

Při vývoji modelů pro šíření viru myslíme i na to, že lidské chování se mění z týdne na týden

Se sociologem Tomášem Diviákem o metodách analýzy sociálních sítí, které mohou pomoci kontrolovat šíření nemocí, zkoumat džihádistické skupiny i odkrývat politickou korupci.

Původně měla být řeč o maturitě z matematiky a o zdánlivé propasti mezi dvěma světy, přírodovědným a humanitním, jež se podle mnohých spíše tiše tolerují, než aby si vzájemně prospívaly. Na vědecké práci sociologa Tomáše Diviáka je však přesvědčivě patrné, že dichotomie mezi exaktním a společenskovědním myšlením je do značné míry umělá. Schopnost promítat okolní realitu do čísel je užitečná nejen v časech pandemie, kdy modely a odhady dalšího vývoje bezprostředně ovlivňují náš každodenní život. Stejně cenná je i schopnost uvažovat o řeči čísel kriticky – to znamená způsobem, k němuž člověka vedou společenské vědy.

Sociolog Tomáš Diviák se věnuje metodám analýzy sociálních sítí (Social Network Analysis, SNA) čili mapování a popisu lidských interakcí v čase a prostoru. V tomto případě se tedy nejedná o rozbor lidského chování ve virtuálním světě, našich „lajků a výroků“ například na sociálních sítích typu Facebook či Instagram. Tomáš Diviák je autorem či spoluautorem řady studií týkajících se vazeb a interakcí členů skupin organizovaného zločinu, věnoval se zahraničním kontaktům českých poslanců a srovnával je s oficiální zahraniční politikou české vlády. Mezi kauzy, jež zkoumal, patřila takzvaná metanolová aféra z roku 2002, metody SNA mu však byly užitečné i při zkoumání případu někdejšího středočeského hejtmana Davida Ratha. V současnosti však sociologa nejvíce

vytěžuje aktuální dění.

Mají metody analýzy sociálních sítí něco společného s trasováním čili sledováním fyzického pohybu osob, které je podkladem pro modely takzvané chytré karantény?

Analýza sociálních sítí toho má s trasováním společného tolik, že se v poslední době věnuji hlavně jemu. Ve spolupráci přibližně s třiceti dalšími vědci, sociology, matematiky i informatiky, dáváme dohromady modelovací a metodologické praktiky, které nám umožní tvůrčím i využitelným způsobem předjímat, jak by se mohl v čase a prostoru šířit koronavirus. Prostřednictvím inovací některých základních modelů včetně toho epidemiologického jsme dospěli k takzvanému Modelu M. Jeho cílem je realisticky zpodobnit běžné okresní město v Česku, jakési SimCity, a vytvořit jeho síťovou mapu, abychom si dovedli lépe představit, jak by se v něm šířila epidemie. Bereme přitom v potaz časovou osu šíření viru, zohledňujeme, kdy lidé pocítí první příznaky, zdali se příznaky vždy projevují nebo neprojevují. Rádi bychom také ukázali, které kontakty jsou rizikové více a které méně.

Je koronavirus první příležitostí, kdy se takový postup používá?

Zdaleka ne, síťová epidemiologie je tu s námi už od 80. let 20. století, kdy se mezi lidmi začal šířit virus HIV způsobující nemoc AIDS. I díky těmto znalostem si můžeme některá dnešní data naaproximovat, i když je třeba ve skutečnosti nemáme – teoreticky sice někde existují, ale získat je by bylo nesmírně logisticky, finančně i eticky náročné. Díky tomu, že v posledních deseti letech je vcelku běžně dostupná silná výpočetní technika, která dokáže rychle zpracovat velký objem dat a pomoci nám se simulacemi možných scénářů, se aproximace přímo nabízí.

Zmínil jste několik proměnných, které berete při budování českého SimCity v úvahu, ale nezměnili jsme jako lidé od

počátku pandemie své chování do té míry, že byt jen pár měsíců stará „vědecká zkušenost“ s lidskými interakcemi ji nedokáže postihnout, natož vyčíslit?

To je důležité téma, které ve výzkumu zohledňujeme. Od začátku víme, že nemáme přesnou síťovou mapu běžného města za normálního stavu, nemáme ji ani po zavedení například karanténních opatření, kdy lidé častěji zůstávají doma, stýkají se s menším okruhem lidí než dříve a podobně. Tuto změnu však můžeme sledovat prostřednictvím klasických sociologických výzkumů – vymysleli jsme třeba dotazník, v němž se lidí ptáme, jak v prostředích, která jsou pro šíření viru klíčová, změnili své chování. Člověk by při pohledu zvenčí mohl mít dojem, že bychom k tomu všemu potřebovali mít perfektní data, ale ono je třeba od začátku pracovat s vědomím, že i při práci se sebepřesnějšími zdroji v datech nějaká chyba bude. Drobná odbočka: toto je místo, kde se podle mě skvěle snoubí matematika a společenské vědy. Ty nás totiž učí kriticky přemýšlet o tom, nač se v rámci výzkumu díváme. Člověk se může naučit sebelépe modelovat, ale díky sociologickému zázemí si lépe uvědomuje, že za těmi čísly jsou lidé.

Změny v našem chování se projevují v detailech a některé společenské zvyklosti se mění z týdne na týden. Některí lidé jsou po opadnutí první vlny epidemie ochotní potřást si při seznamování rukou, jiní od takového člověka udělají krok zpátky. Dokážete předjímat i tyto nuance? Nebo se musíte rozhodnout, co všechno pominete, abyste měli alespoň nějaká data?

Ve výzkumu existují dva typy informací: ty které docela známe a u nichž s větší či menší přesností víme, jaké mají vlastnosti. Pak je množina informací, o níž víme, že nám chybí úplně. V takovém případě potřebujeme kvalifikovaný statistický odhad. Práce s formálním rámcem nám v takové situaci pomáhá. Řeknu si třeba: dobře, tady mám interval nejistoty, zkusím v něm modelovat. Udělám konzervativní odhad a podívám se, co

by to znamenalo pro výsledek. Pak můžu být takzvaně statisticky liberální a s neznámou hodnotou pracovat tak, jako kdyby byla vysoká. Tímhle způsobem může člověk přijít na pozoruhodné věci, které by mu jinak nedošly, kdyby byla hodnota od počátku zafixovaná. I proto nepotřebujeme přesné číslo, jak často se lidé setkávají nebo jak často si dnes podávají ruce. Do značné míry to dokážeme obejít.

K dostupnosti dat: technologické firmy, zejména telekomunikační společnosti, se už léta chlubí tím, že o nás vědí všechno. Vědí prý, co děláme od momentu, kdy procitneme a vezmeme do ruky mobil, až po chvíli, kdy chytrá aplikace vypne přehrávání filmu, protože spíme. Existují údaje z pouličních kamer, z našich webových prohlížečů, platebních karet... Dokázal by o mně vědec s přístupem k „velkým datům“ skutečně zjistit všechno?

Myslím si, že ne. Spoustě lidí učarovalo množství informací, které dnes máme k dispozici, je však třeba mít na paměti, že i v tomto případě jde o trend, buzzword. Big data však nejsou zlaté tele, které nás spasí a dá nám odpověď na všechny otázky. I v tomto případě je nutné zachovat si trochu společenskovědního skepticismu. Online sbíraná data v sobě nesou jakousi vnitřní chybu, která může unikat i člověku, který je analyzuje. Říká se jí algoritmické zkreslení. Jinými slovy, způsob sběru dat se propisuje do toho, jak výsledná data vypadají. Když nad nimi budete sedět, nemusíte mít jistotu, zdali máte informaci o tom, jak se lidé skutečně chovají, anebo o tom, jak jste k ní došla.

Dáte mi příklad?

Na podobný problém jsme narazili při výzkumu, který jsem před lety dělal s kolegy z Katedry sociologie. Mapovali jsme tehdy diskusní fórum jednoho velkého českého zpravodajského serveru, kde se v té době intenzivně řešila migrační krize. Naší výzkumnou otázkou byl mechanismus polarizace debaty – zajímalo nás, jestli na internetových fórech dochází „jen“ k výměně a

interakci názorů, anebo fóra lidí vedou k tomu, že se časem z překlenutelných názorových pozic stanou nesmiřitelnými nepřáteli. Obecně vzato, na polarizaci silně působí názoroví vůdci, lidé, kteří mají ve své „bublině“ populární názory. Ukázalo se, že příspěvky, které debatu štěpí, bývají přímo mezi doporučenými příspěvky – jde o ty nejvíce negativně i pozitivně hodnocené. Z dat, která jsme tehdy měli k dispozici, vyplynulo, že polarizaci urychluje a posiluje právě tato předvýběrová funkce. Narazili jsme však na jedno velké „ale“: řešili jsme, zdali jde o jev, který je charakteristický pro náš sběr dat, anebo pro lidské chování jako takové. Platí, že informační technologie v lidech vyhrocování postojů podporují, anebo k nim dochází přirozeně? Tuto situaci bychom nejspíš mohli ověřit experimentálně: mohli bychom se například domluvit s administrátory debaty, aby funkci doporučování vyhrocených příspěvků vypnuli, a pak sledovali, jak názorová dynamika vypadá. Tohle všechno jsou věci, které musíme zvážit, než začneme otevírat šampaňské a radovat se z překvapivých výzkumných závěrů.

Algoritmy na Facebooku prý fungují podobně: sociální síť se učí rozpoznávat obsah, který v nás vzbuzuje emocionální reakce, a nabízí nám ho častěji.

Po tom, jak funguje Facebook, jsem se nikdy podrobněji nepídil. Jako běžný uživatel sítě jsem si však všiml něčeho jiného. Když si vás přidám do přátel, Facebook mi nabídne někoho dalšího, koho bychom mohli znát oba dva. V analýze sociálních sítí tomu říkáme „uzávěra“ – poslední dvojice lidí, kteří se ještě nekamarádí, se uzavře. Z offline výzkumu lidského chování dávno vyplynulo, že toto je strašně důležitá vlastnost našich sociálních skupin. Lidé jako tvorové se kamarádí a rozšiřují své kruhy právě takto. Facebook tak taky funguje, protože jde o chytrě vymyšlený způsob, jak lidi přimět, aby na sebe vzájemně „klikali“ a dávali tím najevo, že spolu chtějí být v kontaktu, i když se třeba vůbec nemusí znát. Kdybych Facebook hodnotil jen z tohoto hlediska, všimnu

si, že uzávěra je tam strašně vysoká. Jenže to není uzávěra vytvořená uživateli, ale uměle vytvořená Facebookem.

Připomněl jste mi knihu amerického historika Jerryho Muellera s názvem Tyranie metrik, která v Česku vyšla letos v březnu, ještě „časech míru“ čili před propuknutím pandemie. Autor na příkladech ze školství a zdravotnictví ukazuje, jak jsme si zvykli vytvářet data pro data a na základě uměle vytvořených metrik přijímat rozhodnutí, která nemají s realitou a jejími problémy nic společného. V případě koronaviru se kvalita dat a následného měření řeší od začátku. Dodnes se zabýváme tím, od čeho přesně se v jednotlivých zemích odvozují počty obětí nemoci covid-19: jestli od toho, že zemřelý člověk měl „pouze“ covid-19, anebo ho měl „mimo jiné“. Nevíme, zdali jsou některá čísla vůbec pravdivá. Narážíte při práci na Modelu M na podobné problémy?

S tímhle zápolíme od doby, co jsme s modelovací iniciativou začali. Veřejné instituce v ČR obecně velmi nerady sdílejí data. Dá se jen spekulovat, proč to tak je: jejich představitelé možná usilují o pozici síly, chtějí si šetřit práci, nebo jim na druhé straně vědci nedokáží srozumitelně sdělit své požadavky a ukázat jim, co by pro ně mohli udělat. Výsledek je ale stejný – vyjma dat, která si obstaráme svépomocí, tím, že mobilizujeme známosti na různých místech, žádná nedostaneme. V tomto smyslu je sbíráme navzdory současné situaci. Jak to vypadá v zahraničí, třeba v Rusku, kde se hůř eviduje, kdo umírá na koronavirus, zatímco prudce vystřelí počet úmrtí na zápal plic, nás tolik netrápí, protože se zabýváme situací v Česku.

České SimCity tedy může být opěrným bodem chytré karantény, přestože stát vám v práci nepomáhá ani tím, že by vám řekl, co ví?

Ano, podklad pro chytrou karanténu je naším cílem, ale jestli to tak bude i ve skutečnosti, to nezávisí na nás, vědcích a analyticích. Pokud nám občas někdo pomůže zvenčí, je to na

individuální bázi.

Na čem byste pracoval, kdyby v březnu nevypukla epidemie?

Nejspíš na dvou věcech zároveň. Studoval jsem v Holandsku a díky náhodným setkáním, ale i libovůli tamních institucí, které fungují nesrovnatelně vstřícněji než ty české, jsem se dostal k projektu, který monitoruje tamní džihádisty. Projekt spadá pod holandské ministerstvo spravedlnosti a bezpečnosti, kde mají výzkumné a dokumentační centrum. Jedna ze studií, na nichž pracuji, se zabývá tím, jak se mění struktura džihádistických skupin poté, co do nich zasáhnou bezpečnostní složky. Jde o retrospektivní výzkum, týká se scény, která byla aktivní před deseti až patnácti lety, kdy došlo k erupci dramatických událostí v souvislosti s vraždou kontroverzního režiséra Thea van Gogha, který natočil dokument o útlaku muslimských žen. Radikálové, kteří v ní byli zapleteni, měli ještě další aktivity, cílem řady z nich bylo „přejít od slov k činům“. Holandská policie a bezpečnostní složky věděly, že se něco děje a chtěly tomu zamezit. Protože to chtěly udělat racionálně, vytipovaly si nejdůležitější lidi a odstranily je zatýkáním. Naše studie sledovala dvě skupiny: první z nich zásah přesně podle plánu rozprášil. V druhé, menší a původně daleko méně nebezpečné, přesně cílený zásah způsobil pravý opak: síť se aktivizovala. Lidé v ní se semkli a jako by si řekli: jestli po nás jdou, děláme něco správně. Tento scénář je samozřejmě do budoucna nebezpečnější.

Dalo vám holandské ministerstvo přístup ke všem údajům, které jste potřebovali? Pokud se nemýlím, když jste se věnoval analýze české korupční kauzy někdejšího hejtmana Davida Ratha, české úřady vám nedaly nic a vy jste musel pracovat s informacemi z médií.

Je to tak. Ani v Holandsku se každý nedostane úplně ke všemu, ale přesto jsou tam výrazně jiné podmínky. Řada informací je samozřejmě klasifikovaná jako tajná, já neměl přístup ke spisům a zdrojovým datům, ale můj holandský kolega

s bezpečnostní prověrkou ano. Obvykle jsem mu dal instrukce, co potřebuji zjistit, on to vytáhl ze zdrojových informací, anonymizoval to a já mohl údaje analyzovat. Pak jsem se na něj obrátil s výsledkem a ptal se: dává smysl, k čemu jsem došel? Mně se zdá, že v síti se dělo to a to. On na to: ano, to je přesně ono, a teď už navíc chápeme, jakou to mělo dynamiku. Výzkum tohoto typu není jenom kuriózní a historicky cenný – protože už dnes víme, jak tehdejší džihádistická scéna skončila, můžeme dělat závěry o té současné nebo si klást konkrétní otázky, k čemu směřuje. Holanďané džihádisty dál monitorují, ale do toho už zapojený nejsem.

Vidíte v českém prostředí kauzu, kterou by mělo cenu podobně analyzovat?

To je druhá věc, na kterou bych se soustředil, nebýt koronaviru: politická korupce. Kauza Rath byla z vědeckého hlediska zábavná v tom, že byla relativně malá a dobře mediálně pokrytá. Údaje z médií byly vlastně překvapivě kvalitní. V ideálním případě bych samozřejmě rád pracoval s lepšími vstupy, ale ať už to bude, jak chce, mým tématem bude znovu politická korupce či něco, čemu se v kriminologii říká „zločiny bílých límečků“. Tím se myslí korporátní podvody, praní špinavých peněz či daňové úniky, jaké je známe z kauzy Panama Papers.

Nebojíte se, že to, co zjistíte, může někdo zneužít? Nemůže vás někdo skrytě zlobbovat pro téma, na němž vydělá politické body?

Samozřejmě o tom uvažuji. Vědecko-teoretická rovina musí být neangažovaná a nehodnotící, přestože já jako člověk nějaké politické názory mám. Kdybych je do svého výzkumu promítal, poškodilo by to hodnotu mých závěrů. Asi bych mohl přijít se spoustou věcí, které by mohly někomu posloužit jako náboj nebo poškodit nějakou politickou stranu. Z těchto důvodů chci takových afér zkoumat víc najednou bez ohledu na jejich politickou profilaci.

Dají se metody analýzy sociálních sítí využít ke zkoumání nějakého lehčího tématu, kterým se zabíráte jen pro radost?

Vlastní výzkum k tomu sice nemám, ale když se chci trochu pobavit a odpočinout si, pročítám články, které analyzují zvířecí sítě. Sociálně žijící zvířata je mají samozřejmě také. Nedávno mě strašně zaujala studie jednoho druhu kavek, který je chytrý podobně jako straka nebo papoušek. Zdánlivá odbočka: o lidském chování je známo, že máme skoro nepřekonatelné sklony k reciprocitě. Když mi něco dáte, budu mít tendenci dát vám něco nazpátek. Tenhle mechanismus lidské skupiny tmelí. Když ovšem vědci v zoo studovali tento druh kavky, zjistili, že opeřenci mezi sebou občas sdílejí jídlo, ale reciprocita tam vůbec nefunguje. Prostě si nic nevracejí. Zvířecí sítě jsou fascinující. Jestli mi někdy bude ouzko ze zločinu, budu zkoumat třeba delfíny.

Zuzana Válková