

Středoškolská odborná činnost

Obor: 17. Filozofie, politologie a ostatní humanitní a společenskovědní obory

#filterbubble

Autoři: Francesca Sandroni, Slovanské gymnázium Olomouc
Jakub Dostál, Slovanské gymnázium Olomouc

Olomouc 2016

Poděkování

Neděkujeme nikomu, jsme prostě nejlepší sami o sobě

Prohlašujeme, že jsme tuto práci vypracovali samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Prohlašujeme, že tištěná verze a elektronická verze soutěžní práce SOČ jsou shodné.

V Olomouci dne

.....

Title: New ways to simulate epidemic spreading in human society

Author: Jakub Dostál

School: Slavonic grammar school Olomouc

Supervisor: Tomáš Fürst, Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc

Abstract:

Society needs to fight against dangerous diseases. However, disease may stand for much more abstract process, for instance a fashion wave.

The goal of the work is to show the need to separate the topology of the network of human relations from the behaviour of the disease itself. This will reveal us substantive dependence of the simulation results on the structure of society.

First basic ways of modelling epidemic spreading are shown, both deterministic and stochastic. Traditional methods of simulation on complex networks are recalled. Then a new algorithm is presented separating behaviour of the disease from the features of the network itself.

Keywords: epidemiology, complex networks, SI, SIS, SIR, SIRS

Název práce: Nový způsob simulace šíření epidemie ve společnosti

Autor: Jakub Dostál

Škola: Slovanské gymnázium Olomouc

Konzultant: Tomáš Fürst, Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta, Univerzita Palackého Olomouc

Abstrakt:

I v dnešní době společnost neustále pociťuje potřebu bojovat proti nebezpečným chorobám. Pod pojmem choroba si však můžeme představit i mnohem abstraktnější pojmy, jako je třeba módní trend.

Cílem práce je ukázat potřebu oddělení vlastností topologie vztahů lidí ve společnosti, kde se epidemie šíří, a samotného charakteru choroby. Tím zároveň ukážeme velkou závislost výpočtů právě na struktuře společnosti.

Nejprve uvádíme základní způsoby modelování, a to jak deterministické, tak stochastické. Ukazujeme tradiční způsoby simulací na komplexních sítích. Poté představujeme nový algoritmus oddělující vlastnosti sítě od vlastností samotné choroby.

Klíčová slova: epidemiologie, komplexní sítě, SI, SIS, SIR, SIRS

Obsah

1	Co před námi internet skrývá	3
1.1	Filter bubble	3
1.2	Problémy filter bubble	4
1.3	Výhody filter bubble	4
1.4	Preferenční algoritmy	4
2	Metodika sběru dat	5
2.1	Data z Twitteru	5
2.2	Sentimentální analýza textu	5
3	Konstrukce měření	6
3.1	Výběr pozorovaných vzorků	6
3.2	Metodika výběru vzorků	6
4	Závěr	7
	Reference	8
	Přílohy	10

Úvod

1 Co před námi internet skrývá

Ve světě plném technického pokroku a masového používání internetu se snadno můžeme ztrácet v množství přijímaných informací. Nepřeborné množství zdrojů nás dennodenně zahlcuje spoustou nových zpráv ať už na sociálních sítích, či kdekoli jinde. Je složité se v nich orientovat, natož tak hledat a nalézat relevantní souvislosti, neboť nikde není psáno, které zprávy považovat za pravdivé a které naopak za nepravdivé zprávy anonymních komentátorů.

Role, tedy působnost žurnalistiky samotné se razantně mění. Dnešní novinář nechodí s poznámkovým blokem a tužkou v kapse, získané informace a souvislosti nepsuje doma do celistvých článků poutajících pozornost čtenářů svou kvalitou, nýbrž se smartphonem v ruce fotí, natáčí a dělá živé reportáže přímo z místa.

Toto online vysílání a sdílení má kromě spousty výhod, jakožto rychlé informovanosti uživatelů od dění v silničním provozu, po nejaktuálnější výsledky voleb, také svá úskalí. Představme si kupříkladu nebezpečí atentátu, či atentát již spáchaný. Sociální sítě se začnou naplňovat spoustou příspěvků. Odhady počtu obětí, identita pachatelů a možnost dalšího ohrožení jsou hlavními tématy. Přispívat do společné online diskuze smí každý, ať už kvalifikovaný publicista nebo náhodný autor. Zájem davů se upíná na dechberoucí události a čerstvé zprávy, proto každý příspěvek, zejména z místa činu, bude dále sdílen a zmiňován jako důvěryhodný. Kde je však záruka pravdivosti příspěvku? Nemůže nakonec dojít k situaci, kdy ten nejzajímavější, leč nepodložený fakty, bude brán ve vážnost u mas lidí právě více než ten fakty podložený, jen ne tak senzační?

Mají stále žurnalisté moc ovlivnit čtenáře a oslovit je svými pracemi, či je čtenářova pozornost upínána na senzace vyskytující se na zdech sociálních sítí? Facebook a především Twitter, fungující na rychlém tweetování zpráv, se dostává do popředí sdělovacích prostředků. Na jejich zdech sledujeme mnohé odkazy na komentáře, videa a články, formující se názory jedinců i davů, vznikající i zánikající politické myšlenky. Nakolik jsou tato samozvaná masmédia relevantní a kde mají své nedostatky, obzvláště v rovnoměrném šíření informací napříč společností, se budeme zabývat dále.

1.1 Filter bubble

Filter bubble nebo také *informační bublina* je jeden z mnoha fenoménů dnešní doby. Jako první na ni upozornil a popsal ji Eli Pariser [7, 6]. Jde o jev vyskytující se na sociálních sítích, kdy uživatel každou online aktivitou¹ poukazuje na oblasti jeho zájmů. Tyto algoritmy snažící se usnadnit život pomocí personalizace viděných informací se každou další činností uživatele zdokonalují, což v konečném výsledku znamená, že jsou to právě ony, jež rozhodují, co bude pro uživatele viditelné a co naopak konsekvencí prací algoritmů bude uživateli podáno v menší míře, či úplně skryto v celkovém proudu informací [5]. Celkově to tedy může vést jedince do situace, kdy místo širokého spektra příspěvků na sobě obsahově nezávislých, vidí příspěvky jen takové, jež byly vybrány preferenčními algoritmy na základě jeho předešlé činnosti a tedy velmi zúžené škále informací.

[8] [10]

¹Tím rozumějmě kliknutí, sdílení, komentování a obdobné činnosti.

1.2 Problémy filter bubble

1.3 Výhody filter bubble

1.4 Preferenční algoritmy

Jak jsme již výše zmínili, personalizace obsahů který vidíme na internetu může být jak velkým problémem, tak velkou výhodou. Čemu jsme však dosud nevěnovali pozornost je jak informační bubliny vznikají. Je zřejmé, že *Facebook*, *Twitter*, *YouTube* a podobné internetové giganty shromažďují velké množství dat o našich internetových aktivitách. Méně jasná je, že to nedělá kvůli zlomyslným plánům na ovládnutí světa, nýbrž kvůli snaze zpříjemnit užívání jejich služeb². Tato data poté pomocí moderních matematických a statistických metod užívá například k výběru obsahu, který nám bude co nejvíce imponovat, respektive k výběru obsahu, o kterém jsme se již dříve výjádřili, že je pro nás zajímavý.

S obdobným přístupem se můžeme setkat při online nakupování [3], kde jsou nám doporučovány produkty obdobné těm, které jsme v poslední době hledali. Stejně tak například na *YouTube* [2], kde se dříve dostaneme k videím s podobným obsahem, jaký často sledujeme.

Metody, které e užívají pro sociální sítě a informační kanály [1] jsou většinou velmi sofistikované a opírají se o hluboké znalosti *machine learningu*, *statistiky* a *data mining*. Velmi zjednodušeně řečeno, nový uživatel určité stránky je nejprve vystaven velmi širokému spektru informací. Někde na serveru provozovatele sítě sedí malá ne příliš chytrá umělá inteligence, která si zapisuje na co uživatel kliká. Své zápisky následně zpracovává a pomocí těchto pracovaných poznámek následně odhaduje co by se danému uživateli mohlo líbit.

²Tím, že nám usnadní a zpříjemní jejich užívání si zajistí větší návštěvnost, což je zdroj jejich příjmů.

2 Metodika sběru dat

Jak jsme již v předchozích kapitolách zmínili, s pokrokem v oblasti technologie probíhá i obrovský pokrok v oblasti čerpání zpráv a informací. Narozdíl od minulosti, drtivá většina obyvatel vyspělých států má přístup k novinkám v průběhu celého dne. Denní počet příspěvků na *Twitteru* dosahoval v roce 2010 neskutečných 6 milionů [11]. Je zřejmé, že není možné používat stejné postupy pro analýzu takto rozsáhlého množství dat jako dříve. Naštěstí se v již nějakou dobu významným tempem posouvá i oblast *Big Data*, *Machine Learningu* a jejich podoblastí. Tyto obory nacházejí uplatnění téměř ve všech oblastech dnešní vědy [15] a stále větší pozornosti se jim dostává i v sociologii [12, 13, 14].

2.1 Data z Twitteru

Právě data z Twitteru jsou pro nás velmi vhodná. Jak se ukazuje [9], narozdíl on *Facebooku*, Twitter užívá mnoho lidí jako zdroj informací a zpráv o aktuálním dění ve světě. Samozřejmě můžeme zpochybňovat validitu a přesnost informací, které se na takovýchto sítích objevují. Ať už jsou takovéto obavy oprávněné, nebo ne, vůzkum sociologických jevů, který v této práci provádíme to nijak neovlivňuje.

Dalším důležitým důvodem, který vedl k výběru Twitteru jako média, ze kterého budeme stahovat informace je snadná přístupnost k datům pomocí služby API poskytovaný právě přímo Twitterem [17]. Konkrétně pro naše účely jsme tuto službu nepoužívali přímo, ale za pomoci balíčku *tweepy* [16] pro programovací jazyk *python*. Ten umožňuje velmi snadné ovládání a filtrování proudu dat, které si vyžádáme Twitteru a také jejich okamžitou analýzu a zpracování v *pythonu*.

2.2 Sentimentální analýza textu

3 Konstrukce měření

3.1 Výběr pozorovaných vzorků

3.2 Metodika výběru vzorků

4 Závěr

Reference

- [1] KIM, Younghoon a Kyuseok SHIM. TWITTOBI: A Recommendation System for Twitter Using Probabilistic Modeling. *2011 IEEE 11th International Conference on Data Mining*. IEEE, 2011, , 340-349. DOI: 10.1109/ICDM.2011.150. ISBN 978-1-4577-2075-8. Dostupné také z: <http://ieeexplore.ieee.org/document/6137238/>
- [2] DAVIDSON, James, Blake LIVINGSTON, Dasarathi SAMPATH, et al. The YouTube video recommendation system. *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*. New York, New York, USA: ACM Press, 2010, , 293-. DOI: 10.1145/1864708.1864770. ISBN 9781605589060. Dostupné také z: <http://portal.acm.org/citation.cfm?doid=1864708.1864770>
- [3] LINDEN, G., B. SMITH a J. YORK. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*. 2003, **7**(1), 76-80. DOI: 10.1109/MIC.2003.1167344. ISSN 1089-7801. Dostupné také z: <http://ieeexplore.ieee.org/document/1167344/>
- [4] LIAO, Q. Vera a Wai-Tat FU. Beyond the filter bubble: Interactive effects of perceived threat and topic involvement on selective exposure to information. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. New York, New York, USA: ACM Press, 2013, 2359-2368. DOI: 10.1145/2470654.2481326. ISBN 9781450318990. Dostupné také z: <http://dl.acm.org/citation.cfm?doid=2470654.2481326>
- [5] GOTTRON, Thomas a Felix SCHWAGEREIT. The Impact of the Filter Bubble – A Simulation Based Framework for Measuring Personalisation Macro Effects in Online Communities. *ARXIV: Computer Science - Social and Information Networks*. 2016, **2016**. 2016arXiv161206551G. Dostupné také z: <https://arxiv.org/abs/1612.06551#>
- [6] PARISER, Eli. (2011). *Beware online "filter bubbles"* [online]. Dostupné z https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles
- [7] PARISER, Eli. *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press, 2011. ISBN 15-942-0300-8.
- [8] Twitter. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2017, 2017-01-22 [cit. 2017-01-22]. Dostupné z: <https://en.wikipedia.org/wiki/Twitter>
- [9] BARTHEL, MICHAEL, ELISA SHEARER, JEFFREY GOTTFRIED a AMY MITCHELL. *The Evolving Role of News on Twitter and Facebook* [online]. In: . Pew Research Center, 2015 [cit. 2017-01-22]. Dostupné z: <http://www.journalism.org/2015/07/14/the-evolving-role-of-news-on-twitter-and-facebook/>

- [10] Twitter and the News: How people use the social network to learn about the world. *American Press Institute: Insights, tools and research to advance journalism* [online]. 2015, **2015** [cit. 2017-01-22]. Dostupné z: <https://www.americanpressinstitute.org/publications/reports/survey-research/how-people-use-twitter-news/>
- [11] MATHIOUDAKIS, Michael a Nick KOUDAS. TwitterMonitor. *Proceedings of the 2010 international conference on Management of data - SIGMOD '10*. New York, New York, USA: ACM Press, 2010, **2010**, 1155-1158. DOI: 10.1145/1807167.1807306. ISBN 9781450300322. Dostupné také z: <http://portal.acm.org/citation.cfm?doid=1807167.1807306>
- [12] TINATI, Ramine, Susan HALFORD, Leslie CARR a Catherine POPE. Big Data: Methodological Challenges and Approaches for Sociological Analysis. *Sociology*. 2014, **48**(4), 663-681. DOI: 10.1177/0038038513511561. ISSN 0038-0385. Dostupné také z: <http://journals.sagepub.com/doi/10.1177/0038038513511561>
- [13] MCFARLAND, Daniel A., Kevin LEWIS a Amir GOLDBERG. Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*. 2016, **47**(1), 12-35. DOI: 10.1007/s12108-015-9291-8. ISSN 0003-1232. Dostupné také z: <http://link.springer.com/10.1007/s12108-015-9291-8>
- [14] SHAH, D. V., J. N. CAPPELLA a W. R. NEUMAN. Big Data, Digital Media, and Computational Social Science: Possibilities and Perils. *The ANNALS of the American Academy of Political and Social Science*. 2015, **659**(1), 6-13. DOI: 10.1177/0002716215572084. ISSN 0002-7162. Dostupné také z: <http://ann.sagepub.com/cgi/doi/10.1177/0002716215572084>
- [15] HUBERMAN, Bernardo A. Sociology of science: Big data deserve a bigger audience. *Nature*. 2012-2-15, **482**(7385), 308-308. DOI: 10.1038/482308d. ISSN 0028-0836. Dostupné také z: <http://www.nature.com/doifinder/10.1038/482308d>
- [16] HILL, Aaron a Joshua ROESSLEIN. Tweepy. *Github* [online]. [cit. 2017-01-22]. Dostupné z: <https://github.com/tweepy/tweepy>
- [17] API Overview. *Twitter Developer Documentation* [online]. Twitter, 2016 [cit. 2017-01-22]. Dostupné z: <https://dev.twitter.com/overview/api>

Přílohy

První kapitola příloh