

ANALÝZA SOCIÁLNÍCH SÍTÍ – VYUŽITÍ V PRAXI

Jan Panuš

Univerzita Pardubice, Fakulta ekonomicko-správní, Ústav systémového inženýrství a informatiky

Abstract: *Social networks are phenomenon of these days. We can usually use it for web application like a Facebook etc. Subject of this paper is to bring different perspective on social networks and its analysis. We will show some basic metrics of social networks analysis on particular problem which is dynamic analysis of government of the Czech Republic.*

Keywords: *Social Networks, Analysis, Methods, Government of the Czech Republic*

1. Úvod

Sociální sítě se objevují v současném slovníku stále častěji. Velice oblíbeným se stalo v souvislosti s rozvojem moderních technologií spojených s internetem. Jako příklad je možné uvést používání webových aplikací typu Facebook nebo využívání tzv. chytrých telefonů (smart phones). Z toho hlediska je tedy velice vhodné zaměřit se na analýzu sociálních sítí a hledání souvislostí mezi jednotlivými prvky těchto sítí [4][1]. Sociální sítě je možné použít na analýzu chování zákazníků [2] nebo i firem mezi sebou. Jedním z kladných efektů sociálních sítí, který byl popsán, je tzv. efekt „Long tail“ [1], [5]. Ten říká, že někdy je lepší prodávat menší objemy zboží po delší čas a zaměřit se na různé skupiny zákazníků, které mají specifitější zájmy, než prodávat velké objemy unifikovaného produktu.

Cílem tohoto článku je ukázat jednu z možných aplikací analýzy sociální sítě na vybraný problém za pomoci nástroje, jímž je vyhledavač Google.

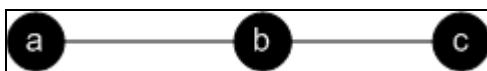
2. Formulace problematiky

2.1. Reprezentace a měření sítí

Sociální sítě hrají svoji roli v přenosu informací při hledání nového zaměstnání, také v oblasti obchodu ve výměně zboží a služeb. Sociální sítě jsou také důležité při rozpoznávání jak se šíří choroby, které produkty se kupují, jakým jazykem mluví lidé v určité oblasti, jak lidé volí, proč se lidé rozhodují zabývat se kriminalitou, jakého dosáhnou lidé vzdělání nebo jak úspěšní budeme po profesionální stránce. Je nespočet způsobů a důvodů proč využít sociální sítě a jejich analýzu. Důležité je porozumět (i) jaký vliv má struktura sítě na další vývoj a chování jedince a (ii) jaká struktura sítě se nejspíše ve společnosti objeví.

Sítě se vyjadřují převážně jako neorientované grafy, kde pokud jsou dva body spojeny, pak to znamená, že existuje vztah mezi těmito body. Vztahy jako přátelství, aliance, obchodní spojení jsou vyjádřeny tímto typem grafů. Avšak existují i speciální případy, kdy je potřeba vztahy reprezentovat pomocí orientovaných grafů, kdy tento

graf vyjadřuje směr ze kterého bodu je vztah směřován k jinému bodu. Typickým příkladem jsou citace v rámci vědeckých článků, nebo zda existují odkazy z jedné webové stránky na druhou. Nyní tedy bude následovat formální definice sítě. Graf (N, g) obsahuje množinu bodů $N = \{1, \dots, n\}$ a matici reálných čísel $n \times n$, která obsahuje prvky $g_{i,j}$, jež reprezentují vztah mezi body i a j . Jednotlivé prvky g mohou obsahovat váhu, která má vztah mít nebo směr, kterým je vztah tvořen. Jednoduchý graf je vidět na obr. 1.



Obrázek 1 - Vizualizace jednoduché sociální sítě

2.2. Měřené metriky pro síťovou analýzu

Pro síťovou analýzu se používá celá řada metrik, které vychází z teorie grafů[4]. Pro potřeby tohoto článku se budeme zabývat následujícími.

Vstupní stupeň vyjadřuje počet hran, které vstupují do daného uzlu. Používá se pouze u orientovaných grafů.

Výstupní stupeň vyjadřuje počet hran, které vystupují z daného uzlu. U neorientovaných grafů se nepoužívá.

Centralita měřená blízkostí ke středu (Closeness centrality) [4] vyjadřuje průměrnou vzdálenost mezi daným uzlem a všemi ostatními uzly, kterých může dosáhnout. Průměrná vzdálenost je vypočítána z nejkratších cest, které se v síti vyskytují.

Centralita měřená středovou mezípolohou (Betweenness centrality) [4] vyjadřuje, zda se skrz daný uzel je možno dostat do jiného uzlu v síti. Uzel, který má největší hodnotou středové mezípolohy je centrální uzel a skrze tento uzel pak proudí nejvíce informací. Tento uzel je klíčovým hráčem v dané síti.

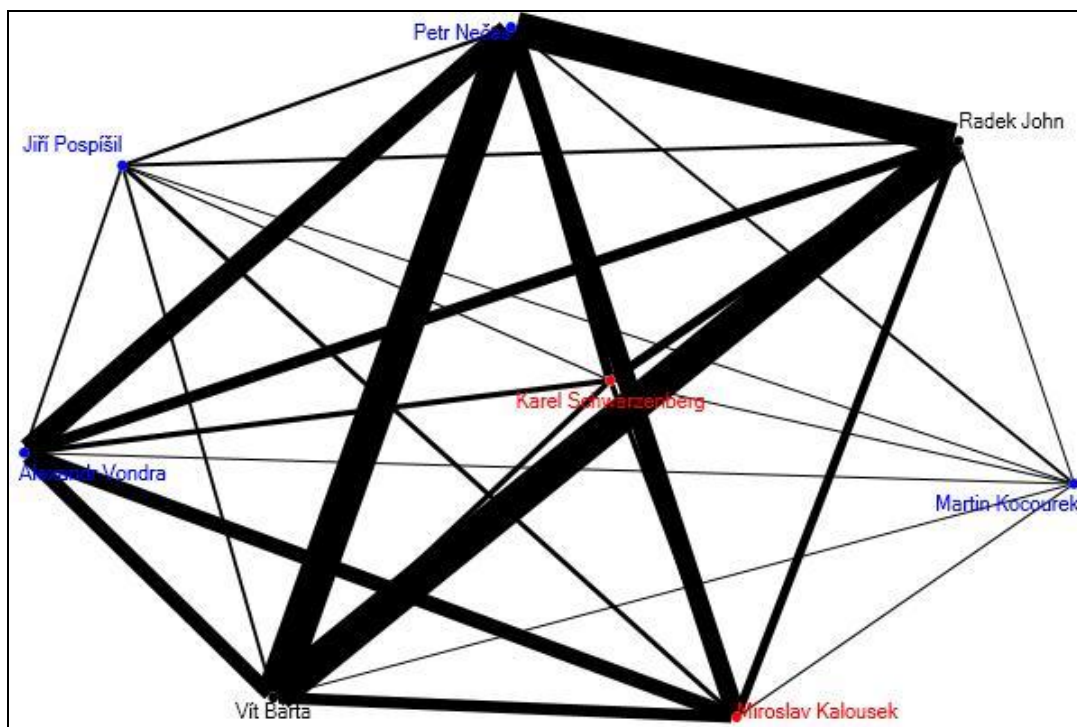
Centralita měřená koeficientem eigenvektoru [4] představuje důležitost uzlu v síti. Čím vyšší hodnota koeficientu, tím vyšší napojení daného uzlu na jiné, důležité uzly v síti. To znamená, že čím je vyšší spojení s důležitými uzly v síti (důležitost je dána např. vyšší hodnou stupně uzlu), tím je i důležitější samotný uzel. S tím souvisí i hodnota Page Rank, kterou vyvinula firma Google. Hodnota Page Rank [6] udává důležitost uzlu v síti. Tuto metriku vyvinul majitel firmy Google, Larry Page.

3. Případová studie – vybraní ministři vlády ČR

Analýza sociálních sítí slouží k nalezení struktury vztahů ve skupině nebo v organizacích. Konvenčně se řeší dotazníkovým šetřením, které je však zdlouhavé a složité. V současné době je možné si pomáhat moderními nástroji, jako jsou vyhledavače na internetu nebo je možné naprogramovat aplikaci, kterou používají členové zkoumané skupiny. U některých skupin je dostupnost jednotlivých členů obtížná a proto je nutné využít podpůrných prostředků. Z celé vlády ČR byli vybráni pouze následující ministři: Petr Nečas, Radek John, Karel Schwarzenberg, Jiří

Pospíšil, Alexandr Vondra, Miroslav Kalousek a Vít Bárta. Vzhledem k tomu, že není v silách autora ptát se jednotlivých ministrů, s kým komunikují, nebo s kým se radí, byl využit princip tzv. lingvistické koincidence.

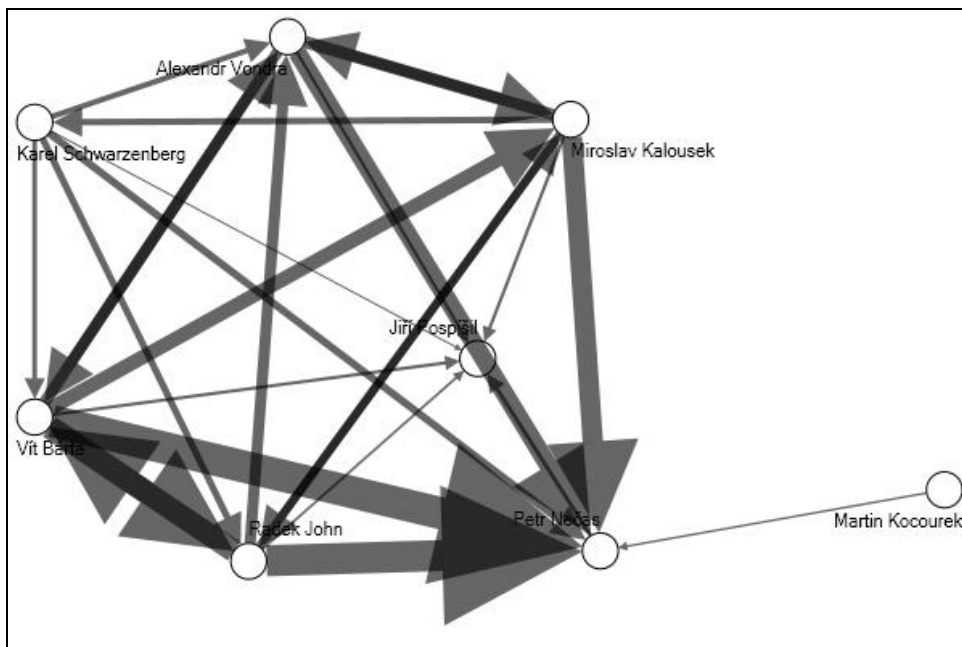
Koincidence v lingvistickém smyslu vyjadřuje, jak se určité vybrané pojmy spolu vyskytují. V případě vyslovení určitého pojmu zaznamenáváme slovo, které účastníka zkoumání napadne. Podobným způsobem zkoumáme i chování jednotlivých ministrů. Jako nástroj pro měření koincidence poslouží internetový vyhledávač, konkrétně vyhledávač od firmy Google. Byly vyhledávány vždy dva ministři a to v jednotném čísle, prvního pádu. Vyhledávání probíhalo v aplikaci Google News. Zajímá nás tedy, jak se vyskytují dvojice jmen ministrů v článkách médií. Počet nalezených odkazů vyjadřuje počet nalezených dokumentů, kde se obě jména vyskytují zároveň. Autor článku přiznává, že tento počet je pouze přibližný, jelikož v analýze nejsou zahrnuta tištěná média, rozhlas nebo televize. Poté vznikne matice, která obsahuje vazby mezi vybranými ministry. Výsledek je patrný na obr. 2. Z obrázku je patrné, že nejsilnějším článkem je logicky předseda vlády ČR, další silné vazby vznikají mezi členy strany Věci veřejné Radkem Johnem a Vítem Bártou a právě Petrem Nečasem. Tato situace je dána současnou vládní krizí. Nejslabší vazby na ostatní ministry pak mají Martin Kocourek a Jiří Pospíšil.



Obrázek 2 - Lingvistická koincidence vybraných ministrů vlády ČR ke dni 10. 4. 2011

Z výše uvedeného pak vyplývá, že Petr Nečas je jak formálním, tak i skutečným lídrem. Pro specifičtější analýzu jsme provedli nové hledání a to tak, abychom dostali

orientovaný graf. Zajímalo nás, zda po zadání určitého jména se v článku objeví i druhý ministr. Do grafu (viz obr. 3) pak byly zaznamenány silnější vazby mezi ministry. Rozdíl oproti obr. 2 je patrný především v tom, že více vazeb probíhá směrem k Petru Nečasovi, čímž se potvrzuje, že předseda vlády je nejsilnějším prvkem této skupiny.



Obrázek 3 - Specifičtější analýza dynamického chování vlády ČR ke dni 10. 4. 2011

Z výše uvedeného je možné provést přehlednou tabulku jednotlivých hodnot, které se pro síťovou analýzu mohou spočítat (viz tab. 1). Hodnota vstupního stupně říká, že pokud se v článku objeví jméno konkrétního ministra, pak se tam spíše objeví i jméno předsedy vlády, nežli by tomu bylo naopak. Tedy, že Petr Nečas má s různými kauzami, které se objevují v médiích hodně, co do činění. Jiří Pospíšil má vstupní stupeň také relativně vysoký, ale vzhledem k jeho nulové hodnotě mezilehlosti se dá spíše říci, že nebude tím hlavním článkem sítě. V analýze také byly vynechány takové hodnoty spojení ministrů, které byly nižší než 10. Důvodem bylo zamezení nízkých hodnot, které nic nevyovídají o daném vztahu.

Hodnota mezilehlosti u Petra Nečase udává jeho důležitost v síti. To také určuje i hodnota blízkosti. Petr Nečas je klíčovým hráčem a převážná většina informací, jde právě přes něj. To z něj na druhou stranu dělá i velmi vytíženého článku.

Tabulka 1: Přehled jednotlivých metrik pro dynamickou analýzu vlády ČR

Ministr	Vstupní stupeň	Výstupní stupeň	Mezilehlost	Blížkost	Eigenvektor	Page rank
Petr Nečas	6	1	12	0,143	0,142	1,316
Karel Schwarzenberg	1	5	0	0,125	0,139	1,062
Radek John	3	5	0	0,125	0,139	1,062
Miroslav Kalousek	3	5	0	0,125	0,139	1,062
Alexandr Vondra	4	4	0	0,125	0,139	1,062
Jiří Pospíšil	6	0	0	0,125	0,139	1,062
Vít Bárta	3	5	0	0,125	0,139	1,062
Martin Kocourek	0	1	0	0,077	0,024	0,31

4. Závěr

Cílem článku bylo ukázat, že je možné pomocí nástrojů analýzy sociálních sítí popsat ve vybrané skupině a to i přesto, že tuto skupinu neznáme podrobně a tudíž nemáme relevantní informace, avšak jisté informace získat dokážeme. Těmito informacemi byly data z médií, která jsme získali pomocí nástroje Google News. Jistým nedostatkem tohoto článku je analýza pouze části vlády ČR, ale cílem článku je ukázat využití základních metrik pro analýzu vybrané sociální sítě. V případě analýzy celé vlády bychom došli k podobným výsledkům. V důsledku byli jednotliví ministři vybíráni takovým způsobem, který odráží současnou situaci a víceméně ji i potvrzuje. Jistou nevýhodou je nemožnost vidět do vnitřních pochodů v rámci dané skupiny a tudíž se dá popsat pouze podle toho, co je možné získat. Dalším problémem je samotné získávání potřebných dat. Je jistě diskutabilní odvolávat se na média současného stavu, kdy v nich chybí hlubší analytická práce redaktorů, témata nejsou probírána s patřičným rešeršním aparátem. I přesto je patrné, že nástroje pro analýzu sociálních sítí jsou vhodné pro popis vybrané skupiny. Otázkou pouze je, jakým způsobem získat potřebná data a co se vlastně vyžaduje analyzovat.

Použité zdroje:

- [1] ANDERSON, C. *The Long Tail: Why the Future of Business is Selling Less of More*. New York, USA: Hyperion, 2008. 288 s. ISBN 978-1401309664.
- [2] EASLEY, D., KLEINBERG, J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge, UK: Cambridge University Press, 2010. 744 s. ISBN 978-0-511-77675-5.
- [3] GLADWELL, M. *The Tipping Point: How Little Things Can Make a Big Difference*. New York, USA: Back Bay Books, 2002. 301 s. ISBN 978-0316346627.
- [4] JACKSON, M. *Social and Economic Networks*. Princeton: Princeton University Press, 2008. 648 s.

- [5] SUROWIECKI, J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York, USA: Doubleday, 2004. 320 s. ISBN 978-0-38550-3860.
- [6] www.wikipedia.com [online]. 2011-04-10 [cit. 2011-04-10]. PageRank. Dostupné z WWW: <<http://en.wikipedia.org/wiki/PageRank#Algorithm>>.

Kontaktní adresa:

Ing. Jan Panuš, Ph.D.
Univerzita Pardubice
Fakulta ekonomicko–správní
Studentská 84,
53210 Pardubice
e-mail: jan.panus@upce.cz
tel. č.: 46 603 6001