



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ  
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ  
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY  
DEPARTMENT OF INTELLIGENT SYSTEMS

# VYHODNOCOVÁNÍ DŮVĚRY NEBO REPUTACE NA ZÁKLADĚ ANALÝZY SOCIÁLNÍ SÍTĚ

EVALUATING TRUST OR REPUTATION BASED ON SOCIAL NETWORK ANALYSIS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

TOMÁŠ ŠVEC

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. JAN SAMEK, Ph.D.

BRNO 2012

## **Abstrakt**

Tato práce si klade za cíl aplikovat pojem důvěra a reputace z oblasti umělé inteligence na metody analýzy sociální sítě. Obsahuje zhodnocení současných dostupných interakcí pro model důvěry z pohledu zástupců sociálních sítí. Následně popisuje navržený matematický model důvěry pro sociální síť Facebook. Tento model byl implementován v jazyce Python. Experimentální výsledky práce na vzorku uživatelů sítě Facebook jsou následně rozebrány jak z pohledu umělé inteligence, tak z pohledu sociální psychologie.

## **Abstract**

This thesis' goal is to apply the terms trust and reputation from the point of view of artificial intelligence to social network analysis methods. It evaluates current available interactions for a model of trust considering various social networks. A mathematical model of trust for Facebook is designed and described in further parts. This model had been implemented in Python programming language. Experiments regarding this thesis are conducted on a sample amount of Facebook users and furthermore analyzed from the perspective of both artificial intelligence and social psychology.

## **Klíčová slova**

Důvěra, reputace, sociální síť, analýza sociální sítě, kontext důvěry, Open Graph, Facebook, Python, access token.

## **Keywords**

Trust, reputation, social network, social network analysis, context of trust, Open Graph, Facebook, Python, access token.

## **Citace**

Tomáš Švec: Vyhodnocování důvěry nebo reputace na základě analýzy sociální sítě, bakalářská práce, Brno, FIT VUT v Brně, 2012

# Vyhodnocování důvěry nebo reputace na základě analýzy sociální sítě

## Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Jana Samka, Ph.D.

.....  
Tomáš Švec  
17. května 2012

## Poděkování

Na tomto místě bych chtěl poděkovat vedoucímu mé práce, Ing. Janu Samkovi, Ph.D., za trpělivost při konzultacích, odborné vedení práce, poskytnutí výborného přehledu o literatuře a také za možnost zpracovat pro mne velmi atraktivní téma.

© Tomáš Švec, 2012.

*Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.*

# Obsah

<b>1 Úvod</b>	<b>3</b>
<b>2 Analýza sociálních sítí, důvěry a reputace</b>	<b>4</b>
2.1 Důvěra . . . . .	4
2.2 Reputace . . . . .	5
2.3 Sociální síť . . . . .	8
2.4 Metody analýzy sociální sítě . . . . .	8
2.4.1 Kompletní metoda . . . . .	9
2.4.2 Metoda Sněhové koule . . . . .	9
2.4.3 Detekce komunit . . . . .	9
2.4.4 Hierarchické shlukování . . . . .	10
<b>3 Analýza současného stavu</b>	<b>11</b>
3.1 Vlastnosti preferované pro aplikaci . . . . .	11
3.2 Zástupci sociálních sítí . . . . .	11
3.2.1 Myspace . . . . .	12
3.2.2 Google+ . . . . .	12
3.2.3 Facebook . . . . .	12
3.2.4 Twitter . . . . .	12
3.2.5 Spolužáci.cz . . . . .	12
3.3 Interakce . . . . .	13
3.3.1 Příspěvky . . . . .	13
3.3.2 Reakce na příspěvky . . . . .	13
3.3.3 Fotografie . . . . .	13
3.3.4 Skupiny . . . . .	13
3.3.5 Společní přátelé . . . . .	14
3.3.6 Soukromé zprávy . . . . .	14
3.4 Shrnutí . . . . .	14
<b>4 Návrh modelu důvěry pro sociální síť Facebook</b>	<b>16</b>
4.1 Vyčlenění jednotlivých kontextů důvěry . . . . .	16
4.1.1 Časové rozpětí interakcí . . . . .	16
4.1.2 Počet interakcí . . . . .	17
4.1.3 Počet znaků . . . . .	17
4.1.4 Pravidelnost interakcí . . . . .	18
4.1.5 Označení na fotografiích . . . . .	19
4.1.6 Členství ve skupinách . . . . .	19
4.1.7 Společné zájmy . . . . .	20

4.1.8	Další možné rozšíření modelu . . . . .	20
4.2	Agregace kontextů . . . . .	21
<b>5</b>	<b>Implementace modelu</b>	<b>23</b>
5.1	Výběr programovacího jazyka . . . . .	23
5.2	Přístup k datům . . . . .	23
5.3	Potřebná práva pro access token . . . . .	24
5.4	Opakované dotazy na API . . . . .	24
5.5	Struktura aplikace . . . . .	25
5.5.1	Samostatné funkce . . . . .	25
5.5.2	Třída Friend . . . . .	25
5.5.3	Třída Analysis . . . . .	26
5.5.4	Volající funkce . . . . .	26
<b>6</b>	<b>Experimentální výsledky</b>	<b>27</b>
6.1	Návrh průzkumného šetření . . . . .	27
6.2	Průběh šetření . . . . .	27
6.3	Otázky v dotazníku . . . . .	29
6.3.1	Otázka 1 . . . . .	29
6.3.2	Otázka 2 . . . . .	29
6.3.3	Otázka 3 . . . . .	30
6.4	Výsledky průzkumného šetření . . . . .	30
6.4.1	Výsledky pro otázku 1 . . . . .	30
6.4.2	Výsledky pro otázku 2 . . . . .	30
6.4.3	Výsledky pro otázku 3 . . . . .	32
6.5	Vyhodnocení šetření . . . . .	32
<b>7</b>	<b>Závěr</b>	<b>34</b>
<b>A</b>	<b>Výsledky průzkumného šetření</b>	<b>37</b>
<b>B</b>	<b>Obsah CD</b>	<b>38</b>
<b>C</b>	<b>Manuál k použití aplikace</b>	<b>39</b>
C.1	Požadavky pro spuštění skriptu . . . . .	39
C.2	Adresářová struktura . . . . .	39
C.3	Generování přístupového tokenu . . . . .	40

# Kapitola 1

## Úvod

*In a networked world, trust is the most important currency.*

Eric Schmidt

Slova výkonného ředitele společnosti Google z proslovu na Pensylvánské univerzitě vystihují současný stav lidských emocí ve světě síťových technologií. V uplynulých letech docházelo a stále dochází k radikálním změnám definice společenské interakce mezi lidmi. Na internetu o sobě odhalujeme více a více informací a v okruhu našich přátel přibývá lidí, se kterými komunikujeme pouze přes počítačovou síť. Ačkoliv je umělá inteligence ještě na míle daleko od úspěšného zvládnutí Turingova testu [23], nabízí se otázka, zda je skutečně možné pomocí nalezených vzorců v lidském chování simulovat emoce.

Tato práce se zabývá otázkou, zda lze vytvořit model důvěry z oblasti umělé inteligence, který by pracoval s důvěrou z pohledu sociální psychologie na sociálních sítích. Základním předpokladem je fakt, že pokud byl princip důvěry ve svých začátcích přenesen do informatiky ze sociologie a sociální psychologie, pak lze po téměř dvaceti letech vývoje tento princip opět aplikovat v původním oboru a studovat odlišnosti.

V úvodu práce budou definovány termíny potřebné k pochopení a zorientování se v problematice důvěry a reputace jak z hlediska informatiky, tak sociální psychologie. Také budou popsány dostupné metody analýzy sociálních sítí a jejich vhodné využití pro následnou implementaci.

Pro implementaci modelu je nutné vybrat jednu konkrétní sociální síť. Bude popsána provedená analýza možných interakcí mezi entitami z pohledu zástupců jednotlivých sociálních sítí. Analýza navíc bere v potaz další faktory, jako např. rozšíření v českém a mezinárodním prostředí, bezpečnost dat uživatelů, kteří se model rozhodnou vyzkoušet, nebo API přístupné pro vývojáře.

Samotný navržený model bude popsán z matematického hlediska s co nejmenší možnou měrou závislosti na konkrétní implementaci, aby jej bylo možné v případě potřeby realizovat v libovolné sociální síti. Budou také rozebrány předpokládáné důvody korelace jednotlivých typů interakcí s důvěrou mezi entitami.

Část o implementaci modelu popisuje praktické poznatky z implementace v jazyce Python a v sociální síti Facebook. Budou vysvětlena programátorská omezení dané sítě, jejich řešení v konkrétním jazyce a zajištění bezpečnosti dat uživatelů. Nastíněn bude také výsledný objektový návrh.

Ověření platnosti celého modelu a jeho implementace bude následně popsáno a shrnuto ve formě průzkumného šetření mezi uživateli zvolené sociální sítě.

## Kapitola 2

# Analýza sociálních sítí, důvěry a reputace

Teoretická část práce je zaměřena především na definování pojmů používaných ve zbytku textu. Jelikož téma důvěry a reputace v sociálních sítích stojí na rozhraní dvou vědních oborů mezi oblastmi sociologie a umělé inteligence v informatice, není vždy možné držet se striktně technického akurátního popisu a matematických definic. Pro vymezení lidských emocí nebyl dosud definován žádný matematický aparát. Přesto je orientace práce spíše technického než humanitního rázu. Pokud se vyskytne pojem, který lze definovat více způsoby, budou vždy uvedeny alespoň dvě všeobecně uznávané definice k porovnání.

Podkapitoly jsou vystavěny na principu duálního paralelismu právě mezi sociologií a informatikou. Převod důvěry a reputace do digitálního světa mělo velký význam, stejně jako přenesení principu sociálních sítí z fyzicky existujících komunit do komunit v počítačových sítích a na internetu. Právě v tomto paralelismu lze hledat argumenty pro výhody spojení důvěry se sociální sítí.

### 2.1 Důvěra

Z hlediska sociologie je důvěra dlouho diskutovaným pojmem. Bruce Schneier [25] je toho názoru, že právě schopnost budovat důvěru mezi jedinci, kteří s námi nejsou pokrevně spřízněni, byla základem moderní společnosti. Evoluční biologie popisuje mechanismus, který lze vysledovat i k dnešním živočišným druhům, který nazývá *reciproční altruismus*. Princip je často uváděn na příkladu skupiny zvířat (stáda, hejna, regionální jednotky), která spolu sdílí potravu v případě, že pro jedno nebo více z nich dopadl lov neúspěšně. Tento jev jde na první pohled proti Darwinově teorii evoluce. Pokud by se jednalo o dělení potravy v konkrétní okamžik, pak by to byla samozřejmě pravda. Zvíře však očekává, že pokud bude neúspěšný pro změnu jeho lov, dostane možnost přežít právě díky ostatním.

Nebylo by vhodné pouštět se do rozsáhlého rozboru pojmu *důvěra* z hlediska jiných vědních oborů (a není to ani cílem této práce). Uvádím zde proto definici vycházející z moderních poznatků jednoho z nejuznávanějších sociologů současnosti, Anthonyho Giddense [13]:

Trust is related to absence in time and in space. There would be no need to trust anyone whose activities were continually visible and whose thought processes were transparent, or to trust any system whose workings were wholly known and understood. It has been said that trust is *a device for coping with*

*the freedom of others*, but the prime condition of requirements for trust is not lack of power but lack of full information.

Ve volném překladu:

Důvěra je spojena s vlastnostmi času a prostoru. Kdyby jednání všech lidí a jejich myšlenkové pochody byly vždy a všude transparentní, nebo pokud by existoval systém, jehož principy by byly zcela pochopitelné a transparentní, nebylo by důvěry vůbec třeba. Důvěra je *prostředkem k vyrovnaní se se svobodou ostatních*, ale primárním předpokladem její existence není nedostatek moci, nýbrž nedostatek informací.

V roce 1994 se objevila v akademických kruzích teorie, která zásadním způsobem ovlivnila odvětví umělé inteligence. Stephen Paul Marsh tehdy ve své práci [21] zavedl princip důvěry v multiagentních systémech. Jeho chápání důvěry bylo založeno na mnoha definicích z humanitních oborů. Ačkoliv je v jeho práci dodržována technická akurátnost a navržený matematický aparát lze považovat za dostatečně obsáhlý, pojem „důvěra“ se v informatice nikdy nedočkal přesné definice. Dokonce ani Marsh si ji netroufl ve své práci sám definovat. Jako nejpopulárnější uvádí definici psychologa Deutsche [11]:

1. The individual is confronted with an ambiguous path, a path that can lead to an event perceived to be beneficial (Va +) or to an event perceived to be harmful (Va -).
2. He perceives that the occurrence of Va + or Va - is contingent on the behaviour of another person.
3. He perceives the strength of Va - to be greater than the strength of Va +.

If he chooses to take an ambiguous path with such properties, I shall say he makes a trusting choice; if he chooses not to take the path, he makes a distrustful choice.

V překladu pak:

1. Jedinec má před sebou výběr cesty, která vede k události vnímané jako prospěšná (Va +) nebo k události vnímané jako škodlivá (Va -).
2. Je si vědom toho, že výskyt Va + nebo Va - závisí na chování jiného člověka.
3. Vnímá možnost Va - silněji než Va +.

Pokud si zvolí nejednoznačnou cestu, říkáme, že je volba založena na důvěře. Pokud tuto cestu nezvolí, je volba založena na nedůvěře.

## 2.2 Reputace

Intuitivně si pod pojmem *reputace* představujeme názor na nějakou společenskou entitu (člověka, skupinu lidí, organizaci), který je často výsledkem hodnocení mnoha kritérií. Zajímavé je, že v humanitních vědách není reputace jako taková definována. Tomuto termínu se nejvíce přibližuje termín *sociální prestiž*. Nejlépe rozdíl mezi těmito dvěma termíny zřejmě vystihuje práce [27]:



Podstatným problémem každého člena skupiny je jeho *prestíž*, která souvisí s osobnostními vlastnostmi člověka a jeho uplatněním ve skupině. Jde o ocenění člověka skupinou a tato osobní hodnota je označována jako status (hovorově právě „*reputace* – pověst, vážnost, úcta” [18]). Status a pozice se vzájemně propojují a vytvářejí hierarchizovaný systém. Se systémem statusů a jim odpovídající sociální prestiže se setkáváme všude, kde lidé žijí ve skupinách. Čím větší provoz, čím větší podnik nebo instituce, tím propracovanější systém hierarchie je lidmi vytvořen.

Z hlediska sociálních věd je možné ještě zmínit pojem *pozice*. Je důležitý zejména z pohledu kontextů důvěry, jejichž použití je vysvětleno v části 4.1. Podle Řezáče [28] označujeme místo, které jedinec ve skupině zaujímá, pojmem *pozice*. Pozice určuje postavení jedince ve vztahu k ostatním členům skupiny v dimenzích: přiřazenost, nadřazenost a podřazenost. Pozice jedince ve skupině je určována:

- mírou sociální přitažlivosti jedince
- mírou prestiže
- způsobem prosazování při začleňování se do skupiny
- podílem jedince na dosahování skupinových cílů

Dle práce [25] je důvěra přirozeným obranným mechanismem lidské společnosti. Reputace nám pomáhá zjistit, komu můžeme věřit, v koho vložit svou důvěru. Pokud se ve společnosti vyskytne jedinec, který zneužívá důvěru druhých a opakovaně jedná pouze ve svůj osobní prospěch, reputace je nejpřirozenějším nástrojem k jeho označení a varování ostatních. Dlouhodobě nepříznivá reputace pak často vede ke stigmatizaci a k úplnému vyloučení z dané společenské skupiny.

V tomto kontextu považuji za důležité zmínit princip, který dále rozvádím v kapitole 4.1.8 již v konkrétním implementačním případě. Jedná se o fakt, že v průběhu vývoje lidské společnosti nebylo z evolučního hlediska možné, abychom se adaptovali na novou škálu modelu reputace. Tzv. *Dunbarovo číslo* [12] označuje počet lidí, se kterými jsme schopni udržovat sociální vztahy. Dle výzkumů toto číslo dlouhodobě stagnuje, zatímco měřítko společnosti, ve které žijeme, vzrostlo několikanásobně.

Navážeme-li na předchozí odstavec, pro vývoj společnosti bylo nutné reputaci určitým způsobem formalizovat. Zde lze spatřit paralelu s digitálním světem, byť se tato formalizace udála tisíce let před vynálezem mikročipu. Formalizováním reputace lze dosáhnout rozsahu, pro který důvěra ani reputace nejsou v prvoplánovém principu vůbec stavěny. Nevěříme už člověku, který vždy léčil naše příbuzné a sousedy, důvěřujeme lidem, kteří mají na zdi pověšený lékařský diplom. Přestože jsme toho člověka patrně nikdy neviděli, věříme např. státu nebo lékařské komoře, že tohoto jedince dostatečným způsobem proškolili a dohlíží na to, že bude dodržovat standardy spojené s danou profesí. Jinými slovy správu reputace přenecháváme jiným entitám, zpravidla institucionalizovaným.

V digitálním světě má reputace význam velmi podobný světu fyzickému. Spousta internetových komunit je založena právě na reputaci, jmenujme např. fórum známé ve výzkumech v oblasti důvěry Slashdot.org nebo stránku eBay [2], která reputaci přímo využívá pro hodnocení jednotlivých uživatelů/prodejců/zákazníků (často vše v jednom). Právě na stránkách eBay je zaveden formalismus reputace, který umožňuje tento systém škálovat do

zcela jiných měřítek, než pro která byl navržen, i když jde pochopitelně o větší riziko zneužití. Jako další příklad mohou posloužit i sociální sítě, přestože nejde tak úplně o reputaci jako takovou, což bude vysvětleno v následujících odstavcích.

Elizabeth Gray definuje reputaci z infromatického hlediska takto [14]:

A reputation is a collection of recommendations, i.e., personal observations recommended by one or more third parties, about an entity's past behaviour which are accumulated in such a way as to characterise an entity's nature with regard to ability or reliability in potential future interactions in a given context. If the accumulated recommendations are evidence of behaviour for a given trust purpose, then the resultant reputation characterising an entity's trustworthiness can be used as input to a trust-based decision-making system.

V překladu:

Reputace je souborem doporučení, např. soukromá pozorování jedné nebo více třetích stran, o dosavadním jednání entity, která jsou shromažďována takovým způsobem, aby charakterizovala vlastnosti entity ve vztahu ke schopnostem nebo spolehlivosti v potenciálních budoucích interakcích v daném kontextu. Pokud jsou shromážděná doporučení důkazem jednání v konkrétním kontextu důvěry, lze výslednou reputaci charakterizující důvěryhodnost entity použít jako vstup rozhodovacího systému založeného na důvěře.

Tato práce však není zaměřena na reputaci, nýbrž na důvěru. Na závěr kapitoly by tedy bylo vhodné osvětlit rozdíl mezi důvěrou a reputací. Výstižné porovnání použil doktor Audun Josang a jeho kolegové [17]:

The difference between trust and reputation can be illustrated by the following perfectly normal and plausible statements:

1. „I trust you because of your good reputation.”
2. „I trust you despite your bad reputation.”

Assuming that the two sentences relate to identical transactions, statement (1) reflects that the relying party is aware of the trustee's reputation, and bases his trust on that. Statement (2) reflects that the relying party has some private knowledge about the trustee, e.g. through direct experience or intimate relationship, and that these factors overrule any reputation that a person might have. This observation reflects that trust ultimately is a personal and subjective phenomenon that is based on various factors or evidence, and that some of those carry more weight than others. Personal experience typically carries more weight than second hand trust referrals or reputation, but in the absence of personal experience, trust often has to be based on referrals from others.

V překladu:

Rozdíly mezi důvěrou a reputací lze demonstrovat na následujících zcela přijatelných výrocích:

1. „Důvěřuji ti, protože máš dobrou reputaci (pověst).”

2. „Důvěřuji ti, přestože máš špatnou reputaci (pověst).”

Předpokládejme, že se obě věty vztahují na stejnou situaci. Věta (1) odráží fakt, že důvěřující strana si je vědoma reputace příjemce důvěry a zakládá svou důvěru na ní. Věta (2) odráží fakt, že druhá strana disponuje informacemi o příjemci důvěry, např. přímá zkušenost nebo osobní vztah, a tyto faktory předčí reputaci, kterou daná osoba má. Lze tedy vyzorovat, že důvěra je zcela osobním a subjektivním jevem, který je založen na mnoha faktorech nebo důkazech, a že některé mohou mít větší váhu než ostatní. Osobní zkušenost má často větší váhu než doporučení nebo reputace z druhé ruky, ale pokud osobní zkušenost chybí, důvěra se často musí zakládat na doporučení od ostatních.

## 2.3 Sociální síť

Pokud uvažujeme o použití sociální sítě pro tvorbu modelu důvěry nebo reputace, je třeba nejprve „sociální síť” definovat. Tomuto pojmu dal vzniknout doktor Barnes ve své hojně citované práci, kde popsal svůj pobyt v norské vesnici Bremnes [8]. Vzhledem k rodovým tradicím a izolovanosti této vesnice od zbytku světa byl Barnes schopen zkoumat určité třídní jevy a kategorizovat obyvatelstvo do skupin. Tyto relativně autonomní skupiny pak popsal pomocí nového pojmu „social network”. Za více než 50 let své existence byl pojem zobecněn takto:

A social network is a social structure made up of individuals (or organizations) called „nodes”, which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige.[3]

Ve volném překladu tedy

Sociální síť je sociální struktura tvořená jednotlivci (nebo organizacemi) nazývanými „uzly”, které jsou vázány (spojeny) jedním nebo více typy propojení, např. přátelství, příbuzenství, společný zájem, finanční vztahy, odpor, sexuální vztahy nebo vztahy na základě víry, znalostí nebo prestiže.

Pro srovnání zde uvádím rovněž definici tohoto pojmu v oblasti sociologie uvedenou v Ottově všeobecné encyklopedii: [19].

sepjetí i kontakty mezi lidmi; v běžné řeči tzv. známosti

V užším slova smyslu dnes sociální sítí rozumíme internetovou službu určenou ke komunikaci mezi uživateli. Definice mají v tomto případě poměrně široký rozptyl.

## 2.4 Metody analýzy sociální sítě

Po definování pojmu sociální síť již můžeme podrobněji rozebrat rovněž jejich analýzu, která se rozvíjela zejména na začátku druhé poloviny 20. století. S rozšířením globalizace a masové výroby dávalo smysl analyzovat síť zákazníků a jejich potřeby. Toto období zavdalo záminku k vytvoření a rozvoji mnoha vědních oborů, např. nové využití teorie grafů, sociogramy, matematická teorie relací. Vznikl nový druh umělé inteligence, který

na základě sběru a dolování dat (*Data Mining*) je schopen předpovídat vzájemné vztahy a potřeby aktérů. Vzniklý obor zaměstnal mnoho špičkových matematiků z oblasti statistiky a zpracování dat. Forma inteligence, která sítě analyzuje, se nazývá Numerati. Více informací o tomto fenoménu lze nalézt např. v publikaci [7].

Pro tuto práci jsou však klíčové metody, které lze na analýzu sociální sítě použít. Následuje krátký přehled nejpoužívanějších metod, který zpracoval již můj předchůdce ve své bakalářské práci [20]. U každé metody je rovněž uvedeno vysvětlení, proč je tato metoda vhodná či nevhodná pro konkrétní analýzu v daném kontextu této práce.

### 2.4.1 Kompletní metoda

Tato metoda vyžaduje sběr informací o všech vazbách, které se v sociální síti vyskytují. Je ekvivalentem kartézského součinu v matematice, kdy pro každé dva účastníky musíme vyjádřit vzájemný vztah. Jak už název napovídá, jedná se o nejsilnější metodu analýzy, která je však také velmi náročná na čas a zdroje. Tato metoda je pro plánovanou analýzu sociální sítě nevhodná, neboť by vyžadovala spolupráci všech účastníků a také přístupové údaje nebo pověření od každého z nich zaráz. Můžeme pouze spekulovat, do jaké míry tato analýza probíhá na straně serverů sociálních sítí samotných.

### 2.4.2 Metoda Sněhové koule

Sněhová koule (z anglického *Snowball method*) funguje podobně jako rekurzivní systém dotazování DNS. Stopuje vazby počátečního aktéra, nalezne podle nich další aktéry a dále se rekurzivně dotazuje na ně. To lze provádět do té doby, dokud nám výsledky nestačí, popř. dokud nevyprší tolerovaný výpočetní čas. Touto metodou lze analyzovat propojené komunity nejen uvnitř sociální sítě, americká FBI údajně ve složitější podobě tuto metodu používá k izolaci teroristických buněk [7]. Samozřejmě je třeba zmínit i nevýhody, které spolu souvisejí. Touto metodou není možné nalézt osamocené entity v grafu sociální sítě, jelikož neexistuje garantovaný způsob nalezení všech propojení. Izolované entity mohou přitom být pro výzkum velmi významnou oblastí k analýze. Nabízí se otázka, zda si mezi sebou členové izolované entity věří více než lidé, kteří jsou otevřeni okolnímu světu.

Pro analýzu v rámci této bakalářské práce je tedy tato metoda vhodná, v širším nasazení by však bylo vhodné kombinovat ji s individuálním určením konkrétních uzlů, které je třeba prozkoumat. To ale vyžaduje hlubší znalosti struktury sítě a vylučuje to možnost plně strojového zpracování.

### 2.4.3 Detekce komunit

Nalezení komunit, klastrů nebo soudržných podskupin (terminologie se různí) je jedním z klíčových úkolů analýzy sociálních sítí. Tato metoda se snaží analyzovat skupiny uživatelů sociální sítě, které mezi sebou výrazně častěji komunikují. Detekce skupin může probíhat zjišťováním platnosti kritérií mezi uzly (vzájemnost, dosažitelnost nebo stupeň). Lze rozpracovat teorii, ve které můžeme na základě symetrické důvěry mezi uzly určovat příslušnost k určité skupině. Do této oblasti však tato bakalářská práce příliš nezasahuje, byť se jedná o jedno z možných využití zde navrženého modelu.

#### **2.4.4 Hierarchické shlukování**

Hierarchické shlukování využívá komplexní matematickou teorii pro vytvoření shluků na různých úrovních. Pro tuto metodu není nutné na vstupu specifikovat předem očekávaný počet skupin, které analýzou mají vzniknout. Ačkoliv se jedná o používanou metodu např. v rámci marketingu sociálních sítí, je pro aplikaci v této práci příliš komplexní a kolidovala by s transparentním použitím modelu.

## Kapitola 3

# Analýza současného stavu

Tato kapitola si klade za cíl analyzovat současný stav na trhu se sociálními sítěmi s důrazem na podporované interakce mezi uživateli, které mohou dobře posloužit pro matematický model. V části 3.1 popisuje požadavky pro konkrétní aplikaci. Část 3.2 se zabývá jednotlivými zástupci sociálních sítí. Část 3.3 pak popisuje typy interakcí, které dnešní sociální sítě z velké části podporují. Závěrem kapitola prezentuje shrnutí získaných poznatků a vybraného kandidáta pro implementaci.

### 3.1 Vlastnosti preferované pro aplikaci

Pro vytvoření modelu důvěry nebo reputace bylo stanoveno pět základních kritérií, která by měla vybraná služba splňovat. Pro konstrukci modelu je výhodné mít k dispozici co nejpočetnější množinu operací, které mohou mít zásadní vliv na vytváření vztahů důvěry mezi jednotlivými uživateli dané služby. Nestačí však, aby tato množina existovala, musí být také aktivně využívána, aby se promítla do výsledku. Z programátorského hlediska bylo požadováno pokud možno co nejjednodušší a nejintuitivnější API (Application Programming Interface) ideálně přímo od poskytovatele služby, aby bylo možné se soustředit spíše na podstatu problému než na implementační detaily. Posledním požadavkem ovlivnitelným poskytovatelem pak byla poměrně nová vlastnost nazvaná Location Awareness, která umožňuje zjišťovat skutečnou geografickou polohu uživatelů. Za předpokladu, že funkci uživatel povolí, bude možné v budoucnu do modelu zahrnout rovněž vztahy v rámci lokality.

Poslední dva požadavky jsou čistě pragmatické. Mělo by se jednat o rozšířenou službu, aby bylo co nejjednodušší získat reprezentativní a pokud možno co nejpočetnější vzorek dat. Ideální je pak co největší rozšířenost v českém prostředí, jelikož sběr dat bude probíhat u českých uživatelů.

### 3.2 Zástupci sociálních sítí

Na základě průzkumu firmy InSites Consulting [9] jsou v Evropě nejznámějšími sociálními sítěmi Facebook, Twitter a Myspace. Vlastnosti těchto tří sítí jsem zahrnul do porovnání. Navzdory těmto údajům však největší procento uživatelů v Evropě využívá služby sítí Facebook, Twitter a VKontakte. Síť Myspace je známá jako jedna z prvních sociálních sítí a nyní je v konfrontaci se sítí Facebook spíše na ústupu. VKontakte je pak síť oblíbená zejména v Rusku a na Ukrajině, v České republice je poměrně neznámá. V rámci porovnání jsem do seznamu přidal novou sociální síť firmy Google, G+, kterou sice nevyužívá takové

procento uživatelů, ale může být zajímavá z hlediska použitého API. Pro české prostředí je pak typická služba Spolužáci.cz. Následuje přehled jednotlivých sítí a jejich hlavních rysů.

### 3.2.1 Myspace

Síť Myspace byla první masově rozšířenou sociální sítí na internetu. Roku 2006 předčila návštěvností v USA i stránku Google.com. Umožňuje zveřejňovat zprávy pro všechny přátele v kontaktech (funkce Bulletin). Tato funkce má pro výzkum poměrně zásadní nevýhodu – zprávy se po deseti dnech automaticky mažou. V roce 2008 Myspace představil API pro vývoj aplikací založené na API sociální sítě Facebook. I přesto ztratil ve stejném roce vedoucí postavení na trhu a dnes počet uživatelů konstantně klesá. V českém prostředí je služba málo rozšířená.

### 3.2.2 Google+

V loňském roce společnost Google oznámila vývoj a uvedení přímého konkurenta sociální sítě Facebook. Služba G+ byla spuštěna 20. září 2011 pro veřejnost. Prezentovala některé inovativní koncepty, které by bylo možné začlenit do modelu důvěry, např. *kruhy (circles)*. Jedná se o skupiny, do kterých si uživatelé člení své kontakty a nastavují tak zároveň viditelnost svých příspěvků. Klíčovým nedostatkem je přístup ke kruhům pouze z účtu uživatele, který je vytvořil. Vznikl by pak pouze jednostranný model důvěry. Navzdory rozsáhlým možnostem interakce a intuitivnímu API v jazyce JavaScript však zůstává největším problémem sítě nedostatečná penetrace jak na světovém, tak českém internetu.

### 3.2.3 Facebook

Podle statistik společnosti InSites Consulting se jedná o nejrozšířenější sociální síť na světě. Daří se jí rovněž v českém prostředí. Vzhledem k velkému počtu uživatelů, přes 800 milionů k červenci 2011, jsou některé funkce již notoricky známé. Jedná se např. o tlačítko „Like“ (česká lokalizace „To se mi líbí“) nebo „Become a fan“ („Stát se fanouškem“). Facebook v roce 2010 spustil nové API pro vývojáře pod názvem Open Graph API. Pouze pomocí adresy a využití OAuth autorizace lze získat informace ve formátu JSON.

### 3.2.4 Twitter

Twitteru se týká jeden zajímavý paradox v uvedeném průzkumu. 80% respondentů jej znalo, ale jen 16% používalo. Je to dáno především zaměřením sítě. Zatímco ostatní zmíněné služby slouží spíše ke komunikaci mezi jednotlivci nebo sociálními skupinami, Twitter slouží k šíření krátkých informací o délce maximálně 140 znaků mezi tzv. „followers“. Často tyto možnosti využívají firmy nebo známé osobnosti. Pro některé uživatele se stal Twitter náhradou RSS kanálů. Výhodou je kvalitní API využívající, stejně jako Facebook, protokolu OAuth 2.0. Největší nevýhodou je však absence více druhů interakce. Lze pouze zjistit, kdo koho „sleduje“, koho zmiňuje ve zprávách a na koho reaguje.

### 3.2.5 Spolužáci.cz

V českém prostředí se jedná o populární variantu sociální sítě pro většinou současné, občas však také bývalé spolužáky, kteří se snaží zůstat v kontaktu. Problémem této sítě není ani tak rozšířenost, jako spíše absence použitelného API. Byť jsou možnosti interakce poměrně široké, většina uživatelů je nevyužívá.

### 3.3 Interakce

Model důvěry je pro evaluaci nutno aplikovat na konkrétní sociální síť a porovnat výsledky s realitou. I přes tento požadavek se však budu snažit řešení zobecnit. Za tímto účelem zde předkládám popis některých typů interakcí, které se budou objevovat v mnou navrhnutém modelu. Pro vybraných pět sítí je třeba popsat specifické rysy jednotlivých interakcí a jejich vliv na použití v modelu.

#### 3.3.1 Příspěvky

Příspěvky na sociální síti obecně slouží k šíření informace mezi přáteli. Všechny popsané sítě tuto základní funkci podporují. Rozdíl se týká hlavně omezení viditelnosti pro skupiny přátel. Inovativní přístup zde nabízí hlavně Google+ s principem *kruhů*, do kterých má uživatel přátele kategorizované. Pokud chce zveřejnit zprávu pouze pro blízké přátele, je možné nastavit viditelnost pouze na tento kruh. Z hlediska analýzy důvěry však tato vlastnost postup paradoxně komplikuje, neboť je nutné ve vyhodnocení reakce uživatele brát v potaz faktor, zda příspěvek vůbec viděl. To vyžaduje zásah do soukromí uživatele, neboť kruhy jsou pro ostatní skryty. Pro úplnost ještě dodávám, že zprávy na sociální síti Twitter jsou omezeny délkou 140 zpráv, aby bylo možné zasílat tyto informace též textovou zprávou SMS. Je to dáno historickým pozadím vzniku Twitteru. Ostatní služby umožňují zveřejňovat příspěvky podstatně delší.

#### 3.3.2 Reakce na příspěvky

Možnost příspěvek komentovat nabízí kromě Twitteru všechny vyjmenované služby. V modelu důvěry stavím reakci na příspěvek v koeficientech velmi vysoko, neboť se jedná v podstatě o jakousi náhražku běžného rozhovoru mezi dvěma lidmi. Facebook a Google+ vnášejí do této oblasti navíc subjektivní princip tlačítka „To se mi líbí“, popř. Google varianta „+1“. I když nedojde k žádné faktické výměně informací, tuto událost zaznamenám jako vzájemnou interakci a tedy se posílí důvěra mezi aktéry. Twitter neumožňuje přímou reakci nebo komentování příspěvku, je však možné ve vlastním *tweetu* pomocí tzv. *tagů* reagovat na konkrétní příspěvek.

#### 3.3.3 Fotografie

Facebook, Google+ i MySpace nabízí v podstatě identickou možnost označovat na zveřejněných fotografiích přátele v seznamu. Tuto interakci rovněž stavím v žebříčku velmi vysoko, neboť přítomnost obou aktérů na jedné fotografii indikuje setkání v reálném životě. Je ovšem nutné filtrovat obrázky, ve kterých je označen větší počet uživatelů. S velkou pravděpodobností se jedná např. o přání k Vánocům nebo řetězové fotografie, na kterých uživatel označí téměř všechny své přátele. I kdyby se ovšem jednalo o společnou fotografii většího počtu lidí, jejím vyloučením z modelu důvěry žádnou důležitou informaci neztratíme. Může se jednat např. o třídní fotku nebo fotku fotbalového mužstva, což je zahrnuto v dalším bodu.

#### 3.3.4 Skupiny

MySpace rozdělení uživatelů do skupin podle určitých společných rysů (což mimochodem odpovídá definici sociální sítě na začátku kapitoly) řeší jako „přátelství“ s danou skupinou.



Google+ ani Twitter skupiny nemodelují vůbec. Nejbohatší základy pro analýzu pokládá Facebook, na kterém postupem času vznikly skupiny prakticky na všechno a jsou často terčem kritiky zarytých odpůrců této sociální sítě. Pro tvorbu modelu důvěry jsou však nesmírně cenné. Empiricky bylo zjištěno, že váha faktu „dva aktéři přísluší do stejné skupiny“ je nepřímou úměrnou velikosti dané skupiny. Skupina pokrývající globální fenomén, která obsahuje několik milionů fanoušků, uživatele spojuje s mnohem menší pravděpodobností než skupina, ve které se sdružují např. členové lokální organizace.

### 3.3.5 Společní přátelé

Tento faktor lze u všech uvedených sítí modelovat zcela bez problémů. Stačí porovnat seznamy přátel jednotlivých uživatelů. Nemusí se nutně jednat o určující faktor. Pokud však uživatel v nastavení soukromí explicitně informaci neskryje, je brána jako veřejná a lze k ní tedy přistupovat bez asistence uživatele, což je pro počáteční odhad struktury vzniklé sítě důvěry nedocenitelné. Je třeba také dodat, že Twitter používá odlišný koncept „přátelství“ a je tedy nutné porovnávat symetrické vlastnosti množin *followers* a *following*.

### 3.3.6 Soukromé zprávy

Mohlo by se zdát, že soukromé zprávy jsou ideálním materiálem pro výstavbu modelu důvěry. V této práci je ovšem vypouštím z několika klíčových důvodů:

- Ochrana soukromí uživatelů – zatímco příspěvky, fotografie, členství ve skupinách nebo společní přátelé jsou údaje, které uživatel zveřejňuje poměrně ochotně, obsah soukromých zpráv mezi uživateli je často citlivého charakteru a ne každý je ochoten poskytnout k nim přístup. Razantně by se tedy snížil počet ochotných účastníků experimentu.
- Charakter komunikace – nejedná se zde již o sociální síť jako takovou, ale spíše o Instant Messaging. Soukromé zprávy začínají nahrazovat služby typu ICQ, Jabber nebo MSN. Nejedná se tedy o komunikaci se skupinou, jak ji definuje Barnes, ale o komunikaci výhradně mezi dvěma aktéry.
- Množství dat – z čistě pragmatického hlediska by se jednalo o velmi podstatné objemy dat pro každého uživatele v řádu jednotek MB, což by znemožňovalo analýzu provést ve větším měřítku.

Z těchto tří důvodů nejsou soukromé zprávy v modelu důvěry ve finální verzi práce zahrnuty vůbec. Pro zájemce o tuto problematiku doporučuji prostudovat zásuvný modul HistoryStats pro komunikátor Miranda, jenž je schopen z historie programu generovat přehledovou tabulku XML i grafy.[\[6\]](#)

## 3.4 Shrnutí

Každá ze jmenovaných sociálních sítí má některé kýžené vlastnosti odpovídající požadavkům. Pro přehlednost přikládám ještě jednou seznam požadavků spolu se zkratkou, pod kterou jsou uvedeny v souhrnné tabulce.

- Interakce – početná množina operací pro interakci

- API – jednoduché a přístupné API
- GPS – Location Awareness
- Penetrace – dostatečná penetrace na trhu
- České – rozšíření v českém prostředí

	Interakce	API	GPS	Penetrace	České
Myspace	x		x		
Google+	x	x	x		
Facebook	x	x	x	x	x
Twitter		x	x	x	
Spolužáci	x			x	x

Tabulka 3.1: Přehled vlastností jednotlivých sociálních sítí.

Jak je patrné z tabulky 3.1, nastavené podmínky splnila pouze sociální síť Facebook. Je jak v českém, tak ve světovém prostředí dostatečně rozšířená. Zpřístupňuje uživatelům bohatou množinu operací, kterou můžeme v práci analyzovat. Open Graph API pak představuje zřejmě nejintuitivnější z rozhraní. Služby navázané na Location Awareness jsou již implementovány a je možné je využívat explicitním nastavením.

## Kapitola 4

# Návrh modelu důvěry pro sociální síť Facebook

Tato kapitola popisuje matematickou teorii vztahující se k navrženému modelu důvěry. Text je uveden se snahou o abstrakci veškerých implementačních detailů, aby bylo možné tento model naprogramovat v podstatě v libovolném programovacím jazyce. Z hlediska algoritmizace je pak důležité zejména uvažovat nad složitostí jednotlivých metod, jelikož některé testované profily mohou vyžadovat počty dotazů v řádu stovek nebo tisíců. I když Facebook API nabízí možnost agregace dotazů do dávky, vzniklé množství dat je nutné analyzovat a především porovnávat mezi sebou takovým způsobem, že se v některých případech dostáváme na hranici kubické složitosti.

### 4.1 Vyčlenění jednotlivých kontextů důvěry

Marsh důvěru ve své práci [21] rozděluje do tzv. *kontextů* v závislosti na oblastech, ve kterých zkoumané entitě důvěřujeme. Ve volném překladu „věřím svému bratrovi, že mě v pořádku odveze na letiště, ale rozhodně bych mu nesvěřil pilotování letadla”. Aby vůbec bylo možné matematickým modelem obsáhnout tak širokou oblast, jakou důvěra mezi lidmi bezesporu je, je nutné striktně vymežit konkrétní kontexty důvěry i v tomto oboru. Uplatňování tohoto principu v návrhu může vést ke zkreslenému pohledu na důvěru při pohledu na izolované kontexty (příklad hledejte např. v kontextu závislém na rozložení interakcí), v konečném důsledku však umožňuje flexibilnější práci s výsledným modelem a především jeho další vývoj pomocí vektoru priorit popsaného v následující části práce 4.2.

Postupem času při tvorbě vykrytalizovalo sedm oblastí, ve kterých lze pozorovat závislosti mezi interakcemi a důvěrou mezi entitami. I když se zde již nelze zcela oprostít od implementačních specifik sítě Facebook, stále jsou vybrány typy interakcí, u kterých lze nalézt ekvivalenty v ostatních sítích (viz 3). Aby byla zachována univerzálnost řešení, je třeba výsledek každého kontextu normalizovat. Pro jednoduchost byl zvolen uzavřený interval od 0 do 1. Uvádím zde seznam zmíněných oblastí (*kontextů*) i s krátkým popisem každé z nich. Mnoho poznatků je zde převzato z práce [24].

#### 4.1.1 Časové rozpětí interakcí

- **Důvod zařazení:** Časové rozpětí vzájemného kontaktu je zřejmě intuitivně první věcí, která každého při tvorbě modelu napadne. Čím déle se s někým znám, tím déle jsme mohli budovat vzájemnou důvěru. Autor je si však vědom možnosti zavádějících

případů, ve kterých mohlo dojít např. k interakci krátce po přidání do seznamu a poté k nedávné interakci, což by v tomto kontextu vedlo k důvěře konvergující k číslu 1. Proto je tomuto kontextu ve finální verzi modelu přiřazena jen nejnižší priorita.

- **Implementace v aplikaci:** Pro každého člověka v seznamu přátel vybere aplikace nejstarší a nejnovější interakci. Spočítá mezi nimi časový interval v jednotkách dnů a ten poté uloží a normalizuje.
- **Postup normalizace:** Tento kontext má nejjednodušší postup normalizace. Pouze dojde k podělení výsledné hodnoty délkou intervalu, pro který analýzu provádíme.

#### 4.1.2 Počet interakcí

- **Důvod zařazení:** Počet interakcí je velmi důležitou součástí modelu. *Interakcí* se v daném případě myslí jednostranný tok informací, v případě sítě Facebook tedy příspěvky na zdi, komentáře a označení „To se mi líbí“. Je zde však třeba dávat pozor na uživatele, kteří jsou výrazně aktivnější než ostatní. Jedná se např. o přátele, kteří sociální sítí skutečně žijí a proto by neměl nadměrný počet interakcí zvyšovat důvěru, spíše naopak. Zatím neexistuje spolehlivý způsob, jak tyto jedince odhalit (a jedná se o námět pro další výzkum), je však možné stanovit určitý limit, nad který už důvěra daného jedince nestoupne a neovlivní tak ani ostatní při normalizaci.
- **Implementace v aplikaci:** Algoritmus bere v potaz všechny komentáře, příspěvky na zdi a označení „To se mi líbí“ na zdi analyzovaného uživatele a na jeho fotkách. Pouze je spočítá a uloží si jejich počet.
- **Postup normalizace:** Normalizaci lze popsat rovnicemi

$$A = \frac{1}{n} \cdot \sum_{x=1}^n I_x \quad (4.1)$$

$$T_N(x) = \frac{I_x}{A + \frac{1}{n} \cdot \sum_{x=1}^n |A - I_x|} \quad (4.2)$$

kde  $I_x$  je počet interakcí s určitou osobou  $x$ ,  $A$  je průměrný počet interakcí a dělitel v rovnici 4.2 lze vyjádřit jako součet průměru a absolutní odchylky všech naměřených hodnot. Tento způsob výpočtu byl stanoven empiricky, jelikož poskytoval nejakurátnější výsledky v dané oblasti.

#### 4.1.3 Počet znaků

- **Důvod zařazení:** Práce [26] se do velké míry zabývala právě vlivem počtu znaků vyměněných mezi uživateli v roli ukazatele vzájemné důvěry. Ačkoliv lze tento faktor poměrně jednoduše a efektivně vypočítat a poskytuje i jakousi vypovídající hodnotu, v práci mu není přikládána tak velká váha. Je to zejména z důvodu odlišnosti aplikační domény. Zmíněná práce zkoumala vliv počtu znaků na odborném internetovém fóru, na kterém uživatelé diskutovali o hardware. Obecně lze opravdu říci, že pokud je komentář dlouhý, uživatel skutečně o dané problematice větší přehled může mít. Dle autorova názoru jde však na sociální sítí o zcela jiný typ interakce a délka komentářů nehraje takovou roli.

- **Implementace v aplikaci:** Aplikace eviduje všechny vyměněné texty mezi uživateli a sečte jejich celkovou délku. Z důvodu zachování soukromí uživatelů není samotný obsah zpráv nikde použit, přestože by se opět jednalo o potenciální téma pro další výzkum, namátkou např. využití emotikonů nebo slov z hovorového jazyka.
- **Postup normalizace:** Normalizace používá podobný způsob výpočtu jako v kontextu 4.1.2, pouze je zjednodušen použitím průměrné hodnoty. Stanovíme průměrný počet znaků, počítáme procentuální podíl. Pokud je číslo vyšší než průměr, je výsledek zaokrouhlen na hodnotu 1. Narozdíl od kontextu 4.1.2 je zde patrný nižší důraz na důležitost této oblasti. Přičtením absolutní odchylky bychom získali vyšší horní mez a tudíž i nastavili kritérium délky příspěvků výše.

#### 4.1.4 Pravidelnost interakcí

- **Důvod zařazení:** Pravidelnost interakcí odlišuje uživatele, kteří přispívají na zed' jen v určitém období, a uživatele, se kterými pravidelně komunikujeme na denní bázi. Může se stát, že s někým potřebujeme projednat detaily koncertu nebo oslavy narozenin a v konečném výsledku vyprodukuje s daným člověkem více interakcí než s některými přáteli, se kterými sice komunikujeme často, ale vyměníme si pouze pár komentářů a vidíme se místo toho v reálném životě. Tento faktor se snaží tato metrika zohlednit.

Za zmínku stojí ještě jedna věc. Projevuje se zde negativní jev popsáný na začátku kapitoly. Subjektivně se zdá, že tato metrika vrací zcela nerelevantní výsledky důvěry. Pokud má např. někdo jedinou interakci a ta proběhla zhruba v polovině analyzovaného období, jeho hodnota důvěry v tomto kontextu se blíží číslu 1. Je třeba si ale uvědomit, že se jedná o vypovídající hodnotu v tomto kontextu. Pokud chceme důvěru spočítat s větší abstrakcí, je třeba ji zasadit do kontextu ostatních částí modelu.

- **Implementace v aplikaci:** Původní návrh aplikace počítal s využitím metriky použité v práci [26]. Ve zkratce se jedná o převedení intervalů mezi po sobě jdoucími interakcemi do jednotek vteřin a poté jejich vynásobení. Daný součín nabývá maximálních hodnot pro sady interakcí, které jsou nejrovnoměrněji rozloženy po celé časové ose. Narážíme zde však na problém rozsahu datových typů, kdy ani po převedení do jednotek dnů nebylo možné vyvinout takovou metodu, aby v součinech nevracela velmi vysoká čísla. Tato metrika by tedy byla ze všech nejvhodnější, bylo by však nutné ji provozovat např. na procesorech schopných pracovat s rozšířeným rozsahem hodnot. Pro srovnání uvádím vzorec převzatý z citované publikace:

$$x_v^{\Delta T}(A, B) = \prod_{i=1}^{n-1} |t_{i+1} - t_i| \quad (4.3)$$

kde  $n$  patří do množiny přirozených čísel a značí počet intervalů, které počítáme (pokud došlo ke dvěma interakcím v určitém časovém období, rozdělí časovou osu na tři díly atd.) a  $t$  značí čas interakce. Laicky řečeno tak zde vynásobíme za sebou jdoucí intervaly, přičemž maximální hodnoty dosáhneme právě tehdy, pokud jsou interakce rovnoměrně rozděleny po časové ose. Platí zde ovšem důležité omezení, že součín obou činitelů musí být větší než 1. V modelu důvěry TrustNet [26] je tato podmínka vyřešena výpočtem v milisekundách.

Do výsledného modelu tedy byla zakomponována vlastní metoda. Pro každého přítele je nutné spočítat absolutní odchylku intervalů mezi interakcemi (jinými slovy, jak se v průměru liší od ideálního rozložení interakcí).

- **Postup normalizace:** Z absolutních odchylek všech přátel se spočítá průměr. Tento průměr je poté vynásoben číslem 2 (empiricky zjištěný koeficient) a použije se jako maximální hodnota pro normalizaci.

#### 4.1.5 Označení na fotografiích

- **Důvod zařazení:** Společnou fotografii lze označit jako jednu z nejvíce vypovídajících složek celkové důvěry. Do určité míry jde o krok zpět ze sociální sítě do reálného světa, neboť většinu fotografií bylo nutné pořídit při osobním kontaktu obou účastníků. Jelikož tedy označení na fotografii indikuje osobní kontakt, ve výsledném modelu je mu přiřazena poměrně velká důležitost. Facebook nedávno začal nabízet novou funkci, ve které navrhuje označení na fotkách i pokud se uživatel sám neoznačí. Jedná se opět o zajímavou, byť ne úplně etickou možnost rozšíření analýzy.

V reálné situaci na sociální síti je zde třeba vyřešit ještě jeden zkrslující element. Tím jsou fotky, na kterých je označeno více uživatelů, „než je zdrávo“. Jedná se například o vánoční přání, hromadné označování apod. Nejjednodušší variantou je zřejmě stanovení hranice počtu označení, po kterém už je fotografie z analýzy vyřazena. Druhou variantou je použití pouze fotografií, které nahrál nebo schválil jen sám uživatel. Sice tak přicházíme o určitou část relevantních dat, na druhou stranu máme k dispozici fotky s větším počtem účastníků (např. ze sportovních soutěží, projektů apod.). Tato možnost na použitém vzorku dat poskytovala relevantnější výsledky, při nasazení v širší oblasti uživatelů by však bylo moudré zvážit nasazení limitu označení, kontext je pak robustnější a ne tolik ovlivnitelný extrémními výkyvy.

- **Implementace v aplikaci:** V principu jde pouze o procházení seznamu metadat o fotografiích. Procházíme jednotlivá označení na fotce a každému uživateli, kterého v seznamu spolu s analyzovanou osobou najdeme, zvýšíme čítač o 1.
- **Postup normalizace:** Projdeme seznam všech přátel a najdeme nevyšší počet fotek. Ten použijeme pro porovnání s ostatními. I když se zdá, že může jedna vybraná hodnota poměrně zkrslvat, jedná se o záměr. V člověka, se kterým má uživatel nejvíce fotek, má zpravidla také nejvyšší důvěru. V širším nasazení by pak bylo možné např. vypočítat průměr tří nejvyšších hodnot.

#### 4.1.6 Členství ve skupinách

- **Důvod zařazení:** Před osvětlením tohoto kontextu je nejprve třeba definovat a blíže popsat pojem „skupina“. Po vzniku sociální sítě Facebook totiž skupiny převzaly funkci sdílení nejrůznějších vtipů, obrázků a rad do života. V dnešním modelu sociální sítě však reprezentují jakési komunity, ve kterých se sdružují lidé se společnými zájmy. Jako příklad uvedu skupinu lidí, kteří píšou inzeráty na společné dojíždění mezi dvěma městy. Tyto skupiny jsou vesměs neveřejné a členy musí pozvat buďto přímo správce skupiny, nebo ostatní členové.

Nabízí se otázka, do jaké míry lze brát v potaz velikost daných skupin. V modelu je uplatněn princip nepřímé úměry velikosti skupiny a její důležitosti. Pokud jsem

členem skupiny o 6 lidech, pravděpodobně jsme ji společně založili a máme toho více společného, než 250 členů onoho zmíněného „dojížděcího kroužku”.

- **Implementace v aplikaci:** Každý uživatel by měl mít ve svém profilu uvedeny skupiny, jejichž je členem. Přístupové API se však již liší, pokud jde o počet členů skupiny. Pokud počet členů chybí, zpravidla je možné toto číslo získat jinou cestou. Projdeme tedy seznam uživatelových skupin, zjistíme, které má s analyzovanou osobou společné, vypočítáme převrácenou hodnotu na základě počtu členů (skupina se dvěma členy bude mít hodnotu 0,5). Sečtením může vzniknout číslo v intervalu od 0 do nekonečna, které je třeba normalizovat.
- **Postup normalizace:** Normalizace probíhá stejně jako normalizace v kontextu označených fotografií. Platí pro ni i stejná doporučení ohledně robustnosti modelu v širším použití.

#### 4.1.7 Společné zájmy

- **Důvod zařazení:** Tento kontext jako jediný z celého modelu nemá nic společného se vzájemnými interakcemi mezi lidmi. Slouží k reprezentování společných zájmů, které však lidé nesdílí pomocí skupin, nýbrž pomocí tlačítka „To se mi líbí”, např. oblíbená značka elektroniky, oblíbený učitel ve škole nebo třeba podpora politických hnutí. Ačkoliv se může zdát, že se tato metrika do analýzy nehodí, s reálnými daty byla empiricky zjištěna překvapivá korespondence mezi hodnotami ostatních koeficientů. Laicky řečeno (a je tomu tak i v reálném životě, pouze je nečekané, že se tento princip tak výrazně promítá rovněž do virtuálního světa) lze říci, že čím více společných zájmů se člověkem máme, tím větší je pravděpodobnost, že si budeme navzájem věřit [25].
- **Implementace v aplikaci:** Implementace je prakticky totožná jako v předchozím kontextu. Pouze zde narážíme na problém mnohonásobně vyššího počtu položek, které potřebujeme zkontrolovat. Jedná se o seznam našich zájmů, seznam přátel a seznam zájmů pro každého z nich zvlášť. Dostáváme se tak na hranici kubické složitosti, což pro průměrně využívaný profil může lehce vygenerovat přes 2 000 000 porovnání. Jako určitá optimalizace se jeví vyřazení skupin, jejichž výsledná hodnota příliš neovlivní výsledek (toto číslo lze v analýze dynamicky nastavovat, v konkrétní implementaci jsou vyřazeny všechny skupiny nad 5000 členů).

V reálném nasazení skriptu by však bylo rozumné uvažovat o vypuštění tohoto kontextu úplně, popř. o jeho použití pouze v omezené míře. Ve vektoru priorit je mu totiž přiřazena nejnížší důležitost, zabere však přibližně polovinu výpočetního času celého modelu.

- **Postup normalizace:** Normalizace je totožná s normalizací v kontextu společných skupin.

#### 4.1.8 Další možné rozšíření modelu

Jako návrh na rozšíření modelu zde uvádím ještě osmý kontext důvěry. Jedná se o kontext, který nebyl nakonec implementován z důvodu omezení Graph API sociální sítě Facebook. Při migraci ze staršího rozhraní dosud nebyla tato funkce přenesena. Jedná se o zjištění celkového počtu přátel daného uživatele. Nejde ale jen o to, kolik přátel má námi analyzovaný

uživatel. Tento počet lze zjistit poměrně snadno, koneckonců pro každého z nich sestavíme rozsáhlou analýzu. Problém je v počtu přátel, které mají naši přátelé. Vycházím z předpokladu uvedeného v [25], a sice že lidé, kteří mají více společných vlastností, mají i větší pravděpodobnost vzájemné důvěry.

Profesor Dunbar ve své práci [12] zavádí pojem tzv. *Dunbarova čísla*. Jedná se o číselný údaj, který reprezentuje počet lidí, se kterými je člověk schopen udržovat vztahy na širší než společenské bázi. Společnost se podle něj vyvinula mnohem vyšším tempem, než vůbec evoluce stihla zaznamenat. Stále tak podle něj máme neokortex uzpůsobený pro vztahy s velmi omezeným počtem lidí (údaj se má pohybovat v rozmezí 100 – 230, nejnovější průzkumy hovoří o čísle 150, což je příhodně jak průměrná velikost neolitické vesnice, tak také současný průměrný počet přátel na Facebooku).

Navzdory poměrně přesné definici i průměrným údajům může být Dunbarovo číslo pro každého člověka značně individuální. Právě odchylka od všeobecně uznávaného průměru, ať už v kladném nebo záporném smyslu, může sloužit jako určitý indikátor podobné mentality uživatelů sociální sítě. Společenské dopady těchto odchylek a extrémní případy nejsou předmětem této práce.

## 4.2 Agregace kontextů

Pro uvedených sedm kontextů důvěry byla zjištěna jistá korespondence s reálnými emocemi v reálném světě. Otázkou zůstává, zda je vůbec možné tyto kontexty zkombinovat tak, aby odrážely celkovou důvěru. Marsh ve své zakládající práci [21] kombinuje jednotlivé kontexty důvěry prostým násobením mezi sebou. Tento způsob získávání výsledné důvěry je zde nevhodný zejména z důvodu nestejného přispění kontextů k výsledku. Již při formulování částí modelu bylo jasné, že různé části budou mít různou míru důležitosti. Tato míra je odvislá od mnoha faktorů, jmenovitě např. od dosaženého vzdělání, kultury, výchovy, společenského postavení nebo pohlaví. Model byl tedy pojat od začátku jako určitá skládanka, která umožňuje variabilní skládání jednotlivých součástí modelu pomocí tzv. *vektoru priority*. Matematicky se jedná pouze o uspořádanou množinu koeficientů, kterými lze přiřadit váhu každému kontext zvlášť:

$$P = (T_S, T_N, T_C, T_F, T_P, T_G, T_L) \quad (4.4)$$

kde  $T_{\{S,N,C,F,P,G,L\}}$  je koeficient priority pro daný kontext důvěry. Celkovou důvěru  $T$  lze pak vypočítat na základě následujícího vzorce:

$$T_x = \frac{S \cdot T_S + N \cdot T_N + C \cdot T_C + F \cdot T_F + P \cdot T_P + G \cdot T_G + L \cdot T_L}{S + N + C + F + P + G + L} \quad (4.5)$$

Písmena označující jednotlivé kontexty mají význam seřazený podle uvedeného pořadí kontextů v této práci, tedy:

- **S (Span):** časové rozpětí interakcí
- **N (Number):** počet interakcí
- **C (Characters):** počet znaků
- **F (Frequency):** pravidelnost interakcí



- **P (Photos):** označení na fotografiích
- **G (Groups):** členství ve skupinách
- **L (Likes):** společné zájmy
- **T (Trust):** celková důvěra ve člověka  $x$

V případě adaptace tohoto modelu do dalšího výzkumu je samozřejmě možné tento vektor priorit modifikovat a přiřadit tak kontextům vlastní váhy. Samotná implementovaná aplikace využívá následující hodnoty, které jsou podrobně vysvětleny vždy u popisu konkrétního kontextu.

$$P = (1, 3, 2, 2, 1, 2, 3) \tag{4.6}$$

## Kapitola 5

# Implementace modelu

Ačkoliv je celý matematický model stavěn tak, aby byl co možná nejméně implementačně závislý na konkrétním programovacím jazyce nebo platformě, v krátkosti zde uvádím popis samotné implementace a úskalí, která mohou vyvstat při případné implementaci do jiného jazyka. Rovněž jsou zde popsána některá specifika rozhraní OpenGraph [1], která je třeba brát v potaz.

### 5.1 Výběr programovacího jazyka

Požadavkem na programovací jazyk pro tvorbu aplikace pro výpočet modelu důvěry byla především existence jednoduché knihovny pro komunikaci HTTP, rychlé zpracování formátu JSON a efektivní práce s textovými řetězci. Uvedené požadavky splňuje prakticky jakýkoliv skriptovací jazyk. Zvolil jsem jazyk Python ve verzi 3. Důvodem byla zejména existence modulu využívajícího přímo API Facebooku. Nebyl sice použitelný přímo v aplikaci, bylo však možné se inspirovat ve veřejně přístupném zdrojovém kódu funkce pro přihlášení ke službě. Python také efektivně pracuje s velkými datovými strukturami a je možné využitím slovníků a množin rychle přistupovat k velkému množství dat, což bylo pro aplikaci klíčové.

### 5.2 Přístup k datům

Síť Facebook nabízí pro vývojáře nové API nazvané Open Graph. Narozdíl od dřívějších metod přístupu k datům je možné Open Graph napojit v podstatě na jakýkoliv programovací jazyk, jelikož komunikuje pouze prostřednictvím HTTP dotazů. Skládáním jednotlivých částí řetězce je možné přesně specifikovat, jaké informace nás zajímají, popř. také upřesnit časový rozsah. Server potom vrací objekty v jazyce JSON, které lze lehce analyzovat ve skriptovacím jazyce. Autentizaci řeší protokol OAuth 2.0. Pro aplikaci je podstatný fakt, že k některým informacím uživatele nelze přistupovat bez jeho povolení. Je nutné vygenerovat tzv. *access token*, který je rovněž vložen do dotazu na Open Graph API a funguje jako forma hashovaného hesla. Při generování access tokenu je možné podrobně nastavit, jakými právy bude držitel tokenu disponovat.

Jednou možností sběru informací bylo zaregistrovat aplikaci jako součást centra Developers, přičemž by uživatelé mohli přistupovat přímo k aplikaci a automaticky jí potvrzovat oprávnění. Tato možnost není vhodná ze dvou důvodů. Tím nejzávažnějším je patrně fakt, že Facebook formy sběru dat zaregistrovanou aplikací přímo na svých stránkách potlačuje a

dané aplikace staví mimo podmínky užívání služby. Druhým důvodem je nutnost placeného hostingu pro aplikaci.

Druhou možností sběru je pak spolupráce se samotnými uživateli. Facebook na stránkách pro vývojáře nabízí možnost generování access tokenu s podrobným nastavením oprávnění daného tokenu. Je tak zcela transparentní, jaká práva bude aplikace využívat (např. jak již bylo zmiňováno nebude mít přístup k soukromým zprávám). Access token lze jednoduše předat pomocí Instant Messagingu. Z hlediska podmínek služby Facebook povoluje generovat access token pro vlastní výzkumné potřeby. Odpadají tedy nevýhody komerční aplikace. Skript lze navíc spouštět vždy na stejném počítači a výsledky nemusí putovat internetem – zůstanou na jednom veřejně nepřístupném místě.

### 5.3 Potřebná práva pro access token

V této podkapitole budou v krátkosti uvedena práva potřebná pro bezproblémový běh aplikace pro nového uživatele. Uživatelé sami v nabídce zatrhávají práva, která jsou ochotni aplikaci poskytnout (nabídka je zobrazena na obrázku 6.1). Je tak zcela transparentní, co všechno vygenerovaný token umožňuje. Následuje seznam práv podle terminologie Facebook API:

- **user\_about\_me:** Základní informace o profilu uživatele. Umožňuje např. získat seznam přátel nebo jméno uživatele pro zařazení analýzy do databáze.
- **user\_photos:** Přístup k fotografiím, které uživatel sám nahrál nebo na kterých je označen. Pomocí odkazu lze také zobrazit fotografie samotné, tato možnost však není v aplikaci využívána. Důležité jsou pouze komentáře a označení „To se mi líbí“ na fotkách.
- **user\_groups:** Skupiny, do kterých uživatel patří. Obecně se jedná o menší seskupení než v *Likes*, jelikož jde spíše o skupiny lokálního charakteru. Jako příklad lze uvést taneční spolky, skupinu dojíždějících do stejného města nebo třeba pracovní tým.
- **user\_likes:** Stránky, na kterých uživatel zvolil označení „To se mi líbí“. Pro některé uživatele mohou podobné dotazy dosahovat datového objemu přes 300 kB.
- **user\_photo\_video\_tags:** Označení uživatele na fotografiích a ve videonahrávkách. Tato práva mají omezenou míru působnosti, pokud se uživatel na fotkách neoznačuje (popř. jej neoznačují přátelé).
- **user\_status:** Přístup k příspěvkům na zdi. Jedná se o jádro textové části analýzy a je proto pro aplikaci velmi důležité, aby tato možnost byla do tokenu zahrnuta.
- **read\_stream:** Nachází se v kolonce „Extended Permissions“, je nicméně klíčové tuto možnost zaškrtnout pro profily, které dosud nepřešly na typ profilu Timeline [4].

### 5.4 Opakované dotazy na API

Velkou nevýhodou aplikace v počátečních fázích vývoje byla nutnost při každém spuštění analýzy stahovat všechna potřebná data. Jednotlivé dotazy sice negenerují příliš velké soubory dat, dotazů se však při rozsáhlejší analýze může vyskytnout i přes 1000. Dostáváme se

tedy do řádů megabajtů dat, což může na pomalejším připojení poměrně drasticky snižovat efektivitu.

Bylo tedy nutné funkci pro zpracování dotazu zapouzdřit tak, aby nejprve zkontrolovala, zda dotaz už nemá uložený někde do souboru. Při prvním spuštění tomu tak pochopitelně není, funkce tedy dotaz provede a výsledek uloží nejen do své proměnné, ale rovněž do souboru nazvaného podle adresy, která dotaz reprezentuje. Při vývoji a ladění je výhodnější ukládat data v této čitelné formě. V nasazení v reálné síti by pak bylo ideální názvy souborů hashovat pomocí algoritmu MD5, aby nebylo zřejmé, který soubor obsahuje která data.

Kešování dat také přináší výhodu při analýze několika blízkých přátel. Předpokládá se, že některé dotazy zde mohou být stejné (např. jsou členy stejné skupiny nebo mají společné přátele) a tím pádem se čas potřebný pro analýzu s vzrůstajícím počtem uživatelů rovněž snižuje.

## 5.5 Struktura aplikace

Při návrhu aplikace byla použita metodika objektového programování, byť se jedná v podstatě o sekvenční zpracování dat a otázku, kam výsledná data uložit a kde je zpracovat. Základním stavebním kamenem návrhu je třída `Analysis`, která zapouzdřuje analýzu pro jednoho uživatele. V aplikaci se sice vyskytují globální proměnné a samostatné funkce, ty však mají obecný charakter a jsou společné pro všechny analýzy dohromady. Vysvětleny budou podrobněji v kapitole 5.5.1. Každá instance objektu `Analysis` potom obsahuje seznam objektů třídy `Friend`, které reprezentují přátele analyzovaného uživatele. Právě pro ně je počítána důvěra. Návrh uzavírá ekvivalent funkce `main` z jazyka C [16], v té době k vytvoření instance třídy `Analysis` a přípravě kontextu pro aplikaci.

### 5.5.1 Samostatné funkce

Samostatné funkce v aplikaci lze rozdělit do dvou základních skupin. Do první skupiny lze zařadit funkce, které byly implementovány z praktického hlediska – totiž aby skript byl přenositelný bez instalace zvláštních knihoven, aby fungoval pouze na čisté instalaci jazyka Python 3. Jedná se zejména o funkce použité pro statistické výpočty (medián, absolutní odchylka, rozptyl a průměr [22]). Dále pak funkce pro převody časových značek do Unixového formátu [10] nebo převod data ve formátu řetězce do formátu vhodného pro práci v programovacím jazyce.

Druhou skupinu tvoří funkce implementačně závislé na této konkrétní aplikaci, je však možné je poměrně jednoduše zobecnit na jakoukoliv práci s API Facebooku. Setkávají se zde dva principy z návrhu, kešování již jednou zadaných výsledků a pak také omezení na 600 dotazů za 600 vteřin pro běžné použití. Funkce tedy slouží k zasílání dotazů, případně k načtení již hotových dat ze souboru, dále pak ke kontrole dosud zaslaného počtu dotazů a případného čekání, dokud se limit zaslaných dotazů opět neuvolní.

### 5.5.2 Třída `Friend`

Jedna instance této třídy reprezentuje jednoho přítele v sociální síti Facebook. Atributy třídy zde reprezentují vlastnosti a interakce s danou osobou, tedy např. společné fotografie, počet označení „To se mi líbí“, záliby, skupiny nebo komentáře. Každá instance pak obsahuje 7 atributů reprezentujících důvěru v příslušném kontextu a k tomu jeden atribut reprezentující celkovou agregovanou důvěru. Metody třídy se omezují na práci s atributy

(aby bylo dosaženo objektového zapouzdření) a na výpočet kontextové důvěry, přičemž je nutné metodám dodávat při volání dodatečné informace z volajícího objektu. Jednoduše řečeno to znamená, že důvěra je odvozena od vlastností konkrétní analýzy a snaží se tak reflektovat důvěru ve vztahu ke konkrétní analyzované osobě. Pokud někdo nemá mnoho zveřejněných fotografií, budou v tomto kontextu znevýhodněni všichni přátelé podobně.

### 5.5.3 Třída `Analysis`

Třída `Analysis` zapouzdřuje veškerý kontext pro konkrétní analýzu jednoho uživatele. V případě, že budeme provádět analýzu více uživatelů, vytvoří se pro každého nová instance tohoto objektu. Globální proměnné (počet dotazů na Facebook API, popisovače souborů, názvy souborů) však zůstávají neměnné pro běh aplikace, neboť jsou vázány na konkrétní souborový systém a připojení k síti. Jsou tedy společné pro všechny analyzované uživatele.

V instanci třídy `Analysis` jsou tedy uložena data spojená s konkrétním uživatelem, u kterého analýzu provádíme, nikoliv jeho přátel. Jedná se např. o soubor statusů, fotky uživatele samotného, jeho zájmy. Součástí je také slovník dvojic *id – instance objektu Friend*, který umožňuje snazší přístup ke všem přátelům. Instanci objektu lze po jejím vytvoření aktivovat metodou, která spustí veškerý sběr dat, vytvoří instance třídy `Friend`, pomocí jejích metod pro ně spočítá důvěru a výsledek dle zadání vypíše na příkazovou řádku nebo do souboru.

### 5.5.4 Volající funkce

Jedná se o vstupní bod celé aplikace. Stará se pouze o vytvoření adresářové struktury a otevření souborů (důležitý je zde soubor pro ukládání již použitých dotazů, ze kterého lze odvodit počet dosavadních dotazů v 600sekundovém intervalu), analýzu vstupního souboru, který obsahuje *access token* (viz 5.2), o vytvoření instance objektu `Analysis` a o její následné spuštění. O výpis se již stará tento objekt sám.

## Kapitola 6

# Experimentální výsledky

Funkčnost modelu důvěry na sociálních sítích byla ověřena průzkumným šetřením na reálném vzorku dat ze sociální sítě Facebook. Vzhledem k plánovanému počtu respondentů nebylo možné využít metod kvantitativního výzkumu. Počet respondentů byl odvislý právě od důvěry respondentů k osobě provádějící průzkum. Zde byla ověřena teorie v praxi, jelikož při nedostatku institucionalizované důvěry je poměrně složité respondenty přesvědčit o korektním zacházení s osobními údaji.

### 6.1 Návrh průzkumného šetření

Podle Hendla jsou metody přírodních věd vzorem kvantitativního přístupu k výzkumu v sociálních vědách. Hendl předpokládá, že lidské chování lze do jisté míry měřit a předpovídat (následující odstavce čerpají z díla [15]). Průzkumné šetření se týkalo 18 respondentů vybraných náhodně ze skupiny méně známých uživatelů této sociální sítě ve věkové kategorii od 17 do 30 let. Byla zde rovnoměrně zastoupena obě pohlaví (9 mužů a 9 žen). Analýza byla prováděna pro časové období od 1.4.2011 do 1.5.2012. Výsledek funkčnosti modelu zhodnotili respondenti v kvantitativním průzkumném šetření s využitím dotazníkové metody. Dotazník byl sestaven pomocí uzavřených otázek s využitím škálování.

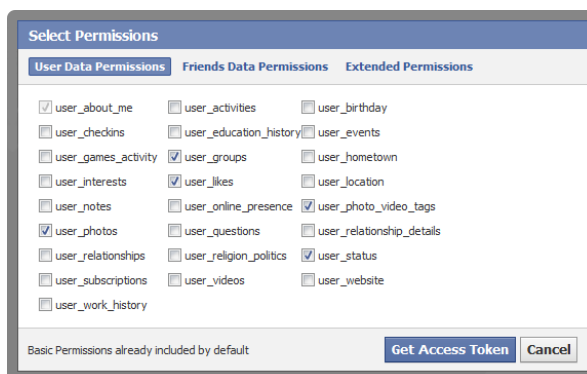
Kvantitativní průzkumné šetření i výzkum využívají náhodné výběry, experimenty a strukturovaný sběr dat pomocí testů, dotazníků nebo pozorování. Ve výzkumu lze získaná data statistickými metodami analyzovat, popisovat a případně ověřovat pravdivost našich představ o vztahu sledovaných proměnných, v tomto konkrétním případě verifikovat korepondenci reálných výsledků s očekávanými výsledky produkovanými navrženým modelem. Verifikace zde byla provedena metodou experimentu s následným zaznamenáním zkušenosti jedinců. Výhodou kvantitativního výzkumu je reprezentativnost, přesnost měření a do budoucna možnost využít velkých datových souborů, přičemž lze statisticky ověřit spolehlivost, pozorování a měření.

Výsledek kvantitativního výzkumu lze při validním vzorku zobecnit na populaci, a tak validovat teorii. Nevýhodou zvolené metody je přílišná abstraktnost a omezení na reduktivní způsoby získávání dat.

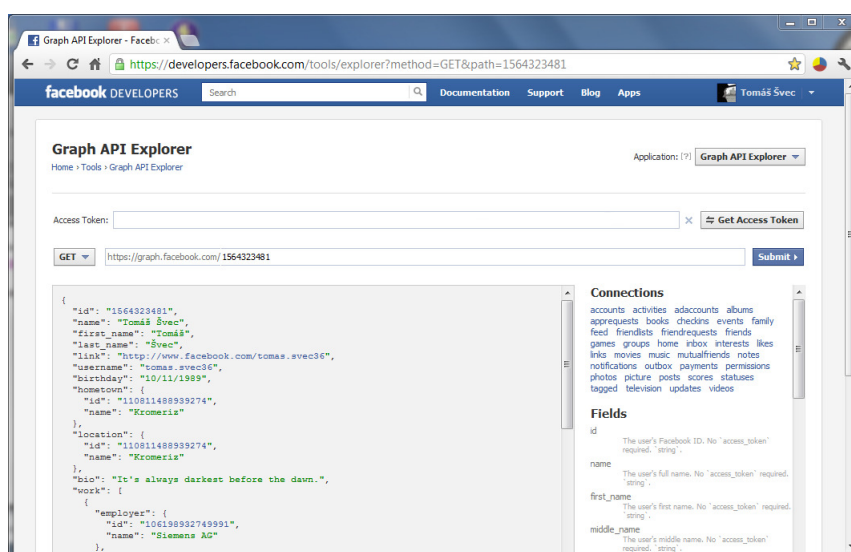
### 6.2 Průběh šetření

Na sociální síti Facebook byli pro šetření vybráni méně známí uživatelé, kteří projevíli ochotu se do šetření zapojit. Dalším požadavkem pro účast v šetření byla také míra aktivity daného

profilu. Byl sestaven krátký návod, podle kterého si uživatelé mohli vygenerovat vlastní access token (teoretický podklad k tomuto tématu lze nalézt v části 5.2, nabídka na obrázku 6.1). Obrázek 6.2 představuje rozhraní, ve kterém uživatelé pracovali. V zasílaném dokumentu byl přítom kladen důraz na informování uživatelů o bezpečnosti a o nakládání s jejich údaji. Krátce byl vysvětlen rovněž princip access tokenu, který bylo nutné srozumitelně popsat jen pomocí velmi omezené technické terminologie.



Obrázek 6.1: Výběr oprávnění přístupového tokenu.



Obrázek 6.2: Rozhraní pro vývojáře.

Po vygenerování tokenu začíná běžet časový limit pro jeho použití, bylo tedy nutné analýzu provádět dynamicky spolu s danými uživateli. Po dobu běhu skriptu (v řádu minut) dostali uživatelé za úkol si v duchu odpovědět na následující otázku:

Kterým deseti lidem na Facebooku nejvíce věříte?

Dalším krokem bylo zapsání odpovědi na tuto otázku na kousek papíru. Velice důležitým faktorem zde byl fakt, že uživatelé nemuseli přímo sdělovat svůj počáteční odhad. Zřejmě by

bylo obtížné nalézt mnoho účastníků, kteří by člověku svěřili osoby, kterým na Facebooku důvěřují (i z tohoto důvodu by měly výsledky běhu skriptu bez ohledu na přesnost korespondence s realitou uchovány pouze pro účastníky průzkumu). Po dokončení běhu skriptu byl textový soubor s výsledky analýzy vždy zaslán danému uživateli v archivu zabezpečeném heslem. Obsahem souboru byl seřazený seznam přátel podle vypočtené hodnoty důvěry v uzavřeném intervalu od 0 do 1. Uživatel dostal za úkol porovnat výsledky s realitou.

Finálním krokem průzkumného šetření pak bylo vyplnění dotazníku vytvořeného na službě Google Docs [5], jehož znění a jednotlivé otázky jsou rozebrány v následující části práce 6.3.

## 6.3 Otázky v dotazníku

Jak již bylo řečeno, dotazník byl sestaven pro průzkumné šetření, nikoliv pro výzkum. Za účelem výzkumu by bylo nutné sestavit širší okruh otázek, získat pro vyplnění dotazníku mnohem více respondentů a také jej odpovídajícím způsobem vyhodnotit s ohledem na důvěru z hlediska sociální psychologie. Jelikož tato kapitola není stěžejním bodem práce, byly využity právě otázky pro průzkumné šetření.

### 6.3.1 Otázka 1

- **Znění otázky:** Kolik lidí na Vašem předem připraveném seznamu se opravdu vyskytlo v první desítce?
- **Možné odpovědi:** 0 - 10
- **Odůvodnění zařazení:** Jedná se o klíčovou otázku dotazníku. Jelikož by bylo obtížné porovnávat důvěru všech jednotlivých přátel daného respondenta, byla vybrána alespoň reprezentativní část stupnice. Model funguje tím lépe, čím aktivněji respondent svůj profil využívá. Může tak vycházet z více informací. Předpokladem tedy je, že právě pro prvních deset uživatelů budou výsledky nejpřesnější a vhodné k porovnání.

### 6.3.2 Otázka 2

- **Znění otázky:** U kolika lidí pro Vás model výrazně selhal?
- **Doplňující komentář:** Např. jde o kamarády, kteří Vám píšou neustále na zeď otravné komentáře a model je zařadil mezi vysoce důvěryhodné.
- **Možné odpovědi:**
  - 0 - 2,
  - 3 - 5,
  - 6 - 8,
  - 9 - 11,
  - 12 a více.
- **Odůvodnění zařazení:** Tato otázka je spíše doplňujícího charakteru. Zajímavostí vytvořeného modelu totiž byl poměrně nízký výskyt nevhodných hodnot v obou směrech. Jinými slovy, pokud někdo nadšeně uživateli komentoval statusy, model jej stále vyloučil z množiny nejdůvěrnějších přátel díky absenci interakcí v jiném kontextu.



### 6.3.3 Otázka 3

- **Znění otázky:** Jakým akcím mezi přáteli na Facebooku přikládáte jaký význam? Číslo 1 znamená nedůležité interakce, číslo 5 nejdůležitější.
- **Doplňující komentář:** Svou odpovědí můžete ovlivnit celkový model důvěry.
- **Možné odpovědi:** Hodnoty 1 - 5 pro následující kategorie:
  - soukromé zprávy,
  - komentáře,
  - označení „To se mi líbí”,
  - společné fotografie,
  - společné skupiny,
  - pravidelnost interakce.
- **Odůvodnění zařazení:** Důležitost této otázky vyplývá z části 4.2. Model důvěry je navržen tak, aby bylo možné jej parametrizovat pro danou situaci, věkovou skupinu, nebo jej obecně přizpůsobit kritériím, která by hodnotu důvěry mohla ovlivnit. Problémem je fakt, že priority jednotlivých koeficientů určuje vždy osoba provádějící analýzu na základě svého přesvědčení. Tato otázka se snaží odhalit mínění respondentů a zakomponovat je do modelu, přestože respondentům princip modelu nemusí být jasný, nebo se o něj zkrátka nezajímají.

## 6.4 Výsledky průzkumného šetření

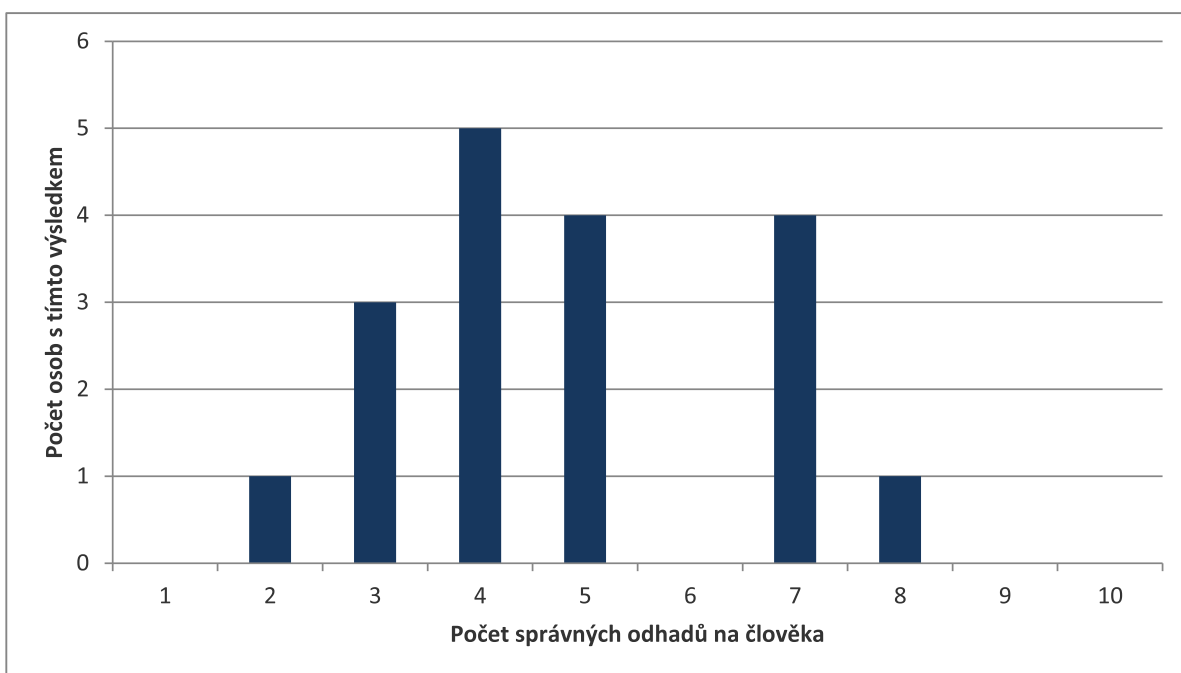
Výsledné odpovědi všech 18 respondentů na každou z položených otázek jsou shrnuty do následujících grafů. Pro každý graf jsou rovněž stručně popsány získané poznatky.

### 6.4.1 Výsledky pro otázku 1

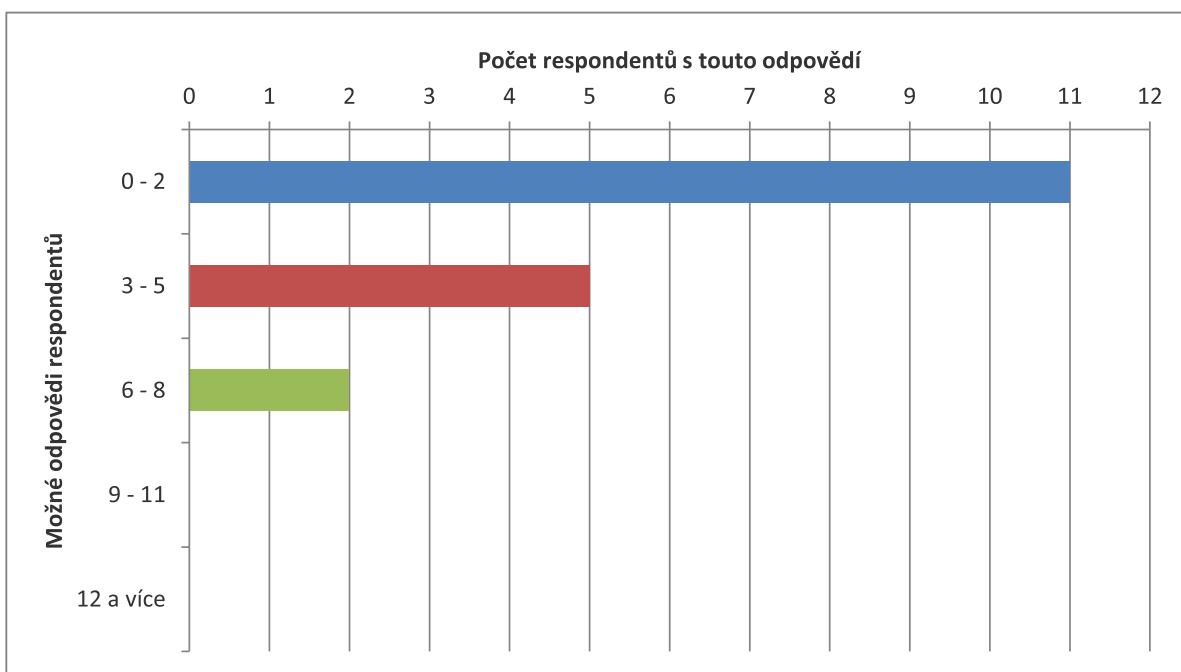
Otázka 1 byla nejdůležitější otázkou celého šetření. Hodnotila přímo korespondenci výsledků implementovaného modelu s realitou. Výsledné hodnoty zanesené do histogramu 6.3 tvoří mírně deformovanou Gaussovu křivku. Ukazuje se, že lze vysledovat fluktuaci modelu do obou směrů, ve většině případů se však blíží k průměru, který zde činí 4,83 správně odhadnutých nejdůvěryhodnějších přátel na respondenta.

### 6.4.2 Výsledky pro otázku 2

V otázce 2 šlo zejména o případy extrémního selhání modelu. Respondenti měli možnost určit množství osob, u kterých výsledná důvěra, kterou analýza určila, byla vyloženě chybná. Vzhledem k odpovědím ve formě intervalů, nikoliv konkrétních čísel, nebylo možné spočítat spolehlivě průměr ze získaných hodnot. V tomto případě by však efekt byl sporný. Jak ukazuje histogram 6.4, většina respondentů (v 11 případech byla odpověď „0 - 2”) nenalezla v modelu téměř žádné nepřírozené výchyly, ve kterých by model selhával.



Obrázek 6.3: Úspěšnost aplikace podle hodnocení respondentů.



Obrázek 6.4: Znatelné odchyly od reálné důvěry.

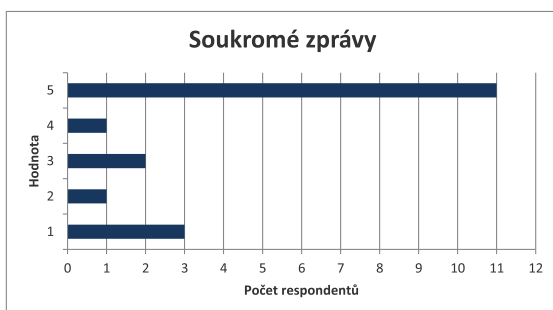
### 6.4.3 Výsledky pro otázku 3

Vyhodnocení odpovědí na otázku číslo 3 není tak jednoznačné, jako u předchozích otázek. Výsledné hodnoty měly sloužit zejména ke kontrole správně zvolených hodnot vektoru priorit. Kolektivní názor zde však nemusí znamenat, že původní hypotéza byla chybná. Nelze předpokládat, že se důvěra řídí zcela podle všeobecně uznávaných norem a z nich odvozených koeficientů. Pro úplnost následuje tabulka 6.1 s porovnáním odpovídajících kontextů ve vektoru priorit a průměrných hodnot získaných z průzkumného šetření.

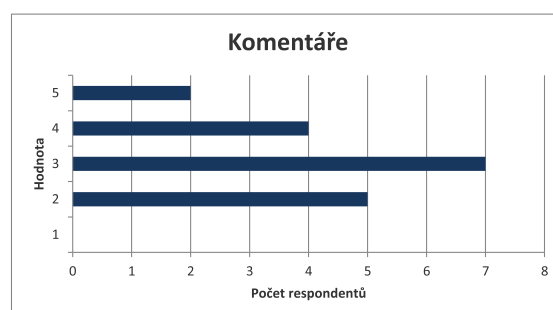
Název akce v šetření	Průměrná důležitost	Odpovídající kontext	Vektor priorit
Soukromé zprávy	3,89	–	–
Komentáře	3,17	Number	3
„To se mi líbí“	2,61	Likes	1
Společné fotky	3,28	Photos	2
Společné skupiny	2,67	Groups	2
Pravidelnost interakce	3,61	Frequency	3

Tabulka 6.1: Srovnání názorů účastníků šetření a použitého vektoru priorit.

V tabulce 6.1 jsou nejdůležitější hodnoty s nejvyšší průměrnou důležitostí. Ty přibližně odpovídají kontextům s nejvyšší hodnotou ve vektoru priorit. Nejlepší ohodnocení obdržely soukromé zprávy. Odpovědi respondentů tedy podporují teorii, že by soukromé zprávy do modelu přinesly zpřesnění hodnot důvěry. Pro srovnání následují histogramy jednotlivých akcí v šetření (6.5, 6.6, 6.7, 6.8, 6.9 a 6.10).



Obrázek 6.5: Soukromé zprávy.



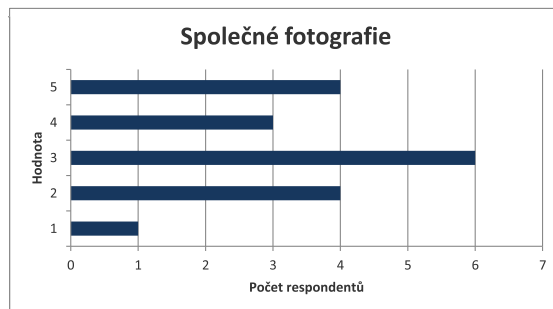
Obrázek 6.6: Komentáře.

## 6.5 Vyhodnocení šetření

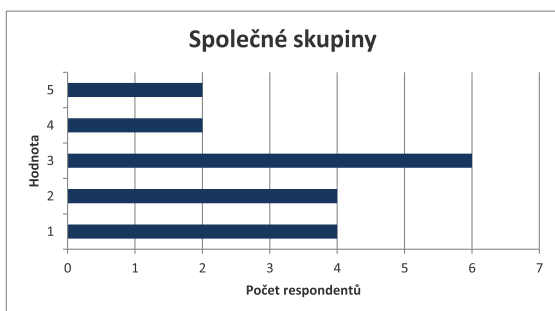
Provedené průzkumné šetření na vzorku uživatelů sociální sítě ve většině případů potvrdilo stanovené hypotézy. V otázce č. 1 je evidentní prostor pro zlepšení modelu, neboť mnohdy vychází ve výsledku aktivnější uživatelé, kteří si ale tímto chováním nutně nemusí získat přízeň analyzovaného respondenta. Na druhou stranu se jedná o určitou paralelu se situacemi v reálném životě, ve kterých se lidé chovají přátelsky, jen aby něčeho dosáhli, popř. protože se nejedná o reciproční vztah. Odhalit tento způsob chování představuje mnohdy



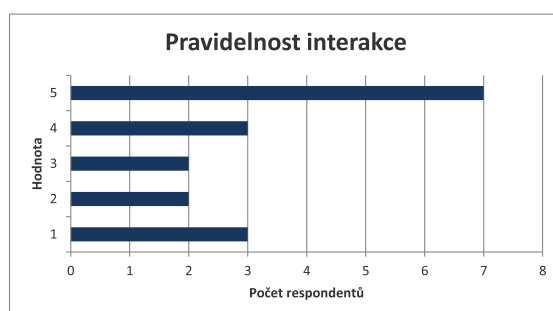
Obrázek 6.7: Označení „To se mi líbí”.



Obrázek 6.8: Společné fotografie.



Obrázek 6.9: Společné skupiny.



Obrázek 6.10: Pravidelnost interakce.

problém i mimo digitální svět, tudíž je krajně nepravděpodobné, že by tuto problematiku byla schopna obsáhnout oblast umělé inteligence.

Z odpovědí na otázku č. 2 vyplývá, že model zásadním způsobem neselhává a skutečně poskytuje relevantní výsledky, byť mohou být v některých případech nepřesné a plně neodrážet realitu. Porovnáním odpovědí na otázku č. 3 byla ověřena alespoň rámcová relevance zvolených koeficientů do vektoru priorit. V případě dostatku ochotných respondentů by bylo zajímavé sesbírat větší množství odpovědí na otázku 3 a neřídít se při parametrizaci modelu pouze vlastní intuicí, nýbrž brát v potaz veřejné mínění o dané problematice, což má svá pro a proti a nejrozumnějším způsobem je pravděpodobně kombinace obou zmíněných postupů.

Jelikož jsem se s respondenty znal osobně, narazil jsem v závěru šetření na zajímavý poznatek. Přesnost výsledků modelu byla ve velké většině případů přímo úměrná tomu, jak dotyčný sociální síť vnímá z osobního hlediska. Nejedná se zde o počet interakcí nebo objem informací zveřejněných na síti (byť by se tato domněnka zdála na první pohled logická), nýbrž o míru adaptace sociálních sítí jako prostředku reálné komunikace, nikoliv jako služby pro zábavu. Ověření tohoto poznatku však zasahuje hluboko mimo rámec technických věd.

# Kapitola 7

## Závěr

Cílem práce bylo analyzovat současnou situaci v sociálních sítích z hlediska interakcí, navrhnout model důvěry pro sociální síť, implementovat jej a otestovat jeho korespondenci s reálným světem. Jednotlivé kapitoly práce dokumentují kroky, které byly úspěšně provedeny. Nejlépe pro zhodnocení práce poslouží experimentální výsledky, jelikož jsou přímo odvislé od úspěšného splnění předchozích stanovených bodů.

Za hlavní výstup provedeného průzkumného šetření lze považovat následující fakt: Respondenti hodnotili úspěch modelu ve sledovaném výřezu reality průměrnou úspěšností 48,3%. Na jednu stranu se může zdát, že model v polovině případů selhal i s přihlédnutím k faktu, že byl hodnocen v oblasti, pro kterou byl stvořen a optimalizován, tedy při použití k nalezení deseti nejdůvěryhodnějších přátel.

Na druhou stranu byly modelu dodávány pouze informace, které lze neopatrným nastavením zabezpečení na sociální síti získat pro velké procento uživatelů zcela bez přístupových práv. V nejbezpečnějším možném případě tyto informace vidí „jen“ přátelé. Jejich průměrný počet se pohybuje okolo hodnoty 150 [25]. Je otázkou, nakolik by uživatele sociální sítě zneklidnil fakt, že polovinu jeho nejdůvěryhodnějších přátel lze odhadnout pouze z informací, které dává k dispozici minimálně všem svým ostatním přátelům.

Za nejprínosnější principy pro analýzu sociální sítě lze v práci považovat využití sociálních věd (zejména v oblasti průzkumného šetření), flexibilitu a škálovatelnost vytvořeného modelu. Díky analýze interakcí v sociálních sítích byl kladen důraz na nezávislost vytvořeného modelu na konkrétní síti. V oblasti aplikace důvěry a reputace pak model zavádí tzv. vektor priorit, který umožňuje agregovat jednotlivé kontexty důvěry s různou vahou v závislosti na jejich důležitosti.

Pro další vývoj projektu existuje mnoho podnětů. Nabízí se rozšíření modelu o další kontexty, např. hodnocení důvěry na základě podobného počtu přátel, pokud bude tato funkce do Graph API v nejbližší době implementována. Soukromé zprávy by rovněž podle výsledků průzkumného šetření mohly výrazným způsobem přispět k přesnosti modelu, stírá se zde však výhoda dostupnosti veřejných informací pro analýzu.

Dalším návrhem na zlepšení modelu je adaptace vektoru priorit pro konkrétního uživatele. Na základě aktivity v daném kontextu by byla kontextu přiřazena odpovídající důvěra. Poslední podnět směřuje k formátu výstupu aplikace. Místo informací v běžném textovém souboru by bylo vhodné generovat data ve formátu HTML pro snazší orientaci uživatelů, kteří třeba nepatří do informatické obce. Z osobní komunikace s respondenty vyplynulo, že poptávka po rozšíření modelu mezi veřejnost rozhodně existuje.

# Literatura

- [1] Graph API [online]. <https://developers.facebook.com/docs/reference/api/>, 2011-04-02 [cit. 2012-05-01].
- [2] eBay [online]. <http://www.ebay.com/>, 2011-05-01 [cit. 2011-05-02].
- [3] Social network [online]. [http://en.wikipedia.org/wiki/Social\\_network](http://en.wikipedia.org/wiki/Social_network), 2011-11-04 [cit. 2011-11-12].
- [4] Timeline [online]. <http://www.facebook.com/about/timeline>, 2012-01-03 [cit. 2012-05-06].
- [5] Google Docs [online]. <http://docs.google.com/?hl=cs>, 2012-04-30 [cit. 2012-05-06].
- [6] Afanasjew, M.: HistoryStats Miranda Plugin [online]. <http://addons.miranda-im.org/details.php?action=viewfile&id=2534>, 2008-04-16 [cit. 2012-01-31].
- [7] Baker, S.: *Numerati*. Brno: Computer Press Brno, 2009, iISBN 978-80-251-2444-4.
- [8] Barnes, J. A.: Class and committees in a Norwegian island parish. In *Human Relations* 7, London School of Economics, University of London, England, 1954, s. 39–58.
- [9] van Belleghem, S.: Social media around the world 2011 [online]. <http://www.slideshare.net/stevenvanbelleghem/social-media-around-the-world-2011>, 2011-09-13 [cit. 2011-11-12].
- [10] Braden, R.: Requirements for Internet Hosts – Application and Support [online]. <http://www.faqs.org/rfcs/rfc1123.html>, 1989-10-01 [cit. 2012-05-01].
- [11] Deutsch, M.: Cooperation and Trust: Some Theoretical Notes. In *Jones, M. R. (ed), Nebraska Symposium on Motivation*, Nebraska University Press, 1962.
- [12] Dunbar, R.: *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1998, iISBN 0674363361.
- [13] Giddens, A.: *The Consequences of Modernity*. Polity Press, Oxford, UK, 1990, iISBN 0745607934.
- [14] Gray, E. L.: *A Trust-Based Reputation Management System*. Dizertační práce, Trinity College, University of Dublin, Dublin, 2006.
- [15] Hendl, J.: *Kvalitativní výzkum: Základní metody a aplikace*. Praha: Portál s.r.o, 2005, iISBN 80-7367-040-2.

- [16] Herout, P.: *Učebnice jazyka C*. Kopp, 2009, iISBN 978-80-7232-383-8.
- [17] Josang, A.; Ismail, R.; Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. In *Decision Support Systems Volume 43, Issue 2*, Amsterdam: Elsevier Science Publishers B. V. Amsterdam, březen 2007.
- [18] Kolektiv autorů: *Akademický slovník cizích slov*. Praha: Academia, 2001, iISBN 80-200-0607-9.
- [19] Kolektiv autorů: *Ottova všeobecná encyklopedie ve dvou svazcích*. Ottovo nakladatelství Praha, 2003, iISBN 80-7181-959-X.
- [20] Machulka, T.: *Metody analýzy sociální sítě pro dolování znalostí*. Bakalářská práce, FIT VUT v Brně, Brno, 2011.
- [21] Marsh, S. P.: *Formalising Trust as a Computational Concept*. Dizertační práce, University of Stirling, Stirling, UK, 1994.
- [22] Novovičová, J.: *Pravděpodobnost a matematická statistika*. Vydavatelství ČVUT, 2006.
- [23] Oppy, G.; Dowe, D.: The Turing Test [online].  
<http://plato.stanford.edu/archives/spr2011/entries/turing-test/>,  
 2011-01-26 [cit. 2012-03-02].
- [24] Samek, J.: *Důvěra a reputace v distribuovaných systémech*. Dizertační práce, FIT VUT v Brně, Brno, 2011.
- [25] Schneier, B.: *Liars and Outliers*. Wiley, 2012, iISBN 978-1118143308.
- [26] Zhang, L.; Tan, C. P.; Li, S.; aj.: The Influence of Interaction Attributes on Trust in Virtual Communities. In *International Conference on User Modeling, Adaptation and Personalization (UMAP)*, Nanyang Technological University, Singapore, 2011.
- [27] Šimíčková Čížková, J.: *Přehled sociální psychologie*. Olomouc: Univerzita Palackého v Olomouci, 2004, iISBN 80-244-0929-1.
- [28] Řezáč, J.: *Sociální psychologie*. Brno: Paido, 1998, iISBN 80-85931-48-6.

# Příloha A

## Výsledky průzkumného šetření

Pohlaví	Otázka 1	Otázka 2	Otázka 3 <sup>1</sup>					
			a	b	c	d	e	f
M	7	0 - 2	4	3	1	3	2	4
M	5	0 - 2	5	4	2	4	2	4
M	5	0 - 2	5	3	3	2	1	2
Ž	4	0 - 2	5	2	1	4	3	3
Ž	5	0 - 2	5	4	3	1	3	1
Ž	8	0 - 2	5	5	4	4	1	5
Ž	4	3 - 5	5	2	4	5	1	4
Ž	7	3 - 5	2	2	2	2	3	1
M	3	3 - 5	1	3	3	3	4	5
M	2	0 - 2	3	5	4	5	3	5
Ž	5	6 - 8	5	3	2	3	2	3
M	3	0 - 2	5	4	3	2	1	5
M	4	3 - 5	5	3	3	5	5	5
Ž	4	0 - 2	5	3	2	5	4	5
M	3	3 - 5	5	3	3	3	3	5
Ž	7	0 - 2	3	4	2	3	2	1
M	7	0 - 2	1	2	3	3	3	2
Ž	4	6 - 8	1	2	2	2	5	5

Tabulka A.1: Plné znění otázek lze nalézt v části práce 6.3.

<sup>1</sup> Pro otázku č. 3 jsou vysvětlivky pro kontexty následující:

- **a:** soukromé zprávy
- **b:** komentáře
- **c:** označení „To se mi líbí“
- **d:** společné fotografie
- **e:** společné skupiny
- **f:** pravidelnost interakce



# Příloha B

## Obsah CD

Příložené CD obsahuje následující adresářovou strukturu:

- **src**: adresář se skriptem, jehož použití je vysvětleno v příloze C, spolu s nezbytným konfiguračním souborem; zbytek adresářové struktury a využitých souborů si skript vytvoří sám
  - `trust.py`: skript v jazyce Python3, kódování i výstupní soubory v UTF-8
  - `conf.csv`: jednoduchý konfigurační soubor
- **text**: adresář se zdrojovými soubory sázecího systému LaTeX; šablona poskytnutá fakultou nebyla nijak modifikována
  - `xsvect00.pdf`: zkompilevaný soubor ve formátu PDF
  - `xsvect00.zip`: komprimovaný archiv s potřebnými soubory pro překlad

## Příloha C

# Manuál k použití aplikace

### C.1 Požadavky pro spuštění skriptu

Pro spuštění skriptu je zapotřebí pouze interpret jazyka Python v minimální verzi 3.2. Není nutné instalovat žádné knihovny. Jediným souborem nezbytným pro jeho fungování je soubor `conf.csv`, okolní adresářovou strukturu si skript vytvoří sám. Z toho vyplývá požadavek na práva zápisu ve složce, ve které je skript spouštěn (nelze jej spouštět přímo z CD). Vzhledem k síťové povaze skriptu je nutné připojení k internetu, pokud však jednou analýza proběhla, lze jej spustit pro stejného uživatele i bez síťového připojení (dotazy se ukládají do paměti cache). Výstupem analýzy je textový soubor s názvem:

```
<vlastní_jméno_uživatele>_<čas_analýzy>.txt
```

V konfiguračním souboru `conf.csv` je nutné zadat vlastní jméno uživatele (pod tímto názvem bude evidován v databázi paměti cache), access token (generování popisuje část [C.3](#)) a datum, od kterého si přejete analýzu provádět. Formát vstupních dat je následující:

```
<vlastní_jméno_uživatele>;<access_token>;<datum_ve_formátu_DD:MM:YYYY>
```

### C.2 Adresářová struktura

Skript vytváří vlastní adresářovou strukturu:

- `cache`: adresář pro ukládání dotazů, název souboru odpovídá dotazu
- `output`: adresář pro ukládání výstupních souborů
- `log.csv`: soubor s odeslanými dotazy, omezuje počet dotazů za časovou jednotku
- `id_numbers.txt`: soubor s ID přiřazeným jednotlivým uživatelům

## C.3 Generování přístupového tokenu

Pro přístup k analyzovanému profilu je nutné vygenerovat tzv. přístupový token. To lze provést na následující adrese:

<https://developers.facebook.com/tools/explorer>

Pro vygenerování je nutné být přihlášený na profilu sítě Facebook, což lze ověřit v pravém horním rohu stránky. Také je důležité poznamenat, že po celou dobu analýzy nesmí dojít k odhlášení, jinak platnost tokenu okamžitě vyprší. Po stisknutí tlačítka „Get Access Token” se zobrazí nabídka s právy. Pro činnost skriptu jsou nezbytná následující práva:

- **User Data Permissions:**
  - `user_about_me`: základní informace o uživateli
  - `user_photos`: přístup k fotografiím (použity jsou pouze komentáře)
  - `user_groups`: skupiny, kterých je uživatel členem
  - `user_likes`: oblíbené stránky
  - `user_photo_video_tags`: označení na fotkách a videomateriálech
  - `user_status`: přístup ke zveřejněným statusům
- **Extended Permissions:**
  - `read_stream`: čtení statusů, pokud není aktivován profil Timeline

Vygenerovaný přístupový token má platnost přibližně 30 minut (intervaly se v průběhu testování lišily). Stačí jej vložit do souboru `conf.csv` a spustit skript.