

Modelování komplexních sítí pomocí heuristických algoritmů

Bc. Tomáš Trčka

Diplomová práce
2016



Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky
akademický rok: 2015/2016

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Tomáš Trčka**
Osobní číslo: **A14451**
Studijní program: **N3902 Inženýrská informatika**
Studijní obor: **Informační technologie**
Forma studia: **prezenční**

Téma práce: **Modelování komplexních sítí pomocí heuristických algoritmů**
Téma anglicky: **Modelling of Complex Networks by Means of Heuristic Algorithms**

Zásady pro vypracování:

1. Vypracujte literární rešerši na dané téma, zaměřte se zejména na dynamiku tzv. hejnových algoritmů, komplexních, sociálních a citačních sítí.
2. Zvolte vhodný hejnový algoritmus a popište jeho principy.
3. Provedte transformaci dynamiky hejnového algoritmu do formy komplexních sítí.
4. Nalezněte takové nastavení těchto algoritmů a odpovídající testovací problémy, aby dynamika algoritmu transformována do formy sítě odpovídala modelu skutečné sítě.
5. Ověřte možnosti predikce dynamiky komplexní sítě.

Rozsah diplomové práce:

Rozsah příloh:

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

Seznam odborné literatury:

1. **ESTRADA, Ernesto.** The structure of complex networks: theory and applications. New York: Oxford University Press, 2012, xii, 465 stran. ISBN 978-0-19-959175-6.
2. **ZELINKA, Ivan.** Evoluční výpočetní techniky: principy a aplikace. 1. vyd. Praha: BEN - technická literatura, 2009, 534 s. ISBN 978-80-7300-218-3.
3. **DE JONG, Kenneth A.** Evolutionary computation: a unified approach. Cambridge: MIT Press, 2006, ix, 256 s. ISBN 02-620-4194-4.
4. **DOROGOVCEV, Sergej Nikolajevič.** Lectures on complex networks. First published. Oxford: Oxford University Press, 2010, ix, 134 stran. ISBN 978-0-19-954893-4.
5. **BARRAT, Alain, Marc BARTHÉLEMY a Alessandro VESPIGNANI.** Dynamical processes on complex networks. First published. Cambridge: Cambridge University Press, 2008, xvii, 347 stran. ISBN 978-1-107-62625-6.
6. **LU, Linyuan a Fan R CHUNG.** Complex graphs and networks. Providence, RI: American Mathematical Society, 2006, vii, 264 s. ISBN 9780821836576.
7. **KENNEDY, James, Russell C EBERHART a Yuhui SHI.** Swarm intelligence. San Francisco: Morgan Kaufmann, c2001, xxvii, 512 s. ISBN 1-55860-595-9.
8. **Handbook of Optimization: From Classical to Modern Approach.** 2013. vyd. Editor Ivan Zelinka, Václav Snášel, Ajith Abraham. Berlin: Springer, 2013, xii, 1100 s. Intelligent systems reference library, 38. ISBN 978-3-642-30503-0.

Vedoucí diplomové práce:

doc. Ing. Roman Šenkeřík, Ph.D.

Ústav informatiky a umělé inteligence

Datum zadání diplomové práce:

5. února 2016

Termín odevzdání diplomové práce:

20. května 2016

Ve Zlíně dne 5. února 2016


doc. Mgr. Milan Adámek, Ph.D.
děkan




doc. Mgr. Roman Jašek, Ph.D.
ředitel ústavu

Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen připouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

- že jsem na diplomové/bakalářské práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně, dne 20.5.2016


.....
podpis diplomanta

ABSTRAKT

Diplomová práce si klade za cíl seznámit čtenáře s modelováním reálných komplexních sítí pomocí heuristického hejnového algoritmu Firefly. Úvod teoretické části je věnován historii, praktickým příkladům, typickým vlastnostem a centralitám komplexních sítí. Dále jsou vysvětleny vybrané pojmy z oblasti teorie grafů a možnosti jejich elektronické reprezentace, neboť právě pomocí nich bývají komplexní sítě často interpretovány. Dále jsou obecně charakterizovány heuristické a hejnové algoritmy s podrobným vysvětlením principů světluščího algoritmu (algoritmu Firefly). Na konci teoretické části jsou nastíněny základní informace o použitých softwarových nástrojích. Praktická část popisuje nalezené vhodné reálné komplexní sítě, implementované testovací funkce, převod dynamiky hejnového algoritmu na komplexní síť, implementaci algoritmu Firefly a samotné modelování a predikci dynamiky komplexních sítí.

Klíčová slova: komplexní síť, modelování, predikce, algoritmus Firefly

ABSTRACT

The diploma thesis aims to acquaint the reader with modelling of real complex networks using the heuristic swarm algorithm Firefly. The beginning of the theoretical part is devoted to the history, practical examples, typical characteristics and centralities of complex networks. Furthermore, selected terms of the theory of graphs and possibilities of electronic representation of graphs are explained because graphs are often used to interpret complex networks. Moreover, heuristic and swarm algorithms are generally characterized whereas principles of the Firefly algorithm are explained in detail. At the end of the theoretical part, basic information about the used software tools is outlined. The practical part describes found suitable real complex networks, implemented test functions, conversion of the dynamics of the swarm algorithm to the complex network, the implementation of the Firefly algorithm and modelling and prediction of the dynamics of complex networks itself.

Keywords: complex networks, modelling, prediction, Firefly algorithm

Rád bych tímto poděkoval své nejbližší rodině za psychickou i materiální podporu během celého studia, panu vedoucímu doc. Ing. Romanu Šenkeříkovi, Ph.D. a panu konzultantovi Ing. Michalu Pluháčkovi, Ph.D. za cenné podněty, rady a připomínky při zpracovávání této diplomové práce, protože bez nikoho z nich by nikdy nevznikla.

OBSAH

ÚVOD	10
I TEORETICKÁ ČÁST	12
1 KOMPLEXNÍ SÍTĚ	13
1.1 HISTORIE.....	13
1.2 KOMPLEXNÍ SÍTĚ V REÁLNÉM SVĚTĚ	14
1.2.1 Technologické, informační a citační sítě	14
1.2.2 Biologické sítě.....	16
1.2.3 Sociální sítě	17
1.2.4 Shrnutí příkladů komplexních sítí	18
1.3 TEORIE GRAFŮ.....	18
1.3.1 Podgraf	19
1.3.2 Orientace hran	19
1.3.3 Váha hran	20
1.3.4 Stupeň vrcholu	20
1.3.5 Reprezentace grafu.....	21
1.4 TYPICKÉ VLASTNOSTI KOMPLEXNÍCH SÍTÍ	23
1.4.1 Malé vzdálenosti (vlastnost malého světa, small world fenomen)	23
1.4.2 Shlukování (clustering)	24
1.4.3 Bezškálovost (scale-free)	25
1.5 CENTRALITY KOMPLEXNÍCH SÍTÍ	25
1.5.1 Degree centralita (DC)	25
1.5.2 Closeness centralita (CC).....	25
1.5.3 Betweenness centralita (BC).....	26
2 HEURISTICKÉ ALGORITMY	30
2.1 HEJNOVÉ ALGORITMY	30
2.1.1 Základní pojmy	31
2.1.2 Firefly algoritmus (světluščí algoritmus)	32
2.2 NO FREE LUNCH TEORÉM.....	34
3 POUŽITÉ SOFTWAREVÉ NÁSTROJE	35
3.1 WOLFRAM MATHEMATICA.....	35
3.2 CYTOSCAPE.....	35
3.2.1 aMatReader	36
3.2.2 CytoNCA.....	37
II PRAKTICKÁ ČÁST	39
4 HLEDÁNÍ VHODNÝCH KOMPLEXNÍCH SÍTÍ	40
4.1 CITAČNÍ SÍTĚ	40
4.1.1 Scopus	40
4.2 SOCIÁLNÍ SÍTĚ	40
4.2.1 SNAP.....	41
5 TESTOVACÍ FUNKCE	42

5.1	NEKONEČNÁ NAKLONĚNÁ ROVINA (RAMPA)	42
5.2	PRVNÍ DE JONGOVA FUNKCE (SFÉRA)	42
5.3	DRUHÁ DE JONGOVA FUNKCE (ROSENBROCKOVO SEDLO, BANÁNOVÁ FUNKCE)	43
5.4	RASTRIGINOVA FUNKCE	44
5.5	SCHWEFELOVA FUNKCE	45
5.6	GRIEWANGKOVA FUNKCE	45
5.7	MASTERSOVA FUNKCE	46
6	PŘEVOD DYNAMIKY HEJNOVÉHO ALGORITMU NA KOMPLEXNÍ SÍŤ.....	47
7	IMPLEMENTACE ALGORITMU FIREFLY	49
7.1	FUNKCE MAIN[]	49
7.1.1	Preset params	50
7.1.2	Import adj mat	50
7.1.3	Directed	52
7.1.4	Nodes.....	52
7.1.5	Edges	52
7.1.6	Test function.....	53
7.1.7	Dimensions.....	54
7.1.8	Lower bound	54
7.1.9	Upper bound.....	54
7.1.10	Init positions.....	54
7.1.11	Act probability	54
7.1.12	Rnd probability.....	54
7.1.13	BetaMin.....	55
7.1.14	Action.....	55
7.1.15	Edge no.....	56
7.1.16	Export adj mat	56
7.1.17	Exit app	57
7.2	FUNKCE FIREFLY[]	57
7.2.1	Vstupní parametr alfa.....	57
7.2.2	Vstupní parametr gama	58
7.2.3	Začátek těla funkce	58
7.2.4	Začátek smyčky While.....	58
7.2.5	Výběr act a rnd jedince	58
7.2.6	Aktualizace relY.....	58
7.2.7	Výpočet beta.....	58
7.2.8	Výpočet r	59
7.2.9	Určení eps.....	59
7.2.10	Výpočet xn	59
7.2.11	Konec těla smyčky While	59
7.2.12	Návratové hodnoty funkce	60
8	MODELOVÁNÍ KOMPLEXNÍCH SÍŤÍ POMOCÍ ALGORITMU FIREFLY	61
8.1	CITAČNÍ SÍŤ ÚSTAVU INFORMATIKY A UMĚLÉ INTELIGENCE NA FAI UTB	61
8.1.1	Modelování citační sítě UIUI v roce 2015	65
8.1.2	Modelování citační sítě UIUI v roce 2014.....	69

8.1.3	Modelování citační sítě UIUI v roce 2013	70
8.1.4	Modelování citační sítě UIUI v roce 2012	70
8.1.5	Modelování citační sítě UIUI v roce 2011	71
8.1.6	Modelování citační sítě UIUI v roce 2010	72
8.2	SOCIÁLNÍ SÍŤ SPOLUAUTORSTVÍ NAD FYZIKÁLNÍMI VĚDECKÝMI ČLÁNKY	73
8.2.1	Modelování sociální sítě coAuth v časovém rámci 4.....	76
8.2.2	Modelování sociální sítě coAuth v časovém rámci 3.....	77
8.2.3	Modelování sociální sítě coAuth v časovém rámci 2.....	78
8.2.4	Modelování sociální sítě coAuth v časovém rámci 1.....	79
9	PREDIKCE DYNAMIKY KOMPLEXNÍCH SÍTÍ POMOCÍ ALGORITMU FIREFLY	81
9.1	CITAČNÍ SÍŤ ÚSTAVU INFORMATIKY A UMĚLÉ INTELIGENCE NA FAI UTB	81
9.1.1	Predikování citační sítě UIUI na rok 2015.....	81
9.1.2	Predikování citační sítě UIUI na rok 2016.....	83
9.2	SOCIÁLNÍ SÍŤ SPOLUAUTORSTVÍ NAD FYZIKÁLNÍMI VĚDECKÝMI ČLÁNKY	83
9.2.1	Predikování sociální sítě coAuth na časový rámec 4	84
9.2.2	Predikování sociální sítě coAuth na časový rámec 5	85
	ZÁVĚR	87
	SEZNAM POUŽITÉ LITERATURY.....	89
	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....	91
	SEZNAM OBRÁZKŮ	93
	SEZNAM TABULEK.....	95
	SEZNAM PŘÍLOH.....	97
	PŘÍLOHA P I: OBSAH PŘILOŽENÉHO CD	98

ÚVOD

Diplomová práce využívá moderních trendů v analýze populační dynamiky hejnových a evolučních algoritmů, zejména v oblasti používání komplexních sítí pro zvýšení výkonu hejnových optimalizačních algoritmů. Tento princip spočívá ve vytváření modelu komplexní sítě na základě dynamiky hejnového algoritmu. Každý jedinec je reprezentován jedním uzlem sítě, přičemž jedinci, jejichž interakce s ostatními často vedla k nalezení lépe ohodnoceného místa v rámci hyperplochy, vytvářeli v síti více vazeb (resp. hran) s ostatními uzly. Jedinci s nejmenším počtem hran tedy pouze spotřebovávají výpočetní čas a ke zlepšování výsledku přispívají jen málo, proto je vhodné tyto jedince vyměnit za nové (např. resetováním jejich souřadnic na náhodné hodnoty).

Ve spolupráci FAI UTB a FEI VŠB-TUO došlo k realizaci myšlenky reverzního procesu, tedy využít tuto techniku nikoliv pro zlepšení vlastností optimalizačního algoritmu pomocí zpětnovazebních zásahů, nýbrž pro predikci chování reálných komplexních sítí. Postup se tedy obrací a místo hejnového algoritmu na vstupu a komplexní sítě na výstupu bude na vstupu struktura komplexní sítě a na výstupu nastavení parametrů hejnového algoritmu modelujícího tuto komplexní síť co možná nejpřesněji.

Vývojovým prostředím aplikace byla zvolena Wolfram Mathematica, se kterou mám již dostatečné zkušenosti a dle ukazatele Score Index na webu Project Euler vyjadřujícím vhodnost daného programovacího jazyka pro řešení tamějších matematicko-programátorských úloh se Mathematica umístila na 3. místě z 93 jazyků, přičemž lepšího hodnocení dosáhly pouze jazyky PARI/GP a Stata a ještě před pár lety byla Mathematica na pozici nejvyšší. Kompletní statistika je pro přihlášené uživatele dostupná na adrese <https://projecteuler.net/languages>. Registrace je zdarma, ale lze také použít můj testovací účet s přihlašovacím jménem *jenomzkousim* a heslem *jenomzkousim*.

Mezi výhody SW Mathematica patří obrovská knihovna integrovaných funkcí, které pokryjí potřeby začínajícího i profesionálního programátora, výborně zpracovaná dokumentace v rámci „offline“ nápovědy samotného programu i „živé“ komunity na portále <http://mathematica.stackexchange.com/>, dále značné množství vestavěných funkcí pro práci s grafickým uživatelským rozhraním. Přenositelnost vytvořené aplikaci již není problémem, neboť kromě plné verze SW Mathematica lze skript (musí být uložen ve formátu *cdf*, nikoliv *nb*) spustit také v samostatném CDF přehrávači (CDF Player), jehož základní verze je ke

stažení zdarma k dispozici na adrese <https://www.wolfram.com/cdf-player/> a je kompatibilní s platformou Windows, Linux i Mac.

I. TEORETICKÁ ČÁST

1 KOMPLEXNÍ SÍTĚ

Z pohledu teorie grafů představuje komplexní síť graf s jistými topologickými rysy, které se v jednoduchých sítích, jako jsou například náhodné grafy, nevyskytují. Tyto rysy jsou dány způsobem propojení jednotlivých prvků, které v rámci celé sítě vytváří určité vzory, které nejsou ani čistě pravidelné (deterministické), ale ani čistě náhodné (stochastické).[1][2]

Typickými představiteli komplexních sítí jsou reálné systémy jako technologické sítě (např. počítačové, telekomunikační, elektrorozvodné, silniční sítě), biologické sítě (např. neuronové sítě, potravinové řetězce) nebo sociální sítě.[1][2]

Pro modelování komplexních sítí je možné použít metody z oblasti jednoho z největších pilířů diskrétní matematiky, a to teorie grafů.[1][2]

1.1 Historie

Za zakladatele teorie grafů je považován švýcarský fyzik a matematik Leonhard Euler, který v roce 1736 vyřešil slavnou úlohu „Jak projít přes sedm mostů v městě Königsberg (Královec)“, přičemž každý z mostů bude navštíven právě jednou.[3][4]

Roku 1845 zveřejnil německý fyzik Gustav Kirchhoff fyzikální zákony, které v elektrických obvodech slouží k určení hodnot elektrického napětí a proudu v jednotlivých větvích obvodu. Tyto zákony jsou v teorii grafů dodnes využívány při určování takzvaných toků v síti, které v praxi řeší úlohy o propustnosti například železniční, silniční nebo vodovodní sítě.[3][4]

Roku 1852 publikoval jihoafrický matematik a botanik Francis Guthrie „Problém čtyř barev“, který spočíval v domněnce, zdali je možné pomocí pouhých čtyř různých barev obarvit libovolnou politickou mapu takovým způsobem, aby žádné dva sousední státy neměly stejnou barvu. Domněnka byla potvrzena až roku 1976 počítačovým programem amerických matematiků.[4]

Ve 30. a 40. letech 20. století probíhaly první sociologické dotazníkové výzkumy např. za účelem zmapování mezilidských vztahů na pracovišti a následného identifikování sociálního vůdce v rámci skupiny pracovníků na stejné profesní úrovni.[3]

Teorie grafů se od doby svého vzniku v 18. století zabývala klasickými grafy s desítkami, nejvýše stovkami vrcholů. V 50. letech minulého století však byly rozlehlé technologické sítě, které v globálním měřítku vznikaly spíše přirozeně než podle nějakého celosvětového

plánu, popsány jako náhodné grafy. Vznikli tak první zástupci nejjednodušších komplexních sítí.[3]

Do 90. let minulého století tak byly modelem komplexních sítí právě náhodné grafy. Ve druhé polovině 90. let již byl dostatek sítí převeden do elektronické podoby, a tak bylo překročeno k jejich počítačové analýze. Koncem 90. let již existovaly první modely složitějších komplexních sítí s miliony, později dokonce i s miliardami vrcholů.[2]

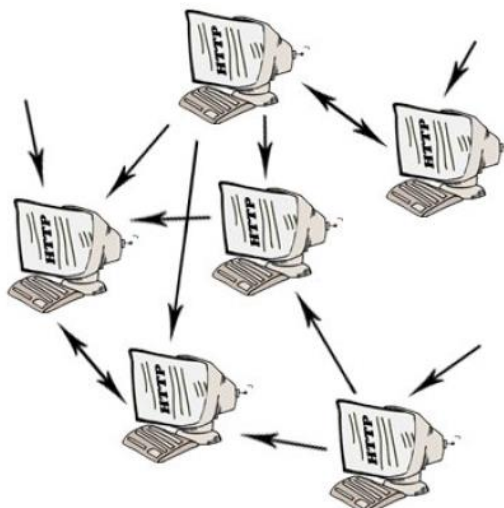
1.2 Komplexní sítě v reálném světě

Mnoho sítí z různých oborů lidské činnosti bylo namodelováno a podrobena matematické analýze. Výsledky ukázaly, že neohledě na obor, ze kterého síť pocházela, se hodnoty určitých parametrů velmi nápadně podobaly. Následující typy sítí je tedy možné zařadit do oblasti komplexních sítí.[5][6]

1.2.1 Technologické, informační a citační sítě

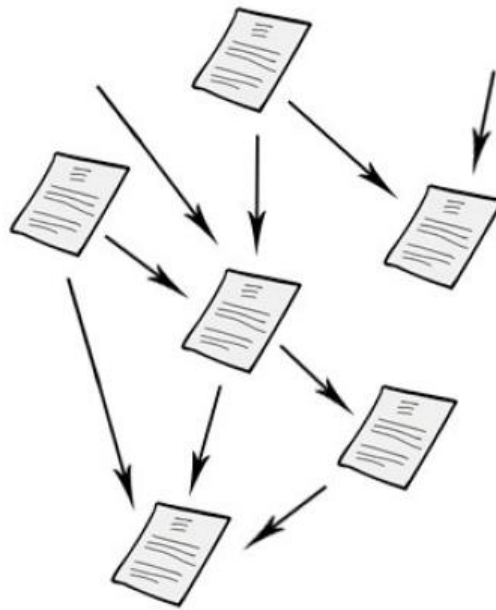
Jedná se o sítě vytvořené člověkem, a to zejména za účelem distribuce určité komodity jako je ropa, plyn, voda, elektřina, nebo teplo. Dále jsou zde řazeny komunikační sítě, jako jsou telefonní rozvodné sítě, železnice, silnice, trasy letadel či lodí, a informační sítě jako např. počítačové sítě (v čele s celosvětově nejrozšířenější sítí zvanou Internet) nebo sítě citací mezi vědeckými články.[5][6]

Obrázek 1 znázorňuje malou část Internetu, tedy několik HTTP serverů, z nichž některé na sebe odkazují pomocí hypertextových odkazů. Vrcholy tedy představují WWW stránky a hrany hypertextové odkazy. Internetová síť je orientovaná a mohla by být také vážená, pokud by byl brán v úvahu například počet odkazů na jiný server, který daný server obsahuje. Tato hodnota by potom reprezentovala váhu hrany.[3]



Obrázek 1 Síť dokumentů na Internetu[3]

Na Obrázku 2 lze vidět příklad citační sítě. Drtivá většina kvalitních akademických publikací cituje dříve napsaná díla s podobnou tematikou. Vrcholy v citační síti na Obrázku 2 představují vědecké články a hrany znázorňují, že článek, ze kterého hrana vychází, cituje článek, do kterého hrana směřuje. Citační sítě jsou tedy orientované a mohou být také vážené, pokud budou například uzly sítě představovat autory článků a hrany budou vyjadřovat, že články daného autora citují články jiného autora, pak lze hranám přiřadit váhu podle toho, kolik článků jiného autora články daného autora citovaly. Citační sítě jsou vždy acyklické, což znamená, že nově publikovaný článek ocituje své zdrojové články jednorázově v momentě svého vydání. Tyto citace již do budoucna zůstávají stejné a nemohou se měnit. Na druhou stranu již publikovanému článku může počet vstupních citací pouze narůstat, není tedy možné, aby článek, který určitý jiný starší článek citoval, v budoucnu nějakým způsobem zanikl nebo citaci zrušil.[3][6]

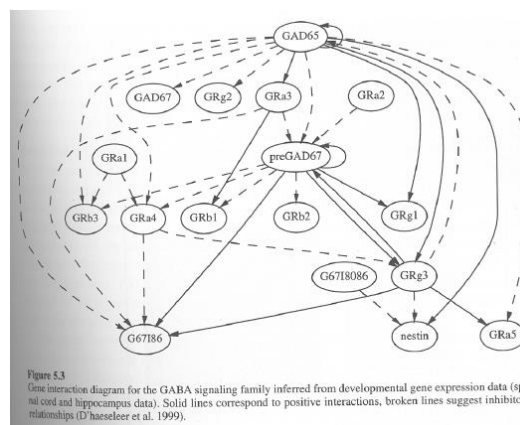


Obrázek 2 Citační síť[3]

1.2.2 Biologické sítě

Zkoumání biologických sítí začalo výzkumem DNA a genetických řetězců v 80. letech minulého století. Biologická síť vždy nějakým způsobem reprezentuje biologický systém. Zástupcem může být například neuronová síť, kde jsou uzly tvořeny neurony a hrany přenášením signálu mezi danými neurony.[3][6]

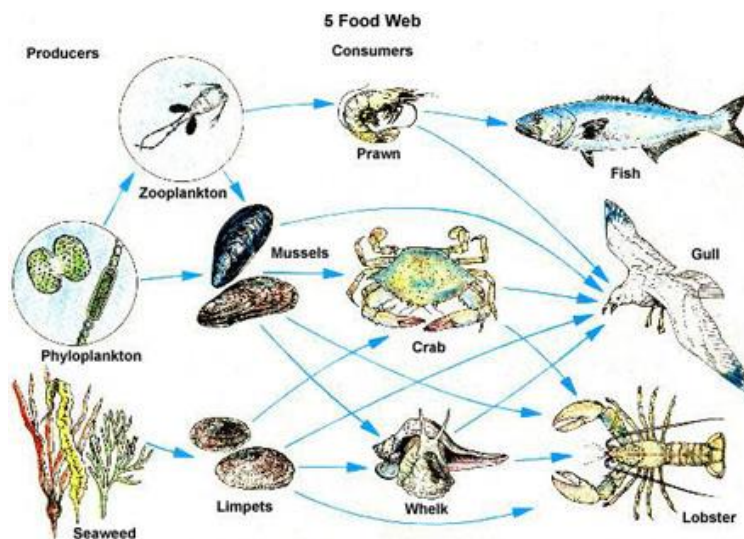
Obrázek 3 představuje síť bílkovin, kde uzly jsou jednotliví zástupci bílkovin a hrany znamenají, že dané bílkoviny spolu reagují.[7]



Obrázek 3 Síť bílkovin[7]

Dalším příkladem může být potravní řetězec, kde na Obrázku 4 uzly představují druhy živočichů a hrany spojují živočichy, kde jeden je potravou druhého. Hrany jsou orientované,

přičemž začátek hrany označuje živočicha, který je potravou pro živočicha, v němž hrana končí.[8]

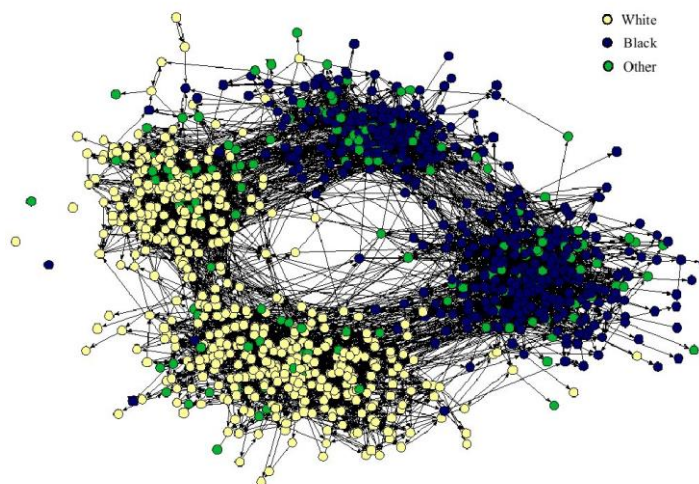


Obrázek 4 Potravní řetězec[8]

1.2.3 Sociální síť

Sociální síť tvoří množina lidí, kteří mezi sebou udržují nějaké vztahy nebo vazby, což může být pracovní, rodinný, přátelský nebo sexuální vztah, ekonomická vazba nebo prostě jen obecný zájem. Lidé v rámci sociální sítě spolu určitým způsobem komunikují a na základě svého chování se vzájemně ovlivňují. Hlavním cílem analýzy sociálních sítí je identifikace míry vlivu jednotlivých osob na ostatní, což může vést k optimalizaci pracovního týmu za účelem jeho vyššího pracovního výkonu i větší spokojenosti zaměstnanců, zamezení šíření poplašné zprávy na Internetu s minimálním vynaloženým úsilím a náklady, anebo rozbití zločineckého gangu či teroristické skupiny.[3][6]

Obrázek 5 znázorňuje síť známostí mezi bělochy, černochoy a ostatními rasami. Bílé uzly představují zástupce bělochů, černé uzly zástupce černochoy a zelené uzly zástupce jiných ras. Hrany pak představují známosti mezi danými lidmi. Z Obrázku 5 na první pohled vyplývá, že ve sledované skupině spolu udržují více kontaktů vždy lidé v rámci jedné rasy, což ovšem za několik desítek let vlivem globalizace už nemusí být pravda.[8]



Obrázek 5 Síť známostí mezi bělochy a černochoy[8]

1.2.4 Shrnutí příkladů komplexních sítí

Tabulka 1 shrnuje základní zmíněné oblasti, které spadají do komplexních sítí. Sítě jsou rozděleny do tří výše popsaných kategorií, i když rozdělení není vždy jednoznačné, například telefonní hovory či vědecká spolupráce by mohli patřit také do skupiny informačních sítí.

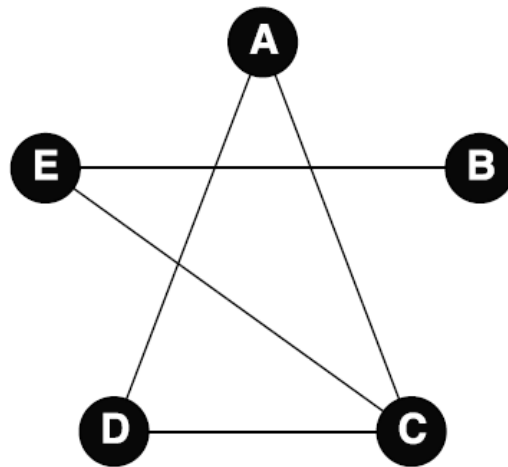
Tabulka 1 Shrnutí příkladů komplexních sítí[7][3]

Typ	Oblast	Orientovaná?	Uzly	Hrany
Technologické, informační, citační	elektrické rozvody	ne	transformátory	dráty
	vlakové trasy	ne	stanice	kolejiště
	Web	ano	stránky	odkazy
	internet	ne	servery	dráty
	citační sítě	ano	vědecké články	citace
	lingvistika	ne	slova	synonyma
Biologické	potravní řetězce	ano	druhy zvířat	vztah lovec-kořist
	metabolismus	ne	chemické látky	reakce
	neuronové sítě	ne	neurony	synaptické spojení
Sociální	síť herců	ne	herci	hráli v jednom filmu
	síť sexuálních kontaktů	ne	lidé	sex
	telefonní hovory	ano	telefony	volání
	vědecká spolupráce	ne	vědci	spoluautorství

1.3 Teorie grafů

Komplexní sítě bývají nejčastěji zobrazovány (modelovány) jako grafy (viz Obrázek 6), které lze definovat jako množinu vrcholů (někdy též uzlů; anglicky „nodes“), z nichž některé

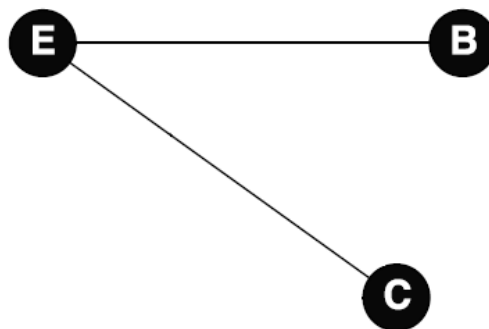
jsou vzájemně propojeny hranami (anglicky „edges“). Méně formálně se tedy jedná o strukturu objektů a vazeb mezi nimi.[2]



Obrázek 6 Jednoduchý graf[9]

1.3.1 Podgraf

Podgraf na Obrázku 7 představuje část původního celého grafu na Obrázku 6. Formálně lze tedy říci, že podgraf vznikne odebráním některých vrcholů a souvisejících hran z původního grafu při zachování podmínky, že každá hrana spojuje právě dva vrcholy.[2][9]

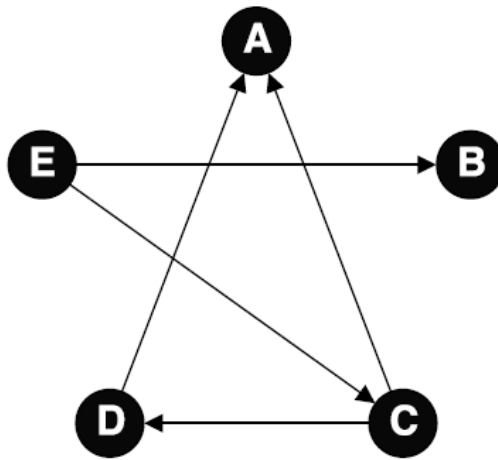


Obrázek 7 Podgraf[9]

1.3.2 Orientace hran

Grafy lze rozdělit dle rozlišování orientace hran na orientované (viz Obrázek 8) a neorientované (též jednoduché, viz Obrázek 6). Neorientovaná hrana značí průchodnost cesty oběma směry, orientovaná značí průchodnost pouze jedním směrem (např. v Obrázku 8 se lze z uzlu C přímo dostat do uzlu D, ale z uzlu D do uzlu C nikoliv). V případě orientovaného grafu lze průchodnost cesty oběma směry zaznačit dvěma opačně orientovanými hranami. Orientovaná hrana bývá obvykle zakončena šipkou značící směr

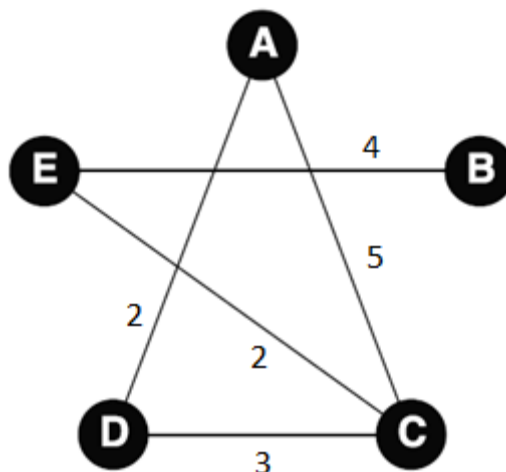
průchodnosti. V případě neorientovaného grafu není orientace hran známa nebo důležitá.[2][9]



Obrázek 8 Orientovaný graf[9]

1.3.3 Váha hran

Grafy lze také rozdělit dle rozlišování váhy u hran na vážené (viz Obrázek 9) a nevážené (viz Obrázek 6). V neváženém grafu mají všechny hrany stejnou váhu (důležitost, propustnost, počet propojení). Ve váženém grafu vyjadřujeme kvalitu jednotlivých propojení číslem. U neváženého grafu není váha hran známa nebo důležitá.[2][9]

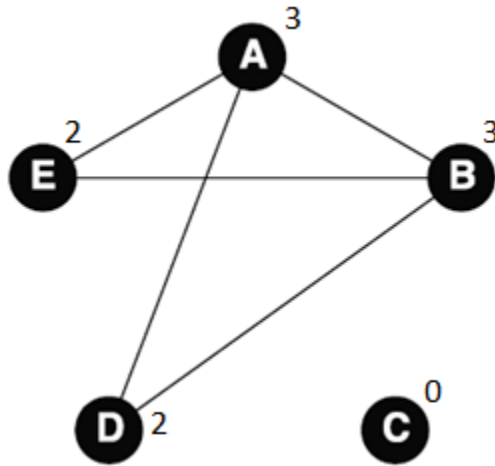


Obrázek 9 Vážený graf[9]

1.3.4 Stupeň vrcholu

Tato hodnota vyjadřuje, s kolika jinými uzly (resp. vrcholy) daný uzel (resp. vrchol) sousedí, resp. je přímo propojen hranou (viz Obrázek 10). V případě orientovaného grafu se rozlišují

dvě hodnoty. Vstupní stupeň vrcholu určuje, kolik hran směřuje do daného uzlu, zatímco výstupní stupeň vrcholu určuje, kolik hran směřuje z daného uzlu.[2][9]



Obrázek 10 Graf s označenými stupni vrcholů[9]

1.3.5 Reprezentace grafu

Pro možnost zpracování grafu výpočetní technikou a zjištění jeho vlastností je nutné umět graf bezztrátově převést na nějakou vhodnou strukturu.[9]

Seznam sousedů

Tato metoda je vhodná spíše pro vizualizaci nebo manuální vyhledávání sousedů daného uzlu než pro výpočet vlastností komplexní sítě. U hustě propojených sítí se ztrácí přehlednost a tabulka (viz Tabulka 2) také obsahuje polovinu redundantních dat.[9]

Tabulka 2 Seznam sousedů grafu na Obrázku 6

Uzel	Sousedé
A	C, D
B	E
C	A, D, E
D	A, C
E	B, C

Incidenční matice

Počet řádků incidenční matice odpovídá počtu uzlů v grafu. Každý sloupec matice pak představuje jednu hranu v grafu. Matice tedy bude mít rozměry $n \times m$, kde n odpovídá počtu

uzlů a m odpovídá počtu hran v grafu. Hodnoty „1“ v rámci sloupce svými řádky označují uzly, které daná hrana spojuje. V případě váženého grafu jsou fixní hodnoty „1“ v matici nahrazeny čísly reprezentujícími váhu dané hrany.[9]

Incidenční matice na Obrázku 11 představuje neorientovaný graf na Obrázku 6. Sloupce zleva doprava postupně zachycují propojení uzlů A-C, A-D, B-E, C-D a C-E.

A	1	1	0	0	0
B	0	0	1	0	0
C	1	0	0	1	1
D	0	1	0	1	0
E	0	0	1	0	1

Obrázek 11
Incidenční matice
grafu na Obrázku 6

Pro zachycení orientovaného grafu se kromě hodnoty „1“ navíc využívá i hodnoty „-1“. Uzel, odkud hrana vychází, je označen hodnotou „1“, zatímco uzel, do kterého hrana směřuje, je označen hodnotou „-1“.[9]

Incidenční matice na Obrázku 12 představuje orientovaný graf na Obrázku 8. Sloupce zleva doprava postupně zachycují propojení uzlů C→A, D→A, E→B, C→D a E→C.

A	-1	-1	0	0	0
B	0	0	-1	0	0
C	1	0	0	1	-1
D	0	1	0	-1	0
E	0	0	1	0	1

Obrázek 12
Incidenční matice
grafu na Obrázku 8

Matice sousednosti (vstupně-výstupní matice, adjacency matrix)

Počet řádků stejně jako počet sloupců matice sousednosti odpovídá počtu uzlů v grafu. Matice tedy bude vždy čtvercová o rozměrech $n \times n$, kde n odpovídá počtu uzlů v grafu. Hodnota „1“ v matici představuje jednu hranu, přičemž pozice hodnoty „1“ v rámci matice označuje svým řádkem a sloupcem uzly, které tato hrana spojuje. Propojení uzlu „sám na sebe“ (zpětné vazby) se tedy budou v matici nacházet na hlavní diagonále. U

neorientovaného grafu (viz Obrázek 13) bývá matice kvůli snadnějšímu vyhledávání propojení obvykle symetrická podle hlavní diagonály, tedy součet prvků matice představuje dvojnásobek skutečného počtu hran. V případě váženého grafu jsou fixní hodnoty „1“ v matici nahrazeny číslem reprezentujícím váhu dané hrany.[9]

	A	B	C	D	E
A	0	0	1	1	0
B	0	0	0	0	1
C	1	0	0	1	1
D	1	0	1	0	0
E	0	1	1	0	0

Obrázek 13 Matice susednosti grafu na Obrázku 6

V případě orientovaného grafu (viz Obrázek 14) představuje odpovídající řádek hodnoty „1“ uzel, z něhož hrana vychází, a odpovídající sloupec hodnoty „1“ uzel, do něhož hrana směřuje. Součet prvků matice tedy odpovídá skutečnému počtu hran v síti. Matice tak bude symetrická pouze ve speciálních případech, kdy by všechna propojení byla obousměrná.[9]

	A	B	C	D	E
A	0	0	0	0	0
B	0	0	0	0	0
C	1	0	0	1	0
D	1	0	0	0	0
E	0	1	1	0	0

Obrázek 14 Matice susednosti grafu na Obrázku 8

1.4 Typické vlastnosti komplexních sítí

Těmito typickými vlastnostmi se komplexní sítě liší právě od náhodných grafů, které tyto vlastnosti postrádají.[1][10]

1.4.1 Malé vzdálenosti (vlastnost malého světa, small world fenomenon)

Vlastnost také zvaná jako „Problém malého světa“ spočívá v tom, že každý uzel je se všemi ostatními uzly v síti propojen přes malý počet hran. Méně formálně je tedy možné se

z libovolně vybraného uzlu dostat na libovolný jiný uzel přes velmi malý počet hran. Obecně lze tedy říci, že čím větší počet propojení v rámci sítě existuje, tím je síť komplexnější a tím je také průměrná vzdálenost mezi uzly menší.[1][10]

Milgramův experiment

Tato vlastnost byla prokázána u známostí lidí v celosvětovém měřítku, a to v populárním experimentu amerického sociálního psychologa Stanleyho Milgrama z roku 1967 (nezaměňovat s experimentem téhož psychologa z roku 1963 ohledně elektrošoků a poslušnosti k autoritě).[1][2]

Bylo dokázáno, že kterýkoliv člověk se „zná“ s libovolným jiným člověkem prostřednictvím řetězce průměrně šesti známostí. Tedy kamarád kamaráda kamaráda kamaráda kamaráda mého kamaráda může být americký prezident stejně jako vůdce teroristické organizace. Milgramův experiment spočíval v tom, že Milgram rozdál spoustu dopisů svým známým a poprosil je, aby se na dopis podepsali a předali jej dále svým známým, kteří se na dopis také podepíší, a to tak, aby se dopisy bez využití poštovních služeb dostaly co nejdříve k Milgramovu příteli do Bostonu. Milgram následně vyhodnotil, že průměrný počet předání dopisu byl šest. Výsledek tohoto experimentu následně potvrdilo mnoho dalších nezávislých experimentů.[1][2]

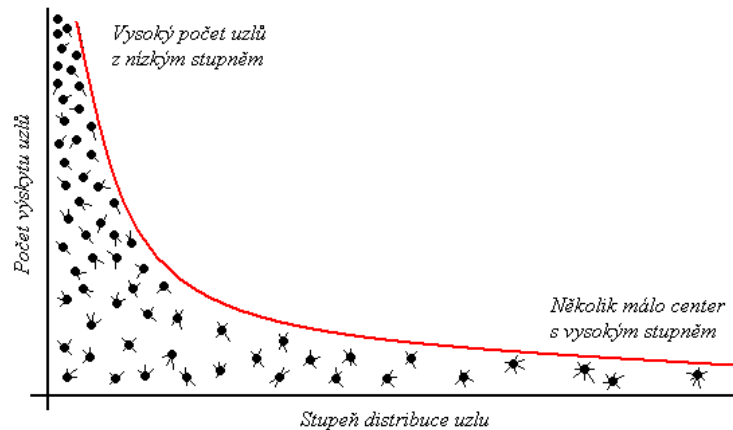
1.4.2 Shlukování (clustering)

Shlukem se rozumí jakási „komunita“ několika uzlů s množstvím vzájemných intenzivních propojení, zatímco počet propojení mezi jednotlivými komunitami v síti je velmi malý. V rámci komunity tedy narůstá pravděpodobnost vzájemného propojení doposud nepropojených uzlů. Tento jev je velmi dobře patrný na Obrázku 5, kde lze na první pohled rozlišit dvě komunity bělochů a dvě komunity černochů. Počet velkých komunit bývá malý, zatímco počet malých komunit velký.[1][2]

Síť je teoreticky možné „rozřezat“ na jednotlivé komunity a v rámci každé z těchto komunit je opět možné nalézt shluky. Platí zde tedy fraktální princip, který je však dobře patrný pouze u rozsáhlých sítí. S praktickým návrhem algoritmu pro vyhledávání komunit je však problém, neboť se lidé zatím na definici pojmu „komunita“ neshodli ani formálně, tedy logicky zatím nemůže existovat ani žádný deterministický algoritmus.[1][2]

1.4.3 Bezškálovost (scale-free)

Bezškálovost sítí volně navazuje na shlukování, neboť říká, že několik málo uzlů v síti (tzv. „hubů“) má velký počet propojení, zatímco ostatní uzly mají propojení málo. Distribuce stupňů uzlů v komplexních sítích se tedy řídí mocninným zákonem (viz Obrázek 15).[1][10]



Obrázek 15 Mocninný zákon distribuce stupňů uzlů jako důkaz bezškálovosti komplexních sítí[3]

1.5 Centrality komplexních sítí

Centrality komplexních sítí představují vlastnosti, které ohodnocují každý uzel v síti podle toho, jak moc je v rámci sítě důležitý. Názvy centralit se většinou v české literatuře nepřekládají, proto budou i zde ponechány anglické termíny. Definice výpočtů těchto centralit pro nevážené sítě popisuje Tabulka 3 a pro vážené sítě Tabulka 4 a Tabulka 5.[11]

1.5.1 Degree centralita (DC)

Jednoduchá degree centralita je založena na myšlence, že důležité uzly musí být v rámci sítě velmi aktivní, a to v tom smyslu, že mají velký počet přímých propojení s dalšími uzly. Degree centralita tedy odpovídá počtu sousedů daného uzlu, tj. jeho stupni.[11]

1.5.2 Closeness centralita (CC)

Closeness vyjadřuje jak blízko je daný uzel ke všem ostatním uzlům. Jinými slovy, jak rychle dokáže komunikovat se všemi ostatními uzly v síti, ne jen se svými přímými sousedy. Vzdálenost či rychlost komunikace je dána nejmenším počtem hran (nejkratší cestou), které dva zkoumané uzly spojují.[11]

1.5.3 Betweenness centralita (BC)

Interakce mezi dvěma nesousedními uzly závisí zejména na struktuře hran mezi nimi. Z tohoto důvodu mají prostřední uzly strategickou kontrolu a velký vliv na ostatní. Uzel je tedy pro síť důležitý, pokud leží na cestě mezi velkým množstvím jiných uzlů. Betweenness centralita pak vyjadřuje počet nejkratších cest, které daným uzlem prochází.[11]

Tabulka 3 Definice centralit pro nevážený graf[12]

Centrality Name	Description	Reference
Betweenness Centrality (BC)	$C_B(u) = \sum_{s \neq u \neq t} \frac{\rho(s, u, t)}{\rho(s, t)}$ <p>$\rho(s, t)$ is the total number of shortest paths from node s to node t $\rho(s, u, t)$ is the number of those paths that pass through u</p>	Anthonisse, et al., 1971
Closeness Centrality (CC)	$C_C(u) = \frac{ N_u - 1}{\sum_{v \in V} dist(u, v)}$ <p>N_u is the number of node u's neighbors. $dist(u, v)$ is the distance of the shortest path from node u to node v.</p>	Sabidussi, et al., 1966
Degree Centrality (DC)	$C_D(u) = N_u $ <p>N_u is the number of node u's neighbors.</p>	Jeong, et al., 2001
Eigenvector Centrality (EC)	$C_E(u) = \alpha_{max}(u)$ <p>α_{max} is the eigenvector corresponding to the largest eigenvalue. of the adjacency matrix A's</p>	Bonacich, et al., 1987
Local Average Connectivity-based method (LAC)	$LAC(u) = \frac{\sum_{\omega \in N_u} deg_{C_u}(\omega)}{ N_u }$ <p>C_u is the subgraph induced by N_u, $deg_{C_u}(\omega)$ is the number of nodes is directly connected in C_u.</p>	Li, et al., 2011
Network Centrality (NC)	$C_N(u) = \sum_{v \in N_u} ECC(u, v) = \sum_{v \in N_u} \frac{ N_u \cap N_v }{\min(N_u - 1, N_v - 1)}$ <p>N_u is the node set containing all the neighbors of node u. $N_u \cap N_v$ is the node set containing all the common neighbors of node u and node v.</p>	Wang, et al., 2012
Subgraph Centrality (SC)	$C_S(u) = \sum_{l=0}^{\infty} \frac{\mu_l(u)}{l!} = \sum_{v=1}^N (v_v^u)^2 e^{\lambda_v}$ <p>$\mu_l(u)$ is the uth diagonal entry of the lth power of the adjacency matrix of network. v_1, v_2, \dots, v_N is an orthonormal basis of R^N composed by eigenvectors of A associated to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$ v_v^u is the uth component of v_v</p>	Estrada, et al., 2005
Information Centrality (IC)	$C_I(u) = \left[\frac{1}{ V } \sum_{v \in V} \frac{1}{I_{u,v}} \right]^{-1}$ $I_{u,v} = (c_{u,u} + c_{v,v} - 2 * c_{u,v})^{-1}$ $C = (c_{u,v}) = (D - A + J)^{-1}$ <p>A is the adjacency matrix of network, D is the diagonal matrix of all nodes' degree, J is a matrix whose all elements are 1.</p>	Stephenson, et al., 1989

Tabulka 4 Definice centralit pro vážený graf, 1. část[12]

Centrality Name	Description
Betweenness Centrality (with weight)	$cost(s, t) = \frac{1}{w_{(s,t)}} (w_{(s,t)} > 0)$ <p>$w_{(s,t)}$ is the weight of edge from node s to node t,</p> $dist_w(u, v) = \sum_{(s,t) \in P} cost(s, t)$ <p>The shortest path from node s to node t is the path whose $dist_w(u, v)$ is smallest.</p> $C_{B_w}(u) = \sum_{s \neq u \neq t} \frac{\rho(s, u, t)}{\rho(s, t)}$ <p>$\rho(s, t)$ is the total number of shortest paths from node s to node t.</p> <p>$\rho(s, u, t)$ is the number of those paths that pass through u</p>
Closeness Centrality (with weight)	$cost(s, t) = \frac{1}{w_{(s,t)}} (w_{(s,t)} > 0)$ <p>$w_{(s,t)}$ is the weight of edge from node s to node t,</p> $dist_w(u, v) = \sum_{(s,t) \in P} cost(s, t)$ <p>$dist_w(u, v)$ is the weighted distance of the shortest path from node u to node v.</p> $C_{C_w}(u) = \frac{ N_u - 1}{\sum_{v \in V} dist_w(u, v)}$ <p>N_u is the number of node u's neighbors.</p>
Degree Centrality (with weight)	$C_{D_w}(u) = \sum_{v \in N_u} w(u, v)$ <p>N_u is the node set containing all the neighbors of node u.</p> <p>$w(u, v)$ is the weight of the edge connected node u and node v.</p>
Eigenvector Centrality (with weight)	$C_{E_w}(u) = \alpha_{max}(u)$ <p>α_{max} is the eigenvector corresponding to the largest eigenvalue of the weighted adjacency matrix A_w</p>
Local Average Connectivity-based method (with weight)	$LAC_w(u) = \frac{\sum_{s \in N_u} \sum_{t \in N_s \cap N_u} w(s, t)}{ N_u }$ <p>N_u is the node set containing all the neighbors of node u.</p> <p>$w(s, t)$ is the weight of the edge connected node s and node t.</p>

Tabulka 5 Definice centralit pro vážený graf, 2. část[12]

<p>Network Centrality (with weight)</p>	$C_{N_w}(u) = \sum_{v \in N_u} \frac{\sqrt{\sum_{s \in N_u \cap N_v} w(u, s) * \sum_{t \in N_u \cap N_v} w(v, t)}}{\min(\sum_{s \in N_u} w(u, s) - 1, \sum_{t \in N_v} w(v, t) - 1)}$ <p>N_u is the node set containing all the neighbors of node u.</p> <p>$N_u \cap N_v$ is the node set containing all the common neighbors of node u and node v.</p> <p>$w(u, s)$ is the weight of the edge connected node u and node s.</p>
<p>Subgraph Centrality (with weight)</p>	$C_{S_w}(u) = \sum_{l=0}^{\infty} \frac{\mu_l(u)}{l!} = \sum_{v=1}^N (v_v^u)^2 e^{\lambda_v}$ <p>$\mu_l(u)$ is the u^{th} diagonal entry of the l^{th} power of the weighted adjacency matrix of network.</p> <p>v_1, v_2, \dots, v_N is an orthonormal basis of R^N composed by eigenvectors of A associated to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$</p> <p>$v_v^u$ is the uth component of v_v</p>
<p>Information Centrality (with weight)</p>	$C_{I_w}(u) = \left[\frac{1}{ V } \sum_{v \in V} \frac{1}{I_{u,v}} \right]^{-1}$ $I_{u,v} = (c_{u,u} + c_{v,v} - 2 * c_{u,v})^{-1}$ $C = (c_{u,v}) = (D_w - A_w + J_w)^{-1}$ $\frac{1}{I_{u,u}} = 0$ <p>A is the weighted adjacency matrix of network, D is the diagonal matrix of all nodes' degree, J is a matrix where all elements are 1.</p>

2 HEURISTICKE ALGORITMY

Heuristické algoritmy v rámci svého běhu několikrát volí, kterou cestou bude výpočet dále pokračovat, a tak algoritmus pro stejné vstupy vrací pouze podobné výstupy. Heuristika je tedy protipólem deterministických algoritmů, kde výpočet probíhá vždy podle předem jasné daných pravidel, tudíž pro stejné vstupy deterministický algoritmus vrací vždy stejné výstupy.[13]

Motivací pro uchýlení se k heuristickému řešení problému bývá většinou velká časová náročnost deterministického řešení nebo k danému problému dokonce (zatím) žádný deterministický algoritmus neexistuje. Pro využití heuristiky navíc autor nemusí znát vnitřní strukturu problému a může pracovat pouze na úrovni vstupů a výstupů, což umožňuje například programátorům bez hlubších fyzikálních znalostí hledat přibližná řešení komplikovaných fyzikálních problémů.[13]

Heuristické metody nezaručují nalezení optimálního řešení a výsledek bývá často jen přibližný, nicméně při zprůměrování několika výsledků se odhad řešení může výrazně zlepšit, nebo ideálně po ohodnocení výsledků účelovou funkcí (pokud je k dispozici), je za výsledek prohlášen nejlepší pokus.[13]

Heuristika je tedy rychlá a relativně jednoduchá metoda pro nalezení dostatečně kvalitního řešení bez nutnosti znalosti vnitřní struktury problému, a to za uspokojivý čas, což se hodí zejména při krizových situacích, kde čas hraje velmi významnou roli.[13]

2.1 Hejnové algoritmy

Inspirací pro vývoj těchto algoritmů bylo přirozené chování hejn ryb, ptáků nebo stád nejrůznějších zvířat, které, aniž by byla centrálně řízena nějakým vůdcem, jsou často schopny se pohybovat synchronizovaně, a například se tak vymanit predátorovi. Dalším vzorem mohly být mravenčí nebo včelí kolonie, které jsou bez globálního dohledu schopny přežívat a rychle si mezi jedinci vyměňovat důležité informace o tom, kde se například nachází nové zdroje jídla.[14]

Hejnové algoritmy spadají do oblasti umělé inteligence. Hejno se skládá z populace agentů, kteří nejsou řízeni centrálně, nicméně dle předem daných jednoduchých pravidel reagují na své okolí a na sebe navzájem, což se zvenčí může jevit, jakoby hejno centrálně řízeno bylo.[14]

2.1.1 Základní pojmy

V této části budou vysvětleny základní pojmy týkající se hejnových algoritmů jako je hyperplocha, účelová funkce, testovací funkce, jedinec nebo populace.

Hyperplocha

Hyperplocha představuje n -dimenzionální prostor. Poloha každého objektu v hyperploše je tedy vyjádřena pomocí vektoru n souřadnic. Hodnoty jednotlivých souřadnic mohou být nezávisle na sobě shora, zdola, či jiným způsobem omezeny.[15]

Účelová funkce

Účelová funkce je tvořena funkčním předpisem, který ohodnocuje kvalitu umístění objektu v hyperploše. Pro definici účelové funkce se nekladou žádné podmínky a zcela tedy závisí na řešiteli problému. Podoba účelové funkce hraje v optimalizačních úlohách klíčovou roli. Ohodnocování objektu účelovou funkcí často bývá úzkým hrdlem algoritmu, proto bývá právě počet ohodnocení účelovou funkcí srovnávacím kritériem výkonnosti algoritmů, anebo podmínkou ukončující běh algoritmu.[15]

Testovací funkce

Testovací funkce představují množinu několika standardních účelových funkcí, na nichž bývá testována správná funkčnost či výkonnost nově vyvíjeného algoritmu, neboť je u těchto funkcí obecně známa hodnota i souřadnice jejich ideálního řešení.[15]

Jedinec

Jedinec je objekt nacházející se v hyperploše. Ohodnocením jeho souřadnic účelovou funkcí se získá jakási kvalita tohoto jedince. Každý jedinec je kandidátem na optimální řešení daného problému popsaného právě účelovou funkcí. V průběhu řešení daného problému jsou méně kvalitní jedinci nahrazováni novými nebo jsou jejich parametry upravovány tak, aby se více podobaly kvalitnějším jedincům, přičemž se předpokládá, že některý z upravených jedinců se stane nejvyšší kvality ze všech a bude vzorem pro další úpravy.[15]

Populace

Populace je tvořena souborem jedinců snažících se o nalezení ideálního či alespoň dostatečně dobrého řešení daného problému. V rámci iterativního procesu snahy o upravování horších jedinců za lepší si populace prochází jednotlivými generacemi odpovídající jednotlivým iteracím. Čím více jedinců se v populaci nachází, tím rychleji a

tím lepší řešení může být nalezeno, avšak daní za tyto výhody bývají zvýšené nároky na výpočetní čas.[15]

2.1.2 Firefly algoritmus (světlučí algoritmus)

Algoritmus Firefly byl inspirován principem výběru ideálního světlučího partnera pro zplození potomstva. Pro většinu druhů světlušek je při vábení partnera charakteristické vydávání krátkých a rytmických záblesků světla. Jev vydávání světla se u živých organismů nazývá bioluminiscence a je výsledkem jistých chemických reakcí v tělech těchto živočichů. Jedná se o studené světlo, neboť proces jeho vytváření je natolik efektivní, že pouze 8 % energie se přemění na teplo, zatímco například u klasické žárovky se na ztrátové teplo přemění celých 96 %.[16]

Světlušky nemusí světlo využívat pouze při námluvách, ale může sloužit i pro výměnu informací mezi jedinci, jako je blízkost dravce, vyznačování teritoria, nebo prostě jen pro zlepšení viditelnosti v prostoru. Své uplatnění tato vlastnost najde také u některých dravých druhů světlušek při lákání kořisti. Tyto světlušky totiž napodobují vábící světelné signály samiček jiného druhu. V neposlední řadě lze bioluminiscenci využít jako nejzazší způsob ochrany pro oslepení predátora zvyklého na tmou, anebo pro trvalou kamufláž, která skryje jejich vlastní stín. Larvy mohou pomocí světla do okolí vysílat varovný signál upozorňující predátora na jedovatost nebo odpornou chuť.[16]

Technické aspekty

Algoritmus vymyslel roku 2007 Xin-She Yang a zjednodušeně lze říci, že se řídí následujícími třemi pravidly:

- 1) Svítivost světlušky je dána hodnotou účelové funkce vyhodnocující kvalitu jejího umístění na hyperploše.
- 2) Atraktivita světlušky pozitivně koreluje s její svítivostí, která ovšem postupně klesá se zvětšující se vzdáleností.
- 3) Každá světluška je na základě pravidel 1) a 2) přitahována ke všem ostatním.[17]

Z výše uvedených pravidel vyplývá, že méně svítivé (resp. přitažlivé) světlušky se budou pohybovat směrem k těm svítivějším (resp. přitažlivějším). Do každého pohybu světlušky je zanesena jistá míra náhody. Pohyb nejzářivější světlušky, která není nikým jiným přitahována, bude tedy ryze náhodný.[17]

Pohyb i -té světlušky k j -té světlušce je popsán následující rovnicí:[17]

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha \varepsilon_i^t$$

Parametr x představuje vektor souřadnic světlušky v hyperploše, t v horním indexu značí stav před pohybem, zatímco $t+1$ stav po pohybu, β_0 je relativní atraktivita světlušky j z pohledu světlušky i při nulové vzdálenosti, e je Eulerovo číslo, konstanta γ odpovídá míře absorpce světla prostředím (čím bude hodnota γ větší, tím méně se budou světlušky přitahovat), r_{ij} představuje vzdálenost světlušky i od světlušky j v hyperploše, α ovlivňuje míru náhody při pohybu světlušky (větší hodnota α znamená větší míru náhody) a ε je vektor náhodných čísel -1 a 1 pro určení směru náhodného pohybu v jednotlivých osách.[17]

Vzdálenost r je v Euklidovském prostoru definována jako:[17]

$$r_{ij} = \sqrt{\sum_{k=1}^{\dim} (x_{i,k} - x_{j,k})^2}$$

Pseudokód

Vygeneruj počáteční populaci o velikosti n světlušek.

Světelná intenzita $I_i = f(x_i)$ světlušky x_i je dána hodnotou účelové funkce $f(x_i)$.

Definuj koeficient světelné absorpce γ .

dokud ($t <$ maximální počet generací) **opakuji**

pro i od 1 do n pro všechny světlušky **opakuji**

pro j od 1 do n pro všechny světlušky **opakuji**

jestliže $I_i < I_j$ **pak**

Posuň světlušku i blíž k j .

Spočítej novou světelnou intenzitu I_i .

konec

konec j

konec i

Najdi nejlepší řešení g .

konec t [17]

Složitost algoritmu

Algoritmus obsahuje jeden cyklus pro iterování generací t , další vnitřní cyklus pro iterování aktivních posouvajících se světlušek i a další vnořený cyklus opět pro iterování všech světlušek v populaci j , díky nimž se může aktivní světluška posouvat. Tento algoritmus má tedy časovou složitost $\Theta(tn^2)$. Populace ve většině případů skýtá relativně malý počet jedinců (např. 50), zatímco počet generací bývá vysoký (např. 1000), aby se stačila rozběhnout vnitřní mechanika algoritmu. V případě nutnosti většího počtu jedinců v populaci bývá algoritmus optimalizován tak, že se na konci každého vytváření nové populace tato nová populace seřadí podle atraktivity, tudíž pak není třeba procházet všechny světlušky a časová složitost klesá na $\Theta(tn \log n)$. [18]

2.2 No Free Lunch teorém

Hejnových algoritmů, stejně jako heuristických algoritmů obecně existuje celá řada. Každý typ algoritmu obsahuje velké množství nejrůznějších variací a optimalizací, nicméně tvrzení známé jako No Free Lunch teorém říká, že neexistuje a ani v budoucnu nemůže existovat univerzální algoritmus, který by nejlépe, nejrychleji a s co nejmenší časovou a paměťovou náročností řešil všechny typy problémů. Na každý problém nebo i množinu dat se totiž lépe hodí jiný algoritmus. [15]

3 POUŽITÉ SOFTWARE NÁSTROJE

V této kapitole jsou popsány softwarové nástroje Wolfram Mathematica, Cytoscape a jeho zásuvné moduly aMatReader a CytoNCA, které byly použity při vypracování této diplomové práce.

3.1 Wolfram Mathematica

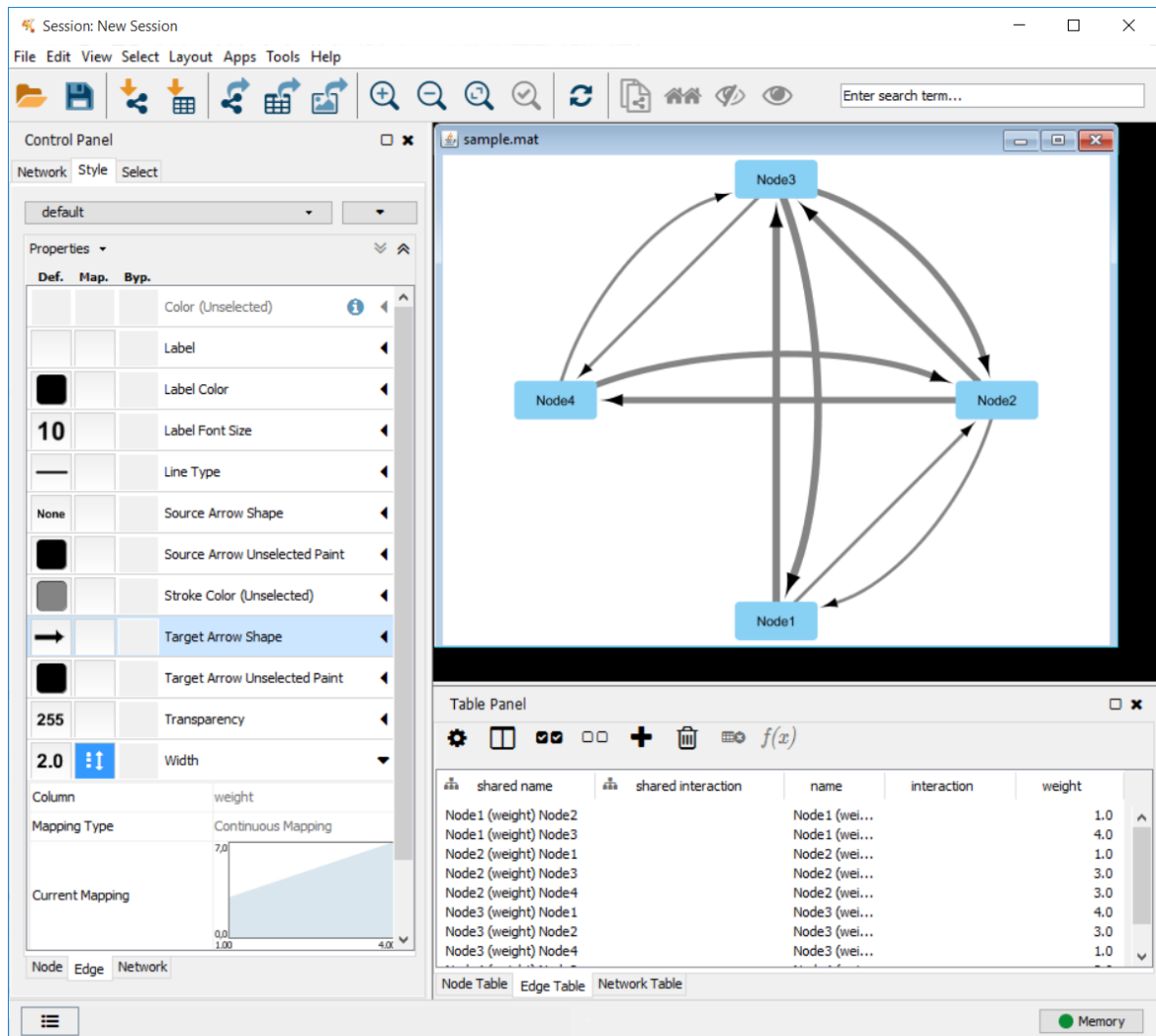
Mathematica představuje proprietární počítačovou aplikaci využívanou, jak již sám název napovídá, pro matematické, vědecké nebo technické výpočty, jejich vizualizaci a také pro názornou výuku na akademické půdě. Mathematica byla napsána v jednoúčelově vytvořeném funkcionálním programovacím jazyku Wolfram, výpočet zde tedy spočívá ve vyhodnocení matematických funkcí a není tedy problém provádět symbolické výpočty jako je například derivování či integrování.[19]

Mathematica se skládá z výpočetního jádra, které vyhodnocuje výrazy a vrací výsledky, a front endu, který zajišťuje grafické uživatelské rozhraní (GUI), které uživatelům umožňuje vkládat výrazy a následně zobrazovat nebo vykreslovat výsledky. Mathematica je multiplatformní aplikace, je tedy dostupná na všech hlavních desktopových platformách, mezi něž patří Microsoft Windows, Linux a Macintosh.[19]

Autorem programu je britský fyzik a matematik Stephen Wolfram, přičemž aplikace je v současné době spravována společností Wolfram Research. Z části podobným konkurenčním programem je populární Matlab od společnosti MathWorks.[19]

3.2 Cytoscape

Cytoscape představuje bioinformační softwarovou platformu s otevřeným zdrojovým kódem pro vytváření, úpravu, výpočet parametrů a vizualizaci nejen biologických sítí (viz Obrázek 16). Ačkoliv byl Cytoscape původně navržen pro biologický výzkum, postupem času se vyvinul v platformu pro analýzu a vizualizaci komplexních sítí obecně. Možnosti tohoto nástroje jsou rozšiřitelné pomocí mnoha zásuvných modulů dostupných na Cytoscape App Store (<http://apps.cytoscape.org/>), které například mohou upravovat způsob vizualizace, přidávat podporu pro další souborové formáty, propojit aplikaci s externími databázemi nebo dle zadaných parametrů vyhledávat v rozsáhlých sítích.[20]

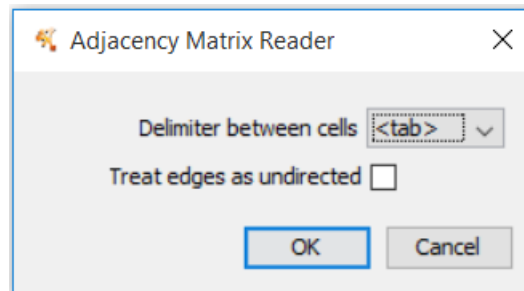


Obrázek 16 Screenshot aplikace Cytoscape s načtenou orientovanou a váženou komplexní sítí

3.2.1 aMatReader

Tento zásuvný modul slouží pro načítání matic sousednosti z textových souborů do aplikace Cytoscape. Vstupní textový soubor musí mít příponu „mat“, aby nedošlo ke konfliktu s vestavěnými nástroji pro čtení softwaru Cytoscape. Oddělovač jednotlivých buněk lze zvolit mezi tabulátorem, mezerou, čárkou a rourou (viz Obrázek 17). Zaškrtnutím příslušného políčka při načítání souboru lze hrany považovat za neorientované, tedy matici za symetrickou (viz Obrázek 17). Aplikace chápe celý první řádek a celý první sloupec textového souboru jako názvy jednotlivých uzlů. Objekt na prvním řádku a prvním sloupci je aplikací ignorován. Hodnoty v buňkách matice sousednosti jsou přímo brány jako váhy příslušných hran. Neexistující propojení mezi danými uzly je nutno interpretovat chybějící

hodnotou, nikoliv hodnotou „0“. Stejně pořadí řádků a sloupců nemusí být dodrženo. Zásuvný modul vyvinul John Morris v roce 2015.[21][22]



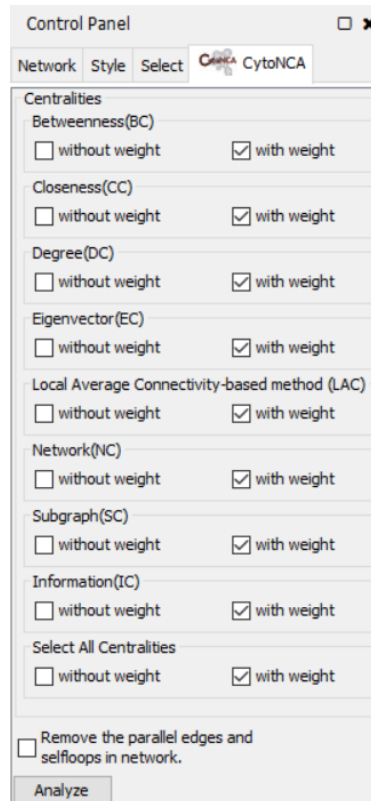
Obrázek 17 Volba oddělovače buněk a orientace hran při načítání vstupního souboru zásuvným modulem aMatReader

Struktura vzorového vstupního souboru

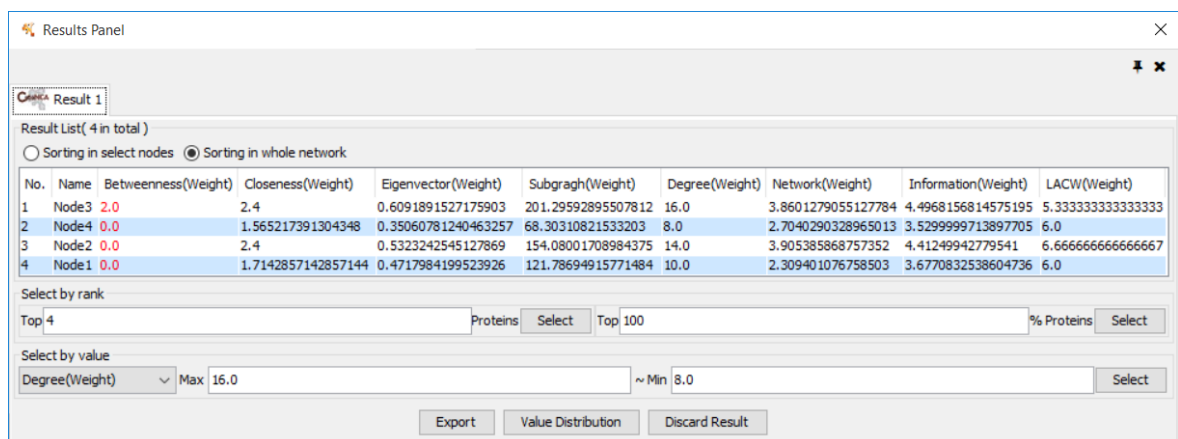
```
#Simple tab-delimited adjacency matrix
      Node1 Node2 Node3 Node4
Node1      1     4
Node2  1     3     3
Node3  4     3     1
Node4      3     1     [22]
```

3.2.2 CytoNCA

Zásuvný modul CytoNCA je nástrojem pro výpočet a vizualizaci osmi typů centralit pro vážené i nevážené komplexní sítě (viz Obrázek 18 a Obrázek 19). Vyhodnocované centrality jsou *Betweenness*, *Closeness*, *Degree*, *Eigenvector*, *Local Average Connectivity-based method*, *Network*, *Subgraph* a *Information*. Výpočty u nevážených sítí probíhají podle vzorců v Tabulce 3, u vážených sítí pak podle Tabulky 4 a Tabulky 5.[23]



Obrázek 18 Ovládací panel pro volbu vyhodnocovaných centralit pomocí CytoNCA



Obrázek 19 Výsledkový panel pro vizualizaci vypočtených centralit pomocí CytoNCA

II. PRAKTICKÁ ČÁST

4 HLEDÁNÍ VHODNÝCH KOMPLEXNÍCH SÍTÍ

Před možností samotného modelování a predikce chování komplexních sítí bylo zapotřebí získat vhodná data popisující komplexní sítě, na kterých bude probíhat následný vývoj a testování heuristického algoritmu modelujícího tyto vzorové komplexní sítě.

4.1 Citační síť

Data komplexní sítě odpovídající citační síti by měla ideálně mít hrany jednak orientované, aby bylo zřejmé, který autor (resp. článek) citoval kterého autora (resp. článek), a jednak vážené, aby bylo možné rozlišovat počet citací a ne pouze zdali takovýto citační pár někdy existoval. Pro možnost predikce dynamiky citační sítě je navíc nutné znát časový okamžik vzniku každé existující hrany.

4.1.1 Scopus

Tato mezinárodní citační databáze vědeckých článků se nachází na Internetové adrese <https://www.scopus.com/> a obsahuje všechny potřebné informace pro modelování a predikci citačních sítí, tedy orientaci, váhu i časové razítko všech hran. Jediným nedostatkem je absence nástroje pro export těchto dat v textové podobě. Data je zde možné zobrazit pouze v tabulce či si je nechat vykreslit v mapě, avšak samotné hodnoty je odtud nutné opisovat ručně, což je pro malé sítě například o 10 uzlech otázka několika minut, pro větší sítě by to však mohl být problém, neboť časová složitost procesu získávání dat tímto způsobem narůstá kvadraticky. Autorovi této diplomové práce se však ani po konzultaci s jedním z nejerudovanějších lidí v tomto oboru v České republice, panem docentem Milošem Kudělkou působícím na Vysoké škole báňské – Technické univerzitě Ostrava, nepodařilo najít lepší zdroj dat, který by splňoval všechny výše uvedené podmínky, a zároveň by z něho bylo možné data elegantně exportovat, proto byla jako zdroj dat citačních sítí použita databáze Scopus.

4.2 Sociální síť

Oproti citačním sítím není třeba, aby sociální síť obsahovaly informaci o orientaci hran. Jinými slovy není důležité, kdo přátelství či spolupráci mezi dvěma lidmi inicioval, ale pouze to, zda toto přátelství existuje či nikoliv. Rovněž není zvykem, aby měly sociální síť vážené hrany, tedy aby nějakým způsobem vyjadřovaly kvalitu či četnost kontaktu dvou objektů. Jedinou podmínkou pro možnost predikce tedy zůstává nutnost informace o čase, kdy hrana

vznikla, avšak čím více informací bude o síti známo, tím snadnější a kvalitnější predikce bude.

4.2.1 SNAP

Na webové stránce <http://snap.stanford.edu/data/index.html> Stanfordské univerzity, kterou spravuje Jure Leskovec a Andrej Krevl se nachází velké množství datových souborů komplexních sítí od sociálních sítí jako je Facebook, Twitter, Google+, přes citační sítě amerických patentů, spoluautorských sítí týkajících se fyzikálních vědeckých článků až po silniční komunikační sítě amerických států Kalifornie, Pensylvánie, Texas, a mnoho dalších typů komplexních sítí. U datových souborů internetových sociálních sítí jako je Facebook však je problémem absence časového razítka vzniku hran v síti, tudíž jsou tato data pro predikci zcela nevhodná. Z tohoto důvodu byla jako sociální síť zvolena komplexní síť spoluautorství nad fyzikálními vědeckými články *ca-HepPh*.

5 TESTOVACÍ FUNKCE

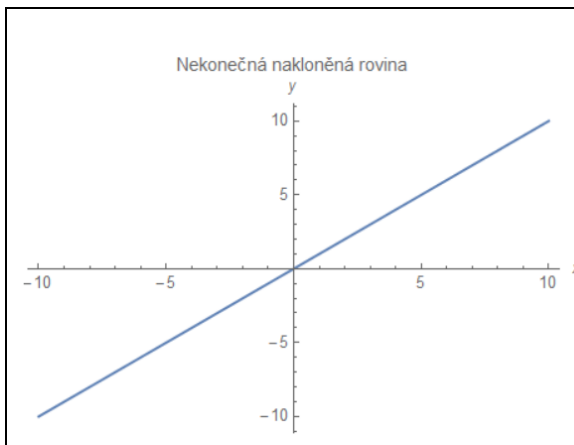
Pro potřebu vývoje a testování nově vznikajícího heuristického algoritmu bylo naprogramováno několik testovacích funkcí, na kterých bude možné ověřovat správnou funkčnost a ladit výkonnost nového algoritmu.

5.1 Nekonečná nakloněná rovina (rampa)

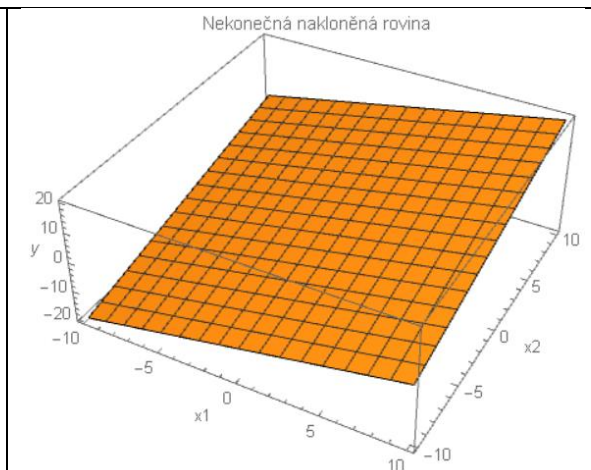
$$\text{Funkční předpis: } y = \sum_{i=1}^{\dim} x_i$$

$$\text{Globální minimum: } y(-\infty, -\infty, \dots, -\infty) = -\infty$$

Grafy pro jednu (resp. dvě) nezávislé proměnné nekonečné nakloněné roviny lze vidět na Obrázku 20 (resp. Obrázku 21).



Obrázek 20 Nekonečná nakloněná rovina, 2D graf na intervalu $\langle -10; 10 \rangle$



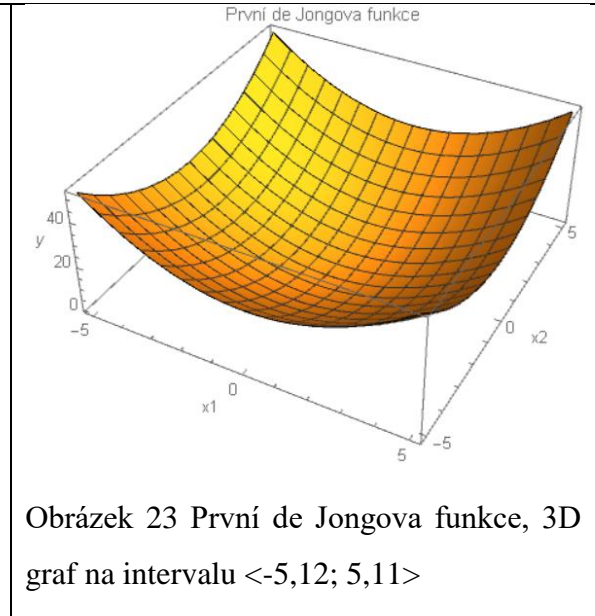
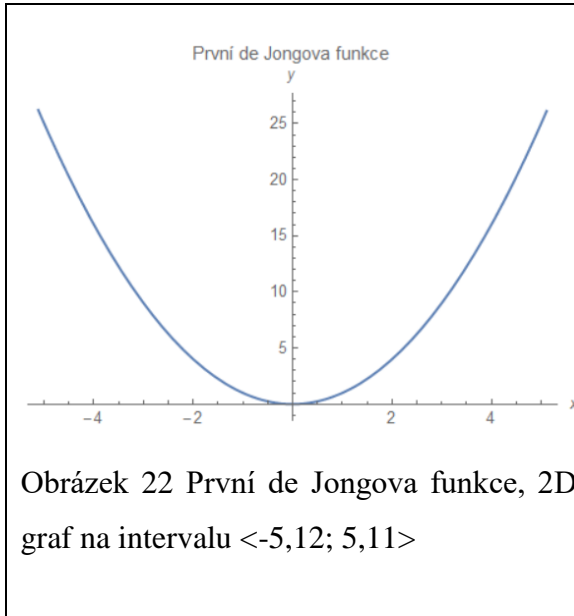
Obrázek 21 Nekonečná nakloněná rovina, 3D graf na intervalu $\langle -10; 10 \rangle$

5.2 První de Jongova funkce (sféra)

$$\text{Funkční předpis: } y = \sum_{i=1}^{\dim} x_i^2 \quad [15]$$

$$\text{Globální minimum: } y(0,0,\dots,0) = 0 \quad [15]$$

Grafy pro jednu (resp. dvě) nezávislé proměnné První de Jongovy funkce lze vidět na Obrázku 22 (resp. Obrázku 23).

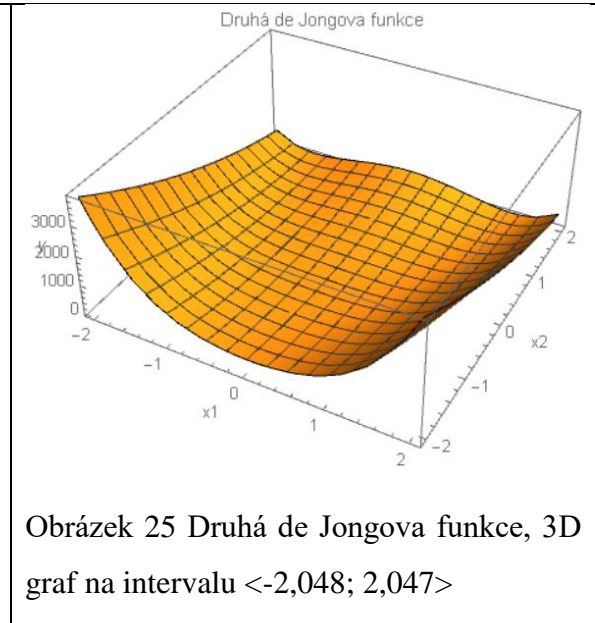
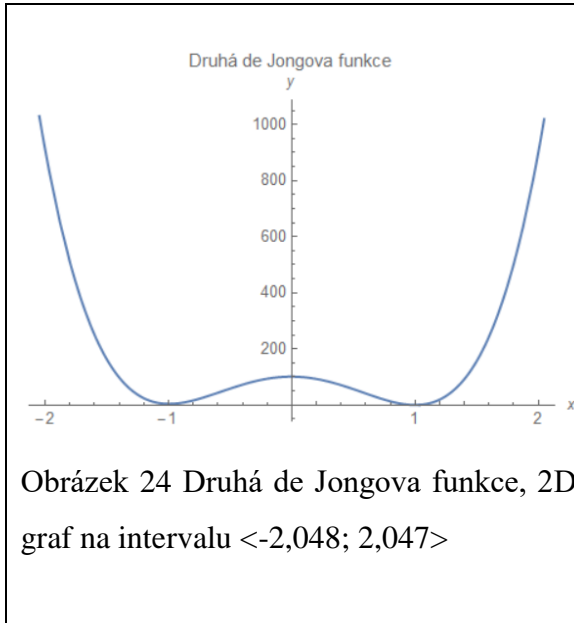


5.3 Druhá de Jongova funkce (Rosenbrockovo sedlo, Banánová funkce)

$$\text{Funkční předpis: } y = \sum_{i=1}^{\text{dim}-1} 100(x_i^2 - x_{i+1})^2 + (1 - x_i)^2 \quad [15]$$

$$\text{Globální minimum: } y(1,1,\dots,1) = 0 \quad [15]$$

Grafy pro jednu (resp. dvě) nezávislé proměnné Druhé de Jongovy funkce lze vidět na Obrázku 24 (resp. Obrázku 25). Funkční předpis je však definován pro minimálně dvě nezávislé proměnné, a tak bylo nutné pro možnost vykreslení grafu jedné nezávislé proměnné zafixovat druhou nezávislou proměnnou na určité hodnotě. V tomto případě byla zvolena souřadnice globálního minima, tedy hodnota „1“.

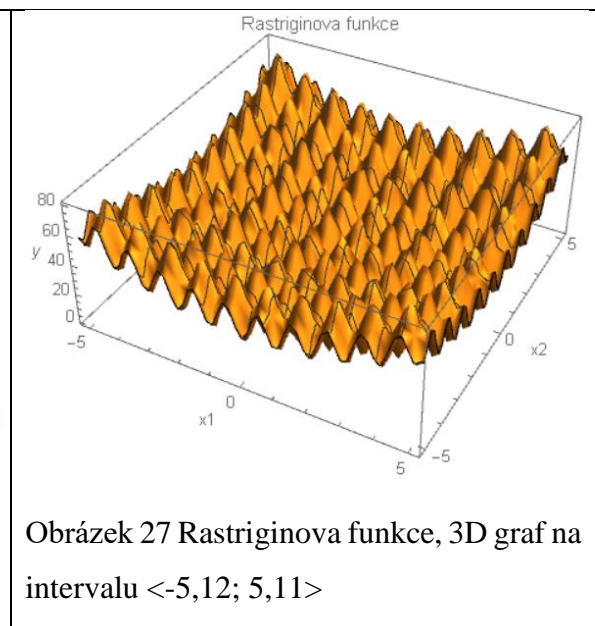
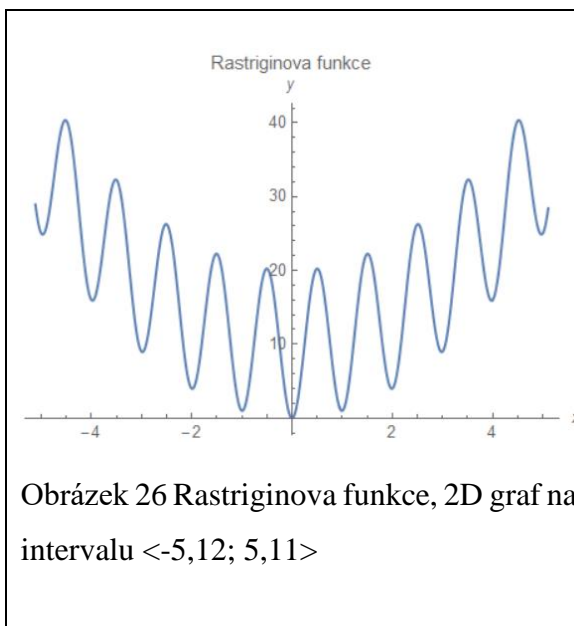


5.4 Rastriginova funkce

$$\text{Funkční předpis: } y = 10 * \text{dim} + \sum_{i=1}^{\text{dim}} [x_i^2 - 10 * \cos(2\pi x_i)] \quad [24]$$

$$\text{Globální minimum: } y(0,0,\dots,0) = 0 \quad [24]$$

Graf pro jednu (resp. dvě) nezávislé proměnné Rastriginovy funkce lze vidět na Obrázku 26 (resp. Obrázku 27).

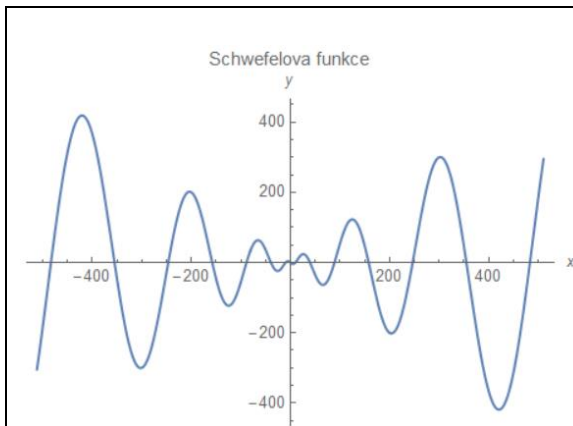


5.5 Schwefelova funkce

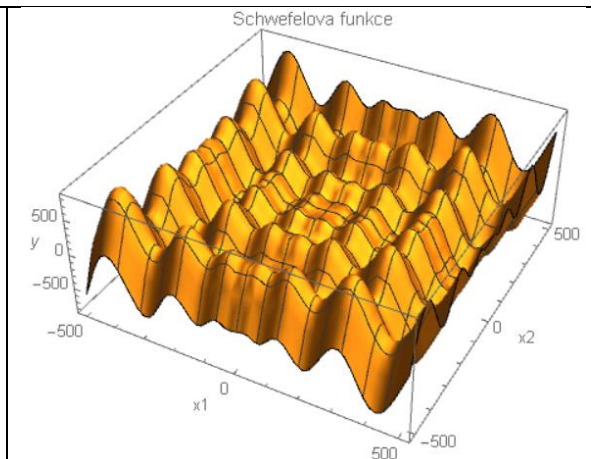
$$\text{Funkční předpis: } y = \sum_{i=1}^{\text{dim}} -x_i * \sin\left(\sqrt{|x_i|}\right) \quad [15]$$

$$\text{Globální minimum: } y(420,969;420,969;\dots;420,969) = -418,983 * \text{dim} \quad [15]$$

Graf pro jednu (resp. dvě) nezávislé proměnné Schwefelovy funkce lze vidět na Obrázku 28 (resp. Obrázku 29).



Obrázek 28 Schwefelova funkce, 2D graf na intervalu $\langle -512, 511 \rangle$



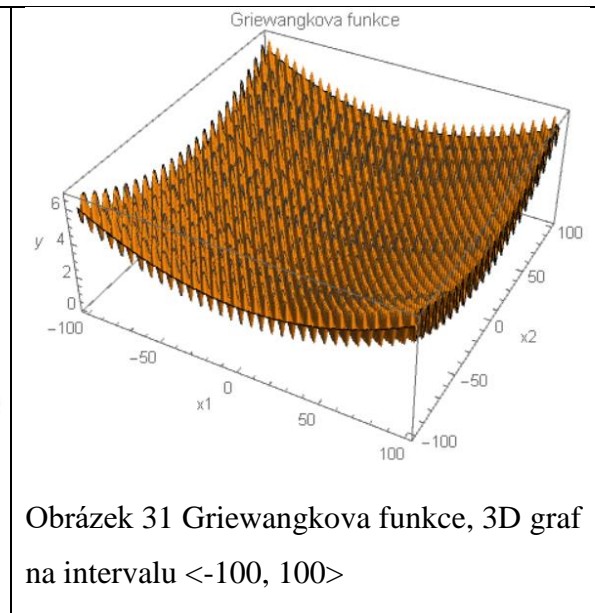
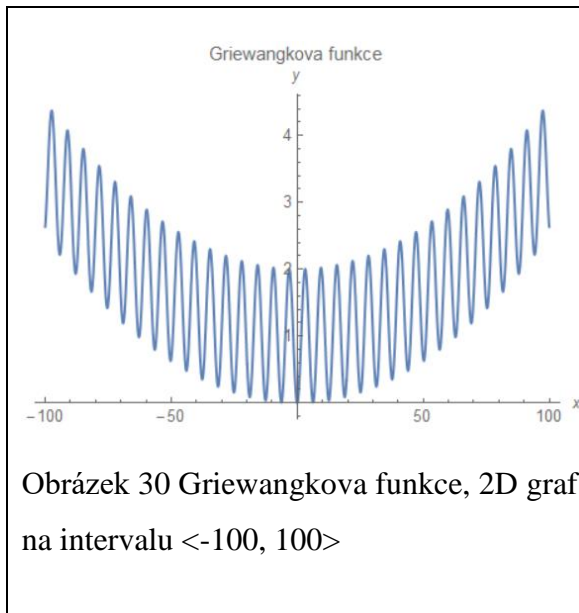
Obrázek 29 Schwefelova funkce, 3D graf na intervalu $\langle -512, 511 \rangle$

5.6 Griewangkova funkce

$$\text{Funkční předpis: } y = 1 + \sum_{i=1}^{\text{dim}} \frac{x_i^2}{4000} - \prod_{i=1}^{\text{dim}} \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad [15]$$

$$\text{Globální minimum: } y(0,0,\dots,0) = 0 \quad [15]$$

Graf pro jednu (resp. dvě) nezávislé proměnné Griewangkovy funkce lze vidět na Obrázku 30 (resp. Obrázku 31).

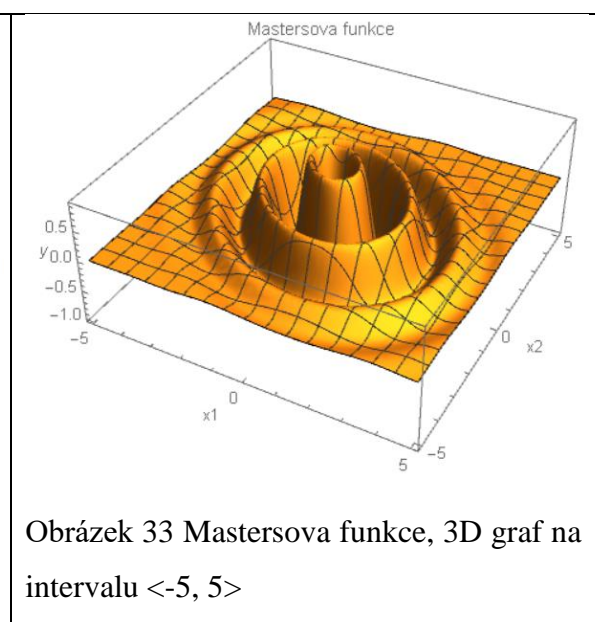
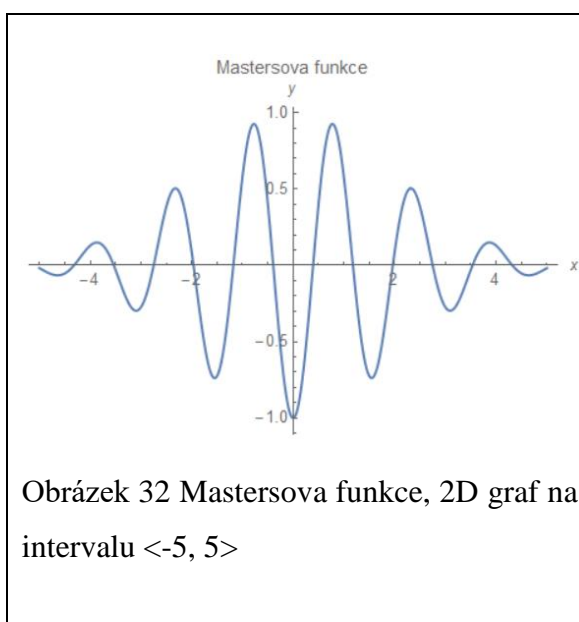


5.7 Mastersova funkce

$$\text{Funkční předpis: } y = \sum_{i=1}^{\dim-1} e^{\frac{-(x_i^2 + x_{i+1}^2 + 0,5x_i x_{i+1})}{8}} \cos\left(4\sqrt{x_i^2 + x_{i+1}^2 + 0,5x_i x_{i+1}}\right) \quad [15]$$

$$\text{Globální minimum: } y(0,0,\dots,0) = -(\dim-1) \quad [15]$$

Graf pro jednu (resp. dvě) nezávislé proměnné Mastersovy funkce lze vidět na Obrázku 32 (resp. Obrázku 33). Funkční předpis je však definován pro minimálně dvě nezávislé proměnné, a tak bylo nutné pro možnost vykreslení grafu jedné nezávislé proměnné zafixovat druhou nezávislou proměnnou na určité hodnotě. V tomto případě byla zvolena souřadnice globálního minima, tedy hodnota „0“.

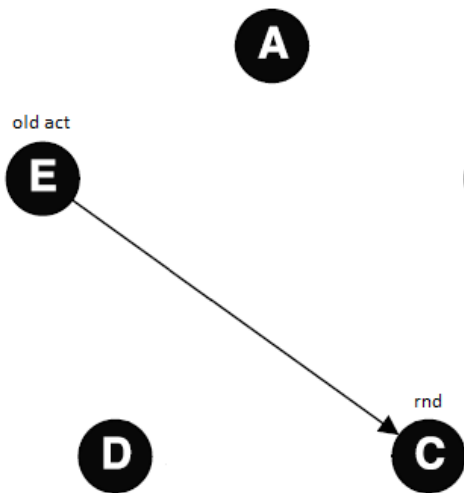


6 PŘEVOD DYNAMIKY HEJNOVÉHO ALGORITMU NA KOMPLEXNÍ SÍŤ

Pro modelování komplexních sítí pomocí dynamiky hejnového algoritmu byl zvolen algoritmus Firefly, neboť má pro tento proces vhodné vlastnosti. Jednak se vytváření nové pozice daného jedince účastní pouze dva jedinci z populace, jejichž případnou „úspěšnou“ interakci lze v komplexní síti dobře zaznamenat, a jednak není jedním ze dvou interagujících jedinců vždy zatím nejlepší jedinec v rámci celé populace jako je tomu u řady ostatních hejnových algoritmů, což by mohlo mít za důsledek vytváření pouze „nevyvážených“ sítí či sítí s náhodnou a špatně „řiditelnou“ strukturou.

Každý jedinec algoritmu Firefly je reprezentován právě jedním uzlem komplexní sítě. Pokud je v rámci běhu algoritmu výsledkem interakce dvou jedinců nový jedinec, jehož pozice v rámci hyperplochy má dle účelové funkce lepší ohodnocení než původní pozice posouvaného (aktivního) jedince, dojde v matici sousednosti popisující komplexní síť k inkrementování hodnoty prvku matice na řádku odpovídajícímu aktivnímu jedinci *act* a sloupci odpovídajícímu „náhodnému“ jedinci *rnd*, čímž v komplexní síti vznikne nová hrana mezi těmito uzly anebo se zvětší váha této již existující hrany.

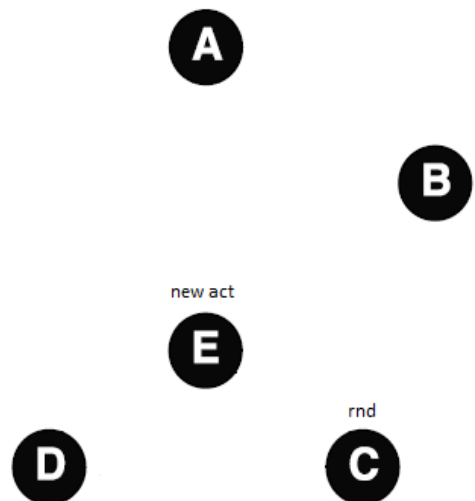
Obrázek 34 představuje pět jedinců algoritmu Firefly, přičemž zde právě dochází k interakci mezi aktivním jedincem E a „náhodným“ jedincem C. Pozice jedince C je ohodnocena lépe než pozice jedince E, a tak se jedinec E posouvá směrem k jedinci C (viz Obrázek 35). Pokud bude nová pozice jedince E ohodnocena účelovou funkcí lépe než původní pozice jedince E, pak lze tuto interakci prohlásit za „úspěšnou“ a inkrementovat příslušnou hodnotu v matici sousednosti komplexní sítě, neboť se jedinec E posunul na lepší pozici právě díky jedinci C (viz Obrázek 36 a Obrázek 37).



Obrázek 34 Původní pozice jedinců algoritmu Firefly před „úspěšnou“ interakcí aktivního jedince E s „náhodným“ jedincem C

	A	B	C	D	E
A	0	0	0	0	0
B	0	0	0	0	0
C	0	0	0	0	0
D	0	0	0	0	0
E	0	0	0	0	0

Obrázek 36 Matice sousednosti před „úspěšnou“ interakcí



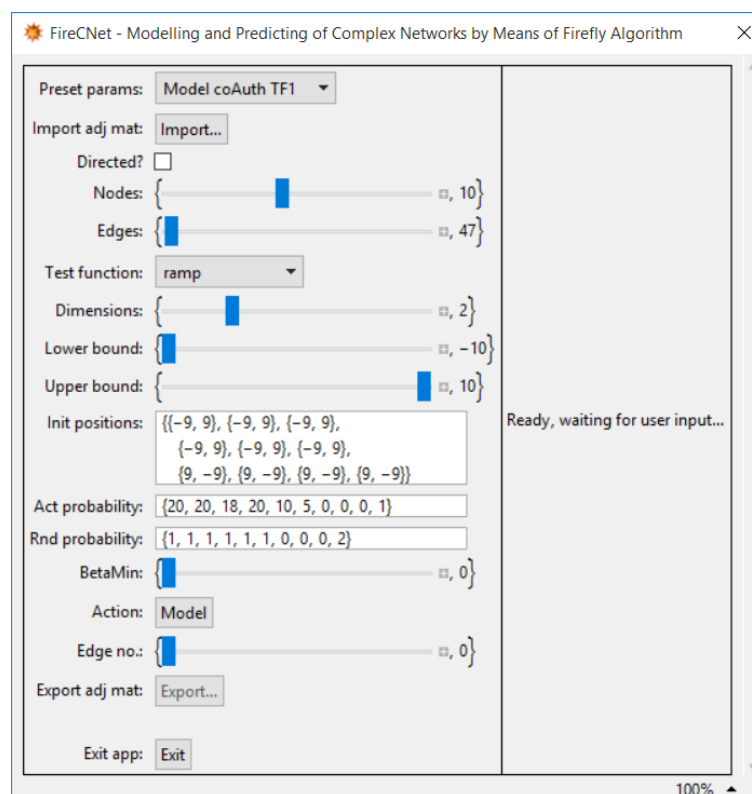
Obrázek 35 Nové pozice jedinců algoritmu Firefly po „úspěšné“ interakci aktivního jedince E s „náhodným“ jedincem C

	A	B	C	D	E
A	0	0	0	0	0
B	0	0	0	0	0
C	0	0	0	0	0
D	0	0	0	0	0
E	0	0	1	0	0

Obrázek 37 Matice sousednosti po „úspěšné“ interakci

7 IMPLEMENTACE ALGORITMU FIREFLY

Aplikace FireCNet pro modelování a předpovídání chování komplexních sítí pomocí algoritmu Firefly je implementována v prostředí Wolfram Mathematica. Jedná se o modální okenní aplikaci s grafickým uživatelským prostředím s vertikálním a horizontálním posuvníkem a nástrojem lupy, tedy možností zvětšení či zmenšení obsahu dle potřeb uživatele. Levá část okna slouží pro zadávání vstupních informací uživatelem, zatímco v pravé části jsou zobrazovány výsledky modelování či chybové hlášky. Vzhled aplikace po spuštění vyjadřuje Obrázek 38.



Obrázek 38 Vzhled aplikace FireCNet po spuštění

Kód aplikace se skládá z front-endové funkce `main[]`, back-endové funkce `firefly[]`, implementace sedmi testovacích funkcí popsanych v kapitole 5 a několika globálních konstant obsahujících zejména krajní hodnoty definičních oborů testovacích funkcí a nalezené ideální počáteční pozice jedinců populace pro modelování zkoumané citační a sociální sítě.

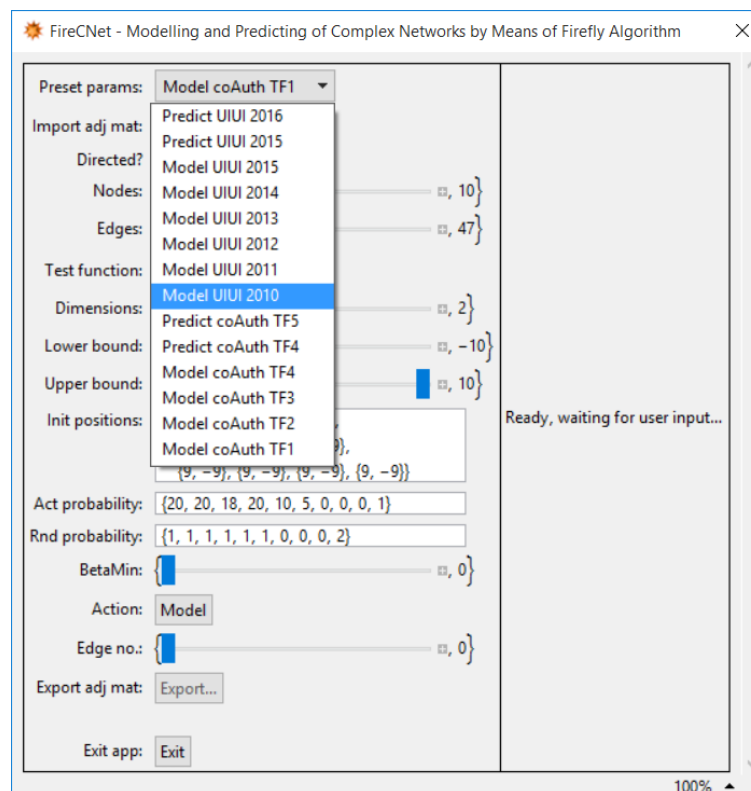
7.1 Funkce `main[]`

Dynamickou interakci s uživatelem zajišťuje funkce `main[]`, která je vstupním bodem aplikace a má za úkol vykreslit okno aplikace, hlídat platnost a vzájemnou integritu

vstupních parametrů, vypisovat chybové hlášky, vykreslovat výsledky a na žádost uživatele okno zavřít a aplikaci řádně ukončit.

7.1.1 Preset params

Položky tohoto rozevřacího seznamu (viz Obrázek 39) slouží pro přednastavení vstupních parametrů na hodnoty ideální pro danou činnost, které také sloužily jako finální techniky pro modelování a predikování komplexních sítí popsanych v kapitolách 8 a 9. Přednastavení hodnot odstraňuje z paměti aplikace případnou importovanou matici sousednosti, aby byla zaručena integrita přednastavených hodnot. Přednastavené hodnoty je samozřejmě možné dále upravovat.

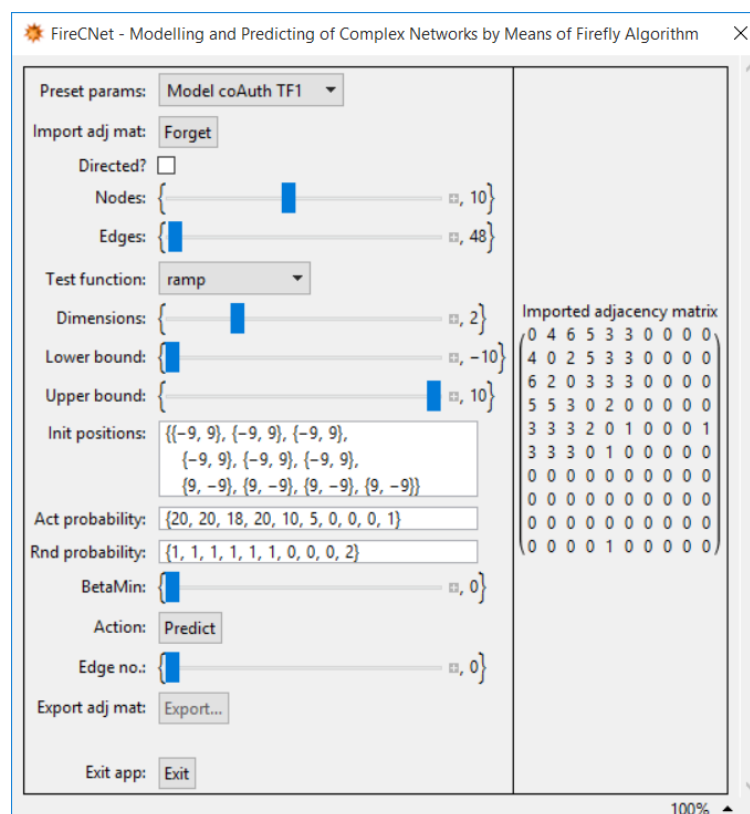


Obrázek 39 Aplikace FireCNet s rozevřeným seznamem pro přednastavení parametrů

7.1.2 Import adj mat

Tlačítko *Import* slouží pro možnost nahrání matice sousednosti se záměrem předpovídat chování komplexní sítě vyjádřené touto maticí. Po stisknutí tlačítka se objeví systémové dialogové okno s možností procházení adresářové struktury na dostupných médiích pro výběr textového souboru obsahujícího matici sousednosti. Soubor musí mít příponu *.mat*, aby byl při procházení viditelný. Oddělovačem hodnot v souboru musí být tabulátor.

Obsahem souboru musí být čtvercová matice o rozměru v rozsahu od 3 do 21, aby byl dodržen rozsah počtu uzlů v síti od 2 do 20. První řádek a první sloupec matice je ignorován, obvykle totiž obsahuje názvy jednotlivých uzlů. Prvky matice musí být nezáporná celá čísla vyjadřující váhy hran spojujících uzly na příslušném řádku a sloupci. Součet prvků matice musí být menší než 3000, aby byl dodržen maximální počet hran v síti. V případě orientované komplexní sítě odpovídá řádek výstupnímu uzlu a sloupec vstupnímu uzlu, v případě neorientované sítě musí být matice symetrická, tj. hodnoty vah musí být zaznamenány jak nad hlavní diagonálou, tak pod hlavní diagonálou. Hodnoty 0 značící absenci vazby mezi danými uzly mohou být vynechány. Pokud vybraný soubor neobsahuje platnou matici sousednosti, je v pravé části aplikace zobrazena chybová hláška *Error: Imported file is not valid*. Při úspěšném nahrání matice sousednosti je tato v aplikaci ihned zobrazena, popisek tlačítka *Import* je změněn na *Forget* a popisek tlačítka v řádku *Action* je změněn z *Model* na *Predict* (viz Obrázek 40). Funkcí tlačítka *Forget* je odstranění importované matice z paměti aplikace.



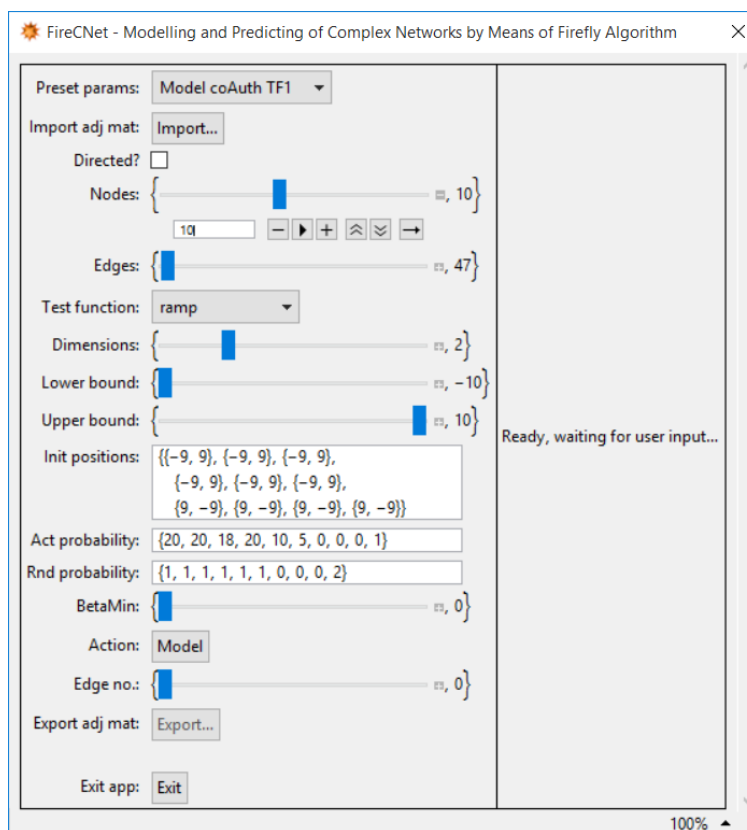
Obrázek 40 Aplikace FireCNet po úspěšném importu matice sousednosti

7.1.3 Directed

Toto zaškrtačací políčko aplikaci indikuje, zda má být komplexní síť považována za orientovanou či nikoliv. V případě, že byla importována matice, která není symetrická, zůstane toto políčko vždy zaškrtnuté.

7.1.4 Nodes

Tento manipulátor (viz Obrázek 41) slouží pro nastavení počtu uzlů v komplexní síti (resp. počtu jedinců v populaci algoritmu Firefly). Hodnotou musí být celé číslo v rozsahu od 2 do 20. Pokud byla importována matice sousednosti, bude tato hodnota vždy odpovídat rozměru importované matice.



Obrázek 41 Aplikace FireCNet s otevřeným manipulátorem pro nastavení počtu uzlů v komplexní síti (resp. počtu jedinců v populaci)

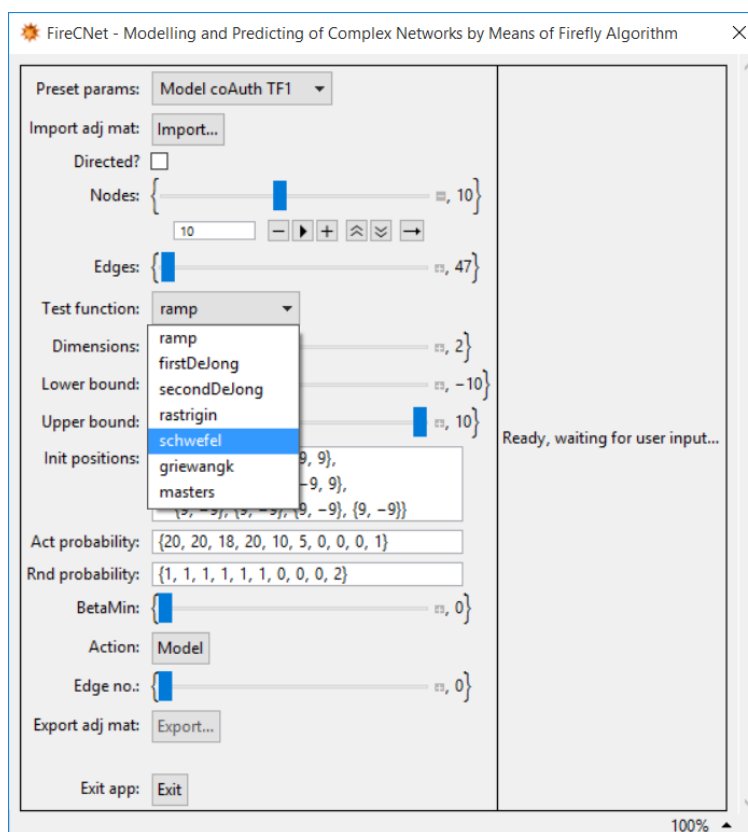
7.1.5 Edges

Manipulátor *Edges* nastavuje při dosažení jakého počtu hran v síti má být algoritmus Firefly ukončen. Hodnotou musí být celé číslo v rozsahu od 1 do 3000. V případě, že byla importována matice sousednosti, bude tato hodnota vždy alespoň o 1 větší než je součet

prvků matice v případě orientované sítě, a alespoň o 1 větší než polovina součtu prvků matice v případě neorientované sítě.

7.1.6 Test function

Položky tohoto rozevřacího seznamu (viz Obrázek 42) slouží pro výběr funkce, po jejíž hyperploše se budou jedinci populace algoritmu Firefly pomyslně pohybovat. Vybírat je možno ze sedmi funkcí, což jsou nekonečná nakloněná rovina (rampa), První de Jongova funkce (sféra), Druhá de Jongova funkce (Rosenbrockovo sedlo, Banánová funkce), Rastriginova funkce, Schwefelova funkce, Griewangkova funkce a Mastersova funkce. Každá funkce má jiný definiční obor vyjádřený pomocí krajních mezí parametrů *Lower bound* a *Upper bound*. Bližší charakteristiky všech zmíněných testovacích funkcí je možné nalézt v kapitole 5.



Obrázek 42 Aplikace FireCNet s rozevřeným seznamem pro výběr testovací funkce

7.1.7 Dimensions

Tento manipulátor slouží pro nastavení počtu dimenzí zvolené testovací funkce. Hodnotou musí být celé číslo v rozsahu od 1 do 5. V případě, že je zvolena jiná hodnota než 2, nebude pohyb jedinců po hyperploše při modelování či predikci vykreslován.

7.1.8 Lower bound

Manipulátor *Lower bound* nastavuje dolní mez definičního oboru zvolené testovací funkce. Hodnotou musí být reálné číslo v rozsahu krajních mezí definičního oboru zvolené testovací funkce, ne však větší nebo rovno hodnotě parametru *Upper bound*.

7.1.9 Upper bound

Manipulátor *Upper bound* nastavuje horní mez definičního oboru zvolené testovací funkce. Hodnotou musí být reálné číslo v rozsahu krajních mezí definičního oboru zvolené testovací funkce, ne však menší nebo rovno hodnotě parametru *Lower bound*.

7.1.10 Init positions

Toto textové pole slouží pro určení počátečních pozic jedinců populace algoritmu Firefly v hyperprostoru, a to pomocí kartézských souřadnic. Textové pole musí obsahovat matici zapsanou pomocí jazyka Wolfram, jejíž počet řádků odpovídá zvolenému počtu uzlů v komplexní síti (resp. jedinců algoritmu Firefly) a jejíž počet sloupců odpovídá zvolenému počtu dimenzí testovací funkce. Prvky matice musí být reálná čísla v rozmezí zvoleného definičního oboru testovací funkce.

7.1.11 Act probability

Textové pole *Act probability* slouží pro určení rozložení pravděpodobnosti výběru *act* jedince v rámci populace algoritmu Firefly pro každou iteraci algoritmu. Textové pole musí obsahovat vektor (jednorozměrné pole) zapsaný v jazyce Wolfram, jehož počet prvků odpovídá zvolenému počtu uzlů v komplexní síti (resp. jedinců algoritmu Firefly). Prvky vektoru musí být nezáporná reálná čísla.

7.1.12 Rnd probability

Textové pole *Rnd probability* slouží pro určení rozložení pravděpodobnosti výběru *rnd* jedince v rámci populace algoritmu Firefly pro každou iteraci algoritmu. Textové pole musí obsahovat vektor (jednorozměrné pole) zapsaný v jazyce Wolfram, jehož počet prvků

odpovídá zvolenému počtu uzlů v komplexní síti (resp. jedinců algoritmu Firefly). Prvky vektoru musí být nezáporná reálná čísla.

7.1.13 BetaMin

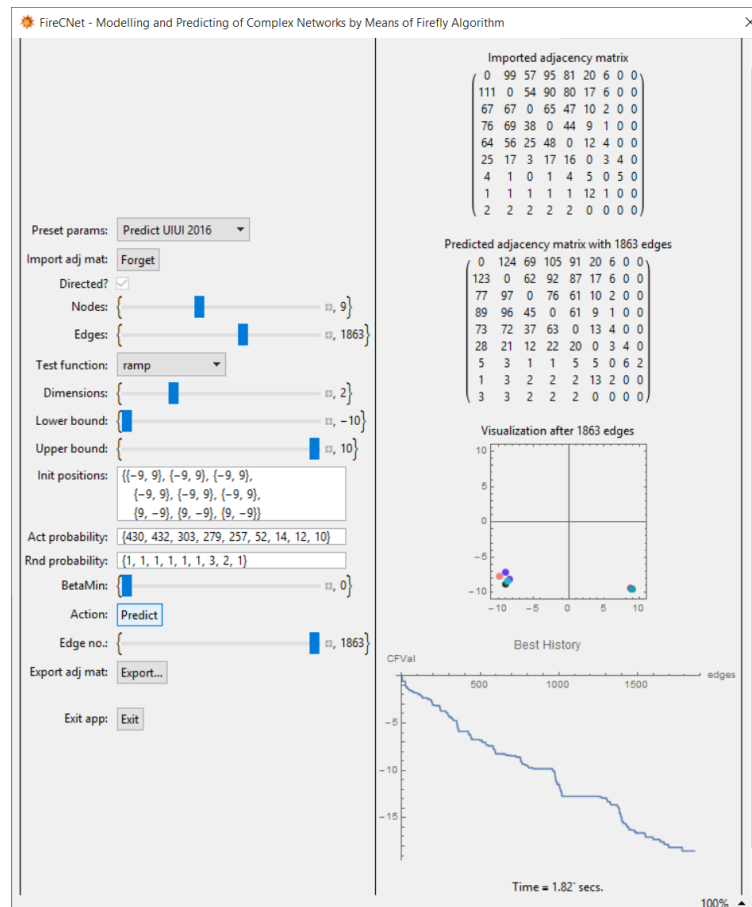
Tento manipulátor slouží pro nastavení minimální hodnoty parametru beta, tedy minimální míry atraktivity, kterou bude *act* jedinec přitahován k *rnd* jedinci, a to i v případě, že je pozice *rnd* jedince ohodnocena hůře než pozice *act* jedince. Hodnotou musí být reálné číslo v rozmezí 0 až 0,1.

7.1.14 Action

Tlačítko v řádku *Action* má ve výchozím stavu aplikace popisek *Model*, po úspěšném importu matice sousednosti pak *Predict*. V obou případech dojde po kliknutí na toto tlačítko k zavolání funkce `firefly[]` s patřičnými vstupními parametry, změření času do obdržení výstupů této funkce a následnému zobrazení získaných výsledků v pravé části aplikace (viz Obrázek 43).

Ve výsledcích je vždy zobrazena modelovaná či predikovaná matice sousednosti, doba trvání výpočtu a graf historie nejlépe ohodnoceného nalezeného místa jedinci populace, která v tomto případě, kdy se nejedná o optimalizační úlohu, plní spíše doprovodnou roli, která by měla obsluhu aplikace případnou dlouhou vodorovnou částí křivky upozornit na případné uvíznutí populace v lokálním či globálním extrému testovací funkce, což může mít za následek velkou odchylku ve vytvořeném modelu; v případě opakování této situace může být pro odvrácení tohoto jevu nutná změna vstupních parametrů.

V případě importované matice sousednosti tato zůstává ve výsledcích zobrazena a v případě dvoudimenzionální testovací funkce je mezi maticí sousednosti a grafem historie nejlepších nalezených funkčních hodnot zobrazena také vizualizace pohybu jedinců v hyperploše.



Obrázek 43 Aplikace FireCNet zobrazující výsledky predikování dynamiky importované komplexní sítě

7.1.15 Edge no

Manipulátor *Edge no* slouží pro analýzu průběhu modelování či predikování dynamiky komplexní sítě. Po vytvoření modelu sítě je možné nahlížet na stav matice sousednosti, polohu jedinců v hyperploše a stav vývoje grafu historie nejlepších nalezených funkčních hodnot, a to po každé další vytvořené hraně v modelované komplexní síti. Hodnotou musí být celé číslo v rozsahu od 0 do počtu hran vytvořeného modelu, v případě predikování však od počtu hran importované matice do počtu hran vytvořeného predikčního modelu.

7.1.16 Export adj mat

Tlačítko *Export* slouží pro uložení vytvořeného modelu komplexní sítě vyjádřeného pomocí matice sousednosti. Po stisknutí tlačítka se objeví systémové dialogové okno s možností procházení adresářové struktury na dostupných médiích pro výběr umístění textového souboru obsahujícího matici sousednosti. Při procházení jsou viditelné pouze soubory s příponou *.mat*, které je možné přepsat, anebo ve vybraném umístění vytvořit soubor nový.

Oddělovačem hodnot ve vytvořeném souboru je tabulátor. První řádek a první sloupec souboru obsahuje záhlaví s názvy jednotlivých uzlů, a to *Node 1*, *Node 2* atd., přičemž hodnota na prvním řádku a prvním sloupci je vynechána. Prvky matice jsou nezáporná celá čísla vyjadřující váhy hran spojujících uzly na příslušném řádku a sloupci. V případě orientované komplexní sítě odpovídá řádek výstupnímu uzlu a sloupec vstupnímu uzlu, v případě neorientované sítě je matice symetrická, tj. hodnoty vah jsou zaznamenány jak nad hlavní diagonálou, tak pod hlavní diagonálou. Hodnoty 0 značí absenci vazby mezi danými uzly jsou vynechány. Struktura vytvořeného souboru odpovídá požadavkům zásuvného modulu `aMatReader` aplikace Cytoscape pro analýzu a výpočet centralit příslušné komplexní sítě.

7.1.17 Exit app

Tlačítko *Exit* slouží pro zavření okna aplikace a její řádné ukončení. Stejně tak bezpečně může být aplikace ukončena kliknutím na křížek v pravém horním rohu jejího okna nebo stisknutím klávesy *Escape*.

7.2 Funkce `firefly[]`

Funkce `firefly[]` představuje implementaci světluščího hejnového algoritmu a je volána funkcí `main[]` v případě, že uživatel klikne na tlačítko v řádku *Action*. Většina vstupních parametrů této funkce již byla popsána v kapitole 7.1. Výjimku tvoří parametry *alfa* a *gama*.

7.2.1 Vstupní parametr *alfa*

Hodnota *alfa* vyjadřuje míru náhody při pohybu světlušky, závisí na zvolené velikosti definičního oboru testovací funkce a je dána vztahem:

$$alfa = 0,005 * (UpperBound - LowerBound)$$

Každá světluška se tedy může v rámci jedné iterace vlivem náhody posunout až o 0,5 % velikosti hyperplochy, a to v každé dimenzi.

7.2.2 Vstupní parametr gama

Hodnota gama představuje míru absorpce světla prostředím. Čím vyšší hodnota gama, tím méně se budou světlušky přitahovat. Parametr gama také závisí na zvolené velikosti definičního oboru testovací funkce a je dán vztahem:

$$gama = \frac{1}{\sqrt{UpperBound - LowerBound}}$$

7.2.3 Začátek těla funkce

Na začátku těla funkce jsou ohodnoceny (y) počáteční pozice jedinců v hyperprostoru (x) a následně vytvořeny první záznamy (pro stav sítě při počtu hran 0 v matici sousedenosti) v historii pozice a ohodnocení nejlepšího jedince (*bestHist*), v historii pozic a ohodnocení všech jedinců populace (*interHist*) a v historii vývoje matice sousedenosti (*adjMatHist*).

7.2.4 Začátek smyčky While

Následuje hlavní smyčka *While*, ve které se nachází výpočetní jádro algoritmu Firefly a která běží, dokud je číslo právě vytvářené hrany v matici sousedenosti (*in*) menší nebo rovno požadovanému počtu vytvořených hran (*inters*).

7.2.5 Výběr act a rnd jedince

V těle smyčky *While* se nejprve zvolí index aktivního jedince (*act*) a index tzv. „náhodného“ jedince (*rnd*) v populaci, a to dle zvoleného pravděpodobnostního rozložení *actProb* a *rndProb*. Jedinec *act* bude následně přitahován k jedinci *rnd*. Index jedince *rnd* nesmí být stejný jako index jedince *act*, jinými slovy aktivní jedinec nesmí být zároveň náhodný.

7.2.6 Aktualizace relY

Poté dojde k aktualizaci hodnoty *relY* představující relativní ohodnocení jedinců v populaci, tedy ohodnocení přemapovaná do rozmezí od 0 do 1, přičemž hodnota 0 představuje jedince s nejhorším ohodnocením (v případě hledání minima testovací funkce se jedná o jedince s nejvyšší funkční hodnotou) a hodnota 1 představuje jedince s nejlepším ohodnocením (v případě hledání minima testovací funkce se jedná o jedince s nejnižší funkční hodnotou).

7.2.7 Výpočet beta

Následuje výpočet hodnoty beta, která je dána odečtením hodnoty *relY* jedince *act* od hodnoty *relY* jedince *rnd*. V případě, že je vypočtená hodnota beta menší nebo rovna

parametru *betaMin*, je hodnota *beta* upravena na hodnotu *betaMin* a proměnná *rndHelped* nastavena na logickou hodnotu *False*, v opačném případě je hodnota *beta* ponechána a proměnná *rndHelped* nastavena na logickou hodnotu *True*, což znamená, že tato interakce může vytvářet další hranu v matici sousednosti.

7.2.8 Výpočet *r*

Dále je vypočtena hodnota *r* představující Euklidovskou vzdálenost v hyperprostoru mezi jedincem *act* a jedincem *rnd*, a to podle vztahu:

$$r = \sqrt{\sum_{k=1}^{\dim} (x_{act,k} - x_{rnd,k})^2}$$

7.2.9 Určení *eps*

Následně je určena hodnota *eps*, a to jako vektor náhodných reálných čísel v rozmezí od -1 do 1 o rozměru *dims*.

7.2.10 Výpočet *x_n*

Poté dojde k samotnému výpočtu nové pozice jedince *act* (*x_n*) podle vzorce:

$$x_n = x_{act} + beta * e^{-gamma * r^2} * (x_{rnd} - x_{act}) + alfa * eps$$

Následuje ošetření nové pozice aktivního jedince proti přesáhnutí hranic definičního oboru testovací funkce, a to tím způsobem, že pokud je v nějaké dimenzi hranice přesáhnutá, je souřadnice v této dimenzi upravena na hodnotu hranice.

7.2.11 Konec těla smyčky *While*

Dále je nová pozice aktivního jedince ohodnocena (*yn*). Pokud je ohodnocení nové pozice lepší než ohodnocení původní pozice, je vypočtena hodnota *rn* představující Euklidovskou vzdálenost v hyperprostoru mezi starou a novou pozicí aktivního jedince, a to podle vztahu:

$$r_n = \sqrt{\sum_{k=1}^{\dim} (x_{act,k} - x_{n,k})^2}$$

Nová pozice a její ohodnocení je zaneseno do populace a v případě, že hodnota proměnné *rndHelped* byla nastavena na logickou hodnotu *True*, dojde v matici sousednosti k navýšení hodnoty na řádku *act* a sloupci *rnd* o desetinásobek hodnoty *rn*, maximálně však o 1.

V případě neorientované komplexní sítě pak také ke stejnoměrnému navýšení hodnoty na řádku *rnd* a sloupci *act*.

Pokud se po tomto navýšení na celá čísla zaokrouhlený počet hran v matici sousednosti rovná hodnotě *in*, dojde k vytvoření dalšího záznamu v proměnné *bestHist*, jejíž hodnoty musí být neklesající, což znamená, že v případě horšího ohodnocení nejlepšího jedince v populaci oproti poslední hodnotě je na novou pozici zkopírována tato minulá hodnota. Rovněž dojde k vytvoření dalšího záznamu v proměnné *interHist*. Pokud se jedná o vytváření predikčního modelu a hodnota *in* se rovná počtu hran v importované matici, pak je v tento okamžik dosud modelovaná matice sousednosti vyměněna za importovanou matici sousednosti. Do proměnné *adjMatHist* je také zanesen nový záznam, a to matice sousednosti s prvky zaokrouhlenými na celá čísla. Nakonec je hodnota *in* inkrementována o 1.

V případě, že je ohodnocení nové pozice horší nebo stejné oproti ohodnocení původní pozice, dojde pouze k zanesení nové pozice a jejího ohodnocení do populace.

7.2.12 Návrátové hodnoty funkce

Po vystoupení z hlavní smyčky *While* vrátí funkce *Firefly* hodnoty proměnných *bestHist*, *interHist* a *adjMatHist*.

8 MODELOVÁNÍ KOMPLEXNÍCH SÍTÍ POMOCÍ ALGORITMU FIREFLY

Kapitola 8 popisuje vývoj a aplikaci modelovacích metod pro vytváření komplexních sítí, jejichž struktura co možná nejvíce odpovídá vzorovým komplexním sítím. Techniky modelování byly testovány na reálných komplexních sítích, konkrétně citační síti UIUI a sociální síti coAuth. Hrany v modelované síti byly vytvářeny na základě chování hejnového algoritmu Firefly, jehož nastavitelné parametry ovlivňovaly podobu výsledného modelu.

8.1 Citační síť Ústavu informatiky a umělé inteligence na FAI UTB

Vývoj citační sítě UIUI v letech 2010 až 2015 získaný z internetové databáze na adrese <https://www.scopus.com/> zachycují postupně Tabulka 6, Tabulka 7, Tabulka 8, Tabulka 9, Tabulka 10 a Tabulka 11. Uzly sítě představují zaměstnance UIUI, přičemž někteří z nich v průběhu zkoumaného období na UIUI teprve přišli (tj. začali navazovat spojení s ostatními uzly sítě), jiní změnili pracoviště (nicméně stále navazovali další spojení) nebo měli práci z jakéhokoliv důvodu dlouhodoběji přerušenu. Váha hran vyjadřuje počet článků daného autora, které citují alespoň jeden článek příslušného jiného autora. Přestože jsou tyto informace veřejně dostupné, byla jména příslušných zaměstnanců UIUI částečně skryta pod jejich iniciály. Pro začátek roku 2010, odkdy byl výskyt citací sledován, byly simulovány nulové počáteční hodnoty u všech autorů. Hodnoty autocitací nebyly v tomto případě zaznamenávány, proto je hlavní diagonála vždy prázdná. Nulové hodnoty jsou v tabulkách pro přehlednost vynechány. Řádek příslušející dané hodnotě odpovídá autorovi, který citoval, zatímco sloupec odpovídá autorovi, který byl citován. Např. hodnota „3“ v Tabulce 6 tedy vyjadřuje, že autor DD citoval autora RS do konce roku 2010 celkem třikrát. Napravo od tabulek jsou součty hodnot v jednotlivých řádcích, které udávají, kolikrát daný autor citoval, tedy kolikrát by měla odpovídající světluška být při vytváření nové hrany v komplexní síti v roli aktivního jedince, zatímco pod tabulkami jsou součty hodnot v jednotlivých sloupcích, které udávají, kolikrát byl daný autor citován, tedy kolikrát by měla odpovídající světluška být při vytváření nové hrany v komplexní síti v roli „náhodného“ jedince.

Tabulka 6 Stav citační sítě UIUI na konci roku 2010

2010	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		4		2	5					11
RS	8			1	4					13
MP										0
DD	5	3								8
ZO	8	4		1						13
RJ	1									1
PV										0
BC										0
JJ										0
<i>rnd</i>	22	11	0	4	9	0	0	0	0	

Tabulka 7 Stav citační sítě UIUI na konci roku 2011

2011	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		12		9	13					34
RS	17			8	12					37
MP										0
DD	10	7			4					21
ZO	15	12		7						34
RJ	12	7		6	8		1	1		35
PV	2				2					4
BC						2				2
JJ										0
<i>rnd</i>	56	38	0	30	39	2	1	1	0	

Tabulka 8 Stav citační sítě UIUI na konci roku 2012

2012	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		26		24	20	2	2			74
RS	32			24	19	2	2			79
MP	6	6		6	1		1			20
DD	22	18			9		1			50
ZO	27	23		18		2	2			72
RJ	15	10		10	10		1	2		48
PV	3	1		1	3	1		1		10
BC						4				4
JJ										0
<i>rnd</i>	105	84	0	83	62	11	9	3	0	

Tabulka 9 Stav citační sítě UIUI na konci roku 2013

2013	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		57	18	55	43	7	5			185
RS	69		17	54	44	6	5			195
MP	30	30		30	17	1	1			109
DD	45	39	11		23	1	1			120
ZO	44	37	7	31		4	3			126
RJ	20	12		12	13		1	2		60
PV	4	1		1	4	5		5		20
BC						10	1			11
JJ										0
<i>rnd</i>	212	176	53	183	144	34	17	7	0	

Tabulka 10 Stav citační sítě UIUI na konci roku 2014

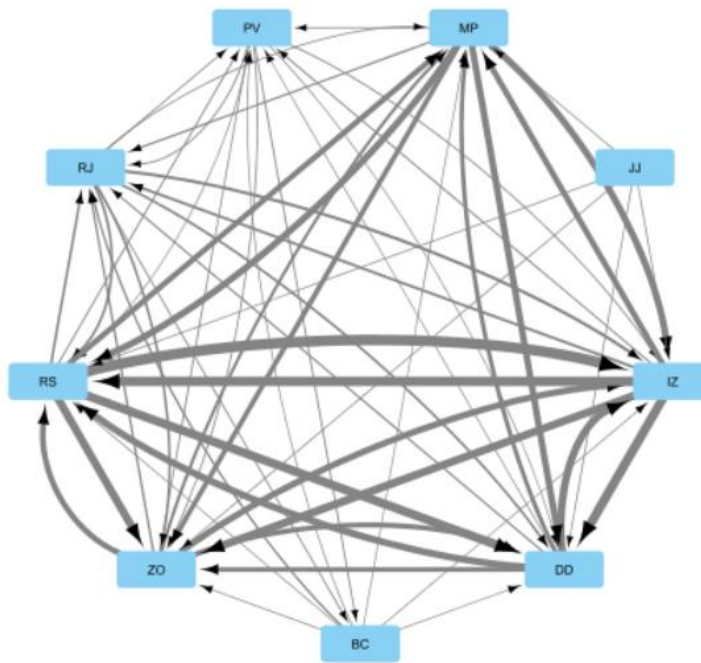
2014	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		87	45	84	72	18	6			312
RS	99		42	80	70	15	6			312
MP	55	55		55	39	9	2			215
DD	64	58	28		38	7	1			196
ZO	60	50	20	44		10	4			188
RJ	22	14	2	14	15		1	2		70
PV	4	1		1	4	5		5		20
BC	1	1	1	1	1	12	1			18
JJ										0
<i>rnd</i>	305	266	138	279	239	76	21	7	0	

Tabulka 11 Stav citační sítě UIUI na konci roku 2015

2015	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		99	57	95	81	20	6			358
RS	111		54	90	80	17	6			358
MP	67	67		65	47	10	2			258
DD	76	69	38		44	9	1			237
ZO	64	56	25	48		12	4			209
RJ	25	17	3	17	16		3	4		85
PV	4	1		1	4	5		5		20
BC	1	1	1	1	1	12	1			18
JJ	2	2	2	2	2					10
<i>rnd</i>	350	312	180	319	275	85	23	9	0	

V Obrázku 44 lze vidět grafické znázornění stavu citační sítě UIUI na konci roku 2015 vytvořené aplikací Cytoscape. Tloušťka čar reprezentujících hrany sítě odpovídá váze těchto hran. Na první pohled je tedy zřejmé, že dominantními uzly sítě jsou MP, IZ, DD, ZO a RS,

zatímco uzly JJ, BC, RJ a PV jsou významně méně. Šipky udávají orientaci hran, a to tím způsobem, že uzel, odkud šipka vychází, citoval uzel, kam šipka směřuje. Porovnáním s Tabulkou 11 lze tedy například vidět, že z uzlu JJ několik šipek vychází, ale žádné do něj nesměřují, tzn., že tento uzel citoval ostatní, ale sám citován nebyl.



Obrázek 44 Vizualizace citační sítě UIUI na konci roku 2015

Tabulka 12 zobrazuje průměrné centrality citační sítě UIUI v jednotlivých letech vypočítané pomocí CytoNCA. Hodnota SC vycházela vždy velmi vysoká nebo byla už přímo programem zaokrouhlována na nekonečno, proto při následném modelování nebude tato hodnota sledována. Rovněž tak hodnota DC nebude sledována, neboť v rámci daného roku pro stejný počet hran v síti musí vyjít vždy stejně.

Tabulka 12 Vypočtené průměrné centrality citační sítě UIUI pro jednotlivé roky

rok	hran (citací)	BC	CC	DC	EC	LAC	NC	SC	IC
2010	46	0,74	0,17	10	0,215	8	2,6	Inf	1,3
2011	167	1,56	0,33	37	0,253	28	4,4	Inf	2,2
2012	357	2,00	0,66	79	0,251	63	10,9	Inf	3,7
2013	826	1,11	0,72	184	0,254	150	12,9	Inf	4,6
2014	1331	0,00	0,74	296	0,261	271	19,2	Inf	5,3
2015	1553	0,74	6,39	345	0,262	364	25,3	Inf	12,6

Při modelování citační sítě UIUI byly pro finální modely všech časových období použity následující hodnoty parametrů:

Directed = True

Population_size = 9

Test_function = ramp

Dimensions = 2

Lower_bound = -10

Upper_bound = 10

Init_positions = {{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{9,-9},{9,-9},{9,-9}}

BetaMin = 0

Hodnoty zbylých parametrů se pro jednotlivá časová období lišily a jsou uvedeny v příslušných kapitolách.

8.1.1 Modelování citační sítě UIUI v roce 2015

Tabulka 13 obsahuje údaje o kvalitě pokusů o namodelování stavu citační sítě UIUI na konci roku 2015. První sloupec popisuje techniku modelování, dalších šest sloupců představuje průměrné centrality dané sítě a poslední sloupec vyjadřuje v bodech, jak moc se namodelovaná síť podobá té skutečné, přičemž čím nižší hodnota ve sloupci *pts*, tím lepší výsledek. Hodnota *pts* se vypočítá jako:

$$pts = \left[\left(\frac{BC_r - BC_m}{BC_r} \right)^2 + \left(\frac{CC_r - CC_m}{CC_r} \right)^2 + \dots + \left(\frac{IC_r - IC_m}{IC_r} \right)^2 \right] * 1000$$

kde *r* v dolním indexu znamená hodnotu skutečné sítě, zatímco *m* znamená hodnotu namodelované sítě. Tento způsob výpočtu zajišťuje, že všechny centrality mají při měření kvality modelu stejnou váhu i přesto, že se jejich hodnoty pohybují řádově v jiných mezích. Kvadráty jednotlivých členů zajišťují jednak stejná znaménka pro jejich sčítání a jednak, že malé odchylky od žádaných hodnot budou zvyšovat hodnotu *pts* jen nepatrně, zatímco velké odchylky budou zvyšovat hodnotu *pts* mnohem více oproti lineárním členům. Model shodující se se skutečnou sítí zcela ve všech sledovaných hodnotách by tedy měl *pts* = 0.

Tabulka 13 Vývoj modelování citační sítě UIUI v roce 2015

<i>modelling 2015</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,74	6,39	0,262	364	25,3	12,6	0
uniform_1	0,07	21,20	0,333	302	14,5	18,1	6 660
uniform_2	0,00	21,63	0,333	302	14,3	16,9	7 101
uniform_3	0,00	21,04	0,333	302	14,3	18,7	6 783
expo_1	0,89	24,86	0,239	429	22,2	7,9	8 599
linear_1	0,00	0,78	0,277	286	10,1	24,8	3 106
linear_2	0,00	0,78	0,285	244	9,3	6,8	2 497
linear_3	0,00	0,78	0,284	323	11,7	3,7	2 577
exact_1	0,50	15,22	0,268	364	23,2	10,1	2 069
exact_2	0,67	14,25	0,278	369	16,5	9,4	1 716
exact_3	0,89	15,61	0,274	408	20,2	9,8	2 236
exact_position_1	0,44	10,77	0,266	353	21,1	10,3	693
exact_position_2	1,11	11,05	0,263	404	21,5	9,4	883
exact_position_3	0,44	11,33	0,275	364	19,4	10,9	836
manual_position_1	0,44	6,74	0,268	288	22,1	13,6	229
manual_position_2	0,44	6,88	0,264	318	20,6	11,2	230
manual_position_3	0,44	7,79	0,264	301	19,8	7,3	464

V prvním řádku Tabulky 13 se nacházejí hodnoty průměrných centralit komplexní sítě, která má být namodelována. Cílem tedy bylo, aby se hodnoty modelu blížily těmto hodnotám co možná nejvíce.

První model této sítě byl vytvořen prostou metodou *uniform*, kdy byly počáteční pozice jedinců v rámci hyperplochy náhodné a pravděpodobnost výběru *act* i *rnd* jedince rozložena rovnoměrně. Hodnoty *pts* se pohybovaly kolem 7000 a měly být jakýmsi odrazovým můstkem pro možnost porovnávání dalších technik.

Druhý pokus (*expo*) spočíval ve využití řádkových a sloupcových součtů v Tabulce 11, přičemž byly tyto hodnoty seřazeny sestupně a následně aproximovány exponenciálou (rozložení stupňů uzlů komplexní sítě by se totiž mělo řídit mocninným zákonem, viz Obrázek 15) a touto exponenciálou se následně řídila pravděpodobnost výběru *act* i *rnd* jedinců. Výsledek dopadl ovšem ještě hůře než prvotní nástřel, neboť aproximace exponenciálou byla velmi nepřesná a rozložení stupňů uzlů by mnohem lépe vystihovala lineární funkce.

Při pravděpodobnosti výběru *act* a *rnd* jedinců odpovídající aproximované lineární funkci (*linear*) již bylo dosahováno výrazně lepších hodnot *pts*, a to okolo 2500. Dalším posunem bylo upravení pravděpodobnosti výběru *act* a *rnd* jedinců tak, že přesně odpovídala

žádanému konečnému stupni příslušných uzlů v síti (*exact*). Hodnota *pts* se zde pohybovala okolo 2000 a jediným výrazným problémem zůstávala velká odchylka hodnoty CC. Model zachycuje Tabulka 14.

Tabulka 14 Model *exact_1* citační sítě UIUI na konci roku 2015

2015	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		86	63	76	38	5				268
RS	111		67	50	37	10				275
MP	91	89		73	34	11	2			300
DD	88	77	47		31	6				249
ZO	68	66	48	53		6				241
RJ	36	30	36	26	17					145
PV	9	9	4	7	5	2		1		37
BC	3	5	4	5	2	1				20
JJ	1	4	2	2	8	1				18
<i>rnd</i>	407	366	271	292	172	42	2	1	0	

Pro upravení hodnoty CC bylo nutné implementovat vylepšení v podobě zásahu do programového kódu, jehož výsledkem bylo, že se při posunu jedince *act* na lépe ohodnocené místo v hyperploše za pomoci jedince *rnd* neinkrementovala příslušná hodnota v matici sousednosti o 1, ale o hodnotu odpovídající vzdálenosti, o kterou se jedinec *act* posunul (maximálně však o 1). Hodnoty v matici sousednosti jsou pak při zobrazování v aplikaci nebo exportu zaokrouhlovány na celá čísla dle matematických pravidel.

Po této úpravě pak mělo smysl rozdělit populaci na dvě skupiny o 6 a o 3 jedincích, mezi kterými byl výrazný rozdíl v žádaném konečném stupni odpovídajících uzlů, a tyto skupiny umístit na počátku deterministicky do dvou protilehlých rohů hyperplochy jednak proto, aby obě skupiny měly kvalitativně přibližně stejné počáteční podmínky a mohly tak vytvářet nové hrany v matici sousednosti i mezi sebou, a jednak proto, aby byly obě skupiny dost daleko od sebe a nespojily se postupem času do jedné. Výsledkem pro tuto techniku *exact_position* (viz Tabulka 15) pak bylo ohodnocení *pts* okolo 800, a to zejména díky snížení hodnoty CC.

Hodnoty parametrů modelu *exact_position* citační sítě UIUI pro modelování roku 2015:

Interactions = 1553

Act_probability = {358,358,258,237,209,85,20,18,10}

Rnd_probability = {350,312,180,319,275,85,23,9,0}

Tabulka 15 Model *exact_position_2* citační síť UIUI na konci roku 2015

2015	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		113	40	94	78	10		1		336
RS	75		38	80	65	10	1			269
MP	78	78		58	64	6				284
DD	61	80	22		48	8				219
ZO	78	56	33	58		5				230
RJ	30	31	15	32	30					138
PV	7	5	5	7	3	2				29
BC	11	4	3	11	4	2				35
JJ		5	1	5	2					13
<i>rnd</i>	340	372	157	345	294	43	1	1	0	

Posledním zásahem byla manuální úprava pravděpodobností výběrů *act* a *rnd* jedinců, přičemž byla za účelem snížení hodnot ve spodních čtyřech řádcích matice sousednosti zmenšena pravděpodobnost výběru odpovídajících *act* jedinců na polovinu. Pravděpodobnost výběru *rnd* jedinců byla ve větší skupině pro zjednodušení sjednocena na stejnou hodnotu, neboť relativně malé rozdíly v konečném žádaném stupni odpovídajících uzlů nehrály pro kvalitu modelování významnou roli, zatímco pravděpodobnost výběru *rnd* jedince PV byla trojnásobkem pravděpodobnosti výběru jedince z větší skupiny, jedince BC dvojnásobkem a jedince JJ ponechána na nule, aby došlo ke zvýšení hodnot v sedmém a osmém sloupci matice sousednosti. Hodnoty *pts* těchto modelů se pohybovaly okolo 300. Výsledek lze vidět v Tabulce 16.

Hodnoty parametrů modelu *manual_position* citační síť UIUI pro modelování roku 2015:

Interactions = 1553

Act _ probability = {358,358,258,237,209,43,10,9,5}

Rnd _ probability = {1,1,1,1,1,1,3,2,0}

Tabulka 16 Model *manual_position_1* citační síť UIUI na konci roku 2015

2015	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		99	84	71	62	18	3			337
RS	77		48	64	41	23	2			255
MP	62	81		50	51	24	1	1		270
DD	65	68	54		46	15	1	2		251
ZO	64	71	51	54		18	1			259
RJ	21	28	28	21	17		1			116

2015	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
PV	2	4	5	4	4	4		5		28
BC	4	1	4	5	6	3	1			24
JJ	2			3	2	3	2	1		13
<i>rnd</i>	297	352	274	272	229	108	12	9	0	

8.1.2 Modelování citační sítě UIUI v roce 2014

Pro modelování citační sítě v roce 2014 (viz Tabulka 17 a Tabulka 18) byla úspěšně aplikována finální technika použitá pro rok 2015, tedy populace je rozdělena do dvou skupin o 6 a o 3 jedincích, umístěných v protilehlých rozích hyperplochy. Pravděpodobnost výběru *act* jedinců vychází z konečného žádaného stupně odpovídajících uzlů, přičemž u posledních čtyř jedinců s nejnižším stupněm je tato pravděpodobnost snížena na polovinu, zatímco u *rnd* jedinců je pravděpodobnost výběru prvních šesti jedinců s nejvyšším stupněm stejná, sedmý jedinec má pravděpodobnost trojnásobnou, osmý dvojnásobnou a devátý nulovou.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro modelování roku 2014:

Interactions = 1331

Act_probability = {312,312,215,196,188,35,10,9,0}

Rnd_probability = {1,1,1,1,1,1,3,2,0}

Tabulka 17 Modelování citační sítě UIUI v roce 2014

<i>modelling 2014</i>	BC	CC	EC	LAC	NC	IC	<i>pts</i>
real	0,00	0,74	0,261	271	19,2	5,3	0
manual_position_1	0,00	0,74	0,262	264	14,1	4,8	81
manual_position_2	0,00	0,75	0,270	280	17,7	5,1	10
manual_position_3	0,22	0,75	0,269	271	17,2	4,9	66

Tabulka 18 Model *manual_position_2* citační sítě UIUI na konci roku 2014

2014	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		74	57	48	65	18	2			264
RS	70		44	49	55	15	1	1		235
MP	66	73		66	69	14				288
DD	56	59	25		55	13				208
ZO	57	53	29	41		19				199
RJ	15	21	12	12	16		1			77
PV	4	4	5	2	8	5		4		32
BC	3	5	3	7	4	5	1			28
JJ										0
<i>rnd</i>	271	289	175	225	272	89	5	5	0	

8.1.3 Modelování citační sítě UIUI v roce 2013

Pro modelování citační sítě v roce 2013 (viz Tabulka 19 a Tabulka 20) platí stejná pravidla jako pro modelování citační sítě v roce 2014 a 2015.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro modelování roku 2013:

Interactions = 826

Act_probability = {185,195,109,120,126,30,10,6,0}

Rnd_probability = {1,1,1,1,1,1,3,2,0}

Tabulka 19 Modelování citační sítě UIUI v roce 2013

<i>modelling 2013</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	1,11	0,72	0,254	150	12,9	4,6	0
<i>manual_position_1</i>	0,44	0,72	0,263	144	12,1	8,1	959
<i>manual_position_2</i>	0,22	0,71	0,278	133	9,6	4,2	732
<i>manual_position_3</i>	0,44	0,73	0,270	137	11,0	4,2	397

Tabulka 20 Model *manual_position_3* citační sítě UIUI na konci roku 2013

2013	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		61	47	38	27	30				203
RS	35		45	37	24	13	1			155
MP	25	33		19	8	13				98
DD	30	31	32		17	20	1			131
ZO	33	39	38	19		19		1		149
RJ	9	16	9	6	7					47
PV	4	4	1	4	7	4		4		28
BC	1			2	2	2	8			15
JJ										0
<i>rnd</i>	137	184	172	125	92	101	10	5	0	

8.1.4 Modelování citační sítě UIUI v roce 2012

Pro modelování citační sítě v roce 2012 (viz Tabulka 21 a Tabulka 22) byla použita stejná technika jako při modelování sítě z roku 2013, 2014 a 2015, ovšem s tím rozdílem, že pravděpodobnost výběru *rnd* jedince byla stejná pouze pro první čtyři jedince dle stupně, pátý v pořadí měl kvůli neobvykle (asi pětkrát) většímu výstupnímu stupni oproti vstupnímu pravděpodobnost výběru *rnd* jedincem pouze desetinovou, šestý jedinec měl

pravděpodobnost trojnásobnou, sedmý dvojnásobnou, osmý a devátý nulovou, neboť v síti existují dva uzly s nulovým vstupním stupněm.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro modelování roku 2012:

Interactions = 357

Act_probability = {74,79,10,50,72,48,5,2,0}

Rnd_probability = {1,1,0,1,1,0.1,3,2,0}

Tabulka 21 Modelování citační sítě UIUI v roce 2012

<i>modelling 2012</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	2,00	0,66	0,251	63	10,9	3,7	0
manual_position_1	0,89	0,69	0,257	83	10,7	3,8	415
manual_position_2	1,11	0,67	0,261	77	9,5	3,3	277
manual_position_3	1,04	0,68	0,263	64	9,8	3,8	246

Tabulka 22 Model *manual_position_3* citační sítě UIUI na konci roku 2012

2012	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		20		20	36	1	3			80
RS	23			25	23	1	2			74
MP	10	2		6	6	2	1			27
DD	5	11			25		3	1		45
ZO	10	22		17			2			51
RJ	15	11		15	17		1	2		61
PV	5	3		3	1			3		15
BC	1						3			4
JJ										0
<i>rnd</i>	69	69	0	86	108	4	15	6	0	

8.1.5 Modelování citační sítě UIUI v roce 2011

Pro modelování citační sítě v roce 2011 (viz Tabulka 23 a Tabulka 24) platí stejná pravidla jako pro modelování citační sítě v roce 2012.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro modelování roku 2011:

Interactions = 167

Act_probability = {34,37,0,21,34,35,2,1,0}

Rnd_probability = {1,1,0,1,1,0.1,3,2,0}

Tabulka 23 Modelování citační sítě UIUI v roce 2011

<i>modelling 2011</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	1,56	0,33	0,253	28	4,4	2,2	0
manual_position_1	1,11	0,33	0,256	24	5,1	2,1	131
manual_position_2	0,89	0,31	0,253	34	6,8	2,4	559
manual_position_3	0,89	0,33	0,241	25	6,9	2,4	540

Tabulka 24 Model *manual_position_1* citační sítě UIUI na konci roku 2011

2011	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		5		5	15		1			26
RS	18			12	4	2	1			37
MP										0
DD	12	8			9	1				30
ZO	7	8		5		2				22
RJ	11	6		11	13			1		42
PV	2	1						2		5
BC	3	1					1			5
JJ										0
<i>rnd</i>	53	29	0	33	41	5	3	3	0	

8.1.6 Modelování citační sítě UIUI v roce 2010

Pro modelování citační sítě v roce 2010 (viz Tabulka 25 a Tabulka 26) platí stejná pravidla jako pro modelování citační sítě v roce 2011 a 2012, ovšem s tím rozdílem, že pravděpodobnost výběru *rnd* jedince byla pro posledních pět jedinců dle konečného žádaného stupně odpovídajících uzlů v síti nulová, neboť v síti existuje pět uzlů s nulovým vstupním stupněm.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro modelování roku 2010:

Interactions = 46

Act_probability = {11,13,0,8,13,1,0,0,0}

Rnd_probability = {1,1,0,1,1,0,0,0,0}

Tabulka 25 Modelování citační sítě UIUI v roce 2010

<i>modelling 2010</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,74	0,17	0,215	8	2,6	1,3	0
manual_position_1	0,52	0,17	0,233	6	2,1	1,1	266
manual_position_2	0,44	0,17	0,232	7	2,6	2,1	592
manual_position_3	0,44	0,17	0,229	6	1,8	1,3	351

Tabulka 26 Model *manual_position_1* citační sítě UIUI na konci roku 2010

2010	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		1		3	3					7
RS	7			1	2					10
MP										0
DD	8	2			4					14
ZO	6	3		3						12
RJ		2			1					3
PV										0
BC										0
JJ										0
<i>rnd</i>	21	8	0	7	10	0	0	0	0	

8.2 Sociální síť spoluautorství nad fyzikálními vědeckými články

Ze staženého textového souboru *out.ca-cit-HepPh* byly vybrány pouze řádky týkající se prvních deseti autorů. Tyto řádky byly následně seřazeny dle časového razítka a rozděleny do čtyř časových rámců. První rámeček zahrnoval časy 825 – 875, druhý 876 – 925, třetí 926 – 975 a čtvrtý 976 – 1025. Vývoj sociální sítě zachycují postupně Tabulka 27, Tabulka 28, Tabulka 29 a Tabulka 30. Jména dotyčných autorů se nepodařilo dohledat, proto jsou označeni písmeny A až J. Na začátku časového rámce 1 jsou předpokládány nulové hodnoty ve všech buňkách matice sousednosti. Sociální síť je neorientovaná, proto jsou matice sousednosti vždy symetrické podle hlavní diagonály. Váhy hran představují počet fyzikálních vědeckých článků, na kterých se podíleli příslušní dva autoři zároveň. Napravo od tabulek jsou součty hodnot v jednotlivých řádcích, které udávají, kolikrát daný autor figuroval ve spoluautorství, tedy kolikrát by měla odpovídající světluška být při vytváření nové hrany v komplexní síti v roli aktivního jedince, zatímco pod tabulkami jsou součty hodnot v jednotlivých sloupcích, které stejně tak udávají, kolikrát daný autor figuroval ve spoluautorství, tedy kolikrát by měla odpovídající světluška být při vytváření nové hrany v komplexní síti v roli náhodného jedince.

Tabulka 27 Stav sociální sítě coAuth na konci časového rámce 1

TF1	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		8	4	5	2	1					20
C	8		4	5	2	1					20
F	4	4		7	2	1					18
E	5	5	7		2	1					20
G	2	2	2	2		1				1	10
I	1	1	1	1	1						5
H											0
A											0
J											0
D					1						1
<i>rnd</i>	20	20	18	20	10	5	0	0	0	1	

Tabulka 28 Stav sociální sítě coAuth na konci časového rámce 2

TF2	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		13	10	8	5	4					40
C	13		8	8	4	3					36
F	10	8		10	5	4					37
E	8	8	10		3	2					31
G	5	4	5	3		4				1	22
I	4	3	4	2	4						17
H								2	2		4
A							2		1		3
J							2	1			3
D					1						1
<i>rnd</i>	40	36	37	31	22	17	4	3	3	1	

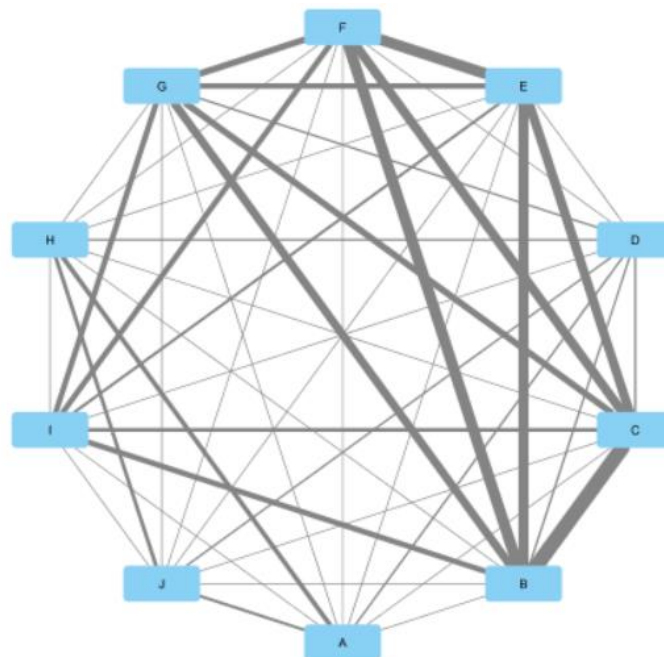
Tabulka 29 Stav sociální sítě coAuth na konci časového rámce 3

TF3	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		15	11	10	8	5	1	1	1	1	53
C	15		9	9	6	4	1	1	1	1	47
F	11	9		11	6	5	1	1	1	1	46
E	10	9	11		5	3	1	1	1	1	42
G	8	6	6	5		5	1	1	1	2	35
I	5	4	5	3	5		1	1	1	1	26
H	1	1	1	1	1	1		4	3	1	14
A	1	1	1	1	1	1	4		2	1	13
J	1	1	1	1	1	1	3	2		1	12
D	1	1	1	1	2	1	1	1	1		10
<i>rnd</i>	53	47	46	42	35	26	14	13	12	10	

Tabulka 30 Stav sociální sítě coAuth na konci časového rámce 4

TF4	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		16	12	11	9	6	1	1	1	2	59
C	16		10	10	7	4	1	1	1	2	52
F	12	10		12	7	6	1	1	1	1	51
E	11	10	12		6	3	1	1	1	1	46
G	9	7	7	6		6	1	1	1	2	40
I	6	4	6	3	6		1	1	1	1	29
H	1	1	1	1	1	1		5	4	2	17
A	1	1	1	1	1	1	5		3	2	16
J	1	1	1	1	1	1	4	3		2	15
D	2	2	1	1	2	1	2	2	2		15
<i>rnd</i>	59	52	51	46	40	29	17	16	15	15	

V Obrázku 45 lze vidět grafické znázornění stavu sociální sítě coAuth na konci časového rámce 4 vytvořené aplikací Cytoscape. Tloušťka čar reprezentujících hrany sítě odpovídá váze těchto hran. Na první pohled je tedy zřejmé, že dominantními uzly sítě jsou F, E, C, B, I a G, zatímco uzly D, A, J a H jsou významné méně.



Obrázek 45 Vizualizace sociální sítě coAuth na konci časového rámce 4

Tabulka 31 zobrazuje průměrné centrality sociální sítě coAuth v jednotlivých časových rámcích vypočítané pomocí CytoNCA. Hodnota SC vycházela vždy velmi vysoká nebo byla už přímo programem zaokrouhlována na nekonečno, proto při následném modelování

nebude tato hodnota sledována. Rovněž tak hodnota DC nebude sledována, neboť v rámci daného časového rámce pro stejný počet hran v síti musí vyjít vždy stejně.

Tabulka 31 Vypočtené průměrné centrality sociální sítě coAuth pro jednotlivé časové rámce

časový rámec	hran	BC	CC	DC	EC	LAC	NC	SC	IC
TF1	47	1,00	0,21	9,4	0,229	7,1	3,9	Inf	1,60
TF2	97	1,05	0,23	19,4	0,237	14,6	3,6	Inf	2,82
TF3	149	0,00	1,70	29,8	0,266	26,5	13,3	Inf	7,48
TF4	170	0,50	1,85	34,0	0,267	30,2	12,3	Inf	6,99

Při modelování sociální sítě coAuth byly pro finální modely všech časových období použity následující hodnoty parametrů:

Directed = False

Population_size = 10

Test_function = ramp

Dimensions = 2

Lower_bound = -10

Upper_bound = 10

Init_positions = {{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{9,-9},{9,-9},{9,-9},{9,-9}}

BetaMin = 0

Hodnoty zbylých parametrů se pro jednotlivá časová období lišily a jsou uvedeny v příslušných kapitolách.

8.2.1 Modelování sociální sítě coAuth v časovém rámci 4

Při posuzování kvality vytvořených modelů byla aplikována stejná pravidla jako při modelování citační sítě UIUI, která jsou popsána v kapitole 8.1.1 Modelování citační sítě UIUI v roce 2015.

Pro modelování sociální sítě v časovém rámci 4 (viz Tabulka 32 a Tabulka 33) byly úspěšně aplikovány velmi podobné techniky jako u citační sítě UIUI, tedy populace je rozdělena do dvou skupin o 6 a o 4 jedincích, umístěných v protilehlých rozích hyperplochy. Pravděpodobnost výběru *act* jedinců vychází z konečného žádaného stupně odpovídajících uzlů, přičemž u posledních čtyř jedinců s nejnižším stupněm je tato pravděpodobnost snížena na polovinu, zatímco u *rnd* jedinců je pravděpodobnost výběru prvních šesti jedinců

s nejvyšším stupněm (celá první skupina) stejná, zatímco u posledních čtyř jedinců (celá druhá skupina) je pravděpodobnost dvojnásobná.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro modelování časového rámce 4:

Interactions = 170

Act_probability = {59,52,51,46,40,29,9,8,7,7}

Rnd_probability = {1,1,1,1,1,1,2,2,2,2}

Tabulka 32 Modelování sociální sítě coAuth v časovém rámci 4

<i>modelling TF4</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,50	1,85	0,267	30,2	12,3	6,99	0
<i>manual_position_1</i>	0,45	2,86	0,282	29,9	11,1	6,14	340
<i>manual_position_2</i>	0,45	2,86	0,278	29,5	11,4	7,16	319
<i>manual_position_3</i>	0,60	2,45	0,269	28,5	11,1	7,01	163

Tabulka 33 Model *manual_position_3* sociální sítě coAuth na konci časového rámce 4

TF4	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		10	12	10	10	4	1	1	1	1	50
C	10		7	11	12	9	2	1	2	1	55
F	12	7		10	12	5	1		1	1	49
E	10	11	10		6	5	1	1	1	2	47
G	10	12	12	6		6	3		1	1	51
I	4	9	5	5	6		2	1		2	34
H	1	2	1	1	3	2		1	2	3	16
A	1	1		1		1	1		3	3	11
J	1	2	1	1	1		2	3		1	12
D	1	1	1	2	1	2	3	3	1		15
<i>rnd</i>	50	55	49	47	51	34	16	11	12	15	

8.2.2 Modelování sociální sítě coAuth v časovém rámci 3

Pro modelování sociální sítě v časovém rámci 3 (viz Tabulka 34 a Tabulka 35) platí stejná pravidla jako pro modelování sociální sítě v časovém rámci 4.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro modelování časového rámce 3:

Interactions = 149

Act _ probability = {53,47,46,42,35,26,7,7,6,5}

Rnd _ probability = {1,1,1,1,1,1,2,2,2,2}

Tabulka 34 Modelování sociální sítě coAuth v časovém rámci 3

<i>modelling TF3</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,00	1,70	0,266	26,5	13,3	7,48	0
manual_position_1	0,55	2,21	0,275	25,3	11,1	6,99	426
manual_position_2	0,45	2,39	0,279	25,8	12,0	7,25	382
manual_position_3	0,73	2,62	0,281	25,1	10,6	6,46	895

Tabulka 35 Model *manual_position_2* sociální sítě coAuth na konci časového rámce 3

TF3	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		9	6	9	7	6	1	2	2	2	44
C	9		10	8	4	8	2	1	1	1	44
F	6	10		10	8	7	1	1	1	1	45
E	9	8	10		6	4	2	1	2		42
G	7	4	8	6		5	2	1	2	1	36
I	6	8	7	4	5		1	1	2	2	36
H	1	2	1	2	2	1		2	1	1	13
A	2	1	1	1	1	1	2		1	1	11
J	2	1	1	2	2	2	1	1		3	15
D	2	1	1		1	2	1	1	3		12
<i>rnd</i>	44	44	45	42	36	36	13	11	15	12	

8.2.3 Modelování sociální sítě coAuth v časovém rámci 2

Pro modelování sociální sítě v časovém rámci 2 (viz Tabulka 36 a Tabulka 37) platí stejná pravidla jako pro modelování citační sítě v časovém rámci 3 a 4. Ani při různých obměnách modelovací techniky se nepodařilo dle ukazatele *pts* dosáhnout uspokojivého výsledku, a to zejména kvůli přibližně pětkrát převyšným hodnotám BC a CC, nicméně při vizuálním porovnání žádané matice sousednosti (viz Tabulka 28) a namodelované matice sousednosti (Tabulka 37) si matice podobné jsou.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro modelování časového rámce 2:

Interactions = 97

Act _ probability = {40,36,37,31,22,17,2,2,2,1}

Rnd _ probability = {1,1,1,1,1,1,2,2,2,2}

Tabulka 36 Modelování sociální sítě coAuth v časovém rámci 2

<i>modelling TF2</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	1,05	0,23	0,237	14,6	3,6	2,82	0
manual_position_1	4,80	1,25	0,250	13,0	3,8	3,98	32 607
manual_position_2	6,40	1,03	0,246	14,0	3,6	3,78	38 310
manual_position_3	6,80	1,25	0,247	14,0	3,3	3,25	49 876

Tabulka 37 Model *manual_position_1* sociální sítě coAuth na konci časového rámce 2

TF2	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		6	10	4	7	7					34
C	6		8	5	5	7	1			1	33
F	10	8		4	6	6		1			35
E	4	5	4		4	4				1	22
G	7	5	6	4		4	1				27
I	7	7	6	4	4						28
H		1			1			1	1		4
A			1				1		2	1	5
J							1	2			3
D		1		1				1			3
<i>rnd</i>	34	33	35	22	27	28	4	5	3	3	

8.2.4 Modelování sociální sítě coAuth v časovém rámci 1

Pro modelování sociální sítě v časovém rámci 1 (viz Tabulka 38 a Tabulka 39) byla použita stejná technika jako při modelování sítě v časovém rámci 2, 3 a 4, ovšem s tím rozdílem, že pravděpodobnost výběru *rnd* jedince byla pro sedmého, osmého a devátého jedince nulová, neboť v síti mají tyto tři uzly nulový vstupní stupeň.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro modelování časového rámce 1:

Interactions = 47

Act _ probability = {20,20,18,20,10,5,0,0,0,1}

Rnd _ probability = {1,1,1,1,1,1,0,0,0,2}

Tabulka 38 Modelování sociální sítě coAuth v časovém rámci 1

<i>modelling TF1</i>	<i>BC</i>	<i>CC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	1,00	0,21	0,229	7,1	3,9	1,60	0
manual_position_1	1,20	0,22	0,243	6,6	2,5	1,54	175
manual_position_2	1,20	0,22	0,242	7,1	2,9	1,57	108
manual_position_3	0,90	0,22	0,240	7,2	4,0	1,67	16

Tabulka 39 Model *manual_position_3* sociální sítě coAuth v časovém rámci 1

TF1	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		6	4	6	3	1					20
C	6		3	6	2	2				1	20
F	4	3		2	3	2				1	15
E	6	6	2		2	2					18
G	3	2	3	2		1					11
I	1	2	2	2	1						8
H											0
A											0
J											0
D		1	1								2
<i>rnd</i>	20	20	15	18	11	8	0	0	0	2	

9 PREDIKCE DYNAMIKY KOMPLEXNÍCH SÍTÍ POMOCÍ ALGORITMU FIREFLY

Tato kapitola popisuje predikce pro citační síť UIUI a sociální síť coAuth, a to vždy na jedno časové období, pro které jsou data dostupná, a je tedy možné kvalitu predikce posoudit, a na jedno období, pro které data zatím dostupná nejsou, a tak bude možné predikci ohodnotit až v budoucnu.

9.1 Citační síť Ústavu informatiky a umělé inteligence na FAI UTB

Kapitola 9.1 popisuje predikci citační sítě UIUI jednak pro rok 2015, kde je možné výsledek porovnat se skutečnými daty a ohodnotit tak kvalitu predikce, a jednak pro rok 2016, kde data zatím dostupná nejsou a kvalitu predikce tak bude možné posoudit až na konci tohoto kalendářního roku.

Při predikování citační sítě UIUI byly pro všechna časová období použity stejné hodnoty parametrů jako při modelování této sítě, tedy:

Directed = True

Population_size = 9

Test_function = ramp

Dimensions = 2

Lower_bound = -10

Upper_bound = 10

Init_positions = {{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{9,-9},{9,-9},{9,-9}}

BetaMin = 0

Hodnoty zbylých parametrů se pro jednotlivá časová období lišily a jsou uvedeny v příslušných kapitolách.

9.1.1 Predikování citační sítě UIUI na rok 2015

Při posuzování kvality vytvořených predikčních modelů pro rok 2015, kde jsou skutečná data již známa, a je tedy možné ohodnotit kvalitu predikce, byla aplikována stejná pravidla jako při modelování citační sítě UIUI, která jsou popsána v kapitole 8.1.1 Modelování citační sítě UIUI v roce 2015.

Predikování citační sítě na rok 2015 (viz Tabulka 40 a Tabulka 41) vycházelo z technik používaných pro modelování této sítě. Jako u všech modelů byla populace rozdělena do dvou skupin o 6 a o 3 jedincích, umístěných v protilehlých rozích hyperplochy, pravděpodobnost

výběru *act* jedinců vychází z konečného žádaného stupně odpovídajících uzlů, který byl pro každý uzel zvlášť odhadnut lineární regresí dat známých z let 2010 až 2014, přičemž u posledních čtyř jedinců s nejnižším stupněm je tato pravděpodobnost snížena na polovinu. U *rnd* jedinců je pravděpodobnost výběru prvních šesti jedinců s nejvyšším stupněm stejná, sedmý jedinec má pravděpodobnost trojnásobnou, osmý dvojnásobnou a devátý nulovou, tedy přesně tak, jak tomu bylo při modelování sítě ve dvou předcházejících letech 2013 a 2014.

Predikce není příliš přesná, neboť v roce 2015 začal vytvářet vazby i uzel JJ, což z dat dostupných do roku 2014 nebylo možné předpokládat.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro predikování roku 2015:

Interactions = 1514

Act_probability = {350,354,231,222,219,46,14,11,0}

Rnd_probability = {1,1,1,1,1,1,3,2,0}

Tabulka 40 Predikování citační sítě UIUI na rok 2015

<i>predicting 2015</i>	<i>BC</i>	<i>CC</i>	<i>DC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,74	6,39	345,11	0,262	364	25,3	12,6	0
<i>manual_position_1</i>	0,00	0,75	336,44	0,260	288	19,6	5,2	2 217
<i>manual_position_2</i>	0,00	0,75	336,44	0,263	304	18,8	5,1	2 228
<i>manual_position_3</i>	0,00	0,75	336,44	0,262	290	19,5	5,2	2 224

Tabulka 41 Model *manual_position_1* predikce citační sítě UIUI na rok 2015

2015	<i>IZ</i>	<i>RS</i>	<i>MP</i>	<i>DD</i>	<i>ZO</i>	<i>RJ</i>	<i>PV</i>	<i>BC</i>	<i>JJ</i>	<i>act</i>
IZ		98	52	93	75	19	6			343
RS	112		51	103	81	16	6			369
MP	65	58		66	45	9	2			245
DD	71	63	30		40	7	1			212
ZO	68	58	25	49		10	4			214
RJ	26	15	3	16	19		1	2		82
PV	4	3	1	2	6	5		6		27
BC	1	2	2	2	2	12	1			22
JJ										0
<i>rnd</i>	347	297	164	331	268	78	21	8	0	

9.1.2 Predikování citační sítě UIUI na rok 2016

Pro predikování citační sítě na rok 2016 (viz Tabulka 42 a Tabulka 43) platí stejná pravidla jako pro predikování citační sítě na rok 2015, ovšem s tím rozdílem, že pravděpodobnost výběru devátého jedince jako *rnd* jedince byla z nuly navýšena na stejnou hodnotu jako u prvních šesti jedinců, neboť se předpokládá, že když odpovídající autor JJ v loňském roce vydal vědecké články citující ostatní autory, budou pravděpodobně v tomto roce i ostatní autoři citovat články autora JJ, stejně jako to probíhalo u ostatních začínajících autorů.

Skutečná data budou dostupná až na konci roku 2016, zatím tedy není možné ohodnotit kvalitu predikce.

Hodnoty parametrů modelu *manual_position* citační sítě UIUI pro predikování roku 2016:

Interactions = 1863

Act_probability = {430,432,303,279,257,52,14,12,10}

Rnd_probability = {1,1,1,1,1,1,3,2,1}

Tabulka 42 Predikování citační sítě UIUI na rok 2016

<i>predicting 2016</i>	<i>BC</i>	<i>CC</i>	<i>DC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>
manual_position_1	0,54	8,25	414	0,264	391	24,6	14,4

Tabulka 43 Model *manual_position_1* predikce citační sítě UIUI na rok 2016

2016	IZ	RS	MP	DD	ZO	RJ	PV	BC	JJ	<i>act</i>
IZ		122	77	115	98	31	6			449
RS	119		69	102	87	23	6			406
MP	74	79		76	51	11	2			293
DD	86	87	52		51	10	1			287
ZO	71	74	33	59		17	4			258
RJ	25	23	5	22	21		3	4		103
PV	5	3		2	4	6		5		25
BC	1	2	3	1	1	13	4		1	26
JJ	2	2	2	3	3		3	1		16
<i>rnd</i>	383	392	241	380	316	111	29	10	1	

9.2 Sociální síť spoluautorství nad fyzikálními vědeckými články

Kapitola 9.2 popisuje predikci sociální sítě coAuth jednak pro časový rámec 4, kde je možné výsledek porovnat se skutečnými daty a ohodnotit tak kvalitu predikce, a jednak pro časový

rámec 5, kde data zatím dostupná nejsou a kvalitu predikce tak bude možné posoudit až v budoucnu.

Při predikování sociální sítě coAuth byly pro všechna časová období použity stejné hodnoty parametrů jako při modelování této sítě, tedy:

Directed = False

Population_size = 10

Test_function = ramp

Dimensions = 2

Lower_bound = -10

Upper_bound = 10

Init_positions = {{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{-9,9},{9,-9},{9,-9},{9,-9},{9,-9}}

BetaMin = 0

Hodnoty zbylých parametrů se pro jednotlivá časová období lišily a jsou uvedeny v příslušných kapitolách.

9.2.1 Predikování sociální sítě coAuth na časový rámec 4

Při posuzování kvality vytvořených predikčních modelů pro časový rámec 4, kde jsou skutečná data již známa, a je tedy možné ohodnotit kvalitu predikce, byla aplikována stejná pravidla jako při modelování citační sítě UIUI, která jsou popsána v kapitole 8.1.1 Modelování citační sítě UIUI v roce 2015.

Predikování sociální sítě na časový rámec 4 (viz Tabulka 44 a Tabulka 45) vycházelo z technik používaných pro modelování této sítě. Jako u všech modelů byla populace rozdělena do dvou skupin o 6 a o 4 jedincích, umístěných v protilehlých rozích hyperplochy. Pravděpodobnost výběru *act* jedinců vychází z konečného žádaného stupně odpovídajících uzlů, který byl pro každý uzel zvlášť odhadnut lineární regresí dat známých z časových rámců 1 až 3, přičemž u posledních čtyř jedinců s nejnižším stupněm je tato pravděpodobnost snížena na polovinu, zatímco u *rnd* jedinců je pravděpodobnost výběru prvních šesti jedinců s nejvyšším stupněm (celá první skupina) stejná, zatímco u posledních čtyř jedinců (celá druhá skupina) je pravděpodobnost dvojnásobná.

Predikce vychází velmi přesně, dokonce srovnatelně s výsledky modelování této sítě podle skutečných dat. Drobnými nepřesnostmi byl odhad počtu hran na konci časového rámce 4, který byl lineární regresí odhadnut na 200, přičemž ve skutečnosti bylo hran pouze 170, a

také přibližně 50% odchylka parametru CC, ke které ale docházelo již při modelování sítě v tomto časovém rámci.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro predikování časového rámce 4:

Interactions = 200

Act_probability = {71,61,62,53,47,37,10,9,9,7}

Rnd_probability = {1,1,1,1,1,1,2,2,2,2}

Tabulka 44 Predikování sociální sítě coAuth na časový rámec 4

<i>predicting TF4</i>	<i>BC</i>	<i>CC</i>	<i>DC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>	<i>pts</i>
real	0,50	1,85	34	0,267	30,2	12,3	6,99	0
manual_position_1	0,62	2,73	40	0,267	35,6	13,4	7,65	362
manual_position_2	0,37	2,66	40	0,265	35,6	13,0	6,77	332
manual_position_3	0,60	2,58	40	0,267	35,6	12,5	6,48	265

Tabulka 45 Model *manual_position_3* predikce sociální sítě coAuth na časový rámec 4

TF4	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		18	14	13	8	8	1	1	1	1	65
C	18		11	12	7	7	1	1	2	1	60
F	14	11		15	8	8	2	1	1	1	61
E	13	12	15		8	5	1	1	2	1	58
G	8	7	8	8		8	1	1	1	2	44
I	8	7	8	5	8		2	1	1	2	42
H	1	1	2	1	1	2		6	4	2	20
A	1	1	1	1	1	1	6		3	3	18
J	1	2	1	2	1	1	4	3		2	17
D	1	1	1	1	2	2	2	3	2		15
<i>rnd</i>	65	60	61	58	44	42	20	18	17	15	

9.2.2 Predikování sociální sítě coAuth na časový rámec 5

Pro predikování sociální sítě na časový rámec 5 (viz Tabulka 46 a Tabulka 47) platí úplně stejná pravidla jako pro predikování sociální sítě na časový rámec 4, neboť není důvod očekávat výraznější změnu v trendu vývoje této sítě a měnit tedy metodu osvědčenou jak při modelování, tak při předešlé predikci.

Skutečná data časového rámce 5 pro časy 1026 – 1075 snad budou časem zveřejněna na webových stránkách <http://snap.stanford.edu/data/index.html>, zatím tedy není možné ohodnotit kvalitu této predikce.

Hodnoty parametrů modelu *manual_position* sociální sítě coAuth pro predikování časového rámce 5:

Interactions = 221

Act_probability = {76,66,65,57,53,40,12,12,11,10}

Rnd_probability = {1,1,1,1,1,1,2,2,2,2}

Tabulka 46 Predikování sociální sítě coAuth na časový rámec 5

<i>predicting TF5</i>	<i>BC</i>	<i>CC</i>	<i>DC</i>	<i>EC</i>	<i>LAC</i>	<i>NC</i>	<i>IC</i>
manual_position_1	0,45	2,97	44,2	0,270	39,3	12,1	4,28

Tabulka 47 Model *manual_position_1* predikce sociální sítě coAuth na časový rámec 5

TF5	B	C	F	E	G	I	H	A	J	D	<i>act</i>
B		19	16	12	10	8	1	2	1	3	72
C	19		13	13	11	7	1	2	1	2	69
F	16	13		15	8	9	1	1	1	2	66
E	12	13	15		8	6	1	1	1	1	58
G	10	11	8	8		7	2	2	2	2	52
I	8	7	9	6	7		2	1	2	1	43
H	1	1	1	1	2	2		6	5	2	21
A	2	2	1	1	2	1	6		4	3	22
J	1	1	1	1	2	2	5	4		3	20
D	3	2	2	1	2	1	2	3	3		19
<i>rnd</i>	72	69	66	58	52	43	21	22	20	19	

ZÁVĚR

Diplomová práce navazuje na současné trendy v metaheuristických algoritmech, tedy využití informací o dynamice populace jedinců pro její převod na komplexní síť a následnou analýzu pro možné zpětnovazební zásahy adaptace a učení. Práce staví na těchto základech, ale cílem bylo provést reverzní proces, tedy předpovídat vývoj komplexních sítí v reálném světě, tedy namísto hejnového algoritmu na vstupu a komplexní síť na výstupu je na vstupu žadaná struktura komplexní síť a na výstupu vhodné nastavení parametrů hejnového algoritmu pro co možná nejpřesnější modelování žadané komplexní síť.

Zdrojem vhodných datových souborů popisujících reálné komplexní síť byla nakonec vybrána databáze Scopus (<https://www.scopus.com/>), odkud je možné získat váženou orientovanou citační síť s časovými razítky vzniku všech hran, nicméně databáze postrádá nástroj pro elegantní export dat v textové podobě, takže data bylo nutné z vykreslených tabulek vypisovat ručně, což by už pro síť o několika desítkách autorů nebylo časově únosné. Nespornou výhodou zůstává, že se nejedná o anonymní data s nejistým původem, ale o zpětně ověřitelné informace o konkrétních vědeckých pracovnících FAI UTB ve Zlíně.

V případě sociálních sítí jsou na webových stránkách SNAP (<http://snap.stanford.edu/data/index.html>) k dispozici datové soubory pro populární sociální síť jako jsou Facebook či Twitter, nicméně všechny tyto zdejší datové soubory postrádají informaci o době vzniku hran sítě, tudíž nejsou vhodné pro predikci dynamiky těchto sítí. Datové soubory s časovými razítky je sice možné od uvedených společností zakoupit, nicméně by se na nákup během necelého půl roku řešení této diplomové práce nestihl vyspat grant. Místo toho byla zvolena sociální síť nad fyzikálními vědeckými články ze stejného zdroje, která všechny potřebné informace obsahuje.

Finální parametry pro modelování v jednotlivých časových rámcích vykazují u obou sítí velmi podobné rysy (dokonce i mezi oběma sítěmi navzájem), proto bylo možné pokusit se dynamiku sítí na krátký časový okamžik předpovídat. K výraznějším úpravám je nutné přistoupit pouze v momentech, kdy dosud neaktivní uzel začíná vytvářet vazby s okolím. Okamžiky, kdy se dosud samostatný uzel začlení do sítě, je však prakticky nemožné předpovídat. Vytvořené modely a predikce, které již bylo možné ohodnotit, kopírovaly skutečnost velmi věrohodně.

Vzhledem k modelování pomocí heuristického algoritmu bude k menším odchýlkám docházet vždy, neboť ani dva průběhy heuristického algoritmu s totožnými hodnotami

parametrů obecně nebudou stejné. Lze předpokládat, že při modelování či predikování rozsáhlejších sítí s větším počtem uzlů a hran, ke kterým by navíc byly dostupné informace o vývoji sítě v delším časovém období, by kvalita vytvořených modelů byla ještě vyšší, neboť připojení dalšího uzlu by v tomto měřítku nemělo tak výrazný dopad na změnu celkové struktury sítě.

Námětem na vylepšení je zahrnutí výpočtů centralit komplexních sítí přímo do vytvořené aplikace FireCNet, což by mělo za následek oproštění od nutnosti přesouvání dat do aplikace Cytoscape. Tak by bylo možné celý proces modelování více zautomatizovat. Bylo by možné spustit několik opakování běhu algoritmu Firefly se stejnými parametry dávkově a jednoduše pak v rámci aplikace FireCNet srovnat dosažené hodnoty centralit s požadovanými a výsledky následně zprůměrovat a zobrazit.

Ještě ambicióznějším námětem na vylepšení by bylo implementování „vnějšího“ evolučního algoritmu, který by měl za úkol hledat optimální nastavení parametrů „vnitřního“ evolučního algoritmu (tj. současná implementace algoritmu Firefly) jako je typ testovací funkce, matice počátečních pozic jedinců, vektor rozložení pravděpodobnosti výběru *act* i *rnd* jedince nebo hodnotu parametru *betaMin*.

Pro mezinárodní srozumitelnost byla jazykem uživatelského prostředí aplikace FireCNet, ale i názvů proměnných a funkcí ve zdrojových kódech, zvolena angličtina. Zdrojové kódy jsou přehledně členěny a jejich důležité části stručně okomentovány. Názvy funkcí a proměnných byly volena tak, aby i za cenu delšího názvu co možná nejvíce vystihovaly jejich účel.

Diplomovou práci i zdrojový kód aplikace FireCNet jsem se snažil psát stručně, jasně a výstižně. Kde to bylo možné, tam jsem text doplňoval tabulkami a obrázky, tak jak je v technické dokumentaci dobrým zvykem. Na tuto práci bych na FAI UTB ve Zlíně rád navázal v doktorském studijním programu.

Aplikaci FireCNet, stejně jako textové soubory všech reálných, modelovaných a predikovaných matic sousednosti zkoumané citační a sociální komplexní sítě je možno nalézt v přílohách této diplomové práce.

SEZNAM POUŽITÉ LITERATURY

- [1] BARABÁSI, Albert-László. *V pavučině síti*. Vyd. 1. V Praze: Paseka, 2005. Fénix (Paseka). ISBN 80-718-5751-3.
- [2] PELÁNEK, Radek. *Modelování a simulace komplexních systémů: jak lépe porozumět světu*. Vyd. 1. Brno: Masarykova univerzita, 2011. ISBN 978-80-210-5318-2.
- [3] GÁLIK, Martin. *Modely a metody v oblasti komplexních síti*. České Budějovice, 2012. Bakalářská práce. Jihočeská univerzita v Českých Budějovicích. Vedoucí práce Ladislav Beránek.
- [4] Teorie grafů. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2016-01-27]. Dostupné z: https://cs.wikipedia.org/wiki/Teorie_graf%C5%AF
- [5] DUNCAN J. WATTS. *Small worlds: the dynamics of networks between order and randomness*. 8th print., and 1st pbk. print. Princeton, N.J: Princeton University Press, 1999. ISBN 06-911-1704-7.
- [6] JANOŠTÍK, Jakub. *Aplikace pro analýzu a vizualizaci komplexních síti*. Zlín, 2014. Diplomová práce. Univerzita Tomáše Bati ve Zlíně. Vedoucí práce Roman Šenkeřík.
- [7] PELÁNEK, Radek. *Modelování komplexních síti* [online]. [cit. 2016-01-30]. Dostupné z: <http://www.fi.muni.cz/~xpelane/IV109/slidy/networks.pdf>
- [8] ZELINKA, Ivan. *Complex Networks* [online]. 2015 [cit. 2016-01-31]. Dostupné z: <http://dataanalysis.vsb.cz/data/Vyuka/NAVY/ComplexNetworks.pdf>
- [9] TOMASZEK, Lukáš. *Dynamika komplexních síti*. Ostrava, 2015. Diplomová práce. VŠB - Technická univerzita Ostrava. Vedoucí práce Pavla Dráždilová.
- [10] BORNHOLDT, Stefan a Heinz Georg SCHUSTER (eds.). *Handbook of graphs and networks: from the genome to the internet*. 1st ed. Weinheim: Wiley-VCH, c2003. ISBN 35-274-0336-1.
- [11] LATORA, Vito a Massimo MARCHIORI. *A measure of centrality based on the network efficiency* [online]. [cit. 2016-02-05]. Dostupné z: <http://arxiv.org/pdf/cond-mat/0402050v1.pdf>
- [12] TANG, Yu. *CytoNCA: a Cytoscape apps for network centralities analysis (Version 2.1)* [online]. 2014 [cit. 2016-02-14]. Dostupné z:

<https://drive.google.com/file/d/0B13gZNIVofBmcVZld1pjY0tDeDg/edit?pref=2&pli=1>

- [13] PEARL, Judea. *Heuristics: intelligent search strategies for computer problem solving*. Reading, Mass.: Addison-Wesley Pub. Co., c1984. ISBN 02-010-5594-5.
- [14] KENNEDY, James F, Russell C EBERHART a Yuhui SHI. *Swarm intelligence*. San Francisco: Morgan Kaufmann Publishers, c2001. ISBN 15-586-0595-9.
- [15] ZELINKA, Ivan. *Evoluční výpočetní techniky: principy a aplikace*. 1. vyd. Praha: BEN - technická literatura, 2009. ISBN 978-80-7300-218-3.
- [16] MOČUBOVÁ, Dominika. *Bioluminiscenční organizmy*. Brno, 2006. Bakalářská práce. Masarykova univerzita. Vedoucí práce Pavel Hyršl.
- [17] YANG, Xin-She. *Swarm intelligence and bio-inspired computation: theory and applications*. 1st ed. Boston: Elsevier, 2013. Elsevier insights. ISBN 01-240-5163-4.
- [18] YANG, Xin She a Xingshi HE. Firefly algorithm: recent advances and applications. *International Journal of Swarm Intelligence* [online]. 2013, 1(1), 36- [cit. 2016-02-28]. DOI: 10.1504/IJSI.2013.055801. ISSN 2049-4041. Dostupné z: <http://www.inderscience.com/link.php?id=55801>
- [19] *Wolfram: Computation Meets Knowledge* [online]. [cit. 2016-03-07]. Dostupné z: <http://www.wolfram.com/index.php>
- [20] *Cytoscape: Network Data Integration, Analysis, and Visualization in a Box* [online]. [cit. 2016-03-09]. Dostupné z: <http://www.cytoscape.org/>
- [21] aMatReader: App to read adjacency matrix (.mat) files. *Cytoscape App Store* [online]. [cit. 2016-03-09]. Dostupné z: <http://apps.cytoscape.org/apps/amatreader>
- [22] aMatReader: A Cytoscape Reader for Adjacency Matrices. *Resource for Biocomputing, Visualization, and Informatics* [online]. [cit. 2016-03-09]. Dostupné z: <http://www.cgl.ucsf.edu/cytoscape/utilities3/amatreader.shtml>
- [23] CytonCA: Providing calculation, evaluation and visualization analysis for several centralities of weighted and unweighted network. *Cytoscape App Store* [online]. [cit. 2016-03-10]. Dostupné z: <http://apps.cytoscape.org/apps/cytonca>
- [24] Optimization Test Problems: Rastrigin Function. *Virtual Library of Simulation Experiments: Test Functions and Datasets* [online]. 2015 [cit. 2016-03-18]. Dostupné z: <http://www.sfu.ca/~ssurjano/rastr.html>

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

2D	Dvoudimenzionální, dvourozměrný.
3D	Trojdimenzionální, trojrozměrný.
act	Aktivní jedinec, který je v dané iteraci přitahován k „náhodnému“ jedinci <i>rnd</i> .
adj mat	Adjacency matrix, matice sousednosti.
app	Application, aplikace.
BC	Betweenness centrality.
CC	Closeness centrality.
CDF	Computable document format, formát souboru pro aplikaci CDF Player.
CFVal	Cost function value, hodnota testovací funkce.
coAuth	Co-authorship, spoluautorství.
DC	Degree centrality.
dim	Dimension, dimenze.
e	Eulerovo číslo, přibližně 2,71828.
EC	Eigenvector centrality.
FAI	Fakulta aplikované informatiky.
FEI	Fakulta elektrotechniky a informatiky.
IC	Information centrality.
Inf	Infinity, nekonečno.
init	Initialize, nastavit na počáteční hodnotu.
LAC	Local average connectivity-based method.
NC	Network centrality.
no	Number, číslo.
pts	Points, body.
rnd	„Náhodný“ jedinec, který v dané iteraci přitahuje aktivního jedince <i>act</i> .
SC	Subgraph centrality.

secs	Seconds, sekundy.
SW	Software, programové vybavení počítače.
TF	Time frame, časový rámec.
UIUI	Ústav informatiky a umělé inteligence.
UTB	Univerzita Tomáše Bati.
VŠB-TUO	Vysoká škola báňská – Technická univerzita Ostrava.

SEZNAM OBRÁZKŮ

Obrázek 1 Síť dokumentů na Internetu[3]	15
Obrázek 2 Citační síť[3]	16
Obrázek 3 Síť bílkovin[7]	16
Obrázek 4 Potravní řetězec[8]	17
Obrázek 5 Síť známostí mezi bělochy a černochoy[8]	18
Obrázek 6 Jednoduchý graf[9]	19
Obrázek 7 Podgraf[9]	19
Obrázek 8 Orientovaný graf[9]	20
Obrázek 9 Vážený graf[9]	20
Obrázek 10 Graf s označenými stupni vrcholů[9]	21
Obrázek 11 Incidenční matice grafu na Obrázku 6	22
Obrázek 12 Incidenční matice grafu na Obrázku 8	22
Obrázek 13 Matice sousednosti grafu na Obrázku 6	23
Obrázek 14 Matice sousednosti grafu na Obrázku 8	23
Obrázek 15 Mocninný zákon distribuce stupňů uzlů jako důkaz bezškálovosti komplexních sítí[3]	25
Obrázek 16 Screenshot aplikace Cytoscape s načtenou orientovanou a váženou komplexní sítí	36
Obrázek 17 Volba oddělovače buněk a orientace hran při načítání vstupního souboru zásuvným modulem aMatReader	37
Obrázek 18 Ovládací panel pro volbu vyhodnocovaných centralit pomocí CytoNCA	38
Obrázek 19 Výsledkový panel pro vizualizaci vypočtených centralit pomocí CytoNCA	38
Obrázek 20 Nekonečná nakloněná rovina, 2D graf na intervalu $\langle -10; 10 \rangle$	42
Obrázek 21 Nekonečná nakloněná rovina, 3D graf na intervalu $\langle -10; 10 \rangle$	42
Obrázek 22 První de Jongova funkce, 2D graf na intervalu $\langle -5,12; 5,11 \rangle$	43
Obrázek 23 První de Jongova funkce, 3D graf na intervalu $\langle -5,12; 5,11 \rangle$	43
Obrázek 24 Druhá de Jongova funkce, 2D graf na intervalu $\langle -2,048; 2,047 \rangle$	44
Obrázek 25 Druhá de Jongova funkce, 3D graf na intervalu $\langle -2,048; 2,047 \rangle$	44
Obrázek 26 Rastriginova funkce, 2D graf na intervalu $\langle -5,12; 5,11 \rangle$	44
Obrázek 27 Rastriginova funkce, 3D graf na intervalu $\langle -5,12; 5,11 \rangle$	44

Obrázek 28 Schwefelova funkce, 2D graf na intervalu $\langle -512, 511 \rangle$	45
Obrázek 29 Schwefelova funkce, 3D graf na intervalu $\langle -512, 511 \rangle$	45
Obrázek 30 Griewangkova funkce, 2D graf na intervalu $\langle -100, 100 \rangle$	46
Obrázek 31 Griewangkova funkce, 3D graf na intervalu $\langle -100, 100 \rangle$	46
Obrázek 32 Mastersova funkce, 2D graf na intervalu $\langle -5, 5 \rangle$	46
Obrázek 33 Mastersova funkce, 3D graf na intervalu $\langle -5, 5 \rangle$	46
Obrázek 34 Původní pozice jedinců algoritmu Firefly před „úspěšnou“ interakcí aktivního jedince E s „náhodným“ jedincem C.....	48
Obrázek 35 Nové pozice jedinců algoritmu Firefly po „úspěšné“ interakci aktivního jedince E s „náhodným“ jedincem C.....	48
Obrázek 36 Matice sousednosti před „úspěšnou“ interakcí.....	48
Obrázek 37 Matice sousednosti po „úspěšné“ interakci.....	48
Obrázek 38 Vzhled aplikace FireCNet po spuštění	49
Obrázek 39 Aplikace FireCNet s rozevřeným seznamem pro přednastavení parametrů	50
Obrázek 40 Aplikace FireCNet po úspěšném importu matice sousednosti.....	51
Obrázek 41 Aplikace FireCNet s otevřeným manipulátorem pro nastavení počtu uzlů v komplexní síti (resp. počtu jedinců v populaci)	52
Obrázek 42 Aplikace FireCNet s rozevřeným seznamem pro výběr testovací funkce	53
Obrázek 43 Aplikace FireCNet zobrazující výsledky predikování dynamiky importované komplexní sítě	56
Obrázek 44 Vizualizace citační sítě UIUI na konci roku 2015	64
Obrázek 45 Vizualizace sociální sítě coAuth na konci časového rámce 4.....	75

SEZNAM TABULEK

Tabulka 1 Shrnutí příkladů komplexních sítí[7][3]	18
Tabulka 2 Seznam sousedů grafu na Obrázku 6	21
Tabulka 3 Definice centralit pro nevážený graf[12]	27
Tabulka 4 Definice centralit pro vážený graf, 1. část[12]	28
Tabulka 5 Definice centralit pro vážený graf, 2. část[12]	29
Tabulka 6 Stav citační sítě UIUI na konci roku 2010.....	62
Tabulka 7 Stav citační sítě UIUI na konci roku 2011.....	62
Tabulka 8 Stav citační sítě UIUI na konci roku 2012.....	62
Tabulka 9 Stav citační sítě UIUI na konci roku 2013.....	63
Tabulka 10 Stav citační sítě UIUI na konci roku 2014.....	63
Tabulka 11 Stav citační sítě UIUI na konci roku 2015.....	63
Tabulka 12 Vypočtené průměrné centrality citační sítě UIUI pro jednotlivé roky	64
Tabulka 13 Vývoj modelování citační sítě UIUI v roce 2015	66
Tabulka 14 Model <i>exact_1</i> citační sítě UIUI na konci roku 2015.....	67
Tabulka 15 Model <i>exact_position_2</i> citační sítě UIUI na konci roku 2015	68
Tabulka 16 Model <i>manual_position_1</i> citační sítě UIUI na konci roku 2015	68
Tabulka 17 Modelování citační sítě UIUI v roce 2014	69
Tabulka 18 Model <i>manual_position_2</i> citační sítě UIUI na konci roku 2014	69
Tabulka 19 Modelování citační sítě UIUI v roce 2013	70
Tabulka 20 Model <i>manual_position_3</i> citační sítě UIUI na konci roku 2013	70
Tabulka 21 Modelování citační sítě UIUI v roce 2012	71
Tabulka 22 Model <i>manual_position_3</i> citační sítě UIUI na konci roku 2012	71
Tabulka 23 Modelování citační sítě UIUI v roce 2011	72
Tabulka 24 Model <i>manual_position_1</i> citační sítě UIUI na konci roku 2011	72
Tabulka 25 Modelování citační sítě UIUI v roce 2010	72
Tabulka 26 Model <i>manual_position_1</i> citační sítě UIUI na konci roku 2010	73
Tabulka 27 Stav sociální sítě coAuth na konci časového rámce 1	74
Tabulka 28 Stav sociální sítě coAuth na konci časového rámce 2	74
Tabulka 29 Stav sociální sítě coAuth na konci časového rámce 3	74
Tabulka 30 Stav sociální sítě coAuth na konci časového rámce 4	75
Tabulka 31 Vypočtené průměrné centrality sociální sítě coAuth pro jednotlivé časové rámce	76

Tabulka 32 Modelování sociální sítě coAuth v časovém rámci 4	77
Tabulka 33 Model <i>manual_position_3</i> sociální sítě coAuth na konci časového rámce 4	77
Tabulka 34 Modelování sociální sítě coAuth v časovém rámci 3	78
Tabulka 35 Model <i>manual_position_2</i> sociální sítě coAuth na konci časového rámce 3	78
Tabulka 36 Modelování sociální sítě coAuth v časovém rámci 2	79
Tabulka 37 Model <i>manual_position_1</i> sociální sítě coAuth na konci časového rámce 2	79
Tabulka 38 Modelování sociální sítě coAuth v časovém rámci 1	80
Tabulka 39 Model <i>manual_position_3</i> sociální sítě coAuth v časovém rámci 1	80
Tabulka 40 Predikování citační sítě UIUI na rok 2015	82
Tabulka 41 Model <i>manual_position_1</i> predikce citační sítě UIUI na rok 2015	82
Tabulka 42 Predikování citační sítě UIUI na rok 2016	83
Tabulka 43 Model <i>manual_position_1</i> predikce citační sítě UIUI na rok 2016	83
Tabulka 44 Predikování sociální sítě coAuth na časový rámeček 4	85
Tabulka 45 Model <i>manual_position_3</i> predikce sociální sítě coAuth na časový rámeček 4	85
Tabulka 46 Predikování sociální sítě coAuth na časový rámeček 5	86
Tabulka 47 Model <i>manual_position_1</i> predikce sociální sítě coAuth na časový rámeček 5	86

SEZNAM PŘÍLOH

Příloha P I: Obsah přiloženého CD.

PŘÍLOHA P I: OBSAH PŘILOŽENÉHO CD

CD obsahuje:

[Networks]	Složka s textovými soubory reálných, modelovaných a predikovaných matic sousednosti zkoumané citační a sociální komplexní sítě.
FireCNet.cdf	Vytvořená aplikace FireCNet ve formátu pro spouštění ve freeware SW CDF Player.
FireCNet.nb	Vytvořená aplikace FireCNet ve formátu pro spouštění v komerčním SW Wolfram Mathematica.
fulltext.pdf	Úplný text diplomové práce.