

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 103003/I/2024/36114651040459780

ANALÝZA SOCIÁLNYCH SIETÍ
Diplomová práca

2024

Bc. Silvia Kmeťková

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

ANALÝZA SOCIÁLNYCH SIETÍ

Diplomová práca

Študijný program: Informačný manažment

Študijný odbor: Ekonómia a manažment

Školiace pracovisko: Katedra operačného výskumu a ekonometrie FHI

Vedúci záverečnej práce: Ing. Pavel Géžik, PhD.

Bratislava 2024

Bc. Silvia Kmet'ková

Čestné vyhlásenie

Vyhlasujem, že diplomovú prácu som vypracovala samostatne, a že som uviedla všetky použité zdroje literatúry súvisiace so zameraním tejto práce.

Dátum: 4. mája 2024

.....

Podpis študenta

Pod'akovanie

Rada by som sa pod'akovala za spoluprácu a odborné vedenie môjho školiteľa, Ing. Pavla Géžika, PhD., ktorý mi bol veľkou oporou a pomocou pri písaní tejto záverečnej práce. Vďaka jeho prístupu, vedeniu a radám, rovnako ako konštruktívnej kritike voči názorom prvotne mienených v práci, bolo možné vytvoriť a spracovať prácu a jej problematiku kvalitnejšie a na vyššej úrovni.

ABSTRAKT

KMEŤKOVÁ, Silvia: Analýza sociálnych sietí – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky; Katedra operačného výskumu a ekonometrie – Vedúci záverečnej práce: Ing. Pavel Gežík, PhD., Bratislava: FHI, 2022, počet strán: 59.

Záverečná diplomová práca je zameraná na sociálne siete a ich analýzu. Jej cieľom je hlavne aplikovanie niekoľkých známych ale aj menej známych aplikácií na ľubovoľne vybranú sociálnu sieť, ktorou je v tejto záverečnej práci sociálna sieť Twitch. Práca využíva pre teóriu grafov výsledné grafy a modely, spojené s odbornou terminológiou, nadväzujúcou na danú tému. Približuje a snaží sa o podporu a propagáciu využitia analýzy a nástrojov pre jej vykreslenie. Osobité časti záverečnej práce sú zamerané na definíciu, analyzovanie, aplikovanie a posúdenie aplikácií pre lepšie zhodnotenie a následne použitie jednotlivých nástrojov podľa potreby. Výsledkov riešenia tejto problematiky sú ilustračné obrázky z použitých nástrojov pre analýzu a ich subjektívny pohľad autorky na ich počiatočné a nasledujúce použitie.

Kľúčové slová: Analýza sociálnych sietí, Teória grafov, Gephi, NetworkX, NetMiner, SocNetV

ABSTRAKT

KMEŤKOVÁ, Silvia: *Social Network Analysis* – University of Economics in Bratislava. Faculty of Economic Informatics; Department of Operations Research and Econometrics – Supervisor: Ing. Pavel Géžik, PhD., Bratislava: FHI, 2022, number of pages: 59.

The final thesis is focused on social networks and their analysis. Its main goal is to apply several well-known but also less well-known applications to an arbitrarily selected social network, which in this thesis is the social network Twitch. The thesis uses the resulting graphs and models for the theory of graphs, connected with the technical terminology related to the given topic. It approaches and tries to support and promote the use of analysis and tools for its rendering. Parts of the final thesis are focused on the definition, analysis, application and assessment of applications for better evaluation and subsequent use of individual tools as needed. The results of solving this problem are illustrative pictures of the used tools for analysis and their subjective view of the author on their initial and subsequent use.

Key Words: Social Network Analysis, Graph Theory, Gephi, NetworkX, NetMiner, SocNetV

OBSAH

Úvod	10
1 Súčasný stav riešenej problematiky doma a v zahraničí	12
1.1 Počiatky analýzy sociálnych sietí	12
1.2 Sociálne siete	13
1.3 Pojmy sociálnych sietí	13
1.3.1 Účastník	14
1.3.2 Vzťahová väzba	14
1.3.3 Vzťah	14
1.3.4 Dyáda	14
1.3.5 Triáda	15
1.3.6 Podskupina	15
1.3.7 Skupina	15
1.3.8 Sociálna sieť	16
1.4 Charakteristika analýzy sociálnych sietí	16
1.5 Príklad sociálnej siete: Analýza a Interpretácia	17
1.6 Teória grafov	19
1.6.1 Aplikácie teórie grafov	19
1.7 Sociogram	20
1.8 Vplyv nástrojov na analýzu a vizualizáciu sietí	21
1.8.1 EgoNet	21
1.8.2 Gephi	21
1.8.3 NetMiner	21
1.8.4 NetworkX	22
1.8.5 SocNetV (Social Networks Visualizer)	22
1.8.6 R	22
2 Cieľ práce	23
3 Metodika práce a metódy skúmania	24
3.1 Pracovný postup	24
3.2 Spôsob získavania dát, ich modifikácia a zdroje	24
3.3 Metódy analýzy	26
3.3.1 Gephi:	26
3.3.2 NetMiner:	26

3.3.3	SocNetV:	27
3.3.4	NetworkX:	27
4	Výsledky práce a diskusia	29
4.1	SocNetV	30
4.1.1	Vizuálna stránka programu	30
4.1.2	Používateľská prívetivosť	30
4.1.3	Spracovanie dát v rámci programu	31
4.1.4	Vizualizáciu a výsledky dát	31
4.2	Python a NetworkX	32
4.2.1	Vizuálna stránka programu	32
4.2.2	Používateľská prívetivosť	33
4.2.3	Spracovanie dát v rámci programu	33
4.2.4	Vizualizácia a výsledky dát	37
4.3	Gephi	40
4.3.1	Vizuálna stránka programu	40
4.3.2	Používateľská prívetivosť	41
4.3.3	Vizualizácia a výsledky dát	45
4.4	NetMiner 4	49
4.4.1	Vizuálna stránka programu	49
4.4.2	Používateľská prívetivosť	50
4.4.3	Vizualizácia a výsledky dát	51
	Záver	56
	Zoznam použitej literatúry	58

Úvod

Pojem sociálne siete sú v dnešnej spoločnosti ovládanej internetom veľmi dobre známy a zaužívaný pojem. Internetová doba nás zaplavila akonáhle to bolo možné, a takmer každý jedinec na planéte žije svoj život neodmysliteľne spolu so svojou online komunitou, v ktorej si vytvoril jedinečnú identitu, známosti či konexie.

Značnú zásluhu na ešte väčšej inklinácii k online svetu mala aj nedávna pandemická situácia a práve jej zásluhou sa spoločnosť stala takmer závislou na sociálnych sieťach a online svete. Zatiaľ čo niektorí ľudia túto dobu označujú za negatívne obdobie vplývajúce záporne na fyzickú ako aj psychickú stránku, pre sociálne siete znamená táto situácia veľký rozmach.

Vďaka digitalizácii a komunikácii s okolitým svetom prostredníctvom online platformy získavame stále viac informácií, ktoré nám pomáhajú dopĺňať často nekompletné informácie, alebo ich schopnosť aktualizovať.

Takto získané informácie napomáhajú analýze sociálnych sietí k zisťovaniu informácií o rozmanitých prepojeniach, formovaní podskupín a ich prepojení s okolím. Analýza sociálnych sietí poskytuje príležitosť rozboru sociálnych sietí, ktoré sú v súčasnosti často používané. Jej aplikácia je možná aj v odvetviach a odboroch, kde by sme využitie týchto poznatkov na prvý pohľad neočakávali.

Táto záverečná práca má za záujem uviesť čitateľov do jedného z mnohých spôsobov, aktívne vizualizovať dáta z datasetu vo vybraných nástrojoch. Má priblížiť základné poznatky pre vizualizáciu sociálnej siete ako aj prácu s nástrojmi, ktoré sa v danej záverečnej práci nachádzajú.

Zároveň sa snaží povzbudiť budúce generácie, ktoré sa obávajú, že ich aktuálne vedomosti nie sú dostačujúce pre praktizovanie metódy, o využitie dostupných prostriedkov vykresľovania grafov a ich ľahké použitie.

Prvá kapitola je komplexný prehľad analýzy sociálnych sietí (SNA), ktorá zahŕňa jej históriu, koncepty, aplikácie a nástroje. Začína sa sledovaním koreňov SNA až do začiatku 20. storočia, pričom zdôrazňuje prínosy sociológov ako Auguste Comte, Émile Durkheim a Georg Simmel. Ďalej sa zaoberá kľúčovými konceptmi SNA, vrátane aktérov, vzťahov, diád, triád, podskupín, skupín a sietí.

Kapitola zdôrazňuje dôležitosť SNA pri pochopení štruktúry a dynamiky sociálnych sietí, ktoré sa pohybujú od online komún po organizačné hierarchie. Zvýrazňuje jeho aplikácie v rôznych oblastiach, ako je sociológia, antropológia, psychológia, informatika a podnikanie.

Na ilustráciu praktickej aplikácie SNA text uvádza príklad analýzy sociálnej siete Star Wars pomocou softvéru Gephi. Tento príklad demonštruje, ako SNA dokáže identifikovať ústredné postavy, kliky a vzorce interakcie v sieti.

Teoretická čas sa uzatvára diskusiou o rôznych nástrojoch a softvérových balíkoch dostupných pre SNA, ako napríklad EgoNet, Gephi, NetMiner, NetworkX, SocNetV a R. Je potrebné pochopiť dôležitosť výberu vhodného nástroja na základe konkrétnej výskumnej otázky a súboru údajov.

Napokon sa vo výsledkoch práce zameriame na vytvorené zobrazenia a reporty a zhodnotíme určité aspekty každého z nástrojov, ako napríklad vizuálnu stránku, nájdenie správneho rozloženia dát či spracovanie väčšieho obnosu dát a ďalšie.

Cieľom je komplexné pochopenie funkcionality, možností a limitov rôznych nástrojov pre analýzu sociálnych sietí. Práca sa preto zameriava na celý proces od zadávania a tvorby dát až po vymodelovanie sociogramu a generovanie reportov.

V jednotlivých kapitolách budú predstavené a analyzované vybrané aplikácie pre analýzu sociálnych sietí. Budú definované spôsoby zadávania dát do programov, metódy vykreslenia sociogramu a tvorby reportov.

V časti metodika práce a metódy skúmania budú podrobne popísané postupy práce s jednotlivými nástrojmi. Budú definované úkony potrebné pre vytvorenie grafického zobrazenia a reportu.

Záverečná časť práce sa zameria na vyhodnotenie vytvorených zobrazení a reportov. Budú porovnané vizuálne aspekty, možnosti rozloženia dát, spracovanie veľkých objemov dát a ďalšie relevantné vlastnosti analyzovaných nástrojov.

1 Súčasný stav riešenej problematiky doma a v zahraničí

Kapitola analýza sociálnych sietí je venovaná pojmom, ktoré sa často vyskytujú v sociálnych sieťach. Spomínané pojmy sa pokúsime priblížiť aj pre nezainteresovaného čitateľa, vysvetlíme tvorbu grafov a ich rôzne podoby vyobrazenia. Taktiež si priblížime cestu tejto analýzy rokmi a jej uplatnenie.

1.1 Počiatky analýzy sociálnych sietí

Korene analýzy sociálnych sietí nesiahajú do tak hlboko do minulosti, ako pri iných pozorovacích praktikách. Jedná sa relatívne o novú možnosť skúmania. Jej využitie je prevažne v sociologickej sfére, no našla aj svoje uplatnenie v iných odvetviach a oboroch.

Počas minulého storočia sa sociológii venovalo výrazné množstvo odborníkov. Z ich radov niekoľko pochopilo potrebu nazrieť na spoločnosť z iného uhla pohľadu a vedomky formovať základy práve k spomínanej analýze sociálnych sietí.

V prvých počiatkoch analýzy sociálnych sietí sa vraciame až na začiatok 20. storočia, kde niekoľko sociológov, medzi ktorých patril aj Auguste Comte (celým menom Isidore Auguste Marie François Xavier Comte), Émile Durkheim, či Georg Simmel, položilo počiatočnú bázu pre analýzu sociálnych sietí.

Postupom času a aj príchodom nových odborníkov sa v 30. rokoch minulého storočia objavujú dve najhlavnejšie osobnosti, ktoré sa zaslúžili o vytvorenie sociometrie, inak nazývanej aj analýzy sociálnych sietí. Jacoba Levy Moreno môžeme prirovnať ku hnacej sile, ktorá sa zaslúžila o vývoj sociometrie. Pôvodom z Viedne, tento mladý sociálny vedec, hoci pôvodne študent medicíny, aplikoval sociometriu a jej parametre na štruktúru divadla v jednej zo svojich publikácií (Freeman, 2004).

Počas svojho účinkovania v divadle v New Yorku spoznal ďalšieho človeka, Helenu Hall Jennings, s ktorou zásluhy za vybudovanie základov analýzy sociálnych sietí. Vďaka nej, vyštudovanej psychiatricky, spoznal Gardenera Murphyho, talentovaného psychológa, ktorý sa taktiež zaujímal a špecializoval aj vo výskumných metódach a štatistike. Taktiež Helen Jennings pomohla Jacobovi Morenovi pri realizácii sociálnych štúdií. Moreno ako prvý vo svojej publikácii využil pojem “sieť” pre význam, v akom ho dnes používame aj my v tejto diplomovej práci (Freeman, 2004).

Po úspechu Morena a Jennings prichádza v povojnovom období na scénu mnoho osobností, ktoré sa začali venovať sociálnym sieťam. Medzi tieto figúry patril aj John Arundel Barnes, profesor sociológie na Cambridge University. Barnes, spolu s ďalšími nadšencami

pre sociológiu prispeli veľkou mierou ku pokroku analýzy sociálnych sietí do pôvodu, v akom ju dnes poznáme my.

Napriek svojej mladosti sa analýza sociálnych sietí rýchlo adaptovala na dnešnú dobu a získala svoje postavenie v rôznych odvetviach.

1.2 Sociálne siete

Mnohí z nás už počuli o sociálnych sieťach resp. ich online verziách, ako sú Facebook alebo Twitter, či ich tradičných verziách, akými sú sociálne vzťahy medzi členmi susedskej komunity. Taktiež môžeme sociálne siete pozorovať vo firmách či iných zamestnaniach, ako externé prepojenia firmy s okolím tak aj interné v rámci firmy, ako prepojenie rôznych oddelení či priamo zamestnancov.

Sociálne siete sa skladajú z objektov, ktorými sú vo väčšine prípadov jednotlivci či organizácie a vzťahov medzi týmito objektami. Rôzne typy vzťahov, ktoré objekty nadväzujú, taktiež nazývané aj hrany, či interakcie, sú zodpovedné za rôzne typy sietí a to aj v prípadoch, keď je skúmaný rovnaký súbor objektov.

Existuje mnoho typov sociálnych sietí, ktoré sa líšia podľa ich štruktúry, charakteru vzťahov a účelu ich vytvorenia. Niektoré siete sa zameriavajú na komunikáciu a zdieľanie informácií medzi jednotlivcami, ako je to napríklad pri sociálnych sieťach Facebook alebo LinkedIn. Iné siete sa zaoberajú prepojeniami medzi organizáciami, ako je to v prípade obchodných sietí alebo sietí partnerských vzťahov.

Okrem toho môžeme pozorovať sociálne siete aj v rámci rôznych komúní alebo skupín, kde sa jednotlivci spojili na základe spoločných záujmov, záujmov alebo cieľov. Tieto siete často slúžia ako platforma na zdieľanie skúseností, spoluprácu alebo podporu medzi členmi skupiny.

1.3 Pojmy sociálnych sietí

V rámci sociálnych sietí existuje niekoľko základných a hlavných pojmov, ktoré neodmysliteľnou časťou pre definovanie a porozumenie grafom vytvorených práve pomocou analýzy sociálnych sietí. Jedná sa o pojmy ako: účastník, vzťahová väzba, dyáda, triáda, podskupina, skupina vzťah a sieť. Pre lepšie porozumenie, a tým aj ľahšie identifikovanie grafov si ich v nasledujúcich podkapitolách priblížime.

1.3.1 Účastník

Účastník, často nazývaný aj aktér, je typ sociálnej entity vyskytujúci sa v každej sociálnej sieti. Parametre účastníka môžu byť rôzne: diskkrétne individuálne, podniky alebo kolektívne sociálne jednotky. Za príklady môžeme uviesť ľudí v skupine, oddelenia v rámci firmy alebo národné štáty vo svetovom systéme. V sociálnej sieti sa zväčša zaoberáme účastníkmi, ktorých parametre sú totožné a ich sociálnu sieť nazývame jednorežimové siete (Wasserman, 1994).

1.3.2 Vzťahová väzba

Ako sme už vyššie spomenuli, účastníci sú navzájom prepojení sociálnymi väzbami. Ich hlavnou charakteristiku je vytváranie vzťahov medzi účastníkmi. Vzhľadom na situáciu, vyplývajúcu z údajov a ich parametrov, môžeme definovať niekoľko typov vzťahov ako napríklad:

- Posudzovanie jednej osoby druhou. Do tejto kategórie môžeme zaradiť napríklad priateľstvo, náklonnosť alebo rešpekt.
- Prevody materiálnych zdrojov, teda obchodné transakcie, požičiavanie peňazí či vecí
- Združenie alebo príslušnosť, teda účasť na sociálnom podujatí alebo klubu
- Behaviorálna interakcia,
- Fyzické spojenie (cesta, rieka alebo most spájajúci dva body)
- Formálne vzťahy, akými je dodržiavanie (napríklad autorita)
- Biologický vzťah a mnoho ďalších (Wasserman, 1994).

1.3.3 Vzťah

Súhrn väzieb určitého druhu medzi členmi sociálnej siete sa nazýva vzťah. Jedná sa napríklad o vzťahy medzi skupinami tvoriacimi sa v škole, alebo aj formálne vzťahy diplomatických väzieb vo svete. Pre akúkoľvek skupinu aktérov môžeme zaznamenať niekoľko rôznych vzťahov, ktoré sme si spomenuli pri vzťahovej väzbe. Vzťahy samy osebe existujú len medzi konkrétnymi dvojicami aktérov (Wasserman, 1994).

1.3.4 Dyáda

Za najzákladnejšiu úroveň prepojenia alebo vzťahu medzi účastníkmi patrí dyáda, vytvárajúca spojenie medzi dvoma aktérmi. Tento vzťah je neodmysliteľnou časťou akéhokoľvek spojenia účastníkov. Mnoho druhov sieťových analýz sa zaoberá práve

porozumením väzieb medzi dvojicami a berú dyádu ako jednotka analýzy. Dyáda pozostáva z dvojice účastníkov a ich (možnej) väzby. Dyadické analýzy sa zameriavajú na vlastnosti vzťahov medzi párom účastníkov, napríklad či sú väzby opätované alebo nie, alebo či špecifické typy viacerých vzťahov majú tendenciu sa vyskytovať spoločne (Wasserman, 1994).

1.3.5 *Triáda*

Triádu môžeme definovať ako vzťahy medzi väčšími podskupinami účastníkov sietí. Veľa podstatných sociálnych sietí sa zameriava na triády: podskupinu, ktorú tvoria traja účastníci a ich (možné) väzby. Analýza sa prenáša z dvojice na jedinca a naspäť na trojicu účastníkov a ich prepojenie, ktoré opísal teoretik Simmel v roku 1908 (Wasserman, 1994).

1.3.6 *Podskupina*

Z vyššie spomínaných kapitol vieme, že dyády sú dvojice účastníkov prepojené pomocou väzieb a triády sú trojice prepojené väzbami. Z toho vyplýva, že môžeme v rámci sietí možno definovať podskupiny účastníkov a vytvorené väzby medzi nimi. Lokalizácia a štúdium podskupín pomocou špecifických kritérií je jedným z dôležitých problémov pri analýze sociálnych sietí (Wasserman, 1994).

1.3.7 *Skupina*

Sieťová analýza nepozoruje iba súbory dyád, triád alebo podskupín. Do veľkej miery sila sieťovej analýzy spočíva v schopnosti modelovať vzťahy medzi sieťovými účastníkmi. Tieto systémy sú zložené z väzieb medzi členmi (viac či menej obmedzenej) skupiny.

Sociálni vedci poskytli široký rozsah definície pojmu skupina. Pre naše účely môžeme definovať skupinu ako súbor všetkých účastníkov, na ktorých sa majú merať väzby. Musíme byť schopní získať teoretické, empirické alebo koncepčné kritériá o účastníkoch, ktorí k sebe v skupine patria a tvoria viac či menej ohraničené súbory. Skupina teda pozostáva z konečnej množiny účastníkov, ktorí sa považujú za konečný súbor jednotlivcov.

Obmedzenie na konečnú množinu alebo množiny účastníkov je analytickou požiadavkou. Modelovanie konečných skupín predstavuje niekoľko ďalších problematík v sieťovej analýze vrátane špecifikácie hranice siete, vzorkovanie a definícia skupiny (Wasserman, 1994).

1.3.8 Sociálna sieť

Po definovaní vzťahov vyskytujúcich sa v rámci sociálnej siete, môže teraz poskytnúť jasnejšiu definíciu sociálnej siete. Sociálna sieť pozostáva z konečnej množiny alebo množín účastníkov a definovaného vzťahu alebo vzťahov medzi nimi (Wasserman, 1994).

1.4 Charakteristika analýzy sociálnych sietí

V posledných rokoch sa postupnou modernizáciou a nárastom popularity online sociálnych sietí dostávajú do popredia nástroje ako možno analyzovať tieto siete. Jeden z najkomplexnejších nástrojov ako skúmať takéto siete sa nazýva analýza Sociálnych sietí. Týmto nástrojom je možno skúmať mnoho aspektov týchto sietí, či už sa jedná o rôzne formy a zmeny interakcií a zmeny správania.

Zo získaných dát je možné skúmať modeli správania, hľadať prepojenia jednotlivých objektov či skupín objektov, alebo sledovať korelácie správania objektov, ktoré by bez analýzy nebolo možné dokázať. Konkrétnym príkladom je Facebook algoritmus, ktorý používa nástroje na analyzovanie siete používateľov pre následne odporúčanie či už priateľov ktorých by používateľ mohol poznať, skupiny z ktorými zdieľa rovnaké záľuby, alebo zobrazovanie aktuálne relevantných reklám pre používateľa.

Až keď sa na svet začneme pozeráť ako na spleť rôznych navzájom prepletených sietí, ktoré prepájajú a prenášajú priateľstvá, informácie, peniaze či moc, začneme si uvedomovať, ako analýza sociálnych sietí môže viesť k novým objavom v oblastiach ako sú kultúra, história, či politika. Analýzu sociálnych sietí je možno použiť v obrovskom počte oblastí a preto je do budúcnosti dôležité venovať tejto oblasti pozornosť a rozvíjať naše vedomosti ohľadom nej.

Analýza sociálnych sietí, je podľa viacerých názorových prúdov definovateľná ako určitá teória, a zároveň nástroj na vytváranie relevantných analýz, súvisiacich s týmto javom. V skutočnosti sa však jedná o proces, využívajúci siete a teórie grafov, pre skúmanie tejto formy interakcie v rámci sociálnych sietí (Otte, 2002).

Pri analýze sociálnych sietí sa primárne koncentrujeme na sledovanie priamych spojení, medzi jednotlivými aktérmi týchto interakcií, využívajúcich sociálne siete. Charakteristika a typológia týchto aktérov ostáva ako druhotný parameter analýzy. Hoci je táto charakteristika a typológia aktérov pri analýze druhotným typom informácií, je neoddeliteľnou súčasťou tejto analýzy, pretože napomáha k získaniu uceleného obrazu, čo vo výraznej miere prispieva k porozumeniu sociálnych sietí a je taktiež potrebné sledovať správanie sa individuálnych aktérov, ich spojenia a vzájomné interakcie.

Analýza sociálnych sietí sa časom rozšírila aj do iných modelov. V nich sa aktívne zúčastňuje pri sledovaní a analyzovaní jednotlivých prepojení medzi uzlami.

Praktické využitie tohto sledovania je možné pozorovať napríklad v týchto oblastiach:

- Telekomunikácii, teda sledovanie komunikačnej mapy medzi účastníkmi,
- Dynamiku prenosu civilizačných chorôb v skupinách a podskupinách,
- Kriminálnou aktivitou a podobne.

Táto forma analýzy má vďaka svojej komplexnosti svoje využitie aj v rámci iných vedných odborov, dokonca aj v biológii. V rámci tejto vedeckej sféry, má svoje využitie nesporne napríklad v antropológii (Tringali a kol., 2020).

Analýza sociálnych sietí je vďaka svojej univerzálnosti využiteľná nielen na sledovanie konexií medzi rôznymi subjektami, no môžeme ňou sledovať aj efektivitu dostupných a využívaných zdrojov. Svoje uplatnenie si nájde aj pri skúmaní pripravenosti zamestnancov na núdzové situácie, stres, časový aj mentálny nátlak, s akým sa môžeme stretnúť v zdravotníckych zariadeniach či nemocniciach na pohotovosti, alebo vo vypätých situáciách v rámci represívnych zložiek, kde efektívne využitie dostupných zdrojov zohráva kľúčovú rolu pri zaistení úspechu. (Harris, 2007).

1.5 Príklad sociálnej siete: Analýza a Interpretácia

Pre lepšie porozumenie analýzy sociálnych sietí si uvedieme príklad. Po tomto príklade by mala vizualizácia a analýza sociálnych sietí byť zrozumiteľná aj pre začiatočníka a laika.

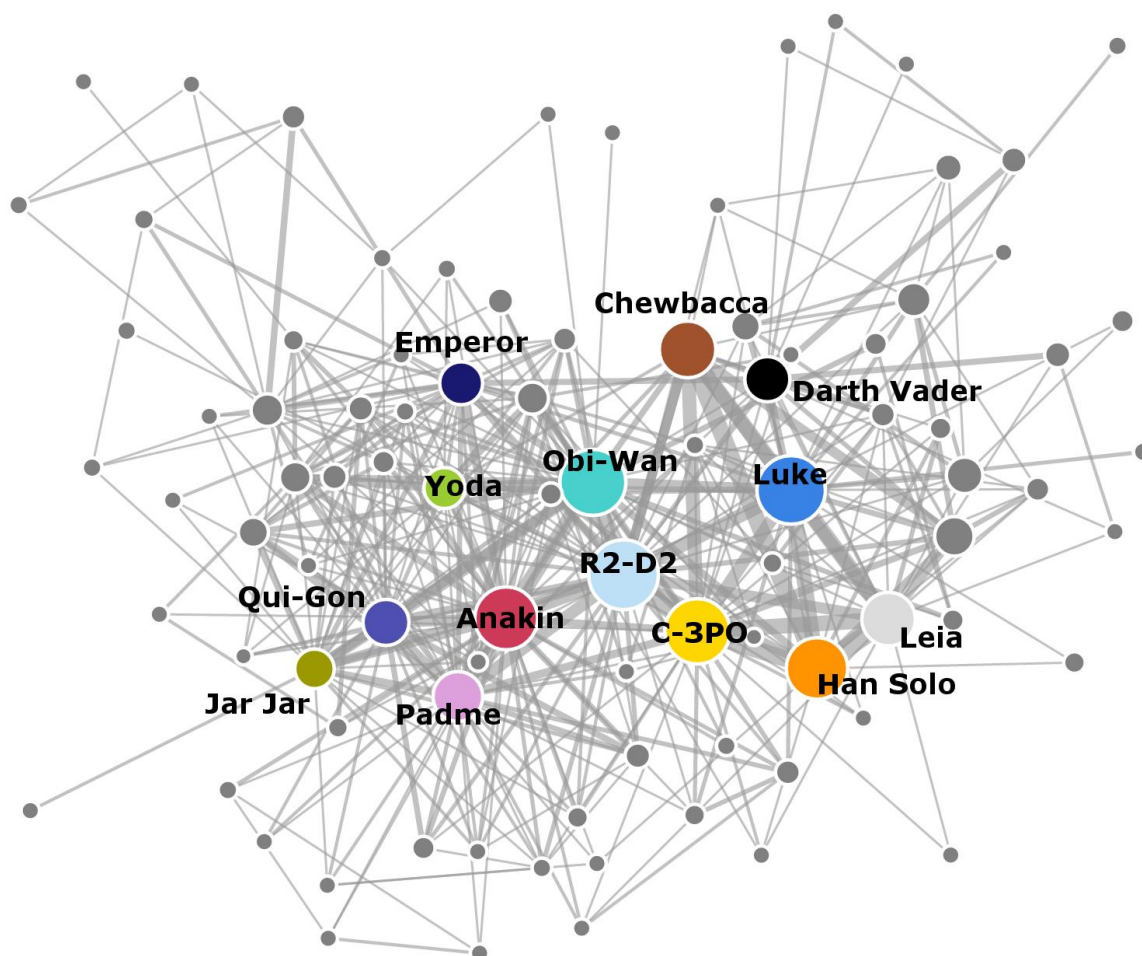
V dnešnej technologickej dobe sa na internete nachádza nespočetne veľa sociálnych sietí, ich spracovaní aj interpretácii. Pre aktuálnu záverečnú prácu autorku zaujala analýza sociálnej siete Star Wars, na ktorej bude demonštrovať interpretáciu a dôležitosť tvorby sociálnych sietí.

Ako sa definovalo vo vyššie spomenutých podkapitolách, na vytvorenie sociálnej siete je potrebné mať aspoň 2 body (takiež nazývané uzly alebo účastníci) a vzťah medzi nimi, teda hranu. Uzlom sociálnej siete môže byť takmer všetko, nie len ľudia alebo filmové postavy, ale aj mestá, krajiny, knihy a mnoho ďalších. Hrana vo vizualizácii sociálnej siete nemusí indikovať len vzťahy ako osobné vzťahy. Tieto vzťahy môžu byť založené napríklad aj na interakciách, spoločných záujmoch, alebo jednoducho na známostiach.

Ako je možné vidieť, **obrázok 1** vizualizuje sociálnu sieť filmov Star Wars. Uzly boli definované ako postavy z filmov, zatiaľ čo ich hrany reprezentujú množstvo konverzácií v roznych scénach. Najzákladnejšou vizualizáciou sú uzly s rovnakou veľkosťou a hrany

s rovnakou hrúbkou. Pri vizualizácii socialnej siete môžeme tieto dva aspekty modifikovať podľa potreby interpretácie sociálnej siete.

Obrázok 1: Star Wars sociálna sieť všetkých 6 filmov



Zdroj: <https://evelinag.com/blog/2015/12-15-star-wars-social-network/>

V **obrázku 1** je teda definovaná veľkosť uzlu podľa počtu scén, v ktorých sa daná postava objavila. Farba v tomto prípade nemá žiadnu definíciu, v iných vizualizáciách však pomocou farieb možno identifikovať napríklad skupiny, triedy v škole a pod. Taktiež možno pozorovať, že vo vizualizácii sociálnej siete sa nachádzajú aj uzly bez osobitej farby a v menšej veľkosti. Tieto postavy teda v danej vizualizácii neboli významné natoľko, aby boli špecificky definované, avšak vo vizualizácii ostali kvôli svojim vzťahom s ostatnými uzlami.

V danej vizualizácii socialnej siete možno hrany na **obrázku 1** interpretovať ako počet interakcií, ktoré dané body mali medzi sebou. Možno teda poznamenať, že tieto hrany sú vážené, resp. ich hrúbka je závislá na počte interakcií. Hrany môžeme navyše rozdeliť vo vizualizáciách aj na smerované a nesmerované. V danej vizualizácii sú hrany nesmerované, avšak napríklad pokiaľ by sme chceli pozorovať, kto začal konverzáciu, mohli by sme dané hrany smerovať z iniciátora konverzácie na druhý uzol.

Teda keď sa pozrieme na postavy Luke a Leia, môžeme konštatovať, že tieto dve postavy mali medzi sebou veľa scén s interakciami a taktiež sa objavili v mnohých scénach. Práve vďaka tejto vizualizácii teda môžeme pozorovať, ktoré z postáv z filmov Star Wars sú hlavné postavy a ktoré naopak vedľajšie.

1.6 Teória grafov

Graf je abstraktný matematický nástroj, ktorý slúži na reprezentáciu vzájomných vzťahov medzi objektmi. Skladá sa z vrcholov a hrán, pričom vrcholy môžu byť spojené pomocou hrán. Graf môže mať rôzne vlastnosti a využitia v rôznych oblastiach, vrátane informatiky, sociálnej siete a dopravnej siete.

Vrcholy a hrany sú základné stavebné bloky grafov. Vrcholy predstavujú objekty alebo prvky, ktoré sú vzájomne prepojené v grafe. Hrany sú spojenia medzi vrcholmi, ktoré môžu mať rôzne vlastnosti, ako napríklad váhu alebo smer. Vrcholom a hranám je možné pridať rôzne atribúty a využiť ich na reprezentáciu špecifických vlastností objektov a vzťahov v grafe.

Orientované grafy sú grafy, v ktorých hrany majú zadaný smer. To znamená, že pripojenie medzi dvoma vrcholmi je jednosmerné. Naopak, neorientované grafy majú hrany, ktoré nemajú definovaný smer. Vrcholy sú spojené bez ohľadu na smer, čo umožňuje vzájomné spojenie medzi vrcholmi v oboch smeroch. Orientované a neorientované grafy sa využívajú v závislosti od konkrétnych potrieb a vlastností problému, ktorý sa s nimi rieši.

Jednoduchý graf je graf, v ktorom medzi dvoma vrcholmi existuje najviac jedna hrana a žiadna hrana nie je pripojená k vrcholu samému sebe. Tým sa zabezpečuje, že v grafe nie sú žiadne násobné hrany ani slučky. Multigrafy, na druhej strane, môžu mať medzi dvoma vrcholmi viacero hrán a umožňujú existenciu slučiek. Jednoduché grafy majú jednoduchšiu štruktúru a sú často používané pri modelovaní problémov, ktoré nevyžadujú viackrát prepojené vrcholy.

1.6.1 Aplikácie teórie grafov

Plánovanie dopravy je jednou z aplikácií teórie grafov, kde sa využíva grafová reprezentácia pre modelovanie cestovných sietí. Pomocou vrcholov a hrán je možné zachytiť jednotlivé dopravné trasy a vytvárať tak plány pre optimalizáciu prepravy tovaru či osôb. S využitím algoritmov teórie grafov možno napríklad vypočítať najkratšie cesty, minimalizovať prekážky na cestách, sledovať nákladné vozidlá a synchronizovať ich pohyb. Plánovanie dopravy sa aplikuje v rôznych odvetviach, ako je logistika, verejná doprava a šírenie

sietí. Táto aplikácia je nevyhnutná pre efektívne riadenie dopravných tokov a zabezpečenie optimálneho využitia dostupných zdrojov.

Optimalizácia sietí je ďalšou významnou aplikáciou teórie grafov, ktorá sa zaoberá hľadaním najlepších riešení v rámci rôznych typov sietí. Tieto siete môžu zahŕňať napríklad telekomunikačné siete, energetické siete, dodávateľské siete alebo internetové siete. Grafy sa využívajú na reprezentáciu spojení, vzťahov a tokov v týchto sieťach. Optimalizačné algoritmy teórie grafov slúžia k minimalizácii nákladov, maximalizácii výkonu a efektívnemu využitiu zdrojov v sieti. Takáto aplikácia je neoceniteľná pre poskytnutie spoľahlivých, rýchlych a kvalitných služieb v sieťovom prostredí a prispieva k celkovému zlepšeniu efektívnosti sietí a ich správy.

Sociálne siete sú obľúbeným a dôležitým predmetom výskumu v kontexte teórie grafov. Tieto siete predstavujú vzťahy a interakcie medzi jednotlivcami alebo organizáciami. Grafy sa využívajú na modelovanie rozsiahlych sociálnych sietí a analýzu ich štruktúry. Pomocou teórie grafov je možné identifikovať centrálnych aktérov v sieti, skúmať šírenie informácií, predikovať trendy a vyvodiť dôsledky sociálnych vzťahov. Sociálne siete majú veľký vplyv v oblasti marketingu, sociológie, psychológie a informačnej vedy. Jej štúdium a analýza pomáha lepšie porozumieť vzťahom medzi jednotlivcami v spoločnosti a zlepšuje efektívnosť komunikácie a interakcie v sociálnom prostredí.

1.7 Sociogram

Sociogram je grafické znázornenie vzťahov medzi jednotlivými bodmi (uzlami) v rámci sociálnej siete. Tieto vzťahy sa zvyčajne získavajú prostredníctvom prieskumov, ako sú rozhovory, dotazníky alebo iné formy zberu dát. Na základe týchto údajov sa vytvára matica, ktorá je potom vizualizovaná v podobe sociogramu.

Pri vytváraní sociogramu môžeme identifikovať rôzne typy vzťahov medzi bodmi, ktoré definujú vzájomné interakcie v sieti. Niektoré z týchto typov vzťahov môžu zahŕňať:

- Priateľský vzťah - obojsmerný pozitívny vzťah medzi dvoma bodmi
- Nepriateľský vzťah - obojsmerný negatívny vzťah medzi dvoma bodmi
- Hviezda, vodca – do bodu prichádzajú pozitívne vzťahy
- “Čierna ovca” - dostáva predovšetkým negatívne vzťahy
- “Ježko” – body, z ktorého vychádzajú negatívne vzťahy
- Izolát – bod, ktorý neobsahuje žiadne vzťahy

Sociogramy poskytujú užitočný nástroj na vizuálne porozumenie vzťahov a štruktúry v sociálnych sieťach. Pomáhajú identifikovať vzory, zoskupenia a centrálné postavy, čo umožňuje lepšie porozumieť dynamike a fungovaniu komunity.

1.8 Vplyv nástrojov na analýzu a vizualizáciu sietí

Ako bolo už spomenuté, analýza a vizualizácia sociálnych sietí zohráva kľúčovú úlohu v mnohých odboroch, zamierajúcich sa nie len na informatické a technologické odvetvie, ale aj na sociológiu, biomedicínu, či zabezpečenie optimálneho využitia zdrojov. Nástroje disponujú mnohými vlastnosťami a funkciami, ktoré môžu modifikovať interpretáciu a spôsob skúmania siete.

V dnešnej technologickej dobe a možnosťami, akými technický boom priniesol poznáme nespočetne nástrojov, ktoré skúmajú a vizualizujú sociálne siete.

1.8.1 EgoNet

EgoNet je nástroj navrhnutý na získavanie a analýzu egocentrických sietí. Pomáha vytvárať dotazníky, zhromažďovať dáta a poskytuje všeobecné globálne sieťové metriky. Okrem toho je možné použiť dátové matice na ďalšiu rozsiahlejšiu analýzu v iných softvéroch. Po ukončení prieskumu je sieť okamžite vizualizovaná, čo uľahčuje ďalšie otázky týkajúce sa siete. Ďalšie možnosti zahŕňajú zlučovanie osobných sietí do jednej celkovej siete.

1.8.2 Gephi

Gephi je interaktívna platforma na vizualizáciu a prieskum všetkých druhov sietí a komplexných systémov, vrátane dynamických a hierarchických grafov. Je kompatibilný so systémami Windows, Linux a Mac OS X, a je dostupný ako open-source software zdarma. Gephi je nástroj pre tých, ktorí sa potrebujú zorientovať v grafoch a porozumieť im. Jeho používatelia interagujú s reprezentáciou dát podobne ako v programe Photoshop, ale na manipuláciu so štruktúrami, tvarmi a farbami s cieľom odhaliť skryté vlastnosti.

1.8.3 NetMiner

NetMiner je sofistikovaný nástroj slúžiaci na exploratívnu analýzu a vizualizáciu rozsiahlych sieťových dát na základe metód analýzy sociálnych sietí (SNA). Jeho funkcionality zahŕňa interný skriptovací engine, ktorý je postavený na jazyku Python a disponuje automatickým generátorom skriptov. Licencie na NetMiner 4 sú poskytované študentom a učiteľom na účely vzdelávania.

1.8.4 *NetworkX*

NetworkX je softvérový balík pre jazyk Python, určený na vytváranie, manipuláciu a štúdium štruktúry, dynamiky a funkcií komplexných sietí. S NetworkX môžete načítať a ukladať siete v rôznych formátoch dát, generovať rôzne typy náhodných a klasických sietí, analyzovať štruktúru siete, budovať modely siete, navrhovať nové algoritmy a vykresľovať siete. Tento nástroj je využívaný napríklad matematikmi, fyzikmi, biológmi, informatikmi a sociálnymi vedcami.

1.8.5 *SocNetV (Social Networks Visualizer)*

SocNetV je multiplatformový nástroj určený na analýzu a vizualizáciu sociálnych sietí, ktorý sa vyznačuje používateľsky prívetivým prostredím. Tento nástroj umožňuje vytváranie sietí (matematických grafov) na virtuálnom plátne alebo načítavanie sietí vo formátoch ako GraphML, GraphViz, Adjacency, Pajek, UCINET a ďalšie. Okrem toho SocNetV poskytuje možnosť modifikovať sociálne siete, analyzovať ich sociálne a matematické vlastnosti a aplikovať rôzne vizualizačné rozloženia.

1.8.6 *R*

R je všeobecný nástroj určený na analýzu údajov, no existuje niekoľko knižníc špeciálne navrhnutých pre analýzu sociálnych sietí. Medzi ne patria Degreenet, RSeina, PAFit, Igraph, SNA network, tnet, ergm, Bergm, hergm, latentnet a networksis. Každá z týchto knižníc poskytuje špecializovanú funkcionálnu a pre používateľov oboznámených s R predstavujú bohatý zdroj nástrojov.

Toto je len niekoľko z mnohých dostupných nástrojov, ktoré sú vhodné na skúmanie a vizualizáciu sociálnych sietí. Práve zo spomenutých nástrojov si autorka tejto závernej práce zvolil pre podrobnejšie oboznámenie sa.

2 Cieľ práce

Ako sme naznačili v obsahu tejto záverečnej práce, hlavným cieľom práce je analyzovanie a následne porovnanie vizualizačných a performačných aspektov vybraných aplikácií pre modelovanie analýz sociálnych sietí. Práca sa zmeriava na internetovo dostupné aplikácie, ktoré umožňujú vizualizáciu a následnú analýzu sociálnych sietí. Pre lepšie porozumenie aplikáciám, ktoré budú definované v nižších kapitolách, sa pozrieme ne od samostatného začiatku, teda od zadávania či vytvárania dát až po vymodelovanie sociogramu.

V nižšie uvedených kapitolách sa oboznámime s funkcionalitou každého nástroja, definujeme spôsoby zadávania údajov do programov a metódy vykreslenia sociogramu. Taktiež sa pozrieme na tvorbu reportov, resp. správ ohľadom skúmaných dát. V časti metodika práce a metódy skúmania si definujeme každý z nástrojov na analýzu sociálnych sietí, ako v ňom pracovať a jednotlivé úkony, ktoré je potrebné realizovať pre vytvorenie grafického zobrazenia ako aj jeho reportu.

Napokon sa vo výsledkoch práce zameriame na vytvorené zobrazenia a reporty a zhodnotíme určité aspekty každého z nástrojov, ako napríklad vizuálnu stránku, nájdenie správneho rozloženia dát či spracovanie väčšieho obnosu dát a ďalšie.

3 Metodika práce a metódy skúmania

V teoretickej časti záverečnej práce sme si ozrejmili poznatok, že analýzu sociálnych sietí môžeme považovať za proces, vďaka ktorému je možné sledovať spojenia a podobnosti celku alebo aj selektívne vybraných subjektov, vzťahov, či iných parametrov v rámci siete.

Vytvoriť však prínosnú, zreteľne čitateľnú a pochopiteľnú sociálnu sieť je náročné hlavne vďaka faktu, že takéto siete neraz obsahujú veľké množstvo dát, ktoré nemusí byť pre sieť relevantné. Taktiež je potrebné v rámci nástrojov vybrať ten vhodný nástroj, či už pre vizualizáciu sociálnej siete priamo alebo hlbšej analýzy, ako napríklad zistenie najrýchlejšej trasy medzi uhlami či závislosť medzi uzlami.

3.1 Pracovný postup

Predtým, než sme sa dokázali posunúť ku spracovávaniu akýchkoľvek dát a ďalšej práce s nimi, bolo potrebné definovať, pomocou akého typu dát mienime naše pozorovanie uskutočniť. Pre dlhoročnú prítťažlivosť a zainteresovanie autorky do počítačových hier, či už ich hranie alebo sledovanie prostredníctvom videozáznamov, bola vybraná oblasť nášho skúmania a analýzy, ktorá je v dnešnej dobe dostatočne známa a populárna – streamovacia platforma Twitch.

Po úspešnom získaní dát na analýzu sme zhodnotili jednotlivé metódy a spôsoby akými je dáta možné ďalej spracovať. Bolo potrebné definovať veľkosť sledovaných dát, jeho prípravu do správneho formátu, ako aj správny postup pri aplikovaní dát do ilustračne znázorňujúcich grafických aplikácií.

Vďaka dôslednému zberu dát od autora datasetu, ktoré sa v záverečnej práci spracovali a analyzovali, nevznikol žiaden markantný problém, akými by boli napríklad chýbajúce dáta v rámci bodoch, ktoré nám interpretovali streamovacie kanály, či ich prepojenia, teda hrany spájajúce dva alebo viaceré body medzi sebou.

3.2 Spôsob získavania dát, ich modifikácia a zdroje

Analyzovateľné dáta pre záverečnú prácu boli získané z verejnej stránky Stanford University, vystupujúcou aj pod skratkou SNAP. Stránku vytvoril docent informatiky na Stanfordskej univerzite, Jure Leskovec. Stanfordská univerzita obsiahla na svojej stránke enormnú kolekciu datasetov, ktoré sú dostupné pre širokú verejnosť.

Na výber je hneď niekoľko možností, neorientované alebo orientované grafy, bipartitné grafy, multigrafy či dočasne zaradené body či hrany do grafov. Taktiež sa tu nachádzajú grafy, ktorých hrany majú definovanú váhu.

Stránka obsahuje hneď niekoľko typov datasetov, napríklad dáta zbierané zo stránky Facebook, Twitter, či Wikipédia, dokonca aj cestovateľskú sieť z amerických štátov. Zo škály dostupných možností sa zvolil dataset obsahujúci informácie ohľadom sociálnej siete Twitch streamerov. Jednalo sa o pomerne obsiahli dataset, aj napriek tomu, že na stránke Stanfordskej univerzity možno nájsť veľké množstvo datasetov, ktoré sú omnoho väčšie.

Za príklad môžeme zobrať zber dát pre dataset Pokec, ktorého množstvo bodov prerástlo cez 1 milión a jeho hrany boli ohodnotené na viac ako 30 miliónov záznamov, zatiaľ čo dataset, ktorý bol vybraný pre túto záverečnú prácu obsahoval pod 200 tisíc bodov a takmer 7 miliónov hrán, ktoré ich spájali.

Dáta obsiahnuté v datasete boli zozbierané pomocou rozhrania pre programovanie aplikácií, poznaného aj pod skratkou API (application programming interface). Body, taktiež inak nazývané aj ako profily alebo kanály, ktoré budeme v rámci našej záverečnej práce sledovať, obsahli v sebe hneď niekoľko charakteristík, akými sú napríklad ich Id, počet sledovateľov či jazyk, ktoré budeme pozorne sledovať. Ich hrany, ktoré znázorňujú spoločenské vzájomné prepojenie medzi nimi, sú neorientované, teda netvorí sa len jednosmerný vzťah, čo môžeme prirovnať ku vzťahom vytvárajúcich sa na stránke Facebooku.

Dáta v surovom stave sme stiahli zo stránky v .zip súbore. Po otvorení súboru súboru sme získali prístup ku dvom súborom a poznámkach od autora uvedených v textovom poli. V iných prípadoch sa jednalo o súbory typu .csv (Comma-separated Values), ktoré obsahovali potrebné dáta ku vizualizácii grafov a ich analyzovaniu.

Tieto dáta bolo potrebné extrahovať pomocou programu Microsoft Excel do tabuľky. Nakoľko sa jednalo o súbory .csv, tabuľka obsahovala všetky potrebné informácie len v jednom stĺpci. Práve kvôli tejto skutočnosti bolo potrebné tieto dáta rozdeliť tak, aby každá informácia oddelená znakom, alebo medzerou, bola obsiahnutá v samostatnom stĺpci.

Taktiež vznikol problém pri zobrazení hrán, nakoľko ich počet presiahol konečný počet Microsoft Excel hárku, preto bolo v prípade zobrazenia dát potrebné čiastočne transformovať daný súbor do textovej aplikácie, kde sa dáta zobrazili bez chýb.

Pri začiatkoch používania dát sme predpokladali, že veľkosť dát a ich charakteristiky či počet prepojení nebude stáť v ceste nášmu analyzovaniu tejto sociálnej siete. Postupom času a procesom spracovania dát do grafov, ktoré sme uviedli vo výsledkoch práce sa ukázalo,

že pre prehľadnejšie a podrobnejšie pozorovanie nami zvolených cieľov bude potrebné modifikovať základné tabuľky.

3.3 Metódy analýzy

Na analýzu dát boli použité nasledovné nástroje:

3.3.1 Gephi:

Gephi je komplexný open-source softvér pre vizualizáciu a analýzu sietí. Ponúka širokú škálu funkcií pre vizualizáciu a analýzu dát, vrátane vizualizácia grafov. Gephi umožňuje vytvárať rôzne typy vizualizácií grafov, vrátane 2D a 3D grafov, kruhových grafov a stromov. Vizualizácie grafov sa dajú upravovať pomocou rôznych farieb, veľkostí a tvarov uzlov

Gephi taktiež ponúka rôzne možnosti layoutu pre usporiadanie uzlov v grafe. Tieto layouts dokážu odhaliť skryté vzory a komunity v sieti. Pomocou filtrov je možné zameriavať sa na konkrétne časti siete. Môžeme filtrovať uzly podľa rôznych kritérií, ako je napríklad počet sledovateľov, jazyk alebo úroveň prepojenia.

Tento nástroj taktiež dokáže vypočítať rôzne štatistické ukazovatele sietí, ako je hustota, priemerný stupeň uzlov, priemer najkratšej cesty a modularita komunity. Tieto štatistiky pomáhajú kvantifikovať vlastnosti siete a pochopiť jej štruktúru.

Rovnako ako všetky aspekty, je dobré spomenúť aj pluginy. Gephi podporuje širokú škálu pluginov, ktoré rozširujú jeho funkcionality. Tieto pluginy umožňujú vykonávať pokročilejšie analýzy sietí, ako je napríklad analýza prúdenia informácií alebo detekcia komunít.

V našej práci sme využívali Gephi predovšetkým na vizualizáciu grafov sociálnych sietí na platforme Twitch. Vytvorili sme rôzne typy vizualizácií, ktoré nám pomohli identifikovať komunity streamerov, analyzovať vzťahy medzi kanálmi a skúmať vplyv jazykových bariér na interakciu.

3.3.2 NetMiner:

NetMiner je ďalší open-source softvér určený na analýzu sociálnych sietí. Zameriava sa najmä na kvantitatívnu analýzu sietí a ponúka nástroje pre identifikáciu komunít.

NetMiner dokáže automaticky identifikovať komunity v sieti pomocou rôznych algoritmov. Dokáže vypočítať rôzne miery centrality pre uzly v sieti. Centralita uzla určuje jeho dôležitosť v sieti v závislosti od jeho prepojenia s ostatnými uzlami.

Nástroj rovnako umožňuje analyzovať tok informácií v sieti. Môžeme sledovať, ako sa informácie šíria medzi rôznymi komunitami a uzlami. Podobne ako Gephi aj NetMiner ponúka rôzne štatistické analýzy sietí.

V našej práci budeme NetMiner používať predovšetkým na identifikáciu komunit streamerov na Twitchi a na analýzu centrálnych uzlov v sieti. Zisťovali sme, ktoré kanály sú najvplyvnejšie na základe ich prepojenia a počtu sledovateľov.

3.3.3 SocNetV:

SocNetV je jednoduchší nástroj pre vizualizáciu a analýzu sociálnych sietí vhodný pre začiatočníkov. Ponúka základné funkcie pre vizualizácia grafov, filtre, analýzu komunit, či štatistické ukazovatele.

SocNetV umožňuje vytvárať jednoduché vizualizácie grafov s uzlami a hranami. Umožňuje aj základné úpravy farieb a veľkostí uzlov. Rovnako tak poskytuje jednoduché filtrovanie uzlov podľa rôznych kritérií.

Nástroj taktiež ponúka základné nástroje na identifikáciu komunit v sieti a základné štatistické ukazovatele sietí ako hustota a priemerný stupeň uzlov.

V našej práci sme využívali SocNetV najmä na začiatkové skúmanie dát a jednoduchú vizualizáciu grafov. Pomocou SocNetV sme identifikovali základné štruktúry siete a získali prvý prehľad o prepojeniach medzi kanálmi na Twitchi.

3.3.4 NetworkX:

Na rozdiel od predchádzajúcich troch nástrojov, NetworkX nie je samostatný softvér, ale programovacia knižnica pre Python. Knižnica NetworkX umožňuje manipulovať s dátami o sieťach a vykonávať pokročilé analýzy sietí.

NetworkX umožňuje konštruovať rôzne typy grafov, a to ako smerované, tak nesmerované, vážené aj nevážené. Vďaka flexibilnej štruktúre knižnice je možné jednoducho implementovať aj zložitejšie modely sietí s dodatočnými atribútmi a vlastnosťami.

Taktiež ponúka širokú škálu algoritmov na komplexnú analýzu sietí. Medzi kľúčové funkcie patria hľadanie najkratšej cesty medzi uzlami, ktorá je dôležitá pre pochopenie prepojenia a toku informácií v sieti.

Ďalej umožňuje výpočet centrality uzlov, ktorá kvantifikuje ich dôležitosť a vplyv v sieti na základe rôznych metrík (stupeň uzla, medziľahlá centralita, vlastná centralita). Pomocou algoritmov detekcie komunit knižnica NetworkX dokáže identifikovať skupiny uzlov úzko prepojených medzi sebou, čo je užitočné pre pochopenie štruktúry a dynamiky siete.

NetworkX umožňuje výpočet rôznych štatistických ukazovateľov sietí, ktoré boli spomenuté aj pri predchádzajúcich nástrojoch. Tieto ukazovatele, ako napríklad hustota, priemerný stupeň uzla a priemerná dĺžka najkratšej cesty, poskytujú dôležité informácie o celkovej štruktúre a vlastnostiach siete.

Tento nástroj sa bezproblémovo integruje s ostatnými knižnicami Pythonu, ako sú NumPy a Matplotlib. Táto integrácia umožňuje pokročilé vedecké výpočty a komplexnú vizualizáciu dát. NumPy sa používa na efektívne spracovanie veľkých numerických dát, zatiaľ čo Matplotlib umožňuje vytvárať pútavé a informatívne vizualizácie grafov a štatistických ukazovateľov.

Vďaka kombinácii týchto funkcií a integrácii s inými nástrojmi sa NetworkX stáva výkonným nástrojom pre analýzu komplexných sietí v rôznych oblastiach, vrátane sociálnych vied, informatiky a systémovej biológie.

4 Výsledky práce a diskusia

V záverečnej práci sme sa venovali analýze sociálnych sietí a pre tento účel sme selektovali štyri nástroje, ktoré nám umožnia efektívne skúmať a analyzovať nadobudnuté dáta. Neupravené dáta, ktoré sme získali, sú často neštruktúrované a neprehľadné, a preto je nevyhnutné použiť nástroje, ktoré nám pomôžu ich spracovať a vizualizovať tak, aby sme z nich mohli vyvodiť relevantné závery.

Práve nižšie uvedené nástroje nám umožnia graficky reprezentovať sociálne siete pomocou sieťových grafov, čo nám uľahčí analýzu a porozumenie vzťahom medzi jednotlivými bodmi v sieti. Medzi vybrané nástroje patria:

- Python, predovšetkým jeho knižnica Networkx,
- Gephi,
- NetMiner 4,
- SocNetV.

Na základe stanovených cieľov tejto práce budeme každý z týchto nástrojov hodnotiť podľa niekoľkých aspektov, aby sme mohli porozumieť ako nástroju, tak aj jeho dizajnu a využitiu počas vizualizovania a analýzy sociálnych sietí.

Stanovené aspekty hodnotenia sme rozdelili na:

- Vizualnú stránku programu, teda subjektívny pohľad na daný nástroj pre vizualizáciu sociálnej siete,
- Používateľskú prívetivosť, taktiež subjektívny pohľad na prácu v programe,
- Spracovanie dát v rámci programu, respektíve vkladanie dát do programov
- Vizualizáciu a výsledky dát, kde sme subjektívne zhodnotili pohľad na finálnu vizualizáciu sociálnej siete v danom programe.

Takýmto štruktúrovaným hodnotením nástrojov budeme schopní poskytnúť objektívnu analýzu ich schopností a prispieť tak k lepšiemu porozumeniu ich využitia pri skúmaní sociálnych sietí.

Získané záznamy dát boli pre naše hodnotenie enormnej veľkosti, pozostávali z 168 114 bodov a vzajomným prepojením medzi nimi, ktoré sa šplhalo ku 7 miliónov (konkrétne 6 797 557 hrán). Hoci takto veľké záznamy sú veľmi významné a dajú sa medzi nimi dobre pozorovať prepojenia, pre grafické zobrazenie je takto veľký záznam neprehľadný a nie dobre čitateľný. Na základe tohto poznatku sme sa rozhodli pracovať v prvom nástroji s výrezmi datasetu.

4.1 SocNetV

Ako prvý z nástrojov sme siaholi po programe SocNetV. Social Networks Visualizer je multiplatformový, používateľsky prívetivý nástroj pre analýzu a vizualizáciu sociálnych sietí. Umožňuje vytvárať siete na virtuálnom plátne alebo načítať siete rôznych formátov (GraphML, GraphViz, Adjacency, Pajek, UCINET atď.). SocNetV tiež poskytuje možnosť modifikovať sociálne siete, analyzovať ich sociálne a matematické vlastnosti a aplikovať vizualizačné rozloženia.

4.1.1 Vizualná stránka programu

SocNetV sa vyznačuje vysokou úrovňou vizuálnej prezentácie, čo z neho robí atraktívny nástroj pre výskum a analýzu sociálnych sietí. Jeho vizualizačné schopnosti sú dôkladne premyslené a ponúkajú používateľom viacero možností vizualizácie.

Program SocNetV umožňuje používateľom zobrazit' sociálne siete pomocou rôznych typov grafických rozložení. Môžu si vybrať medzi rôznymi algoritmami pre rozmiestnenie uzlov v sieti, čo umožňuje identifikovať vzory a štruktúry v sieťach s väčšou presnosťou. Okrem toho je možné prispôbiť farby, štýly zobrazenia a veľkosť uzlov a hrán, čo zvyšuje prehľadnosť a estetickosť výslednej vizualizácie.

4.1.2 Používateľská prívetivosť

Jednou z výhod je aj interaktivita. Počas tvorby vizualizácie sociálnej siete sme sa efektívne oboznámili aj s modifikáciou výsledného zobrazenia priamo vo vizualizácii. Program ponúka jednoducho vyberať a presúvať uzly, meniť veľkosť a farbu uzlov, pridávať a mazať hrany, a ďalšie. Tento interaktívny prístup umožňuje rýchle a efektívne skúmanie štruktúry a vlastností sociálnych sietí.

Používatelia majú k dispozícii prehľadné nástroje na manipuláciu so sieťami a analýzu ich vlastností. Napríklad môžu jednoducho pridávať a mazať uzly a hrany, vykonávať rôzne analýzy sietí, a aplikovať vizualizačné efekty bez zbytočnej komplikácie.

SocNetV taktiež umožňuje načítavanie sietí vo viacerých formátoch dát, čo znamená, že používatelia nie sú obmedzení na konkrétne formáty. Môžu pracovať s dátami vo formátoch ako GraphML, GraphViz, Adjacency, Pajek, UCINET a ďalšie, čo zvyšuje flexibilitu a univerzálnosť nástroja.

4.1.3 Spracovanie dát v rámci programu

Vloženie dát do programu SocNetV je jednoduché a intuitívne. Pokiaľ používateľ nemá predpripravenú žiadne body a hrany, je možné ich jednoducho vytvoriť. Pokiaľ už jestvujú predošlé podklady, ako aj v prípade tejto záverečnej práce, do programu SocNetV jednoducho stačí vložiť Excel súbor s hranami, resp. s počiatočnými a konečnými bodmi. Na základe obdržaných údajov SocNetV vizualizuje sociálnu sieť

4.1.4 Vizualizáciu a výsledky dát

Ako je možné vidieť na **obrázku 2**, pre lepšie grafické zobrazenie sociálnej analýzy sme sa rozhodli o vizualizáciu výrezu dát. Vizualizáciu v tvare lietajúceho taniera sme získali pomocou modelu Kamada-Kawai.

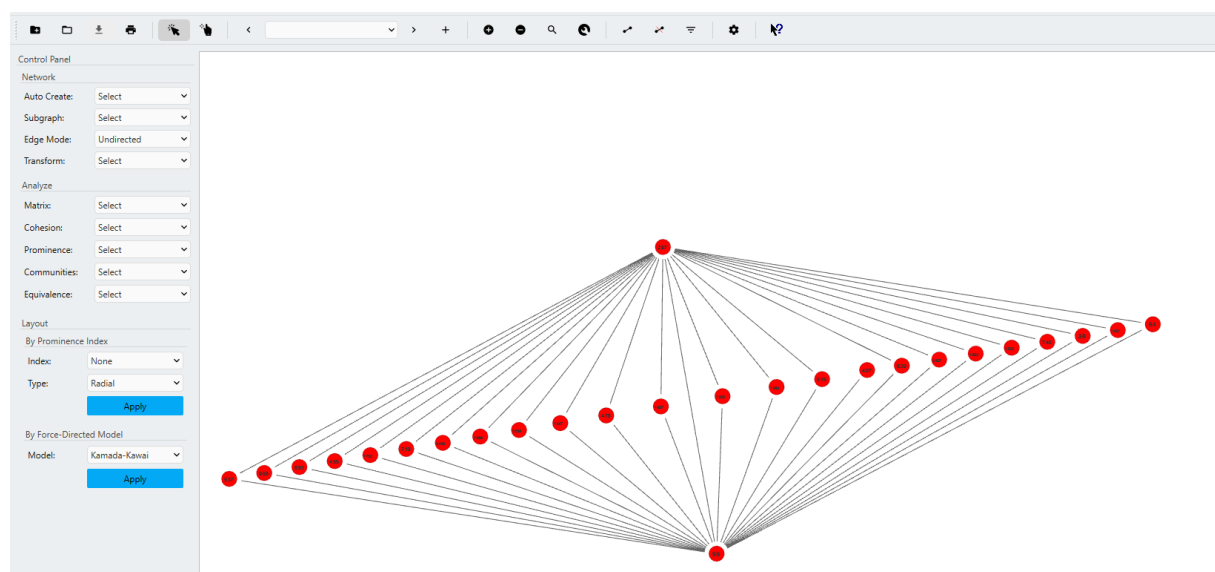
Kamada-Kawai model je algoritmus používaný pri vizualizácii grafu v analýze sociálnych sietí. Tento model sa snaží reprezentovať graf tak, aby výsledná vizualizácia bola esteticky príjemná. Základným princípom Kamada-Kawai modelu je umiestniť uzly grafu v rovine tak, aby sa minimalizovala celková dĺžka hran. Tento postup je založený na predpoklade, že uzly, ktoré sú spojené hranami, by mali byť blízko seba vo vizualizácii.

Medzi výhody Kamada-Kawai modelu môžeme začleniť esteticky príjemnú vizualizáciu. Poskytuje vizuálne atraktívne výsledky, ktoré sú prehľadné a ľahko interpretovateľné. Taktiež medzi výhody tohto modelu môžeme zaradiť zachovanie štruktúry grafu. Model sa snaží zachovať pôvodnú štruktúru grafu, čo je dôležité pre analýzu sociálnych sietí.

Ako nevýhody Kamada-Kawai modelu môžeme uviesť výpočtovú náročnosť. Proces výpočtu môže byť pre program náročný, hlavne pokiaľ sa tento model aplikuje na vyššie množstvo uzlov a hrán. Taktiež medzi nevýhodu môžeme zaradiť závislosť na počiatočnom rozmiestnení uzlov a hrán. Výsledky modelu môžu byť ovplyvnené počiatočným umiestnením uzlov, čo môže viesť k rôznym výsledkom pri opakovaných behoch algoritmu.

Možno konštatovať, že hrany vo vizualizácii sociálnej siete sú nesmerované a uzly nie sú nijak špecificky definované. Pre vylepšenie vizuálizácie by bolo možné aplikovať modifikáciu vrchného a spodného uzlu. Práve tieto dva uzly indikujú profily Twitch streamovacej platformy, ktoré sú sledované ostatnými uzlami, resp. osobami vlastniacimi profily na tejto platforme. Taktiež možno konštatovať, že zároveň aj tieto dva profily sú navzájom spojené, teda v rámci sociálnej siete vznikajú nie len vzťahové dyady ale aj triády.

Obrázok 2: Vizualizácia výrezu dát v programe SocNetV



Zdroj: SocNetV org.

Hrany v sociálnej sieti sú nesmerované teda nevieme špecifikovať, ktorý púrofil inicioval spojenie. Váženie hrán v danej vizualizácii by bolo nepodstatné, nakoľko vzťah medzi profilmi na streamovacej platforme Twitch možno definovať len binárne, teda či sa vzťah, sledovanie, medzi profilmi nachádza, alebo nie.

4.2 Python a NetworkX

Začiatkom zostavovania sociálnej siete pre analýzu sme zvolili pre svoje dáta metódu vykreslenia pomocou knižnice Networkx.

Networkx je knižnica pre programovací jazyk Python navrhnutá pre vytváranie, manipuláciu a štúdium štruktúry, dynamiky a funkcií komplexných sietí.

4.2.1 Vizuálna stránka programu

Vizuálna stránka pri práci s nástrojom Python a NetworkX silno závisí od preferencie používateľa. Vďaka svojej popularite je Python podporovaný v rôznych vývojových prostrediach a editoroch. Spracovali a posudzovali sme dáta nástroju Python v editore Visual Studio Code, editor poskytovaný a udržiavaný spoločnosťou Microsoft. Zabezpečuje bohatú sadu rozšírení pre Python, ako aj nástroje na ladenie a správu projektov.

Taktiež sme pracovali s použitím Jupyter Notebook, čo je interaktívne prostredie na vytváranie a zdieľanie dokumentov obsahujúcich kód, vizualizácie a vysvetľujúce texty. Pre analýzu bol využitý hlavne ako prostriedok na modularizáciu kódu, čím nám umožnil upravovať a znovu spúšťať len určité časti kódu bez potreby spúšťania celého kódu. Taktiež je

dôležité spomenúť, že prostredníctvom Jupyter Notebook sme mohli efektívne demonštrovať postupnosť krokov a výsledky analýzy, čo prispelo k lepšiemu pochopeniu procesu výskumu.

4.2.2 *Používateľská prívetivosť*

Python a NetworkX ponúkajú jednoduché použitie a flexibilitu. Python je považovaný za jeden z najjednoduchších programovacích jazykov na učenie a následne jeho použitie v praxi, čo uľahčuje prácu aj začiatočníkom. Jeho priateľská syntax a bohatá dokumentácia robia prácu s ním pohodlnou. NetworkX umožňuje tvorbu atraktívnych grafických reprezentácií sietí, čo uľahčuje vizualizáciu vzťahov medzi jednotlivými členmi siete a pomáha lepšie porozumieť jej štruktúre a dynamike.

Napriek týmto výhodám však Python a NetworkX majú aj svoje nevýhody. Pre človeka nie veľmi zdatného v programovaní alebo len úplneho začiatočníka môže byť náročné nastavenie vývojového prostredia pre prácu s týmito nástrojmi. Inštalácia potrebných knižníc a modulov a konfigurácia prostredia môžu byť taktiež výzvou, najmä ak chýba skúsenosť s programovaním.

Okrem toho, niektoré pokročilé analýzy alebo požiadavky môžu vyžadovať vlastné implementácie alebo použitie ďalších knižníc, čo môže mať za následok spomalenie vývoju a tvorby analýzy. Ďalej, vytvorená vizuálna reprezentácia sietí v NetworkX môže byť menej estetická alebo obmedzená v porovnaní s niektorými inými nástrojmi na vizualizáciu sietí. To môže ovplyvniť dojem z výslednej analýzy a zrozumiteľnosť pre ostatné zainteresované osoby.

4.2.3 *Spracovanie dát v rámci programu*

Ako prvý pokus sme ponechali súbor dát v plnom počte. Takýto obsiahli súbor dát však nekorešpondoval s výkonom priemerného domáceho počítača. Zároveň vizualizácia sociálnej siete bola nezrozumiteľná a nečitateľná, ako bude možné vidieť v podkapitole 4.2.4 – Vizualizácia a výsledky dát. Vzhľadom na túto skutočnosť sme sa rozhodli načítať len výrez dát z dátového súboru.

V programe Visual Studio Code sme pomocou jazyka Python vytvorili kód. Pre lepšie pochopenie je možné nižšie sledovať celú implementáciu kódu a jeho laické vysvetlenie.

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
```

```

features_df = pd.read_csv('large_twitch_features.csv')

original_size = len(features_df)

filtered_features_df = features_df[(features_df['created_at'] > '2018-04-01') &
(features_df['language'] == 'FR')]

filtered_size = len(filtered_features_df)

reduction = original_size - filtered_size

reduction_percent = (reduction / original_size) * 100

print(f"Original size of dataset: {original_size} rows")
print(f"Size of dataset after filtering: {filtered_size} rows")
print(f"Reduction in size: {reduction} rows ({reduction_percent:.2f}% reduction)")

edges_df = pd.read_csv('large_twitch_edges.csv')

filtered_edges_df =
edges_df[edges_df['numeric_id_1'].isin(filtered_features_df['numeric_id']) |

edges_df['numeric_id_2'].isin(filtered_features_df['numeric_id'])]

G = nx.from_pandas_edgelist(filtered_edges_df, 'numeric_id_1', 'numeric_id_2')

```

V rámci prvej časti kódu na riadkoch 01 až 04 je možné sledovať, že pre vytvorenie grafu je potrebné importovať viacero knižníc, nielen knižnicu NetworkX, ktorá je zodpovedná za vytvorenie štruktúry grafu. Taktiež je potrebné importovať knižnicu Matplotlib presnejšie pyplot modul tejto knižnice, ktorú sme použili na grafickú vizualizáciu a následné uloženie nášho grafu ako obrázok typu jpg.

Ďalšou importovanou knižnicou je knižnica Pandas. Pandas je open source knižnica, ktorá poskytuje množstvo nástrojov na analýzu a správu údajov. V našom prípade sme ju používali pre načítavanie datasetu údajov z csv súboru.

Poslednou použitou knižnicou bola knižnica NumPy. Knižnica NumPy (Numeric Python) je jednou z najdôležitejších knižníc v jazyku Python pre vedecké výpočty a prácu s numerickými dátami. Poskytuje vysoko výkonné nástroje na manipuláciu s poliami

a maticami, ako aj širokú škálu matematických funkcií na vektorizované operácie s dátami. V našom kóde sme ju použili na získanie nezápornej odmocniny póla bodov.

V poradí nasledujúcou časťou je načítanie dát. Kód na `načítava` dáta zo súboru s názvom `"large_twitch_features.csv"` do premennej s názvom `"features_df"`. Súbor `"large_twitch_features.csv"` je tabuľkový súbor, konkrétne CSV súbor (Comma Separated Values), ktorý obsahuje nejaké dáta. Funkcia `pd.read_csv()` je funkcia z knižnice `pandas` (používaná aj ako `pd`), ktorá umožňuje načítať dáta zo súboru CSV a uložiť ich do dátového rámca (`DataFrame`).

Ďalší riadok zachytáva kód ktorý možno interpretovať ako tvorbu premennej s názvom `"original_size"`, do ktorej sa uloží počet riadkov dátového rámca `"features_df"`. Dátový rámec `"features_df"` obsahuje tabuľku dát, ktorá bola načítaná z počiatočného dátového súboru, v CSV tabuľkovom formáte. Táto premenná teda obsahuje počet riadkov, ktoré boli načítané z tohto súboru do dátového rámca.

Na siedmom riadku kódu nájdeme selekciu, ktorú sme spravili pre zníženie počtu dát v súbore a teda pre vytvorenie dátového výrezu. Tento kód vyberá iba niektoré riadky z dátového rámca `"features_df"` a ukladá ich do nového dátového rámca s názvom `"filtered_features_df"`.

V nižších odsekoch sme sa pokúsili podrobnejšie rozobrať časti kódu:

`features_df['created_at'] > '2018-04-01':` Táto časť porovnáva hodnoty v stĺpci `"created_at"` v dátovom rámci `"features_df"` s dátumom 1. apríla 2018. Vyberá všetky riadky, ktoré majú v tomto stĺpci dátum v neskoršom čase ako 1. apríla 2018.

`features_df['language'] == 'FR':` Táto časť porovnáva hodnoty v stĺpci `"language"` v dátovom rámci `"features_df"` s hodnotou `"FR"`. Vyberá všetky riadky, ktoré majú v tomto stĺpci hodnotu `"FR"`, čo zvyčajne znamená, že ide o záznamy v francúzštine.

`&:` Tento symbol znamená logický operátor `"a"`, čo znamená, že chceme vybrať riadky, ktoré spĺňajú obe podmienky: že dátum v stĺpci `"created_at"` je po 1. apríli 2018 a že hodnota v stĺpci `"language"` je `"FR"`.

Táto časť kódu teda vyberá iba tie riadky z dátového rámca `"features_df"`, ktoré boli vytvorené po 1. apríli 2018 a majú jazyk nastavený na francúzsky. Zároveň ich ukladá do nového dátového rámca `"filtered_features_df"`.

V ďalších šiestich riadkoch (riadok 8 až 13) kód identifikuje rozdielnosť medzi pôvodným dátovým setom a filtrovanou časťou. Výsledkom tohto kódu bol výpis, ktorý konštatuje, že originálna veľkosť datasetu bola 168 114 riadkov, resp. uzlov. Po vyfiltrovaní

dát sa veľkosť, a teda dáta s ktorými sme pokračovali, zmenšili na 32 uzlov. Veľkosť redukcie bola teda 168 082, čo činí 99,98% redukciu datasetu.

V posledných štyroch riadkoch kódu sme upravili hrany na základe predvybraných uzlov. V najvrchnejšej časti kódu prebehlo načítanie súboru "large_twitch_edges.csv", ktorý obsahuje dáta o hranách siete, do DataFrame s názvom "edges_df".

Hneď pod tým sme vyfiltrovali dáta hrán na základe existencie numerických identifikátorov (numeric_id) v stĺpci 'numeric_id_1' alebo 'numeric_id_2', ktoré sú prítomné v DataFrame filtered_features_df. Výsledkom toho je DataFrame "filtered_edges_df", ktorý obsahuje len tie hrany, ktoré sú prepojené s filtrovanými uzlami.

V poslednej časti kódu sme vytvorili graf s pomenovím G zo zoznamu hrán v DataFrame "filtered_edges_df", pričom stĺpce 'numeric_id_1' a 'numeric_id_2' slúžia ako zdrojové a cieľové uzly. Funkcia nx.from_pandas_edgelist() je z knižnice NetworkX a slúži na vytvorenie grafu zo zoznamu hrán v pandas DataFrame. Táto čas predpripravila graf, no pred vizualizáciou sme pridali modifikáciu parametrov.

```
node_size = [  
v * 5000 for v in degree centrality.values()  
]  
plt.figure(figsize=(14, 9))  
nx.draw_networkx(G, pos=pos, node_size=node_size, with_labels=False, width=0.5)  
plt.axis("off")
```

Posledná časť kódu je zameraná na vizualizáciu grafu pomocou knižnice NetworkX a matplotlib v Pythone. V prvom riadku sme definovali veľkosť uzlov ("node_size") na základe centrality stupňa uzlov (degree centrality). Centrálita stupňa uzlov meria, ako dôležitý je každý uzol v grafe podľa toho, koľko hrán je s ním spojených. V ďalšom riadku pomocou kódu vytvára nástroj novú obrazovkovú plochu s rozmermi 14x9 palcov, na ktorej sa vizualizuje sociálna sieť.

V predposlednom riadku nástroj vykresľuje graf G pomocou funkcie nx.draw_networkx(), kde:

- G je graf, ktorý sa má vykresliť,
- Pos sú pozície uzlov v grafe,
- node_size sú veľkosti uzlov, ktoré sa vypočítali na predchádzajúcom riadku,
- with_labels=False – táto funkcia umožňuje vykresliť graf bez popisov uzlov,
- width=0.5 je šírka hrán v grafe.

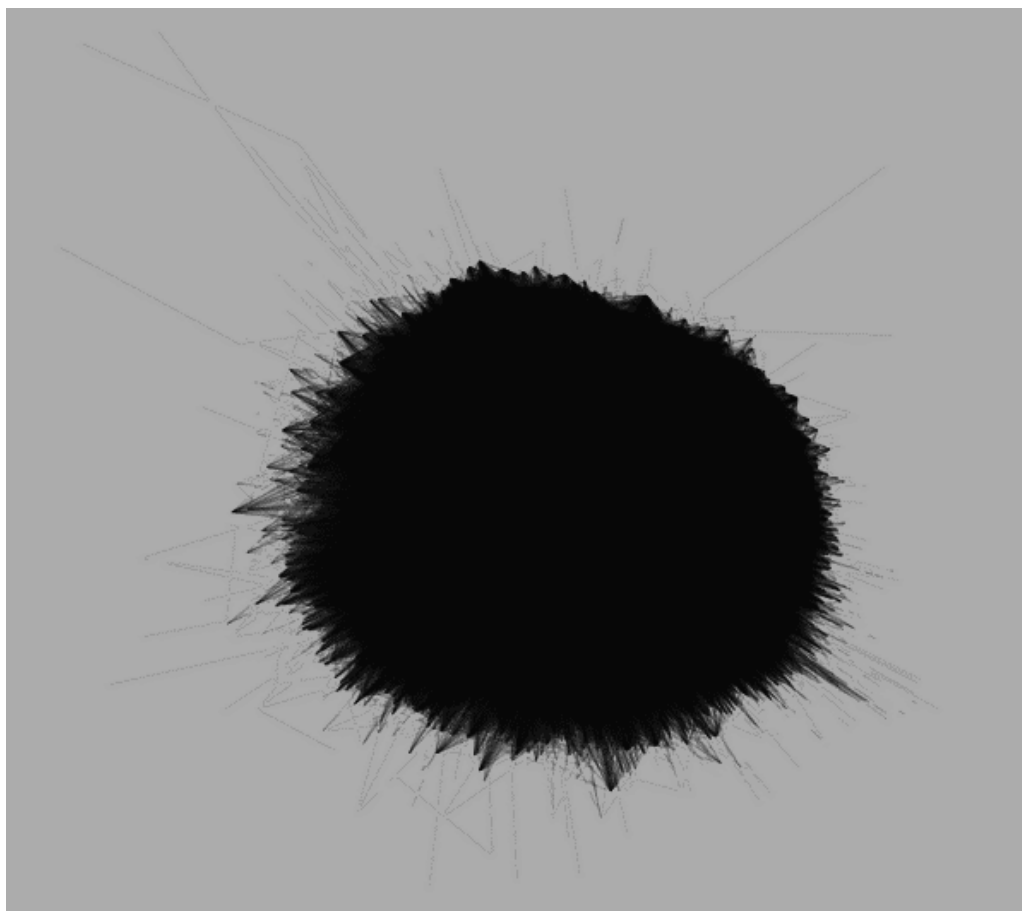
Ako po poslednú časť kódu sme vypli zobrazenie osí, čím sa zabezpečilo, že na grafe nebudú žiadne viditeľné osy, čo vytvára čistejšiu a estetickjšiu vizualizáciu.

4.2.4 Vizualizácia a výsledky dát

Ako sme definovali v predchádzajúcej podkapitole, v počiatočnej vizualizácii sociálnej siete boli ponechané dáta z datasetu bez zmeny. V predpripravenej databáze sa vyskytovalo 168 114 uzlov. Tento obnos dát a ich definovaných parametrov bol v rámci databázy prepojený pomocou 6 797 557 hrán.

Takto významný počet dát sme nechali vykresliť pomocou programovacieho jazyka Python a jeho knižnice NetworkX grafický obrazec, ktorý na pohľad neposkytuje priestor a ani možnosť analyzovať spoločné prepojenia, tvoriace sa komunity alebo iné vlastnosti grafu, ktoré by bolo možné sledovať. Celkový čas pre vizualizáciu celého datasetu presiahol na výkonnejšom počítači 11 hodín.

Obrázok 3: Grafické zobrazenie Twitch profilov a ich prepojení v programe Python



Zdroj: Vlastné spracovanie

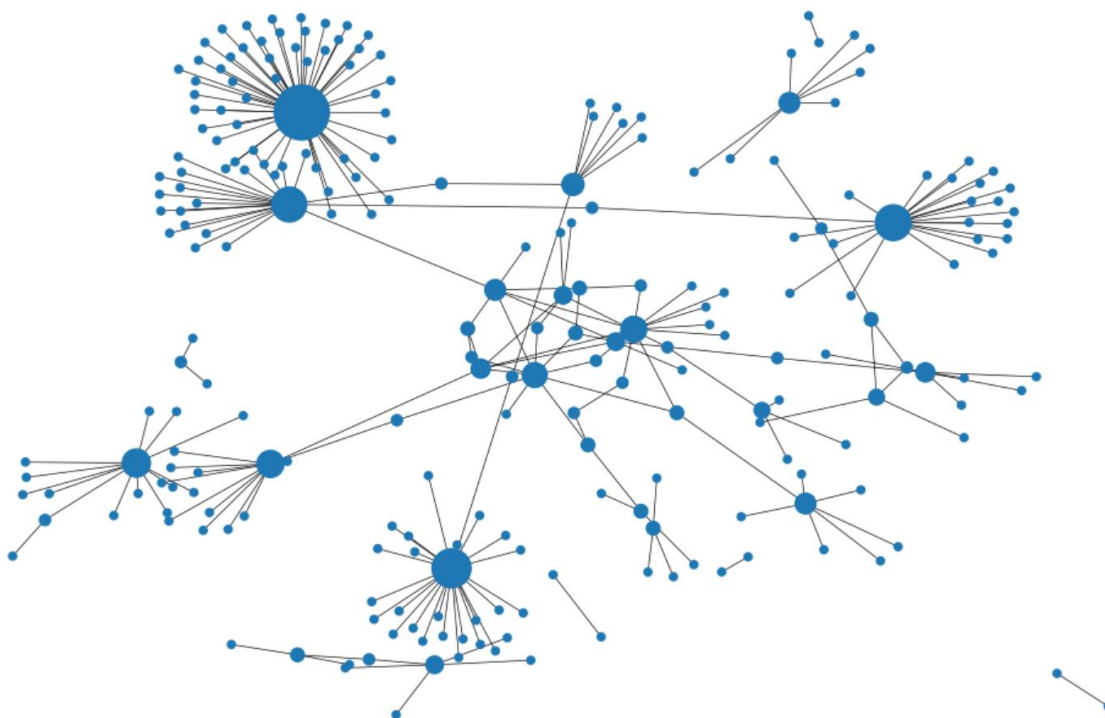
Pri pohľade na **obrázok 3**, ktorý pozostáva zo spojených bodov, je zreteľné, že vytvorenie grafického znázornenia prepojení medzi profilmi bolo pri takto veľkom obsahu neidentifikovateľné. Negatívnych bodov pri tvorení tejto vizualizácie bolo toľko, že sme sa rozhodli vyfiltrovať len výrez datasetu, ktorý tvoril 0,02% z celkového datasetu. Kód pre túto vizualizáciu sociálnej siete sme uviedli v predchádzajúcej podkapitole 4.2.3 - Spracovanie dát v rámci programu.

Vizualizácia na **obrázku 4** je podstatne väčšia ako v predchádzajúcom nástroji. Je to aj z toho dôvodu, že pri menších dátových setoch je síce vizualizácia podstatne jednoduchšia, avšak následne odvíjajúce sa pozorovania v histogramoch sú ovplynené, resp. ich interpretácia na malom počte dát je veľmi zložitá.

Na **obrázku 4** je možné vidieť sieť zloženú z profilov získaných zo streamovacej platformy Twitch, ktorých primárnym jazykom je francúzština. Rovnako ako v predchádzajúcej vizualizácii sociálnej siete, aj tu možno pozorovať tvorbu určitých skupín a prepojení medzi nimi. Väčšie uzly, ako už bolo spomínané, odrážajú počet hrán smerujúcich ku nim.

Laicky to možno opísať ako počet sledovateľov, ktorý daný profil má. Hrany vo vizualizácii sú binárne, teda nevieme konštatovať, či je sledovanie spoločné, alebo len jednostranné.

Obrázok 4: Modifikácia zobrazenia grafu a jeho bodov a hrán



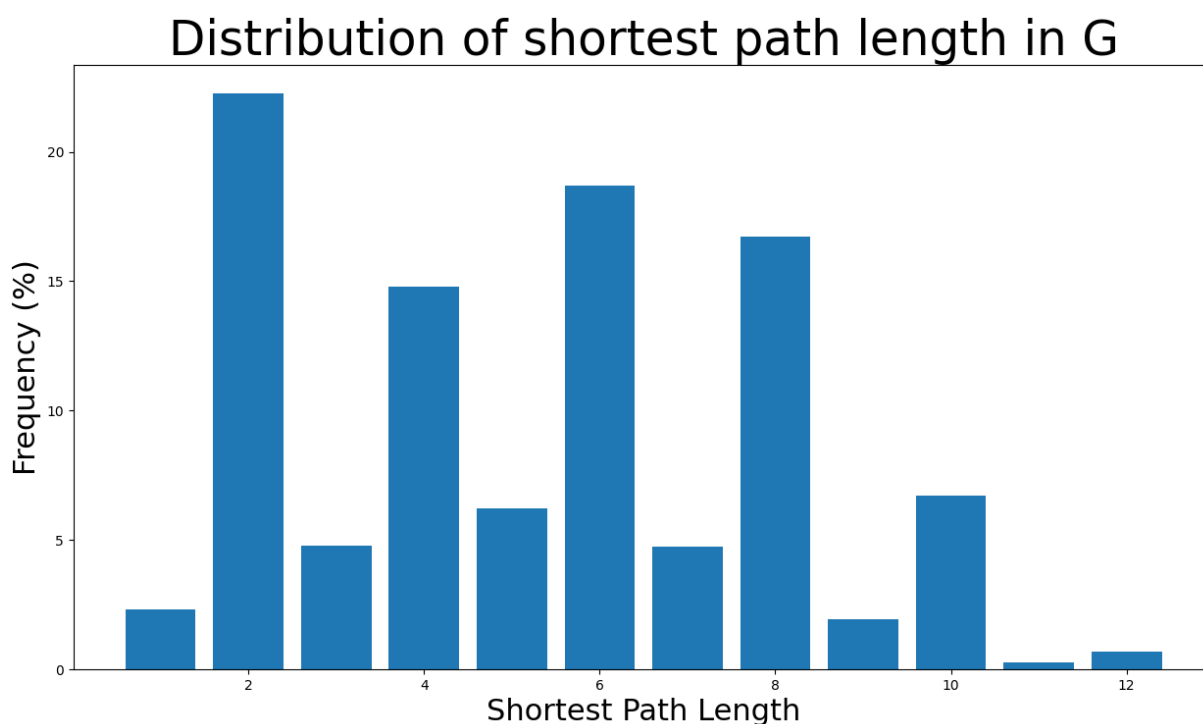
Zdroj: Vlastné spracovanie

Vo vizualizácii sociálnej siete taktiež môžeme pozorovať, že nie všetky uzly sú spojené v jednu sociálnu sieť. Je tu možnosť určiť niekoľko dyád, triád či samostatných skupín, ktoré nie sú žiadnym spôsobom spojené s hlavnou sociálnou sieťou. Túto skutočnosť možno interpretovať tak, že dané profily nijak nesúvisia s veľkou sociálnou sieťou.

Avšak taktiež treba poznamenať, že sa jedná o výrez dát a teda nie je možné toto tvrdenie potvrdiť, kým sa v nástroji pre sociálne siete nevytvorí vizualizácia celého datasetu. Vzniká tu možnosť, že uzly a samostatné skupiny môžu byť napojené na francúzsku sociálnu sieť, avšak kvôli filtrovaniu ostatných prítomných jazykov, akými v datasete boli angličtina, čínština, taliančina a mnoho ďalších, sa toto prepojenie nezobrazilo vo vizualizácii.

Z filtrovaného datasetu sme tiež siahli po tvorbe histogramu a na základe dostupných dát vytvoril histogram na **obrázku 5**.

Obrázok 5: Rozloženie dĺžky najkratších trás v grafe



Zdroj: Vlastné spracovanie

Histogram teda znázorňuje rozloženie dĺžok najkratších trás v grafe. Os x histogramu zobrazuje dĺžku najkratších trás, pričom každá hodnota na osi predstavuje počet krokov potrebných na prechod medzi dvoma uzlami. Os y zobrazuje počet najkratších trás s danou dĺžkou.

Taktiež ukazuje, že v grafe existujú najkratšie trasy rôznych dĺžok. Najčastejšia dĺžka najkratšej trasy sú 2 hrany, čo znamená, že medzi väčšinou párov uzlov v grafe existuje cesta, ktorá trvá 2 hrany alebo menej.

Z histogramu je tiež zrejmé, že existujú aj dlhšie najkratšie trasy. Napríklad existujú aj najkratšie trasy s dĺžkou 12, 14 a 16 hrán. To naznačuje, že graf má určité oblasti, ktoré sú menej prepojené ako iné časti.

Celkovo histogram poskytuje prehľad o rozložení dĺžok najkratších trás v grafe. Z informácií v histograme je možné vyvodit' závery o štruktúre a prepojení v rámci grafu.

Taktiež môžeme uviesť, že na základe histogramu je najčastejšia dĺžka najkratšej trasy 2 hrany. Rozsah dĺžok najkratších trás: 1 až 12 hrán. Môžeme tu dedukovať, že ak počet najkratších trás s dĺžkou 2 hrany je najvyšší, potom počet najkratších trás klesá s rastúcou dĺžkou.

Okrem informácií z histogramu je možné analyzovať aj iné vlastnosti grafu, ako napríklad priemernú dĺžku najkratšej trasy, medián dĺžky najkratšej trasy a módu dĺžky najkratšej trasy. Tieto informácie môžu poskytnúť ďalšia tvorba histogramov a ich interpretácia.

4.3 Gephi

Pre ilustrovanie a následnú analýzu dát sme sa rozhodli použiť program Gephi. Gephi je široko používaný softvérový nástroj, ktorý sa vyznačuje svojou schopnosťou vizualizovať a analyzovať komplexné siete. Jeho užívateľské rozhranie je intuitívne a jednoduché na ovládanie, čo umožňuje rýchle a efektívne spracovanie veľkých dátových sád.

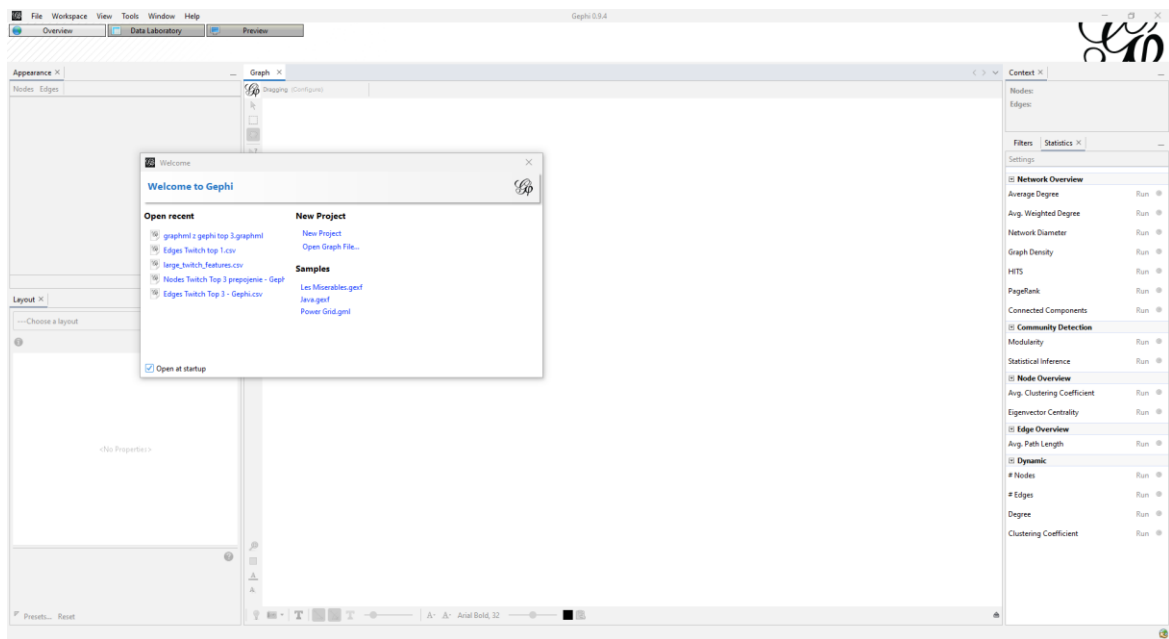
4.3.1 Vizuálna stránka programu

Pri používaní programu Gephi si môžeme všimnúť jeho vizuálne príjemný dizajn. Hoci logo programu je viac umelecky založené a nie veľmi čitateľné, samotný program je prehľadný a intuitívny. Po otvorení nás okamžite osloví estetickým dizajnom grafického rozhrania s požiadavkou na načítanie dát, ako je zachytené na **obrázku 6**.

Na vrchu obrazovky sa nachádzajú záložky s ďalšími možnosťami, zatiaľ čo najpoužívanéjšie časti programu sú umiestnené v popredí. Pre ľahkú orientáciu je nižšie umiestnené menu pre prácu s dátami, ich databáza a prehľad modelu, ktoré budú podrobnejšie vysvetlené v ďalších častiach.

V ľavom paneli máme prístup k transformácii vzhľadových aspektov uzlov a hrán. Tu môžeme upravovať veľkosť, farbu a tvar uzlov a hrán podľa našich preferencií. Nižšie sa nachádza možnosť zvoliť si požadované rozloženie modelu a ďalšie úpravy, ako je nastavenie prítlačlivosti uzlov a hrán.

Obrázok 6: Dizajn programu Gephi



Zdroj: Gephi.org.

V pravom paneli sú umiestnené filtre a štatistické záložky, kde môžeme vytvárať štatistické grafy a analyzovať dáta z rôznych perspektív. Tento panel nám umožňuje detailne preskúmať štatistické vlastnosti nášho grafu a identifikovať vzory a trendy.

Zo subjektívneho hľadiska je tento bod hodnotený pozitívne, aj preto, že v rámci programu je navigácia a hľadanie potrebných požiadaviek používateľa takmer úplne prirodzená. Pre hlbšie využitie programu sa ďalší postup nezaobišiel bez potrebného návodu, ktorý však stránka pre Gephi sama ponúka. Rovnako je k dispozícii množstvo online videí, ktoré používateľa prevedú programom.

4.3.2 Používateľská prívetivosť

Program Gephi disponuje vysokou používateľskou prívetivosťou. Práve tento aspekt ho robí atraktívnym nástrojom pre široké spektrum používateľov, od začiatočníkov až po pokročilých analytikov. Ako nový používateľ tohto programu môžeme zo svojho subjektívneho pohľadu konštatovať, že Gephi je zameraný na obecnosť, ktoré nemá znalosť programovacích jazykov, avšak má aspoň minimálny matematický poznatok a poznatok ohľadom analýzy sociálnych sietí.

Používateľská prívetivosť programu Gephi je dosiahnutá vďaka niekoľkým kľúčovým faktorom, ktoré uľahčujú jeho použitie a zabezpečujú plynulý a intuitívny pracovný proces.

Jedným z hlavných prvkov, ktoré prispievajú k používateľskej prívetivosti Gephi, je jeho jednoduché užívateľské rozhranie. Po spustení programu sa používateľovi ponúka prehľadné prostredie s logickým usporiadaním nástrojov a možností.

Každé tlačidlo a typ rozloženia je jednoducho predstavený a počas navigácie po aplikácii je možné kurzorom myši na nich na chvíľu spočinúť, čím sa zobrazí krátky popis funkcie a hodnotenie ohľadom kvality a rýchlosti rozloženia. Tieto informácie poskytujú používateľom jasný pohľad na účel a význam jednotlivých prvkov programu a prispievajú k ich lepšiemu porozumeniu a efektívnejšiemu využitiu.

Ďalším dôležitým prvkom je dostupnosť podporných materiálov a návodov. Aj keď program Gephi je navrhnutý tak, aby ho bolo možné používať aj bez hlbokých znalostí o analýze sietí, je k dispozícii rozsiahla dokumentácia a online zdroje, ako sú tutoriály a videá, ktoré poskytujú užitočné tipy, triky a postupy.

Tieto zdroje sú veľkou pomocou pre nových používateľov, ktorí sa chcú rýchlo naučiť základy a začať pracovať s programom.

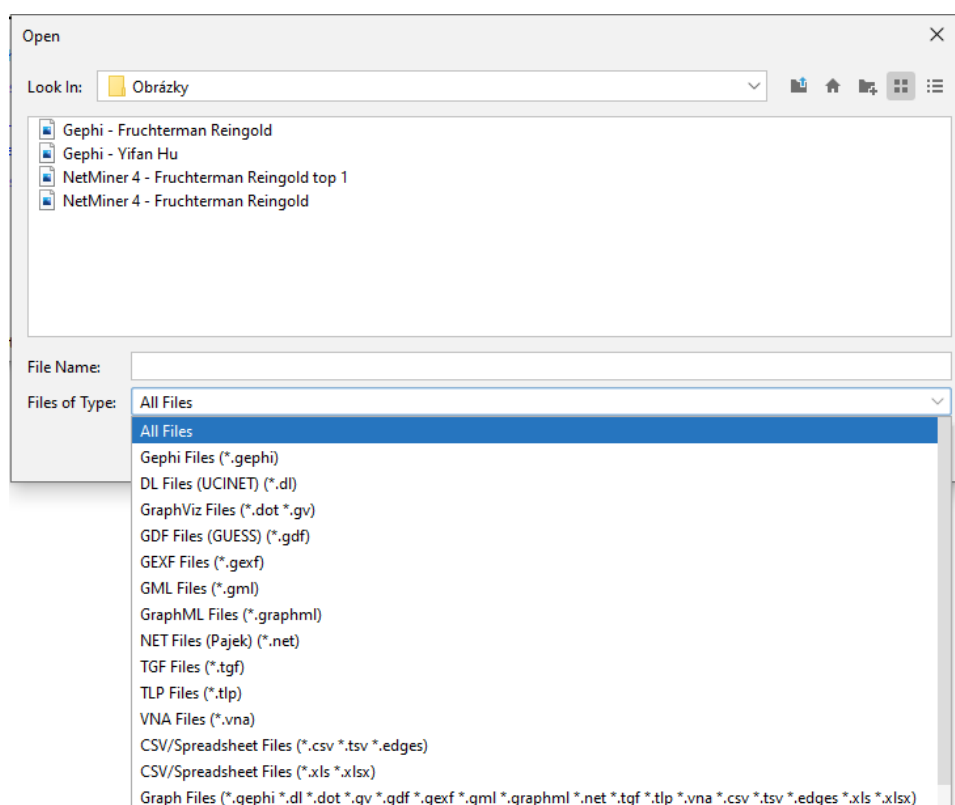
Okrem toho je program Gephi vybavený aj súborom jednoduchých, ale výkonných nástrojov, ktoré umožňujú rýchlo a efektívne manipulovať s dátami a vytvárať atraktívne vizualizácie sietí. Prehľadné ovládacie prvky a intuitívne ovládanie umožňujú používateľom plynule prechádzať medzi rôznymi funkciami a vykonávať potrebné úpravy a analýzy bez zbytočných komplikácií.

Celkovo je používateľská prívetivosť programu Gephi jedným z jeho najväčších prínosov. Jeho jednoduché užívateľské rozhranie, dostupnosť podporných materiálov a efektívne nástroje robia z Gephi vhodný nástroj nielen pre skúsených analytikov, ale aj pre nových používateľov, ktorí sa chcú zapojiť do analýzy a vizualizácie sociálnych sietí. Spracovanie dát v rámci programu

Vďaka svojmu vzhľadu a orientácii na publikum, ktoré nemá znalosti programovania, aké boli očakávané pri jazyku Python a jeho knižnici NetworkX, sa dáta do Gephi vkladajú pomerne jednoducho. Na výber je hneď z niekoľko možností:

- Ponuka – Gephi hneď pri spustení programu odporučí niekoľko vozriek dát v prípade, že človek nemá svoj vlastný dataset, alebo má záujem si program najskôr vyskúšať na iných dátach;
- Nový projekt a otvorenie súboru – Samozrejme nechýba ani klasická ponuka a teda otvorenie vlastných súborov alebo priame vytvorenie nového projektu;
- Nedávno otvorené – pokiaľ človek už pracoval s programom, Gephi si pamätá päť posledných súborov, ktoré sa v ňom otvárali bez ohľadu na ich typ.

Obrázok 7: Vkládanie dát do Gephi



Zdroj: Gephi org.

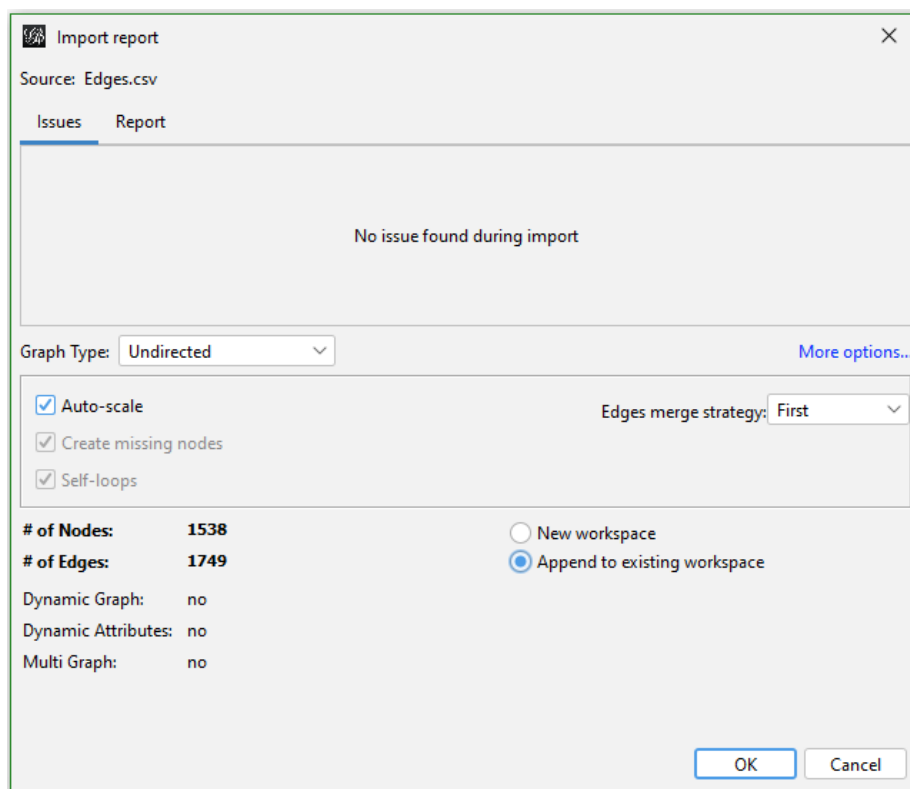
Gephi má v ponuke otvorenia hneď niekoľko typov súborov, cez ktoré je možné dáta importovať, ako je viditeľné na **obrázku 7**. Pokiaľ by si však používateľ rozhodol o zadanie svojich dát vlastnoručne, túto akciu je možné previesť priamo v Gephi, v záložke Data Laboratory.

Nakoľko sme disponovali vlastnými dátovými súbormi, dáta sme importovali do Gephi využitím otvorenia súborov. Pri vkladaní dát nám Gephi zobrazil správu, kde sa následne dalo upraviť o aký typ grafu sa jedná, či má danú databázu zahrnúť do aktuálneho alebo nového pracovného zošitu alebo ako má reagovať na duplicitné dáta tak, ako je zobrazené na **Obrázku 7**.

V rámci spracovania dát môžeme konštatovať, že Gephi mierne zaostáva. Zatiaľ čo mu nerobilo problém vizualizovať menšie dátové sety, akými boli dáta v množstve približne 2 600 uzlov a 5100 hrán, pri prvotnom testovaní dátového modelu o jeho celej veľkosti, teda 168 114 uzlov a 6 797 557 hrán sa analýza pri modelovaní zasekla hneď na začiatku a program odmietal pokračovať v modelovaní.

O trochu lepšie výsledky sme pozorovali pri dátovom modeli o veľkosti 100 000 uzlov a 2 416 904 hrán. Výkonnosť Gephi zastala na modelácii, kde sme mali možnosť pozorovať trhaný pohyb hrán pri inicializácii akéhokoľvek rozloženia.

Obrázok 8: Finalizácia implementácie dát do Gephi



Zdroj: Gephi org.

Analýza dátových modelov bola prvotne testovaná na osobnom počítači s nasledovnými parametrami:

- CPU: AMD Ryzen 5 3600
- Pamäť: 16GB DDR4
- Disk: NVMe PCIe Gen3 500GB SSD
- GPU: NVIDIA GeForce GTX 1660 Ti

Pre objektívnejšie posúdenie sme sa analýzu pokúsili namodelovať na výkonnejšom počítači s parametrami:

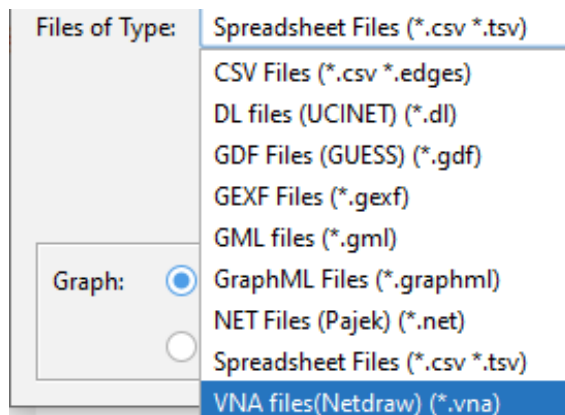
- CPU: AMD Ryzen 7 3700X
- Pamäť: 16GB DDR4
- Disk: NVMe PCIe Gen3 1TB SSD
- GPU: NVIDIA GeForce GTX 1080

Výsledok pri druhom počítači sa minimálne líšil a analyzovanie väčšieho množstva uzlov a hrán bol stále problém.

Pozitívom Gephi je skutočnosť, že konečnú analýzu možno uložiť nie len v jednom type súboru. Gephi tak ako pri prijímaní súborov podporuje hneď niekoľko typov súborov zobrazených na **obrázku 9**, pod ktorými vie prácu uložiť. Tento poznatok nám v ďalších krokoch bol

prínosom pri tvorbe analýz v nasledujúcich programov aj vďaka tomu, že nebolo potrebné opätovne vytvárať sociálnu sieť od začiatku. Práve kvôli svojej rozmanitosti sme mohli súbor uložiť pod takým typom, ktorý je použiteľný aj v iných programoch.

Obrázok 9: Dostupné typy pre uloženie súborov z programu Gephi



Zdroj: Gephi org.

4.3.3 Vizualizácia a výsledky dát

Prvotný náčrt po vložení dát do programu bol neprehľadný a nečitateľný. Program umiestnil uzly a hrany do vymedzeného priestoru v tvare štvorca, nehládajac na ich množstvo. Prvým krokom teda bola opätovný výrez dát do takého množstva, aby bol nástroj schopný pracovať bez problémov. Následne sme aplikovali analýzu dát podľa spomínaného rozloženia Yifan Hu.

Nakoľko sme pracovali s najnižším počtom uzlov a hrán, Gephi nasimulovalo dané rozloženie bez problémov a aj počas procesu bolo možné sledovať nepretržitý pohyb bodov na stanovené miesto.

Algoritmus Yifan Hu v Gephi sa snaží o vykreslenie grafu tak, aby bol čo najviac prehľadný a estetický. Tento model sa snaží minimalizovať prekrývanie uzlov a hrán a zároveň zachovať vzdialenosti medzi nimi čo najrozumnejšie. Tým sa zabezpečuje, že uzly, ktoré sú v reálnom svete bližšie k sebe, budú vizualizované aj blízko seba.

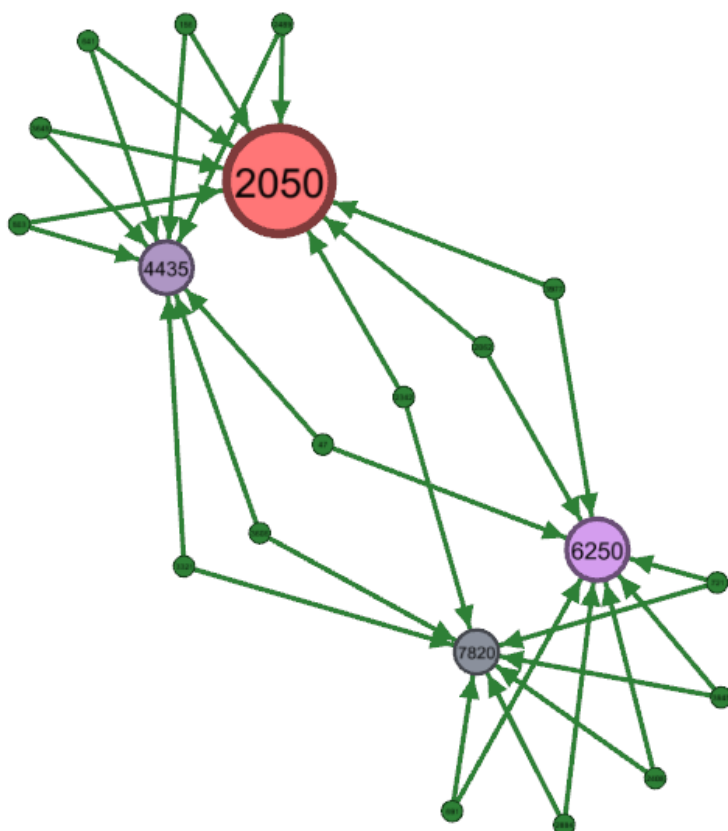
Jednou z výhod tohto modelu je jeho schopnosť efektívne spracovať veľké siete s tisíckami uzlov a hrán. Algoritmus Yifan Hu umožňuje používateľovi rýchlo a jednoducho vykresliť grafy aj z veľkých dátových sád, čo je veľmi užitočné pre analýzu sociálnych sietí a komplexných informačných sietí.

Tento bod môže potvrdiť aj sami, keď sme v modeli Yifan Hu vizualizovali sociálnu sieť o veľkosti 2 513 uzlov a ich hrán. Model perfektne, bez pomoci iných modifikácií

socigramu, bol schopný identifikovať uzly s najväčším počtom sledovateľov, teda uzly s najväčším počtom hrán.

Model Yifan Hu je v Gephi široko využívaný pre svoju schopnosť poskytnúť esteticky príjemné a prehľadné vizualizácie sietí. Používatelia majú možnosť prispôsobiť rôzne parametre algoritmu, ako je napríklad veľkosť uzlov, vzdialenosť medzi nimi a ďalšie. Tieto možnosti umožňujú užívateľom prispôsobiť výstup ich špecifickým potrebám a preferenciám.

Obrázok 10: Yifan Hu rozloženie uzlov a hrán



Zdroj: Gephi org.

Po ukončení modelovania sociálnej siete možno definovať osobitne hrany a aj body. Stále však pretrvával problém veľkého množstva čiernej farby, vďaka ktorej nebolo možné identifikovať jednotlivé body. Pokračovalo sa teda v grafickej úprave a na analýzu sa použilo farebné rozlíšenie, ktoré program ponúkal v ľavej časti obrazovky, pod piktogramom maliarskej palety.

Rozlíšenie uzlov bolo možné riešiť hneď podľa niekoľkých aspektov, akými boli napríklad informácie v rámci dát (počet sledovateľov, jazyk, dostupnosť pre adolescentov a podobne). Uzlom bolo možné modifikovať farbu ľubovoľne podľa subjektívneho výberu,

nechať program nech priradí farbu konkrétnemu bodu či bodom, alebo zvoliť vlastnú paletu troch farieb, ktorých prechodom sa značili aj uzly na **obrázku 10**.

Pre finálnu verziu, zobrazenú na **obrázku 10**, sme ďalej využili možnosť diferencovať uzly podľa veľkosti. Aplikáciu tejto možnosti bolo možné nájsť v piktograme troch kruhov umiestnených v sebe, ktoré sa nachádzali hneď vedľa farebného rozpoznania. Rovnako ako pri farbe, aj tu sme rozdelili uzly podľa ich informácií ohľadom sledovanosti, pričom najmensej hodnote sme udelili veľkosť 5 a najväčšej hodnotu 25.

Ako záverečnú možnosť rozpoznanie uzlov sme nechali vypísať názvy v centre každého z uzlov. V tomto nástroji sme sa zamerali na vizualizáciu a sledovanie počtu sledovateľov jednotlivých uzlov. Čím viac mal profil streamovacej platformy sledovateľov, tým je jeho veľkosť väčšia a farba sa približovala ku červenej.

Možno teda konštatovať, že vďaka vizualizácii sociálnej siete sme boli schopný jednoducho nájsť uzol, ktorý symbolizuje profil s najvyšším počtom sledovateľov na streamovacej platforme Twitch.

Narozdiel od vizualizácii v SocNet a NetworkX, sociálna sieť v gephi má hrany smerované a nie binárne. V tejto vizualizácii sme chceli poukázať na dôležitosť vedieť určiť, či pri vizualizácii socialnej siete je potrebné hrany smerovať alebo nechať binárne.

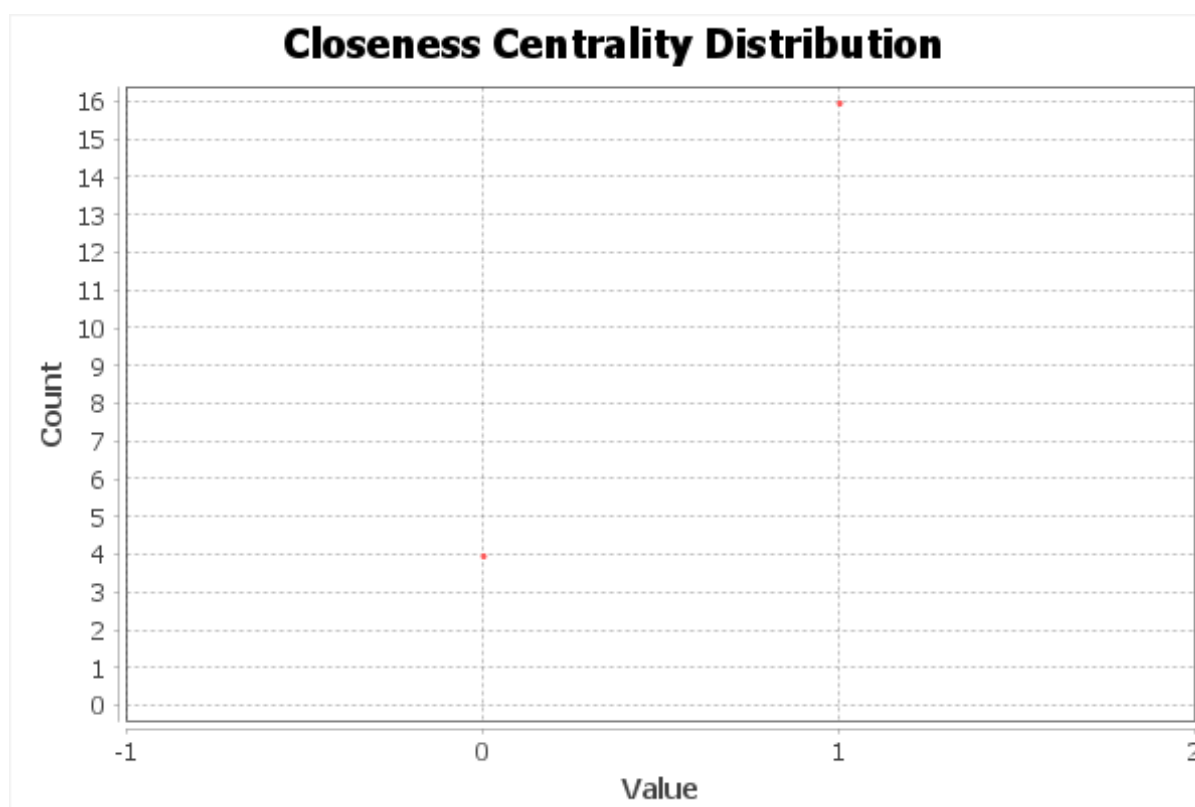
Ako bolo už mnohokrát spomenuté, binárne hrany, teda hrany bez smerovania, určujú v sociálnej sieti, či existuje vzťah medzi uzlami alebo nie. Naopak smerované hrany môžu už určiť, kto bol iniciátorom prepojenia. Vo vizualizácii sociálnej siete na **obrázku 10** smerované hrany indikujú sledovateľa a sledovaného. Hrana smeruje od sledovateľa ku sledovanému.

Streamovacia platforma Twitch má rozdielnu možnosť sledovania ako napríklad sociálna sieť Facebook. V rámci vzťahov na sociálnej sieti Facebook, nie je definové, kto inicializoval spojenie, teda toto spojenie je binárne a uzly sú touto hranou spojené. Ak si osoba na sociálnej sieti Facebook pridá medzi svojich priateľov druhú osobu, druhá osoba taktiež získa nového priateľa.

Streamovacia platforma Twitch však nevyžaduje, aby sledovateľ bol zároveň aj sledujúcim svojho sledovateľa, teda neopätuje jeho iniciáciu. Preto vo vizualizácii sociálnej siete v nástroji Gephi pozorujeme hrany so smerom.

Čo sa týkalo matematicko–štatistického výstupu Gephi ponúka v záložke Statistics, umiestnenej na pravej strane obrazovky, možnosť analyzovať vizualizáciu a dáta sociálnej siete. Pomocou viacerých štatistických a matematických nástrojov je v Gephi možné vytvoriť histogramy, ktoré vedia podrobnejšie analyzovať sociálnu sieť.

Obrázok 11: Distribúcia centrálnosti blízkosti



Zdroj: Gephi org.

Histogram, ktorý sme získali z nástroja Gephi, používa os y na zobrazenie počtu uzlov a os x na zobrazenie hodnoty centrálnosti blízkosti. Centrálnosť blízkosti je mierou toho, ako blízko je jeden uzol ku všetkým ostatným uzlom v sieti. Uzly s vysokou centrálnosťou blízkosti sú blízko mnohých ďalších uzlov v sieti. Červené body v histograme indikujú, že hodnoty majú tendenciu zhlukovať sa okolo stredy, čo znamená, že väčšina uzlov má centrálnosť blízkosti blízko priemeru.

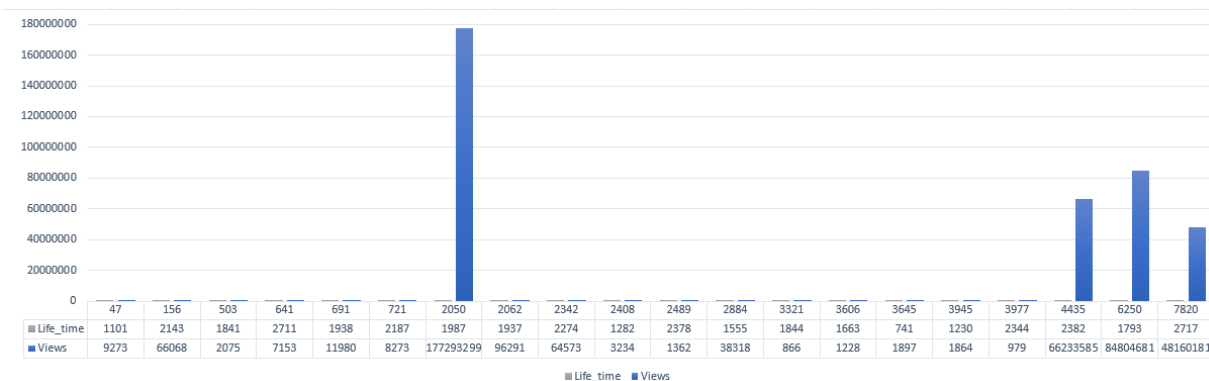
Textové popisky na osi y sa pohybujú od 0 do 16, zatiaľ čo os x siaha od -1 do 2. Je dôležité poznamenať, že centrálnosť blízkosti môže nadobúdať aj záporné hodnoty, preto sa os x rozširuje na -1.

Ďalší z histogramov sme zvolili pre schopnosť poukázať na závislosť medzi počtom sledovateľov a dĺžkou života profilu. V histograme na **obrázku 12** sme spojili do jedného histogramu hodnoty sledovateľov (Views) a dĺžkou aktivity profilu (Life_time).

Pri porovnaní histogramu z **obrázku 12** a vizualizácii sociálnej siete v gephi z **obrázku 10** je možné konštatovať, že viditeľné časti histogramu prislúchajú uzlom s najväčším počtom sledovateľov. Práve na týchto 4 uzloch môžeme pozorovať, že uzol s najväčším počtom sledovateľov má kratšiu životnosť ako uzol s pomenovaním 4 435 alebo 7 820, ktorých životnosť presiahla 2 000 dní.

Na tomto výreze dát sme teda dospeli k záveru, že závislosť medzi počtom sledovateľov a životnosťou nie je. Taktiež sa tu domnievame, že na aspekt počtu sledovateľom vyplýva mnoho externých faktorov, ako napríklad osobné preferencie osoby, kvalita obsahu, ale aj marketingové kampane v danej oblasti.

Obrázok 12: Distribúcia počtu zobrazení a dĺžky života



Zdroj: Vlastné spracovanie

4.4 NetMiner 4

Operujúci od roku 2000, NetMiner sme spolu s Gephi zaradili do rebríčka tejto záverečnej práce, aj vzhľadom na jeho priateľské prostredie pre používateľa bez znalostí programovacích jazykov. NetMiner je sofistikovaný nástroj slúžiaci na exploratívnu analýzu a vizualizáciu rozsiahlych sieťových dát na základe metód analýzy sociálnych sietí (SNA).

Jeho funkcionality zahŕňa interný skriptovací engine, ktorý je postavený na jazyku Python a disponuje automatickým generátorom skriptov. Licencie na NetMiner 4 sú poskytované študentom a učiteľom na účely vzdelávania.

4.4.1 Vizualná stránka programu

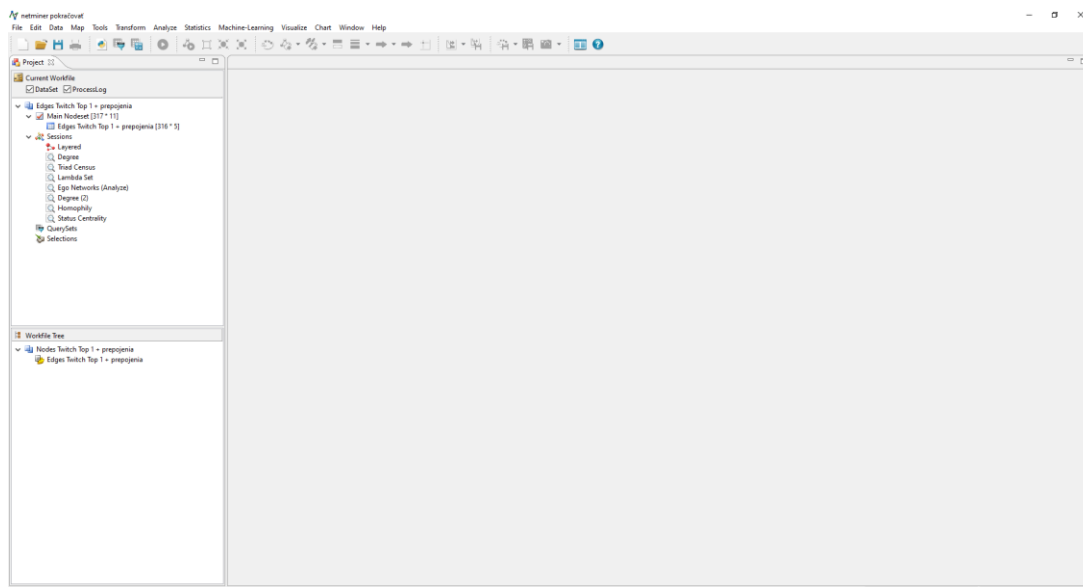
NetMiner 4 je softvérový nástroj, ktorý sa vyznačuje svojou jednoduchou a prehľadnou vizuálnou stránkou. Jeho užívateľské rozhranie je intuitívne a dobre organizované, čo uľahčuje prácu s veľkými množstvami dát. Pri pohľade na prostredie NetMineru sa okamžite všimne jeho minimálny dizajn a jednotná farebná schéma, ktorá je prirodzenejšia pre užívateľa. Zároveň však poskytuje dostatočné množstvo funkcií a možností pre analýzu sietí.

Na vrchnej lište programu sa nachádzajú hlavné navigačné prvky, ako sú možnosti na načítanie dát, tvorbu grafu a ďalšie úpravy. Tento dizajn umožňuje užívateľom rýchly prístup k základným funkcionalitám programu.

Okrem toho sa na ľavej strane nachádza panel s detailnými informáciami o dátach a modeloch, čo umožňuje užívateľovi rýchlu orientáciu a pochopenie štruktúry svojich dát.

Logo programu NetMiner jasne komunikuje jeho hlavný účel - analýzu a vizualizáciu sietí. Jeho jednoduchý, no výstižný dizajn dodáva programu profesionálny a spoľahlivý vzhľad. Pomáha to aj novým používateľom rýchlo sa zorientovať a začať pracovať s nástrojom.

Obrázok 13: Dizajn programu NetMiner 4



Zdroj: Cyram NetMiner 4

Za pozitívum je možné brať skutočnosť, že umiestnenie detailnejších nastavení a úprav v ľavom rohu panela ako je zobrazené na obrázku 13. To zabráňuje zbytočnej záplave možností a zjednodušuje používanie programu, pretože užívateľ sa nemusí navigovať cez veľké množstvo možností, ktoré nie sú relevantné pre jeho aktuálnu úlohu.

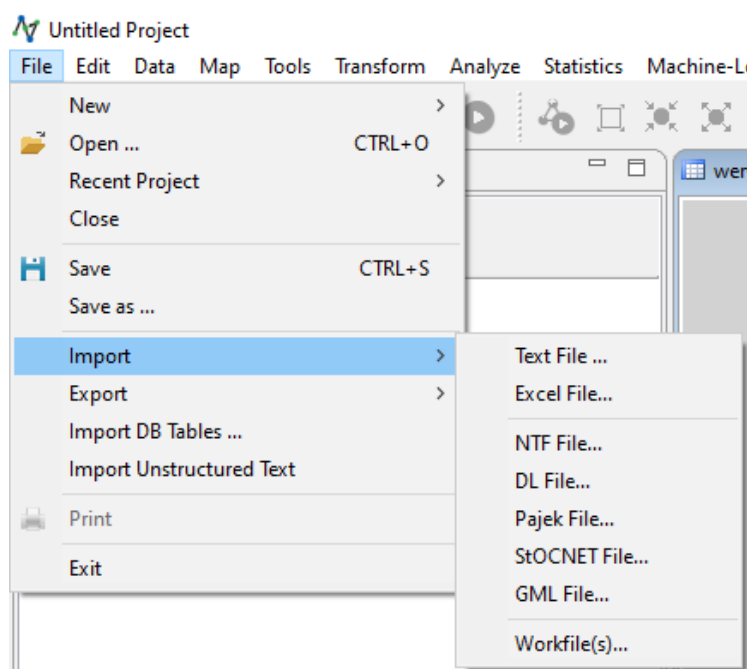
4.4.2 Používateľská prívetivosť

Rovnako ako predošlý program na tvorbu analýz sociálnych sietí, aj NetMiner 4 je navrhnutý tak, aby bol prístupný a použiteľný pre širokú škálu užívateľov, bez ohľadu na ich technické znalosti. Jeho jednoduché užívateľské rozhranie a intuitívne ovládanie robia z programu vhodný nástroj pre začiatočníkov aj pokročilých analytikov.

Jedným z kľúčových prvkov používateľskej prívetivosti je možnosť použitia programu aj bez znalosti programovacích jazykov. NetMiner 4 poskytuje množstvo vstavaných nástrojov a funkcií, ktoré umožňujú užívateľom vykonávať rôznorodé analýzy sociálnych sietí bez potreby programovania.

Okrem toho, program podporuje aj užívateľov so znalosťou programovacích jazykov, pretože ponúka možnosť rozšírenia funkcionalít pomocou jazyka Python.

Obrázok 14: Vkládanie dát do NetMiner 4



Zdroj: Cyram NetMiner 4

Vďaka prehľadným a intuitívnym tlačidlám a panelom je používanie programu NetMiner pohodlné a efektívne. Názvy tlačidiel sú jasné a zrozumiteľné, čo umožňuje užívateľom rýchlo sa orientovať a vykonávať požadované úlohy.

Taktiež kurzorom nad tlačidlami je možné získať dodatočné informácie o ich funkciách, čo prispieva k zlepšeniu používateľského zážitku.

Rovnako ako pri predošlom nástroji analýzy, aj NetMiner disponuje niekoľkými možnosťami, ako vložiť dáta do programu. Používateľ má, rovnako ako v programe Gephi, možnosť vložiť vlastné dáta z určených typov súborov, alebo vytvoriť priamo svoju sieť údajov. Ako je však možné vidieť na **obrázku 14**, vkladanie súborov je možné cez import.

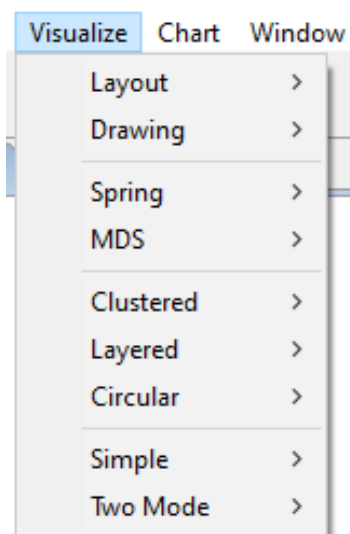
4.4.3 Vizualizácia a výsledky dát

Podľa publikácie Hansen a kol., 2011, model FR funguje na princípe odpudivých a príťažlivých síl medzi uzlami grafu. Každý uzol si predstavuje ako náboj, ktorý sa snaží od seba odháňať. Cieľom algoritmu je nájsť takú pozíciu pre každý uzol, aby sa dosiahlo rovnovážne rozmiestnenie, kde sa odpudivé a príťažlivé sily vyvažujú.

Odpudivé sily medzi uzlami sú definované ako funkcia ich vzdialenosti. Čím bližšie sú si uzly, tým silnejšia je odpudivá sila medzi nimi. Tento princíp zaisťuje, že uzly sa od seba odchýlia, aby sa zabránilo ich prekrytiu. Na druhej strane, príťažlivé sily pôsobia medzi

spojenými uzlami. Uzly, ktoré sú spojené hranou, sa pritiahnu k sebe, aby sa zabezpečilo, že spojené uzly zostanú blízko seba.

Obrázok 15: Možnosti rozloženia grafu

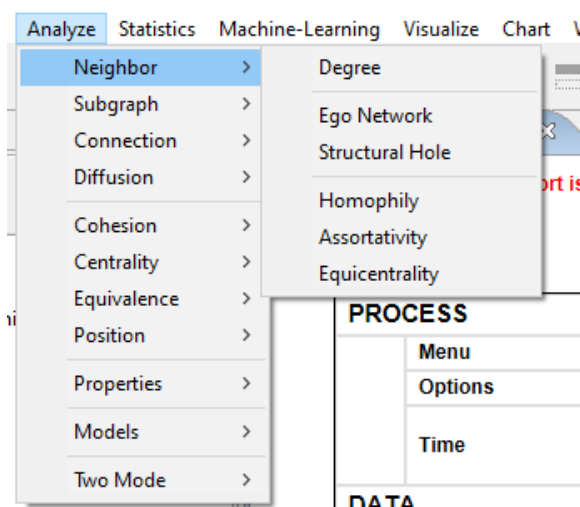


Zdroj: Cyram NetMiner 4

Algoritmus postupuje iteratívne, pričom v každej iterácii sa vypočítajú sily pôsobiace na každý uzol a aktualizujú sa ich pozície v grafe. Proces sa opakuje, kým systém nedosiahne rovnováhu, ktorá je definovaná ako stav, v ktorom sa žiadna sila nevyvíja na žiadny uzol.

Model FR je efektívny na vizualizáciu grafov, pretože dokáže zobrazit' štruktúru grafu tak, aby boli spojené uzly blízko seba a nepríbuzné uzly od seba vzdialené. Tento algoritmus sa často používa vo vizualizačných nástrojoch, ako je napríklad Gephi alebo NetworkX, na tvorbu esteticky príjemných a prehľadných vizualizácií sociálnych sietí a iných komplexných grafov.

Obrázok 16: Možnosti analýzy dát



Zdroj: Cyram NetMiner 4

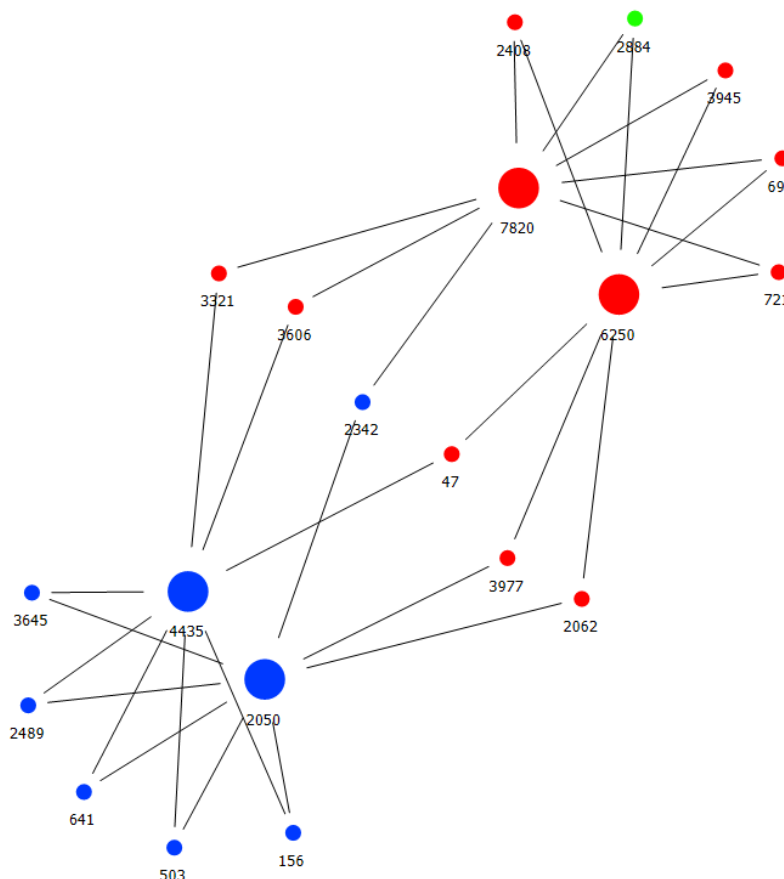
Zároveň sme sa pozreli na správu stupňovej analýzy. Rovnako ako pri modelovaní sociogramu, aj túto možnosť analýzy sme našli vo vrchných záložkách v časti Analyze, znázornenej na **obrázku 16**.

Vo finálnej vizualizácii sociálnej siete sa v programe NetMiner 4 možno pozorovať, že vizualizácia sociálnej siete sa nápadne podobá vizualizácii v nástroji Gephi. Avšak narozdiel od nástroja Gephi, kde sa vizualizácia riadila podľa modelu Yifan Hu, v tomto prípade sme využili model Fruchterman-Reingold.

Aj keď majú tieto modely podobné princípy a dosahujú podobné výsledky, nie sú úplne rovnaké. Existujú medzi nimi určité rozdiely v tom, ako vypočítavajú sily medzi uzlami a ako sa vykonáva proces rozmiest'ovania.

Napríklad Kamada-Kawai model môže mať tendenciu vyvolať viac "prepruženia" v grafe, čo môže viesť k vyššej hustote uzlov v niektorých oblastiach. Na druhej strane, Fruchterman-Reingold model môže byť efektívnejší pri rozmiest'ovaní grafov s väčším počtom uzlov.

Obrázok 17: Fruchterman-Reingold model



Existuje niekoľko podobností medzi nimi, ktoré spôsobujú, že tieto dva modely sú často považované za podobné:

- Fyzikálne modelovanie: Oba modely sa zakladajú na fyzikálnych princípoch, konkrétne na princípe sil, ktoré pôsobia medzi uzlami v grafe. Tieto sily sú využité na rozmiest'ovanie uzlov tak, aby sa dosiahol rovnovážny stav, kde sú sily vyvažované.
- Odpudivé a príťažlivé sily: Oba modely využívajú odpudivé sily medzi uzlami, ktoré sú definované ako funkcia ich vzdialenosti. Príťažlivé sily sa pôsobia medzi spojenými uzlami, pričom spojené uzly sa pritiahnu k sebe.
- Iteratívny proces: Oba modely postupujú iteratívne, pričom v každej iterácii sa vypočítajú sily medzi uzlami a aktualizujú sa ich pozície v grafe. Proces sa opakuje, kým sa nedosiahne rovnovážny stav.
- Výsledný layout: Obidva modely produkujú podobné vizuálne výsledky, kde sú uzly grafu rozmiestnené tak, aby boli spojené uzly blízko seba a nepríbuzné uzly od seba vzdialené.

Na **obrázku 17**, zobrazujúcom vizualizáciu sociálnej siete, je možné pozorovať binárne hrany medzi uzlami, podobne ako v prípade vizualizácie sociálnej siete v nástroji SocNetV (kapitola 4.1 SocNetV). Týmto je potvrdené, že vzťahy medzi uzlami existujú, avšak nie je možné určiť, či sledovanie profilov je vzájomné alebo nie.

Aplikovali sme modifikáciu uzlov aj v tejto vizualizácii. Veľkosť uzlov bola zobrazená na základe počtu sledovateľov daného profilu na streamovacej platforme Twitch.

Farebnosť uzlov bola determinovaná jazykom, ktorý si používateľ zvolil pri vytváraní profilu na platforme. Vo výreze dát, na ktorom sme aplikovali vizualizáciu a analýzu sociálnej siete, existujú tri jazyky: angličtina (červená), čínština (modrá) a taliančina (zelená).

Vizualizácia na **obrázku 17** ukazuje, že taliansky determinovaný uzol je vo vzťahu s uzlami, ktoré sa radia k anglicky hovoriacim. Taktiež je možné vidieť, že napriek jazykovej bariére, anglicky a čínsky hovoriace profily sa spájajú viackrát v danej sieti.

V tejto oblasti možno pokračovať v polemizácii o tom, prečo jazyková bariéra nebráni nadviazaniu vzťahov medzi uzlami s rôznymi jazykmi, prípadne skúmať bližšie dáta a zistiť, čo viedlo k týmto vzťahom.

Nakoľko dataset, z ktorého bol výrez dát získaný a následne bol použitý na vizualizáciu, neobsahuje informácie o vzťahoch medzi jazykmi, nie sme schopní dospieť k záveru, prečo

tieto vzťahy vznikli. Na odpoveď na túto otázku by bolo potrebné pokračovať v získavaní dát a ich následnej analýze.

Záver

Analýza sociálnych sietí na streamovacej platforme Twitch priniesla cenné poznatky o prepojení a interakcii medzi kanálmi a ich sledovateľmi. Aj keď sme na začiatku narazili na technické problémy a museli sme upravovať dáta, s kombináciou nástrojov SocNetV, NetMiner, Gephi a Pythonu s knižnicou NetworkX sa nám podarilo dosiahnuť úspešné výsledky.

NetMiner nám pomohol analyzovať sociálne siete s dôrazom na identifikáciu komúní, centrality a ďalšie štatistické metódy. SocNetV a Gephi sa ukázali ako užitočné nástroje pre vizualizáciu a analýzu dát. SocNetV ponúkol jednoduché rozhranie a umožnil rýchlu identifikáciu komúní a vzťahov medzi kanálmi. Gephi zasa poskytoval silné vizualizačné možnosti a umožnil detailnú analýzu prepojení a vzťahov.

NetworkX nám pomohol manipulovať s dátami v jazyku Python a implementovať algoritmy pre analýzu sociálnych sietí. Vďaka knižnici NetworkX sme dokázali vytvoriť grafy, identifikovať komunity a analyzovať centrálnosť jednotlivých uzlov v sieti.

Analýza odhalila, že medzi sledovanými kanálmi existujú rôzne typy prepojení, a to aj napriek jazykovým bariéram. Identifikovali sme komunity a sledovali ich rozvoj. Zistili sme, že prepojenia medzi kanálmi sú komplexné a ovplyvňujú sa navzájom.

Napriek úspechom sme sa stretli aj s obmedzeniami pri práci s veľkým množstvom dát. Systém síce umožňoval sledovať stiahnuté dáta, avšak vizualizáciu celej sociálnej siete nebolo možné vykonať na priemernom domácom počítači tak, aby vizualizácia prebehla bez sekania či padania systému.

Na základe našich skúseností môžeme konštatovať, že neexistuje jeden "najlepší" nástroj pre analýzu sociálnych sietí. Optimálny prístup spočíva v kombinácii vhodných nástrojov a systematickom prístupe k analýze. Takáto kombinácia nám umožňuje porozumieť komplexnej štruktúre a vzťahom na platforme Twitch a otvára cestu k ďalším možnostiam výskumu a propagácie obsahu na tejto platforme.

Analýza sociálnych sietí na Twitchi má široké uplatnenie v rôznych oblastiach. Môže slúžiť na pochopenie správania sledovateľov, identifikáciu trendov, analýzu konkurencie a optimalizáciu marketingových stratégií.

Výsledky analýzy nám taktiež pomôžu pri vývoji algoritmov pre personalizované odporúčanie obsahu a pri budovaní efektívnejších komúní na platforme Twitch.

Veríme, že naša práca prispeje k lepšiemu pochopeniu fungovania sociálnych sietí na platforme Twitch a k efektívnejšiemu využívaniu tejto platformy pre rôzne účely.

Zoznam použitej literatúry

Literatúra v tlačenej podobe:

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

Literatúra v internetovej podobe a odborné články:

FREEMAN, Linton. (2004). The Development of Social Network Analysis [online]. Dostupné na:

https://www.researchgate.net/publication/239228599_The_Development_of_Social_Network_Analysis

OTTE, Evelien & Rousseau, Ronald. (2002). Social Network Analysis: A Powerful Strategy, also for the Information Sciences. Journal of Information Science [online]. 28. 441-453. 10.1177/016555150202800601. Dostupné na:

https://www.researchgate.net/publication/242401176_Social_Network_Analysis_A_Powerful_Strategy_also_for_the_Information_Sciences/link/0c960537a4bc5c6efa000000/download

TRINGALI, Angela & Sherer, David & Cosgrove, Jillian & Bowman, Reed. (2020). Life history stage explains behavior in a social network before and during the early breeding season in a cooperatively breeding bird. PeerJ [online]. 8. e8302. 10.7717/peerj.8302. Dostupné na: https://www.researchgate.net/publication/339149346_Life_history_stage_explains_behavior_in_a_social_network_before_and_during_the_early_breeding_season_in_a_cooperatively_breeding_bird/citation/download

HARRIS JK, Clements B (2007). Using social network analysis to understand Missouri's system of public health emergency planners. Public Health Rep [online].;122(4):488-498. doi:10.1177/003335490712200410. Dostupné na:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1888499/>

B. Rozemberczki and R. Sarkar (2021). Twitch Gamers: a Dataset for Evaluating Proximity Preserving and Structural Role-based Node Embeddings [online]. Dostupné na: https://snap.stanford.edu/data/twitch_gamers.html

W3Schools. Matplotlib Tutorial [online] W3Schools. Dostupné na: https://www.w3schools.com/python/matplotlib_intro.asp

NumPy. NumPy documentation [online] NumPy. Dostupné na: <https://numpy.org/doc/stable/>

Sociogram. In: ManagementMania.com [online]. Wilmington (DE) 2011-2022, 18.05.2016 [cit. 20.04.2022]. Dostupné na: <https://managementmania.com/sk/sociogram>

TOMOV, Patrik (2022). Webové rozhraní pro procházení úložiště RDF. Online. Bakalářská práce. Brno: Vysoké učení technické v Brně. Dostupné na: <https://theses.cz/id/ut2yzt/>. [cit 2024-04-23]

VARGA, Ladislav. Modelování HW designu v UGE. Online. Bakalářská práce. Brno: Vysoké učení technické v Brně. 2020. Dostupné na: <https://theses.cz/id/tdi6g9/>. [cit. 2024-04-23]

DESALE, Devendra. (2015). Top 30 Social Network Analysis and Visualization Tools [online]. Dostupné na: <https://www.kdnuggets.com/2015/06/top-30-social-network-analysis-visualization-tools.html/2> [cit. 2024-04-25]

Gabasová, E. (2015). How Star Wars Characters Connect [Blog Post]. Získané z <https://evelinag.com/blog/2015/12-15-star-wars-social-network/>

CrashCourse. (2017). Why Should You Care About Social Networks? [Video]. YouTube. https://youtu.be/PJOM0m_WeTA

NetworkX Developers (2021). "Facebook Graph." NetworkX.. https://networkx.org/nx-guides/content/exploratory_notebooks/facebook_notebook.html.