

Univerzita Pardubice

**Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky**

**Analýza vybrané skupiny pomocí nástrojů využívaných pro měření
sociálních sítí**

Michaela Ceplová

**Bakalářská práce
2013**

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2012/2013

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Michaela Ceplová**
Osobní číslo: **E10538**
Studijní program: **B6209 Systémové inženýrství a informatika**
Studijní obor: **Informatika ve veřejné správě**
Název tématu: **Analýza vybrané skupiny pomocí nástrojů využívaných pro měření sociálních sítí**
Zadávací katedra: **Ústav systémového inženýrství a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Na vybrané skupině jedinců bude provedena analýza pomocí nástrojů, které se využívají pro měření sociálních sítí.

- 1) Výběr vhodné skupiny jedinců
- 2) Pozorování dané skupiny
- 3) Analýza získaných údajů

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

BARABÁSI, A.-L. V pavučině sítí. Praha: Paseka, 2005. 274 s. ISBN 80-7185-751-3.

DISMAN, M. Jak se vyrábí sociologická znalost. Praha: Karolinum, 2000, ISBN 80-246-0139-7.

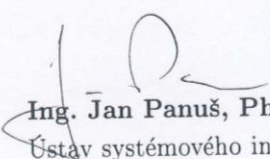
GIDDENS, A. Sociologie. Praha: Argo, 1999, ISBN 80-7203-124-4.

JACKSON, M. Social and Economic Networks. Princeton: Princeton University Press, 2008. 648 s.

KELLER, J. Úvod do sociologie. Praha: Sociologické nakladatelství, 2004, ISBN 80-86429-39-3.

WASSERMAN, S., FAUST, K. Social network analysis: methods and applications. UK: Cambridge University Press, 1994. 825 s. ISBN 0-521-38707-8.

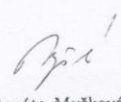
Vedoucí bakalářské práce:


Ing. Jan Panuš, Ph.D.

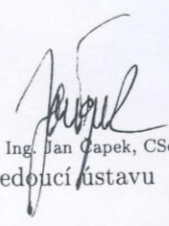
Ústav systémového inženýrství a informatiky

Datum zadání bakalářské práce: **1. října 2012**

Termín odevzdání bakalářské práce: **30. dubna 2013**


doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


prof. Ing. Jan Capek, CSc.
vedoucí ústavu

V Pardubicích dne 1. října 2012

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 30. 4. 2013

Michaela Ceplová

PODĚKOVÁNÍ:

Tímto bych ráda poděkovala svému vedoucímu práce Ing. Janu Panušovi, Ph.D. za jeho odbornou pomoc, cenné rady a poskytnuté materiály, které mi pomohly při zpracování bakalářské práce. Dále bych chtěla poděkovat své rodině za trpělivost a podporu při studiu.

ANOTACE

Tato práce se zabývá analýzou vybrané skupiny osob pomocí nástrojů využívaných pro měření sociálních sítí. Cílem této práce je na vybrané skupině uživatelů sociální sítě provést analýzu pomocí základních metrik, které se využívají pro měření sociálních sítí a popsat strukturu vztahů v této skupině. Vstupní data jsou získána importem ze sociální sítě do počítačového programu, ve kterém jsou data zpracována a poskytnou vizuální výstupy k vyhodnocení vztahů ve skupině.

KLÍČOVÁ SLOVA

sociální síť, analýza sociální sítě, graf, vrchol, hrana, centrality

TITLE

Analysis of selected group with the tools used to measure social networks

ANNOTATION

This work deals with analysis the selected group of people with the tools used for measuring social networks. The aim of this work is on a selected group of users of social network to analyze using basic metrics that are used for measuring social networks and describe the structure of relationships in this group. Input data are obtained by importing from the social network into a computer program in which the data are processed and provide visual outputs to evaluate the relationships in the group.

KEYWORDS

social network, social network analysis, graph, vertex, edge, centrality

OBSAH

ÚVOD.....	9
1 SOCIÁLNÍ SÍTĚ.....	10
1.1 PŘÍKLADY SOCIÁLNÍCH SÍTÍ	11
1.2 SOCIÁLNÍ SÍTĚ Z POHLEDU WEBU	11
1.2.1 Profil uživatele.....	12
1.3 HISTORIE WEBOVÝCH SOCIÁLNÍCH SÍTÍ.....	13
1.3.1 Vznik a rané počátky.....	13
1.3.2 Hlavní proud.....	15
1.3.3 Globální fenomén doby.....	16
1.4 SOCIÁLNÍ SÍTĚ V ČESKÉ REPUBLICE	18
2 TEORIE GRAFŮ.....	20
2.1 HISTORIE TEORIE GRAFŮ.....	20
2.1.1 Postupný vývoj v čase.....	21
2.2 ZÁKLADNÍ POJMY	22
2.2.1 Graf.....	23
2.2.2 Stupeň vrcholu a cyklus v grafu.....	25
2.2.3 Cesta, sled a souvislost grafu	25
2.2.4 Další pojmy.....	26
3 ANALÝZA VYBRANÉ SKUPINY	28
3.1 ANALÝZA SOCIÁLNÍCH SÍTÍ.....	28
3.1.1 Vývoj analýzy sociálních sítí.....	29
3.2 VYBRANÉ METRIKY PRO MĚŘENÍ SOCIÁLNÍCH SÍTÍ.....	30
3.3 VÝBĚR SOFTWARE PRO SÍŤOVOU ANALÝZU	32
3.4 VÝBĚR VHODNÉ SKUPINY	34
3.5 POPIS POSTUPU ŘEŠENÍ.....	35
3.6 ANALÝZA ZÍSKANÝCH ÚDAJŮ	38
3.6.1 Analýza následovníků uživatele	39
3.6.2 Analýza následujících uživatelem.....	42
3.6.3 Vyhledávání slovního spojení v tweetu	44
3.6.4 Analýza mapy následovníků uživatele.....	46
ZÁVĚR.....	48
POUŽITÁ LITERATURA	49

SEZNAM TABULEK

Tabulka 1: Metriky analýzy následovníků uživatele	39
Tabulka 2: Metriky analýzy následujících uživatelem	42
Tabulka 3: Metriky analýzy vyhledávání konkrétního spojení v tweetu.....	45

SEZNAM ILUSTRACÍ

Obrázek 1: Svět sociálních sítí - nejpoužívanější sociální sítě v daném regionu.....	18
Obrázek 2: Plánek sedmi mostů ve městě Královec před rokem 1875	21
Obrázek 3: Příklad obecného grafu	23
Obrázek 4: Příklad neorientovaného grafu.....	24
Obrázek 5: Příklad orientovaného grafu.....	24
Obrázek 6: Pracovní prostředí nástroje NodeXL	35
Obrázek 7: Graf síťové analýzy následovníků uživatele.....	40
Obrázek 8: Graf síťové analýzy následujících uživatelem	44
Obrázek 9: Graf síťové analýzy vyhledávání slovního spojení v tweetu	46
Obrázek 10: Mapa analýzy následovníků uživatele	47

SEZNAM ZKRATEK

AIM	Instant Messaging
ČSR	Československá republika
GPL	General Public License
HTML	HyperText Markup Language
ICQ	I Seek You
IRC	Internet Relay Chat
MS	Microsoft
SMS	Short Message Service
SNA	Social Network Analysis (analýza sociálních sítí)
USD	Americký dolar
WPF	Windows Presentation Foundation

ÚVOD

Aniž by si to lidé uvědomovali, tak se sítě objevují všude kolem nich. Existovaly mnohem dříve před vznikem internetu. Jedinci jsou součástí několika sítí, které ovlivňují jejich životy. Nic se neděje izolovaně od celku, o tom se lidé přesvědčují dnes a denně. Všichni žijí v malém světě, kde je všechno spojeno se vším ostatním. Každý je svědkem probíhající revoluce a poznání, jak jsou sítě důležité. S nástupem internetu dostal pojem sociální síť jiný význam, než byl znám daleko dříve ze sociologie. Životy lidí jsou s vývojem internetu a služeb s ním spojených ovlivňovány stále více. V současné době je to nejrychlejší a nejefektivnější způsob, jak reflektovat co největší počet lidí, komunikovat s nimi a sdílet informace. Webové sociální sítě sdružují čím dál více uživatelů a mnozí si život bez nich neumí ani představit. Ať už jsou to jedinci, kteří se zde pouze sdružují a komunikují spolu, nebo firmy, které propagují svoji společnost. Z těchto důvodů je tedy velice vhodné zaměřit se na analýzu sociálních sítí a na hledání souvislostí mezi jednotlivými uživateli těchto sítí.

Tato práce je rozdělena do tří částí. První část nahlíží na sociální síť nejprve z pohledu sociologie a až poté je vysvětlen pojem webových sociálních sítí. Je zde také přiblížen vývoj webových sociálních sítí a uvedeny jsou i současné sociální sítě, které se staly fenoménem doby. Nejsou opomenuty ani sociální sítě z českého prostředí. Druhá část stručně popisuje historický vývoj a základní pojmy z oblasti teorie grafů, které jsou základem pro analýzu sociálních sítí. Třetí část je zaměřena na popis analýzy sociálních sítí a vybraných metrik jako nástrojů pro měření sociálních sítí. Dále jsou popsány jednotlivé analýzy na vybrané skupině osob sociální sítě vytvořené pomocí počítačového programu, které jsou následně vyhodnoceny.

Cílem této bakalářské práce je na vybrané skupině osob provést analýzu pomocí nástrojů, které se využívají pro měření sociálních sítí. Autorka chce ukázat využití základních metrik pro analýzu vybrané skupiny osob sociální sítě, nalézt a popsat strukturu vztahů v této skupině i za předpokladu, že tuto skupinu nezná osobně. K analýze jsou využity informace v dostatečném rozsahu, které se dají získat ze sociální sítě, a tak není pochyb o relevantnosti vstupních dat. Tato data jsou importována do počítačového programu, kde jsou zpracována a poskytnou potřebné vizuální výstupy pro vyhodnocení.

1 SOCIÁLNÍ SÍTĚ

Sociální síť má dnes mnoho různých významů. Proto je tento pojem vysvětlen nejprve z pohledu sociologie a až poté podle definice webových sociálních sítí, kde je zmíněna historie webových sociálních sítí a její následný vývoj až do současnosti. Nejsou opomenuty ani webové sociální sítě z českého prostředí.

Sociální sítě zde byly mnohem dříve, než vznikl samotný internet. Tento pojem byl poprvé použit na přelomu 50. a 60. let 20. století v britské sociální antropologii. Zasloužil se o něj zejména J. A. Barnes (1972). Sociální síť lze charakterizovat jako množinu sociálních subjektů, které jsou navzájem propojeny směnnými vztahy. Subjekty jsou zde vrcholové body sítě a vztahy jsou vyjádřeny spojnicemi těchto bodů. Tento pojem zachycuje komunitní vztahy překračující uzavřené hranice skupin (rodin, společenských organizací, sociálních vrstev). Osobní sítě jsou tvořeny vztahy jednotlivce k druhým osobám a institucím v jeho sociálním okolí. Výzkum těchto vztahů sociální reality se nazývá analýza sociálních sítí. Pojem sociální síť byl původně použit k výzkumu komunity a rodiny, později v epidemiologii, politologii, sociologii vědy a kultury [20].

Každý člověk má potřebu se v průběhu celého života začleňovat do společnosti. Proces socializace nás provází už od narození [9]. Urbanizace přispěla k rozbití tradičních komunit a vytvořila prostor pro setkávání neznámých lidí. Došlo tak k vytvoření prostor s extrémně nízkou sociální kontrolou [15]. Slovo globalizace pochází z počátku 60. let 20. století a stalo se fenoménem souvisejícím s vývojem informačních, komunikačních a dopravních technologií. Současná postindustriální neboli informační společnost ovlivňuje život a rozhodování jednotlivců, skupin, firem i celé společnosti [2].

Spolu s rychle se rozvíjejícími poznatky o sítích se objevují také jevy, které jsou velmi přínosné. Sítě ovlivňují náš život mnohem více, než si je spousta lidí schopna reálně připustit. Internet, považovaný za čistě lidský výtvar, se začíná pomalu podobat spíše jakémusi ekosystému, který demonstruje svoji sílu pomocí zákonů pro sítě. Sítě jsou všude kolem nás, stačí se jen správně dívat. Albert Barabási pohlíží na společnost, přírodu a ekonomiku jako na komplexní sociální pavučinu a uvědomuje si malost tohoto velkého světa, ve kterém všichni žijí [1].

1.1 Příklady sociálních sítí

Příklady sociálních sítí lze v životě jednotlivců pozorovat všude kolem nich. Z pohledu teorie grafů lze tyto příklady reprezentovat pomocí vrcholů a hran. Na setkání několika lidí na večírku se lze podívat tak, že hosté jsou vrcholy grafu a jejich vzájemná setkání vytvářejí sociální vazby, tedy hrany grafu. Vzniká tak pavučina známostí, ve které jsou skupiny vrcholů spojeny hranami. Dalšími příklady jsou počítače propojené telefonními linkami, molekuly v těle ovlivňované biochemickými reakcemi, nervové buňky spojené axony, firmy a zákazníci spojení obchodem, ostrovy propojené mosty a spousty dalších. Matematicky jsou tyto příklady z pohledu grafu či sítě chápány vždy stejně. Problém nastává při lidském pohledu, kdy se společnost, internet, buňka či mozek navzájem velmi liší. Cílem je nalézt nejjednodušší možné vysvětlení velmi složitých jevů. Každý z nás je součástí jediného shluku, celosvětové sociální sítě, z něhož nikdo nemůže uniknout. Ačkoli nikdo nezná všechny lidi na planetě, je jisté, že v lidské pavučině existuje spojnice mezi každými dvěma z nich. Nic nestojí mimo tuto provázanou síť života. Paul Erdős a Alfréd Rényi tvrdí, že stačí jen jedna vazba na jeden vrchol a všichni jsou spojeni [1].

Existence komplexní sítě je důležitým klíčem k úspěchu a přináší příznivé prostředí pro naši činnost. Příkladem mohou být dva na první pohled odlišní muži, kteří mistrně ovládali síť. Prvním z nich je apoštol Saul (Svatý Pavel), který se významně podílel na šíření křesťanství. Aniž by si to uvědomoval, používal ve svém poselství teologii velmi dobře a účinně. Ovládal vlákna sociální a náboženské sítě již v 1. století. Druhým příkladem je patnáctiletý mladík skrývající se pod pseudonymem „MafiaBoy“, který v roce 2002 hackerským útokem zastavil provoz společnosti Yahoo!¹ s pomocí sítě počítačů. Důvodem jejich úspěchu byla znalost struktury, topologie sítí a jejich schopnost se v síti jako takové velice dobře pohybovat [1].

1.2 Sociální síť z pohledu webu

Pojem sociální síť se také dostal do středu zájmu internetového světa. Od této části je na sociální síť pohlíženo z hlediska webu. Sociální sítě se začaly vyvíjet jako vědecká koncepce ve druhé polovině 20. století. Předpokladem rozvoje bylo z počátku masivní rozšíření internetu a následně technologie Web 2.0, která umožnila tvorbu obsahu z pohledu uživatele. O koncepci Web 2.0 se poprvé ve svém článku zmínil Tim O'Reilly v září roku 2005 [10].

¹Yahoo! je označení pro celosvětový internetový portál.

Miliony uživatelů jsou od zavedení webových sociálních sítí v každodenním kontaktu na stránkách Facebook, MySpace, Twitter, Cyworld, Bebo, LinkedIn a dalších. Tyto stránky se v mnoha pohledech liší, ať už jde o uživatele či poskytované funkce. Jedna z možných definic pro weby sociálních sítí je, že jsou to webové služby umožňující jednotlivcům vytváření veřejných nebo poloveřejných profilů v omezeném systému. Dále umožňují vytváření seznamu přátel a procházení či prohlížení těchto seznamů. Weby sociálních sítí neslouží jednotlivcům pouze k setkávání se s lidmi, ale uživatelé zde mohou zviditelnit samy sebe, tedy vytvářet svoje aktivity. To může být příčinou toho, že dojde ke spojení jedinců, kteří by se jinak do vzájemného kontaktu nikdy nedostali. Většinou to ale není hlavním cílem, a proto jsou tyto vazby spíše latentní. Webové sociální sítě slouží především pro komunikaci mezi jednotlivci, kteří jsou již součástí rozšířené sociální sítě [5].

Sociální sítě se také liší poskytovanými službami a uživatelskou základnou. Některé slouží ke sdílení fotografií nebo video nahrávek, jiné k vytváření blogů či k rychlému zasílání zpráv. Existují specializované webové sítě, které jsou podporovány aplikacemi mobilních telefonů. Příkladem je síť Dodgeball. Omezená mobilní interakce je také přístupná pro sítě typu Facebook, MySpace a Cyworld. Výjimkou nejsou sítě, které vznikly pro uživatele obývající určitou zeměpisnou oblast či pro různé jazykové skupiny lidí. Příkladem takové sítě je Orkut, která byla spuštěna ve Spojených státech amerických. Byly vytvořeny i speciální sítě z hlediska náboženství, etniky, politického vyznání či sexuální orientace. Existují také sociální sítě, jejichž hlavním tématem jsou kočky (Catster) a psy (Dogster). Mnoho webů ve skutečnosti nevzniklo a nebylo prvotně spuštěno jako sociální síť sloužící ke vzájemné komunikaci lidí [5].

1.2.1 Profil uživatele

Všechny sociální sítě na webu poskytují řadu vlastností. Nejdůležitější z nich je tzv. profil uživatele, který mimo jiné zobrazuje seznam vytvořených přátel ve stejném systému. Seznam přátel obsahuje odkazy na další přátele těchto přátel, čímž se vytváří síťový graf, který lze po kliknutí na seznam dále procházet. Uživatel může různě upravovat vzhled svého profilu, nahrávat fotografie či multimediální obsah. Některé weby umožňují vylepšení profilu pomocí modulu (aplikace). Po vstupu na stránky sociální sítě lze identifikovat přátele, se kterými má uživatel nějaký vztah, přidáním do různých skupin, které se ale web od webu liší. Mezi populární názvy patří skupiny typu přátelé, kontakty či fanoušci. Skoro ve všech sítích je nutné obousměrné potvrzení přátelství, ale ne všude je toto pravidlem. Jednosměrné vazby v přátelství se často označují jako fanoušci či následovníci. Samotné slovo přátelé může být

ovšem velmi zavádějící a nemusí nutně znamenat přátelství jako takové. Na profilové stránce uživatele lze zanechat zprávu či komentář pro příjemce. Zasílání soukromých zpráv je velmi populární, i když nejsou dostupné ostatním uživatelům [5].

Viditelnost každého profilu se liší podle druhu sociálních sítí, důležité je ale podotknout, že každý uživatel si může svoje informace chránit v nastavení webu a omezit tím tedy svoji viditelnost (dostupnost) před ostatními uživateli systému. Weby jako jsou Friendster a Tribe.net mají výchozí nastavení takové, že je možné přes vyhledávač vidět profil uživatele bez ohledu na to, zda je člověk na této stránce sám registrován. Na Facebooku je výchozí nastavení upraveno tak, že si uživatelé mohou prohlížet profily svých přátel stejné sítě, pokud ovšem vlastník profilu nezakáže viditelnost profilu i pro své přátele, tedy jim odepře přístup k jeho osobním informacím. Například LinkedIn kontroluje viditelnost profilu podle typu placeného účtu. Stránky typu MySpace umožňují uživatelům vybrat si, zda chtějí, aby jejich profil byl veřejný, nebo dostupný pouze pro přátele. Seznam přátel je téměř na všech webech přístupný pro každého, kdo má povolený přístup na zobrazení profilu, i když i tady existují výjimky jako je MySpace či LinkedIn. Tyto strukturální změny v přístupu k profilu a v samotné viditelnosti jsou důležité pro vzájemné odlišení webů [5].

1.3 Historie webových sociálních sítí

Tato část je věnována historickému vývoji webových sociálních sítí, jsou zde uvedeny i významné milníky v jejich vývoji a nejsou opomenuty ani sociální sítě z pohledu globálního fenoménu doby.

1.3.1 Vznik a rané počátky

Historicky první sociální sítě byly tvořeny skupinami lidí, kteří používali e-mailly pro prohlubování svých sociálních vztahů. Na vzdálený počítač byl první vzkaz odeslán 2. října roku 1971. Prvními uživateli a odesílateli zpráv v počítačové síti ARPANET, která byla zárodkem dnešního internetu, byli vojáci. Dalším krokem ve vývoji sociálních sítí byl roku 1988 chat přes internet neboli IRC (Internet Relay Chat). Byl to systém, který sloužil pro přímou komunikaci lidí v reálném čase a umožňoval také sdílet soubory a odkazy. Na počátku napomohli vývoji internetu právě první počítače, elektronická pošta a posléze IRC s dalšími technologiemi. Tim Berns-Lee, britský vědec, publikoval 7. srpna roku 1991 první internetové stránky a posunul tak možnosti sociálních sítí o další krok dopředu [10].

Randy Conrad roku 1995 vybudoval první sociální síť **Classmates.com**, která ale ještě neměla všechny znaky současných webových sociálních sítí [10]. Tyto stránky sloužily

registrovaným uživatelům k surfování, vyhledávání a udržování vztahů mezi současnými či bývalými spolužáky, studenty nebo spolupracovníky. Vytváření samotných profilů uživatelů, seznamů přátel a následné procházení těchto seznamů bylo možné až o pár let později [5]. Zpočátku byla tato síť využívána nejvíce v USA a Kanadě, později se rozšířila i do Německa, Švédska, Rakouska a Francie [21].

První rozpoznatelné stránky sociálních sítí, které vyhovovaly výše zmíněné definici, byly poprvé spuštěny v roce 1997 a nazývaly se **SixDegrees.com**. Založil je Andrew Weinreich a byly první, které povolovaly propojení všech dostupných funkcí uživatelům, a to jak vytváření jejich profilů, seznamů přátel, tak i prohlížení těchto seznamů. Pomáhaly jedincům spojit se s jinými lidmi a zároveň si mezi sebou zasílat zprávy. Vytváření osobních profilů bylo podporováno na většině komunitních či seznamovacích stránkách. Služba rychlého zasílání zpráv neboli ICQ zase podporovala vytváření seznamu přátel, ten ale nebyl viditelný pro ostatní uživatele. Stránka SixDegrees přilákala více než jeden milion uživatelů, které se ovšem nepodařilo udržet, a tak byla roku 2001 ukončena. Zakladatel se domnívá, že předběhla svou dobu. Většina lidí neměla rozšířené sítě přátel, kteří by byli k dispozici ke komunikaci [5], [21].

Weby sociálních sítí začaly mezi roky 1997 a 2001 podporovat různé kombinace profilů a veřejné seznamy přátel. Stránky BlackPlanet, AsianAvenue a MiGente umožňovaly jejich uživatelům vytvářet osobní, profesní a seznamovací profily, kde je mohli ostatní uživatelé kontaktovat bez ohledu na to, zda byli nebo nebyli přátelé [5]. V této době bylo také založeno velké sociální médium Wikipedia. Na webu Blogger.com, spuštěném roku 1999, si uživatelé mohli vytvářet vlastní stránky, kde zveřejňovali své příspěvky [21]. Ve stejném roce zavedl **LiveJournal** jednosměrné připojení na uživatelské účty. Lidé si mohli přidávat ostatní uživatele do svých seznamů přátel a sledovat jejich profily [5]. Úkolem bylo přimět jednotlivce ke čtení příspěvků ostatních uživatelů, a tak se vzájemně ovlivňovat. Mezi ostatními weby měl jedinečnou funkci, a to při přidání přítele vybrat status vztahu (tajný, vzájemní přátelé nebo přátelé z jedné strany) [21]. Korejská stránka Cyworld zpřístupnila roku 2001 komplexní služby sociálních sítí [19].

Když **Ryze.com** v roce 2001 umožnil lidem využívat podnikové sítě, odstartoval tím další vlnu oblíbenosti sociálních sítí. Nejdříve se rozšířil mezi přátele, známé a členy podnikatelské a technologické komunity v San Franciscu. Lidé v Ryze.com, Tribe.net, LinkedIn a Friendster mezi sebou měli velmi dobré jak osobní, tak i profesní vztahy. Věřili ve vzájemnou podporu a absenci konkurence. Postupem času ale Ryze.com přestal být tak oblíbený jako Tribe.net.

Silnou službou se stal LinkedIn, avšak nejvýznamnější byl v té době Friendster, který byl po čase bohužel velkým zklamáním [19].

Stránka **Friendster**, založena roku 2002, měla konkurovat webu Match.com jako sociální doplněk Ryze.com. Předpokládala, že přátelé přátel pro sebe budou lepšími partnery než cizí lidé a nebudou se tak seznamovat jen na základě stejných zájmů. Netrvalo dlouho a web postihly vlivem rychlého nárůstu uživatelů značné technické a sociální potíže. Masivní příbytek lidí, kteří se o webu dozvěděli z médií, narušil kulturní rovnováhu. Uživatelé si zde mohli zobrazovat pouze profily přátel, které od nich byly vzdáleny nejvýše čtyři kroky. To jedincům nestačilo, a tak si začali do svých seznamů přidávat i neznámé lidi a různé stránky, aby si svůj dosah co nejvíce zvětšili. Postupně se začaly objevovat i falešné profily známých osobností. Docházelo k postupnému narušování důvěry mezi uživateli a mnoho z nich se z těchto stránek úplně odhlásilo, což předznamenalo jejich konec [5], [19].

1.3.2 Hlavní proud

S odstupem času se sociální sítě staly mainstreamovou záležitostí. Od roku 2003 docházelo ke spouštění nových webových sociálních sítí, které se zaměřovaly na určité skupiny lidí. Mnoho z nich zůstalo populárních dodnes [19]. Vznikali jak seznamovací, tak i podnikové stránky jako je LinkedIn, Visible Path nebo Xing [5].

Nejznámější profesionální sociální síť **LinkedIn** slouží k navazování profesních kontaktů. Výhodou je snadné vyhledávání pracovních příležitostí. Po zaregistrování si uživatel vytvoří profil, na kterém zveřejní osobní životopis a vytváří si síť kontaktů. Samotné založení účtu je zdarma, zpoplatněny jsou až další funkce. Web se stal oblíbeným i mezi manažery a profesionály, kteří zde hledají potenciální zaměstnance [21].

Ve východních zemích Evropy se stala populární síť Hi5, která se funkcemi podobala Facebooku či Myspace [21]. Sociální síť pro psí fanoušky Dogster zase napomáhá setkávání jedinců, kteří mají stejné zájmy. Postupně i weby se zaměřením na sdílení medií vykazovaly rysy typické pro sociální sítě. Příkladem je web Last.FM, na kterém uživatelé mohou poslouchat hudbu, web Flickr, který slouží ke sdílení fotografií nebo YouTube, který je velmi oblíbený z důvodu sdílení a přehrávání několikaminutových videí. S nárůstem množství sítí na trhu bylo složité si udržet pozornost a popularitu u uživatelů. Některé sítě neuspěly ve spuštěném regionu, ale staly se oblíbené v jiných zemích jako například Google Orkut nebo Microsoft Windows Live Spaces [5].

MySpace, založený roku 2003, se stal mezi uživateli velice oblíbeným. Jeho primárním cílem bylo získat uživatele Friendster [5]. Prezentuje se jako alternativní a kreativní síť pro mladé lidi. Zaměřený je především na hudbu, filmy, hry a různá aktuální témata ze světa showbyznysu. Snaží se tak propojit kulturní dění mezi lidmi v reálném čase. Své profily tu mají i známé kapely, zpěváci, herci a další celebrity. Funkčně je podobný s Facebookem, nabízí však navíc přehrávání a nakupování hudby či vytváření playlistů. Na tomto webu má každý uživatel prostor k vlastní prezentaci. Lze si měnit vzhled profilu pomocí motivů, editoru či HTML kódů. Mediální ohlas a nárůst uživatelů zaznamenal roku 2005, kdy ho odkoupil News Corporation za více než půl miliardy USD. Po příchodu Facebooku jeho obliba ale velmi klesala. V létě roku 2011 byl MySpace prodán firmě Specific Media za 38 milionů USD. Její vývoj dobře ilustruje nestálost a proměnlivost nových médií [19].

1.3.3 Globální fenomén doby

Sociální síť **Facebook** byla původně navržena jen jako kolejná síť na Harvardské univerzitě, která se zde začala používat od roku 2004. Autorem je americký programátor a podnikatel Mark Zuckerberg. Aby se jedinec mohl stát novým členem této sítě, musel mít harvardskou emailovou adresu. Netrvalo dlouho a síť se rozšiřovala i na jiné školy. V září roku 2005 se začala rozšiřovat o další studenty a firemní síť. Již o rok později si účet na stránce Facebook mohl založit kdokoli na světě. Po zpřístupnění webu zůstala různá omezení. Noví uživatelé nemohli jen tak přistupovat k uzavřené komunitě [5]. Je to nejrozšířenější a nejpoužívanější služba na světě, která po přihlášení umožňuje uživatelům komunikaci, sdílení informací, fotek, videí, hraní her a plánování akcí. Interakce probíhá především mezi přáteli v závislosti na nastavení soukromí profilu. Uživatelé mohou komentovat různé diskuze, sdílet informace, přidávat se k oblíbeným skupinám a spoustu dalšího [25]. Výhodou Facebooku je, že uživatelé nemusí zveřejňovat své profily všem uživatelům, další předností je vytváření stále nových a mezi uživateli oblíbených aplikací [5].

Roku 2005 vznikla čínská sociální síť **Qzone**, která uživatelům umožňuje vytvářet blogy a diáře, zasílat fotografie a poslouchat hudbu. Má mnoho dalších funkcí, které jsou ale zpoplatněny [21]. **Twitter**, nejúspěšnější mikroblog, byl spuštěn roku 2006. Je to nástroj, pomocí kterého lze sdělovat myšlenky, pocity nebo informace pouze o délce 140 znaků. Tyto krátké vzkazy, označované jako tzv. tweety, se zobrazují na stránce autora a také na stránkách uživatelů, kteří jsou jeho odběratele (followers). Vzkazy jsou přístupné komukoli, pokud je autor neomezí pouze na odběratele. Možná je reakce i na tweety ostatních uživatelů.

Všechno funguje tak, že si každý uživatel zvolí, od koho bude příspěvky odebírat. Stejně to funguje i obráceně, pokud si jedinec svůj profil neoznačí jako privátní. Sdružují se zde i mnozí odborníci, především z oblasti IT a moderních technologií, jejichž diskuze je možné sledovat. Tato síť má stejně jako Facebook velký reklamní potenciál, a proto jsou i tady zaregistrováni známé osobnosti. Službu je možné využívat přes webový prohlížeč, aplikace na mobilních telefonech nebo prostřednictvím SMS zpráv [19], [25].

V roce 2006 byla spuštěna velmi populární síť **Vkontakte.ru**, která je ruskou obdobou Facebooku [21]. Z počátku byla pouze v ruském jazyce, dnes je běžně uváděna jako VK.com. Po vyplnění profilu je zde možné spojit se s dalšími lidmi, komunikovat s nimi, sdílet statusy, odkazy, fotky či domlouvat srazy. Specifikem této stránky je absence her a aplikací [25].

Další roky se vyznačovaly především značným růstem oblíbenosti Facebooku. Původní sociální sítě začaly přicházet o své uživatele a vzniklo už jen pár nových sítí. Google spustil web Buzz, který ale nebyl úspěšný. Napravit si reputaci a konkurovat Facebooku se nyní snaží s novou sociální sítí **Google+** [5]. Tato síť byla spuštěna na konci června roku 2011. Prezentuje se jako služba, která pomůže uživatelům zůstat v kontaktu se známými a poznat nové lidi. Vyznačuje se absencí aplikací a funkcí, které uživatele spíše obtěžují, hry byly ale dodatečně přidány. Znaménko plus v názvu značí možnost rozdělování kontaktů do takzvaných kruhů. Pomocí tlačítka +1 lze označit obsah, který se jedinci líbí, je to tedy indikátor oblíbenosti obsahu, podle kterého se pak mohou zobrazovat i další příspěvky [25].

Existují i sociální sítě, které nelpí na rychlém růstu a expanzi nových uživatelů, ale zaměřují se spíše na uzavřené komunity lidí. Záměrně omezují přístup kohokoli do jejich systému. Příkladem může být síť BeautifulPeople pro krásné lidi. Jsou i sítě omezené z demografického hlediska nebo zaměřené na zájmové aktivity jedinců jako je CouchSurfing², BlackPlanet³ nebo MyChurch⁴. Stále existují i stránky věnované zájmovým záležitostem, které taktéž prosperují. V jejich středu jsou veřejná diskusní fóra a oblíbená témata. Příkladem takové sítě je Usenet. U většiny ostatních je středem pozornosti jednotlivec. V současné době jsou sociální sítě tak oblíbené, že někteří zaměstnavatelé omezují v pracovní době přístup svých zaměstnanců na tyto stránky. Někde dochází i k omezování připojení ve školách a v knihovnách. Vzestup sociálních sítí naznačuje, že se společnost stále více zařazuje do online komunit [5].

² CouchSurfing je internetová služba bezplatného ubytování.

³ BlackPlanet je sociální síť pro uživatele tmavé pleti.

⁴ MyChurch spojuje členy křesťanské církve na internetu.

Na obrázku 1 jsou znázorněny vždy tři nejoblíbenější sociální sítě podle daného regionu na celém světě, které odpovídají následujícím statistikám. Mapa je zpracována podle dat z roku 2011, jinak by dnes byla Brazílie zbarvena také tmavě modrou barvou, protože na první místo se i zde dostala síť Facebook [16]. Podle nejnovějších údajů z roku 2012 má Facebook jednu miliardu aktivních uživatelů a je nejpoužívanější sociální sítí ve 127/137 analyzovaných zemí světa. Z toho v Asii je to 278 milionů, v Evropě 251 milionů, v Jižní Americe 142 milionů, v Africe skoro 52 milionů a v Oceánii 15 milionů uživatelů. V Rusku převládají dvě sociální sítě a to VKontakte se 190 miliony registrovaných uživatelů a Odnoklassniki. V Číně dominuje QZone s 552 miliony uživatelů. Hned za ní následuje síť Tencent Weibo a Sina Weibo. V Íránu, kde je stránka Facebook těžko dostupná kvůli státní cenzuře, vede sociální síť Cloob [24].



Obrázek 1: Svět sociálních sítí - nejpoužívanější sociální sítě v daném regionu

Zdroj: [16]

1.4 Sociální sítě v České republice

Počet uživatelů sociálních sítí roste nejen ve světě, ale i v České republice. Velký vliv na tento růst a současnou popularitu mají i chytré telefony, které jsou stále dostupnější pro větší počet lidí a možnost přístupu na sociální sítě je jedním z lákadel. Nelze opomenout sociální sítě ve firmách, které možnost této komunikace se zákazníkem velice často využívají. Nejoblíbenější sítí s největším počtem uživatelů je Facebook, čím dál více oblíbenou se stává i profesní síť LinkedIn. Svě místo má i Twitter a velmi oblíbený server na sdílení videí

Youtube. V této části jsou připomenuty i české stránky sociálních sítí, které si ve své době získaly řadu uživatelů. Jde především o weby lide.cz, spoluzaci.cz a libimseti.cz [4].

Lidé

Na webu lide.cz si uživatelé mohou vytvořit vlastní profil, na který lze následně nahrávat fotky či videa a komunikovat s ostatními uživateli. Ke komunikaci zde slouží systém vzkazů, které mohou být textové, ale i hlasové. Dále jsou tu k dispozici diskuze na nejrůznější aktuální témata. Oblíbená je i uživatelská nástěnka a samozřejmě chat s ostatními uživateli v několika místnostech. Tato stránka slouží i jako seznamka, kde lze nalézt roztříděné inzeráty do několika kategorií podle pohlaví i zájmů. Vše je propojené, takže si případný zájemce o seznámení může po přečtení inzerátu prohlédnout i osobní profil uživatele [25].

Spolužáci

Sociální síť spoluzaci.cz slouží ke komunikaci, ať už současných či bývalých spolužáků základních, středních i vysokých škol. Uživatelé se po zaregistrování přiřadí do své skupiny neboli třídy. Samotné třídy jsou rozděleny podle jednotlivých škol a měst. Uživatelé zde mohou psát vzkazy, sdílet společné fotografie, vkládat dokumenty, přidávat jména svých učitelů a mají k dispozici i třídní nástěnku, na kterou mohou přidávat různé aktuality [4].

Líbimseti

Stránka libimseti.cz funguje v současné době jako seznamka mezi zaregistrovanými uživateli, kde po přihlášení uživatelé nahrají svoje fotografie, navzájem si je hodnotí a píšou k nim komentáře. Jsou zde diskusní fóra, chat a lze si vytvořit uživatelský blog. Se zvyšující se oblibou Facebooku tento web výrazně ztratil na popularitě. K úpadku oblíbenosti přispěla i kauza s únikem fotografií ze zaheslovaných alb uživatelů, který byl způsoben chybou v zabezpečení [25].

2 TEORIE GRAFŮ

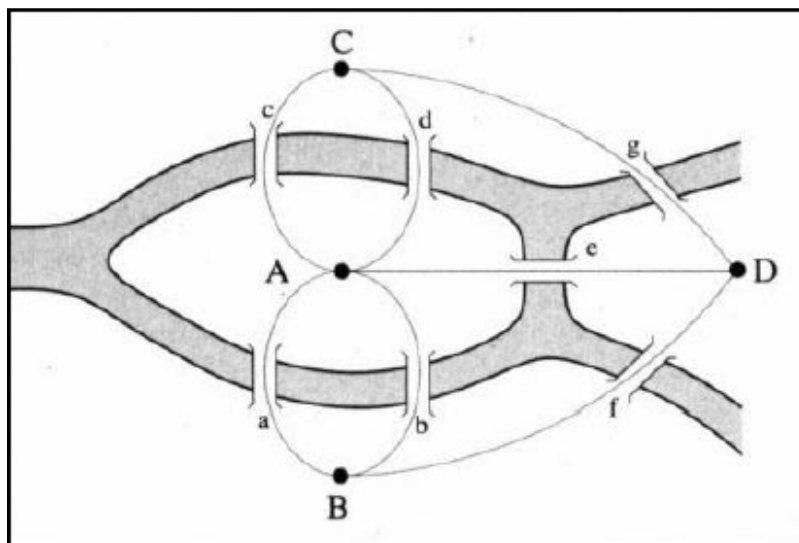
V této kapitole je nejprve zmíněn samotný vznik teorie grafů jako vědní disciplíny a jeho následný vývoj v čase. V další části je objasněno několik důležitých pojmů z oblasti teorie grafů, které jsou nezbytné pro vyhodnocení síťových analýz.

2.1 Historie teorie grafů

Počátek historie teorie grafů se datuje k roku 1736, kdy švýcarský matematik, fyzik a technik Leonhard Paul Euler (1707–1783) sepsal a publikoval krátký článek, ve kterém vyřešil zábavný problém „sedmi mostů“, kterým se zabývali lidé ve městě Královec (Königsberg) [14]. Královec byl v té době kvetoucí metropolí Východního Pruska ležící na břehu řeky Pregel. Prosperující hospodářství přispělo k postavení sedmi mostů, které spojovaly ostrov Kneiphof s ostatními částmi města. Obyvatelé rádi řešili různé hádanky a hlavolamy. Jedním z hlavolamů, který si lidé pokládali, byl tento: „*Je možné přejít všech sedm mostů a nejt po žádném z nich dvakrát?*“ [1]. Řešením královeckého problému bylo nelezení okružní trasy městem, která musela jedince převést přes každý most právě jednou a vrátit ho do výchozího místa [14]. Nikomu z obyvatel se ale tento hlavolam nepodařilo vyřešit. Že taková cesta neexistuje, potvrdil až Euler přesným matematickým důkazem. V uveřejněném článku nevědomky založil matematický obor zvaný teorie grafů. K této rozsáhlé a vyspělé disciplíně přispěla i řada významných matematiků [1].

V Eulerově důkazu byl převratný jeden důležitý krok, který se zapsal do historie. Jeho myšlenka byla taková, že si město Královec se sedmi mosty představil jako graf. Tento graf je reprezentován souborem vrcholů, které jsou navzájem spojeny hranami. Plánek situace královeckých mostů je znázorněn na obrázku 2. Ramena řeky Pregel svírají ostrov Kneiphof (A) a oblasti souše. Každá oblast souše (A až D) je nahrazena vrcholem a všechny mosty hranami (a až g). Výsledkem je graf se čtyřmi vrcholy a sedmi hranami. Při pozorování grafu je zřejmé, že vrcholy, ze kterých vychází lichý počet hran, musí být buď počátečním, nebo koncovým bodem okružní cesty. Souvislá trasa procházející všemi mosty smí mít pouze jeden počáteční i koncový bod. Existence takové trasy není možná v případě, že graf má více než dva vrcholy s lichým počtem hran. Graf města má takové vrcholy čtyři. Tak Euler dokázal, že na královeckém grafu neexistuje žádná cesta, která by procházela každou hranou právě jednou. Až po přečtení článku matematika Leonharda Eulera lidé ve městě Královec uznali, že hlavolam opravdu nemá řešení a postavili v roce 1875 ještě další most mezi

oblastmi B a C. Tímto krokem byl zvýšen počet hran u těchto vrcholů na čtyři. Lichý počet hran měly už jen vrcholy A a D. Nalezení požadované cesty už poté nebylo složité [1].



Obrázek 2: Plánek sedmi mostů ve městě Královec před rokem 1875

Zdroj: [1]

2.1.1 Postupný vývoj v čase

Po Eulerově inspirující práci přispěli k vývoji teorie grafů známí matematici, jako je například Augustin-Louis Cauchy, William Hamilton, Arthur Cayley, Gustav Kirchhoff a George Pólya, kteří se zasloužili o objev skoro všeho, co je dnes známo o velkých uspořádaných grafech [1]. Velice známá úloha teorie grafů, která se vyskytla v 19. století, byla otázka problému čtyř barev. Otázkou bylo, zda je možné každou mapu obarvit pomocí čtyř barev tak, aby sousední státy měly různé barvy. Každý stát si lze představit jako vrchol grafu a hranami spojené sousedící státy. Tento složitý problém byl kompletně vyřešen až roku 1976, kdy byla velká část dokázána pomocí počítače. Praktické využití podobného problému se na složitější úrovni řeší například v mobilních sítích, kde se využívají různé frekvence [14].

Významné poznatky v rozvoji teorie grafů jsou zaznamenány také v bývalé ČSR, kdy roku 1926 Otakar Borůvka (1899–1995) publikoval algoritmus pro nalezení minimální kostry. Úloha hledání minimální kostry popisuje, jak se musí spojit všechny vrcholy grafu hranami s nejnižší váhou (ohodnocením). Jiným algoritmem vyřešil stejný problém minimální kostry v roce 1930 i další český matematik Vojtěch Jarník (1897–1970). Jako praktický příklad lze uvést rozvod elektrické sítě mezi městy, tedy propojení měst co nejmenší délkou elektrického vedení [14].

Cíl teorie grafů byl až do poloviny 20. století jasný, šlo o objevení a katalogizování vlastností různých typů grafů. Příkladem může být hledání únikové cesty z bludiště či labyrintu, které bylo vyřešeno poprvé roku 1873. Až po uplynutí dvou století si vědci položili otázku: „*Jak grafy, či obecněji sítě vznikají?*“ [1]. Revoluci v teorii grafů provedli až dva maďarští matematici, a to Paul Erdős (1913–1996) s Alfrédem Rényim (1921–1970). Vydali osm nadčasových matematických článků zabývajících se stěžejní otázkou, jak sítě vznikají. Řešení bylo základem pro položení teorie náhodných sítí. Tvrdili, že pokud je síť příliš složitá a nedá se vystihnout jednoduchými prostředky, musí být považována za náhodnou. Tato teorie velice ovlivnila myšlení o sítích [1].

V polovině 90. let minulého století se hranice chápání teorie sítí významně posunula. Postarali se o to matematici Duncan Watts (1971) a Steven Henry Strogatz (1959). Oba pohlíželi na problém sítí až za hranicemi pohledu Erdöse s Rényim. Watts si položil jednoduchou otázku: „*Jaká je pravděpodobnost, že dva z mých přátel se znají navzájem?*“ [1]. Jelikož jsou vrcholy pospojovány náhodně, mají dva nejbližší přátelé pravděpodobnost se navzájem znát jako dva lidé žijící na druhém konci světa. Všichni lidé jsou součástí shluků, a aby bylo možné získat důkazy o tom, jak se společnost shlukuje, je nutné toto shlukování změřit. Proto byl zaveden tzv. koeficient shlukování⁵. Tento koeficient nabývá hodnot v rozmezí od nuly do jedné a říká, jak těsné jsou vazby určité skupiny. Vazby skupiny jsou silné, pokud se výsledné číslo blíží k jedné a naopak slabé při přibližování čísla k nule. Shlukování se neomezuje jen na sociální sítě. Shlukování je také charakteristickým rysem webu, pozorovat ho lze ve fyzickém propojení počítačů na internetu, ve spoluvlastnických vztazích firem, v potravních sítích ekosystému i v křehké síti molekul buňky. Díky objevu všudypřítomného shlukování se značně nabourala představa, že reálné sítě jsou víceméně náhodné [1].

2.2 Základní pojmy

Základem úvah o sítích je teorie grafů. Pro pochopení dalšího obsahu práce, tedy následného hodnocení síťových analýz sledované skupiny pomocí vybraných metrik, je nezbytné vysvětlení základních pojmů z oblasti teorie grafů.

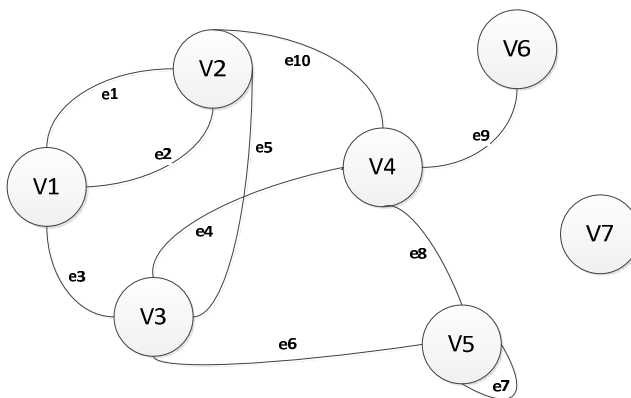
⁵ Koeficient shlukování je přeložen z anglického termínu clustering coefficient.

2.2.1 Graf

Obecný graf

Obecný graf G je v teorii grafů definován jako dvojice dvou množin či vrcholů, které se značí písmenem V a hran, které jsou označeny písmenem E . Někdy je místo dvojice možné mluvit o uspořádané dvojici (V, E) [14]. Obě zkratky V i E pocházejí z angličtiny. Vrcholům či uzlům se v anglickém jazyce říká vertices, nodes, entities, items a podobně. U vazeb neboli hran se lze v anglickém jazyce setkat s názvy typu edges, ties, connections či relationships [22]. Graf lze definovat jako dvojici prvků ve tvaru $G = (V, E)$, kde V je neprázdná množina vrcholů. E je množina neuspořádaných dvojic $\{u, v\}$, kde u, v jsou různé prvky množiny V . Prvky množiny E se nazývají hrany grafu [11].

Na obrázku 3 je znázorněn obecný graf. Jakékoliv dva vrcholy, které jsou spojeny hranou (spojnicí), se nazývají sousední vrcholy. Hrana, která vede z vrcholu u do téhož vrcholu u , se nazývá smyčka. Hrany, které spojují stejnou dvojici vrcholů, se nazývají rovnoběžné hrany (jde o násobnou vazbu). Zobrazení množiny hran do množiny vrcholů je označováno jako incidenční relace. Je-li hrana h v incidenční relaci s vrcholem u , říká se o vrcholu u , že s hranou h inciduje. Vrcholy, které neincidují s žádnou hranou, se nazývají izolované vrcholy (nevstupuje do nich ani nevystupuje žádná hrana) [12].

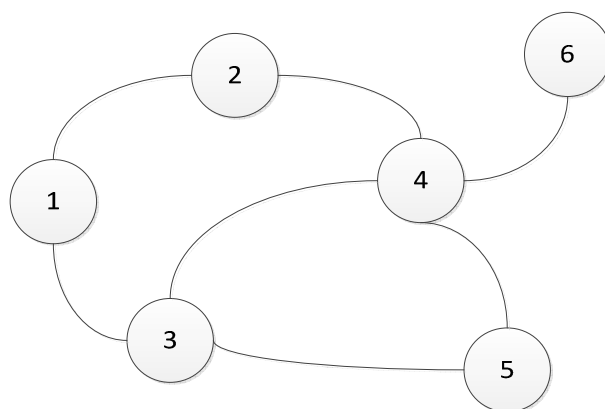


Obrázek 3: Příklad obecného grafu

Zdroj: upraveno podle [12]

Neorientovaný graf

Neorientovaný graf G je z pohledu teorie grafů definován jako uspořádaná dvojice prvků ve tvaru: $G = (V, E)$, kde $V = \{u_1, u_2, \dots, u_n\}$ je konečná množina vrcholů grafu a $E = \{\overline{u_i, u_j}, i, j = 1, 2, \dots, n\}$ je množina některých dvojic vrcholů, kterým se říká hrany grafu. Na obrázku 4 je znázorněna ukázka neorientovaného grafu [8].

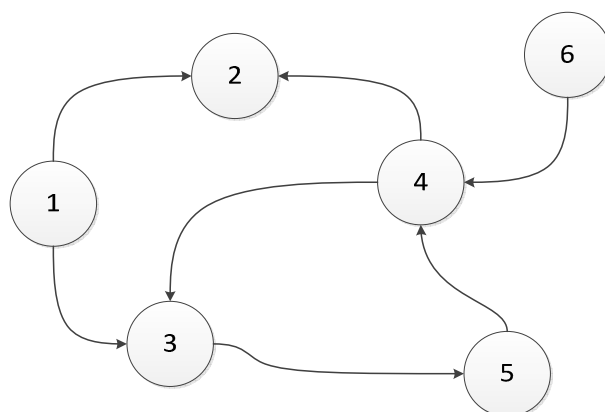


Obrázek 4: Příklad neorientovaného grafu

Zdroj: upraveno podle [8]

Orientovaný graf

Orientovaný graf G je z pohledu teorie grafů definován jako uspořádaná dvojice prvků ve tvaru: $G = (V, E)$, kde $V = \{u_1, u_2, \dots, u_n\}$ je konečná množina objektů, které se nazývají vrcholy grafu a $E = \{\overline{u_i, u_j}, i, j = 1, 2, \dots, n\}$ je množina některých uspořádaných dvojic vrcholů, kterým se říká orientované hrany grafu. Na obrázku 5 je znázorněn příklad orientovaného grafu. Orientace hrany je zde vyobrazena šipkou [8].



Obrázek 5: Příklad orientovaného grafu

Zdroj: upraveno podle [8]

V případě, že jsou v grafu hrany či vrcholy ohodnoceny čísly nebo jinými hodnotami, které mají vnitřní význam, jedná se o graf ohodnocený. Příkladem může být ohodnocení hran vzdáleností mezi městy. Graf, ve kterém mezi jakoukoli dvojicí vrcholů existuje více hran, se nazývá multigraf. Eulerovský graf je takový graf, který lze nakreslit pouze jedním tahem, žádná hrana se v něm tedy neopakuje [8]. Úplný graf je takový graf, ve kterém jsou každé dva vrcholy spojeny hranou [14]. Opakem úplného grafu je graf diskretní, který neobsahuje ani jednu hranu [8]. Bipartitní graf je graf, u kterého lze množinu vrcholů rozdělit na dvě

části, kdy z každého vrcholu jedné části jde hrana pouze do jednoho vrcholu druhé části a naopak. Pokud jde z každého vrcholu jedné části hrana do každého vrcholu druhé části, lze mluvit o úplném bipartitním grafu [14].

Podgraf

Podgraf je jednoduše řečeno část grafu, která má přesné vymezení [11]. Podgraf grafu G je takový graf H , který vznikl odebráním některých vrcholů a hran z původního grafu G . Při odebrání vrcholu je nutné vymazání všech hran vedoucích do nebo z tohoto vrcholu. Při odebrání jen těchto hran se podgraf nazývá indukovaný. Pokud došlo k odebrání i jiných hran, jde obecně o podgraf [14].

2.2.2 Stupeň vrcholu a cyklus v grafu

V grafu může být zajímavé, kolik hran vychází z vrcholů, nebo kolik má vrchol sousedů [8]. Proto je zaveden pojem **stupeň vrcholu**. Necht' G je graf a v je jeho vrchol. Symbolem $deg_G(v)$ je označen počet hran grafu G obsahujících vrchol v . Číslo $deg_G(v)$ je poté stupněm vrcholu v grafu G [14]. Nejvyšší stupeň v grafu G se značí $\Delta(G)$ a nejnižší $\delta(G)$ [11]. U orientovaných grafů lze rozlišit vstupní d^- (indegree) a výstupní d^+ (outdegree) stupeň. Vstupní stupeň udává počet hran, které jsou orientovány směrem do vrcholu. Výstupní stupeň říká, kolik hran je orientováno směrem z vrcholu ven [12].

Kružnice je každá uzavřená cesta v grafu, v orientovaném grafu respektuje orientaci hran a nazývá se **cyklus** [12]. Cyklem v grafu se rozumí posloupnost vrcholů a hran $(v_0, e_1, v_1, \dots, e_t, v_t = v_0)$, kde vrcholy v_0, \dots, v_{t-1} jsou navzájem různé vrcholy grafu G a pro každé $i = 1, 2, \dots, t$ je $e_i = \{v_{i-1}, v_i\} \in E(G)$. Nejkratší kružnicí je trojúhelník, kdy se jedná o úplný graf se třemi vrcholy. Za povšimnutí stojí podobnost definice cyklu s definicí cesty, kdy jde rovněž o posloupnost hran a vrcholů. V cyklu je ale na rozdíl od cesty první a poslední vrchol posloupnosti shodný. Prázdná posloupnost je povolena pouze v cestě s nulovou délkou. Acyklický graf je takový graf, který neobsahuje žádné cykly [14].

2.2.3 Cesta, sled a souvislost grafu

Cesta v grafu je chápána jako posloupnost vrcholů a hran $(v_0, e_1, v_1, \dots, e_t, v_t)$, kde vrcholy v_0, \dots, v_t jsou navzájem různé vrcholy grafu G a pro každé $i = 1, 2, \dots, t$ je $e_i = \{v_{i-1}, v_i\} \in E(G)$. Cesta délky rovna nule je touto definicí přípustná [14]. Cesta, která prochází skrz všechny vrcholy, se nazývá Hamiltonova cesta. Když tato cesta končí v tom samém vrcholu, ve kterém začala, jedná se o tzv. Hamiltonovu kružnici [8].

Posloupnosti vrcholů, kde pro vrcholy u_i a u_{i+1} existuje hrana, se říká **sled**. Pojem sled může být v jiné literatuře uváděn i jako procházka. Sledem lze libovolné vrcholy či hrany navštívit vícekrát [8].

V některých případech je nutné určit, jestli je graf „jedním celkem“. **Souvislost grafu** je definována následující definicí. Graf G je souvislý, jestliže pro každé jeho dva vrcholy x a y existuje v G cesta z x do y . Pokud není graf souvislý, tak se jeho souvislé části nazývají komponenty souvislosti [14].

2.2.4 Další pojmy

Isomorfismus

Při popisování určité situace pomocí grafů může dojít k případu, že dva stejné grafy budou mít různě označené vrcholy. Například jednou jsou vrcholy pojmenovány slovně a podruhé jsou očíslovány. Grafy tedy na první dojem působí, že jsou různé, ale přitom popisují stále stejnou situaci. V takovém případě lze říci, že dané dva grafy jsou isomorfní. Jestli jsou grafy isomorfní, je možné ověřit nalezením způsobu, pomocí kterého by se musely vrcholy jednoho grafu přejmenovat, aby odpovídaly vrcholům jiného grafu. Tomuto přejmenování se říká bijektivní zobrazení. Každý vrchol grafu se zobrazuje na právě jeden vrchol [14]. Pokud mezi grafy G a H existuje isomorfismus, zapíše se tento vztah symbolem $G \simeq H$ [11]. Dva grafy $G = (V, E)$ a $G' = (V', E')$ jsou isomorfní, jestliže existuje bijektivní (vzájemně jednoznačné) zobrazení $f: V \rightarrow V'$ tak, že platí: $\{x, y\} \in E$, právě když $\{f(x), f(y)\} \in E'$ [14].

Vzdálenost

V případě grafu s ohodnocenými hranami je možné určit vzdálenost mezi každými dvěma vrcholy grafu. Podmínkou je, že graf musí být souvislý. Funkce, která dosadí k vrcholům nejkratší vzdálenost, je tzv. metrika. Jednou z vlastností metriky je tzv. trojúhelníková nerovnost, která říká, že v grafu je získána vždy nejkratší vzdálenost. Nechť $G = (V, E)$ je souvislý graf. Pro vrcholy v, v' je definováno číslo $d_G(v, v')$ jako délka nejkratší cesty z v do v' v grafu G . Číslo $d_G(v, v')$ se nazývá vzdálenost vrcholů v a v' v grafu G [14].

Excentricita vrcholu značí nejdelší vzdálenost mezi vrcholy grafu. Průměrem se v grafu G značí největší excentricita vrcholů. Střed grafu naopak značí vrchol s nejmenší excentricitou. Pro nalezení nejkratší cesty mezi dvěma vrcholy se často používá známý tzv. Dijkstrův algoritmus. Tento algoritmus zaručuje výrazně rychlý výsledek hledání nejkratší cesty z jednoho vrcholu. Příkladem z praxe může být nalezení autobusového či vlakového spojení [11].

Strom

Strom je souvislý graf neobsahující kružnice (cykly). Kvůli neexistenci kružnic lze tvrdit, že mezi každými dvěma vrcholy existuje právě jedna cesta. Pro stromy platí Eulerův vzorec, kde jde o vztah mezi počty hran a vrcholů: $|V| = |E| + 1$ [14]. U stromu se rozlišuje jeho kořen, listy a mezilisty. Větev je cesta od kořene k listu. Les je neorientovaný graf, ve kterém jsou libovolné dva vrcholy spojeny nejvýše jednou cestou. Stromy jsou vhodné k popisu hierarchických struktur, schémat, rodokmenů či datových struktur [11]. Datové struktury jsou typické pro binární vyhledávací strom, kde se čísla ukládají pomocí grafu [14].

Matice sousednosti

Pro výpočty na počítači je nutné grafy zadat pomocí čísel, kde se využívá zápis pomocí matic. Postup vytváření matice sousednosti (incidenční matice) u neorientovaných grafů je popsán následovně. Prvním krokem je očíslování všech vrcholů v grafu. Počet všech vrcholů grafu určuje rozměr matice. Pokud mezi i -tým a j -tým vrcholem existuje hrana, pak se na ij -tou pozici napíše jednotka. Na pozici ij se napíše nula, pokud mezi i -tým a j -tým vrcholem hrana neexistuje. Tyto matice jsou symetrické podle hlavní diagonály. Pokud je graf G úplný, obsahuje matice A_G kromě hlavní diagonály pouze jednotky [8], [14]. Necht' $G = (V, E)$ je graf s n vrcholy, které jsou označeny v_1, \dots, v_n . Matice sousednosti grafu G je čtvercová matice $A_G = (a_{ij})_{i,j=1}^n$ definovaná předpisem (1) dle [14]:

$$a_{ij} = \begin{cases} 1 & \text{pro } \{v_i, v_j\} \in E \\ 0 & \text{jinak} \end{cases}, \quad (1)$$

kde: a_{ij} je matice;

v_i, v_j jsou vrcholy.

U orientovaných grafů není matice sousednosti symetrická podle hlavní diagonály. Neplatí, že hodnota prvku na souřadnicích například (1,2) je stejná jako hodnota na pozici (2,1) [14].

Matice vzdálenosti

Dalším důležitým vyjádřením vlastností grafů je matice vzdálenosti označována D_G . U této matice se napíše na pozici ij nula, pokud mezi i -tým a j -tým vrcholem neexistuje procházka. Pokud mezi i -tým a j -tým vrcholem procházka existuje, na ij -tou pozici se napíše počet hran, které obsahuje nejkratší cesta [8].

3 ANALÝZA VYBRANÉ SKUPINY

3.1 Analýza sociálních sítí

Analýza sociálních sítí (anglicky social network analysis) neboli síťová analýza, dále jen SNA, je metoda, která studuje sociální struktury subjektů, zejména však provádí rozbor sociálních vazeb jedinců v podobě tzv. relačních dat [23]. Subjektem nemusí být pouze jednotlivec, který má určité vztahy s jinými subjekty, ale může jím být i organizace, událost či kolektivní sociální entita [3]. Sociální síť je tvořena množinou vazeb, ve které jsou aktéři reprezentováni vrcholy sítě a vazby jsou určeny hranami, které vrcholy spojují. Analyticky je možné pohlížet na sociální síť prostřednictvím matematického vyjádření vlastností vazeb a vrcholů jako na graf, na kterém lze následně provést analýzu [23].

Vazby mohou být orientované nebo neorientované a jednosměrné nebo obousměrné. Stejně jako vazby lze rozdělit i typy sítí, a to na spojitě, nespojitě a egocentrické. Ve spojitě sítí lze projít od jednoho vrcholu k jinému bez ohledu na to, po kolika bodech se projde. V nespojitě grafu tato cesta není možná, protože se zde vyskytuje izolovaný bod (body). U egocentrické sítě existuje vrchol, se kterým jsou propojeny všechny ostatní vrcholy [22]. Reciprocita, intenzita a stálost jsou tři hlavní vlastnosti popisující síť. Často jsou tyto vlastnosti doplňovány o atributy velikosti, hustoty, centralizace, konektivity, zakotvenosti nebo dostupnosti [23]. Hustota sítě označuje počet uskutečněných vazeb ke všem možným vazbám a říká, jak moc nebo málo je síť sociálně propojena [22].

SNA poskytuje statistické nástroje ke zkoumání relačních dat. Zaměřuje se zejména na popisování vzorců a vztahů mezi subjekty a na jejich analýzu struktury [3]. Relační data neboli kontakty, vazby a spojení dvou jedinců jsou uspořádány do sociogramů nebo matic. Základní sociometrická matice v sobě obsahuje zjištění, s kým je každý jednotlivec z n aktérů spojen nebo nespojen vazbou [6]. K samotné analýze sociálních sítí se využívají počítačové programy jako je NodeXL, Ucinet, Pajek, Gephi, Tulip či Network Workbench [23].

Významné postavení si SNA vydobyla v oblastech, kde jsou důležitější vztahy než samotné osoby. Příkladem může být hledání vlivných osob ve skupině [22]. Sféra užití v praxi je velmi rozsáhlá, ať už se jedná o analýzu příbuzenských struktur, sociální mobility, kontaktů v deviantních skupinách, migračních vzorců, mezinárodní obchodní struktury, analýzy citačních indexů, práce tajné služby (teroristické sítě), vzorců šíření nemocí, chování jedinců na sociálních sítích atd. Kromě sociologie má také uplatnění v antropologii, ekonomii, geografii a lingvistice [20].

3.1.1 Vývoj analýzy sociálních sítí

Historii analýzy sociálních sítí lze dohledat již v myšlenkách G. Simmela či M. Webera, kteří pojem sociální síť užívali spíše jako sociální metaforu, která zpočátku neměla analytický koncept. Smysl jí dala až sociometrie, o kterou se zasloužil psychiatr a sociolog Jacob L. Moreno (1889–1974) [20]. Poprvé tento pojem použil již v roce 1916, ale první dílo s tímto tématem vyšlo až roku 1934 s názvem „Who Shall Survive“ [6]. Moreno se pokusil v malé sociální skupině zobrazit a popsat pomocí elementárního dotazování strukturu sociopreferečních vztahů [20]. Nejmenší analytická jednotka je podle Morena systém socioemocionálních vztahů tzv. sociální atom. Vyvinul sociometrický test, ve kterém využil analytický diagram tzv. sociogram zobrazující individua jako body a jejich vztahy jako čáry [6]. Definoval také pojem sociometrická hvězda [7].

Ve 30. a 40. letech minulého století použila na Harvardské univerzitě skupina vědců v čele s W. L. Warnerem a E. Mayem (1880–1949) sociogramy k popisu mezilidských vztahů v reálné pracovní situaci. Provedli výzkum, kde ve větším sociálním uskupení našli kliky, klastry a bloky. Klika je důvěrná nepříbuzenská skupina o 2 až 30 či více lidech. Třetím inspiračním zdrojem ve vývoji SNA v 50. a 60. letech minulého století byl Max Gluckman (1911–1975), který působil na Univerzitě v Manchesteru a tvrdil, že konflikt a moc hraje důležitou úlohu v udržování a transformaci sociální struktury [6].

Britská autorka E. Bootová prokázala existující korelaci mezi typem manželské dvojice a povahou sociálních sítí, do kterých je včleněna. Své myšlenky publikovala roku 1957 ve své práci „Family and Social network“ [7]. S postupným vývojem se zájem přesunul k popisu a analýze větších skupin. Kritérium, na základě kterého lze provést dekompozici sítí pomocí teorie grafů například dle koheze klik nebo strukturální ekvivalence rolí, je sociabilita aktérů, tedy povaha družných vztahů, vliv lokálních podmínek na utváření sítí s přihlédnutím k jejich hustotě a multiplexně⁶. Metodologie SNA, tak jak je v rozvinutější podobě užívána dnes, byla vypracována v 60. letech na Harvardské univerzitě. Pozorování sociální struktury bylo obohaceno matematickým modelováním za přispění Harrisona Whita, které se opíralo o multidimenzionální škálování. Druhým stimulem rozvoje byl návrat k teorii grafů [6].

Silným podnětem pro rozvoj analýzy sociálních sítí bylo také rozlišení vazeb na silné (kontakty s rodinou, blízkými přáteli, spolupracovníky) a slabé (známosti). Autorem známého pojmu „síla slabých vazeb“ je Mark Granovetter, který stál roku 1973 za vytvořením difúzního modelu šíření informací. Síla vazby je dána množstvím času, emocionální

⁶ Multiplexita představuje počet odlišných druhů směřovaných zdrojů.

intenzitou, mírou důvěry a recipročními službami mezi dvojicí ve vazbě. Možnost získání informace, která prochází řetězcem vazeb, závisí na pozici v sociální síti. Stále stejné informace obíhají kolem lidí, kteří jsou spojeni silnou vazbou. Slabé vazby naopak stojí za novými poznatky, které jsou zprostředkovány mezi větším počtem lidí [6]. Granovetter doložil důležitost známostí v dnešní moderní době, které jsou při větším množství sociálně důležitější než existence několika málo intenzivně prožívaných kontaktů [20]. SNA je rozsáhlá sada metod a je nepravděpodobné, že by se mohla vyvinout v novou teorii sociální struktury [6].

3.2 Vybrané metriky pro měření sociálních sítí

V této části jsou zmíněny a blíže popsány vybrané metriky pro měření sociálních sítí, které jsou následně využity v síťových analýzách.

Degree centrality

Degree centrality neboli centralita měřená stupněm vrcholu určuje počet přímých vazeb vrcholu k dalším vrcholům a měří tak aktivitu vrcholů v síti. Pojmem „spojky“ nebo „středy“ se označují vrcholy, které mají vysokou hodnotu degree centrality. V grafu je lze rozpoznat podle tvaru hvězdy, vede k nim mnoho vazeb od ostatních vrcholů v síti. U orientovaných grafů lze rozlišit vstupní „Indegree“ a výstupní „Outdegree“ stupeň centrality, který vyjadřuje počet vazeb do vrcholu vstupujících a z vrcholu vycházejících [22]. Vrchol s maximálním možným počtem vazeb s dalšími vrcholy má nejvyšší stupeň vrcholu [3]. Vrchol se stupněm $n-1$ by měl být přímo napojen na ostatní vrcholy, a tedy zcela centrální k síti. Vrchol, který je spojen pouze s dalšími dvěma vrcholy, by měl být méně centrální. Výsledná hodnota se pohybuje v rozmezí od nuly k jedné a říká, jak je vrchol dobře připojen, pokud jde o přímé spojení. Stupeň centrality vrcholu i je vyjádřen podílem (2) dle [13]:

$$\frac{d_i(g)}{n-1} \quad (2)$$

kde: $d_i(g)$ je stupeň vrcholu i ;

n je počet vrcholů.

Closeness centrality

Closeness centrality je centralita, která měří blízkost polohy ke středu. Tato hodnota je nejvyšší, pokud lze z vrcholu dosáhnout ke všem ostatním vrcholům přímou nezprostředkovanou vazbou. Matematicky odpovídá nejmenší hodnotě součtu vzdáleností k dalším vrcholům. Vrcholy snadno přijímají a přenášejí informace. Velký vliv v síti mají vrcholy s vysokou mírou blízkosti ke středu, jelikož mají nejrychlejší přístup k celé síti. V takových místech dochází nejčastěji k napadení sítě, jelikož mají značný vliv na to, co se v síti odehrává [22]. Spočítat ji lze jako inverzní průměrná vzdálenost mezi vrcholem i a jiným vrcholem j podílem (3) dle [13]:

$$\frac{n - 1}{\sum_{j \neq i} l(i, j)} , \quad (3)$$

kde: n je počet vrcholů;

$l(i, j)$ je počet vazeb v nejkratší cestě mezi vrcholy i a j .

Betweenness centrality

Betweenness centrality neboli centralita měřená středovou mezípolohou je pro určitý vrchol nejvyšší, když cesty mezi libovolnými dvojicemi vrcholů sítě vždy procházejí tímto vrcholem [22]. Středová mezípoloha tedy měří, kolik cest mezi dvěma vrcholy prochází konkrétním vrcholem [3]. Vrcholy s vysokou hodnotou jsou označovány jako mosty (bridges) nebo zprostředkovatelé (brokers). Tok informací v síti je kontrolován body, které mají vysokou hodnotu a umožňují tak dobrou viditelnost všeho důležitého, co se v síti odehrává. Někdy ale mohou být závorou, která viditelnost ztěžuje. Vrcholy s vysokou hodnotou jsou nazývány jako úzká hrdla sítě, která jsou pro síť velmi důležitá, protože brání rozpadu sítě, ale zároveň představují velké riziko [22]. U neorientované sítě je mezípoloha vrcholu x_i , ve které existuje nejvýše jedna geodetická (nejkratší) cesta mezi jakoukoli dvojicí vrcholů, definována počtem cest, které procházejí vrcholem i vztahem (4) dle [17]:

$$x_i = \sum_{st} n_{st}^i , \quad (4)$$

kde n_{st}^i je rovno jedné, když vrchol i leží na nejkratší cestě z s do t , rovno nule je v případě, že tomu tak není nebo cesta vůbec neexistuje.

Eigenvector centrality

Eigenvector centrality je metrika podobná centralitě měřené stupněm vrcholu neboli degree centralitě, navíc ale zohledňuje důležitost vrcholu v síti. Jinak řečeno říká, jak je každý vrchol v síti významný. Podstatou míry centrality eigenvectoru je, že důležitost a hodnota vrcholu roste s počtem vazeb, které má daný vrchol s ostatními důležitými vrcholy v celé síti. Podobná centralita eigenvectoru je Page Rank, která byla uvedena společností Google. Při této metrice je ohodnocen každý vrchol hodnotou, která je rovna součtu hodnot jeho sousedních vrcholů. Míra eigenvector centrality x_i^j je tedy definována jako součet centralit *i-tých* vrcholů, které spolu sousedí vztahem (5) dle [17]:

$$x_i^j = \sum_j A_{ij}x_j \quad , \quad (5)$$

kde: A_{ij} je matice sousednosti;

x_j je vrchol.

3.3 Výběr softwaru pro síťovou analýzu

K analýze sociálních sítí se využívají aplikace, kterých je nepřeberné množství. Záleží na požadavcích a možnostech uživatele. Liší se zejména poskytovanými funkcionalitami, vstupními a výstupními formáty, platformou či licencí. Některé aplikace jsou zcela zdarma, ale mnohé z nich jsou placené. Pro příklad lze uvést již zmíněné softwary jako je NodeXL, Ucinet, Pajek či Gephi. Pro síťovou analýzu vybrané skupiny je v této práci použit nástroj NodeXL, který má několik výhod a práce s ním je dostačující vzhledem k jeho analytickým možnostem.

Network Overview, Discovery and Exploration for Excel neboli NodeXL je volně stažitelnou open source šablonou pro Microsoft Excel šířitelnou pod všeobecně veřejnou licencí GPL. Je to projekt od společnosti Social Media Research Foundation a kompatibilní je s verzemi MS Excel 2007 a 2010. Je to jakási předdefinovaná knihovna tříd, která je postavená na technologii .NET⁷. Autorem tohoto projektu je americký sociolog Marc Smith, který se přímo specializuje na výzkum sociálních sítí. Dříve, než přišel s projektem NodeXL, pracoval na projektu .NetMap, který byl podle všeho průkopnickou verzí popisovaného softwaru. NodeXL zahrnuje široké spektrum nástrojů pro analýzu a vizualizaci dat. Umožňuje tedy snadné vytváření a prozkoumávání síťových grafů ve známém prostředí

⁷ .NET je soubor technologií v softwarových produktech tvořící celou platformu, pochází z anglického dot network.

aplikace MS Excel. Šablona zobrazuje grafy pomocí vlastního systému Windows Presentation Foundation (WPF), které lze znovu použít ve vlastních aplikacích [18].

NodeXL má několik výhod, které uživatelům usnadňují práci. Umožňuje import/export grafů z/do aplikací nebo formátu matic. Podporuje přímé připojení k sociálním sítím a následné importování dat přímo z webů, jako jsou YouTube, Twitter a Flickr. Po stažení a instalaci doplňku lze importovat data i ze sociální sítě Facebook. Umožňuje také import dat z lokálních poštovních účtů. Uživatel se při vizualizaci grafu může přiblížit k oblasti zájmu a podle potřeb si ji zvětšit. Má flexibilní dispozice, které urychlují práci. U výsledných grafů lze snadno měnit jejich vzhled, a to barvu, tvar, velikost, průhlednost nebo přidávat obrázky k vrcholům. Výsledné analýzy lze snadno filtrovat podle zadaných parametrů. Umožňuje seskupovat vrcholy do skupin a následně je vzhledově upravit. NodeXL provádí metrické výpočty jako je degree, betweenness centrality, closeness centrality, eigenvector centrality, koeficient shlukování, hustotu a jiné [18].

Software NodeXL byl pro síťovou analýzu zvolen z důvodu jeho uživatelské dostupnosti, relativní jednoduchosti, grafických výstupů a přívětivosti šablony. Práce se samotnou šablonou by měla být pro běžné uživatele kancelářských balíků MS Office nebo konkrétně aplikace MS Excel bezproblémová, jelikož pracují ve známém prostředí, které jim navíc může usnadnit hromadnou úpravu či filtraci dat. Je to aktivní projekt, který se neustále vyvíjí. Autoři přicházejí s novými vylepšenými verzemi a doplňky. Tato práce je vypracována v současně nejnovější verzi 1.0.1.229. Dalším důvodem je podpora importu dat z online sociálních sítí, který usnadní práci se získáním a předzpracováním vstupních dat.

Za nevýhodu při rozhodování výběru tohoto softwaru by se dala považovat nutnost vlastnit aplikaci MS Excel 2007 nebo 2010 a omezení počtu řádků této aplikace, kterých je něco přes jeden milion. V případě takto velkého počtu vstupních dat by se tento problém dal vyřešit například rozmístěním dat do více listů. Žádná z těchto uvedených nevýhod není komplikací pro použití tohoto softwaru v této práci, jelikož má autorka aplikaci k dispozici a počet vstupních dat není v tak velkém rozsahu.

3.4 Výběr vhodné skupiny

Zvolené softwarové prostředí napomůže nalézt a identifikovat skryté vlastnosti uvnitř skupiny využívající sociální síť. Na vybrané skupině jedinců je provedena analýza pomocí nástrojů, které se využívají pro měření sociálních sítí. Analyticky lze na sociální síť pohlížet prostřednictvím vlastností vazeb a vrcholů jako na graf, na kterém lze následně provést analýzu. Obrazový výstup má v mnohých případech lepší vypovídající hodnotu než pouhý textový popis dat. Na sociální síť je tedy pohlíženo jako na množinu vazeb, ve které jsou uživatelé sítě reprezentováni vrcholy a vazby mezi nimi jsou znázorněny pomocí hran, které tyto vrcholy spojují. Jsou tu využity principy z teorie grafů, které jsou blíže popsány v kapitole 2.2. Vztahy mezi jedinci jsou vyobrazeny pomocí grafu, který je výsledkem zvolené analýzy a jednotlivé výstupy jsou popsány a vysvětleny.

Výběr vhodné skupiny jedinců pro síťovou analýzu proběhl zcela náhodně při tvorbě této práce. Absence osobní znalosti skupiny není pro zkoumání žádnou překážkou. Analyzována je sociální struktura na síti Twitter kolem uživatele „josefslerka“. Josef Šlerka je v reálném životě šéfem vývoje a výzkumu ve společnosti Ataxo Interactive a zároveň vede Studia nových médií na Filozofické fakultě Univerzity Karlovy v Praze. Mimo jiné se podílí na spoustě dalších projektů a je členem několika skupin a asociací. Je také autorem několika článků na internetu zabývajících se analýzou sociálních sítí. Základní charakteristika sociální sítě Twitter je popsána v kapitole 1.3.3.

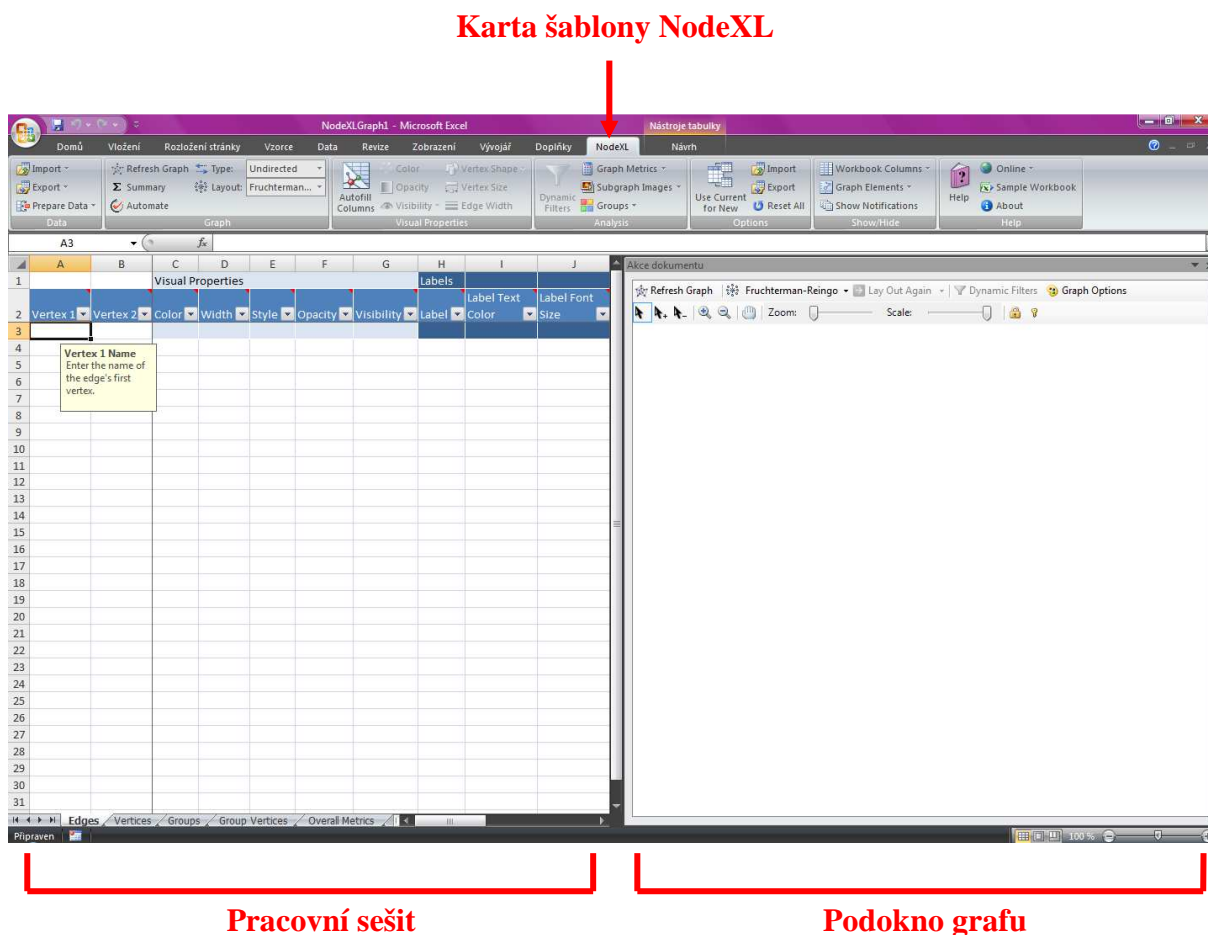
Pozorováním uživatelského účtu na Twitteru bylo zjištěno, že v době psaní této práce má „josefslerka“ celkem 4 873 následovníků (followers) neboli odběratelů jeho příspěvků. On sám následuje, a tedy odebírá příspěvky od 920 uživatelů (following). Napsaných krátkých vzkazů tzv. „tweets“ má celkem 13 947, kterým se v českém jazyce říká tweety. Tyto údaje jsou pouze orientační, protože se jejich počet zvětšuje každým dnem. Lze konstatovat, že vstupních dat pro síťovou analýzu je dostatečné množství k objektivnímu vyhodnocení výsledných grafů. Veškerá dostupná data jsou ze sociální sítě importována do nástroje NodeXL, který tuto funkci nabízí.

V případě Twitteru jsou uživatelé vrcholy a jejich vzájemné následování reprezentují vazby sítě. Vztahy následujících a následovaných kolem uživatelského účtu se zahrnutou úrovní 1, tvoří egocentrický graf. Spojitý graf může vzniknout například při vizualizaci vazeb mezi lidmi hovořícími o nějaké značce či tématu v případě, když o něm mluví celá skupina, pokud tomu tak není, jedná se o graf nespojitý.

3.5 Popis postupu řešení

Všechny síťové analýzy prováděné v této práci mají stejný postup řešení. Součástí této části je i obecný popis pracovního prostředí softwaru NodeXL, který byl vybrán pro potřeby této práce. Nezbytným prvním krokem je stažení nástroje z oficiálních stránek na adrese „<http://nodexl.codeplex.com/>“ a jeho následná instalace. Po otevření pracovního prostředí se šablonou NodeXL v MS Excel se vytvoří nový sešit, který je odlišný od standardních sešitů aplikace MS Excel v tom, že obsahuje více přizpůsobených listů a funkcionalit k tvorbě síťových grafů.

Na obrázku 6 je znázorněno pracovní prostředí nástroje NodeXL, které je rozděleno do tří částí. Vyobrazena je zde karta šablony, pracovní sešit a podokno grafu. Každá z těchto částí je blíže popsána v následujících odstavcích pod obrázkem.



Obrázek 6: Pracovní prostředí nástroje NodeXL

Zdroj: vlastní zpracování

Karta šablony NodeXL, která přibyla na pásu karet aplikace Excel, obsahuje specifické funkcionality potřebné pro síťové analýzy. Všechny příkazy jsou organizovány do skupin se společnými funkcemi. První skupina „Data“ umožňuje import, export a přípravu vstupních i výstupních dat pro síťovou analýzu v několika dostupných formátech. Další skupina „Graph“ nabízí rychlé funkce pro úpravu grafu, a to obnovu aktuálního grafu, změnu typu grafu na orientovaný nebo neorientovaný, změnu jeho celkové dispozice podle předdefinovaných algoritmů či tvarů a různá shrnutí. Vizuelní vlastnosti „Visual Properties“ obsahují položku „Autofill Columns“, která umožňuje vypočítat a vyplnit grafy vizuelními vlastnostmi sloupců, jako je šířka hran a tvar vrcholů, pomocí hodnot z dalších sloupců či vypočítaných metrik. Tato volba je v této práci využita při každé analýze, proto je ještě blíže popsána dále. Tato skupina nabízí i rychlou změnu barev, neprůhlednosti, viditelnosti, tvaru i velikosti vrcholů a šířky hran. Další skupina funkcí síťové analýzy „Analysis“ má položky dynamického filtru, výpočet metrik grafu a nabídku funkcí pro práci se skupinami či podgrafy. Skupina „Options“ nabízí možnosti k importu a exportu dodatečných aktuálních dat. Zobrazit či skrýt sloupce sešitu, elementy grafu nebo různá oznámení obsahuje skupina „Show/Hide“. Nechybí ani nápověda k programu „Help“.

V levé části pracovního prostředí se nachází pracovní sešit, do kterého se vkládají vstupní data. Obsahuje celkem pět předdefinovaných listů. Do listu „Edges“ neboli hrany se zapisují vždy dva vrcholy, které jsou spojeny vazbou a je zde možné doplnit i jejich atributy. Další list pro vrcholy „Vertices“ obsahuje prostor pro ruční doplnění izolovaných vrcholů v grafu. Následující dva listy „Groups“ a „Group Vertices“ umožňují rozdělení síťového grafu do skupin, které napomáhají seskupit a odlišit vrcholy z různých pohledů. Poslední list „Overall Metrics“ obsahuje výpočty celkových metrik. Pro potřeby síťových analýz v této práci je postačující první list z nabídky a to hrany.

Podokno pro graf umístěné v pravé části obrazovky slouží k samotnému zobrazení síťového grafu, který se aktualizuje vždy po stisknutí tlačítka „Refresh Graph“. V této části pracovního prostředí lze také měnit tvary a uspořádání zobrazovaných grafů podle předdefinovaných algoritmů nebo geometrických útvarů. K dispozici je algoritmus „Fruchterman-Reingold“ a „Harel-Koren Fast Multiscale“. Dále je možné změnit dispozici grafu do kruhu, spirály, horizontální či vertikální sinusoidy, mřížky, do náhodného tvaru a dalších. Lze se přiblížit a následně pohybovat v grafu, vybírat jednotlivé vrcholy nebo hrany a provádět s nimi další operace. Je možné změnit celé měřítko grafu, kdy dojde ke zmenšení či zvětšení všech jeho elementů. V nabídce jsou i dynamické filtry „Dynamic Filters“, které uživatelům umožňují filtrovat data pomocí zadaných parametrů. Nechybí ani „Graph

Options“ tedy možnosti grafu, kde lze změnit různé vlastnosti vrcholů, vazeb a pozadí. Konkrétně u vazeb lze měnit jejich barvu, šířku, velikost šipek u orientovaných grafů, průhlednost, zakřivení a lze změnit i barvu u aktuálně vybraných vazeb. U vrcholů je možné upravit jejich barvu, tvar, velikost, průhlednost, efekty i obrázky vybraných vrcholů. V podokně grafu se může změnit i barva pozadí, umožněno je také vložení obrázku na pozadí grafu.

Dalším krokem po otevření nástroje NodeXL je importování analyzovaných dat z online sociální sítě do pracovního sešitu šablony za pomoci ikony „Import“ v kartě NodeXL. Po kliknutí na ikonu se zobrazí několik možností pro import vstupních dat. Jak už bylo zmíněno, importovat data lze z několika sociálních sítí jako je Flickr, Twitter, YouTube a po stažení doplňku i ze sítě Facebook. Pro potřeby této práce je využita volba pro získávání dat ze sociální sítě Twitter kolem konkrétního uživatele, tedy import „From Twitter User’s Network“. Po výběru této možnosti se otevře tabulka, ve které je nutné doplnit informace o importovaných datech. V první řadě se musí doplnit uživatelské jméno osoby, o které je potřeba získat požadovaná data. Poté následuje výběr osob, které budou reprezentovat vrcholy v grafu. Mohou být importována data o následovnících uživatele, data o následujících uživatelem nebo obě dvě možnosti najednou. Po této volbě následuje výběr hran, které spojují jednotlivé vrcholy. Hrany mohou reprezentovat vztahy následující či následovaný, odpovědi na poslední tweety nebo zmínky v posledních tweetech. Lze zde také nastavit zahrnované úrovně v grafu a je možné omezit počet importovaných uživatelů sítě. Poslední položkou je informace, zda osoba, která pracuje s NodeXL, má založený účet na Twitteru a autorizovaný software. Při splnění těchto dvou náležitostí je umožněn rychlejší import vstupních dat.

Po úspěšném načtení vstupních dat ze sociální sítě lze využít jakékoli nástroje šablony. Pomocí tlačítka „Graph Metrics“ se dají vybrat jednotlivé metriky, které je možné ze vstupních dat vypočítat. Na výběr jich je hned několik: „degree, indegree, outdegree, betweenness centrality, closeness centrality, eigenvector, page rank, clustering coefficient, edge reciprocation, group metrics, overall graph metrics a twitter search network top items“. Pro potřeby této práce jsou využity jen metriky jako je „indegree, outdegree, betweenness, closeness a eigenvector centralities“. Dále je možné nastavit a vybrat automatický běh několika činností pomocí tlačítka „Automate“ jako je spojení duplicitních hran, seskupení shluků, výpočet metrik, vyplňování sloupců, zobrazení podgrafů, vizualizace grafu v pravé části pracovního prostředí a jiné. Ještě není nutné zaškrtnout uložení a export grafu, jelikož je to předčasné vzhledem k jeho dalším úpravám do požadovaného grafického výstupu.

V dalším kroku je nutné zvolit položku „Autofill Columns“, kde je možné nastavit zobrazení ohodnocení hran, barvu a šířku hran podle určitých kritérií, jejich neprůhlednost, viditelnost a je možné opatřit je i popiskem. Hrany lze pomocí nástrojů šablony i zaoblit. U vrcholů je možnost nastavování atributů mnohem větší. Lze měnit jejich barvu, tvar, velikost, neprůhlednost a viditelnost podle různých naměřených metrik či popisek s libovolnou barevnou výplní včetně souřadnic umístění elementů v grafu. Jednoduché tvary reprezentující vrcholy grafu lze nahradit například i reálnými obrázky. Nechybějí zde ani možnosti pro skupiny.

Při zobrazení grafu je nutné zvolit, zda jde o graf orientovaný nebo neorientovaný. Pro obrazovou reprezentaci grafu je také nutné vybrat vhodný algoritmus či předdefinovaný tvar, aby byl výsledný graf co nejvíce přehledný. I přes tato opatření je pravděpodobné, že při velkém obsahu vstupních dat bude vzniklý graf velice nepřehledný a přeplněný vrcholy i hranami. V tomto případě je možné využít další nástroj šablony a pomocí dynamického filtru zobrazit hrany a vrcholy podle zadaných parametrů.

3.6 Analýza získaných údajů

V této části jsou na vybrané skupině uživatelů sítě Twitter provedeny síťové analýzy pomocí nástrojů, které se využívají pro měření sociálních sítí. Analyzována je pouze část skupiny kolem uživatele „josefslerka“, která je blíže popsána v kapitole 3.4. Při takto relativně velkém množství vstupních dat je v některých případech nutné při jejich získávání omezovat zahrnuté úrovně sítě i limit uživatelů. Z důvodu omezeného množství stáhnutých dat během jedné hodiny je import se zahrnutou úrovní 1,5 a 2 časově velmi náročný. Za použití nástroje NodeXL jsou vytvořeny síťové grafy, na kterých jsou následně hledány vlastnosti a struktura vzájemných vztahů pomocí vybraných metrik a poznatků z teorie grafů. Výsledné grafy, na kterých jsou popsány a vyhodnoceny jednotlivé závislosti v konkrétních analýzách, jsou součástí této práce.

Nedostatečná znalost cílové skupiny či analýza pouze určité části rozsáhlé sociální sítě není důvodem k obavám o relevantnosti vstupních a výstupních dat. Cílem práce je ukázat využití základních metrik na vybrané skupině osob a je možné se domnívat, že při analýze jiné či větší sociální sítě by byly nalezeny stejné zákonitosti a tedy i podobné výsledky.

3.6.1 Analýza následovníků uživatele

Tato analýza se zabývá nalezením sítě následovníků kolem uživatele „josefslerka“. V době psaní této práce odebírá krátké vzkazy od tohoto uživatele celkem 4 873 uživatelů sociální sítě Twitter. Toto číslo každým dnem roste a pro síťovou analýzu není prvořadé. Při importu následovníků je nutné nastavit některá omezení. V případě importu všech jedinců odebírajících příspěvky od konkrétního uživatele v jedné zahrnuté úrovni je výsledný graf egocentrický a veškeré vazby vedou do středu k jedinci Josefu Šlerkovi. S tímto typem grafu nelze dopočítat potřebné metriky, které se využívají pro měření sociálních sítí. Proto je nutné zahrnout minimálně úroveň 1,5, aby se mezi sebou propojily i ostatní vrcholy se společnými vazbami. Poté je možné rozkrývat struktury uvnitř vybrané skupiny. Dále je nutné nastavit omezený počet stahovaných osob. Při větších hodnotách nastává problém s limitovanou sazbou stahovaných dat v rozmezí jedné hodiny. Proto je limit při této analýze stanoven na 500 náhodně vybraných uživatelů.

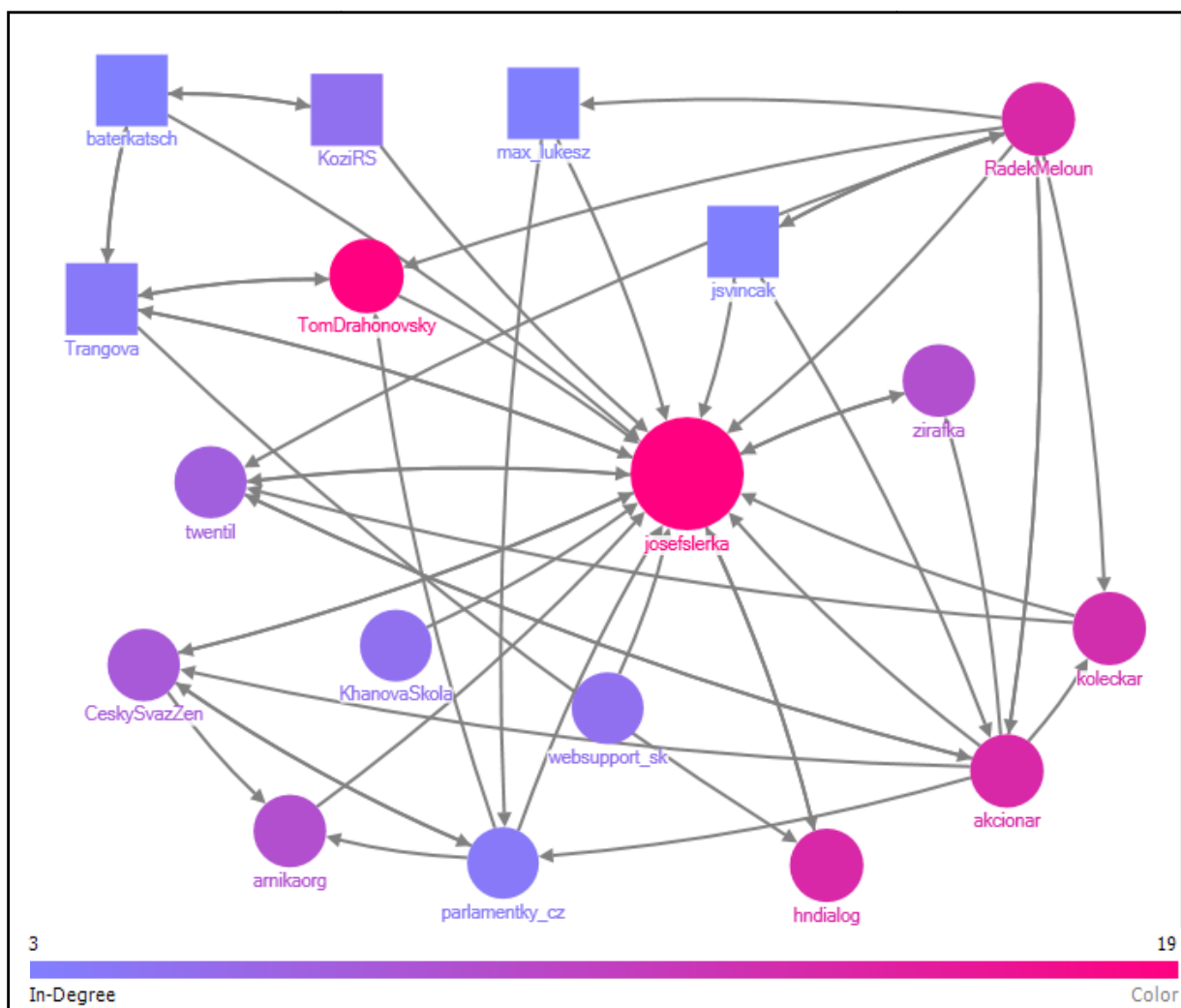
Skutečnost, že síťový graf neobsahuje veškerá dostupná data, není překážkou k provedení analýzy. Zkoumaná je pouze určitá část sociální sítě, na které jsou hledány tytéž zákonitosti vybraných metrik.

Tabulka 1: Metriky analýzy následovníků uživatele

	Vrchol	Velikost vrcholu	Indegree	Outdegree	Betweenness centrality	Closeness centrality	Eigenvector centrality
1	josefslerka	200	300	16	87356,62143	0,003333	0,050675
2	RadekMeloun	84,4	14	19	153,635714	0,00173	0,008334
3	akcionar	84,4	14	30	378,335714	0,001757	0,009744
4	zirafka	82,4	9	2	27,266667	0,001695	0,004974
5	TomDrahonovsky	86,5	19	3	146,466667	0,001724	0,006911
6	pavkahu	80	3	3	4	0,001681	0,004209
7	CeskySvazZen	82	8	5	25,733333	0,001695	0,005048
8	arnikaorg	82,4	9	2	29,333333	0,001695	0,004578
9	Trangova	80,4	4	5	5,5	0,001681	0,003716
10	max_lukesz	80	3	5	7,833333	0,001684	0,003995
11	websupport_sk	80,8	5	3	16,333333	0,001689	0,004141
12	parlamentky_cz	80,4	4	5	15,033333	0,001689	0,004726
13	KhanovaSkola	80,8	5	6	20,666667	0,001689	0,00404
14	twentil	81,6	7	3	11,333333	0,001686	0,004551
15	hndialog	84,4	14	3	99,9	0,001712	0,005685
16	baterkatsch	80	3	4	3	0,001678	0,003251
17	KoziRS	80,8	5	2	9,666667	0,001684	0,003635
18	koleckar	84	13	4	80,733333	0,001712	0,00644
19	jsvincak	80	3	4	4,333333	0,001681	0,004059

Zdroj: vlastní zpracování

Z celkového počtu 622 importovaných hran, které představují vazby mezi jednotlivými vrcholy je nutné na grafu síťové analýzy uplatnit dynamický filtr pro lepší přehlednost a hledání konkrétních struktur v síti. Filtr je nastaven na hodnoty přímých vazeb do vrcholu vstupujících tedy $\text{indegree} > 3$ a vazeb z vrcholu vystupujících $\text{outdegree} > 2$. Po použití těchto filtrů dojde k vizualizaci menšího počtu vrcholů, které jsou předmětem této analýzy. V tabulce 1 jsou uvedeny výsledné hodnoty vypočítaných metrik analýzy následovníků konkrétního uživatele. Jsou zde uvedeny i hodnoty velikostí vrcholu odpovídající hodnotám indegree . Tyto vyfiltrované hodnoty jsou velice podobné, jelikož jsou vypočítány z celkového počtu importovaných dat. Proto je zde využita i barevná škála rozlišující vrcholy podle hodnot indegree v rozmezí od 3 do 19 pro lepší vypovídající hodnotu. Legenda této barevné stupnice je znázorněna pod síťovým grafem. Tvar vrcholů je určen podle hodnot metriky $\text{betweenness centrality}$, která měří, kolik cest mezi dvojicí vrcholů prochází daným vrcholem. Při hodnotách této metriky menší nebo rovno deseti je tvar vrcholu čtvercový, jinak kruhový.



Obrázek 7: Graf síťové analýzy následovníků uživatele

Zdroj: vlastní zpracování

V této části náhodně vybraných uživatelů sítě je za nejvýznamnějšího považován „josefslerka“, což logicky vyplývá z analýzy hledání následovníků tohoto uživatele. Hodnota jeho eigenvector centrality je nejvyšší a dosahuje čísla 0,050675. Dalším důležitým vrcholem po Josefu Šlerkovi je „akcionar“ s hodnotou 0,009744. Třetím je „RadekMeloun“ jehož hodnota této metriky je 0,008334. Důležitost a hodnota vrcholu roste s počtem vazeb, které má daný vrchol s ostatními důležitými vrcholy v celé síti. Na uvedeném grafu síťové analýzy následovníků uživatele vyobrazeného na obrázku 7 je viditelná vazba uživatele „akcionar“ směrem k „josefslerka“. Uživatel „RadekMeloun“ taktéž následuje uživatele „josefslerka“, ale je následovníkem i druhého nejvýznamnějšího jedince „akcionar“. Nejmenší hodnotu eigenvector centrality má „baterkatsch“ a je tak nejméně důležitý článek této konkrétní sítě. Důvodem je jeho malá provázanost s dalšími důležitými uživateli sociální sítě.

Velký vliv na dění v síti má uživatel „josefslerka“ vzhledem k jeho hodnotě metriky closeness centrality 0,003333. Tato skutečnost potvrdila fakt, že následovníci odebírající jeho krátké vzkazy jsou tímto uživatelem ovlivňováni. Z jeho vrcholu lze dosáhnout ke všem dalším přímou nezprostředkovanou vazbou a je tak ideální pro efektivní přenášení informací v síti. V případě napadení sítě je nejzranitelnější právě uživatel „josefslerka“. Nejmenší vliv na to, co se v síti odehrává má již zmíněný „baterkatsch“. To potvrzuje i jeho nejnížší hodnota středové mezipolohy neboli betweenness centrality. Tato metrika byla použita pro zvýraznění tvaru vrcholů v síti, a tak je vrchol tohoto uživatele v grafu reprezentován čtvercem, jelikož je jeho hodnota metriky menší nebo rovna deseti.

Barevné zvýraznění jednotlivých vrcholů je určeno počtem vstupujících vazeb do vrcholu neboli metrikou indegree. Nejmenší hodnoty jsou zvýrazněny modrou barvou, která dále postupuje přes fialovou až k tmavě růžové, a tedy k větším číslům. Analýza opět potvrzuje skutečnost, že nejvíce vazeb vstupuje do vrcholu „josefslerka“ a je tak zvýrazněn tmavě růžovou barvou. Modrými vrcholy s nejnížší hodnotou jsou uživatelé „pavkahu“, „max_lukesz“, „baterkatsch“, „jsvincak“. V této části importované sítě má největší hodnotu outdegree uživatel „akcionar“, z jehož vrcholu vychází nejvíce vazeb k dalším vrcholům v síti, a který je zároveň druhým nejdůležitějším vrcholem sítě. Po něm následuje uživatel „RadekMeloun“ a až na třetím místě je „josefslerka“.

3.6.2 Analýza následujících uživatelem

Předmětem této síťové analýzy je vyhodnotit vztahy v náhodné skupině uživatelů, od kterých odebrá příspěvky „josefslerka“. V tomto případě je právě on následovníkem ostatních uživatelů sociální sítě Twitter, kterých v době psaní této práce následuje 920. I tady se není nutné tohoto proměnného čísla striktně držet. V analýze je nutné jako u předešlého případu nastavit konkrétní omezení. Při vizualizaci vstupních dat je totiž výsledkem opět egocentrický graf, ale orientovaný je v tomto případě ve směru od Josefa Šlerky k ostatním uživatelům, které sám následuje. Pro potřeby této analýzy je nutné při importu dat zahrnout minimálně úroveň 1,5, která poskytne vstupní data k následnému výpočtu vybraných metrik. I zde je limit osob při importu omezen na hodnotu 500, ze stejného důvodu dlouhých časových prodlev při omezeném stahování dat o uživateli. Konečné vyhodnocení struktur uvnitř skupiny v náhodné části sítě není překážkou k objektivnímu zhodnocení vizualizovaného stavu podle vypočítaných metrik.

Tabulka 2: Metriky analýzy následujících uživatelem

	Vrchol	Velikost vrcholu	Indegree	Outdegree	Betweenness centrality	Closeness centrality	Eigenvector centrality
1	josefslerka	200	153	300	75833,89	0,003333	0,035196
2	kasperek	84,8	23	14	130,0606	0,001754	0,009161
3	michalblaha	93,5	23	30	422,8075	0,001799	0,011111
4	hlidka	79,9	9	28	175,2441	0,001754	0,008059
5	Europarl_CS	85,1	11	26	159,9329	0,001761	0,009223
6	CRozhlas	82,4	21	9	80,41548	0,001742	0,008606
7	balsinek	80,1	21	12	93,54393	0,001742	0,008096
8	FLutonsky	104,7	23	48	579,3835	0,001828	0,013667
9	kalenskyj	96,7	28	37	437,9351	0,001802	0,011839
10	veselovskyma	108,2	50	20	591,2075	0,001838	0,014447
11	etabery	102,4	55	2	781,3057	0,001842	0,013133
12	AlexandrMitrofa	94,9	31	23	220,1327	0,001783	0,011455
13	atomsedlacek	87,4	31	3	172,6075	0,001767	0,009745
14	PavlinaAmalie	93,5	29	22	290,5636	0,001786	0,011122
15	jilm	86,4	29	8	144,1139	0,001757	0,009525
16	rozaneek	95,4	43	4	306,8418	0,001795	0,011543
17	Medvidekpu	99,5	46	24	995,2312	0,001848	0,012466
18	ladab	84,7	28	16	286,4778	0,001773	0,009126

Zdroj: vlastní zpracování

I při této analýze je potřebné využít dynamický filtr, jelikož importovaných hran je 2058 a síťový graf je velice nepřehledný vzhledem k počtu vazeb mezi vrcholy. Filtr je aplikován na vstupní data pomocí hodnot metriky eigenvector centrality $> 0,0080$, který určuje významnost vrcholu v síti. Použitím a aktualizováním grafu se počet vrcholů výrazně snížil.

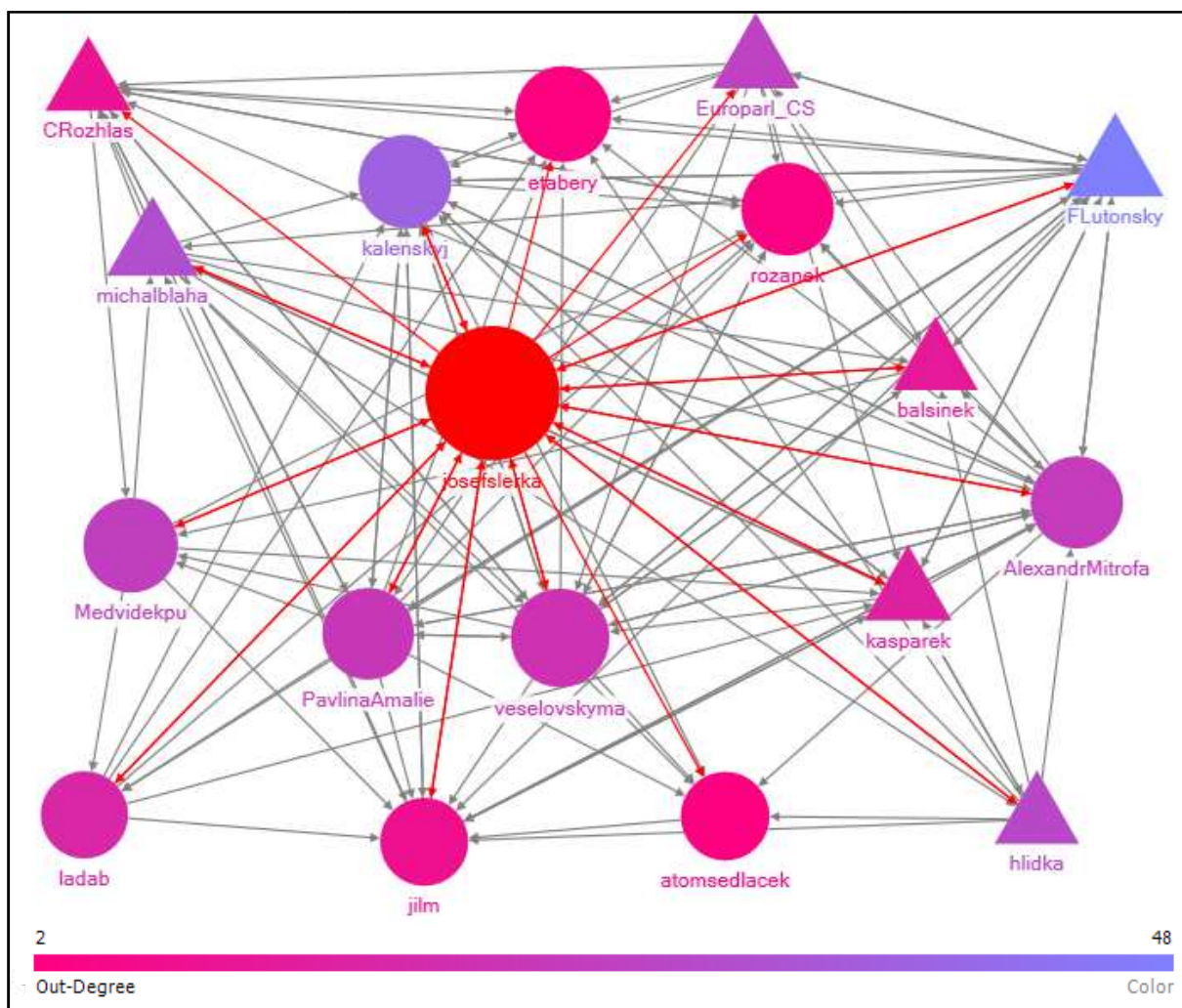
Nyní je možné vypočítané metriky uplatnit na rozlišení vrcholů. Popisky vrcholů reprezentují jednotlivé uživatele a velikost je zde podmíněna taktéž eigenvector centralitou zejména pro zvýraznění nejdůležitějšího jedince sítě, a to Josefa Šlerky. Vzhledem k podobným výsledným hodnotám velikostí vrcholů je zde pro barevné zvýraznění použita důležitější hodnota pro tuto analýzu a to počet výstupních hran z vrcholu, tedy outdegree. Barevná škála se zde pohybuje od 2 do 48 a legenda barevného rozložení je uvedena pod výsledným síťovým grafem. Tvary jsou zde podmíněny hodnotou indegree vrcholů, a to za následující podmínky. Při hodnotách vstupních vazeb do vrcholu větších nebo rovno 23 je tvarem vrcholu kruh, při menších hodnotách je tvarem trojúhelník.

Tabulka 2 obsahuje vypočítané metriky analýzy následujících uživatelem „josefslerka“. Je zde uplatněna velikost vrcholů pomocí metriky eigenvector centrality. Nejvýznamnějším je opět sám „josefslerka“, který je následovníkem ostatních jedinců sociální sítě. Hned po něm následuje „FLutonsky“, „etabery“ a „Medvidekpu“. Uživatel „FLutonsky“ následuje uživatele „etabery“, který jako jediný z těchto jmenovaných jednotlivců nenásleduje Josefa Šlerku. „Medvidekpu“ nemá na grafu této síťové analýzy vyobrazeném na obrázku 8 žádné vazby s těmito uživateli. Nejméně důležitý je v této skupině jedinec „hlidka“, který nemá žádné vazby s dalšími důležitými uživateli sítě kromě uživatele „josefslerka“, který od něho příspěvky odebírá.

Barevné odlišení vrcholů je podmíněné hodnotou outdegree, tedy počtem hran vystupujících z daného vrcholu. Barevná škála se pohybuje od nejmenších hodnot vyobrazených tmavě růžovou barvou, postupuje přes fialovou až k vyšším hodnotám modré barvy. Uživatel „josefslerka“ jako následovník ostatních vrcholů má těchto vazeb nejvíce, a to 300. Všechny hrany na výsledném grafu znázorněny nejsou z důvodu použitého filtru k zřehlednění grafu. Výsledný graf není i tak dostatečně přehledný, aby se daly přesně identifikovat všechny vazby vedoucí od „josefslerka“ k ostatním vrcholům sítě, a to vzhledem k velkému množství znázorněných hran. Jeho vrchol je původně také modrý, ale k docílení znázornění vazeb vedoucích k ostatním jedincům je označen, a proto jsou na výstupním síťovém grafu vrchol i hrany vyobrazeny červenou barvou. Uživatelé „michalblaha, CRozhlas, balsinek, FLutonsky, Europarl_CS, kasperek a hlidka“ jsou znázorněni trojúhelníkem, jelikož je jejich hodnota indegree menší než 23.

Nejvyšší hodnotu metriky closeness centrality má „josefslerka“, který je nejvlivnější osobou sítě k přenášení informací. Druhým vlivným jedincem je „Medvidekpu“, který má zároveň druhou nejvyšší hodnotu betweenness centrality hned po Josefu Šlerkovi a je tak

významným zprostředkovatelem rolí v síti. Kontroluje tok informací v síti a umožňuje dobrou viditelnost všeho, co se v síti odehrává. Dalším takovým uživatelem je „etabery“.



Obrázek 8: Graf síťové analýzy následujících uživatelem

Zdroj: vlastní zpracování

3.6.3 Vyhledávání slovního spojení v tweetu

Tato analýza je zaměřena na vyhledávání tweetů obsahující spojení „josefslerka“. Při importu je nutné zvolit možnost „Import from Twitter search network“, která umožňuje konkrétní slovní spojení v síti vyhledat. Do analýzy jsou zahrnuta veškerá data, kdy bylo toto spojení obsahem, odpovědí nebo zmínkou v tweetu. NodeXL nabízí hledání pouze v nedávných vzkazech, proto je analýza provedena na datech v období od 6. 4. do 14. 4. 2013. Hledání závislostí je vztaženo ke konkrétním událostem, které se odehrály ve vymezeném období a umožní tak nalézt hlubší příčiny jednání a strukturu vztahů výsledné skupiny. Jelikož se importovalo 97 vrcholů a 367 hran, bylo nutné pro větší přehlednost použít dynamický filtr eigenvector centralita $> 0,013$.

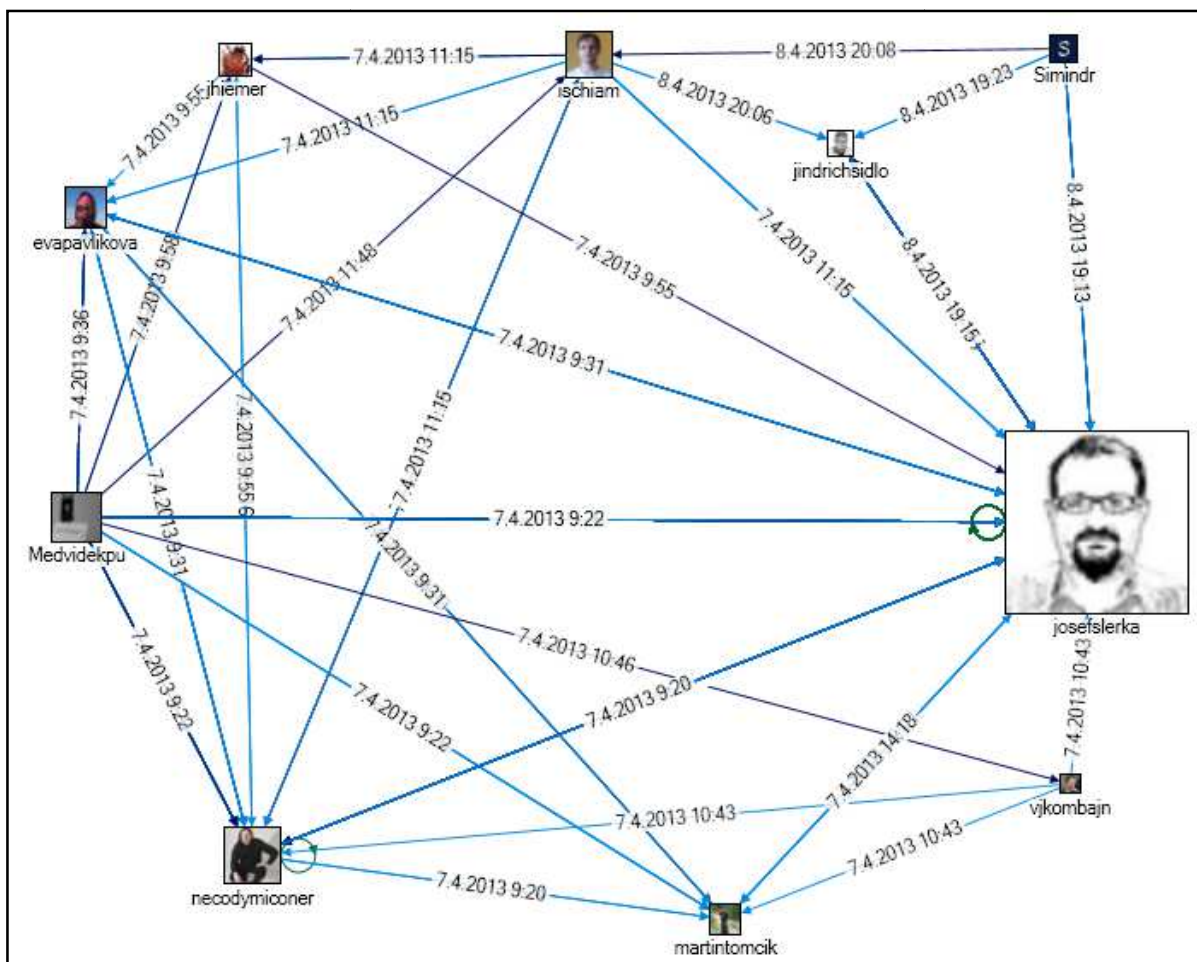
Tabulka 3: Metriky analýzy vyhledávání konkrétního spojení v tweetu

	Vrchol	Velikost vrcholu	Indegree	Outdegree	Betweenness centrality	Closeness centrality	Eigenvector centrality
1	josefslerka	499,5	93	26	8229,167	0,011	0,093
2	Simindr	12,9	0	6	8,667	0,006	0,015
3	jindrichsidlo	11,2	6	1	9,000	0,006	0,015
4	ischiam	32,9	3	5	8,000	0,006	0,018
5	Medvidekpu	38,2	0	8	8,833	0,006	0,019
6	jhiemer	17,4	3	3	0,000	0,006	0,016
7	evapavlikova	25,4	5	3	1,000	0,006	0,017
8	necodymiconer	49,2	8	7	8,833	0,006	0,021
9	martintomcik	14,9	5	1	0,500	0,006	0,015
10	vjkombajn	5,9	1	3	0,000	0,006	0,014

Zdroj: vlastní zpracování

Z výsledků prezentovaných v tabulce 3 je patrné, že nejdůležitějším vrcholem je sám „josefslerka“, jelikož má nejvíce vazeb s dalšími důležitými uživateli. Tato skutečnost byla předpokládána, a proto je nutné hledat zákonitosti hlouběji v síti. Zajímavé je, že hodnota closeness centrality se u všech ostatních vrcholů shoduje, což znamená, že jejich blízkost polohy ke středu je shodná a mají tak v síti relativně stejný vliv. Zjištěné výsledky jsou lépe zřetelné na obrázku 9, kde je znázorněný graf síťové analýzy. Vrcholy nejsou reprezentovány geometrickými tvary jako v předešlých analýzách, ale je zde využita možnost importace obrázků uživatelů, které mají uvedeny na sociální síti. Velikost vrcholů v grafu je podmíněna hodnotou eigenvector centrality pro lepší viditelnost důležitosti v této části sítě.

Druhou nejdůležitější osobou je „necodymiconer“ a hned po něm následuje uživatel „Medvidekpu“. Nejméně vlivný je „vjkombajn“, ačkoliv sám slovní spojení ve svém tweetu uvedl. Důvodem je jeho menší provázanost s dalšími důležitými osobami sítě. Zajímavé je, že uživatel „jindrichsidlo,“ má druhou nejvyšší hodnotu betweenness centrality a je tak důležitým zprostředkovatelem sítě, i když jeho eigenvector centralita dosahuje pouze čísla 0,015. Důvodem tohoto faktu je, že jeho aktivita indegree dosahuje hodnoty 6. Popisky jednotlivých hran reprezentují dny, ve kterých se označení uživatele „josefslerka“ v tweetu objevilo. Je zřejmé, že ve sledovaném období k tomu docházelo nejčastěji ve dnech 7. 4. a 8. 4. 2013. Z této skutečnosti vyplývá zajímavé zjištění, že k takovému označování nedochází pouze nahodile, ale je v mnoha případech podmíněno vnějšími faktory.



Obrázek 9: Graf síťové analýzy vyhledávání slovního spojení v tweetu

Zdroj: vlastní zpracování

3.6.4 Analýza mapy následovníků uživatele

Zmapování geografického rozmístění fanoušků je zajímavou analýzou pro každého uživatele sociální sítě. Otázkou je, zda jsou odběratelé příspěvků uživatele „josefslerka“ převážně z České republiky, nebo jsou rozmístěni po celém světě. V této analýze jsou využita dostupná data, která o sobě ostatní uživatelé poskytují. V případě sociální sítě Twitter je to údaj „Location“, který je možné vyplnit v osobním profilu. Z něho lze následně získat umístění následovníků na mapě s relativní odchylkou od reality. K této analýze postačí import následovníků pomocí nástroje NodeXL a vizualizace pomocí Google Fusion Tables.

Po importu vstupních dat do softwaru je nutné ponechat list „Vertices“ a smazat ostatní přebytečné listy i sloupce. Sešit je vhodné uložit ve formátu „tsv“, který převede data do textového souboru odděleného tabulátory. V dalším kroku je potřeba importovat seznam následovníků do služby Google Fusion Tables, která je primárně určena pro práci a vizualizaci větších souborů dat. V Google Drive se zvolí možnost vytvořit novou tabulku, kam je nutné nahrát soubor s daty. Po úspěšném importu se zobrazí nabídka, kde je potřeba

nastavit u sloupce „Description“ typ na „Location“. Tím je zajištěno, že se budou interpretovat data v uvedeném sloupci, která určí geografickou lokaci následovníků. Poté je už možné vstupní data na mapě vizualizovat.



Obrázek 10: Mapa analýzy následovníků uživatele

Zdroj: vlastní zpracování

Na obrázku 10 je vyobrazena výsledná mapa následovníků uživatele „josefslerka“. Geografické rozmístění jednotlivců je na mapě znázorněno červeným kolečkem. Ne všichni mají údaj o své lokaci na sociální síti vyplněn, proto nedošlo k určení všech uživatelů. Z mapového výstupu je patrné, že nejvíce následovníků je z České republiky. Dalšími významnými oblastmi výskytu jsou Velká Británie a Spojené státy americké. Ostatní výskyty by se daly považovat za více či méně náhodné. Zajímavé je, že v oblasti Ruska nebyl lokalizován ani jeden následovník. Prozkoumání publika na sociálních sítích patří k základům systematické komunikace s ním. Uvedená analýza tak napomohla rozkrýt situaci kolem konkrétního uživatele na sociální síti Twitter.

ZÁVĚR

Všichni jedinci jsou součástí několika sociálních sítí, které ovlivňují jejich životy. S nástupem internetu se pojem sociální síť začal prosazovat i na webových sociálních sítích, které sdružují čím dál více uživatelů. Z tohoto důvodu je analýza sociálních sítí a hledání souvislostí mezi jednotlivými uživateli velice užitečná.

Protože cílem této bakalářské práce je na vybrané skupině osob provést analýzu pomocí nástrojů, které se využívají pro měření sociálních sítí, je první část práce věnována sociálním sítím. Nejprve je objasněn pojem sociální síť z pohledu sociologie a následně přechází k vysvětlení webových sociálních sítí, je zde uvedena její historie i významné sociální sítě současnosti. Druhá část této práce vysvětluje pojmy z oblasti teorie grafů, její historii, podstatu a způsob jakým ji lze k vyhodnocování síťových grafů využít. Třetí část práce se věnuje analýze vybrané skupiny jedinců sociální sítě Twitter. Nejprve je objasněn pojem analýzy sociálních sítí a je zmíněna její historie. Dále je zdůvodněn výběr vhodného softwaru a obecný postup řešení jednotlivých analýz. Výsledky uvedených analýz jsou přehledně graficky znázorněny pomocí síťových grafů, které jsou následně vyhodnoceny.

Síťové analýzy přinesly pomocí nástroje NodeXL zajímavé poznatky o vnitřní struktuře vybrané skupiny a splnily výsledná očekávání. Grafické výstupy znázorňují vzájemnou provázanost s uživatelem „josefslerka“, ale i s ostatními uživateli sociální sítě. Nastiňují možnosti ovlivnění skupiny přes důležité vrcholy této sítě, ale ukazují i slabé články. Analýza pouze určité části rozsáhlé sociální sítě není překážkou k objektivnímu vyhodnocení síťových grafů. Podstatné je ukázat využití základních metrik pro analýzu vybrané skupiny sociální sítě a je možné se domnívat, že při analýze celé sociální sítě by byly nalezeny stejné zákonitosti, a tedy i podobné výsledky. Vzhledem k těmto skutečnostem je patrné, že nástroje pro analýzu sociálních sítí jsou vhodné pro popis struktury vztahů ve vybrané skupině jedinců.

Metoda síťové analýzy umožňuje přehledně a jednoduše graficky prezentovat analyzovaná data. Využitý software napomohl k hledání skrytých a jinak obtížně vysledovatelných informací z importovaných dat. Potenciál této metody je obrovský a její využití je velmi široké. Našla své uplatnění v mnoha oborech, kde lze řešený problém zobrazit pomocí síťového grafu jako vrcholy a hrany, které je propojují. Aplikace by se do budoucna dala využít i například při pozorování dětských skupin v mateřských školách.

POUŽITÁ LITERATURA

- [1] BARABÁSI, A. *V pavučině sítí*. Vyd. 1. Praha: Paseka, 2005. 274 s. ISBN 80-7185-751-3.
- [2] BASL, J. *Podnikové informační systémy: Podnik v informační společnosti*. 1. vyd. Praha: Grada, 2002. 144 s. ISBN 80-247-0214-2.
- [3] BERÁNEK, L. *Síťová analýza v marketingu*. [online]. 2008 [cit. 2013 02-25]. Dostupné z: <<http://znalosti2008.fiit.stuba.sk/download/articles/znalosti2008-Beranek.pdf>>.
- [4] BORNOVÁ, L., ZHOŘ, R., aj. *Využití sociálních sítí v MSP*. [online]. 2011 [cit. 2013-02-16]. Dostupné z: <https://www.ibm.com/developerworks/mydeveloperworks/wikis/home/wiki/W2ee553718f13_4825_b4e6_343b81350b95/page/6%20Soci%C3%A1ln%C3%AD%20s%C3%ADt%C4%9B%20a%20jejich%20vyu%C5%BEit%C3%AD%20v%20MSP?lang=en>.
- [5] BOYD, D., ELLISON, N. Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication* [online]. 2007 [cit. 2013-02-16]. Dostupné z: <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>>.
- [6] BUŠTÍKOVÁ, L. Analýza sociálních sítí. In: *Sociologický časopis* [online]. 1999, roč. 1999, č. 2. [cit. 2013-02-25]. Dostupné z: <<http://www.duke.edu/~lbs11/research/sit.pdf>>.
- [7] ČERNOUŠEK, Š. Využití analýzy sociálních sítí In: *Doktorská vědecká konference*. Praha: Česká zemědělská univerzita, 2011. ISBN 978-80-213-2169-4. Dostupné z: <http://tt.pef.czu.cz/Files/1_printVersion_30.pdf>.
- [8] FRIEBELOVÁ. *Základní pojmy teorie grafů: Graph theory*. [online]. 2008-2009, [cit. 2013-02-11]. Dostupné z: <<http://www2.ef.jcu.cz/~jfrieb/tspp/data/teorie/grafy.pdf>>.
- [9] GIDDENS, A. *Sociologie*. Vyd. 1. Praha: Argo, 1999. 598 s. ISBN 80-7203-124-4.
- [10] Historie sociálních sítí. *Sociální síť* [online]. 2012 [cit. 2013-02-16]. Dostupné z: <<http://www.socialnisite.estranky.cz/clanky/historie-socialnich-siti.html>>.
- [11] HLINĚNÝ, P. *Teorie Grafů* [online]. Verze 1.00. Brno: Masarykova univerzita, 2008 [cit. 2013-02-11]. Dostupné z: <<http://www.fi.muni.cz/~hlineny/Vyuka/GT/Grafy-text07.pdf>>.

- [12] HORDĚJČUK, V. *Teorie grafů*. [online]. 2011 [cit. 2013-02-11]. Dostupné z: <<http://voho.cz/wiki/matematika/graf/>>.
- [13] JACKSON, M. *Social and Economic Networks*. Princeton: Princeton University Press, 2008. 648 s.
- [14] JIROVSKÝ, L. *Teorie grafů* [online]. 2011 [cit. 2013-02-09]. Dostupné z: <<http://teorie-grafu.cz/>>.
- [15] KELLER, J. *Úvod do sociologie*. 1.vyd. Praha: Sociologické nakladatelství, 1992. 192 s. ISBN 80-901059-3-9.
- [16] Marketing v sociálních sítích. *DVORAN: pro krále na internetu* [online]. [cit. 2013-02-17]. Dostupné z: <<http://www.dvoran.com/socialni-site>>.
- [17] NEWMAN, M. *Networks: An Introduction*. New York: Oxford University Press, 2010. 772 s. ISBN 978-0-19-920665-0.
- [18] NodeXL: Network Overview, Discovery and Exploration for Excel. *Microsoft: The Social Media Research Foundation* [online]. 2012 [cit. 2013-02-25]. Dostupné z: <<http://nodexl.codeplex.com/>>.
- [19] PAVLÍČEK, A. *Nová média a sociální sítě*. Vyd. 1. Praha: Oeconomica, 2010. 182 s. ISBN 978-802-4517-421.
- [20] PETRUSEK, M., aj. *Velký sociologický slovník: II. svazek P-Z*. Vyd. 1. Praha: Karolinum, 1996. s. 749-1627. ISBN 80-7184-310-5.
- [21] PROKŮPEK, V. *Z historie sociálních sítí*. [online]. 2012 [cit. 2013-02-16]. Dostupné z: <<http://vaclavprokupek.ano2012.txt.cz/clanky/102529/z-historiesocialnich-siti>>.
- [22] ŠLERKA, J. Social network analysis pro začátečníky. *LUPA CZ: Server o českém internetu* [online]. [cit. 2013-02-11]. Dostupné z: <<http://www.lupa.cz/clanky/social-network-analysis-pro-zacatecniky/>>.
- [23] TOUŠEK, L. *Analýza sociálních sítí*. [online]. [cit. 2013-02-25]. Dostupné z: <<http://www.antropologie.org/cs/vyzkumy-a-analyzy/metodologie/70-analyza-socialnich-siti>>.
- [24] World Map of Social Networks. *Vincoblog* [online]. 2012 [cit. 2013-02-17]. Dostupné z: <<http://vincos.it/world-map-of-social-networks/>>.

[25] ZÁŠKOLNÝ, J. *Sociální síť* [online]. 2011 [cit. 2013-02-16]. Dostupné z: <http://www.socialnisite.123abc.cz/>.