

ŽILINSKÁ UNIVERZITA V ŽILINE
Fakulta riadenia a informatiky

AUTOREFERÁT
DIZERTAČNEJ PRÁCE

(1. strana autoreferátu)

Žilinská univerzita v Žiline
Fakulta riadenia a informatiky

Ing. Alexander Brezáni

Autoreferát dizertačnej práce

2D a 3D vizualizácia s využitím fyzikálnych herných prostredí

na získanie akademického titulu „**philosophiae doctor**“ (v skratke **PhD.**)
v študijnom programe doktorandského štúdia

aplikovaná informatika

v študijnom odbore:
informatika

Žilina, Máj 2024

(2. strana autoreferátu)

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia doktorandského štúdia na katedre Katedre Informatiky Fakulte riadenia a informatiky Žilinskej univerzity v Žiline

Predkladateľ: Ing. Alexander Brezáni
Katedra Informatiky
Fakulta riadenia a informatiky
Žilinská univerzita v Žiline

Školiteľ: doc. Ing. Jozef Kostolný, PhD.
Katedra Informatiky
Fakulta riadenia a informatiky
Žilinská univerzita v Žiline

Oponent: doc. Ing. Marián Hruboš, PhD.
Katedra riadiacich a informačných systémov
Fakulta elektrotechniky a informačných technológií
Žilinská univerzita v Žiline

Oponent: prof. RNDr. Valerie Novitzká PhD.
Katedra počítačov a informatiky
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach

Autoreferát bol rozoslaný dňa: 01.10.2024

Obhajoba dizertačnej práce sa koná dňa 22.10.2024 o 10:30 h. pred komisiou pre obhajobu dizertačnej práce schválenou pracovnou skupinou odborovej komisie v študijnom odbore **informatika v študijnom programe aplikovaná informatika**, vymenovanou dekanom Fakulty riadenia a informatiky Žilinskej univerzity v Žiline.

prof. Ing. Ľudmila Jánošíková, PhD.
predsedníčka pracovnej skupiny odborovej komisie
v študijnom odbore **informatika**
v študijnom programe **aplikovaná informatika**

Fakulta riadenia a informatiky
Žilinská univerzita
Univerzitná 8215/1
010 26 Žilina

Abstrakt

Práca prezentuje súčasné možnosti vizualizácie informácií s ohľadom na efektívny spôsob ich prezentácie a reprezentácia v dvoj a trojrozmernom priestore. Výsledok výskumu predstavuje vytvorenie nových algoritmov a metodológie pre interaktívne spracovanie dát s využitím metód vizualizácie dát a informácií, špecificky vizualizácie rozsiahlych sieťových štruktúr. Dizertačná práca popisuje výskum problému vizualizácie veľkých množín dát a ich štruktúry (prepojení), analyzuje existujúce a nové techniky viacrozmernej vizualizácie dát a zameriava sa na využitie prostredí s implementovaným fyzikálnym prostredím používaných najmä pri vývoji hier a realistických vizualizácií pre potreby vizualizácie sieťových dát. Výsledkom práce sú nové vizualizačné algoritmy pre 3D prostredie použiteľné pre vizualizáciu sieťových štruktúr – grafov a metodológia vytvárania trojrozmerných vizualizácie s ohľadom na použitie prostredí pre vývoj počítačových hier. Podstatnou výhodou takéhoto prístupu je priame zapojenie používateľa-analytika do procesu získavania znalostí z dát.

Kľúčové slová: vizualizácia dáta, viacrozmerné dáta, vizualizačné algoritmy, interaktívne spracovanie dát, analýza dát

Abstract

The work presents the current possibilities of information visualization with regard to the effective way of their presentation and representation in two and three-dimensional space. The result of the research is the creation of new algorithms and methodology for interactive data processing using data and information visualization methods, specifically visualization of large network structures. The dissertation describes research on the problem of visualization of large data sets and their structure (connections), analyzes existing and new techniques of multidimensional data visualization and focuses on the use of environments with an implemented physical environment used mainly in the development of games and realistic visualizations for the needs of network data visualization. The result of the work are new visualization algorithms for a 3D environment that can be used for the visualization of network structures - graphs and a methodology for creating three-dimensional visualizations with regard to the use of environments for the development of computer games. A significant advantage of such an approach is the direct involvement of the user-analyst in the process of obtaining knowledge from data.

Keywords: data visualization, multidimensional data, visualization algorithms, interactive data processing, data analysis

Obsah

ÚVOD.....	6
OPIS DIZERTAČNEJ PRÁCE	7
<i>Vedecké otázky a tézy dizertačnej práce</i>	<i>7</i>
<i>Súčasný stav riešenej problematiky doma a v zahraničí</i>	<i>8</i>
<i>Vizuálna analýza.....</i>	<i>8</i>
<i>Prechod od 2D k 3D vizualizáciám</i>	<i>9</i>
CIEĽ PRÁCE	9
METODIKA PRÁCE A METÓDY SKÚMANIA.....	10
<i>Charakteristika objektu skúmania</i>	<i>10</i>
<i>Pracovné postupy.....</i>	<i>11</i>
<i>Spôsob získavania údajov a ich zdroje.....</i>	<i>12</i>
VÝSLEDKY PRÁCE.....	12
<i>Vizuálna analýza v dvojrozmernom priestore</i>	<i>13</i>
<i>Vizuálna analýza v trojrozmernom priestore</i>	<i>14</i>
<i>Metodika prípravy dát pre vizualizácie v 3D.....</i>	<i>15</i>
<i>Algoritmus 1: Pružinový algoritmus (Spring Layout Algorithm)</i>	<i>17</i>
<i>Algoritmus 2: Silou riadený algoritmus (Force Driven Algorithm)</i>	<i>18</i>
<i>Algoritmus 3: Pružinovo-silový algoritmus</i>	<i>19</i>
<i>Algoritmus 4: Pružinovo-silový algoritmus s kontrolovaným rozmiestňovaním</i>	<i>20</i>
<i>Modifikácie 3D zobrazenia – interaktívne funkcie</i>	<i>25</i>
DISKUSIA	26
<i>Porovnanie časovej zložitosti.....</i>	<i>28</i>
<i>Vizuálne porovnanie</i>	<i>29</i>
<i>Porovnanie zhlukov a priestorovej vzdialenosti</i>	<i>32</i>
ZÁVER	35
ZOZNAM POUŽITEJ LITERATÚRY	37

Úvod

Spracovanie a analýza dát je v súčasnej dobe štandardná činnosť aplikovaná v mnohých odvetviach priemyslu, výskumu, školstva, marketingu a iných. Porozumenie analyzovaným dátam a ich efektívne použitie môže byť v mnohých prípadoch náročný proces, pri ktorom dochádza k mnohým chybám spôsobeným prezentáciou (vizualizáciou) dát, ako aj samotnou interakciou používateľa-analytika s dátami. Zložitosť a štruktúra analyzovaných dát sa neustále zvyšuje, pričom použitie štandardných vizualizačných metód sa v mnohých prípadoch ukazuje ako neefektívne, keďže neposkytujú dostatočnú silu vizualizácie na zachytenie všetkých vlastností analyzovaných dát [1] [2].

Pre získanie informácie o vzťahoch medzi dátami je nutné poznať ich vzájomné prepojenia vychádzajúce z modelovaného konceptu. Okrem štandardných metrík používaných v štatistickej analýze, korelácie a kovariancie, je v prípade tzv. Prepojených dát (angl. *linked data*) možné použiť reprezentáciu dát v podobe grafu a príslušné dobre definované grafové metriky.

Samotná grafová reprezentácia dát má okrem toho viacero výhod. Po prvé, umožňuje vizuálny a intuitívny spôsob pochopenia vzťahov a spojení medzi dátovými bodmi. Po druhé, umožňuje efektívne a flexibilné dopytovanie, pretože konkrétne reprezentácie grafov sú navrhnuté tak, aby sa ľahko prechádzali a prehľadávali. Okrem toho grafová reprezentácia údajov podporuje zložité a dynamické dátové štruktúry, vďaka čomu je vhodná pre scenáre s vyvíjajúcimi sa a vzájomne prepojenými údajmi. Vo výsledku teda grafová reprezentácia dát uľahčuje efektívnu analýzu a modelovanie sietí, ako sú sociálne siete, systémy odporúčaní a dodávateľské reťazce.

Vizualizáciu dát je možné realizovať s využitím viacerých dimenzií priestoru. Každý n-rozmerný priestor má svoje špecifické metódy a algoritmy. Pre človekom spracovateľné výsledky je možné využiť jedno, dvoj, a trojrozmerný priestor napriek skutočnosti, že existujú aj postupy pre vizualizáciu n-rozmerných dát. Základné predpoklady vizualizácie dát v priestore sú [3]:

- **Jednorozmerné (1D) dáta:** Údajové položky môžu byť vizualizované s využitím napr. histogramov. Dáta však majú len jednu premennú a každá dátová položka má jednu hodnotu.
- **Dvojrozmerné (2D) dáta:** Na vizualizáciu údajov s dvoma premennými je možné použiť dvojrozmerné grafy. Tieto dve premenné možno použiť na porovnanie údajov ich vizualizáciou pomocou čiarového grafu. Na vizualizáciu 2D údajov možno použiť aj bodové grafy, koláčové grafy, stĺpcové grafy a mapy.
- **Trojrozmerné (3D) dáta:** Trojrozmerné grafy, ako napríklad 3D stĺpcové grafy a realistické zobrazenia, môžu zobraziť viac informácií ako 1D a 2D grafy. Na vizualizáciu údajov v 3D priestore sa používajú tri premenné. Väčšina softvérových nástrojov je vhodná na vizualizáciu údajov malých rozmerov, ako sú 2D a 3D dáta.
- **N-rozmerné dáta:** N-rozmerné dáta sú dáta n rozmerov, t.j. viacrozmerné dáta s n rozmermi. V niektorých prípadoch je na vizualizáciu potrebných viac premenných, pričom v takýchto situáciách sa pre zrozumiteľnosť výslednej vizualizácie používa technika viacrozmerného grafu. Príkladmi metód sú rozptylové matice, mapy a glyfy¹.

Viacrozmerný priestor a moderné technológie obohatenej a virtuálnej reality, umožňujú nový pohľad na analýzu dát a následnú interakciu. Využitie viacrozmerného priestoru prináša dostatočnú flexibilitu pre manipuláciu s dátami, čo môže výrazne ovplyvniť porozumenie

¹ Glyfy sú grafické entity - zobrazenia, ktoré reprezentujú jednu alebo viac údajových hodnôt prostredníctvom atribútov, ako je tvar, veľkosť, farba a poloha. Používajú sa pri vizualizácii údajov a informácií a sú špeciálne vhodné na zobrazovanie zložitých, viacrozmerných množín údajov.

používateľ-a-analytika zobrazeným dátam. Za tohto predpokladu môžeme uvažovať o využití moderných technológií, ako platformy pre interakciu s vizualizovanými dátami [4]. Trojrozmerný priestor umožňuje zakomponovať reálne prostredie do procesu analýzy dát, pričom je možné využitie fyzikálnych vlastností a zákonitostí pre zložitejšie vizualizácie. Fyzikálne herné prostredia ako napríklad Unity² alebo Unreal Engine³, poskytujú možnosť zakomponovania fyzikálnych vlastností pre vizualizované dáta, pri zachovaní vhodnej úrovne optimalizácie samotnej prípravy výslednej vizualizácie. Výhodou práce v trojrozmernom priestore je možnosť implementácie grafov a ich využitie v procese vizualizácie komplexných dátových štruktúr. Efektívnym spojením spomínaných technológií a metód, môžeme predpokladať zlepšenie porozumenia analyzovaným dátam, ako aj ich spracovania a použitia.

Opis dizertačnej práce

Cieľom dizertačnej práce je využitie prirodzených vzťahov medzi dátami s ohľadom na grafovú reprezentáciu dát v procese vizualizácie. Prepojené dáta sú zvyčajne vizualizované ako zložité grafy alebo siete. Vzhľadom na zložitosť údajov je 2D projekcia založená len na súradniciach x a y komplikovaná a neumožňuje prirodzený interakčný mechanizmus pre ďalšie skúmanie údajov. 3D vizualizácia využíva dodatočný rozmer z , vyžaduje však určitú úroveň interakcie a pohybu kamery na odhalenie prekrytých resp. zakrytých informácií. V dizertačnej práci sa zameriavame na 3D vizualizáciu založenú na fyzickom prostredí poskytovanom nástrojom Unity tak, aby sme rešpektovali požiadavky na interakciu, pohyb kamery a využitie fyzických síl na priestorové rozmiestnenie dát. Ukážeme výhody tohto prístupu spolu s obmedzeniami založenými na prispôbení herného prostredia pre 3D vizualizáciu grafov.

Trojrozmerný priestor však oproti dvojrozmernému prináša nové problémy, s ktorými je potrebné sa vysporiadať, pretože pridaním tretej dimenzie sa mnohé algoritmy používané v 2D priestore stanú neefektívne až nepoužiteľné. V našej práci sme sa zamerali na silou riadené algoritmy, dobre implementovateľné v prostrediach s fyzikálnym modelom, s cieľom usporiadať graf v podobe siete v 3D priestore čo najzrozumiteľnejšie..

Vedecké otázky a tézy dizertačnej práce

V rámcovom projekte boli zadefinované nasledovné výskumné otázky (VO):

- **VO1:** Akým spôsobom je možné využiť fyzikálne herné prostredia na efektívnu vizualizáciu veľkých dátových množín?
- **VO2:** Aké informácie o dátach je možné vizualizovať pomocou fyzikálnych zákonitostí?
- **VO3:** Ako efektívne vybrať a spojiť existujúce vizualizačné metódy využívajúce fyzikálne zákonitosti, pre účel interaktívnej analýzy dát?
- **VO4:** Ako zahrnúť používateľ-a-analytika do procesu získavania znalostí z dát, pri zachovaní efektívneho a rýchleho spracovania dátových množín?
- **VO5:** Akým spôsobom zabezpečiť spoľahlivú interakciu používateľ-a-analytika s vizualizovanými dátami vo viacrozmernom priestore v prostredí obohatenej reality?

Aby sme mohli získať základné informácie pre zodpovedania definovaných výskumných otázok, bolo potrebné realizovať viacero experimentov, vytvoriť niekoľko funkčných prototypov a algoritmov na overenie výsledkov v herných prostrediach.

Podľa výsledkov overenia predpokladov pre získanie odpovedí na formulované výskumné otázky je zrejmé, že samotný prevod algoritmov z 2D do 3D nie je praktický. Vizualizácia

² <https://unity.com>

³ <https://www.unrealengine.com>

veľkého objemu dát sa podľa výsledkov experimentov javí ako realizovateľná, za predpokladu, že dáta sú správne upravené a spracované, a zároveň použité vizualizačné techniky a postupy zodpovedajú štruktúre datasetu.

Pre upresnenie problému **skúmania vizualizácie dát v 3D priestore za použitia herných prostredí s využitím silou riadených algoritmov**, sme zadefinovali nasledovné tézy:

- **Návrh algoritmov pre využitie fyzikálnych herných prostredí na efektívnu vizualizáciu veľkých dátových množín.** Vzhľadom na špecifiká fyzikálneho prostredia je potrebné vytvoriť nové algoritmy, modifikácia existujúcich 2D algoritmov nie je na základe vlastných úvodných experimentálnych zistení postačujúca.
- **Vytvorenie metodiky pre prípravu dát na vizualizáciu v 3D fyzikálnom hernom prostredí.** Vychádzame z predpokladu, že vhodná modifikácia vstupných dát prispieva k efektívnemu fungovaniu vizualizačných algoritmov.
- **Využitie vstupných zariadení poskytujúcich 3D vstupy** (Leap motion, CAVE, rozšírená realita a pod.) pre zabezpečenie efektívnej interakcie s 3D vizualizáciou.

Unifikovaná práca s dátami a ich vizualizáciou v 3D priestore je stále predmetom aktívneho výskumu, pričom mnohé zdroje sú zamerané na úzko špecifikované problémy, ktoré nie sú aplikovateľné v širšom rozsahu. **Tézy dizertačnej práce boli definované tak, aby ich výsledok mohol reprezentovať jednotný postup vytvárania vizualizácie, pre rôzne štruktúry dát.** Interakcia používateľa predstavuje v celom postupe podstatnú rolu, preto sme sa rozhodli zakomponovať nástroje poskytujúce dostatočnú a intuitívnu manipuláciu s 3D priestorom ako jeden z predmetov výskumu.

Súčasný stav riešenej problematiky doma a v zahraničí

Téma dizertačnej práce patrí do vednej oblasti vizualizácie dát a informácií. Táto oblasť je špecifická tým, že z časti presahuje do kreatívnej oblasti a dizajnu. Zameranie na prepojené dáta na druhej strane eliminuje potrebu kognitívnej vedy a kvalitatívnych metód pre posudzovanie výslednej vizualizácie a poskytuje možnosť zamerania sa na grafovú analýzu štruktúry dát.

Problematika analýzy sietí (angl. *network analysis*) je špecifickou oblasťou bádania. Napriek relatívne novej definícii vychádza základ vedy o sieťach z teórie grafov a štatistiky. Napriek skutočnosti, že vizualizácia dát a vizuálna analýza dát je súčasťou veľkého množstva nástrojov, je celkové rozpracovanie danej problematiky zväčša limitované, ako bolo spomenuté, aplikačnou doménou resp. špecificky riešeným problémom. V rámci teoreticko-analytickej časti sme preto postupovali najmä experimentálnym spôsobom s aplikáciou rámcových zistení uvedených nižšie.

Analyzované oblasti výskumu vizualizácie sa zameriavajú na odstránenie limitov konvenčných metód na reprezentáciu údajov (často statických a s úzkym rozsahom) tak, aby bolo možné získať dostupné informácie nachádzajúce sa v analyzovaných údajoch [8] [9].

Optimálny spôsob, ako dosiahnuť takéto pokročilé vylepšenia, je využiť potenciál interaktivity na podporu komplexnej analýzy údajov. Takéto interaktívne systémy obohatené o integrované vizuálne rozhrania (synergia modulov pre simuláciu, analýzu a vizualizáciu) sú navrhnuté s cieľom prekonať obmedzenú ľudskú analytickú kapacitu na syntetizovanie množstva rôznorodých údajov do informačných náhľadov [10].

Vizuálna analýza

Samotný proces interaktívnej vizuálnej analýzy údajov je známy ako *vizuálna analýza* (angl. *Visual Analytics* -VA) [11]. VA spája postupy, ako je dolovanie údajov a štatistické techniky na vizualizáciu a syntetizáciu informácií intuitívnym spôsobom, pričom ďalej využíva vizuálne vnímanie ľudského oka na ich formovanie do základných vedomostí.

Prepojenie vizualizácie s interaktívnymi funkciami je kľúčovým aspektom pri navrhovaní VA systémov. Ťažisko takéhoto prepojenia sa bežne kladlo na 2D zobrazenia obsahujúce abstraktné údaje. Tieto techniky sú celkom vhodné pri zvažovaní vizualizácie spojitých číselných atribútov s cieľom odhaliť potenciálne vzťahy, vzory a korelácie medzi týmito atribútmi. 2D reprezentácie však nie sú vhodné pre všetky úlohy a v určitých kontextoch nefungujú vôbec. Napríklad pri analýze geopriestorových údajov dochádza v takýchto vizualizáciách k prekrytiu (angl. *occlusion*), čo vedie k strate informácií [12]. Je to najmä kvôli veľkému počtu dátových bodov, ktoré sa navzájom prekrývajú, čo sťažuje ľudskému analytikovi vidieť jednotlivé dátové body a robiť závery o potenciálnych vzťahoch. Toto môže potenciálne narušiť vizuálne vnímanie a závery, ktoré sa týkajú: objavovania (lokalizácie geometrického znázornenia), prístupu (získavania tvaru, farby, resp. inej vlastnosti), priestorových vzťahov (vzájomný vzťah geometrických zobrazení) [13].

Prechod od 2D k 3D vizualizáciám

V súčasnosti však narastá potreba kombinovať 2D a 3D vizualizácie a ich vlastné navigačné a prieskumné techniky. Kombinácia 3D pohľadov s tradičnými 2D vizualizáciami dát prostredníctvom viacerých prepojených modulov ponúka viacero výhod. Keďže sa pri skúmaní údajov nespoliehame iba na jednu rovinu reprezentácie, prieskum údajov sa stáva efektívnejším. Pri konceptualizácii takýchto integrovaných vizuálnych systémov je však potrebná aj určitá úroveň sofistikovanosti.

Ciele konceptualizácie je možné zhrnúť do piatich odlišných cieľov dizajnu, ktoré silne závisia od základných údajov a analyzovanej domény [15]:

- *Vysoký výkon a kvalita 3D geometrického vykresľovania* - dosiahnutie vysoko detailnej kontextovej geometrie.
- *Možnosti multivariačnej analýzy* - integrácia rôznych typov vizualizačných diagramov na podporu rôznych analytických metód.
- *Adekvátny výpočet a kvantifikácia priestorových javov* - napr. realistické mapovanie tieňov a odrazové vlastnosti.
- *Spoločný priestorový kontext pre geometrické reprezentácie* - napr. použitie jednotného súradnicového systému.
- *Úzka integrácia medzi geometriou a jej atribútovými aspektmi* - v princípe ide o integráciu cieľov 1 a 2.

Aplikácia vizualizácií sa venujú práce [16] [17] [18], zameriavajú sa však na štandardné 2D vizualizácie informácií prakticky použiteľné v reportingu a business-intelligence aplikáciách. Väčšina publikácií z oblasti 3D vizualizácií sa venuje vedeckým vizualizáciám aplikovaným zväčša ako výstup komplexných modelov z oblasti strojárstva alebo stavebníctva [19]. Pre oblasť sieťovej analýzy a vizualizácie, najmä v nástroji Gephi, sa ako jeden z najkvalitnejších zdrojov informácií javí stránka a publikácie Martina Grandjeana [20].

Cieľ práce

Cieľom práce je vytvorenie metodiky a algoritmov pre trojrozmernú vizualizáciu dát (3D vizualizáciu), ktoré je možné reprezentovať v podobe grafov označovaných aj ako **prepojené dáta** resp. **siete**.

Cieľom 3D vizualizácie je poskytnúť vizuálnu reprezentáciu zložitej vnútornej štruktúry sietí, uľahčiť pochopenie a analýzu sieťových štruktúr a odhaliť vzory a vzťahy v rámci sieťových údajov. 3D vizualizácia umožňuje používateľom skúmať a interagovať so sieťou v trojrozmernom priestore, čím poskytuje komplexnejší a intuitívnejší spôsob práce. Pomocou vizualizácie siete v 3D môžu používatelia získať prehľad o priestorovom usporiadaní uzlov a spojení, identifikovať zhluky alebo komunity v rámci siete a identifikovať akékoľvek vzory

alebo anomálie, ktoré nemusia byť ľahko rozpoznateľné v tradičnom 2D zobrazení. Okrem toho môže 3D vizualizácia siete podporovať prieskum dynamických sietí, kde sa spojenia a vzťahy medzi uzlami časom menia. Celkovo je cieľom 3D vizualizácie siete zlepšiť porozumenie a analýzu zložitých sietí, čo umožní používateľom robiť informované rozhodnutia a získať hlbší prehľad o analyzovaných údajoch.

Predpokladom dizertačnej práce bola hypotéza, že problém analýzy veľkého množstva dát je možné čiastočne zjednodušiť a skvalitniť použitím vizualizačných techník s cieľom vytvorenia lepšej predstavy o charaktere skúmaných dát. V rámci práce sme preto:

- v teoreticko-analytickej časti preskúmali súčasné možnosti vizualizácie informácií s ohľadom na efektívny spôsob ich reprezentácie v 2D a 3D,
- v metodicko-aplikačnej časti vytvorili metodiku a algoritmy na základe výsledkov a zistení a nástrojov pre tvorbu počítačových hier.

Pre implementáciu samotnej vizualizácie je nutné brať do úvahy relatívnu zložitosť vytvárania vizualizácií vo viacrozmerných priestoroch najmä s ohľadom na potenciálny objem spracovávaných dát. Ľudské vnímanie má tiež svoje limity definujúce množstvo informácií, ktoré je človek schopný spracovať zároveň.

Metodika práce a metódy skúmania

Zo štatistického hľadiska existuje metodologický základ pozostávajúci z postupov a nástrojov, ktoré sú spoločné pre špecifickú netriviálnu podskupinu výskumných oblastí súvisiacich so sieťovou vedou. Okrem toho je možné a celkom užitočné kategorizovať špecifické úlohy analýzy sieťových údajov v rôznych doménach podľa štatistickej taxonómie [48]. V súlade s takouto taxonómiou budeme postupovať od deskriptívnych metód k metódam modelovania a vyvodzovania, pričom tieto sú rozdelené do dvoch podoblastí, ktoré zodpovedajú modelovaniu a definícií samotných sietí a procesom v sieťach.

Charakteristika objektu skúmania

Graf $G = (V, E)$ je matematická štruktúra pozostávajúca z množiny vrcholov V (tiež označovaných uzly) a množiny hrán E (nazývaných tiež *linky*), kde prvky E sú neusporiadané dvojice $\{u, v\}$ rôznych vrcholov, kde $u, v \in V$.

Počet vrcholov $N_v = |V|$ a počet hrán $N_e = |E|$ sa nazýva *stupeň* resp. *veľkosť* grafu G .

Vrcholy grafu sú často bez straty všeobecnosti jednoducho označované celými číslami $1, \dots, N_v$ a hranami.

Na základe typu hrán môžeme grafy rozdeliť na dva základné typy:

- *neorientovaný graf* – je graf, v ktorom hrany nie sú orientované a teda neexistuje poradie incidentných uzlov (prislúchajúcich hrane).
- *orientovaný graf* – je graf, v ktorom sú hrany orientované (existuje usporiadanie uzlov prislúchajúcich danej hrane). Pre hrany orientovaného grafu platí, že

$$E \subseteq \{ (x, y) \mid (y, x) \in V^2, \text{ kde } x \neq y \}$$

V prípade, že sú hranám priradené špecifické hodnoty (*váhy*), hovoríme o ohodnotenom grafe (resp. konštantnom v prípade nedefinovaných alebo rovnakých váh pre hrany).

Reprezentácia grafu

Informácie na vytvorenie grafu je potrebné definovať v určitej štruktúre. Je potrebné si uvedomiť, že väčšina sietí, ktoré sa v praxi používajú v procese sieťovej analýzy majú

niekoľko desiatok až miliónov vrcholov a hrán. Príkladom môže byť dataset *Pokec*⁴, ktorý bol v rámci predchádzajúceho výskumu vytvorený pred niekoľkými rokmi, obsahuje viac ako 1,6 milióna vrcholov a viac ako 30 miliónov hrán. Údaje o vrcholoch a hranách je preto potrebné mať uložené v dátovom súbore. Základná reprezentácia grafu vyžaduje existenciu troch zoznamy: zoznam susedov, zoznam hrán a matice príľahlosti.

Matica príľahlosti (angl. *adjacency matrix*) reprezentuje štruktúru, ktorá obsahuje informáciu a príslušnosti uzlov hranám. Alternatívnou štruktúrou je **zoznam príľahlosti** (zoznam hrán).

Charakteristiky vrcholov

Vrchol je základnou štruktúrou, ktorá bude predmetom skúmania. V procese vizualizácie zodpovedá vrchol špecifickej, väčšinou zásadnej skupine charakteristík spracovávaných údajov. Nie je však možné presne definovať, či určitý atribút v údajoch má byť modelovaný ako vrchol alebo hrana.

Charakteristiky hrán

Všetky doteraz diskutované súhrnné miery (t. j. stupeň a iné, všeobecnejšie pojmy centrálnosti) sú definované pre vrcholy. V praxi sa najčastejšie modelujú zásadné vlastnosti vrcholmi, čo je dôvodom definície metrík grafu najmä pre vrcholy. Niektoré otázky na vlastnosti grafu sú však prirodzene spojené s hranami. Príkladom môže byť otázka na to, ktoré väzby na sociálnej sieti sú najdôležitejšie pre šírenie informácií alebo fám. Centrálnosť hrany, ktorá analogicky rozširuje centrálnosť vrcholu priradením hodnoty, ktorá odráža počet najkratších ciest prechádzajúcich danou hranou, je potom vhodná metrika, ktorú môžeme použiť [48].

Charakteristiky sietí

Zložitejšia situácia je v prípade charakteristiky grafu resp. siete ako celku. Väčšina otázok v sieťovej analýze sa obmedzuje na otázky týkajúce sa **súdržnosti siete**, čiže do akej miery sú podmnožiny vrcholov súdržné („zlepené“) vzhľadom na vzťah definujúci hrany v sieťovom grafe. V literatúre sa táto charakteristika označuje anglickým termínom *kohézia* (angl. *network cohesion*).

Pracovné postupy

V tejto kapitole sa budeme zaoberať rôznymi pracovnými postupmi v oblasti analýzy sietí. Sieťová analýza je dôležitým nástrojom pre pochopenie vzťahov a štruktúry v rôznych komplexných systémoch.

Vizualizácia a charakteristika sietí

Deskriptívna (opisná) analýza údajov je základnou metódou skúmania charakteristík sietí a dát. Spoločne s numerickou charakterizáciou siete sú zvyčajne jedným z prvých krokov v sieťovej analýze. V skutočnosti tvoria deskriptívne analýzy pravdepodobne väčšinu publikovaných sieťových analýz [48].

Modelovanie sietí a odvodzovanie

Samotné sieťovému modelovaniu je potrebné venovať dostatočnú pozornosť. Vo všeobecnosti existujú dve triedy sieťových modelov: matematické a štatistické. *Matematickými modelmi* rozumieme modely špecifikované prostredníctvom (typicky) jednoduchých pravdepodobnostných pravidiel na generovanie siete, kde sú pravidlá často

⁴ <https://snap.stanford.edu/data/soc-pokec.html>

definované v snahe zachytiť konkrétny mechanizmus alebo princíp (napr. „bohatí zbohatnú“). Naproti tomu *štatistickými modelmi* rozumieme modely (často aj pravdepodobnostné) špecifikované aspoň čiastočne so zámerom prispôbiť ich pozorovaným údajom – umožňujúce napríklad posúdenie vplyvu určitých premenných na tvorbu hrán v sieti. Aj keď medzi týmito dvoma triedami modelov určite existuje určitý prienik, literatúra venujúca sa konkrétnej problematike je v zásade samostatná pre každú triedu zvlášť [48].

V rámci praktickej časti dizertačnej práce nám vytvorenie nástroja pre vizualizáciu sietí na základe fyzikálnych zákonitostí umožňuje vytvárať siete priamo bez nutnosti matematického resp. štatistického prístupu. Ide teda o tretí spôsob modelovania sietí, ktorý je možný vďaka využitiu fyzikálnych herných prostredí. Pre overovanie výslednej vizualizácie však použijeme štandardné metódy pre overovanie a popis sietí, ktoré popíšeme v časti venujúcej sa vyhodnocovaniu výsledkov.

Procesy v sieťach

Samostatnou oblasťou pre hodnotenie sietí je pohľad na procesy v sieťach. Ako objekty na reprezentáciu interakcií medzi prvkami komplexného systému sú sieťové grafy často hlavným cieľom sieťovej analýzy. Na druhej strane však v mnohých prípadoch ide o konkrétnu veličinu (alebo atribút) spojenú s každým z prvkov v systéme, ktorá je v konečnom dôsledku najzaujímavejšia. Napriek tomu v prostredí sietí môžeme očakávať, že táto veličina bude významným spôsobom ovplyvnená interakciami medzi prvkami, a preto môže byť sieťový graf stále relevantný pre modelovanie a analýzu. Formálnejšie si môžeme predstaviť stochastický proces ako „žijúci“ v sieti a indexovaný podľa vrcholov v sieti. Rôzne otázky týkajúce sa takýchto procesov možno interpretovať ako problémy predikcie buď statických alebo dynamických sieťových procesov [48].

Spôsob získavania údajov a ich zdroje

Údaje potrebné pre dosiahnutie cieľov dizertačnej práce je možné získať viacerými spôsobmi. Pre overovanie použitia navrhnutých metód a algoritmov použijeme základné, porovnávacie/benchmarkové datasety verejne dostupné na stránkach niektorých nástrojov (Gephi), na stránkach Kaggle a pre veľké prepojené dáta použijeme dataset z predchádzajúceho výskumu [49] verejne dostupný na portáli Snap Stanfordskej univerzity. Použitie verejne dostupných datasetov poskytuje možnosť replikácie experimentov a výsledkov uvedených v dizertačnej práci, ako aj možnosť overenia dosiahnutých záverov. Sumárny prehľad datasetov je uvedený v nasledovnej tabuľke.

Tabuľka 1: Zoznam skúmaných datasetov

Názov	Popis	Platforma	Počet	
			záznamov	atribútov
Game of Thrones	Hry o tróny – vzťahy postáv (5 kníh)	Kaggle	682-1008 ⁵	25
Witcher	Zaklínač – fantasy románová séria	Kaggle	2600	12
Lazega	Sieť vzťahov v právnickej firme	R	36	8
Network Science	Sieť spoluautorov	Gephi	1589	3
Pokec	Dáta sociálnej siete	Snap	1632803	7

Výsledky práce

Viaceré najlepšie algoritmy na výpočet rozložení jednoduchých neorientovaných grafov patria do triedy známej ako **silovo riadené algoritmy**. Takéto algoritmy, známe tiež ako pružinové algoritmy, **vypočítajú rozloženie grafu iba pomocou informácií obsiahnutých v samotnej**

⁵ Dataset obsahuje 5 špecifických datasetov, každý zodpovedajúci príslušnej knihe Hry o tróny.

štruktúre grafu, namiesto toho, aby sa spoliehali na znalosti špecifické pre danú doménu definované vrcholmi, hranami resp. atribútmi. Grafy vykreslené pomocou týchto algoritmov sú zo svojej podstaty esteticky príjemné, vykazujú symetriu a majú tendenciu vytvárať rozloženia bez kríženia pre rovinné grafy. Prehľadové práce v tejto oblasti publikoval napr. Di Battista a kol. [59] resp. Brandes [60].

Jednou z prvých silovo orientovaných metód kreslenia (vizualizácie) grafov je algoritmus založený na barycentrických reprezentáciách [61]. Tradičnejšie sa metóda rozloženia pomocou silovej reprezentácie používa podľa Eadesa [62], najčastejšie je však možné pracovať s implementáciou algoritmu podľa Fruchtermana a Reingolda [34]. Obe implementácie sa spoliehajú sily pružiny definované analogicky ako v Hookovom zákone⁶. Pri týchto metódach existujú odpudivé sily medzi všetkými uzlami a zároveň príťažlivé sily medzi uzlami, ktoré sú vzájomne incidentné [63].

Pri riešení dizertačnej práce sme sa zamerali na zmienené silovo riadené algoritmy z dôvodu využitia fyzikálnych možností herných prostredí pre prácu so štruktúrou grafu bez potreby analýzy grafu pomocou štandardných grafových metrík. **Prvú experimentálnu implementáciu sme vytvorili na základe vyššie definovaných pravidiel pre implementáciu silovo riadených algoritmov s tým, že sme použili implementáciu v 3D priestore definovanom natívne v nástroji Unity3D.**

Vizuálna analýza v dvojrozmernom priestore

Výhody silou riadených algoritmov v 2D priestore spočívajú v ich jednoduchosti a zrozumiteľnosti. Táto forma vizualizácie poskytuje rýchle a intuitívne prezentácie dát, čo je výhodné pre rôzne aplikácie vizualizácií sociálnych sietí alebo dátových štruktúr. Napriek týmto výhodám však 2D priestor nesie aj nevýhody. Obmedzený priestor môže viesť k prekrywaniu vrcholov a hrán, čo znižuje čitateľnosť grafu. Taktiež, nedostatok realizmu môže obmedziť schopnosť zachytiť detaily vzťahov a vzdialeností medzi vrcholmi. Sily medzi uzlami možno vypočítať na základe ich teoretických vzdialeností v grafe, určených dĺžkami najkratších ciest medzi nimi. Napríklad algoritmus Kamada a Kawai [64] využíva sily pružiny úmerné teoretickým vzdialenostiam grafu. Vo všeobecnosti metódy orientované na silu definujú účelovú funkciu, ktorá mapuje každé rozloženie grafu na číslo v $\mathbb{R}^+$, ktoré predstavuje energiu rozloženia. Táto funkcia je definovaná tak, že nízke energie zodpovedajú usporiadaniám, v ktorých sú susedné uzly blízko od seba na základe určitej vopred špecifikovanej vzdialenosti a v ktorých sú nesusediace uzly dobre rozmiestnené [63].

Výhody vizualizácie v 2D priestore

Tradičný spôsob vizualizácie grafov a sietí, vrátane veľkých prepojených dát má niekoľko výhod. Na základe charakteru algoritmov a obmedzení zobrazovacej plochy sa však s rastúcou zložitou vzťahov resp. veľkosťou siete tieto výhody postupne strácajú. Medzi zásadné pozitíva 2D vizualizácie patrí:

- *Jednoduchosť a zrozumiteľnosť:* V 2D priestore sú silou riadené grafy často jednoduché a ľahko zrozumiteľné. Ide najmä o prípad grafov do veľkosti desiatok vrcholov.
- *Rýchlosť výpočtu:* Algoritmy pre 2D priestor majú tendenciu byť rýchlejšie v porovnaní s ich 3D ekvivalentmi. Existujú aj algoritmy, ktoré je možné jednoducho paralelizovať, čo ďalej zvyšuje rýchlosť vytvorenia výslednej vizualizácie.
- *Široké využitie:* V 2D priestore sú silou riadené algoritmy vhodné pre širokú škálu aplikácií.

⁶ Hookov zákon opisuje vzťah medzi deformáciou pevného telesa vyvolanou pôsobením napätia a veľkosťou tohto napätia.

Nevýhody vizualizácie v 2D priestore

Pri analýze vizualizácií v 2D priestore sme sa zamerali najmä na identifikáciu nedostatkov dvojrozmernej vizualizácie, ktoré je možné odstrániť pridaním ďalšieho rozmeru a implementáciou silovo orientovaného prístupu s využitím existujúcich možností fyzikálnych prostredí. Medzi základné nedostatky 2D vizualizácie patria:

- *Obmedzený priestor:* V 2D priestore môže dôjsť k prekryvaniu vrcholov a hrán, čo môže zhoršiť čitateľnosť grafu. Tento nedostatok je zásadný v prípade existencie zhlukov, ktorých priemer je malý.
- *Nedostatok realizmu:* 2D vizualizácia nemusí plne zachytiť vzťahy a vzdialenosti medzi vrcholmi. Ide najmä o prípady, kedy je potrebné uvažovať o dĺžke hrán ako o reálnej hodnote vyjadrujúcej napríklad priestorovú dĺžku.
- *Obmedzené možnosti interakcie:* V prípade blízko ležiacich vrcholov nemusí byť zrejmá potreba interakcie s užívateľom napr. vykonanie operácie rotácie scény pre získanie lepšej informácie o vzájomných vzťahoch v grafe.

Metódy a algoritmy vizualizácie v 2D

Základné silovo orientované prístupy majú zásadné obmedzenie dané počtom vykresľovaných/spracovávaných vrcholov. Ich použitie je obmedzená na malé grafy a výsledky sú slabé pre grafy s viac ako niekoľkými stovkami vrcholov. Existuje viacero dôvodov, prečo tradičné algoritmy orientované na silu nefungujú dobre pre veľké grafy. Jednou z hlavných prekážok ich škálovateľnosti je skutočnosť, že fyzický model má zvyčajne veľa lokálnych miním. Aj s pomocou sofistikovaných mechanizmov na vyhýbanie sa lokálnym minimám nie sú základné algoritmy orientované na silu schopné konzistentne vytvárať dobré rozloženia pre veľké grafy. Rovnako barycentrické⁷ metódy nefungujú dobre pre veľké grafy, hlavne kvôli problémom s rozlíšením (pre veľké grafy je minimálna separácia vrcholov veľmi malá, čo vedie k nečitateľným vizualizáciám) [63].

Koncom 90-tych rokov sa objavilo niekoľko techník rozširujúcich funkčnosť silovo riadených metód na grafy s desiatkami tisíc a dokonca stovkami tisíc vrcholov. Hlavnou ideou týchto prístupov je technika viacúrovňového rozloženia, kde je graf reprezentovaný sériou postupne sa zjednodušujúcich štruktúr a usporiadaných v poradí od najjednoduchších po najzložitejšie. Tieto štruktúry môžu byť hrubšie grafy (ako v prístupe Hadanyho a Harela [65], Harela a Korena [66] a Walshawa [67] alebo vertexové⁸ filtrácie ako v prístupe Gajera, Goodricha a Kobourova [68]).

Vizuálna analýza v trojrozmernom priestore

Vizualizácia dát v trojrozmernom priestore zohráva kľúčovú úlohu v analýze a porozumení komplexných vzťahov v rôznych oblastiach výskumu a aplikácií. Nové výzvy v oblasti vizualizácie dát vyžadujú neustále zdokonaľovanie existujúcich algoritmov a vytváranie nových prístupov, ktoré zabezpečia efektívne a intuitívne vnímanie trojrozmerných dátových štruktúr. V tejto kapitole predstavujeme inovatívny algoritmus založený na princípe silou riadených algoritmov, ktorý umožňuje vizualizáciu v 3D priestore, ale tiež integruje sférické zobrazenia, normalizáciu, zhlukovanie, pružinové spoje a gravitačné interakcie pre vytvorenie robustného nástroja na vizualizáciu komplexných dátových sietí.

Algoritmy sú široko používané v oblasti vizualizácie dát a sú založené na princípe simulovania fyzikálnych síl medzi vrcholmi grafu. Ich efektívnosť v dvojrozmerných priestoroch je dobre zdokumentovaná, ale prenos týchto prístupov do trojrozmerného prostredia si vyžaduje nové techniky a optimalizácie. V našom prístupe sme zdokonalili

⁷ Barycentrum je ťažisko dvoch alebo viacerých telies, ktoré sa navzájom obiehajú.

⁸ vertex – vrchol (angl., lat.)

tradičné silou riadené algoritmy o sférické zobrazenia, čo umožňuje lepšiu reprezentáciu vzťahov medzi vrcholmi v trojrozmernom priestore.

Okrem toho sme implementovali normalizačné techniky, ktoré umožňujú efektívnu manipuláciu s rozsahom hodnôt a zlepšujú konzistenciu vizuálnej prezentácie. Využitie zhlukovania nám umožňuje identifikovať a vizualizovať podskupiny dát s podobnými charakteristikami, zatiaľ čo pružinové spoje a gravitačné interakcie pridávajú dodatočnú dynamiku do vizualizácie, čím zabezpečujú lepšiu interpretáciu štruktúry dátového grafu. V nasledujúcich častiach kapitoly sa podrobnejšie zameriame na opis navrhnutých algoritmov, ich implementáciu a výsledky, ktoré sú dosiahnuté pri vizualizácii rôznorodých dátových sietí v trojrozmernom priestore. Tento prístup poskytuje nový pohľad na vizualizáciu dát, zohľadňujúc nielen geometrické aspekty, ale aj hierarchické a štrukturálne vlastnosti dát.

Silou riadené algoritmy v 3D priestore prinášajú výhodu presnejšieho zobrazenia vzdialeností medzi vrcholmi. Táto možnosť umožňuje vytváranie vizualizácií, ktoré sú čitateľnejšie a štruktúrovanejšie, čo je významné pre detailné prieskumy dátových vzorov. S výhodami v 3D priestore prichádzajú aj nevýhody. Algoritmy v 3D priestore sú zložitejšie z výpočtového hľadiska, čo môže spôsobovať väčšiu výpočtovú záťaž. Zvýšená komplexita vizualizácie môže zase viesť k strate detailov a zvýšenej náročnosti interakcií s prostredím.

Výhody v 3D Priestore:

- **Presná vzdialenosť:** V 3D priestore majú silou riadené algoritmy väčší priestor na presné zobrazenie vzdialeností medzi vrcholmi.
- **Lepšia čitateľnosť:** Vďaka možnosti rozmiestňovať vrcholy v trojrozmernom priestore môže byť vizualizácia čitateľnejšia a štruktúrovaná.

Nevýhody v 3D Priestore:

- **Zložitejší výpočet:** Algoritmy pre 3D priestor sú často náročnejšie z výpočtového hľadiska.
- **Zvýšená komplexita vizualizácie:** Zväčšením priestoru sa môže zvýšiť aj komplexita vizualizácie.

Väčší Počet Interakcií: Užívatelia môžu potrebovať väčší počet interakcií na presné pohľady a manipuláciu s trojrozmerným priestorom.

Metodika prípravy dát pre vizualizácie v 3D

Klasické silou riadené algoritmy sú obmedzené na výpočet rozloženia grafu v euklidovskej geometrii, zvyčajne $\mathbb{R}^2$, $\mathbb{R}^3$, niektoré novšie implementácie v $\mathbb{R}^n$ pre väčšie hodnoty n . Existujú však prípady, keď euklidovská geometria nemusí byť tou najlepšou voľbou, čo je aj výsledkom zistení dizertačnej práce. Určité grafy majú štruktúru, ktorú je lepšie vizualizovať s použitím inej geometrie, napríklad na povrchu gule alebo na anuloide. Najmä zobrazenia 3D sietí môžu byť zobrazené na guli za účelom mapovania textúr alebo uloženia na torus bez kríženia. Okrem toho sa zistilo, že určité neeuklidovské geometrie, konkrétne hyperbolická geometria, majú vlastnosti, ktoré sú obzvlášť vhodné na usporiadanie a vizualizáciu veľkých tried grafov [71] [72]. Na základe uvedených zistení Kobourov a Wampler opisujú rozšírenia silovo riadených algoritmov na Riemannove priestory [73].

Predspracovanie dát a siete ako celku je podstatnou časťou vizualizácie, pričom predspracovanie môže byť vykonávané s rôznym zameraním. Pri väčšine prípadov je potrebné dáta predspracovať z dôvodu vyčistenia chybných hodnôt, normalizácie alebo iných operácií, ktoré následne uľahčia prácu z dátami ako opisujú autori [8]. Druhým možným dôvodom vykonania predspracovania dát môže byť snaha o odľahčenie algoritmov slúžiac na výpočet informácií z dát, prehľadávanie alebo odľahčenie hardvérových prostriedkov pri ďalšom spracovaní. Metodika je podrobne spísaná v dizertačnej práci v prílohe A.

Predspracovanie siete

V dnešnej dobe existuje mnoho algoritmov, ktoré majú za úlohu nastaviť zrozumiteľný a prehľadný tvar siete za pomoci algoritmov na výpočet základných charakteristických metrík siete. Výsledné vypočítané hodnoty sú následne použité ako oporný bod pre algoritmy rozloženia siete v priestore. V 2D priestore je tento prístup pomerne bežný a jednoducho aplikovateľný na rôzne druhy dát, v prípade 3D nám ale samotné vypočítanie metrík nestačí. Metriky určujúce vlastnosti siete v 3D priestore môžu byť zavádzajúce, preto je potrebné správne vybrať metriky, s ktorými algoritmy dokážu pracovať efektívne aj v 3D. V prvej verzii našej práce používame silou riadené algoritmy, pre ktoré sú základné metriky len spôsob ako urýchliť výpočet konečnej polohy siete.

Príprava dát

Vizualizácia dát v trojrozmernom priestore zohráva kľúčovú úlohu v analýze a porozumení komplexných vzťahov v rôznych oblastiach výskumu a aplikácií. Nové výzvy v oblasti vizualizácie dát vyžadujú neustále zdokonaľovanie existujúcich algoritmov a vytváranie nových prístupov, ktoré zabezpečia efektívne a intuitívne vnímanie trojrozmerných dátových štruktúr. V tejto kapitole predstavujeme inovatívny algoritmus založený na princípe silou riadených algoritmov, ktorý umožňuje vizualizáciu v 3D priestore, ale tiež integruje sférické zobrazenia, normalizáciu, zhlukovanie, pružinové spoje a gravitačné interakcie pre vytvorenie robustného nástroja na vizualizáciu komplexných dátových sietí.

Prvotným cieľom je zabezpečiť, aby hodnoty hrán vo vizualizovanom grafe boli vhodne normalizované, čím sa maximalizuje interpretovateľnosť a efektívnosť výslednej 3D vizualizácie. Normalizácia dát je predovšetkým potrebná pri parametroch použitých na reprezentáciu síl a pružín, pretože bez normalizácie môžu nastať priestorové komplikácie s vrcholmi, na ktoré je aplikovaná príliš veľká sila.

Škálovanie hodnôt hrán

Škálovanie hodnôt hrán je proces úpravy hodnôt, ktoré reprezentujú vzťahy medzi vrcholmi v grafe. Cieľom tohto procesu je dosiahnuť rovnomerný rozsah hodnôt, ktorý je vhodný pre vizualizáciu a analýzu dátového grafu. Škálovanie hodnôt hrán môže byť dôležité, pretože umožňuje lepšie porovnávanie a interpretáciu vzťahov medzi vrcholmi bez toho, aby sa hodnoty nespravodlivo vynášali v dôsledku ich veľkých variácií. Škálovanie hodnôt hrán je nevyhnutné pre dosiahnutie konzistentných a interpretabilných výsledkov pri vizualizácii v 3D.

Normalizácia parametrov pre hodnoty hrán

Normalizácia parametrov pre hodnoty hrán je proces, ktorý má za cieľ dosiahnuť jednotný a štandardizovaný rozsah hodnôt parametrov, ktoré reprezentujú vzťahy medzi vrcholmi v grafe. Tento postup sa často používa pri príprave dát pre vizualizáciu grafu, či už v 2D alebo 3D priestore. Normalizácia je dôležitá z viacerých dôvodov:

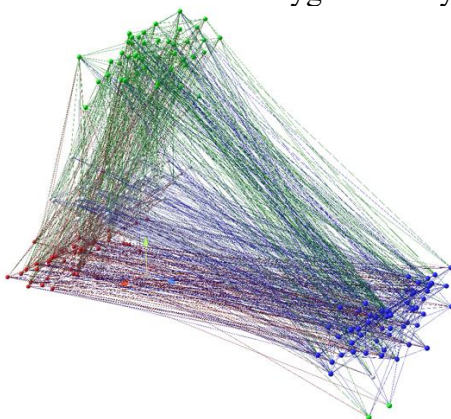
- **Zjednodušenie porovnávania:** Normalizované hodnoty umožňujú jednoduchšie porovnávanie vzťahov medzi vrcholmi, pretože sú škálované na štandardný interval.
- **Minimalizácia vplyvu extrémnych hodnôt:** Normalizácia pomáha eliminovať vplyv extrémnych hodnôt, ktoré by inak mohli nespravodlivo ovplyvniť vizuálne zobrazenie grafu.
- **Zabezpečenie konzistentnosti:** Normalizácia prispieva k dosiahnutiu konzistentného rozsahu hodnôt pre rôzne hrany grafu, čo uľahčuje ich interpretáciu.

Algoritmus 1: Pružinový algoritmus (Spring Layout Algorithm)

Algoritmus reprezentuje graf ako fyzický systém, kde uzly sú považované za nabitú časticu vzájomne spojené pružinami. Každý uzol je priradený dvom typom síl: príťažlivé sily a odpudivé sily. Cieľom algoritmu je znížiť celkovú energiu systému pružín presúvaním uzlov do nových polôh. Napriek tomu, že táto metóda produkuje čisté, symetrické usporiadanie pre grafy strednej veľkosti, považuje sa za náročný algoritmus. Jeho časová zložitosť je v mnohých prípadoch vysoká, čím sa stáva algoritmus nepoužiteľný. Okrem toho trpí nedostatočnou predvídateľnosťou, čo znamená, že opakované spustenie algoritmu môže viesť k rôznym výsledkom. To môže predstavovať problémy pri udržiavaní mentálnej mapy používateľa počas interakcie so nestabilnými usporiadaniami. Napriek týmto nevýhodám sa silou riadené algoritmy široko používajú v mnohých vizualizačných rámcoch. V našej práci bol algoritmus prepracovaný a optimalizovaný s cieľom odstrániť jeho charakteristické nevýhody hlavne pre využitie v 3D priestore.

Modifikácia 1

V prvotných experimentoch boli otestované pružinové spojenia medzi uzlami, ktoré reprezentujú príťažlivé sily. Zároveň zabezpečujú aj prislúchajúce odpudivé sily, pretože jedna z vlastností pružín je sťahovanie svojej dĺžky. Prvotný prototyp môžeme vidieť na Obrázok 1, kde je zobrazená vizualizácia náhodne vygenerovaných dát [74].



Obrázok 1 Zobrazenie záťažového experimentu

Uzly a hrany sú generované s náhodnými vlastnosťami, ako je farba a poloha, aby sa odstránila možná odchýlka spôsobená akýmkoľvek špecifickým súborom údajov. Náhodné pozície používame na odstránenie počiatočného skreslenia spôsobeného nahromadenou silou, ktorá sa aplikuje, keď sú uzly umiestnené na začiatku scény. Okrem toho na nasmerovanie hrán v grafe je uzlom priradená farba, ktorá predstavuje ich typ. Na základe typu počiatočného a koncového uzla sa vytvoria hrany a nastaví sa sila pružinového spoja. Pružinové spojenia umožňujú rýchle a dostatočne prehľadné zobrazenie grafu, pričom s použitím správnych nastavení kontrolujú rozpínanie uzlov v priestore. Ich nevýhoda spočíva v obmedzenosti vyvíjaných síl v jednom smere a zvýšenej záťaži pre výpočtovú jednotku. Každý pružinový kĺb musí byť v scéne reprezentovaný jedným herným objektom, čo zapríčiňuje zvýšenú záťaž nie len na fyzikálny ale aj vykresľovanie vizuálnych elementov (ang. rendering).

Algoritmus 2: Silou riadený algoritmus (Force Driven Algorithm)

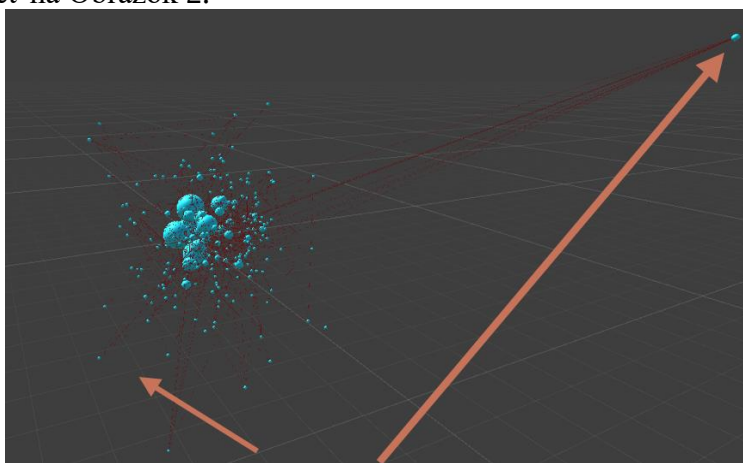
Rozloženie grafu v priestore je náročná činnosť a vyžaduje určité skúsenosti a zručnosti, ktoré je potrebné aplikovať na použité dáta. Mnohé používané algoritmy usporiadania využívajú ako základný parameter, kombináciu metrík opísaných v kapitole 0, pričom je potrebné predspracovanie dát a prípadne manuálne usporiadanie zobrazeného grafu. Silou riadené algoritmy využívajú fyzikálne vlastnosti a procesy na výpočet zrozumiteľného rozmiestnenia uzlov a hrán v priestore, pomocou pôsobiacich síl na jednotlivé elementy grafu.

Algoritmy rozloženia orientované na silu vytvárajú grafické zobrazenie použitím systému emulovaných fyzických síl. Gravitácia priraduje vrcholom hmotnosť v pomere k ich sieťovej centrálnosti, čo umožňuje, aby vrcholy, ktoré sú z hľadiska grafu viac centrálné, boli vizualizované vo fyzicky centrálnych miestach. Zmena mierky mení gravitačnú silu počas simulácie a znižuje prekríženie v porovnaní s gravitáciou bez mierky.

Silou riadené algoritmy poskytujú funkcionalitu, ktorá zabezpečuje rozloženie grafu v priestore za pomoci fyzikálnych vlastností jednotlivých uzlov a hrán. Výsledné rozloženie zobrazuje graf v ustálenom stave, kedy sú sily vyrovnané, pričom zaručuje v 2D priestore neprekrývanie uzlov.

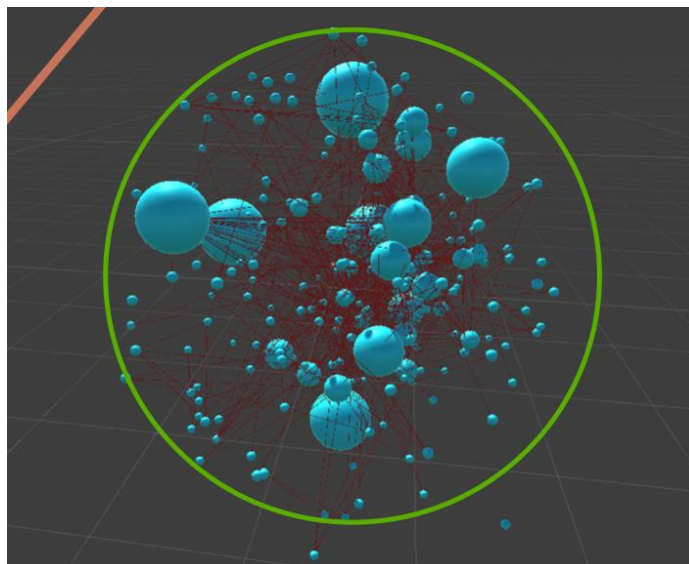
Hlavnou vlastnosťou silou riadených algoritmov je fyzikálna príťažlivosť a odpudivosť medzi jednotlivými uzlami reprezentované hranami grafu. Fyzikálne sily umožňujú vytvárať čitateľné a prehľadné rozmiestnenia hrán a uzlov v 3D priestore, bez nutnosti zásahu používateľa do procesu zostavovania grafu. Zvolená metóda aplikovania sily na jednotlivé hrany grafu, nám umožňuje kontrolovať rozpínanie grafu v 3D priestore, pre lepšie zachovanie prehľadnosti celej vizualizácie.

Keďže na uzly pôsobia rôzne sily, môže nastať situácia, kedy sa uzly zobrazia mimo kamery ako môžeme vidieť na Obrázok 2.



Obrázok 2 Vrcholy odstrčené silou za okraj vizualizácie

V takom prípade je aplikované ukotvenie resp. kotva, ktorá pôsobí na každý uzol v sieti a priťahuje ho smerom k základnému bodu. Aplikovaním kotvy dosiahneme stav ako na Obrázok 3, kde je celá sieť vizualizovaná v zornom poli virtuálnej kamery.



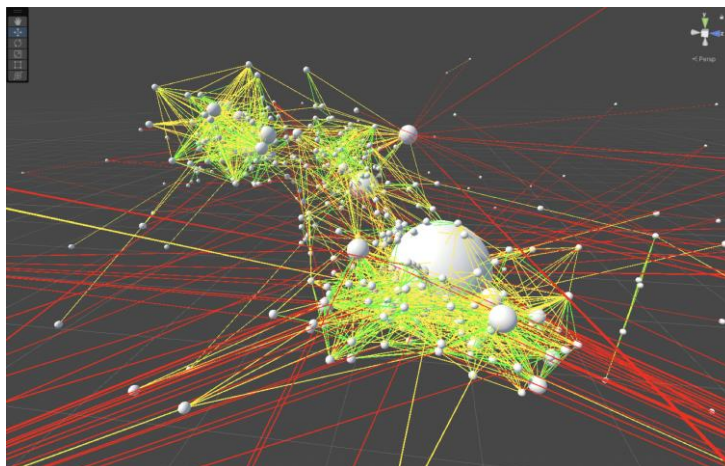
Obrázok 3 Ideálny stav síl

Aplikácia silou riadených algoritmov v 3D priestore je stále predmetom výskumu, pretože v trojrozmernom priestore tieto algoritmy nedokážu sami o sebe zabezpečiť zrozumiteľné zobrazenie z dôvodu pridanej z dimenzie, ktorej odobraním v 2D priestore vytvárame obmedzenie pre pohyb uzlov a hrán. Obmedzenie v podobe pridanej dimenzie je potrebné pretransformovať aj do trojrozmerného priestoru, ale v inej podobe, ktorá zabezpečí, že vizualizácia nebude zakrývať informácie a zároveň bude jasne rozoznateľná pre používateľa [7].

Algoritmus 3: Pružinovo-silový algoritmus

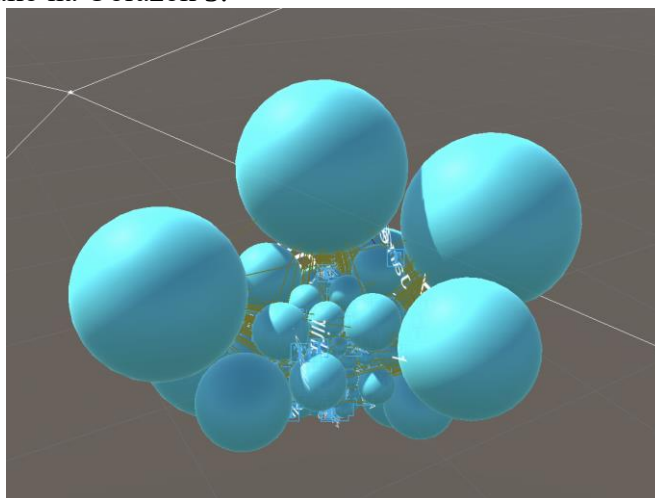
Silou riadený algoritmus a pružinový algoritmus v 3D priestore vykazujú určité nedostatky a obmedzenia. Jednoduchý pružinový algoritmus je extrémne citlivý a nedochádza k jeho konvergencii do stavu, kedy je možné akceptovať finálnu vizualizáciu. Vyznačuje sa nielen neustálym kmitaním, ale aj nevyváženým stavom celého 3D modelu, ktorý sa v čase často mení. Naopak, silou riadený algoritmus trpí nadmerným vytváraním zhlukov a problémom s odpudzovaním vrcholov. Kombináciou týchto dvoch definovaných algoritmov sme boli schopní eliminovať oba pozorované problémy. Vytvorili sme kombinovaný algoritmus pre štandardný euklidovský priestor a jeho modifikáciu so sférickou hranicou, ktorá slúži na zabezpečenie vyššej kompaktnosti a čitateľnosti výslednej vizualizácie.

Použitie pružín (pružinových spojov) vytvára systém, v ktorom sú vrcholy priťahované k sebe pôsobením atraktívnych síl. Tento prístup vedie k rýchlejšej stabilizácii siete, pretože pružiny zabezpečujú stabilnú prevádzku silového systému. V tomto prístupe sú hrany reprezentované atraktívnymi silami, ktoré k sebe ťahajú spojené vrcholy. Tento model je výhodný vďaka rýchlemu spájaniu vrcholov a dobrej internej implementácii vo fyzikálnych prostrediach.



Obrázok 4 Výsledná vizualizácia dát v aplikácii

V prípade kombinácie pružinového a silou riadeného algoritmu dosahujeme vyvážený prístup, kde pružiny a atraktívne sily spolupracujú na stabilizácii grafu. Pri nastavení parametrov síl a pružín je však dôležité nájsť vhodnú rovnováhu medzi rýchlym usadením siete a minimalizáciou zahltenia modelu. Kombinácia týchto postupov tiež zabráni vytvoreniu jedného veľkého zhluku, ktorý by mohol zhoršiť kvalitu výslednej vizualizácie. Spojením týchto dvoch metód docielime harmonickú rovnováhu medzi pružinovými silami a príťažlivými silami, čo vedie k stabilizovanej štruktúre grafu, ako je znázornené na Obrázok 4. Je dôležité vhodne nastaviť silové parametre a pružiny tak, aby dosiahli optimálnu rovnováhu medzi rýchlym ustálením siete a minimalizáciou neprehľadnosti modelu. Kombináciu postupov taktiež zamedzíme vytvoreniu jedného veľkého zhluku, ktorý by mohol narušiť vizualizáciu ako na Obrázok 5.



Obrázok 5 Vytvorený zhluk spôsobený aplikovanými silami

Algoritmus 4: Pružinovo-silový algoritmus s kontrolovaným rozmiestňovaním

Voľné zobrazenie v priestore je štandard pre 3D vizualizáciu, ale v prípade silou riadených algoritmov nastávajú situácie, v ktorých je potrebné nastaviť obmedzenia priestoru, sily a rozostavenia hrán a uzlov. Pri neregulovaných hodnotách síl môžeme pozorovať správanie grafu v dvoch hlavných prípadoch, ktoré sú neoptimálne pre vizualizáciu ako napríklad:

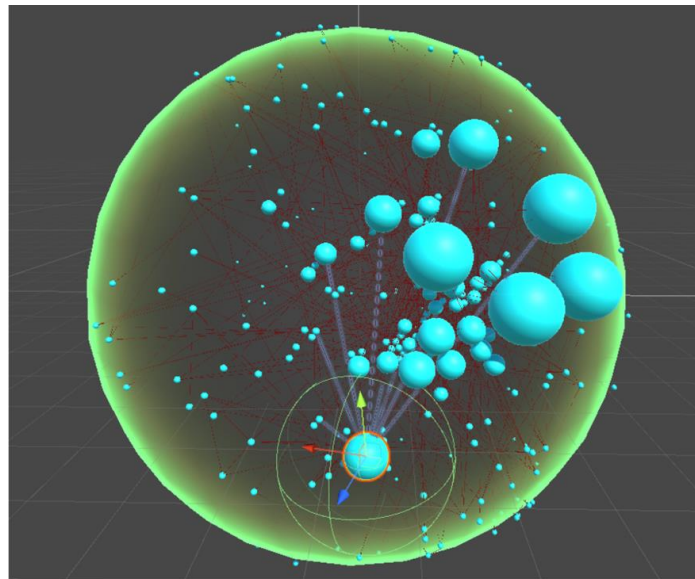
- Zobrazený graf vytvára jeden nečitateľný zhluk

- Vzdialenosť medzi uzlami sa neustále zväčšuje, čo znamená, že nikdy nedosiahne stabilného stavu

V prípade jedného vytvoreného zhluku dochádza k javu, kedy je suma príťažlivých síl medzi uzlami násobne väčšia ako suma odpudivých síl v grafe. Zhluk pozostávajúci zo všetkých alebo z väčšiny uzlov grafu, sa stáva nečitateľným, z dôvodu zakrytia hrán a uzlov ostatnými elementami grafu. V prípade ak sa vzdialenosť uzlov neustále zväčšuje resp. suma odpudivých síl je väčšia ako suma príťažlivých síl, graf nedokáže v reálnom čase dosiahnuť stabilný stav, a z toho dôvodu je vizualizácia nepoužiteľná.

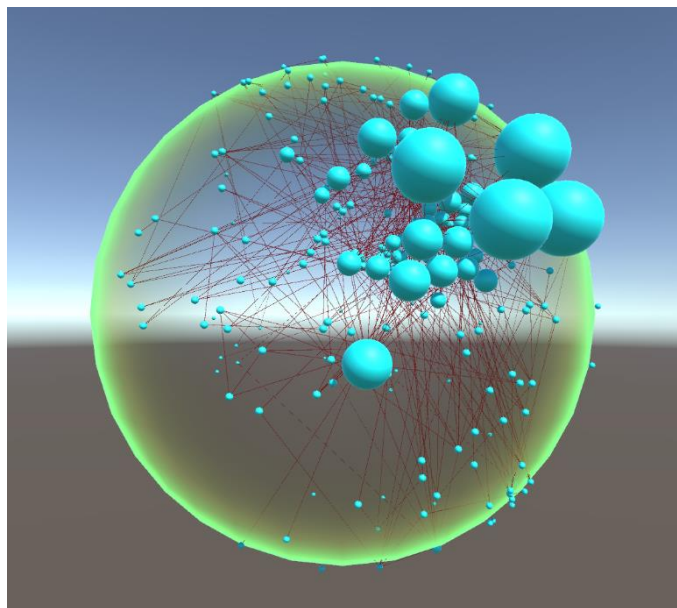
Pre určenie správnej veľkosti obmedzení v 3D priestore, bola vykonaná skupina meraní zameraná na detekciu obmedzení a možností, ktoré poskytuje samotné herné prostredie. Z publikácie [74] vyplývajú doporučené počty použitých herných objektov v jednej scéne, za účelom vytvorenia grafu s udrzaním časovej a výpočtovej zložitosti pod hranicou, kedy sa vizualizácia stáva nepoužiteľnou. Voľné zobrazenie siete v priestore má mnoho nežiadúcich faktorov, ktoré je potrebné detegovať a zabezpečiť. Ako môžeme vidieť na Obrázok 5, vzniknutý zhluk zabráňuje prehľadnosti siete, pretože uzly na ktoré pôsobí menšia príťažlivá sila alebo uzly na ktoré pôsobí priveľká odpudivá sila prekážajú kamere.

Pôsobiace príťažlivé a odpudivé sily v sieti bez zadefinovaných obmedzení, môžu vytvoriť nežiadúce zhľuky, ktoré sú nečitateľné pre používateľa. Aby sme predišli situácií, v ktorej zhluk uzlov zakrýva zvyšok vizualizácie, implementovali sme obmedzenie vzdialenosti uzlov medzi sebou. Obmedzenie je reprezentované sférickým polom, ktoré slúži ako stena pre ostatné uzly. Veľkosť sférického poľa je priamo odvodená od veľkosti samotného uzla t.z. čím väčší uzol tým väčšie sférické pole. Tento prístup umožňuje vytvorenie prehľadnejšej vizualizácie ako môžeme vidieť na Obrázok 6.



Obrázok 6 Sférické obmedzenie

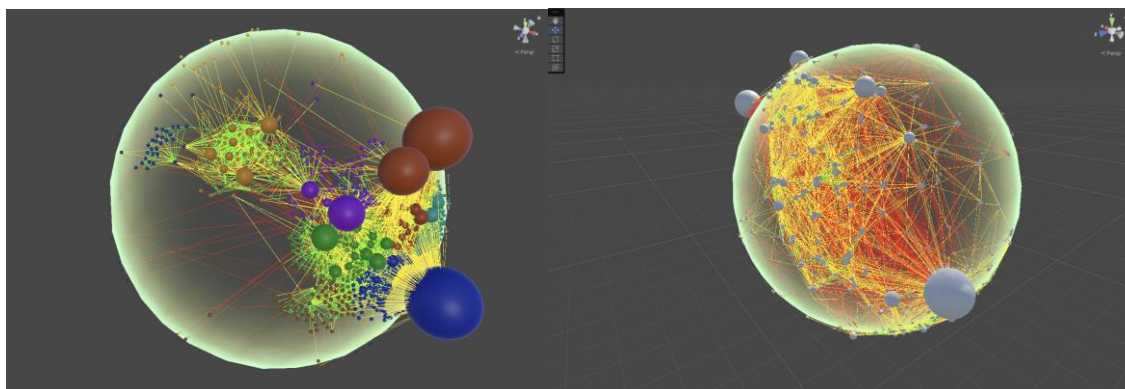
Voľné zobrazenie v priestore prináša nové možnosti ako obohatiť 3D vizualizáciu ale prináša problémy vo forme ustálenia siete v zornom poli kamery. Tento problém je čiastočne vyriešený pre uzly, ktoré sú vytlačené mimo zorného poľa, ale nerieši problém vytvorenia jedného veľkého zhluku. Jeden zhluk vzniká v prípade ak sú príťažlivé sily kotvy väčšie akú suma ostatných pôsobiacich síl v sieti. Aby sme zamedzili tomuto nežiadúcemu stavu implementovali sme sférické obmedzenie, ktoré udržiava uzly neustále v rovnakej vzdialenosti od kotvy, pričom im zároveň zabráňuje sa vzdialiť opačným smerom ako môžeme vidieť na Obrázok 7.



Obrázok 7 Vizualizácia dát so sférickým obmedzením

Na rozdiel od uzlov, hrany majú povolený prechod cez sférickú zábranu, čo vytvára zaujímavý efekt, kde je možné sledovať prepojenia v sieti medzi vybranými uzlami, bez interferencie okolia. Sférická zábrana je z toho dôvodu z väčšiny plochy priehľadná, ale v každom uhle pohľadu si zachováva svoj obrys pre neustálu reprezentáciu zobrazenia na povrchu sféry. Pri vizualizácii grafu v 3D priestore sa často stretávame s potrebou efektívneho zobrazenia vzdialeností medzi vrcholmi. Pre lepšie porozumenie vzdialeností a štruktúry grafu je možné využiť farebné kódovanie hrán na základe vzdialenosti medzi vrcholmi. V tejto vizualizácii sú hrany farebne ovplyvnené podľa vzdialenosti medzi ich koncovými vrcholmi. Konkrétne, hrany sú zafarbené červeno, ak sú vrcholy ďaleko od seba, zeleno, ak sú vrcholy blízko a žltó, ak sa nachádzajú v priemernej vzdialenosti. Toto farebné kódovanie poskytuje vizuálny indikátor vzájomných vzdialeností vrcholov v 3D priestore ako môžeme vidieť na Obrázok 8.

Porovnaním na Obrázok 8 môžeme usúdiť, že výhoda tohto prístupu spočíva v tom, že umožňuje užívateľovi rýchlo identifikovať, aké časti grafu sú vzdialené a ktoré sú blízke. Taktiež umožňuje jednoduchú identifikáciu oblastí grafu, ktoré by mohli vyžadovať ďalšiu pozornosť alebo analýzu. Takéto farebné kódovanie vytvára ďalší vrstevnicový aspekt vo vizualizácii, ktorý zvyšuje informačnú hodnotu pre užívateľa.



Obrázok 8 Rôzna veľkosť sférického obmedzenia

Ukotvenie

Uzly a hrany voľne zobrazení v priestore je potrebné zoskupiť okolo miesta v scéne, na ktoré má používateľ nastavenú kameru. V takom prípade je aplikované ukotvenie resp. kotva, ktorá pôsobí na každý uzol v sieti a priťahuje ho smerom k základnému bodu počas celého priebehu vizualizácie. Aplikovaním kotvy dosiahneme stav, ktorý sme definovali ako žiadaný Obrázok 3, kde je celá sieť vizualizovaná v zornom poli kamery, pričom aplikáciu ukotvenia môžeme pozorovať na Obrázok 7.

Modifikácia 1. Zmena veľkosti sférického obmedzenia

Zvolenie veľkosti sférického obmedzenia závisí od vizualizovaných dát. Je pravdepodobné, že prvotne zvolená veľkosť nebude vyhovovať veľkosti a objemu siete, preto je v aplikácii implementovaná funkcionálna zmena veľkosti sférického obmedzenia počas vykonávania silou riadeného usporiadacieho algoritmu. Zmena veľkosti sféry umožňuje zlepšenie prehľadnosti siete, pretože pri každej zmene, či už zmenšenie sféry alebo jej zväčšenie, sa zároveň mení aj finálne rozloženie uzlov a hrán v priestore.

Modifikácia 2 Ohraničenie priestoru v okolí vrcholov

S narastajúcim počtom vrcholov a hrán narastá aj zložitosť vizualizácie. Aj keď 3D priestor poskytuje dostatočný priestor na vizualizáciu veľkého objemu dát, stále sa musíme vysporiadať s problémom prekrývania vrcholov v jednotlivých zhlukoch. Jedným zo spôsobov je implementácia dynamických záporných hrán, ktoré sú aktívne len v prípade, ak sa dva vrcholy priblížia na určenú vzdialenosť. Táto vzdialenosť je definovaná podľa požiadaviek vizualizácie. Dynamické záporné hrany nám pomáhajú obmedziť prekrývanie vrcholov v priestore a zároveň priamo neovplyvňujú ostatné priťažlivé sily v grafe natoľko, aby zmenili vizualizovanú informáciu.

Modifikácia zobrazenia hrán

Priamo prepojené hrany, ktoré prechádzajú cez sféru, ponúkajú jednoduchú a priamu cestu medzi vrcholmi, čo uľahčuje interpretáciu a pochopenie vzťahov v grafe. Ich kratšia dĺžka môže prispieť k zníženiu celkovej komplexity vizualizácie. Avšak tento prístup môže narušiť kontinuitu priestoru a viesť k dojmu prestávky v vizuálnej reprezentácii. Okrem toho, priame hrany môžu mať tendenciu prekrývať sa s inými prvkami na sfére, čo môže ovplyvniť čitateľnosť grafu.

Na druhej strane, zaoblené hrany, ktoré idú len po povrchu sféry, udržujú kontinuitu priestoru a vytvárajú plynulý vizuálny zážitok. Ich umiestnenie na povrchu sféry môže zlepšiť prehľadnosť vizualizácie tým, že menej zasahujú do významných prvkov na sfére. Avšak ich komplexnejšie modelovanie a manipulácia môžu zvýšiť náročnosť implementácie vizualizačného systému. Okrem toho, tieto hrany môžu mať tendenciu znižovať efektívnu vzdialenosť medzi vrcholmi, čo môže ovplyvniť vnímanie vzájomných vzťahov v grafe. V prípade, keď hrany grafu spájajú vrcholy priamo a prechádzajú cez sféru, môžeme očakávať nasledujúce výhody a nevýhody:

Výhody:

- **Jednoduchosť:** Priama cesta medzi vrcholmi môže byť jednoduchšia na pochopenie a interpretáciu, pretože nevyžaduje žiadne dodatočné zohľadnenie zakrivenia sféry.
- **Krátkosť hrán:** Hrany môžu byť kratšie, čo znižuje celkovú zložitosť vizualizácie a uľahčuje sledovanie spojení medzi vrcholmi.

Nevýhody:

- **Prestávka v kontinuite:** Priame hrany, ktoré prechádzajú cez sféru, môžu vytvoriť dojem prestávky v kontinuite priestoru, čo môže spôsobiť neprirodzený vzhľad vizualizácie.
- **Možné prekrytie:** Priame hrany môžu mať tendenciu prekryvať sa s ostatnými prvkami na sfére, ako sú vrcholy, čo môže spôsobiť zhoršenie čitateľnosti vizualizácie.

V prípade zaoblených hrán, ktoré idú len po povrchu sféry, môžeme očakávať nasledujúce výhody a nevýhody:

Výhody:

- **Kontinuita:** Zaoblené hrany na povrchu sféry môžu zachovať kontinuitu priestoru, čo vytvára plynulý vizuálny zážitok.
- **Väčšia prehľadnosť:** Hrany na povrchu sféry môžu menej prekážať významným prvkom na sfére, čo zlepšuje čitateľnosť vizualizácie.

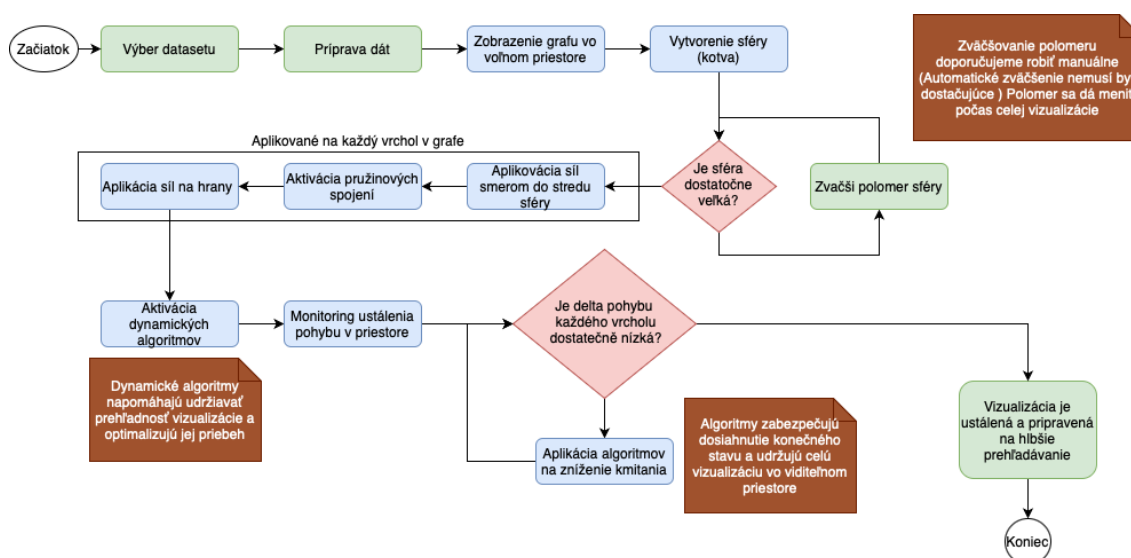
Nevýhody:

- **Zložitejšie modelovanie:** Zaoblené hrany môžu vyžadovať komplexnejšie modelovanie a manipuláciu, čo môže viesť k zvýšenej náročnosti implementácie vizualizačného systému.
- **Zmenšenie vzdialenosti:** Hrany, ktoré sa prekývajú s povrchom sféry, môžu mať tendenciu znižovať efektívnu vzdialenosť medzi vrcholmi, čo môže ovplyvniť vnímanie vzájomných vzťahov medzi nimi.

Vo voľbe medzi týmito dvoma prístupmi je dôležité zvážiť konkrétne požiadavky na vizualizáciu a aplikáciu grafu na sfére, aby sme dosiahli optimálnu reprezentáciu, ktorá zohľadňuje potreby užívateľa a zabezpečuje čitateľnosť a pochopenie dát.

Postup vizualizácie

Postup celej vizualizácie je zaznamenaný na diagrame Obrázok 9. Výber datasetu a príprava dát podlieha vytvorenej metodike v prílohe B. Modré časti diagramu znázorňujú vytvorené algoritmy spolu s ich modifikáciami, ktorými je možné dosiahnuť požadovanú úroveň vizualizácie dát, pričom je potrebné dbať na korektné použitie modifikácií pre vybrané dáta. V diagrame je znázornená zastavovacia podmienka, ktorá je definovaná špeciálne pre vytvorené algoritmy. Pri štandardných 2D algoritmoch je vopred daný presný počet iterácií, ktoré sa majú vykonať, a to aj v prípade ak sa vizualizácia dostane do konečného stavu oveľa skôr. Počet iterácií môže byť aj nedostačujúci, a v takom prípade sa vizualizácia zastaví pred dosiahnutím konečného stavu. Vytvorená zastavovacia podmienka zastaví vizualizáciu v prípade ak je priemer všetkých rýchlostí vrcholov v grafe menší ako 1 za vybranú jednotku času, ktorá je vo väčšine prípadov 0.2 sekundy. Pre väčšie grafy sa priemer vzťahuje len na 90% vrcholov, čo zabezpečuje, že malý počet oscilujúcich vrcholov, ktoré väčšinou disponujú malým počtom hrán, nezabránili zastaveniu vizualizácie.



Obrázok 9 Diagram vizualizácie

Modifikácie 3D zobrazenia – interaktívne funkcie

Interakcia so dátami v podobe siete je v 3D priestore veľmi dôležitá, pretože umožňuje hlbšie preskúmať a pochopiť dáta. Z toho dôvodu sú implementované dve možnosti ako prehľadávať dátový priestor:

- Monitor a myš
- Monitor a zariadenie Leap motion

Klasické ovládanie pomocou myši umožňuje používateľovi rotovať vizualizovanú sieť, vybrať uzly na ktoré sa práve potrebuje zamerať a zároveň môže posúvať zorné pole kamery. Tento prístup je pre používateľa intuitívny pretože je často používaný v iných nástrojoch na vizualizáciu ako napríklad Gephi opísané v kapitole 0. Ovládanie prostredia pomocou myši môže byť v mnohých prípadoch neintuitívne a zavádzajúce, pretože kurzor ako ovládací prvok nedokáže simulovať tretiu dimenziu, čím výrazne redukuje možnosti vizualizácie. Pre zabezpečenie presnej a jednoduchšej interakcie s využitím tretej dimenzie je potrebné použiť zariadenie ako napríklad Leap motion, ktoré spĺňa požiadavky na 3D interakciu v dostatočnej miere pre našu prácu. Zariadenie funguje na princípe snímania rúk pomocou kamier, a následne rozoznáva zadané gestá používateľa, ktoré je možné priradiť jednotlivým úkonom v hernom prostredí. Intuitívne ovládanie pomocou rúk zabezpečuje plynulú a jednoducho ovládateľnú vizualizáciu, pričom umožňuje vizualizačnej aplikácii používať obmedzenia, ktoré majú za úlohu sprevádzať používateľa prehľadávaným priestorom.

Techniky zamerania a kontextu (Focus + Context)

Techniky zamerania a kontextu sa venujú problému strát kontextu pri približovaní sa k určitým údajom. Predpokladajme, že sme priblížili obrázok. Výsledkom bude, že budeme môcť vidieť len oblasť, ku ktorej sme sa priblížili, a stratíme predstavu o okolitých oblastiach obrázka. Práve tu prichádza do hry technika zamerania a kontextu: dáva používateľom schopnosť vidieť hlavný objekt v detailnom pohľade (zameranie) spolu s prehľadom všetkých okolitých informácií (kontext). Všeobecne platí, že strata kontextu je považovaná za problém v aplikáciách vizualizácie informácií. S cieľom odstrániť tento problém, objavili sa techniky zamerania a kontextu, ktoré používateľovi umožňujú sústrediť sa na niektoré detaily bez straty globálneho kontextu. Táto koncepcia nezastupuje metódy približovania a posúvania, ale

skôr ich dopĺňa. Väčšina vizualizačných systémov implementuje obe techniky spolu ako nástroj interakcie.

Priblíženie a posunutie (Zooming and Panning)

Priblíženie a posunutie sú základné nástroje na preskúmavanie veľkého množstva informácií. Posunutie znamená pohyb kamery po scéne, zatiaľ čo priblíženie umožňuje používateľom prepínať medzi abstraktným a detailným pohľadom. Geometrické priblíženie upravuje obrazovú transformáciu a tým umožňuje zväčšovanie alebo zmenšovanie priblíženého grafu. Sémantické priblíženie znamená, že nielen veľkosť objektov, ale aj zobrazené informácie sa môžu zmeniť pri približovaní sa k určitej oblasti grafu.

Priblíženie a posunutie sú základnými nástrojmi pri spracovaní veľkých a komplexných grafov. Tieto techniky umožňujú užívateľom flexibilne manipulovať s vizualizovanými dátami, čím podporujú lepšie porozumenie štruktúry siete. Implementácia týchto techník vytvára interaktívne vizualizačné prostredie, ktoré uľahčuje objavovanie a analýzu vzťahov v grafe. Týmto spôsobom priblíženie a posunutie zohrávajú kľúčovú úlohu v procese efektívnej vizualizácie grafov a sietí.

Diskusia

V našej práci sme vytvorili metodiku pre prípravu dát pre vizualizáciu dát v 3D priestore s použitím silou riadených algoritmov vo fyzikálnom hernom prostredí Unity3D.

Predstavujeme vlastné silou riadené algoritmy a ich modifikácie, postupy a doporučená, výber a použitie správnych interakčných nástrojov, ktoré v konečnom dôsledku napomáhajú vytvoriť zrozumiteľné a čitateľné vizualizácie dát v 3D priestore bez nutnosti skúmania a pohľadávania dát vopred. Vytvorené algoritmy používajú vnútorné vlastnosti grafov, čo napomáha jednoduchšej aplikácii na rôzne dáta. Základ algoritmov vychádza z 2D, a preto je dôležité si uvedomiť, aké problémy a výzvy prináša prenos algoritmov do 3D priestoru. V 2D priestore absencia tretieho rozmeru slúži ako obmedzenie, ktoré odstraňuje niektoré nežiaduce stavy vizualizácie, ako napríklad prekryvanie vrcholov v priestore.

Hlavný algoritmus vizualizácie dát v 3D so všetkými jeho modifikáciami, sme vytvorili iteratívnym spôsobom, pričom pridaním nových postupov a obmedzení sa postupne objavovali nové zistenia, ktoré viedli k súčasnému stavu. V prvej iterácii, sme sa zamerali na reprezentáciu hrán v grafe pomocou pružinových spojení. Výsledky experimentov ukázali obmedzenia 3D priestoru a hlavne samotných fyzikálnych herných prostredí, pričom boli zistené maximálne použiteľné objemy dát, ktoré je možné pomocou týchto nástrojov vizualizovať. Ďalšie podstatné zistenie z vykonaných experimentov v prvej iterácii je, že vrcholy v grafe pri použití pružinových spojení majú tendenciu kmitať do úrovni, kedy už nie je možné graf zastabilizovať. Zároveň sme zistili, že pružinové spojenia, tak ako sú reprezentované v herných prostrediach, sú veľmi náchylné na zmenu ich vnútorných parametrov, ktoré nie sú jednoznačne definované pre druh dát. Preto sme v našej metodike rozhodli o nastavení pružinových spojení ako konštant, pričom jediný parameter ktorý sa mení je normalizovaná sila. V ďalšej iterácii výskumu, sme zadefinovali hrany grafu ako konštantne pôsobiace sily. V tomto experimente sily môžu mať príťažlivý efekt alebo odpudivý efekt, na rozdiel reprezentácie pomocou pružín, v tejto iterácii vytvoríme prepojenia medzi všetkými vrcholmi v grafe. Výsledky experimentov ukázali, že tento prístup má menšiu tendenciu rozkmitať model ale zároveň prináša nové problémy s udrzaním modelu v zornom poli kamery. Aplikované sily na jednotlivé vrcholy závisia od povahy dát a v mnohých prípadoch bol pozorovaný jav, pri ktorom vrcholy, na ktoré bola aplikovaná veľká sila nadobudli také zrýchlenie, že boli nakoniec odtrhnuté od zvyšku modelu. Tento jav môžeme pozorovať aj v 2D prostredí pri algoritmoch ako je napríklad Force Atlas.

Ďalším nežiadúcim javom bolo vytvorenie jedného veľkého zhluku, ktorý obsahoval väčšinu dát, a tým sa stala vizualizácia nečitateľná. Aby sme odstránili spomenuté nedostatky, spojili sme jednotlivé postupy t.z. pružinové spoje a konštantne aplikovanú silu do jedného algoritmu (Pružinovo-silový algoritmus), ktorý je opísaný v kapitole 0. Tento algoritmus odstraňuje problémy s kmitaním, čo zabezpečujú konštantne aplikované sily a zároveň pružinové spojenia držia všetky vrcholy v zornom poli kamery. Z jednoduchej aplikácie algoritmov v 3D, tento algoritmus poskytuje najlepšie výsledky v pomere k jeho zložitosti. Na druhú stranu stále obsahuje určité nedostatky, ktoré môžu zakrývať alebo skresľovať informáciu. Jedným zo závažných problémov je, že tento algoritmus stále nedokáže udržať model vo vybranom bode v priestore. Aj keď odstránime kmitanie jednotlivých vrcholov, model ako celok stále disponuje silou, ktorá ho celý posúva určitým smerom. Model môže nadobudnúť takú rýchlosť, že kamera nedokáže model sledovať v komfortnej rýchlosti. Na odstránenie tohto problému sme v ďalšej iterácii pridali gravitačnú silu t.z. kotvu, ktorá udržuje model v okolí jedného bodu. Kotva pôsobí na každý vrchol v modeli rovnakou konštantnou príťažlivou silou, a zároveň neovplyvňuje samotné vzťahy medzi jednotlivými vrcholmi. Tento prístup je jednoduchý na použitie, ale zároveň pri zle nastavenej konštantnej sile môže celý model skončiť v jednom zhluku ako pri druhej iterácii. Na odstránenie problému jedného veľkého zhluku, bola celá vizualizácia transformovaná na sférické zobrazenie. Sféra v tomto prípade funguje ako zábrana pre vrcholy a zabezpečuje konštantnú vzdialenosť všetkých vrcholov od vybraného bodu v priestore - kotvy. Táto iterácia nášho algoritmu zabezpečuje prehľadnú vizualizáciu pre väčšinu dátových setov a odstraňuje väčšinu nežiadúcich javov z prvých troch iterácií algoritmu.

V ďalších iteráciách algoritmu sme sa zamerali na odstránenie javov, ktoré zabraňujú prehľadnej vizualizácii v 3D ako napríklad prekrývaní vrcholov alebo neoptimálneho zobrazenia dát na sfére. Na rozdiel od 2D priestoru kde nám absencia jedného rozmeru napomáha odstrániť problém prekrývania vrcholov, v 3D sme museli zakomponovať nové postupy:

- Dynamické odpudivé sily
- Konštantné vrcholové obmedzenie

Pri použití konštantného vrcholového obmedzenia zabezpečíme každému vrcholu svoj vlastný priestor, pričom ostatné vrcholy sa nemôžu priblížiť do vzdialenosti, ktorá by spôsobovala prekrývanie vrcholov. Pri tomto postupe nastáva problém pri väčších objemoch dát, pretože vrcholy môžu nepriamo ovplyvňovať vzťahy okolo seba tým, že odtláčajú (neprirodzene) všetko smerom od svojho stredu. Na druhú stranu pri menších objemoch sa podľa experimentov, tento postup osvedčil hlavne vzhľadom na výpočtovú zložitosť a jednoduchosť implementácie. Druhým spôsobom je zapojenie dynamických odpudivých síl, ktoré sú aktívne len v prípade, že sa vrcholy priblížia na definovanú vzdialenosť. Každý vrchol má vzdialenosť definovanú podľa svojho stupňa, čo zabezpečuje, že vrcholy s vysokým stupňom nezabraňujú vrcholom s malým stupňom v pohybe po sfére. Postupy je možné kombinovať, ale treba zachovať podmienku, že dynamické odpudivé sily budú aktívne skôr, ako sa vrchol priblíži do vzdialenosti ako je definované v konštantnom obmedzení. V ďalšej iterácii vývoja sme sa venovali hranám grafu, ich zafarbeniu, tvaru a iným vlastnostiam, ktoré je možné aplikovať na vizualizáciu hrán. Zafarbenie hrán môže byť naviazané na rôzne vlastnosti vyplývajúce z dát, pričom v našej verzii je táto vlastnosť naviazaná na euklidovskú vzdialenosť, pre lepšie zachytenie priestorového vnímania. Taktiež je táto vlastnosť v súlade s myšlienkou silou riadených algoritmov, a to takým spôsobom, že priestorová vzdialenosť je len vnútorná vlastnosť grafu t.z. nie je potrebný zásah z vonku na jej určenie, a preto vytvorený algoritmus môže naďalej pracovať klasickým postupom. Pre ešte lepšie prehľadné priestorového vnímania, bola zadefinovaná hrúbka hrán nasledovne:

1. Zelené hrany (Vrcholy sú bližšie ako $d1$), hrúbka 1
2. Žlté hrany (Vrcholy sú bližšie ako $d2$), hrúbka 0.5
3. Červené hrany (Vrcholy sú bližšie ako $d3$), hrúbka 0.1

Pričom $d1$, $d2$ a $d3$ závisia od polomeru r sférického obmedzenia kde $d1 = r / 3$, $d2 = r$ a $d3 = 2r$. Hrúbka je vyjadrená v jednotkách Unity3D priestoru.

Porovnanie časovej zložitosti

V tejto časti práce sa zameriame na porovnanie výsledkov nášho algoritmu založeného na 3D vizualizácii dát s existujúcimi algoritmi používanými pre 2D vizualizáciu. Naším cieľom je zistiť, ako sa náš prístup k vizualizácii veľkých dátových množín v trojrozmernom priestore porovnáva s tradičnými algoritmi používanými v dvojrozmernom priestore. V tejto kapitole sa pozrieme na čas potrebný k dosiahnutiu ustálenej vizualizácie nášho algoritmu v porovnaní s existujúcimi metodikami a algoritmi, ktoré sú bežne používané na vizualizáciu dát. Tento pohľad na porovnanie nám umožní lepšie pochopiť výhody a obmedzenia nášho prístupu v porovnaní s existujúcimi technikami.

Tabuľka 2 Časové porovnanie algoritmov

Dataset	Počet vrcholov / Počet hrán	Základný alg. [ms]	Fruchterman- Reingold [ms]	Kamada-Kawai [ms]
Game of Thrones	796 / 3909	3510 ± 62,20	3490 ± 48,60	3350 ± 55,30
Witcher	224 / 2600	1120 ± 39,80	1130 ± 21,50	1130 ± 19,00
Lazega	36 / 115	198 ± 03,86	203 ± 11,30	198 ± 08,72
Network Science	1589 / 2742	7030 ± 77,10	6010 ± 42,80	7430 ± 104,00
Pokec5000	5000 / 52182	27800 ± 302,00	21400 ± 396,00	43300 ± 792,00
Pokec1000	1000 / 6297	4980 ± 108,00	4780 ± 130,00	4740 ± 125,00

Časová zložitosť pre vybrané algoritmy je porovnaná v Tabuľka 2. Z výsledkov v Tabuľka 3 je zrejmé, že vybrané algoritmy v 2D priestore sú rádovo rýchlejšie ako vizualizácia v 3D priestore. Zmena nastáva pri rastúcom objeme vizualizovaných dát ako je zobrazené v Tabuľka 3.

Keďže vybrané 2D algoritmy nedisponujú podmienkou na zastavovanie výpočtu vizualizácie, vykonávajú vždy rovnaký počet iterácií na výpočet výsledného stavu grafu. Každá iterácia je výpočtovo náročná a musí byť vykonaná aj keď je graf v ustálenom stave. Predstavený pružinovo silový algoritmus, zastaví vizualizáciu podľa zastavovacej podmienky 0, a tak zabráni zbytočnému prepočtu ustáleného grafu. Pri narastajúcom počte vrcholov a hrán, ako napríklad pri datasete Pokec5000, môžeme pozorovať rýchlejšie časy z toho dôvodu, že aplikované sily v grafe sú dostatočne veľké a nedovolia vrcholom sa rozkmitať, čo má za následok rýchlejšie ustálenie vrcholov v priestore.

Tabuľka 3 Časové porovnanie pružinovo-silového algoritmu s kontrolovaným rozmiestňovaním

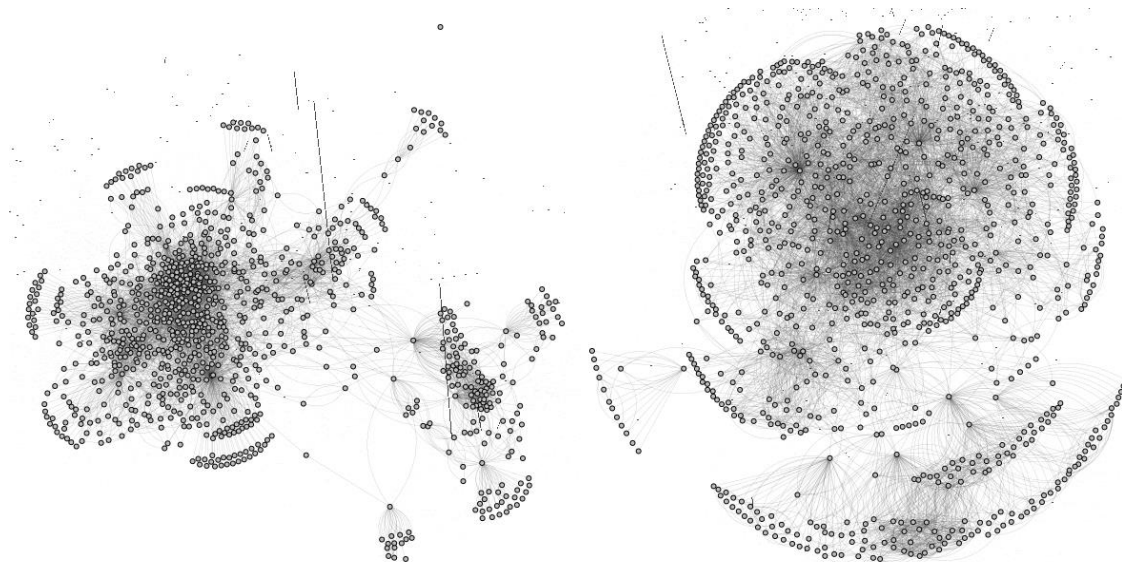
Dataset	Počet vrcholov / Počet hrán	Pružinovo-silový algoritmus [ms]
Game of Thrones	796 / 3909	19244 ± 250
Witcher	224 / 2600	16784 ± 303
Lazega	36 / 115	640 ± 15
Network Science	1589 / 2742	41404 ± 539
Pokec5000	5000 / 52182	16247 ± 1480
Pokec1000	1000 / 6297	83330 ± 15234

Vizuálne porovnanie

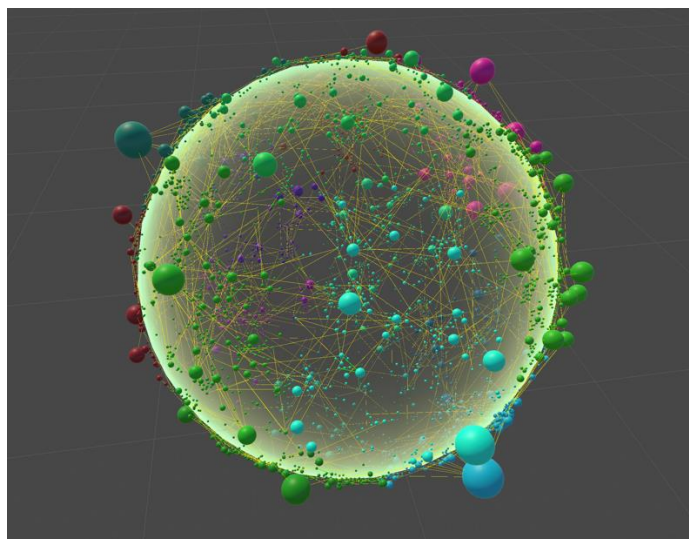
Táto kapitola sa zameriava na vizuálne porovnanie rôznych metód vizualizácie grafov s novo vyvinutou metódou, ktorá kombinuje pružinovo-silový algoritmus s kontrolovaným rozmiestňovaním (PSA-KR). Cieľom je posúdiť využiteľnosť PSA-KR v porovnaní s existujúcimi metódami, ako je jednoduchý pružinový algoritmus (PA), silovo riadený algoritmus (SA) a existujúce vybrané 2D algoritmy.

Vybrané 2D algoritmy na Obrázok 10, majú tendenciu vyvárať jeden veľký zhluk s narastajúcim počtom dát, pričom pozorovateľné pomyselné hranice zhlukov miznú s narastajúcim počtom vrcholov a hrán v grafe. Na Obrázok 10 môžeme vidieť, že vizualizované dáta vytvorili zhluk, ktorý nie je možné hlbšie analyzovať bez vonkajšieho zásahu používateľa alebo bez obmedzenia vizualizovaných dát. PSA-KR na Obrázok 11, vizuálne lepšie odlišuje vytvorené zhluky a umožňuje hlbšiu vizuálnu analýzu. Pozorovaním vizuálneho modelu môžeme usúdiť, že vzniknuté zhluky sú čitateľnejšie a pomyselné hranice zhlukov sú výraznejšie, čo napomáha hlbšej analýze dát. Opísaný jav je výraznejší pri väčšom objeme dát ako môžeme vidieť na Obrázok 15, kde sme pre 2D vizualizáciu pomocou silou riadených algoritmov, pomaly dosiahli hranice použiteľnosti.

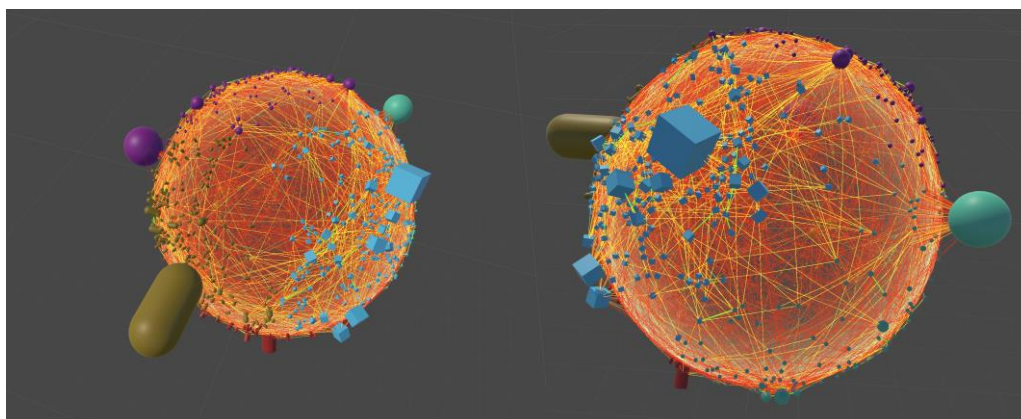
Obrázok 16. Je potrebné podotknúť, že vytvorené algoritmy v 3D priestore disponujú značnou mierou možností interakcie s priestorom, kde je možné dáta pozorovať z rôznych uhlov pohľadu a zároveň interagovať so samotným priestorom. Priestorové vnímanie je pre 3D algoritmy neoddeliteľnou časťou vizualizácie a pre používateľa vytvára mentálne pomôcky, ktoré mu napomáhajú vnímať spojitosti vo vizualizovaných dátach. Pre menšie datasety, ktorých príklad môžeme pozorovať na obrázkoch Obrázok 13 a Obrázok 14, nie je spomínaný fenomén na prvý pohľad zreteľný, ale pri interakcií a manipuláciu s priestorom, môžeme pozorovať na Obrázok 14 zreteľné body záujmu, ktoré pridávajú nové informácie o pozorovanom datasete. Samotné vnímanie priestoru v tomto prípade slúži ako nová skupina vlastností určitých vrcholov, pričom môžeme pozorovať epicentrum pravdepodobne najdôležitejších vrcholov v pozorovanom grafe.



Obrázok 10 Pokec1000 - ForceAtlas2 a Kawada – Kawai



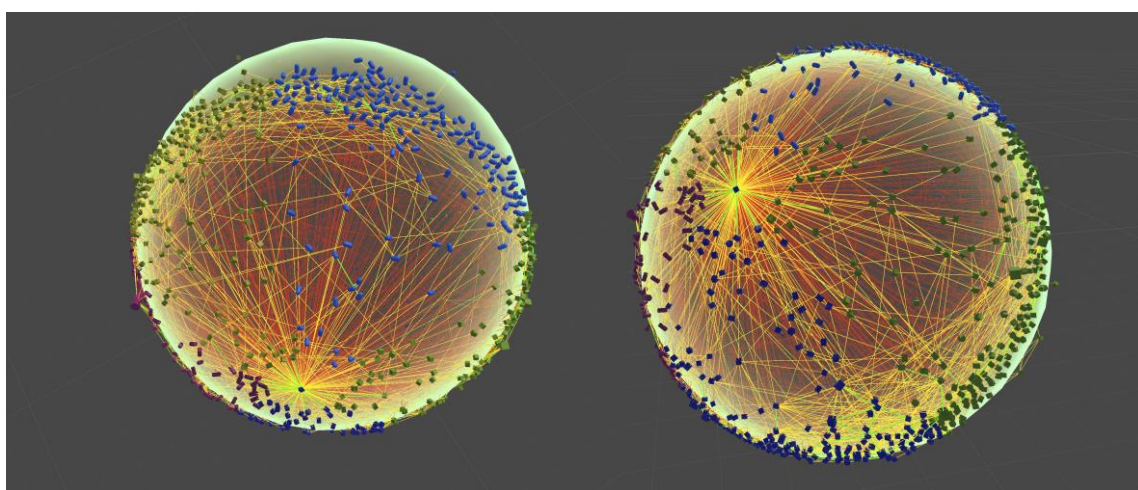
Obrázok 11 Pokec1000 - Pružinovo-silový algoritmus so skrytými hranami



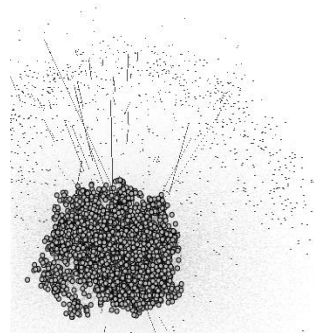
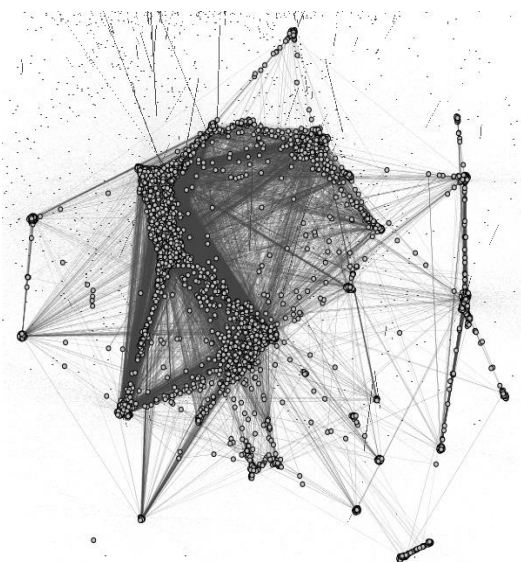
Obrázok 12 Pokec1000 - Pružinovo-silový algoritmus



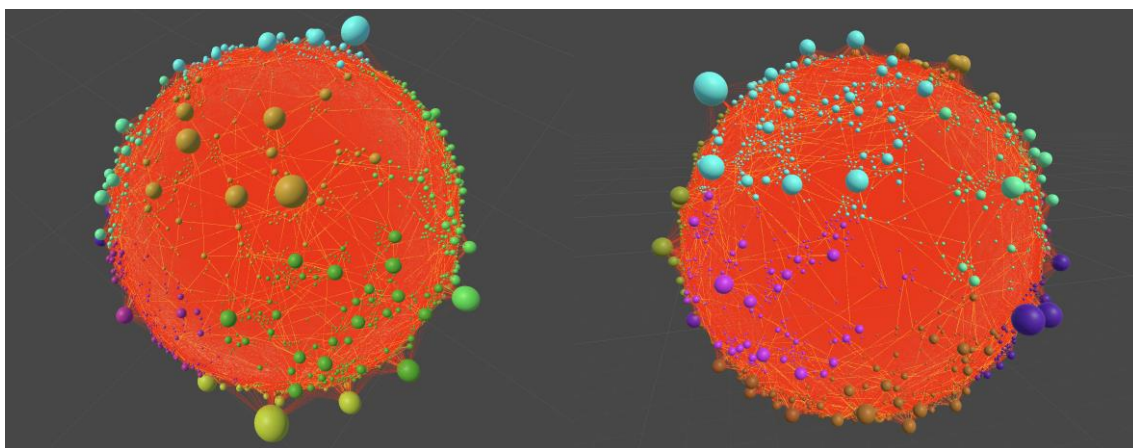
Obrázok 13 Network Science - základný algoritmus a Fruchtreingold algoritmus



Obrázok 14 Network Science - Pružinovo-silový algoritmus



Obrázok 15 Pokec5000 - Fruchterman-Reingold a Force Atlas2



Obrázok 16 Pokec5000 Pružinovo-silový Algoritmus

Porovnanie zhlukov a priestorovej vzdialenosti

Hodnotenie navrhnutých algoritmov si vyžaduje definovanie metrík pre porovnanie. Keďže použitie kritéria estetickej kvality výslednej vizualizácie je bez výskumu používateľov problematické, zvolili sme kvantitatívne kritériá založené na dvoch základných problémoch vizualizačných algoritmov v 3D priestore. Prvým kritériom je priemerná vzdialenosť vrcholov grafu vo výslednej vizualizácii, druhým je priemerná hustota vytvorených zhlukov. Posledným sledovaným parametrom je priemerná vzdialenosť vytvorených zhlukov. *Priemerná vzdialenosť vrcholov* definuje celkovú hustotu grafu. Keďže pre interpretáciu výslednej vizualizácie predstavuje problém ako príliš riedky, tak aj príliš hustý graf, je potrebné sledovať priemernú vzdialenosť vrcholov ako kritérium kvality výslednej vizualizácie.

Priemerná hustota klastrov vyjadruje, nakoľko blízko seba sú vo výslednej vizualizácii vrcholy s podobnými vlastnosťami. V prípade malej vzdialenosti medzi vrcholmi v zhluke je výsledná informácia ťažko čitateľná bez interakcie s používateľom vo forme zväčšovania (ang. zoom). Pri nadmernom zväčšovaní sa následne strácajú informácie o celkovom kontexte vizualizovanej podmnožiny dát. Naopak, pri príliš nízkej hustote nie je zrejmé členstvo vrcholov v rámci zhuku.

Priemerná vzdialenosť klastrov funguje na podobnom princípe ako vzdialenosť medzi vrcholmi, avšak zohľadňuje skupiny vrcholov. Opäť platí, že výsledná vizualizácia by nemala viesť k extrémnej vzdialenosti medzi zhukmi, ani k ich nadmernej blízkosti.

Pružinový algoritmus

Napriek tomu že, táto metóda vytvára čisté a symetrické usporiadania pre grafy strednej veľkosti, považuje sa za náročný algoritmus. Časová náročnosť býva často vysoká, čo robí algoritmus v mnohých prípadoch nepraktickým. Navyše trpí nedostatočnou prediktabilitou, čo znamená, že opakované spustenia algoritmu môžu viesť k rôznym výsledkom. To môže predstavovať výzvu pri udržiavaní mentálnej mapy používateľa počas interakcie s nestabilnými usporiadaniami.

V počiatočných experimentoch sa testovali pružinové spojenia medzi vrcholmi, ktoré predstavujú atraktívne sily. Zároveň zabezpečujú aj zodpovedajúce odpudivé sily, pretože

jednou z vlastností pružín je sťahovanie ich dĺžky. Vrcholy a hrany sú generované s náhodnými vlastnosťami, ako je farba a pozícia, aby sa odstránilo možné skreslenie spôsobené akoukoľvek špecifickou súborom dát. Použili sme náhodné pozície na odstránenie počiatočného skreslenia spôsobeného kumulovanou silou, ktorá sa aplikuje na uzly pri umiestnení na začiatku scény. Na smerovanie hrán v grafe sú vrcholom navyše priradené farby, ktoré reprezentujú ich typ. Na základe typov počiatočného a koncového uzla sa vytvárajú hrany a nastavuje sa pevnosť pružinového spoja. Pružinové spojenia umožňujú rýchlu a dostatočne prehľadnú vizualizáciu grafu, pričom pri použití správneho nastavenia riadia rozloženie uzlov v priestore. Ich nevýhodou sú obmedzené sily pôsobiace v jednom smere a zvýšené zaťaženie výpočtovej jednotky. Každý pružinový spoj musí byť reprezentovaný jedným špecifickým objektom v scéne, čo spôsobuje zvýšené zaťaženie nielen pre fyziku, ale aj pre vykresľovanie [74].

Základná chyba pružinového algoritmu je vysoká náchylnosť na rozkmitanie celého modelu. Toto je základný identifikovaný nedostatok, ktorý viedol k úsiliu o zlepšenie použitej vizualizačnej metódy. Výsledky z experimentov môžeme pozorovať v Tabuľka 4.

Tabuľka 4 - Výsledky experimentov s jednoduchým pružinovým algoritmom (v abstraktných jednotkách používaných v Unity Engine).

Názov datasetu	Priemerná vzdialenosť vrcholov	Priemerná hustota zhlukov	Priemerná vzdialenosť zhlukov
Game of Thrones	348.98	28.65	69.67
Witcher	668.68	27.95	66.71
Lazega	22.01	3.84	26.66
Network Science	166.01	26.89	80.96
Pokec 1000	352.01	26.74	72.03

Silou riadený algoritmus

Navrhnutý silou riadený algoritmus vytvára grafické reprezentácie pomocou emulovaného systému fyzikálnych síl. Hlavným znakom silou riadených algoritmov je fyzikálna príťažlivosť a odpudzovanie medzi vrcholmi reprezentovanými hranami grafu. Fyzikálne sily umožňujú vytváranie prehľadných a zrozumiteľných rozložení vrcholov a hrán v 3D priestore bez nutnosti zásahu používateľa pri procese zostavovania grafu. Nakoľko na uzly pôsobia rôzne sily, môže dôjsť k situáciám, kedy sa vrcholy zobrazia mimo zorného poľa kamery.

Tabuľka 5 - Výsledky experimentov so silou riadeným algoritmom (v abstraktných jednotkách používaných v Unity Engine).

Názov datasetu	Priemerná vzdialenosť vrcholov	Priemerná hustota zhlukov	Priemerná vzdialenosť zhlukov
Game of Thrones	100.22	7.14	18.86
Witcher	228.06	11.70	17.23
Lazega	20.17	2.41	6.42
Network Science	28.66	5.47	15.09
Pokec 1000	113.23	15.17	60.68

Výsledky silou riadeného algoritmu sú uvedené v Tabuľka 5. V porovnaní s pružinovým algoritmom je zrejmé, že finálna vizualizácia vyžaduje menej priestoru na zobrazenie výsledku, má nižšiu priemernú hustotu klastrov a ešte menšiu priemernú hustotu vnútri klastrov. Tým lepšie vyhovuje potrebám pre akceptovateľnú 3D vizualizáciu.

Pružinovo-silový algoritmus s kontrolovaným rozmiestňovaním

Kombináciou dvoch predtým definovaných prístupov sme dosiahli vyvážený spôsob, pri ktorom pružiny a prítlačivé sily spolupracujú na stabilizácii grafu. Vizualizácia ukazuje, aké dôležité je nastaviť parametre síl a pružín tak, aby sa dosiahla optimálna rovnováha medzi rýchlou stabilizáciou siete a minimalizáciou zahltenia modelu. Kombinácia prístupov tiež zabráňuje vytvoreniu jediného veľkého zhluku. Finálna vizualizácia preukazuje dobrú výkonnosť pri identifikácii zhlukov, ale mnohé vrcholy sú zobrazené mimo oblasti snímanej kamerou v 3D priestore. Na základe tohto faktu sme pristúpili k aplikácii kontrolovaného umiestnenia v kombinovanom algoritme, ktorý považujeme za náš výsledný silovo riadený algoritmus.

Vizualizácia siete vo voľnom 3D priestore má niekoľko nevýhod, ktoré je potrebné zohľadniť pri konštrukcii obmedzujúcich podmienok. Napríklad vznik zhlukov zhoršuje prehľadnosť siete, pretože vrcholy ovplyvnené menšími prítlačivými silami alebo vrcholy ovplyvnené nadmernými odpudzivými silami blokujú pohľad kamery.

Nekontrolované prítlačivé a odpudzivé sily v sieti môžu vytvárať nežiadúce zhluky nečitateľné pre používateľov. Aby sme sa vyhli situáciám, kedy zhluk vrcholov zakrýva zvyšok vizualizácie, implementovali sme obmedzenie vzdialenosti medzi vrcholmi. Toto obmedzenie je reprezentované sférickým obmedzením, ktorá slúži ako stena pre ostatné vrcholy. Veľkosť sférického poľa je priamo odvodená od veľkosti samotného vrcholu – väčšie vrcholy disponujú väčšími sférickými obmedzeniami. Tento prístup umožňuje prehľadnejšiu vizualizáciu.

Experimentálne výsledky v Tabuľka 6 ukazujú najlepší výkon na základe použitých metrík. Navyše použitím silou riadeného algoritmu s pružinami a kontrolovaným umiestnením sme boli schopní eliminovať nedostatky pozorované v predchádzajúcich implementáciách založených na pružinách alebo všeobecnom silovo riadenom prístupe.

Tabuľka 6 - Výsledky experimentov s pružinovo-silovo riadeným algoritmom s kontrolovaným rozmiestňovaním (v abstraktných jednotkách používaných v Unity Engine).

Názov datasetu	Priemerná vzdialenosť vrcholov	Priemerná hustota zhlukov	Priemerná vzdialenosť zhlukov
Game of Thrones	230.41	11.32	46.25
Witcher	378.91	16.43	42.34
Lazega	73.96	4.86	12.63
Network Science	92.57	6.04	29.33
Pokec 1000	310.08	14.33	55.17

Celkové experimentálne výsledky získané pre definované metriky sú uvedené v Tabuľka 7. Pre jednoduchšiu identifikáciu sú najlepšie hodnoty zvýraznené tučným písmom, pretože sa snažíme nájsť riešenie bez extrémneho umiestnenia. Ako vidno, iteratívnym vývojom sme sa pre väčšinu súborov dát dokázali prekonať pôvodné nedostatky silou riadeného algoritmu pre 3D prostredie, alebo sme sa k najlepšiemu riešeniu veľmi priblížili.

Tabuľka 7 - Prehľad výkonnostných metrík pre jednoduchý pružinový algoritmus, force-directed algoritmus a force-directed algoritmus s pružinami a kontrolovaným rozmiestňovaním

Názov datasetu	Algoritmus	Priemerná vzdialenosť vrcholov	Priemerná hustota zhlukov	Priemerná vzdialenosť zhlukov
Game of Thrones	Pružinový	348.98	28.65	69.67
	Silou riadený	100.22	7.41	18.86
	Pružinovo-silový SKR*	230.41	11.32	46.25
Witcher	Pružinový	668.68	27.95	66.71
	Silou riadený	228.06	11.70	17.23

	Pružinovo-silový SKR*	378.91	16.43	42.34
Lazega	Pružinový	22.01	3.84	26.66
	Silou riadený	20.17	2.41	6.42
	Pružinovo-silový SKR*	73.96	4.86	12.63
Network Science	Pružinový	166.01	26.89	80.96
	Silou riadený	28.66	5.47	15.09
	Pružinovo-silový SKR*	92.57	6.04	29.33
Pokec1000	Pružinový	352.01	26.74	72.03
	Silou riadený	113.23	15.17	60.68
	Pružinovo-silový SKR*	310.08	14.33	55.17

Pri tisícoch vrcholov a hrán je oveľa jednoduchšie pochopiť internú štruktúru dát a ľahko rozlíšiť rôzne typy dát podľa definovaných klastrov. V 2D je možné zlepšiť vizualizáciu ďalšími iteračnými krokmi a vytvoriť užitočnejší model siete.

V práci celkovo vznikli 3 hlavné algoritmy v 3D spolu s ich modifikáciami a to pružinový algoritmus, silou riadený algoritmus a pružinovo-silový algoritmus. Tieto algoritmy poskytujú intuitívne a prehľadné spôsoby vizualizácie na rôznych typoch dát o rôznej veľkosti, pričom najlepšie výsledky dosahuje pružinovo-silový algoritmus s kontrolovaným rozmiestňovaním, pri väčšine pozorovaných datasetoch. Usudzujeme, že vytvorené algoritmy napĺňajú tézu č.2. Vytvorené algoritmy pracujú s predpokladom, že dáta boli predspracované a pripravené podľa metodiky na prípravu dát opísanú v kapitole 0 a v prílohe A, ktorá nám zabezpečuje vyrovnané fyzikálne vlastnosti v modeli a obmedzuje nežiadúce stavy vizualizácií, pričom napĺňa tézu č.1. Podľa kapitoly 0, usudzujeme, že na interakciu v 3D priestore je potrebné zvoliť vhodné vstupné zariadenia, aby bola zachovaná dostatočná úroveň spätnej väzby vizualizácie a používateľa. Usudzujeme, že opísaný spôsob výberu a použitia vstupných zariadení pre interakciu v 3D napĺňa tézu č.3 a posúva hranice vizualizácie na vyššiu úroveň, ktorú nie je možné dosiahnuť použitím klasických nástrojov ako myš a klávesnica.

Záver

V práci sme preskúmali dôležitosť vizualizácie dát v trojrozmernom priestore a jej úlohu pri analýze a porozumení komplexných vzťahov v rôznych oblastiach výskumu a aplikácií.

Naším hlavným cieľom bolo efektívne vizualizovať veľké objemy dát v 3D priestore pomocou herných prostredí a silou riadených algoritmov.

Na základe našich experimentov sme zistili, že využitie fyzikálnych herných prostredí ponúka efektívnu možnosť vizualizovať veľké dátové množiny. Avšak, prevod algoritmov z 2D do 3D prostredia nie je triviálny proces a algoritmy navrhnuté pre 2D priestor môžu byť neefektívne alebo dokonca nepoužiteľné v 3D. Preto sme sa v prvej rade zamerali na vývoj algoritmov špecificky prispôbených pre vizualizáciu veľkých dátových množín v 3D priestore.

Hlavným prínosom tejto práce je vytvorenie metodiky a algoritmov založených na princípe silou riadených algoritmov, ktoré umožňujú efektívnu vizualizáciu veľkých dátových sietí v 3D. Algoritmy obsahujú modifikácie ako sférické zobrazenia, normalizáciu, zhlukovanie, pružinové spoje a gravitačné interakcie, čím vytvára robustnú techniku na vizualizáciu komplexných dátových sietí. Využitie vytvorených algoritmov prináša inovatívny prístup k vizualizácii dát a môže byť využitý v rôznych oblastiach výskumu a aplikácií, kde je potrebné vizualizovať veľké dátové množiny v trojrozmernom priestore, ako aj v oblastiach analýzy kde vytvorené algoritmy obohacujú pozorované dáta o nové vlastnosti založené na fyzických prepojeniach.

Okrem samotného algoritmu sme sa tiež venovali riešeniu otázky, ako správne pripraviť dáta na vizualizáciu v 3D fyzikálnom hernom prostredí. Zistili sme, že správne upravenie a

spracovanie dát, spolu s vhodnými vizualizačnými technikami a postupmi, sú kľúčové pre efektívnu vizualizáciu veľkých objemov dát v 3D priestore. Pre tento účel vznikla metodika pre prípravu dát na vizualizáciu v 3D priestore, ktorá obsahuje postupy, hodnoty a odporúčenia pre správnu úpravu vstupných dát. Využitím vytvorenej metodiky v spojení s vytvorenými algoritmami je možné dosiahnuť vysokú úroveň vizualizácie dát v 3D priestore.

Ďalšou dôležitou témou našej práce bolo využitie vstupných zariadení poskytujúcich 3D vstupy pre zabezpečenie efektívnej interakcie s 3D vizualizáciou. Využitie týchto zariadení umožňuje používateľom interagovať s dátami a získať nové poznatky o vzťahoch medzi nimi, čím sa zvyšuje užitočnosť vizualizačných nástrojov. Zvolenie správnych vstupných zariadení je kľúčové pre dosiahnutie dostatočnej úrovne interakcie, ktorá poskytuje voľnosť v prehľadávaní dát, a zároveň neobmedzuje používateľa alebo nezakrýva informácie.

Celkovo možno usúdiť, že jednotný postup práce s dátami a ich vizualizáciou v 3D priestore je stále predmetom výskumu. Na základe výsledkov tejto práce sme však zistili, že vizualizácia veľkého objemu dát v 3D priestore je realizovateľná a prináša cenné poznatky v rôznych oblastiach výskumu a aplikácií. Zároveň sme zistili, že správnym spojením prípravy dát, vytvorených algoritmov, výberom vhodných modifikácií a výberom vhodných vstupných zariadení, vytvorené algoritmy dokážu obohatiť dáta o priestorové informácie a nové vlastnosti založené na fyzikálnych vlastnostiach. Obohatené dáta je ďalej možné skúmať a iteratívnym procesom analyzovať a hlbšie porozumieť významu skúmaných dát ako celku. Veríme, že práca v konečnom dôsledku predstavuje nové spôsoby analýzy dát ako aj interakciu dátového analytika s trojrozmerným priestorom, čo považujeme za veľký posun v oblasti vizualizácie dát v 3D priestore a v oblasti silou riadených algoritmov.

Zoznam použitej literatúry

- [1] S. Bassil a R. K. Keller, „Software visualization tools: Survey and analysis,“ rev. *IWPC 2001: Proceedings 9th International Workshop on Program Comprehension*, 2011.
- [2] J. M. Hernández a P. Van Mieghem, „Classification of graph metrics,“ *Delft University of Technology: Mekelweg*, pp. 1-20, 2011.
- [3] R. S. Raghav, S. Pothula, T. Vengattaraman a D. Ponnurangam, „A survey of data visualization tools for analyzing large volume of data in big data platform,“ rev. *International Conference on Communication and Electronics Systems (ICCES)*, 2016.
- [4] West, Douglas a Brent, *Introduction to graph theory*, Prentice hall Upper Saddle River, 2001.
- [5] F. Mcgee, M. Ghoniem, G. Melancom a B. Pinaud, „The State of the Art in Multi-Layer Network Visualization,“ rev. *Computer Graphics Forum 2019*, 2019.
- [6] Wikipedia - The Free Encyclopedia, „Graph drawing,“ [Online]. Available: https://en.wikipedia.org/wiki/Graph_drawing. [Cit. 30 01 2024].
- [7] J. Mathieu, V. Tommaso, H. Sebastien a B. Mathieu, „Force Atlas 2, a Continuous Graph Layout Algorithm for Handy Network Viasualization Design for the Gephi Software,“ *PLOS ONE*, zv. 9, %1. vyd.6, 2014.
- [8] S. M. Anouncia, H. A. Gohel a V. Subbiah, *Data Visualization: Trends and Challenges Toward Multidisciplinary Perception*, Springe Singapore, 2020.
- [9] J. Schmidt, „Visual Data Science,“ rev. *Data Science, Data Visualization, and Digital Twins*, 2022.
- [10] M. Vuckovic, J. Schmidt, T. Ortner a D. Cornel, „Combining 2D and 3D Visualization with Visual Analytics in the Environmental Domain,“ *Information*, zv. 13, %1. vyd.7, 2022.
- [11] B. D. Fisher, „Visual representations and interactions technologies,“ *Illuminating the Path: A Research and Development Agenda for Visual Analytics*, pp. 69-104, 2005.
- [12] F. Amini, S. Rufiange, Z. Hossain, Q. Ventura, P. Irani a M. J. McGuffin, „The impact of interactivity on comprehending 2D and 3D visualizations of movement data,“ *IEEE transactions on visualization and computer graphics*, zv. 21, %1. vyd.1, pp. 122-135, 2014.

- [13] N. Elmqvist a P. Tsigas, „A taxonomy of 3d occlusion management for visualization,“ *IEEE transactions on visualization and computer graphics*, zv. 14, %1. vyd.5, pp. 1095-1109, 2008.
- [14] S. Bleisch a S. Nebiker, „Connected 2D and 3D visualizations for the interactive exploration of spatial information,“ rev. *Proc. of 21th ISPRS Congress*, Beijing, China, 2008.
- [15] T. Ortner, *Tight integration of visual analysis and 3D real-time rendering (Ph. D. thesis)*, Technische Universitat Wien, 2020.
- [16] C. Nussbaumer Knafllic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*, Wiley, 2015.
- [17] K. Healy, *Data Visualization: A Practical introduction*, Princeton University Press, 2018.
- [18] A. Kirk, *Data Visualisation: A Handbook for Data Driven Design*, 2nd Edition, SAGE Publications Ltd, 2019.
- [19] B. R. Kent, *3D Scientific Visualization with Blender (Iop Concise Physics)*, Morgan & Claypoll, 2015.
- [20] M. Grandjean, „Digital History | Data Visualization | Network Analysis,“ [Online]. Available: <https://www.martingrandjean.ch>. [Cit. 01 2024].
- [21] B. Bollobás, *Modern Graph Theory*, New York, NY: Springer, 1998.
- [22] R. Diestel, *Graph Theory*, Berlin: Springer, 2017.
- [23] J. L. Gross a J. Yellen, *Graph Theory and Its Applications (Textbooks in Mathematics)* 2nd Ed., Chapman and Hall/CRC, 2005.
- [24] M. Newman, *Networks*, 2nd Edition, Oxford University Press, 2018.
- [25] A.-L. Barabási a M. Pósfai, *Network Science*, Cambridge University Press, 2016.
- [26] T. H. Cormen, C. E. Leiserson, R. L. Rivest a C. Stein, *Introduction to Algorithms*, 3rd Edition, MIT Press, 2009.
- [27] M.-Y. Kao, *Encyclopedia of Algorithms*, 2nd ed., Springer, 2016.
- [28] S. S. Skiena, *The Algorithm Design Manual (Texts in Computer Science)* 3rd ed., Springer, 2020.

- [29] S. Palúch, *Algoritmická teória grafov*, Žilina: EDIS, 2020.
- [30] J. Janáček, M. Koháni, A. Szendreyová a L. Buzna, *Diskrétna optimalizácia*, Žilina: EDIS, 2015.
- [31] M. B. Fronc, *Kvantitatívne metódy*, Žilina: EDIS, 1998.
- [32] K. Cherven, *Mastering Gephi Network Visualization*, Packt Pub Ltd, 2015.
- [33] K. Cherven, *Network Graph Analysis and Visualization With Gephi*, Packt Pub Ltd, 2013.
- [34] T. M. J. Fruchterman a E. M. Reingold, „Graph drawing by force-directed placement,“ *Journal of Software: Practice and Experience*, zv. 21, %1. vyd.11, pp. 1129-1164, 1991.
- [35] Y. F. Hu, „Efficient and high quality force-directed graph drawing,“ *The Mathematica Journal*, zv. 10, pp. 37-71, 2005.
- [36] J. Barnes a P. Hut, „A hierarchical $O(N \log N)$ force-calculation algorithm,“ *Nature*, zv. 324, %1. vyd.4, pp. 446-449, 1986.
- [37] S. Martin, W. M. Brown, R. Klavans a K. Boyack, „OpenOrd: An Open-Source Toolbox for Large Graph Layout,“ rev. *SPIE Conference on Visualization and Data Analysis (VDA)*, 2011.
- [38] M. Jacomy, T. Venturini, S. Heymann a M. Bastian, „ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software,“ *PLOS ONE*, zv. 9, %1. vyd.6, p. <https://doi.org/10.1371/journal.pone.0098679>, 2014.
- [39] M. Bastian, S. Heymann a M. Jacomy, „Gephi: An Open Source Software for Exploring and Manipulating Networks,“ rev. *International AAAI Conference on Weblogs and Social Media*, 2009.
- [40] S. Heymann, „Gephi Wiki,“ 1 3 2015. [Online]. Available: <https://github.com/gephi/gephi/wiki>. [Cit. 1 2024].
- [41] D. Kaufmann a D. Wagner, *Drawing Graphs*, Berlin: Springer, 1998.
- [42] G. di Battista, P. Eades, R. Tamassia a I. Tollis, *Graph Drawing*, Englewood Cliffs: Prentice Hall, 1999.
- [43] J. Tukey, *Exploratory Data Analysis*, New York: Addison-Wesley, 1977.
- [44] E. Tufte, *The Visual Display of Quantitative Information*, Cheshire: Graphics Press, 1983.

- [45] W. Cleveland, *The Elements of Graphics Data*, Monterey: Wadsworth, 1985.
- [46] W. Cleveland, *Visualizing Data*, Summit: Hobart Press, 1993.
- [47] D. Khokhar, *Gephi Cookbook*, Packt Pub Ltd, 2015.
- [48] E. D. Kolaczyk a G. Csárdi, *Statistical Analysis of Network Data with R*, New York: Springer, 2014.
- [49] L. Takac a M. Zabovsky, „Data Analysis in Public Social Networks,” rev. *International Scientific Conference & International Workshop Present Day Trends of Innovations*, Lomza, Poland, 2012.
- [50] F. Shadhin, „How to Represent a Graph Data Structure in Python,” 5 2022. [Online]. Available: <https://medium.com/geekculture/how-to-represent-a-graph-data-structure-in-python-9f0df57e33a2>. [Cit. 01 2024].
- [51] H. Melcher, „Python Graph Libraries,” 03 12 2022. [Online]. Available: <https://wiki.python.org/moin/PythonGraphLibraries>. [Cit. 1 2024].
- [52] G. Sabidussi, „The centrality index of a graph,” *Psychometrika*, zv. 31, pp. 581-683, 1966.
- [53] L. Freeman, „A set of measures of centrality based on betweenness,” *Sociometry*, zv. 40, %1. vyd.1, pp. 35-41, 1977.
- [54] P. Bonacich, „Factoring and weighting approaches to status scores and clique identification,” *Journal of Mathematical Sociology*, zv. 2, %1. vyd.1, pp. 113-120, 1972.
- [55] L. Katz, „A new status index derived from sociometric analysis,” *Psychometrika*, zv. 18, %1. vyd.1, pp. 39-43, 1953.
- [56] J. Kleinberg, „Authoritative sources in a hyperlinked environment,” *Journal of ACM*, zv. 46, %1. vyd.5, pp. 604-632, 1999.
- [57] E. Lazega, „Lazega lawyers network data,” 2001. [Online]. Available: <https://search.r-project.org/CRAN/refmans/sand/html/lazega.html>. [Cit. 1 2024].
- [58] M. E. J. Newman, „Finding community structure in networks using the eigenvectors of matrices,” *Physical Review E*, zv. 74, 2006.
- [59] G. Di Battista, P. Eades, R. Tamassia a I. G. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs*, Englewood Cliffs: Prentice Hall, 1999.

- [60] U. Brandes, Drawing on physical analogies., Drawing Graphs, volume 2025 of Lecture Notes in Computer Science ed., zv. 2025, M. Kaufmann a D. Wagner, Ed., Springer-Verlag, 2001, pp. 71-86.
- [61] W. T. Tutte, „How to draw a graph.,“ rev. *Proc. London Math. Society*, 1963.
- [62] P. Eades, „A heuristic for graph drawing,“ rev. *Congressus Numerantium*, 1984.
- [63] S. G. Kobourov, „Force-Directed Drawing Algorithms,“ rev. *Handbook of Graph Drawing and Visualization*, R. Tamassia, Ed., CRC Press, 2013, pp. 383-408.
- [64] T. Kamada a S. Kawai, „An algorithm for drawing general undirected graphs,“ *Inform. Process. Lett.*, zv. 31, pp. 7-15, 1989.
- [65] R. Hadany a D. Harel, „A multi-scale algorithm for drawing graphs nicely,“ *Discrete Applied Mathematics*, zv. 113, %1. vyd.1, pp. 3-21, 2001.
- [66] D. Harel a Y. Koren, „A fast multi-scale method for drawing large graphs,“ rev. *Graph Drawing. GD 2000. Lecture Notes in Computer Science*, zv. 1984, J. Marks, Ed., Berlin, Heidelberg, Springer, 2002, pp. 179-200.
- [67] C. Walshaw, „A multilevel algorithm for force-directed graph drawing,“ *Journal of Graph Algorithms and Applications*, zv. 7, %1. vyd.3, pp. 253-285, 2003.
- [68] P. Gajer, M. T. Goodrich a S. G. Kobourov, „A fast multi-dimensional algorithm for drawing large graphs,“ *Computational Geometry: Theory and Applications*, zv. 29, %1. vyd.1, pp. 3-18, 2004.
- [69] G. Csardi a T. Nepusz, „python-igraph,“ 2024. [Online]. Available: <https://python.igraph.org/en/stable/>. [Cit. 01 2024].
- [70] G. Csardi a T. Nepusz, „The igraph software package for complex network research,“ *InterJournal, Complex Systems*, zv. 1695, 2006.
- [71] J. Lamping, R. Rao a P. Pirolli, „A focus+context technique based on hyperbolic geometry for visualizing large hierarchies,“ rev. *Proceedings of Computer Human Interaction*, 1995.
- [72] T. Munzner, „H3: Laying out large directed graphs in 3D hyperbolic space,“ rev. *Proceedings of IEEE Symposium on Information Visualization*, 1997.
- [73] S. G. Kobourov a K. Wampler, „Non-Euclidean spring embedders,“ *IEEE Transactions on Visualization and Computer Graphics*, zv. 11, %1. vyd.6, pp. 757-767, 2005.

- [74] A. Brezáni a M. Zábovský, „3D Graph Visualization Performance using Physical Engine,“ *INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGIES, ENGINEERING AND MANAGEMENT SCIENCE*, zv. 2, %1. vyd.6, 2022.
- [75] E. Lazega, *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership.*, Oxford: Oxford University Press, 2001.