

Posudek k velké obhajobě disertační práce

### **Big Social Data nad the Study of Celebrity Fandom**

Jakub Sedláček, sociologie, FSV UK 2023

Disertační práce Jakuba Sedláčka je příspěvkem do sociologického studia celebrit, ale zejména metodologicko-empirickým průzkumem možností využití digitálních stop a ‚big social data‘ ze sociálních sítí pro hledání odpovědí na sociologicky relevantní otázky. První kapitola práce přináší vstup do sociologického zkoumání celebrit se zaměřením na roli médií. Druhá kapitola se zabývá otázkou využití digitálních stop a ‚big social data‘ v sociálně-vědním výzkumu, přičemž pokrývá technické, metodologické i epistemologické aspekty problematiky. Třetí část pak představuje samotné vlastní empirické zkoumání založené na datech získaných z profilů uživatelů Facebooku. Výzkumné otázky této části – a celé práce – se zaměřují (1) na hledání a posouzení dostupných možností sociologického studia založeném na digitálních stopách, kteří uživatelé zanechávají a zveřejňují v prostoru sociálních sítí, a (2) na zjišťování konzistentních vzorců a vztahů mez kulturní spotřebou, politickými a hodnotovými orientacemi a vztahem k celebritám.

Text práce předložený k velké obhajobě doznal zásadních úprav a strukturálních změn, a to v souladu s doporučeními formulovanými jako výsledek malé obhajoby. Na nové verzi textu je i díky průvodnímu dopisu jasně viditelné vynaložené úsilí při revizi, které výrazně prospělo celkové orientaci v textu a jeho srozumitelnému vyznění. Zároveň se zde ale také naráží na určité limity možného, tedy jak moc lze za relativně krátkou dobu pracovat s textem, který je zásadně empiricky založený a do značné míry ucelený a uzavřený. S vědomím toho, že formulované připomínky byly vhodně zohledněny, se v následujících odstavcích zaměřím především na přetrvávající problémy, které v textu vidím. Mým záměrem přitom není tvrdit, že text není dostatečný, ale otevřít diskuzi nad potenciálně zajímavými odbornými tématy. Obecně myslím, že práce může ilustrovat širší a nesnadný problém datově orientovaných disciplín či výzkumných oblastí, které věnují značnou pozornost praktické práci s daty. Toto zaměření pak jde na úkor promyšlení nejen teoreticko-epistemologických východisek, ale například i standardním a zavedeným nástrojům kvantitativní analýzy dat – matematické statistice (tento problém se odráží dále jak v kritice ‚lifestyle enclave‘, tak v problému ‚lift‘ metody).

---

Na pozadí dominantně metodologického zájmu „jaká data lze vůbec získat ze sociálních sítí a co lze s nimi dělat“ je již v práci zřetelněji formulována sociologická otázka spojená s celebritami, respektive jejich publikem. Jakub si na pozadí základního přehledu existující literatury o tématu klade otázku, zda lze vztah k celebritám chápat jako součást vzorců životního stylu, tzv. ‚lifestyle enclaves‘, které lze spojovat i s politickým a hodnotovými orientacemi. Jedná se o výsostně aktuální otázku, která má vazbu nejen na specifickou tradici bourdieuvského zkoumání kultury, nerovností a politiky, ale i na širší zájem o otázku polarizace politiky a společnosti opřených i o jiná teoretická východiska.

Tato nová argumentace v textu dizertační práce se opírá o nově nastudované texty a studie, čímž výrazně posiluje přímý vztah prezentované empirické analýzy k sociologické debatě. Zároveň je ale zřejmé, že hloubka zpracování tohoto nového tématu mohla být jen omezené, stejně tak jako to, do jaké míry se toto téma mohlo integrálně promítnout do celé práce. Samotný pojem ‚lifestyle enclaves‘ je v dnešní době v zásadě zastaralý. Svou slávu zažil v 80. a raných 90. letech (viz citovaná literatury i např. google ngram), a současná sociologie jej nejspíše z několika dobrých důvodů příliš nepoužívá.

Pokud se snažíme zachytit pohyby a diferenciaci v rámci celé společnosti, a nejde nám o studium specifických, relativně uzavřených nik či subkultur, pak představa jasně vymezených enkláv není příliš realistická. Naopak častěji uvažujeme o určitém prostoru diferencí, které nemají jasně definované hranice, nejsou enklávami v pravém slova smyslu ale pouze analyticky užitečnou parcelací, která ale nemá objektivní, jednoznačnou povahu. Ostatně právě vyjednávání hranic mezi identitami a praxe spojená s hranicemi, kdo kam patří, je bohatým předmětem společenskovedního studia. Pojem ‚lifestyle enclave‘ evokuje jednoznačnost definic, a může být spíše kontraproduktivní.

Jasněji kritická diskuze možností a limitů analýz opírajících se o digitální stopy (2.3) výrazně prospěla celkovému vyznění druhé kapitoly. Ze sociologického hlediska je nicméně zajímavé i to, do jaké míry je v textu stále opomíjen onen „elephant in the room“ – totiž, že digitální stopy zdaleka nezachycují sociální život v jeho úplnosti. Vždy se jedná o neznámým způsobem selektivní odraz jednání dané osoby, a to nejen z hlediska toho, s jakým záměrem ta která platforma měří a shromažďuje digitální data o uživatelích, ale především a hlavně z toho hlediska, jak rozsáhlá oblast sociálního života a našeho jednání není součástí digitálního světa. Lidé se stále dívají, povídají si, přemýšlí, a mnoho dalšího bez digitálních platforem. A neméně často obsahy a svůj způsob používání těchto platforem sami regulují, cenzurují a uzpůsobují svým představám o smyslu a možném dopadu dat, které zde svou přítomností vytváří. Jsem přesvědčen, že průkopnickým výzkumům využívající možnosti dat ze sociálních sítí může opravdové osvojení si kritického náhledu na to, s jakými daty vlastně pracují, jenom prospět.

V práci je nyní také věnováno více prostoru zakotvení použitých metod analýzy dat. Nejvýznamnější je asi pojmové rozpracování nástroje ‚lift bucket‘, konkrétně jeho formální vymezení v rámci ‚association rule mining‘. Tato zdánlivě malá poznámka v práci přitom představuje pomyslnou králičí noru, kterou by celá metodologie práce spojená s data science mohla propadnout do „říše divů“ standardní, všeobecně známé a užívané matematické statistiky. Onen ‚lift‘ je totiž svou definicí nápadně podobný pojmu odds ratio. Na rozdíl od spíše pragmaticky technického ‚liftu‘ má přitom odds ratio za sebou oporu v širokém poli existujícího vědění, metod, implementací ve statistických balíčcích a softwaru. Celá problematika by zasloužila podrobnější promyšlení, a krátká rešerše naznačuje, že už se tak nejspíše děje (viz např. Vu et al. 2019).

Kromě toho v této souvislosti musím upozornit na jednu potenciálně problematickou záležitost. Některé nuance publikační etiky spojené s odkazováním a povahou citací a parafrází nejsou napříč akademickým polem zcela sdílené. Převzít strukturu výkladového odstavce a s drobnou variací použitých slov ji použít ve vlastním textu s náležitým odkazem je pro někoho přijatelnou praxí, pro někoho nikoliv. Já osobně se řadím k druhé skupině. V textu disertační práce se tento problém objevuje právě v pasáži této kapitoly (srov. 3.2.3 a citovaný Praet 2021, p. 194-5), a to včetně přebírání referencí na odkazované texty (např. „McNicholas et al. 2008“). Byl bych raději, kdyby se takovým způsobem v akademických textech nepostupovalo, a kdyby autor odcházel s poučením, že takto „se to nedělá“.

S ohledem na náročnost sběru a přípravy dat je srozumitelné, že samotná empirická analýza v předloženém textu doznala menšího počtu změn. S ohledem na vymezený záměr vyjádřit se k diferenciaci sociálního prostoru a k tzv. ‚lifestyle enclave‘ se ale nyní více nabízí otázka po povaze datového setu. Co vlastně sestavený amalgám aktivních sledovatelů určitých celebrit/stránek představuje? Můžeme vlastně na základě těchto dat říct cokoli o společnosti jako takové? Nebo se

jedná o analýzu údajů o nesourodých skupinách lidí, které dokážeme vymezit operacionálně, tj. jak jsme k této sadě dospěli, ale už ne věcně, tj. co vlastně data představují?

---

Řada témat a zajímavých otázek spojených s prací byla již probírána a probrána během malé obhajoby. Při posuzování druhé verze práce vycházím z toho, jakou pozici k textu zaujala komise při malé obhajobě a do jaké míry se podařilo autorovi tyto požadavky naplnit. Myslím, že je zcela na místě vyzdvihnout a ocenit rozsah změn. Jakub neváhal do značné míry restrukturovat základní půdorys práce, upravit pořadí kapitol, výrazně redukovat části „mediální“ kapitoly, které nedokázaly pro cíle práce být vhodnou oporou, ale i doplnit podstatné a chybějící momenty.

Nová verze disertační práce podle mého názoru naplňuje doporučení formulovaná u malé obhajoby a lze jí tak doporučit k obhajobě. K diskusi během obhajoby lze navrhnout řadu témat otevřených výše v posudku. Další otázkou by mohlo být, zda by při novém otevření obdobného výzkumného záměru autor postupoval stejně, případně by mu přišlo užitečnější postup revidovat, a jak? A to jak z hlediska vymezení záměru, tak i jeho metodiky (sběru dat, analýzy aj.).

V Praze 27. 3. 2023

Ondřej Špaček