která může plynout z neuváženého používání GPS dat, lze ilustrovat na příkladu: GPS zařízení zaznamená, že majitelka sledovacího zařízení tráví dlouhodobě devět hodin denně s kolegy, se kterými sedí v kanceláři, a jen dvě hodiny týdně se svou matkou. Pokud bychom usuzovali na sílu mezilidských vztahů z času, který spolu lidé tráví, dopustili bychom se v tomto případě chyby. Žena sice s kolegy tráví mnohem více času, její skutečná vazba k nim je však výrazně slabší než vazba k vlastní matce [Ruppert, Law, Savage 2013]. Ruppert a kolegyně [2013] upozorňují, že než začneme používat nové technologie, jako je sběr velkých dat prostřednictvím chytrých telefonů, je nutné analyzovat a reflektovat novou sociální situaci samu o sobě a „konceptuálně uchopit specifické vlastnosti digitálních přístrojů a dat, která generují“.

Závěr: Budoucnost velkých dat, internetu a výběrových šetření

V současné době se zdá, že se budoucnost sociálněvědního výzkumu bez velkých dat a dalších nástrojů využívajících internet neobejde. Velká data lze již dnes považovat za potenciálně převratný fenomén, který může do výzkumu přinést mnoho inovací. Někteří nevitají velká data s optimismem a domnívají se, že obrovské nestrukturované datové soubory nemají pro sociální vědy velké využití a že obsahují především šum. Analýza těchto dat bude údajně přinášet především nevalidní výsledky [Barnes 2013]. Pro část výzkumníků však velká data představují výzvu a vzácný zdroj informací o lidských společnostech, který může sociálním vědám pomoci dále se vyvíjet. Americký metodolog Robert Groves vidí zapojení velkých dat do výzkumu jako nevyhnutelné; podle něj bude v následujících deseti letech potřeba začít propojovat data z výběrových šetření s velkými daty a přizpůsobit této nové praxi designy výběrových šetření [Habermann, Kennedy, Lahiri 2016].

Kromě odchylek, chyb, nepřesností a zkrácení ve velkých datech, které jsem představila v tomto textu, existují další chyby, které jsou důsledkem procesu získávání dat ze zdroje, editování a spojování s dalšími datovými zdroji [Japac et al. 2015]. Velké množství odchylek a chyb ve velkých datech zmiňuje jako vysoce problematický aspekt nových dat Robert Groves, podle kterého se však s většinou těchto chyb bude možné vypořádat prostřednictvím již známých modelů pro odhadování chyb měření v datech. Z hlediska nových příležitostí, které umožňují velká data, oceňuje dlouhodobou snahu metodologů v oblasti výzkumu chyb měření v datech z šetření, jež přinesla řadu postupů, které bude možné využít právě v analýze validity velkých dat [Habermann, Kennedy, Lahiri 2016]. Publikace Japacové a kolegů [2015] poskytuje základní rámec pro identifikaci chyb ve velkých datech, inspirovaný Grovesovým konceptem celkové chyby šetření (Total Survey Error).

Velká data mohou být navzdory všem chybám a nedostatkům pro kvantitativní sociálněvědní výzkum v mnohém přínosná. V první řadě mohou sociálním vědám umožnit získávat druh informací, ke kterým vědci dosud neměli přístup nebo je mohli získávat jen s vysokými náklady a s velkým množstvím chyb [Savage, Burrows 2007]. McFarland, Lewis a Goldberg [2016] zdůrazňují, že ideální vytěžení více zdrojů velkých dat umožní v budoucnosti získávat informace na úrovni sociálních systémů, takže bude možné zjišťovat v dříve nemyslitelném rozsahu, jak se vzájemně ovlivňují a proplétají ekonomická, sociální a kulturní sféra společnosti. Tyto nové znalosti podle nich způsobí významný posun v sociologické, politologické a další teorii a zasadí se o rozvoj společenských věd; umožní rovněž zefektivnění veřejných politik. Podle Goldera a Macyho [2014] najdou data ze sociálních médií a transakční data, jako jsou záznamy z telefonních rozhovorů či e-mailů, velké využití ve výzkumu sociálních sítí a interakcí a pomohou významně přispět k oživení těchto oblastí výzkumu. Velká data rovněž nepochybně rozšíří agendu metodologie kvantitativních sociálních věd, neboť bude potřeba najít řešení pro spojování různých druhů dat a identifikaci různých chyb v těchto datech.

Dalším potenciálním přínosem velkých dat v sociálních vědách může být jejich vliv na změnu organizace vědecké práce a větší důraz na týmovou spolupráci. Tradičně existuje v sociologii a v příbuzných oborech poměrně malá dělba práce. Je běžné, že jeden výzkumník či výzkumnice obstará veškeré úkoly v procesu zpracování jedné studie od stanovení výzkumných cílů přes rešerši literatury a statistickou analýzu až po grafickou úpravu obrázků v textu publikovaného článku. Malá či žádná dělba práce se odráží v množství autorů publikovaných studií; autorství v sociálních vědách se většinou připisuje jednomu či dvěma autorům, vícečetné autorství je výjimkou [McFarland, Lewis, Goldberg 2016]. Aby se velká data mohla stát běžným datovým zdrojem sociálních věd, bude nutné stavět výzkumné týmy, jejichž členové budou zodpovědní za různé druhy analýz. Již teď jsou nároky na schopnosti úspěšných výzkumníků poměrně vysoké – ideální kvantitativní sociolog či socioložka by měl/a mít nadprůměrné nadání v humanitní, ale i technické oblasti – a přibude-li mezi vyžadované schopnosti ještě zvládnutí pokročilého programování, nebude lehké takové superženy a supermuže pro sociálněvědní akademický výzkum získat. Pro úspěšné zvládnutí velkých dat v sociálních vědách proto bude zcela nezbytná týmová spolupráce lidí s různými specializacemi. Měly by začít vznikat heterogenní týmy, ve kterých budou kromě sociologů, politoložek, metodologů a datových analytiček i datoví odborníci, zodpovědní za programování databází a analýzu velkých dat. Na tuto poptávku by měly zareagovat i vzdělávací instituce; vysokoškolské fakulty sociálních a humanitních věd by mohly otevřít studijní specializace zaměřené na zpracování a analýzu nových druhů digitálních dat, které budou školit odborníky vzdělané v základech sociálních věd i práci s velkými daty. Bez týmové spolupráce mezi různými specializacemi a školení datových specialistů znalých sociálních věd podle mého názoru k rozšíření velkých dat v sociálních vědách dojde jen těžko.

Třetím možným pozitivem rozšíření analýzy velkých dat by mohlo být osvěžení přezkoumané populace, unavené častými žádostmi o vyplnění dotazníku či účast ve výzkumném rozhovoru. Pokud by soukromé firmy a agentury pro výzkum trhu a veřejného mínění začaly ve větší míře využívat velká data, snížila by se poptávka po respondentech. Pro zjišťování

spotřebitelského chování jsou velká data stejně či možná více vhodná než data z šetření, leckdy mohou být i levnější [Japec et al. 2015], takže pokud k nim budou mít agentury a firmy přístup, sníží se množství realizovaných výběrových šetření. Lidé by dostávali méně nabídek k účasti ve výzkumném rozhovoru, což by se mohlo pozitivně promítnout do návratnosti akademických výběrových šetření, a tím do kvality dat ze šetření. Pokud nás v budoucnosti čeká tento scénář, nebudou nakonec velká data něčím, co výběrová šetření zničí, ale naopak fenoménem, který zvýší jejich kvalitu a důvěryhodnost.

Velmi těžko se odhaduje, jak bude vypadat sociálněvědní kvantitativní výzkum v budoucnosti, ve které zřejmě bude hrát technologie stejnou nebo ještě větší roli než v současnosti. Kvantitativní vědci a vědkyně dosud ještě nevědí, jak kombinovat tradiční zdroje dat s velkými daty, a nemají ani jasnou představu o tom, k čemu a jak velká data využívat. Výhled navíc zastírá fakt, že internet se velice rychle mění; mění se nejsilnější internetové firmy (kdo si dnes vzpomene na vyhledávač Yahoo či síť MySpace?), mění se rovněž prostředí a pravidla sociálních sítí, jako je Facebook, a mění se populace, která využívá nové technologie. Toto dynamické prostředí má za následek, že každé pojednání o velkých datech rychle zestárne. V tuto chvíli lze s určitostí říct jen to, že v současné době nemohou velká data v sociálněvědním výzkumu nahradit výběrová šetření, neboť jejich povaha neumožňuje zodpovídat podstatnou část výzkumných otázek. Sociální vědy by si tedy měly výběrová šetření ještě ponechat.

L I T E R A T U R A :

- Anderson, Chris. 2008. „The end of theory: The data deluge makes the scientific method obsolete.“ *Wired Magazine*.
- Ansolabehere, Stephen, Eitan Hersh. 2012. „Validation: What big data reveal about survey misreporting and the real electorate.“ *Political Analysis* 20 (4): 437–459. <http://dx.doi.org/10.1093/pan/mps023>
- Babbie, Earl R. 2015. *The practice of social research*. Boston: Nelson Education,
- Barnes, Trevor J. 2013. „Big data, little history.“ *Dialogues in Human Geography* 3 (3): 297–302.
- Burrows, Roger, Mike Savage. 2014. „After the crisis? Big Data and the methodological challenges of empirical sociology.“ *Big Data & Society* 1 (1). <http://dx.doi.org/10.1177/2053951714540280>
- Bollen, Johan, Huina Mao, Xiao-Jun Zeng. 2011. „Twitter mood predicts the stock market.“ *Journal of Computational Science* 2 (1): 1–8. <http://dx.doi.org/10.1016/j.jocs.2010.12.007>
- Bollen, Kenneth A. 2009. *Structural equations with latent variables*. Hoboken: Wiley.
- Bessi, Alessandro, Emilio Ferrara. 2016. „Social bots distort the 2016 US Presidential election online discussion.“ *First Monday*, 21 (11).

- Boyd, Danah, Kate Crawford. 2012. „Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon.“ *Information, communication & society* 15, no. 5 (2012): 662–679. <http://dx.doi.org/10.1080/1369118X.2012.678878>
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam DI Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. „A 61-million-person experiment in social influence and political mobilization.“ *Nature* 489 (7415): 295–298. <http://dx.doi.org/10.1038/nature11421>
- Carr, David. 2013. „Giving Viewers What They Want.“ The New York Times. online: <http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html>
- Chandler, David. 2015. „A world without causation: Big data and the coming of age of posthumanism.“ *Millennium-Journal of International Studies*, 43 (3): 833–851. <http://dx.doi.org/10.1177/0305829815576817>
- Couper, Mick P. 2013. „Is the sky falling? New technology, changing media, and the future of surveys.“ *Survey Research Methods* 7 (3): 145–156.
- Crompton, Rosemary. 2008. „Forty years of sociology: Some comments.“ *Sociology*, 42 (6): 1218–1227. <http://dx.doi.org/10.1177/0038038508096942>
- Daas, Piet, Marko Roos, Mark van de Ven, Joyce Neroni. 2012. *Twitter as a potential source for official statistics in the Netherlands*. CBS Discussion Paper.
- Dočekal, Daniel. 2015. Velký pohled Facebooku na české uživatele: co mají nejvíce v oblíbenosti? Lupa online: <http://www.lupa.cz/clanky/velky-pohled-facebooku-na-ceske-uzivatele-co-maji-nejvice-v-oblibe/>
- de Madariaga, Inés Sanchéz. 2012. *Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation. Report of the Expert Group on Structural Change. European Commission*. online: https://ec.europa.eu/research/science-society/document_library/pdf_06/structural-changes-final-report_en.pdf
- Driscoll, Kevin, Shawn Walker. 2014. „Big data, big questions. Working within a black box: Transparency in the collection and production of big twitter data.“ *International Journal of Communication* 8 (2014): 20.
- European Social Survey 2017. webová stránka <http://www.europeansocialsurvey.org/>
- Gobo, Giamietro. 2011. „Back to Likert. Towards the conversational survey“ in *Innovation in Social Research methods* edited by Malcolm Williams, W. Paul Vogt. Sage.
- Golder, Scott A, Michael W. Macy. 2014. Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40: 129–152. <http://dx.doi.org/10.1146/annurev-soc-071913-043145>
- Grimmer, Justin. 2015. We are all social scientists now: how big data, machine learning, and causal inference work together. *PS: Political Science & Politics*, 48 (1): 80–83. <http://dx.doi.org/10.1017/S1049096514001784>

- Griffin, Andrew. 2014. Facebook manipulated users' moods in secret experiment. Independent. Online: <http://www.independent.co.uk/life-style/gadgets-and-tech/facebook-manipulated-users-moods-in-secret-experiment-9571004.html>
- Habermann, Hermann, Courtney Kennedy, Partha Lahiri. 2016. „A Conversation with Robert Groves“. *Statistical Science* 31(4): 128-137. <http://dx.doi.org/10.1214/16-STS594>
- Hanzlová, Radka. 2018. „Tisková zpráva: Důvěra k vybraným institucím veřejného života – březen 2018“. Centrum pro výzkum veřejného mínění. <https://cvvm.soc.cas.cz/cz/tiskove-zpravy/politicke/politicke-ostatni/4584-duvera-k-vybranim-institucim-verejneho-zivota-brezen-2018>
- Harding, Sandra. 2010. „Feminism, science and the anti-Enlightenment critiques“ in *Women, Knowledge and Reality: Explorations in Feminist Philosophy*, eds A. Garry & M. Pearsall, Unwin Hyman, Boston, MA, pp. 298–320.
- Hargittai, Eszter. 2015. „Is bigger always better? Potential biases of big data derived from social network sites.“ *The ANNALS of the American Academy of Political and Social Science*, 659 (1): 63–76. <http://dx.doi.org/10.1177/0002716215570866>
- Hrdina, Matouš. 2016. „Identity, Activism and Hatred: Hate Speech against Migrants on Facebook in the Czech Republic in 2015.“ *Naše společnost* 14 (1): 38–47. <http://dx.doi.org/10.13060/1214438X.2016.1.14.265>
- Hox, Joop, J. K. Roberts (eds.). 2011. *Handbook of advanced multilevel analysis*. London: Psychology Press.
- Japiec, Lilli, et al. 2015. “Big data in survey research: AAPOR task force report.” *Public Opinion Quarterly* 79 (4): 839–880. <https://doi.org/10.1093/poq/nfv039>
- Jutz, R., E. Scholz. 2016. *International Social Survey Programme: ISSP 2014 - Citizenship II. GESIS*
- Kalyvas, James, R., Michael R. Overly. 2015. *Big data business and legal guide*. CRC Press
- Karašćáková, Zuzana. 2013. Prvá priama voľba prezidenta v Českej republike na Twitteri. *Naše společnost* 2 (11): 41–52.
- Kitchin, Rob. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1):1-12. <http://dx.doi.org/10.1177/2053951714528481>
- Kosinski, Michal, David Stillwell, Thore Graepel. 2013. „Private traits and attributes are predictable from digital records of human behavior.“ *Proceedings of the National Academy of Sciences* 110 (15): 5802–5805. <http://dx.doi.org/10.1073/pnas.1218772110>
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2016. „How the Chinese government fabricates social media posts for strategic distraction, not engaged argument.“ *American Political Science Review* 111 (3): 484–501. <http://dx.doi.org/10.1017/S0003055417000144>
- Lazer, David, Ryan Kennedy, Gary King, Alessandro Vespignani. 2014. „The parable of Google Flu: traps in big data analysis.“ *Science* 343 (6176): 1203–1205. <http://dx.doi.org/10.1126/science.1248506>

- Manovich, Lev. 2011. „Trending: The promises and the challenges of big social data.“ *Debates in the digital humanities* 2: 460–475. <http://dx.doi.org/10.5749/minnesota/9780816677948.003.0047>
- Massey, Douglas S, Roger Tourangeau. 2013. „Where do we go from here? Nonresponse and social measurement.“ *The ANNALS of the American Academy of Political and Social Science* 645 (1): 222–236.
- McFarland, Daniel A., Kevin Lewis, Amir Goldberg. 2016. „Sociology in the era of big data: The ascent of forensic social science.“ *The American Sociologist* 47 (1): 12–35. <http://dx.doi.org/10.1007/s12108-015-9291-8>
- Murphy, Joe, Elizabeth Dean, Craig A. Hill, Ashley Richards. 2011. „Social media, new technologies, and the future of health survey research“ in *National Center for Health Statistics (Ed.), Proceedings of the 10th conference on health survey research methods* (pp. 231–241).
- Murphy, Joe, Craig A. Hill, and Elizabeth Dean. 2013. “Social Media, Sociality, and Survey Research.” in *Social Media, Sociality, and Survey Research*, edited by Craig A. Hill, Elizabeth Dean, and Joe Murphy, 1–34. Hoboken, NJ: John Wiley and Sons.
- Murphy, Joe, Michael W. Link, Jennifer H. Childs, Casey L. Tesfaye, Elizabeth Dean, Michael Stern, Paul Harwood. 2014. Social Media in Public Opinion Research Executive Summary of the Aapor Task Force on Emerging Technologies in Public Opinion Research. *Public Opinion Quarterly*, 78(4), 788–794. <http://dx.doi.org/10.1093/poq/nfu053>
- Pew Research Center. 2017. What Low Response Rates Mean for Telephone Surveys. Online <http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys> Citováno 11. 9. 2017.
- Pospíšilová, Marie. 2016. *Facebooková (ne) závislost: Identita, interakce a uživatelská kariéra na Facebooku*. Praha: Karolinum Press.
- Ruppert, Evelyn, John Law, Mike Savage. 2013. „Reassembling social science methods: The challenge of digital devices.“ *Theory, culture & society*, 30 (4): 22–46. <http://dx.doi.org/10.1177/0263276413484941>
- Saris, Willem. E., Gallhofer, Irmtraud N. 2014. *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken: John Wiley & Sons.
- Savage, Mike, Roger Burrows. 2007. „The coming crisis of empirical sociology.“ *Sociology* 41 (5): 885-899. <http://dx.doi.org/10.1177/0038038507080443>
- Savage, Mike, Roger Burrows. 2009. „Some further reflections on the coming crisis of empirical sociology.“ *Sociology*, 43 (4): 762–772. <http://dx.doi.org/10.1177/0038038509105420>
- Snijders, Chris, Uwe Matzat, Ulf-Dietrich Reips. 2012. „Big Data: big gaps of knowledge in the field of internet science.“ *International Journal of Internet Science* 7 (1): 1–5.
- Tabachnick, Barbara G., Linda S. Fidell, Steven J. Osterlind. 2001. Using multivariate statistics.

- Tinati, Ramine, Susan Halford, Leslie Carr, Catherine Pope. 2014. „Big data: methodological challenges and approaches for sociological analysis.“ *Sociology* 48 (4): 663–681. <http://dx.doi.org/10.1177/0038038513511561>
- Thrift, Nigel. 2005. *Knowing capitalism*. London: Sage.
- Vochocová, Lenka, Jaromír Mazák, Václav Štětka. 2016. „Nic pro holky?: Genderové nerovnosti v politické participaci na sociálních sítích.“ *Gender, rovné příležitosti, výzkum (Gender, Equal Opportunities, Research)* 17 (2): 64–75.
- Webber, Richard. 2009. Response to The Coming Crisis of Empirical Sociology: An Outline of the Research Potential of Administrative and Transactional Data. *Sociology*, 43 (1): 169–178. <http://dx.doi.org/10.1177/0038038508099104>

O AUTORECH



Johana Chyliková pracuje v Českém sociálněvědním datovém archivu Sociologického ústavu AV ČR. Doktorské studium absolvovala na Fakultě sociálních věd UK. Zabývá se metodologií sociálněvědního výzkumu, zejména chybami měření v datech ze sociálněvědních šetření.

Lze ji kontaktovat na adrese johana.chylikova@soc.cas.cz.

Recenze knihy: Niels Brügger, Ralph Schroeder (eds.). 2017. *The Web as History: Using Web Archives to Understand the Past and the Present*. London: UCL Press, 296 s.

Martin Vávra

Internet je v životě velké části současných lidí téměř všudypřítomný a současně je tok informací, které „protékají“ webovým prostorem, z celospolečenského hlediska nepředstavitelně mohutný. Ovšem zároveň je internet jen poměrně zřídka využíván jako významný zdroj pro analýzy v sociálních a humanitních vědách. A pokud již užíván je, tak pouze pro zkoumání toho, co se děje právě teď, bez vztahu k minulosti a k dimenzi sociální změny.

Autoři kapitol knihy se snaží na tuto situaci reagovat nejen tím, že na ni z různých hledisek upozorňují, ale také prováděním analýz obsahu internetu a v neposlední řadě prostřednictvím hledání podmínek, institucí a standardů, které je potřebné ustavit, aby se obsah internetu zachoval v podobě využitelné pro vědecký výzkum. Tento obsah se totiž rychle ztrácí, a pokud nejsou internetová data nějakým funkčním způsobem archivována, dochází k mizení, bez nadsázky, kulturního dědictví.

Editoři knihy Ralph Schroeder a Niels Brügger si to uvědomují obzvláště citlivě. Ralph Schroeder, který je profesorem na Oxfordské univerzitě a zároveň součástí Oxford Internet Institute, v roce 2018 publikoval knihu *Social Theory After the Internet: Media, Technology and Globalization*, kterou lze chápat jako více teoreticky zaměřený doplněk recenzované knihy. Niels Brügger je profesorem (Professor in Internet Studies and Digital Humanities) na School of Communication and Culture, která je součástí University of Aarhus, a zároveň stojí v čele Centre for Internet Studies. Oba editoři jsou tedy osobami věci znalými, což dokládá i fakt, že jsou autory nejen předmluvy knihy, ale také spoluautory některých v knize publikovaných analýz.

Pokud projdeme autory jednotlivých příspěvků, jejich vzdělanostní pozadí nebo institucionální zařazení, zjistíme, že jde o skupinu vpravdě různorodou a transdisciplinární. Také metodologické přístupy různých autorů se značně liší – najdeme zde příspěvky spadající do oblasti kvantitativní metodologie využívající vyspělé statistické postupy (často z oblasti analýzy sítí a teorie grafů), stejně jako kvalitativně orientované příspěvky využívající internetová data například pro analýzu diskurzu. To odráží skutečnost, že internet je studován jak v rámci mediálních studií, tak v rámci informační vědy – disciplíny (nebo spíše rodiny subdisciplín), která právě s nástupem internetu získala nové významné teritorium pro své působení.

Příspěvky v knize jsou rozděleny do tří částí. V první z nich se autoři zabývají rozsahem domén a subdomén, strukturami, které jsou definovány doménami jednotlivých států

(konkrétně britskou .uk a dánskou .dk), a jejich vývojem. Důležitým momentem této části je téma samotné archivace webu. Například Scott Hale, Grant Blank a Victoria Alexander v příspěvku „Live versus archive: Comparing a web archive to a population of web pages“ ukazují na případové studii webu TripAdvisor, že v současné době není archivace ani zdaleka úplná a to, co uchováváno je, nepředstavuje nějaký náhodný výběr z celé poluce stránek webu TripAdvisor, ale jde spíše o archivaci stránek popisujících a hodnotících nejprominentnější (nejnavštěvovanější, nejhodnocenější) místa.

V druhé části jsou texty zabývající se vztahem internetu a politické moci. Valérie Schafer například v textu „From far away to a click away: The French state and public services in the 1990s“ ukazuje, jak se politická sféra ve Francii v druhé polovině devadesátých let minulého století pokoušela podpořit rozvoj internetu tak, aby tento podporoval a zároveň usměrňoval demokracii a umožnil Francii být na špičce společenského i technologického vývoje. Schafer ovšem ukazuje, jak politici nebyli schopni porozumět potenciálu infomačních a komunikačních technologií a jak se projevovaly nezamýšlené důsledky provedených opatření i závislost na minulém vývoji (který ve Francii představovala například informační síť Minitel).

V části třetí se pak objevují texty zaměřené na analýzu kultury (v širokém smyslu toho slova). Robert Ackland a Ann Evans v textu „Using the web to examine the evolution of the abortion debate in Australia, 2005– 2015“ využívají analýzu sítí vzájemných odkazů webových stránek, které se věnují interrupcím, stejně jako obsahovou analýzu, aby ukázali proměnu australského diskurzu. Zajímavé nejsou pouze jejich nálezy přímo se týkající tématu (komercializace, vzestupy a poklesy významu jednotlivých aktérů), ale také metodologické problémy a jejich řešení autory – pro svou analýzu nevyužívali dostupné webové archivy (protože neobsahovaly dostatečnou informaci zejména o provázanosti webových stránek odkazy), ale spolehli se místo toho na výsledky, které poskytl vyhledávač Google, a následně na vlastní analýzu sítí odkazů přímo na nalezených stránkách.

Příspěvky se kromě substantivních analýz věnují i metodologickým aspektům, zejména uchování a dostupnosti dat z internetu. Většina zveřejněných analýz využívá v San Francisku sídlící Internet Archive a jeho aplikace (především Wayback machine). Ten, přestože představuje neocenitelný zdroj dat, má také řadu limitů, které autoři nastiňují. Zde je potřeba upozornit, že kniha se bohužel příliš nevěnuje změnám a výzvám, které v oblasti analýz i uchování dat vyvstávají z nástupu „webu 2.0“, především sociálních médií, jejichž obsah je obtížné uchovávat jak z technických, tak především z právních a etických důvodů. Právě sociální sítě a další aplikace, které jsou součástí internetu, budou, zdá se, představovat hlavní výzvu pro archivaci.

Sami editoři knihy v úvodu upozorňují, že je úkolem pro budoucnost propojit zkoumání obsahu internetu samotného s analýzou jeho kontextu a jeho užívání – tj. například se zkoumáním autorů a čtenářů internetových obsahů (což, jak bylo již mnohokrát konstatováno, ale zatím málo podrobněji zkoumáno, jsou role, které se v internetovém prostředí často prolínají).

Knihy má i určité regionální omezení. V rámci prezentovaných analýz jsou ve většině případů analyzovány jednotlivé národní domény. Je to pochopitelné z toho hlediska, že autorům

šlo ve většině případů o analýzy, které mohou přispět k poznání nějakého „národního“ tématu (například debata o interrupcích v Austrálii), ale pravděpodobně bude potřeba zapracovat více na analýzách makroregionálních (například v rámci EU) a globálních. Je jen málo věcí globálnějších, než je internet, byť recenzovaná kniha na některých místech ukazuje, že tato globálnost není bez hranic a významné jsou i lokální clustery webových stránek a aplikací (velmi silně to platí například pro Čínu, ale nejde o jediný případ).

Kniha je rozhodně užitečná pro každého, kdo zamýšlí provádět analýzy obsahu webu, případně se zamýšlí nad tím, jakým způsobem je vůbec možné obsah internetu archivovat. Obohatit však může každého, koho zajímá internet jakožto společenský fenomén.

Data z výzkumů ČT dostupná v Českém sociálněvědním datovém archivu

Marie Pospíšilová

Laická i odborná veřejnost si často klade otázku, jak dopadnou budoucí volby. Do středu pozornosti se tak dostávají předvolební výzkumy, diskutuje se to, jak jsou tyto výzkumy schopny odhadovat, jak volby dopadnou; kritice je podrobována jejich metodika (viz např. [Krejčí 2004]). Česká televize (dále ČT), jako jeden z nejdůležitějších zadavatelů průzkumů veřejného mínění v ČR [Fričová 2013], vychází vstříc těm, kteří chtějí zkoumat kvalitu a povahu dat, která toto veřejnoprávní médium zveřejňuje, a případně dělat hlubší analýzy. ČT proto zveřejňuje data z předvolebních výzkumů a výzkumů pravidelně sledujících volební postoje a názory na aktuální témata. Data slouží České televizi jako základ také pro diskuse (například pro otázky Václava Moravce) – na jejich základě jsou vybírána témata a hosté do vysílání. Aktuality a zprávy založené na datech sebraných pro ČT jsou dostupné na stránkách ČT. [ceskatelevize.cz 2018] V každém publikovaném článku je odkaz na aktuální výzkumnou zprávu, která obsahuje základní analýzy z dat z příslušné vlny.

Tento článek je určen zejména všem, kteří se zajímají o volební data, trendy v oblasti volebního chování a kteří sledují názory obyvatel ČR na aktuální dění. Vhodný je taktéž pro ty, kdo se zajímají o kvalitu výzkumů a metodologické otázky týkající se výzkumů, které si nechává zadávat Česká televize. Článek stručně představuje data České televize, která jsou veřejnosti zpřístupněna prostřednictvím Českého sociálněvědního datového archivu (dále ČSDA nebo Datový archiv)¹ Sociologického ústavu AV ČR v. v. i. Česká televize od roku 2013 systematicky zpřístupňuje data, která si nechala sesbírat výzkumnými agenturami. Data vázaná na konkrétní volby jsou zveřejňována vždy nejpozději tři roky od konání voleb. Data jsou dostupná pro nejširší odbornou veřejnost pro další analytické zpracování. Určena jsou pro nekomerční vědecký výzkum a výuku.² [Fričová, 2013] Data jsou unikátním a jedním z největších veřejně přístupných zdrojů o volebním chování a politických postojích (a nejen o nich) pro sociology, politology, další (nejen) sociálněvědní obory a nejširší veřejnost.

V současné době jsou v ČSDA v katalogu Nesstar³ dostupná data z let 2010 až 2016. Přístupná jsou následující cestou: ČSDA → Česky → Volební výzkumy. Základní frekvence se zobrazí všem uživatelům, složitější analýzy a stahování dat je možné provádět po registraci a přihlášení

¹ <http://archiv.soc.cas.cz/>

² Česká televize jako deponent dat, ani realizátoři jednotlivých výzkumů (SC & C spol. s r.o., SPSS CR s.r.o., Median, Kantar TNS nebo STEM/MARK, a.s.), nenesou žádnou zodpovědnost za případné škody či ztráty způsobené uživateli nebo třetím osobám v souvislosti s využíváním dat.

³ Nesstar je webový systém, který umožňuje prohledávání a prohlížení dat a dokumentace k těmto datům. Prostřednictvím Nesstaru je možné dělat on-line analýzy, datové soubory je možné stahovat v různých formátech. Dostupný je na adrese <http://nesstar.soc.cas.cz/webview/>, více informací o Nesstaru na <http://archiv.soc.cas.cz/elektronicky-datovy-katalog-csda-system-nesstar>.

se. Následující text představí jednotlivé datové soubory dostupné prostřednictvím Nesstaru a na příkladech ukáže, jaké výsledky je možné díky nim získat.

Data ČT dostupná v katalogu Nesstar

Česká televize doposud poskytla Datovému archivu dvojí druh dat. V prvním případě se jedná o předvolební výzkumy před konkrétními volbami. Dále se pak jedná o trackingové výzkumy⁴ zaměřující se zejména na volební model. První skupina zveřejněných dat je z předvolebních výzkumů do Poslanecké sněmovny v letech 2010 a 2013 (v katalogu jsou dostupné jako Volební výzkumy pro Českou televizi – volby 2010, Volební výzkumy pro Českou televizi – volby PS 2013) a z voleb do krajských zastupitelstev⁵ (mimo Prahu) v roce 2012 (v katalogu jsou dostupné jako Volební výzkumy pro Českou televizi – volby 2012). Druhou skupinou výzkumů archivovaných v Českém sociálněvědním archivu jsou trackingové výzkumy ČT (v katalogu jsou dostupné jako Tracking volebního modelu České televize – volby PS 2013, Trendy Česka 2014 – tracking volebního modelu České televize a Trendy Česka 2015–2016).

Předvolební výzkumy

Mezi první skupinu předvolebních výzkumů patří data věnující se volbám do Poslanecké sněmovny 2010 (Volební výzkumy pro Českou televizi – volby 2010) ve dnech 28. a 29. 5. 2010. Prvním datovým souborem dostupným k těmto volbám je Exit poll 2010⁶, který odhadoval volební výsledky v okamžiku uzavření volebních místností 29. 5. 2010 ve 14 hodin. Mimo zjišťování toho, koho dotazovaní volili, a tedy odhadu výsledků voleb, se výzkum zaměřoval na to, jaké tři úkoly by respondenti uložili budoucí vládě a kdo z politiků by byl nevhodnějším kandidátem na premiéra vzhledem k několika parametrům.

Z tabulky 1 je patrné, že přestože ČSSD získala ve volbách nejvíce hlasů, lidé jejího lídra Jiřího Paroubka v dotazovaných parametrech nepovažovali za nejlepšího kandidáta na premiéra. V otázce sestavení vlády s důvěrou poslanecké sněmovny lidé nejvíce důvěřovali Petru Nečasovi. Petr Nečas a Karel Schwarzenberg byli nahlíženi jako nejlepší v případném vyjednávání s ostatními politickými stranami a druhý jmenovaný byl považován za nejlepšího v zahraniční reprezentaci. V roce 2010 byl sestavením vlády nakonec prezidentem Václavem Klausem pověřen Petr Nečas, oproti zvyklosti pověřit lídra nejsilnější strany (viz např. [Kopecký, Kratochvíl 2010]).

⁴ Trackingový výzkum sleduje určité téma v čase skrze opakující se otázky v různých časových obdobích. Díky těmto údajům lze sledovat určité trendy v dané oblasti.

⁵ Volí se krajské zastupitelé ve 13 krajích mimo hlavní město Prahu, tam se volilo v roce 2014 a 2010.

⁶ „Exit-poll je anonymní dotazování respondentů systematicky vybíraných bezprostředně po odchodu z volební místnosti, které probíhá v průběhu samotných voleb.“ (Leontiyeva, 2004, p. 6) Dotazování probíhá velmi krátce po tom, co občané volí, daná situace se tak mnohem více blíží skutečné situaci volby, než jak je tomu u předvolebních průzkumů.

Tabulka 1. Názory na adekvátnost vybraných předsedů stran na premiéra ve vybraných parametrech, srovnání se ziskem jednotlivých stran ve volbách

Strana	Předseda strany	Kdo by v roli premiéra nejlépe reprezentoval ČR v zahraničí?	Kdo by v roli premiéra dokázal nejlépe vyjednávat s ostatními politickými stranami?	Kdo by jako premiér dokázal sestavit většinou vládu s důvěrou Poslanecké sněmovny?	Zisk jednotlivých stran ve volbách
ČSSD	Jiří Paroubek	14,8 %	15,5 %	18,9 %	22,08 %
ODS	Petr Nečas	27,1 %	28,0 %	32,6 %	20,22 %
KSČM	Vojtěch Filip	6,1 %	6,1 %	5,5 %	11,27 %
TOP 09	Karel Schwarzenberg	32,1 %	28,6 %	25,5 %	16,70 %
Věci veřejné	Radek John	14,2 %	16,0 %	12,6 %	10,88 %
KDU-ČSL	Cyril Svoboda	5,7 %	5,8 %	4,8 %	4,39 %

Zdroj: ČSDA (SC & C spol. s.r.o. (Praha), SPSS CR, & Česká televize, 2013) a ČSÚ (2010)

Druhou sadou výzkumných dat dostupných prostřednictvím Datového archivu, která patří mezi *Volební výzkumy pro Českou televizi – volby 2010*, jsou *Předvolební výzkumy v krajích 2010*. Jedná se o výzkumy, které proběhly ve všech 14 krajích v průběhu února až května 2010 před zmíněnými parlamentními volbami. V každém kraji byly zjišťovány stranické preference a otázky týkající se voleb. Mimo to byly zařazeny bloky otázek specificky se vztahujících k jednotlivým krajům. V Jihočeském kraji se specifické otázky týkaly mimo jiné jaderné energetiky, v Jihomoravském zdravotnictví, v Karlovarském se dotazování zaměřovalo na faktory volebního rozhodování specifické v tomto kraji, v Královéhradeckém na postoje k investicím do vzdělávání, v Libereckém na mistrovství světa v běžeckém lyžování, v Moravskoslezském na problémy tohoto kraje a zadluženost, v Olomouckém na rozvoj soukromého podnikání, v Pardubickém na investice, v Plzeňském na kauzu plzeňských práv, v Praze na zadávání státních zakázek, ve Středočeském na rozšiřování letiště Vodochody, v Ústeckém na regulaci prostituce a na názory na těžbu hnědého uhlí, na Vysočině na rozvoj cestovního ruchu a ve Zlínském kraji otázky směřovaly na nezaměstnanost.

Podobně jako v roce 2010 i v roce 2013 probíhaly před parlamentními volbami 25. a 26. října 2013 předvolební výzkumy pro ČT ve všech krajích – v ČSDA dostupné pod záložkou *Volební výzkumy pro Českou televizi – volby PS 2013*. Podobně i v těchto výzkumech byly mimo otázky týkající se volebního modelu pokládány otázky specifické pro jednotlivé kraje. Tabulka 2 prezentuje odpovědi na otázku „*Jste zcela rozhodnut(a), že byste v případě své účasti ve volbách do Poslanecké sněmovny v současnosti volil(a) právě tuto stranu?*“, která byla položena ve všech krajích. Data se v jednotlivých krajích sbírala od 18. 9. do 14. 10. 2013, přičemž volby proběhly 25. a 26. 10. 2013. Z výsledků výzkumu je vidět, že ani v této době těsně před volbami nebyla více než polovina lidí jasně rozhodnuta, koho bude volit. Ukazuje se, proč je obtížné odhadovat výsledky voleb pomocí předvolebních výzkumů.

Tabulka 2. Odpověď na otázku Jste zcela rozhodnut(a), že byste v případě své účasti ve volbách do Poslanecké sněmovny v současnosti volil(a) právě tuto stranu?

	Procenta
Zcela rozhodnutý/á – jinou stranu bych teď nevolil(a)	42,38 %
Takřka rozhodnutý/á – ale zvažoval(a) bych i jiné strany	21,09 %
Nejsem rozhodnutý/á – uvedená strana je jen jedna ze zvažovaných	36,53 %

Zdroj: ČSDA (Median s.r.o. (Praha), STEM/MARK a.s., & Česká televize, 2016)

Posledním výzkumem ze série předvolebních výzkumů, které jsou publikované v Datovém archivu, je výzkum týkající se voleb do zastupitelstev krajů v roce 2012 (volby probíhaly 12. a 13. října společně s prvním kolem voleb do třetiny Senátu). V katalogu Nesstar je tento výzkum pod záložkou *Volební výzkumy pro Českou televizi – volby 2012*. Data byla sbírána ve 13 krajích (všech kromě hlavního města Prahy – volby do Zastupitelstva hlavního města Prahy proběhly v roce 2010). Sběr dat v jednotlivých krajích proběhl vždy těsně před odvysíláním pořadu ČT *Otázky Václava Moravce* z příslušného kraje. Mimo otázek týkajících se volebního chování a volebních preferencí tazatelé opět pokládali otázky specifické pro daný kraj. Příkladem může být zjišťování názorů obyvatel Ústeckého kraje na otázku domobran, jak je uvádí tabulka 3, z níž je patrné, že nejvíce s domobranou souhlasili lidé se základním vzděláním a středním bez maturity, s rostoucím vzděláním pak byl souhlas nižší.

Trackingové výzkumy

Druhou zmiňovanou skupinou výzkumů archivovaných v Českém sociálněvědním archivu jsou trackingové výzkumy ČT (v katalogu jsou dostupné jako *Tracking volebního modelu České televize – volby PS 2013*, *Trendy Česka 2014 – tracking volebního modelu České televize* a *Trendy Česka 2015–2016*). Tyto výzkumy pravidelně s měsíční periodicitou sledují volební model pro volby do poslanecké sněmovny. Součástí dotazování jsou také trendové otázky na obecná společenská témata a ad hoc otázky týkající se aktuální situace. V březnu 2014 se například zkoumal

Tabulka 3. Odpověď na otázku: Souhlasíte s tím, aby obyvatelé možnost legálně zakládat tzv. domobran, tedy občanské hlídky?

Nejvyšší dosažené vzdělání	Základní	Střední bez maturity	Střední s maturitou	Vysokoškolské	Celkem
Ano	65,1 %	67,1 %	59,4 %	45,6 %	62,4 %
Ne	34,9 %	33,0 %	40,6 %	54,5 %	37,7 %

Zdroj: ČSDA (SC & C spol. s.r.o. (Praha), STEM/MARK a.s., & Česká televize, 2015)

Tabulka 4. Odpověď na otázku: Na naší politické scéně se opět diskutuje téma nemocenské v prvních třech dnech nemoci. Bylo by podle Vás správné obnovit náhradu mzdy zaměstnancům v prvních třech dnech nemoci, a to i za cenu zatížení zaměstnavatelů?

	Procenta
Ano	74,29 %
Ne	23,70 %
Neví	2,02 %

Zdroj: ČSDA (Kantar TNS (Praha) & Česká televize, 2017)

názor na úmysl vlády zvýšit minimální mzdu v následujících 4 letech, v květnu 2014 se zjišťoval názor na sankce na Rusko za eskalaci napětí na Ukrajině, v únoru 2015 se zjišťovalo povědomí o úctenkové loterii, v dubnu 2015 lidé hodnotili rekonstrukci dálnice D1. Ad hoc otázky v únoru 2016 se soustředily na téma uprchlictví a v srpnu 2016 na dopady zavedení EET.

Poslední zatím publikovaný výzkum (*Trendy Česka 2015–2016 – tracking volebního modelu České televize. 30. vlna (listopad 2016)*) mimo jiné zjišťoval názor na proplácení nemocenské v prvních třech dnech nemoci. Jak je zřejmé z tabulky 4, většina Čechů a Češek by souhlasila s náhradou mzdy v prvních třech dnech nemoci. Téma náhrady mzdy v prvních třech dnech nemoci bylo v době psaní článku (březen 2018) aktuálním tématem při sestavování nové vlády (viz např. [Bidrmanová, 2018]). Na jednu stranu se toto téma může zdát pro Čechy a Češky jako zásadní, na straně druhé může být návrh zavést tyto tři placené dny chápán jako snaha o získání voličské přízně. Každopádně se ukazuje, že trackingové výzkumy sledují nejen voličské preference v čase, ale zaznamenávají také nejdůležitější společenské otázky v daném období a názory na ně.

Data z výzkumů ČT budou v katalogu Nesstar přibývat

Výše uvedené představení výzkumů nemá být vyčerpávajícím souhrnem témat, která výzkumy pro ČT pokrývají. Mělo by se jednat spíše o ukázkou toho, co data obsahují a nabízejí, jaké další analýzy je možné s daty dělat. Česká televize se zadáváním výzkumů a poskytováním dat Českému sociálněvědnímu datovému archivu nadále pokračuje. Poslední dostupná data z trackingového výzkumu jsou již zmiňované 30. vlny z listopadu 2016. Poslední vlnou, která proběhla a ze které byla publikována výzkumná zpráva (nikoliv data) na webu ČT (ceskatelevize.cz, 2018), je nyní (červen 2018) 39. vlna z dubna 2018. Uživatelé Datového archivu mohou očekávat publikaci dat, jakmile je ČT uvolní pro užívání nejširší veřejnosti.

POUŽITÁ LITERATURA:

- Bidrmanová, M. 2018. Duel: První tři dny nemocenské mají být zase placené, shodlo se ANO s ČSSD. Přibude marodů? Retrieved March 28, 2018, from <https://www.seznamzpravy.cz/clanek/prvni-tri-dny-nemocenske-maji-byt-zase-placene-shodlo-se-ano-s-cssd-pribude-marodu-44042>
- ceskatelevize.cz. 2018. Téma Trendy Česka. Retrieved March 28, 2018, from <http://www.ceskatelevize.cz/ct24/tema/206312-trendy-ceska>
- ČSÚ. 2010. Volby do Poslanecké sněmovny Parlamentu České republiky konané ve dnech 28. 05.–29. 05. 2010. Retrieved February 28, 2018, from <https://www.volby.cz/pls/ps2010/ps2?xjazyk=CZ>
- Fričová, M. 2013. Česká televize uvolní kompletní data z předvolebních průzkumů. Vždy tři roky po konání voleb. Retrieved February 27, 2018, from <http://www.ceskatelevize.cz/vse-o-ct/press/tiskove-zpravy/?id=6647>
- Kantar TNS (Praha), & Česká televize. 2017. Trendy Česka 2015–2016 – tracking volebního modelu České televize [datový soubor] [online]. Ver. 1. 0. Praha: Český sociálněvědní datový archiv. <http://doi.org/10.14473/CSDA00201–221>
- Kopecký, J., & Kratochvíl, L. 2010. Klaus pověřil Nečase, ať jedná o vládě. Sobotku tím prezident zklamal. Retrieved March 2, 2018, from https://zpravy.idnes.cz/klaus-poveril-necase-at-jedna-o-vlade-sobotku-tim-prezident-zklamal-1pf-/domaci.aspx?c=A100604_100635_domaci_klu
- Krejčí, J. (Ed.). 2004. *Kvalita výzkumů volebních preferencí*. Praha: Sociologický ústav AV ČR.
- Leontiyeva, Y. 2004. Volební výzkumy exit-poll. *Socioweb*, (10–12): 6–7.
- Median s.r.o. (Praha), STEM/MARK a.s., & Česká televize. 2016. Předvolební výzkumy pro volby do PS 2013 [datový soubor] [online]. Ver. 1. 0. Praha: Český sociálněvědní datový archiv. <http://doi.org/10.14473/CSDA00181–194>
- SC & C spol. s.r.o. (Praha), SPSS CR, & Česká televize. 2013. EXIT POLL 2010 [datový soubor] [online]. Ver. 1. 0. Praha: Český sociálněvědní datový archiv. <http://doi.org/10.14473/CSDA00058>
- SC & C spol. s.r.o. (Praha), STEM/MARK a.s., & Česká televize. 2015. Krajské volby 2012 [datový soubor] [online]. Ver. 1. 0. Praha: Český sociálněvědní datový archiv. <http://doi.org/10.14473/CSDA00150–162>



Sociologický ústav
Akademie věd ČR



Centrum
pro výzkum
veřejného
mínění



ČSDA

