

**Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky**

**Studijní program: P1301 Geografie
Obor: Geoinformatika a kartografie**

**TEORIE GRAFŮ PŘI VYHODNOCOVÁNÍ
ČASOPROSTOROVÝCH DAT**

Disertační práce

Mgr. Jitka DOLEŽALOVÁ

Školitel: prof. Mgr. Radomír Halaš, Dr.

Olomouc 2018

KLÍČOVÁ SLOVA

teorie grafů, kliky v grafu, podobnost, shluková analýza

Počet stran práce: 100

Počet příloh: 1 (z toho 1 elektronická)

KEYWORDS

Graph Theory, Graph Cliques, Similarity, Cluster Analysis

Number of pages: 100

Number of appendixes: 1 (electronic)

Prohlašuji, že

- disertační práci včetně příloh, jsem vypracovala samostatně a uvedla jsem všechny použité podklady a literaturu.

- jsem si vědoma, že na moji disertační práci se plně vztahuje zákon č. 121/2000 Sb., autorský zákon, zejména § 35 využití díla v rámci občanských a náboženských obřadů, v rámci školních představení, a využití díla školního a § 60 školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, disertační práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk disertační práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé disertační práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé disertační práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé disertační práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci 8. června 2018

Mgr. Jitka Doležalová

Děkuji svému školiteli prof. Radomírovi Halašovi za podněty a připomínky při vypracování práce. Dále děkuji konzultantu Jardovi Burianovi za nápady i elán, Standovi za trpělivost a morální podporu, Alence za precizní korektury a všem kolem za pevné nervy.

A ještě jednou zvláštní poděkování patří Staníkovi, bez kterého by tato práce neměla vůbec šanci vzniknout, rozvíjet se, ani finišovat.

OBSAH

ÚVOD	8
1 CÍLE PRÁCE	9
1.1 DC 1 – Využití grafů v analýze eye-tracking dat	9
1.2 DC 2 – Využití grafů v prostorovém plánování.....	9
1.3 DC 3 – Využití grafů v dopravě.....	10
1.4 DC 4 – Případové studie využívající vytvořené nástroje.....	10
2 METODY A POSTUPY ZPRACOVÁNÍ	12
2.1 DC1 – Využití teorie grafů v analýze eye-tracking dat.....	13
2.2 DC2 – Využití teorie grafů v prostorovém plánování	14
2.3 DC3 – Využití grafů v dopravě.....	14
2.4 DC4 – Případové studie využívající vytvořené nástroje	14
2.5 Software	15
2.6 Data	15
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	16
3.1 Použité pojmy z teorie grafů	16
3.2 Využití teorie grafů v geoinformatice	20
3.2.1 Využití teorie grafů v analýze eye-tracking dat	22
3.2.2 Využití teorie grafů v prostorovém plánování.....	26
3.2.3 Využití grafů v dopravě	30
3.2.4 Další využití teorie grafů v geoinformatice	31
4 DC1 – SCANGRAPH - TEORIE GRAFŮ V ANALÝZE EYE-TRACKING DAT	32
4.1 Důvody vytvoření vlastního nástroje	32
4.2 Popis nástroje ScanGraph	36
4.3 Výpočty ve ScanGraph.....	36
4.3.1 Modifikace algoritmů a popis jejich výpočtu	36
4.3.2 Doporučení pro volbu algoritmu	37
4.3.3 Určení míry podobnosti	38
4.3.4 Vizualizace výsledků	39
4.4 Použití ScanGraph (uživatelé)	39
4.4.1 Import dat	40
4.4.2 Použití ScanGraph.....	42
4.4.3 Ukázka výsledků a jejich interpretace	45
4.5 Limity nástroje ScanGraph	48
5 DC2 – SIMURB – TEORIE GRAFŮ V PROSTOROVÉM PLÁNOVÁNÍ	50
5.1 Důvody vytvoření vlastního nástroje	50
5.2 Popis nástroje SimUrb	50
5.3 Výpočty v SimUrb	51
5.3.1 Metrika.....	51
5.3.2 Určení míry podobnosti	51
5.3.3 Vizualizace výsledků	52

5.4	Použití SimUrb (uživatelé).....	52
5.4.1	Import dat	52
5.4.2	Použití SimUrb	53
5.4.3	Interpretace výsledků	54
6	DC3 – SIMROAD – TEORIE GRAFŮ V DOPRAVĚ	55
6.1	Důvody vytvoření vlastního nástroje	55
6.2	O nástroji SimRoad jak byl vytvořen	55
6.3	Výpočty v SimRoad.....	55
6.3.1	Metrika.....	55
6.3.2	Určení míry podobnosti	55
6.3.3	Vizualizace výsledků.....	56
6.4	Použití SimRoad (uživatelé).....	56
6.4.1	Import dat	56
6.4.2	Použití SimRoad	57
6.4.3	Interpretace výsledků	57
7	DC4 – PŘÍPADOVÉ STUDIE VYUŽÍVAJÍCÍ VYTVOŘENÉ NÁSTROJE.....	58
7.1	Případové studie <i>Scanpath Comparison</i>	58
7.1.1	Vizualizace nejistoty geografických dat	58
7.1.2	Hodnocení nefotorealistickej 3D vizualizace měst	62
7.1.3	Hodnocení kartografické gramotnosti.....	67
7.1.4	Další studie využívající ScanGraph	69
7.2	Případové studie <i>Urban Similarity</i>	71
7.2.1	Hodnocení udržitelného rozvoje území	71
7.2.2	Časoprostorové analýzy typu a stavu zemědělských plodin	80
7.2.3	Porovnání polyline a star glyphů.....	83
7.3	Případová studie <i>Road Similarity</i>	83
7.3.1	Analýza toků veřejné dopravy na základě dat z mýtných bran	83
8	VÝSLEDKY.....	88
8.1	DC1 – Využití grafů v analýze eye-tracking dat	88
8.2	DC 2 – Využití teorie grafů v prostorovém plánování	89
8.3	DC 3 – Využití grafů v dopravě.....	90
8.4	DC 4 – Případové studie využívající vytvořené nástroje.....	90
9	DISKUZE.....	92
10	ZÁVĚR	95
11	POUŽITÉ ZDROJE	96

ÚVOD

Teorie grafů jakožto široce se rozvíjející disciplína sahá svými kořeny až do první poloviny 18. století. Za několik století se na jejím rozvoji podepsala spousta zvučných jmen a inspiraci z ní čerpají výzkumníci dodnes. Výjimkou není ani geoinformatika, která je oproti teorii grafů mladým oborem. Rychlý rozvoj počítačové technologie umožňuje uvedeným oborům své cesty proplétat a ve velké míře tak určuje směr jejich vývoje.

Disertační práce se zabývá použitím teorie grafů při vyhodnocování časoprostorových dat. Charakter časoprostorových dat intuitivně vybízí při svém zpracování k využívání teorie grafů a grafových algoritmů. Přirozeně je tak možné jejich vzájemné aplikace v řadě oblastí. Disertační práce si neklade za cíl vyhledat tyto nespočetné vzájemné interakce, ale doplnit je o další možné pojetí a způsob využití.

Úkolem geoinformatika bývá často vyhodnocení rozsáhlého experimentu či případové studie, kde je potřeba zkoumané subjekty rozčlenit do skupin se stejným charakterem. Jelikož je nemožné velké množství dat jednoznačně roztřídit, je vyžadováno, aby subjekty vykazovaly alespoň jistou (subjektivní) podobnost. Disertační práce se tedy speciálně zaměřuje na vyhledávání skupin podobných sekvencí, ať už v datech z eye-tracking experimentů, datech urbánních či dopravních. Teoretická část práce je tak demonstrována v několika možných aplikacích.

1 CÍLE PRÁCE

Hlavním cílem práce je navrhnout metody pro hledání tříd podobnosti využívající obyčejné grafy, vytvořit nástroje využívající tyto metody a ověřit jejich funkčnost při aplikaci v časoprostorových úlohách v dopravě, urbanismu (prostorovém plánování) a v hodnocení kartografických děl metodou eye-tracking. V dalším textu práce je používán více pojem *prostorové plánování* než *urbanismus*, jelikož blíže vystihuje zaměření výpočetních metod.

V první části práce bude zhodnocen vztah obyčejných grafů a časoprostorových dat a budou zavedeny základní pojmy z teorie grafů pro potřeby tří vytvářených nástrojů. Tyto nástroje budou aplikovány ve výzkumných případových studiích. Následně budou vytvořeny webové nástroje umožňující hledání skupin vzájemně si podobných sekvencí v časoprostorových datových sadách. V první části bude nástroj využit pro analýzu dat o pohybu očí při čtení (nejen) kartografických děl a uzpůsoben pro co nejjednodušší použití. V další části řešení disertační práce bude vytvořený nástroj transformován pro využití k řešení časoprostorových úloh v prostorovém plánování, kde bude možné nalézt skupiny podobných obcí na základě předem definovaných vlastností, a také v dopravě, kde budou hledány úseky dopravní sítě s podobnou průjezdností.

Výstupem disertační práce budou tři vytvořené nástroje pro hledání tříd podobnosti. Práce přinese relevantní výsledky případových studií, ve kterých budou nástroje aplikovány na konkrétní data ze tří tematických oblastí (hodnocení kartografických děl pomocí eye-tracking technologie), prostorové plánování a doprava). Výsledky těchto studií budou obsahovat okomentovaný popis vztahů mezi sledovanými entitami, statistická hodnocení a mapy.

Cíle disertační práce jsou rozděleny na čtyři dílčí části, dále označovány jako DC1 až DC4. Tyto dílčí cíle práce jsou vymezeny v následujících podkapitolách.

1.1 DC 1 – Využití grafů v analýze eye-tracking dat

Náplní prvního dílčího cíle bude návrh metody *Scanpath Comparison* využívající teorie grafů pro hledání podobnosti v eye-tracking datech. Tato metoda bude následně implementována při tvorbě webového nástroje pro časoprostorovou analýzu eye-tracking dat. Tento nástroj s názvem *ScanGraph* bude vyhledávat podobné uživatelské strategie čtení mapy na základě porovnávání řetězců navštívených oblastí zájmu.

Požadavky na funkcionalitu nástroje byly definovány v návaznosti na odborné konzultace s odborníky na využití eye-tracking technologie následovně:

- import dat přímo z prostředí programu OGAMA,
- stanovení míry podobnosti uživatelem nástroje,
- nalezení všech skupin vzájemně si podobných řetězců oblastí zájmu při dané míře podobnosti,
- možnost exportu souboru s výpočty a původními daty,
- možnost permanentního odkazu na výsledek analýzy,
- možnost analýzy více stimulů najednou.

1.2 DC 2 – Využití grafů v prostorovém plánování

Náplní druhého dílčího cíle bude návrh metody využívající teorie grafů pro hledání podobných obcí na základě jejich kvantitativních indikátorů. Tato metoda bude následně implementována při tvorbě webového nástroje *SimUrb* pro analýzu časoprostorových dat v urbánních úlohách, která budou ve formátu uspořádané n -tice ohodnocených indikátorů přiřazených jednotlivým sídelním jednotkám. Nástroj bude vyhledávat

disjunktí skupiny sídelních jednotek, které budou vykazovat míru podobnosti danou uživatelem.

Požadavky na funkcionalitu nástroje byly na základě odborných konzultací stanoveny následovně:

- import dat jako CSV,
- určení disjunktí skupin vzájemně si podobných sídelních jednotek při dané míře podobnosti,
- stanovení míry podobnosti uživatelem,
- rozlišení sídelních jednotek podle větších administrativních celků (např. ORP, okresy),
- stanovení nejmenší velikosti skupiny, která je pro uživatele zajímavá,
- možnost exportu souboru s výpočty a původními daty,
- možnost permanentního odkazu na výsledek analýzy,
- interaktivní vyznačení podobné skupiny v mapě.

1.3 DC 3 – Využití grafů v dopravě

Náplní třetího dílčího cíle bude návrh metody využívající teorie grafů pro hledání disjunktí skupin silničních úseků vykazujících danou míru podobnosti z hlediska počtu průjezdů. Následně bude vytvořen webový nástroj *SimRoad*, využívající tuto nově vytvořenou metodu. Nástroj bude použitelný zejména pro analýzu průjezdnosti dálničních úseků a rychlostních silnic na základě dat z mýtných bran.

Požadavky na funkcionalitu nástroje byly stanoveny následovně:

- import dat jako CSV,
- určení disjunktí skupin vzájemně si podobných silničních úseků při dané míře podobnosti,
- stanovení míry podobnosti uživatelem,
- rozlišení silničních úseků podle označení (např. D1, R46),
- stanovení nejmenší velikosti skupiny, která je pro uživatele zajímavá,
- možnost exportu souboru s výpočty a původními daty,
- možnost permanentního odkazu na výsledek analýzy,
- interaktivní vyznačení podobných úseků v mapě.

1.4 DC 4 – Případové studie využívající vytvořené nástroje

V předchozích třech dílčích cílech byly navrženy metody pro časoprostorovou analýzu dat ze tří oblastí a následně byly vytvořeny webové nástroje tyto metody využívající. Čtvrtým dílčím cílem je proto aplikace těchto nově vytvořených metod a nástrojů ve třech případových studiích:

- Nástroj *ScanGraph* bude použit v případové studii zaměřené vizualizaci nejistoty v geografických datech.
- Nástroj *SimUrb* bude použit v případové studii zaměřené na hodnocení podobnosti obcí v Olomouckém kraji na základě tří tematických skupin indikátorů a porovnání s dosavadním způsobem hodnocení podobnosti obcí.
- Nástroj *SimRoad* bude použit v případové studii zaměřené na časoprostorovou analýzu podobnosti dálničních úseků na základě porovnání počtu průjezdů hromadné dopravy (autobusů) mýtnými branami.

Navržené metody a nástroje byly využity i v dalších případových studiích jiných autorů. Nástroj *ScanGraph* byl využit například při hodnocení nefotorealisticke 3D vizualizace měst či v hodnocení kartografické gramotnosti. Nástroj *SimUrb* byl využit například při časoprostorové analýze typu a stavu zemědělských plodin či při porovnání polyline a star

glyphů. Nástroj *SimRoad*, jakožto nejmladší z těchto nástrojů, nebyl zatím aplikován jinými autory, respektive autorka práce nemá o takovém využití zatím žádné informace.

Všechny dílčí cíle byly navrženy tak, aby jejich naplnění umožnilo aktivní využití teorie grafů při vyhodnocování časoprostorových dat. Současně byly všechny dílčí kroky zdokumentovány tak, aby byl jasný i teoretický postup a přínos práce.

2 METODY A POSTUPY ZPRACOVÁNÍ

Pro úspěšné vypracování disertační práce bylo zapotřebí seznámit se s teoretickými základy nutných pojmů z teorie grafů a speciálně s postupy pro vyhledávání klik v obyčejných grafech (maximal clique problem, maximum clique problem). Dále bylo nutné seznámit se s běžně používanými metodami pro určování podobných skupin v časoprostorových datech, speciálně s metodami *String Edit Distance* a shlukové analýzy a s funkcionalitou běžně dostupných nástrojů s těmito metodami pracujícími.

V disertační práci byly vytvořeny tři metody a tři nástroje využívající obyčejné grafy a kliky v těchto grafech pro vyhledání podobných objektů ve třech oblastech, a to v oblasti eye-tracking výzkumu, urbanismu a dopravy. Tyto nástroje byly vytvořeny pomocí programovacích jazyků PHP, HTML a C# a s využitím knihoven D3.js a jQuery. Každá z metod vyžaduje jiné přístupy a využívá jiné algoritmy.

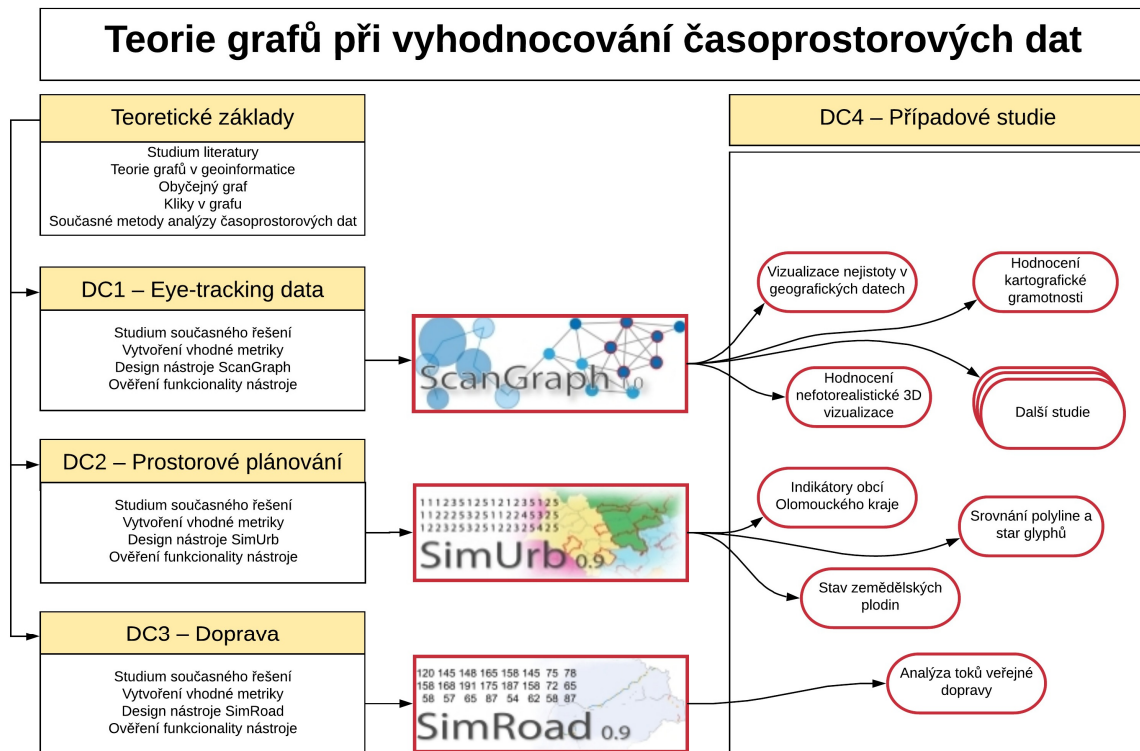
V případě dat z eye-tracking experimentů lze metodu zařadit do problematiky *String Edit Distance*. Metoda vychází z algoritmů pro výpočet vzdálenosti mezi sekvencemi znaků (řetězci písmen). Tyto algoritmy byly optimalizovány speciálně pro aplikaci při určování míry podobnosti mezi eye-tracking daty. Dále byl vytvořen nástroj s názvem *ScanGraph*, jenž vyhodnocuje podobnost mezi eye-tracking daty a využívá tyto navržené algoritmy.

V problematice prostorového plánování byla navržena metoda pro určování podobnosti pro objekty, jejichž vlastnosti jsou charakterizovány uspořádanou n -ticí koeficientů. Metoda vychází z metriky pro výpočet euklidovské vzdálenosti mezi dvěma objekty v n -rozměrném euklidovském prostoru. Byl vytvořen nástroj s názvem *SimUrb*, jenž vyhodnocuje podobnost mezi daty prostorového plánování a využívá navrženou metodu.

Jako posledním okruhem se disertační práce zabývá dopravou, a to konkrétně problematikou průjezdnosti silničními úseky. Jednotlivé silniční úseky mohou být opět chápány jako objekty charakterizovány uspořádanou n -ticí, jež reprezentuje počet průjezdů mýtnými branami za určité časové období. Metoda výpočtu míry podobnosti tedy navazuje na míru podobnosti mezi daty prostorového plánování. Avšak jelikož rozdělení dat průjezdnosti vykazuje nenulovou šikmost, byla metoda uzpůsobena tomuto charakteru s využitím kvantilové funkce. Nástroj s názvem *SimRoad* byl vytvořen tak, aby pracoval s touto navrženou metodou.

Funkcionalita jednotlivých nástrojů byla ověřena na případových studiích ze všech tří oblastí. Součástí vypracování disertační práce byly tři případové studie autorky, které však byly navíc doplněny studiiemi, na nichž bylo spolupracováno s dalšími autory.

Postup práce je schematicky znázorněn na obrázku 1.



Obr. 1 Schéma postupu práce

2.1 DC1 – Využití teorie grafů v analýze eye-tracking dat

První studie je zaměřena na analýzu dat z eye-tracking experimentů. Pro hledání podobných sekvencí v eye-tracking datech je vytvořena metoda ze skupiny *String Edit Distance*. Principem této metody je hledání podobností v sekvencích pohledem navštívených oblastí zájmu. Teoreticky by bylo možné hodnotit přímo podobnost vlastních sekvencí fixací a sakád, avšak výsledky takového srovnání by byly v praxi jen těžko použitelné. Proto je výhodnější agregovat eye-tracking data k oblastem zájmu odpovídajícím logickým částem zobrazovaného stimulu (v případě mapy například ke kompozičním prvkům).

Analyzovaná eye-tracking data byla naměřena pomocí zařízení SMI RED 250 a následně byla pomocí nástroje SMI2OGAMA převedena do prostředí open-source aplikace OGAMA, která slouží pro záznam a analýzu eye-tracking a mouse-tracking dat. Eye-tracking data by bylo možné zaznamenávat přímo v prostředí OGAMA s využitím low-cost zařízení EyeTribe. Jak bylo popsáno v publikaci Popelky a kol. (2016), přesnost tohoto zařízení je pro kartografické účely dostatečná. V aplikaci je však nutné označit oblasti zájmu. Pro korektní výsledek analýzy je vhodné, aby tyto oblasti zájmu odpovídaly logickým celkům analyzovaného stimulu.

Jednou z výhod aplikace OGAMA je funkce *Levenshtein Distance Calculation*, která umožňuje spočítat Levenshteinovu vzdálenost mezi sekvencemi navštívených AOI. Tento výpočet není dostačující pro srovnávání prvků množiny více než dvou řetězců z důvodu jejich rozdílné délky, proto jej nelze přímo použít. Důležité ovšem je to, že je při využití této funkce generována sekvence, kterou lze jednoduše vyexportovat. Vytvořený nástroj *ScanGraph* pracuje přímo s tímto exportem, aby nebylo nutné data nijak upravovat.

Po nahrání dat do prostředí aplikace *ScanGraph* si uživatel zvolí jeden ze tří dostupných algoritmů (viz kapitola 2.2) a pomocí parametru p nebo hustoty grafu zvolí požadovanou míru podobnosti. Výstupem aplikace je obyčejný graf, ve kterém jednotlivé uzly reprezentují uživatele (sekvence). Pomocí Bron-Kerbosch algoritmu (viz kapitola 2.2) jsou

v tomto grafu nalezeny všechny netriviální kliky. Vrcholy nalezených klik jsou v prostředí *ScanGraph* vypsané v uspořádaném seznamu a po kliknutí lze zobrazit sekvence jednotlivých uživatelů obsažených v klikách.

2.2 DC2 – Využití teorie grafů v prostorovém plánování

Pro hledání skupin podobných sídelních jednotek, vykazujících míru podobnosti danou uživatelem, bylo nejprve nutné připravit data, na základě kterých byla podobnost počítána. Jednalo se o hodnoty indexů z oblasti sociální, ekonomické a ekologické, které jsou každoročně počítány v rámci aktualizací Územně analytických podkladů Olomouckého kraje. Každý index může mít přidělenou určitou váhu. Do výpočtu podobnosti pomocí nástroje *SimUrb* vstupuje CSV (Comma Separated Value) tabulka obsahující hodnoty zvolených indexů ve sloupcích a její řádky reprezentují jednotlivé obce. Obce jsou definovány podle kódu Základní územní jednotky (ZUJ). Pro potřeby vizualizace výsledků je v tabulce rovněž obsažen název a příslušnost k vyšší správní jednotce (např. ORP).

Pro vlastní vizualizaci výsledků v mapě je zatím nutné na server dopředu nahrát předem vytvořené vrstvy obcí a ORP ve formátu GeoJSON, se kterými jsou následně data propojena a pomocí nichž jsou výsledky zobrazované v mapě.

Uživatel si v nástroji *SimUrb* volí hodnotu parametru p a minimální velikost skupiny, která jej zajímá.

Nástroj *SimUrb* je navržen tak, aby mohl pracovat se stovkami záznamů. Vyhledávání všech netriviálních klik by v takovémto množství dat bylo obtížné až nemožné a pro danou studii bezúčelné. Proto podobné skupiny se v nástroji *SimUrb* vyhledávají jako disjunktní největší kliky příslušného grafu. Algoritmus pracuje tak, že v prvním kroku vyhledá největší kliku a její vrcholy následně vyřadí ze seznamu. Tento krok opakuje, dokud nezůstanou pouze izolované vrcholy či kliky o menším počtu vrcholů, než bylo požadováno uživatelem. Mohou tedy existovat i kliky, které nebudou výstupem, avšak pro vyhodnocení nejsou významné.

Výsledné netriviální kliky jsou uživateli zobrazeny v pravé části interface programu *SimUrb* obdobně jako v případě *ScanGraph*, ale místo zobrazení pomocí grafu jsou podobné obce zobrazené přímo v mapě.

2.3 DC3 – Využití grafů v dopravě

Metody využití pro tvorbu třetího nástroje s názvem *SimRoad* jsou velmi podobné těm, pomocí kterých byly hledány podobné obce v nástroji *SimUrb*. Data zpracovávaná nástrojem *SimRoad* však mají odlišné rozdělení než data nástroje *SimUrb*. Pro názornější vyhodnocování bude proto určení míry upraveno za pomoci percentilů.

Příprava dat pro výpočet podobnosti úseků silnic na základě dat naměřených mytnými branami je rovněž obdobná jako v případě *SimUrb*. Do výpočtu tentokrát vstupuje CSV tabulka, jejíž řádky reprezentují určité úseky silnic a ve sloupcích jsou uvedeny počty průjezdů vozidel v jednotlivých časových úsecích (např. ve dnech).

Výsledné skupiny podobných obcí jsou vizualizovány jako úseky linií v mapě. Pro znázornění směru vybraného úseku je linie doplněna o šipku.

2.4 DC4 – Případové studie využívající vytvořené nástroje

Vypracované případové studie se dotýkaly témat z oblasti eye-tracking výzkumu, urbanismu a dopravy. Případová studie z oblasti eye-tracking výzkumu se zabývala analýzou podobnosti z experimentu vizualizace nejistoty geografických dat. Pro analýzu byl použit nástroj *ScanGraph* využívající navrženou metodu pro vyhledávání podobných

sekvencí navštívených oblastí zájmu. Případová studie navazovala na diplomovou práci Kučery (2016), v níž byl vytvořen soubor kartografických znaků pro vizualizaci nejistoty v bodovém, liniovém i plošném znaku. Stimuly z tohoto experimentu představovaly mapy s vyznačenými bodovými znaky. Při analýze stimulů v práci Kučery bylo využito eye-tracking metrik jako *Trial Duration* a *Gaze Length*. V případové studii autorky byly analyzovány dva stimuly v podobě ortofotosnímků, a to v prvním případě s možným výskytem prasete divokého, zajíce polního a lišky obecné, a ve druhém případě s možným výskytem pouze lišky obecné.

Pro případovou studii z oblasti urbanismu byly vyhodnocovány územně analytické podklady a byly srovnávány s již existujícími výstupy metodiky používané Olomouckým krajem. Studie byla provedena s použitím nástroje *SimUrb* a navržené metody vyhodnocování podobnosti pro data tohoto charakteru. Podobnost byla hledána mezi obcemi Olomouckého kraje, jež byly charakterizovány systémem atributů sociálního, hospodářského a environmentálního pilíře. Současně byl výstup nástroje *SimUrb* porovnáván i s výstupy nástrojů dostupných v softwaru ArcGIS for Desktop (*Grouping Analysis* a *Multivariate Clustering*), které také vytvářejí shluky objektů na základě předem definovaných atributů.

Ve třetí případové studii, jež se týkala oblasti dopravy, byl aplikován nástroj *SimRoad* s navrženou metodou. Vyhodnocována byla data z mýtných bran s údaji o průjezdech autobusových dopravců. Studie se věnovala srovnání průjezdů v pracovních dnech a o víkendech v rámci jednoho měsíce a to jak pro všechny úseky dálnic, silnic I. třídy a rychlostních komunikací, tak i pouze u dálnic. Z výstupů nástroje byly vybrány ty výstupy, které by mohly být dále studovány v rámci diplomových prací.

2.5 Software

Pro tvorbu webových nástrojů bylo využito programovacích jazyků PHP, HTML a C#. Grafickou vizualizaci obstarává knihovna D3.js, vyšší míru interaktivity zajišťuje knihovna jQuery. Veškeré nástroje popisované v této práci jsou umístěny na serverovém úložišti Univerzity Palackého v Olomouci (konkrétně web-n.upol.cz), pro veřejného uživatele přístupný pod doménou <http://eyetracking.upol.cz>. Server běží na OS Linux s webovým serverem Apache 2.0, PHP verze 5.3.17, databáze MySQL verze 5.0.8. Kompletní konfigurace serveru je dostupná na <http://eyetracking.upol.cz/info.php>.

Příprava geografických dat pro nástroje *SimUrb* a *SimRoad* proběhla v prostředí open-source GIS aplikace QGIS ve verzi 2.18. Všechny nástroje byly primárně testovány pro použití v internetovém prohlížeči Google Chrome.

2.6 Data

Nástroje vzniklé v rámci disertační práce jsou připraveny tak, aby načítaly data ve formátu TXT či CSV. Nástroj *ScanGraph* čte data přímo exportovaná z programu OGAMA ve formě TXT souboru obsahujícího sekvence navštívených oblastí zájmu. Při testování nástroje byla využita data z eye-tracking experimentů vytvořených na Katedře geoinformatiky UP. Nástroj *SimUrb* čte data ve formě CSV tabulky obsahující informace o obcích Olomouckého kraje. Tato data vznikla jako součást územně analytických podkladů Olomouckého kraje. Kromě vlastní tabulky s daty nástroj zobrazuje i mapu obcí Olomouckého kraje, která byla převzata z databáze ArcČR® 500. Nástroj *SimRoad* využívá data z mýtných bran poskytnutá Ředitelstvím silnic a dálnic ČR.

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

V následující kapitole jsou přiblíženy základní použité pojmy z teorie grafů, včetně maticové reprezentace obyčejného grafu (Demel, 2002) a popisu algoritmu pro vyhledávání všech klik v obyčejném grafu (Bron a Kerbosch, 1973).

Dále jsou shrnuty možnosti využití teorie grafů v různých oblastech a přístupech geoinformatiky, včetně nepoužívanějších grafových algoritmů. Jsou také popsány stávající metody pro vyhodnocování časoprostorových dat, a to jak pro určování podobnosti, tak pro vytváření homogenních shluků, včetně stručného popisu aktuálně používaných nástrojů.

3.1 Použité pojmy z teorie grafů

Konečným neorientovaným grafem (dále jen grafem) rozumíme uspořádanou trojici $G = (V, E, p)$, kde V a E jsou konečné množiny, přičemž V je neprázdná množina a p je zobrazení množiny E do množiny všech neuspořádaných dvojic (u, v) , $u, v \in V$. Prvky množiny V nazýváme vrcholy grafu G , prvky množiny E hranami grafu G a zobrazení p incidencí grafu G .

Při studiu geoinformatiky je možné se často setkat i s grafy orientovanými. Orientovaným grafem rozumíme takový graf, kde množina E je množina všech uspořádaných dvojic (u, v) , $u, v \in V$. Rozlišujeme tak počáteční a koncový vrchol hrany. Výstupy této práce se však týkají pouze grafů neorientovaných. Nadále se tedy kapitola 4.1 bude věnovat pouze neorientovaným grafům.

Incidence p přiřazuje každé hraně grafu neuspořádanou dvojici vrcholů. Platí-li pro incidenci hrany $e \in E, p(e) = (u, v)$ říkáme, že hrana e inciduje s vrcholem u i s vrcholem v . Vrcholy u, v nazýváme krajními vrcholy hrany e , nebo také sousedními vrcholy. Říkáme, že tyto vrcholy jsou sousední, nebo také, že spolu sousedí.

Protože každá hrana může být vyjádřena jako neuspořádaná dvojice sousedních vrcholů, je zápis $p(e) = (u, v)$ ekvivalentní se zápisem $p(e) = (v, u)$.

Speciálním případem vrcholu grafu je tzv. izolovaný vrchol, s nímž neinciduje žádná hrana.

Hranu, pro niž platí $p(e) = (u, u)$, nazveme smyčka.

Graf, pro nějž platí, že incidence p je prostým zobrazením (tedy $p(e) = (u, v) = p(e') \Rightarrow e = e'$), nazveme prostým grafem. Prostý graf, jenž neobsahuje smyčky, nazýváme obyčejný graf (obr. 2).

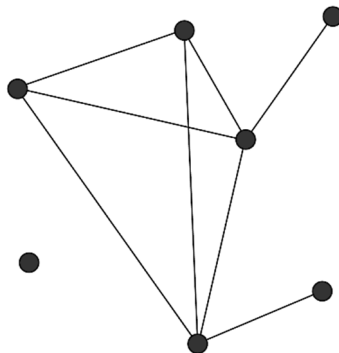
Označme počet vrcholů grafu $|V| = n$ a počet hran grafu $|E| = m$. Počet hran incidujících s vrcholem $v \in V$ nazýváme stupeň vrcholu a označujeme $dg(v)$. Pro každý graf platí následující vztah mezi počtem hran a stupni vrcholů:

$$\sum_{v_i \in V} dg(v_i) = 2m.$$

Každý graf má sudý počet vrcholů lichého stupně.

Grafické znázornění grafu je dáno jeho nakreslením. Nakreslení grafu $G = (V, E, p)$ je dvojice prostých zobrazení f, g takových, že f přiřazuje vrcholům $v \in V$ body euklidovské roviny, g přiřazuje hranám prosté křivky a to tak, že platí

$$p(e) = \{u, v\} \Rightarrow g(e) \text{ má krajní body } f(u) \text{ a } f(v).$$



Obr. 2 Obyčejný graf

Grafické znázornění grafů je užitečné zejména pro svou názornost. K jednomu grafu je možné sestavit mnoho různých nakreslení znázorňujících podle potřeby některé vlastnosti grafu. Na druhé straně grafy, které jsou z formálního hlediska odlišné, mohou být znázorněny stejným nakreslením. Základním prostředkem pro srovnání grafů je pak tzv. izomorfismus grafů.

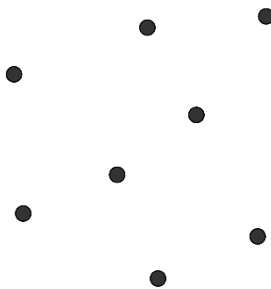
Nechť $G_1 = (V_1, E_1, p_1)$ a $G_2 = (V_2, E_2, p_2)$ jsou dva grafy a φ bijekce množiny $V_1 \cup E_1$ na množinu $V_2 \cup E_2$ taková, že zúžené zobrazení $\varphi|_{E_1}$ je bijekce E_1 na E_2 , zúžené zobrazení $\varphi|_{V_1}$ je bijekce V_1 na V_2 a φ zachovává incidenci, tedy

$$p_1(e) = p_1(u, v) \Rightarrow p_2(\varphi(e)) = (\varphi(u), \varphi(v)).$$

Grafy G_1 a G_2 , pro které lze zobrazení φ nalézt, nazýváme izomorfní grafy. Zapisujeme $G_1 \cong G_2$.

Triviálním grafem nazýváme graf, který se skládá z jednoho vrcholu.

Diskrétním grafem (obr. 3) nazýváme graf $G = (V, \emptyset)$, jehož množina hran je prázdná. Graf je tvořen množinou izolovaných vrcholů.

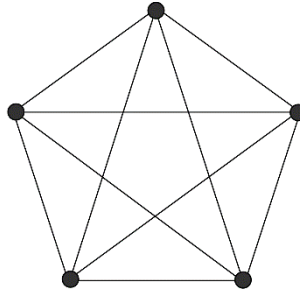


Obr. 3 Diskrétní graf

Kompletním (úplným) grafem (obr. 4) nazýváme graf, v němž je každý vrchol sousední se všemi ostatními vrcholy grafu. Tyto grafy označujeme K_n , kde n je počet vrcholů tohoto grafu. Pro kompletní graf $K_n = (V, E)$ platí

$$m = \frac{n \cdot (n - 1)}{2} \text{ a } dg(v_i) = n - 1, v_i \in V, i = 1, \dots, n,$$

kde m je počet hran daného grafu.



Obr. 4 Úplný graf

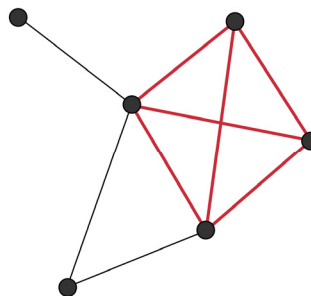
Podgrafem grafu $G = (V, E, p)$ rozumíme graf $\bar{G} = (\bar{V}, \bar{E}, \bar{p})$, pro který platí $\bar{V} \subseteq V, \bar{E} \subseteq E$ a pro každou hranu $e \in \bar{E}$ platí $\bar{p}(e) = p(e)$. Zkráceně zapisujeme $\bar{G} \subseteq G$.

Faktorovým podgrafem (faktorem) grafu $G = (V, E, p)$ rozumíme graf $\bar{G} = (\bar{V}, \bar{E}, \bar{p})$, pro který platí $\bar{V} = V, \bar{E} \subseteq E$ a pro každou hranu $e \in \bar{E}$ platí $\bar{p}(e) = p(e)$

Faktory, případně obecnější podgrafy daného grafu $G = (V, E, p)$, lze určit nějakou podmínkou. Potom řekneme, že jde o podgrafy indukované podmínkou, zkráceně též indukované podgrafy.

Klikou grafu $G = (V, E, p)$ nazveme indukovaný podgraf $\bar{G} = (\bar{V}, \bar{E}, \bar{p})$, který je izomorfní úplnému grafu. Kliku můžeme chápat také jako podmnožinu vrcholů $S \subseteq V(G)$ takovou, že každé dva vrcholy $u, v \in S$ jsou spojeny hranou (obr. 5).

Netriviální klikou pak nazveme takovou kliku, která má $n \geq 2$ vrcholů.



Obr. 5 Kliku grafu

Klikovost grafu G je počet uzlů v největší klice grafu G . Označujeme ji $\omega(G)$.

Jelikož počet možných netriviálních klik v grafu s n vrcholy je roven $\sum_{k=2}^n \binom{n}{k} = 2^n - n - 1$, tedy počtu všech dvou a více prvkových podmnožin, roste časová náročnost prohledání všech těchto podmnožin exponenciálně. Vyhledání všech klik v grafu (maximal clique problem), vyhledání největší kliky v grafu (maximum clique problem) či určení klikovosti grafu se tedy řadí do třídy NP-obtížných úloh. Projít všechny dvou a více prvkové podmnožiny je zvládnutelné v přijatelném čase jen pro grafy s nízkým počtem vrcholů, řádově do 30 vrcholů. V praxi je samozřejmě nutné řešit tyto úlohy pro značně rozsáhlejší grafy. V takovém případě se používají tzv. heuristické algoritmy. Heuristické algoritmy mají polynomiální časovou složitost. Tyto algoritmy však běžně nenaleznou optimální řešení. Vychází zpravidla z intuitivní myšlenky, jak danou úlohu řešit a mnohdy existuje více heuristických algoritmů pro jednu úlohu, které se liší svou časovou náročností i přesností výstupu.

Při vývoji nástroje *ScanGraph* byl původně použit heuristický algoritmus pracující na principu slučování klik. Před první iterací uvažujeme každý vrchol grafu jako kliku

s jedním vrcholem. Při každé iteraci slučujeme kliky do větších klik, až dokud nejdou vytvořit už žádná nová spojení. Dvě různé kliky (A a B) mohou být sloučeny právě tehdy, když každý vrchol kliky A je sousedem každého vrcholu kliky B . Tento algoritmus má pouze lineární náročnost, avšak může selhat v nalezení větší kliky, jestliže již dvě její menší části byly nevhodně sloučeny s jinými vrcholy.

Použití tohoto algoritmu se brzy ukázalo jako nevhodné, právě pro jeho nespolehlivost pro grafy s vyšší hranovou hustotou. Heuristický algoritmus byl proto nahrazen Bron-Kerbosch algoritmem.

Bron-Kerbosch algoritmus byl vytvořen dvojicí Coenraad Bron a Joep Kerbosch z Holandska (1973). Ačkoliv existují jiné algoritmy pro řešení *maximal clique problem*, jejichž časová složitost je nižší, je Bron-Kerbosch algoritmus v praxi považován za účinnější pro grafy s vyšší hustotou hran.

Algoritmus pracuje se třemi množinami R, P a X vrcholů grafu. Množina R je množina vrcholů, které tvoří maximální kliku, množina P je množina vrcholů, které mohou být dále přidány k aktuální klice a množina X je množina vrcholů, které nemohou tvořit kliku a množina $N(v)$ označuje všechny vrcholy sousedící s vrcholem v . Před první iterací jsou množiny R a X prázdné a množina P obsahuje všechny vrcholy grafu (obr. 6).

```

BronKerbosch( $R, P, X$ )
1 if  $P$  and  $X$  are both empty then
2   | report  $R$  as a maximal clique
3 for each vertex  $v$  in  $P$  do
4   | BronKerbosch( $R \cup v, P \cap N(v), X \cap N(v)$ )
5   |  $P \leftarrow P \setminus v$ 
6   |  $X \leftarrow X \cup v$ 

```

Obr. 6 Bron-Kerbosch algoritmus

Průběžný krok je popsán následovně (Obr. 6): Je vybrán vrchol v z množiny P . Vrchol v je přidán do množiny R a všechny vrcholy, se kterými v nesousedí, jsou vyjmuty z množin R a X . Tento krok se opakuje, dokud množina P není prázdná. Jakmile množina P je prázdná, pak pokud množina X je také prázdná, pak množina R obsahuje maximální kliku. Pokud množina X není prázdná, pak množina R obsahuje vrcholy, které jsou podmnožinou vrcholů již dříve nalezené kliky. Algoritmus se vrátí k poslednímu vybranému vrcholu a množiny P, R a X vrátí do stavu před posledním výběrem. Vyjme vrchol v z množiny P , přidá jej do množiny X a pokračuje výběrem dalšího vrcholu.

V mnoha aplikacích teorie grafů reprezentují hrany grafu reálné cesty. Je tedy nutné zavést následující pojmy:

Neorientovaný sled v G je posloupnost vrcholů a hran

$$v_1, e_1, v_2, e_2, \dots, v_{k-1}, e_{k-1}, v_k$$

taková, že pro každé $i = 1, 2, \dots, k - 1$ platí, že vrcholy v_i a v_{i+1} jsou krajní vrcholy hrany e_i . Sled se nazývá uzavřený, jestliže $v_1 = v_k$ a navíc $k > 1$. V opačném případě mluvíme o otevřeném sledu.

Sled nazýváme tahem, jestliže se v něm neopakují hrany.

Tah nazýváme cestou, jestliže se v něm neopakují vrcholy s možnou výjimkou, kdy $v_1 = v_k$, tedy kdy je cesta uzavřená. V takovémto případě nazýváme cestu kružnicí.

Je dán graf $G = (V, E, p)$. Řekneme, že vrchol v je dostupný z vrcholu w , jestliže existuje cesta z w do v .

Řekneme, že graf je souvislý, jestliže pro každé dva vrcholy $u, v \in v$ v grafu existuje cesta z u do v .

Graf G nazveme strom, je-li souvislý a neobsahuje kružnici. Vrchol v stromu, pro který platí $dg(v) = 1$ nazýváme list.

Maticové vyjádření neorientovaných grafů

Maticí sousednosti grafu $G = (V, E, p)$ s n vrcholy nazýváme čtvercovou maticí $A = (a_{ij})_{i,j=1}^n$ definovanou následovně:

- $a_{ij} = 0$, jestliže vrcholy v_i, v_j nejsou sousední,
- $a_{ij} = 1$, jestliže vrcholy v_i, v_j jsou sousední.

Vlastnosti matice sousednosti:

- $\sum_{j=1}^n a_{ij} = dg(v_i)$ pro $j = 1, \dots, n$,
- $\sum_{i=1}^n a_{ij} = dg(v_j)$ pro $i = 1, \dots, n$.

Dva grafy jsou izomorfní právě tehdy, jsou-li jejich matice sousednosti stejné až na permutaci řádků nebo sloupců.

Označme hrany grafu $e \in E$ jako e_1, e_2, \dots, e_m . Graf pak můžeme vyjádřit pomocí **matice incidence** $B = (b_{ij})_{i,j=1}^{n,m}$ definované následovně:

- $b_{ij} = 0$, jestliže vrchol v_i není incidentní s hranou e_j ,
- $b_{ij} = 1$, jestliže vrchol v_i je incidentní s hranou e_j .

Vlastnosti matice incidence:

- $\sum_{j=1}^m b_{ij} = dg(v_i)$ pro $j = 1, \dots, n$,
- $\sum_{i=1}^n b_{ij} = 2$ pro $i = 1, \dots, m$.

Uvažujme hranově ohodnocený graf $G = (V, E, p)$, ve kterém je každé hraně $e \in E$ přiřazena hodnota $o(e)$, která vyjadřuje její délku. **Maticí přímých vzdáleností** grafu G nazýváme čtvercovou maticí $D = (d_{ij})_{i,j=1}^n$, jejíž prvky jsou definovány následovně:

- $d_{ij} = o(h)$, jestliže $\exists e \in E$, pro kterou $p(e) = (v_i, v_j)$,
- $d_{ij} = \infty$, jestliže $\nexists e \in E$, pro kterou $p(e) = (v_i, v_j)$,
- $d_{ij} = 0$, pro $i = j$.

3.2 Využití teorie grafů v geoinformatice

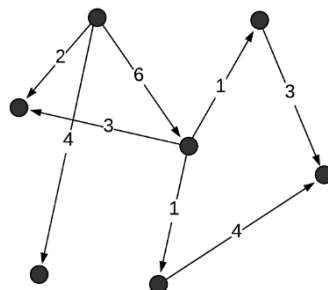
Teorie grafů nachází své aplikace v mnoha vědních oborech. Geoinformatika tak není žádnou výjimkou. Odkazovat se lze již na prvopočátky teorie grafů, kdy Leonhard Euler (1736) řešil úlohu kaliningradských mostů, tedy jak projít sedm mostů přes řeku Pregolu, a to každý z nich právě jednou, a vrátit se do výchozího místa. Tomuto problému odpovídá v teorii grafů pojem eulerovský graf.

Pojem grafu je blízký představě cest mezi sídly a to jak v doslovném tak přeneseném významu. Velký počet úloh lze tak převést na grafové úlohy. Při modelování těchto úloh se nabízí následující možnosti:

- (1) hledání posloupnosti operací
 - a. Vrcholy grafu představují stavy a hrany představují změny (čas, vzdálenost, finance apod.),
 - b. Vrcholy představují operace a hrany představují možnost bezprostřední návaznosti akcí,
- (2) časové výpočty paralelně probíhajících akcí,
- (3) hledání statických operací s využitím cest v grafu.

Příklady hledání posloupnosti operací je možné najít v dopravních problémech, například u nalezení optimální cesty mezi jednotlivými městy.

Časové výpočty paralelně probíhajících akcí se realizují pomocí síťových grafů. Vrcholy grafu pak představují jednotlivé činnosti a hrany, jež jsou ohodnoceny délkou jejich trvání, reprezentují možné přechody mezi danými činnostmi (obr. 7).



Obr. 7 Síťový graf

Hledání statických konfigurací je využíváno v případech umístování důležité budovy (např. nemocnice či požární zbrojnice) do určitého území tak, aby silniční síť umožňovala dostupnost i z nejbližšího místa v co nejkratším čase. Vrcholy takového grafu představují jednotlivé křižovatky a hrany jsou ohodnoceny dobou jízdy.

Při výše popsaných aplikacích jsou často využívány různé grafové algoritmy. Mezi nejpoužívanější se řadí

- *Dijkstrův algoritmus* pro hledání nejkratší cesty v grafu. Algoritmus vypočítá nejkratší cesty z jednoho vrcholu do všech ostatních vrcholů. Lze jej použít pro orientovaný i neorientovaný graf. Jistou nevýhodou může být požadavek na nezáporné ohodnocení všech hran. Dijkstrův algoritmus využil při navrhování silniční sítě pomocí bipartitních sítí Wang a kol. (2012). Pro plánování cest na staveništích hodnotil výkon Dijkstrova algoritmu a genetických algoritmů Soltani a kol. (2002). Precizaci Dijkstrova algoritmu pro plánování tras přinesl Fan & Shi (2010).
- *Bellman-Fordův algoritmus* slouží také k hledání nejkratší cesty v grafu. Na rozdíl od Dijkstrova algoritmu smí graf obsahovat i záporně ohodnocené hrany a jeho časová složitost je vyšší než u Dijkstrova algoritmu ($O(mn)$, kde m je počet hran grafu a n je počet vrcholů grafu). Porovnání Bellman-Fordova algoritmu s jinými vyhledávacími algoritmy při využití pro bezpilotní letouny provedl Sathyaraj a kol. (2008). Srovnáním Bellman-Fordova a multi-modálního Dijkstrova algoritmu v multimodálních dopravních sítích se zabývali Liu a Meng (2009).
- *Floyd-Warshallův algoritmus* vypočítá nejkratší vzdálenosti mezi všemi dvojicemi vrcholů v grafu s paměťovou složitostí $O(n^2)$. Graf může obsahovat i záporně ohodnocené hrany. Výhodou Floyd-Warshallova algoritmu je jeho snadná aplikace v distribuovaném prostředí, mezi autonomními jednotkami, které si mohou informace předávat jen pomocí zaslání zpráv po síti. V dopravních úlohách se speciálními zatáčkami využil Floyd-Warshall algoritmus a Dijkstrův algoritmus Kim a kol. (2003). Stejně algoritmy použili pro řešení rozsáhlých a náročných dopravních problémů v reálném čase Pradhan a Mahinthakumar (2013).
- *Kruskalův algoritmus* pro hledání minimální kostry v grafu, tedy takové kostry, jejíž součet ohodnocení je minimální. Algoritmus pracuje na principu spojování

hran s nejmenším ohodnocením, dokud tyto hrany nespojí vrcholy celého grafu. Časová složitost algoritmu je $O(m \cdot \log(m))$ a v případě předřazených hran podle ohodnocení $O(m \cdot \alpha(m))$, kde α je inverzní Ackermannova funkce. Praktické použití algoritmu je možné najít například při budování elektrických rozvodů ve městě. Snahou je propojit jednotlivá místa co nejmenší délkou elektrického vedení. V problematice plánování a provozu nákladní dopravy využili Kruskalův algoritmus Crainic a Laporte (1997). Při studiu městských uličních sítí využil Kruskalův algoritmus Cardillo a kol. (2006) či Buhl a kol. (2006).

- *Ford-Fulkersonův algoritmus* pro výpočet maximálního toku v síti. Algoritmus je typicky používán pro úlohy typu vodovodního potrubí. Jako „potrubí“ však může vystupovat například silniční či datová síť. Pro plánování rozvodné sítě napětí použil Dijkstrův algoritmus a Ford-Fulkersonův algoritmus Nara (1997) nebo Khodr a kol. (2003), který využil také rozdělení pravděpodobnosti.

Uplatnění teorie grafů je možné také v kartografii. Snad nejznámější úlohou je problém obarvení mapy tak, aby žádná dvě sousední území nebyla obarvena stejnou barvou a současně bylo využito minimálního potřebného množství barev. Tato úloha je aplikací barvení vrcholů rovinného grafu a určení chromatického čísla grafu. Všeobecně známým se stal tzv. *problém čtyř barev*, tedy zda na obarvení libovolné politické mapy stačí čtyři barvy tak, aby žádné dva sousední státy nebyly vykresleny stejnou barvou. Tento problém byl vyřešen až v roce 1976, kdy Appel a Haken (1976) využili pro důkaz počítačový model všech možných konfigurací. Při tomto modelu jsou politické jednotky reprezentovány vrcholem a hrana mezi vrcholy vyjadřuje sdílení společné hranice.

Při generalizaci mapy je možno liniové prvky na mapě reprezentovat orientovanými ohodnocenými hranami grafu. Počátky, rozdělení či křížení liniových prvků jsou pak reprezentovány vrcholy tohoto grafu. Generalizační výběr je pak simulován úlohami o cestách v grafu. Typické zpracování lze najít u autorů Saalfeld (1995) či Mackaness a Beard (1993).

3.2.1 Využití teorie grafů v analýze eye-tracking dat

Teorii grafů lze pro analýzu eye-tracking dat využít zejména při hodnocení podobnosti v průběhu trajektorií pohybu očí (Scanpath Comparison). Tento typ analýzy je velmi vhodný zejména v případech, kdy je požadavkem z eye-tracking dat zjistit odlišnost chování (čtení stimulu) různými skupinami uživatelů. Nejčastěji se jedná o experty a laiky, muže a ženy atd.

Historie studia trajektorie pohybu očí se dá vysledovat v práci Noton a Stark (1971), kteří vyhodnocovali podobnost v prohlížení kreseb respondenty při různých příležitostech.

Jednou z nejběžněji používaných metod *Scanpath Comparison* je takzvaná *String Edit Distance* (Anderson a kol., 2015).

String Edit Distance představená Priviterou a Starkem (2000) porovnává průběh trajektorie pohybu očí na základě zjednodušení dat z linie v prostoru na sekvenci písmen označující navštívené oblasti zájmu (AOI). Podle Duchowski a kol. (2010) je *String Edit Distance* jednou z prvních metod *Scanpath Comparison*, která kvantitativně vyjadřuje nejen pozice fixací, ale i jejich pořadí. Scanpath – záznam pohybu očí – se skládá střídavě ze sakád (přechodů) a fixací v průběhu prohlížení stimulu respondentem.

Metody *Scanpath Comparison* mohou být rozděleny do šesti skupin (*String Edit Distance*, *ScanMatch*, *Samplebased Measures*, *Linear Distance*, *MultiMatch* a *Crossrecurrence Quantification Analysis*). Přehled těchto metod a jejich srovnání jsou dostupné například v práci Anderson a kol. (2015).

Pro použití metody *String Edit Distance* je třeba vyznačit ve sledovaném stimulu tzv. oblasti zájmu (Areas of Interest, AOI). Trajektorii pohledu je pak přiřazen řetězec prvků

reprezentující posloupnost fixací v navštívených oblastech zájmu. Pouze 10 % sakády je tvořeno pohybem okohybných svalů. Fixace oka ve vytvořených oblastech zájmu pak tvoří 90 % z celé sakády (Bahill a Stark, 1979).

Foulsham a kol. (2012) poukázal na nevýhodu *String Edit Distance*, jež spočívá v převedení vzdáleností na sledovaném stimulu v binární klasifikaci (z důvodu nutnosti vytvoření oblastí zájmu). Pro některé aplikace, mezi něž patří i kartografie, tato nevýhoda může být naopak přeměněna ve výhodu. Například při sledování práce respondentů s jednotlivými kompozičními prvky mapy.

Pro vyčíslení vzdálenosti mezi jednotlivými řetězci se běžně používá několik metrik jako například *Levenshteinova vzdálenost* (Levenshtein, 1966), *Damerau-Levenshteinova vzdálenost* (Damerau, 1964) či Needleman-Wunsch algoritmus (Needleman a Wunsch, 1970).

Levenshteinova vzdálenost

Levenshteinovu vzdálenost zavedl v roce 1965 ruský matematik Vladimir Iosifovič Levenshtein pro měření editační vzdálenosti mezi textovými řetězci. Připouští tři editační operace – vložení znaku (insertion), vymazání znaku (deletion) a záměnu jednoho znaku za jiný (substitution). Nejmenší počet operací nutných k převedení jednoho znaku ve druhý pak určuje hodnotu Levenshteinovy vzdálenosti ($Lev(a, b)$).

Hodnota Levenshteinovy vzdálenosti je tedy rostoucí s rostoucí odlišností daných řetězců.

Uvažujme dva textové řetězce (sekvence) $a = (1, \dots, |a|)$ a $b = (1, \dots, |b|)$ s délkami $|a|$ a $|b|$. V průběhu algoritmu je postupně vytvářena matice $M_{|a|+1, |b|+1}$ s řádky $0, \dots, |a|$ a $0, \dots, |b|$. V matici je na pozici $M_{i,j}$ hodnota, která udává Levenshteinovu vzdálenosti podřetězců $a(1, \dots, i)$ a $b(1, \dots, j)$. Hledaná vzdálenost řetězců a a b se tedy nachází v matici na pozici $M_{|a|, |b|}$. Matice M se inicializuje naplněním prvního sloupce a prvního řádku znaky sekvencí a a b . V kroku algoritmu každé i, j vypočítáme $M_{i,j}$ následovně:

if $a(i) = b(j)$ then cost := 0

else cost := 1

$$M_{i,j} = \min \begin{cases} M_{i-1,j} + 1 \\ M_{i,j-1} + 1 \\ M_{i-1,j-1} + \text{cost} \end{cases}$$

Hodnota $Lev(a, b) = 0$ indikuje totožné řetězce, jelikož k převedení jednoho řetězce ve druhý není potřeba žádná operace. Jestliže řetězce nemají žádný společný znak, platí že $Lev(a, b) = \max\{|a|, |b|\}$.

Nevýhodou této metody je vyšší časová složitost. Nejlépe se dá implementovat v čase $O(|a| \times |b|)$.

Damerau-Levenshteinova vzdálenost

Damerau-Levenshteinova vzdálenost pojmenovaná po Fredericku J. Damerauovi a Vladimíru I. Levenshteinovi je mírně upravená myšlenka Levenshteinovy vzdálenosti. Kromě operací vložení, vymazání a substituce přidává ještě operaci transpozice dvou znaků. Transpozicí se rozumí výměna, přesmyčka dvou sousedních znaků (GRAF → GARF). Motivací pro vytvoření Damerau-Levenshteinovy vzdálenosti byla právě detekce překlepů v elektronicky psaném textu, kde překlep ve formě přesmyčky je velmi častou chybou.

Při výpočtu Damerau-Levenshteinovy vzdálenosti ($DL(a, b)$) mezi dvěma sekvencemi a a b se v prvním kroku inicializuje matice o rozměrech $(|a| + 3) \times (|b| + 3)$ a buňky prvního sloupce a prvního řádku se vyplní hodnotou ∞ .

V dalším kroku se hodnota buněk $M_{i,j}$ určí následovně:

$$M_{i,j} = \min \begin{cases} M_{i-1,j} + 1 \\ M_{i,j-1} + 1 \\ M_{i-1,j-1} + 1 \\ M_{i-2,j-2} \end{cases}$$

Stejně jako při výpočtu Levenshteinovy vzdálenosti je možné výslednou hodnotu odečíst z buňky v posledním řádku a posledním sloupci matice, tedy na pozici $M_{|a|+3,|b|+3}$.

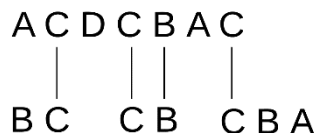
Hodnota Damerau-Levenshteinovy vzdálenosti je stejně jako hodnota Levenshteinovy vzdálenosti rostoucí s rostoucí odlišností daných sekvencí a platí $DL(a,b) = 0$ právě tehdy, když $a = b$ a v případě, že sekvence nemají žádný společný znak, pak $DL(a,b) = \max\{|a|, |b|\}$.

Všechny tři výše zmíněné metody výpočtu vzdálenosti nejsou vhodné přímo pro práci s eye-trackingovými daty. Needleman-Wunsh algoritmus jakožto původně vytvořený pro analýzu genových sekvencí je vhodnější při výpočtech vzdálenosti mezi sekvencemi se stejným počtem znaků, což řetězce navštívených AOI zpravidla nebývají. Levenshtein a Damerau-Levenshtein vzdálenosti byly vytvořeny pro lingvistické potřeby a tedy jejich použití očekává výpočet mezi dvěma sekvencemi, které se od sebe příliš neliší. Jeho odlišnost je dána možným počtem překlepů a jejich délka je buď stejná, nebo minimálně různá. Tento charakter však řetězce eye-trackingových dat zpravidla nemají, byť jistá podobnost mezi nimi v rámci jednoho experimentu je očekávána, nikoliv však zaručena.

Needleman-Wunsch algoritmus

Needleman-Wunsch algoritmus vznikl pro použití v oblasti bioinformatiky, zejména pro vyhledávání podobností mezi sekvencemi nukleotidů v nukleových kyselinách či sekvence aminokyselin v proteinech.

Needleman-Wunsch algoritmus (obr. 8) umožňuje globální zarovnání dvou řetězců a následně nalezení nejdelšího společného podřetězce, přičemž ve svém výsledném zarovnání umožňuje přítomnost mezer. Každá mezera je ohodnocena nezápornou hodnotou gap penalty (GP) a podobnost jednotlivých znaků x a y je ohodnocena funkcí $score(x,y)$. Hodnoty gap penalty i $score(x,y)$ mohou být dány jako libovolná reálná čísla, ale zpravidla jsou voleny celočíselně. Výstupem algoritmu je pak tedy kladná celočíselná hodnota, která udává počet souhlasných znaků.



Obr. 8 Needleman-Wunsch algoritmus

Hodnota Needleman-Wunsh algoritmu ($NW(a,b)$) pro dva řetězce a a b je rostoucí s rostoucí podobností řetězců, tedy s rostoucím počtem shodných znaků v těchto dvou řetězích.

Uvažujme dva vstupní řetězce a a b . Prvním krokem algoritmu je vytvoření matice $M_{k,l}$, kde k a l jsou délky řetězců a a b . Matice $M_{k,l}$ se inicializuje následujícím způsobem:

$$\begin{aligned} M_{i,0} &= -i \cdot GP; \quad 0 \leq i < k \\ M_{0,j} &= -j \cdot GP; \quad 0 \leq j < l \end{aligned}$$

Inicializaci tedy lze chápat jako určení penalty za mezeru, která pokrývá celou jednu sekvenci, tedy že žádné prvky sekvencí se neshodují.

Dalším krokem je naplnění prázdných polí matice směrem z levého horního rohu matice. Určení hodnoty každé buňky vychází z hodnot v sousední levé, horní a levohorní diagonální buňky. K hodnotě levohorní diagonální buňky připočteme hodnotu score $(x(i), y(j))$, v případě levé a horní pozice odpočítáme hodnotu GP . Tímto způsobem můžeme dostat tři hodnoty, ze kterých maximální hodnota naplní aktuálně počítanou buňku matice. Jestli maximální hodnota pochází z horní buňky, znamená to, že v sekvenci x jsme se neposunuli, ale v sekvenci y ano, tedy že se vytvořila mezera. Analogicky se můžeme vyjádřit i pro hodnotu z levé buňky. V případě, že maximální hodnota pochází z diagonální buňky, znamená to, že dva prvky sekvence se úspěšně shodují.

$$M_{i,j} = \max \begin{cases} M_{i-1,j} - GP \\ M_{i,j-1} - GP \\ M_{i-1,j-1} + \text{score}(x(i), y(j)) \end{cases}$$

Posledním krokem algoritmu je zpětně zrekonstruování cesty a nalezení odpovídajících si prvků. Začneme buňkou v pravém dolním rohu matice M a zpětným postupem zjistíme finální zarovnání daných sekvencí. Jestliže zpětný postup umožní více než jedno řešení, pak existuje více zarovnání se stejným ohodnocením. Výstupem algoritmu je zarovnání dvou sekvencí hodnota $NW(a, b)$ mající vztah k podobnosti těchto sekvencí.

Hodnota $NW(a, b)$ je rostoucí s rostoucí podobností sekvencí. V algoritmu používaném nástrojem *ScanGraph* hodnota $NW(a, b) = 0$ indikuje sekvence, které nemají žádný společný znak. V případě, že pro délku sekvencí $|a|$ a $|b|$ platí, že $|a| = |b| = NW(a, b)$, pak jsou sekvence zcela shodné.

Existující nástroje pro analýzu eye-tracking dat

Mezi běžně používané nástroje pro vyhodnocování *String Edit Distance* patří *ClustalG*, *TramineR* a *eyePatterns*. Nástroj *ClustalG* slouží primárně pro účely bioinformatiky a byl vytvořen pro analýzu genových sekvencí (DNA, RNA, proteiny). Jako metodu *String Edit Distance* používá *Multiple Sequence Alignment*. Adekvátní použití této metody očekává sekvence stejné nebo nepříliš různé délky s opakujícími se podsekvencemi. Pro bioinformatické účely je tedy zcela vyhovující, avšak eye-tracking data tento předpoklad nesplňují a tedy *ClustalG* při jejich analýze nebude mít vysokou vypovídající hodnotu.

Nástroj *TramineR* vznikl jako balíček rozšíření prostředí R pro vyhodnocování a vizualizaci sekvencí biografických údajů. Při analýze pomocí nástroje *TramineR* je možné použít několik metod pro hledání podobnosti mezi sekvencemi – *Number of matching positions*, *Longest Common Prefix distances*, *Longest Common Subsequence distances*, *Optimal Matching distances*. Poslední dvě zmíněné metody jsou svým charakterem srovnatelné s metodou Needleman-Wunsch algoritmus a Levenshteinova vzdálenost v tomto pořadí. Použité metriky tedy mohou vyhovovat charakteru eye-tracking dat. Navíc oproti nástroji *eyePatterns* jsou metriky vytvořeny tak, aby zohledňovaly možnou odlišnou délku

porovnávaných sekvencí. Avšak nástroj *TramineR* neumožňuje import eye-tracking dat přímo z prostředí OGAMA a výsledné skupiny vzájemně si podobných sekvencí určuje pomocí *hierarchical clustering algoritmu*. Při této metodě je každá sekvence zařazena do skupiny, která se při jednotlivých iteracích algoritmu může pouze rozšiřovat (skupiny se v jednotlivých iteracích slučují), ale neumožňuje, aby jedna sekvence patřila i do více skupin najednou.

Oproti předchozím nástrojům byl nástroj *eyePatterns* navržen právě pro analýzu eye-tracking dat. Pracuje s algoritmy *String Edit Distance*, *Levenshteinovou vzdáleností* a *Needleman-Wunsch algoritmem*. Přesto se jeho použití ukázalo být nedostatečné pro plnohodnotnou analýzu *Scanpath Comparison*. Na bližší popis nedostatků tohoto nástroje poukazuje kapitola 4.5.

3.2.2 Využití teorie grafů v prostorovém plánování

Teorie grafů v prostorovém plánování příliš uplatnění nenachází. Avšak zmínky lze nalézt například v práci Kong a kol. (2010), kdy byly využity grafy pro plánování zeleně v městských částech. Na propojení přírodní krajiny a měst se zaměřuje Zetterberg a kol. (2010a), rozptýlenost mořských druhů za pomoci teorie grafů řešil Trembl a kol. (2008). Přesto analýzám prostorového plánování dominují metody shlukové analýzy, která grafových algoritmů nevyužívá. V této kapitole disertační práce budou tedy krátce představeny základní pojmy a základní metody shlukové analýzy.

Shluková analýza neboli *klastrování* je souhrnný název, který zahrnuje široké spektrum výpočetních postupů. Jejich společným cílem je hledat vzájemné vztahy uvnitř množiny objektů či jevů. Snahou je pak vytvoření systému, jenž organizuje objekty do silně interně homogenních skupin (objekty ve skupině jsou si co nejvíce podobné) a současně silně externě heterogenních (skupiny mezi sebou si jsou co nejméně podobné).

Úlohu shlukové analýzy formulujeme následovně:

Nechť χ značí množinu n objektů. Rozklad $\Omega = \{C_1, C_2, \dots, C_m\}$ množiny χ je množina disjunktních, neprázdných podmnožin. Pro $i \neq j$ tedy platí

$$C_i \cap C_j = \emptyset, \quad \bigcup_{i=1}^m C_i = \chi.$$

Množiny C_i nazýváme komponenty rozkladu.

Shlukování je pak úloha následujícího tvaru. Její komponenty nazýváme shluky.

Nechť x je náhodný objekt z množiny objektů χ s rozdělením pravděpodobnosti $p_x: \chi \rightarrow \mathbb{R}$. Nechť D je množina rozhodnutí taková, že pro každý objekt $x \in \chi$ se určí jisté rozhodnutí $d \in D$. Nechť $W: \chi \times D \rightarrow \mathbb{R}$ je pokutová funkce, jejíž hodnota $W(x, d)$ představuje ztráty v případě volby rozhodnutí d pro objekt x .

Cílem úlohy je zkonstruovat strategii $Q: \chi \rightarrow D$, která minimalizuje hodnotu

$$\sum_{x \in \chi} p(x) \cdot W(x, Q(x)).$$

Shlukovací metody se mohou dělit dle jejich cíle na hierarchické a nehierarchické.

Hierarchické metody využívají dříve nalezených shluků a vytváří z nich shluky nové, a to buď aglomerativně či divizně. Aglomerativní metody chápou každý objekt množiny χ jako shluk. Tyto shluky dále slučují až vznikne shluk obsahující všechny objekty dané

množiny. Divizní metody chápou vstupní množinu χ jako jeden shluk, jenž dále dělí na shluky nové.

Oproti tomu nehierarchické metody nevytvářejí žádnou hierarchickou strukturu. Rozkládají množinu χ do podmnožin dle předem daného kritéria. První rozklad na komponenty se již dále nedělí ani neslučuje.

Shlukování je často intuitivně chápáno jako hledání skupin vzájemně si podobných objektů a kvantitativní vyjádření této podobnosti, tedy nalezení míry podobnosti a metriky δ splňující podmínky identity, symetrie a trojúhelníkové nerovnosti.

Nejběžnějším způsobem vyjadřování podobností mezi objekty jsou metriky vycházející z geometrického modelu dat. Jednotlivé objekty, jež jsou charakterizovány uspořádanou k -ticí, můžeme chápat body metrického prostoru X_k k -té dimenze.

Při prostorovém plánování se nejčastěji lze setkat s variantami Minkowského metriky definované vztahem:

$$\delta_m(x, y) = \left(\sum_{i=1}^k |x_i - y_i|^m \right)^{\frac{1}{m}}.$$

Speciálně pro $m = 1$ dostáváme Manhattanskou vzdálenost, pro $m = 2$ euklidovskou vzdálenost a pro $m \rightarrow \infty$ Čebyševovu vzdálenost, tedy

$$\delta_\infty(x, y) = \max_{i=1, \dots, k} \{|x_i - y_i|\}.$$

Manhattanská vzdálenost bývá označována jako vzdálenost v městských blocích, tedy počet přímých vzájemně na sebe kolmých ulic, které je nutno projít z bodu x do bodu y . Oproti tomu Euklidovská vzdálenost je intuitivně chápána jako přímá „vzdušná“ vzdálenost mezi body x a y . Použití Čebyševovy vzdálenosti nabývá na významu, když je vzdálenost posuzována více na základě vzdálenosti jednotlivých souřadnic, tedy v aplikacích pro výpočet podobnosti na základě individuální odlišnosti v jednotlivých atributech.

Hierarchické metody shlukování

Shlukování hierarchickou metodou lze znázornit pomocí dendrogramu neboli binárního stromu. Jednotlivé uzly tohoto grafu pak představují samotné shluky. Řezy tohoto binárního stromu reprezentují stupně rozkladů (divizní metody) či slučování (aglomerativní metody) shluků. Nevýhodou hierarchického shlukování je, že jeho snahou je v každé iteraci dosáhnout pouze lokálně nejlepšího řešení bez ohledu na další postup. Při aglomerativním hierarchickém shlukování mohou vzniknout nejednoznačnosti již na začátku shlukování, které se projeví až později ve větších shlucích.

Mezi nejběžnější metody aglomerativního hierarchického shlukování lze zařadit *metodu nejbližšího či nejvzdálenějšího souseda*, *centroidní metodu* či *metodu průměrné vazby*. Divizní hierarchické shlukování zastupuje například *MacNaughton-Smithova metoda*.

Metoda nejbližšího souseda vytváří v každé iteraci shluky z objektů či z již dříve vytvořených shluků, které mají mezi sebou nejmenší vzdálenost (dle zvolené metriky) v porovnání s ostatními shluky. Nevýhodou této metody je, že jestliže existují shluky se stejnou vzdáleností od již existujících shluků, může dojít ke zřetězení. Pokud shluky na okrajích řetězce nevykazují podobnost, může dojít ke špatnému závěru. Dále může dojít k předčasnému sloučení dvou dobře rozlišitelných shluků, jejichž hranice se v jednom místě přibližují.

Metoda nejvzdálenějšího souseda vzdálenost dvou shluků určuje jako největší možnou vzdálenost dvou objektů ze vzdáleností každých dvou objektů z dvou různých shluků. Z takto vypočítaných vzdáleností pak vybere nejkratší.

Centroidní metoda využívá euklidovskou vzdálenost, kterou určí vzdálenosti těžišť shluků a slučuje ty shluky, které mají nejmenší vzdálenost mezi těžišti.

Při *metodě průměrné vazby* se vzdálenost shluků počítá jako průměr vzdáleností mezi každými dvěma objekty patřícími do dvou různých shluků. Slučovány jsou pak shluky s nejmenší průměrnou vzdáleností. Tato metoda bývá méně citlivá na statistické odchylky v datech.

Nevýhodou *divizního hierarchického shlukování* je vysoká výpočetní náročnost. Nalezení optimálního rozkladu množiny vyžaduje prozkoumání $2^{n-1} - 1$ možností, tedy časová složitost s rostoucím počtem objektů n roste exponenciálně. Tyto metody jsou tedy použitelné pouze pro malé počty objektů.

MacNaughton-Smithova metoda divizního hierarchického shlukování snižuje výpočetní náročnost až na kvadratickou, ovšem nemusí nutně nalézt optimální řešení. Jako heuristiku pro dělení používá tato metoda střední vzdálenost objektu od shluku, ve kterém se objekt nachází. Po nalezení objektu, který má tuto vzdálenost největší, je objekt ze shluku vyřazen a je položen jako základ nového shluku. O všech objektech starého shluku se rozhoduje, zda se přefadí do nového shluku či nikoliv.

Nehierarchické metody shlukování

Nehierarchické metody shlukování rozkládají množinu χ do podmnožin podle předem daného kritéria. Jelikož výpočetní náročnost vyzkoušení všech možných uspořádání roste exponenciálně, začíná většina metod podle předem daného rozkladu, který dále upravuje. Nevýhodou tak je, že metody často končí pouze s lokálně optimalizovaným rozkladem.

Jestliže jednotlivým objektům, které byly rozděleny na shluky, můžeme přiřadit jejich geografickou polohu, vytváří pak spolu se svým rozdělením regionalizaci či typizaci území. Pojem *region* lze vymezit jako ohraničené teritoriální jednotky charakterizované ve vztahu k sousedním územím individuálními specifiky. Pro označení území jako regionu musí být splněna podmínka pokrytí a disjunkce. Jinak řečeno, regiony musí pokrývat celé sledované území a nesmějí se navzájem překrývat. Metodou shlukování bychom tedy k regionalizaci území dospěli v případě, kdy určující vlastností pro tvorbu shluků bude jejich geografická poloha. Jestliže jsou však shluky tvořeny na základě jiných atributů, s geografickou polohou nesouvisejících, může jako výstup shlukové analýzy dojít k typizaci území. Tedy k setřídění vybraných jevů ve sledovaném území podle podobných znaků.

Běžně používanými nástroji pro shlukování jsou například *Grouping Analysis* a *Multivariate Clustering*, jež jsou dostupné v programu ArcGIS Pro. Jejich stručný popis následuje v kapitole 7.2.

Dalším nástrojem programu ArcGIS Pro, jenž je používán v prostorovém plánování, a grafových algoritmů využívá, je *Urban Network Analyst*. Nástroj *Urban Network Analyst* jako toolbox pro software ArcGIS for Desktop 10 byl vytvořen na základě projektu, který byl řešen výzkumnou skupinou z Massachusetts Institute of Technology (MIT) pro analýzu urbánních sítí. Hlavními řešiteli byli Andres Sevtsuk (Sevtsuk a Mekonnen, 2012) ze Singapore University of Technology and Design (SUTD), a Michael Mekonnen z MIT. Nástroj je využíván zejména pro účely územního plánování, kde je největší zájem o prostorové rozložení měst a vztahy uvnitř sociálních, ekonomických a environmentálních procesů. Je možné jej využít pro analýzy jak v lokálním, tak v regionálním měřítku.

Urban Network Analyst je volně dostupný ke stažení ve verzi 1.01 na webových stránkách City Form Lab. První verze tohoto nástroje byla uveřejněna v roce 2010.

V rámci bakalářské práce na Katedře geoinformatiky UP nástroj testoval na případových studiích a vytvořil manuál nástroje v českém jazyce Vrkoč (2014).

Nástroj *Urban Network Analyst* umožňuje výpočet pěti základních analýz v prostorových sítích, a to dosažitelnost (Reach), přitažlivost (Gravity), spojitost (Betweenness), blízkost (Closeness) a přímota (Straightness). Při všech analýzách jsou budovy chápány jako prostorové jednotky, tedy jako vrcholy v grafu G . Budovám je možno navíc přiřazovat různé váhy podle jejich specifických vlastností, jako je počet bytů či obyvatel. Tuto volbu umožňuje nástroj UNA jako jediný.

Dosažitelnost

Při výpočtu dosažitelnosti je řešena otázka, kolik okolních budov spadá do stanoveného okruhu od dané budovy v prohledávané prostorové síti. Okruh dosažitelnosti budov určí počet dalších vrcholů (budov) v grafu, které jsou dosažitelné z vrcholu i na základě nejkratší možné vzdálenosti z maximálně stanoveného poloměru r .

Označme G síť s budovami. Pak pro hodnotu dosažitelnosti $Reach[i]^r$ pro daný poloměr r a budovu i platí

$$Reach[i]^r = \sum_{j \in G - \{i\}, d[i,j] \leq r} W[j],$$

kde $W[j]$ je váha budovy j a $d[i,j]$ je nejkratší možná vzdálenost mezi budovami i a j .

Přitažlivost

Míra přitažlivosti je oproti dosažitelnosti ovlivněna přídavnými faktory prostorové impedance potřebnými k cestě do každého cíle. Míra přitažlivosti předpokládá, že dostupnost budovy je přímo úměrná váze budov, které ji obklopují, a nepřímo úměrná vzdálenostem k nim.

Pro hodnotu přitažlivosti sítě G , daný poloměr r a budovu i platí

$$Gravity[i]^r = \sum_{j \in G - \{i\}, d[i,j] \leq r} \frac{W[j]}{e^{\beta \cdot d[i,j]}},$$

kde β je koeficient pro úpravu vzdálenostní funkce (distance decay). Vliv vzdálenosti klesá exponenciálně. Tento vliv lze blíže specifikovat koeficientem β , který je zvolen na základě předpokládaného způsobu transportu mezi budovami. Pro pěší chůzi měřenou v minutách je volena hodnota $\beta = 0,1813$ (Handy a Niemeier, 1997). Jestliže β není zadána, je přednastavena hodnota 0.

Spojitost

Spojitost budovy je určena poměrem počtu nejkratších cest mezi dvojicemi budov v síti v daném poloměru r , které prochází danou budovou, a počtem nejkratších cest v celé síti. Spojitost budovy se často používá pro odhad potenciálu chodců mezi jednotlivými budovami.

Pro hodnotu přitažlivosti sítě G , daný poloměr r a budovu i platí

$$Betweenness[i]^r = \sum_{j,k \in G - \{i\}, d[j,k] \leq r} \frac{n_{jk}[i]}{n_{jk}} \cdot W[i],$$

kde $n_{jk}[i]$ je počet nejkratších cest mezi budovami j a k , které prochází přes budovu i a n_{jk} je počet všech nejkratších cest mezi budovami j a k .

Blížkost

Blížkost budovy je definována jako inverzní kumulativní vzdálenost potřebná k dosažení všech ostatních budov v síti z dané budovy, které spadají do daného poloměru r podél nejkratších cest. Míra blízkosti udává, jak blízká je budova vůči všem ostatním budovám v rámci daného poloměru. Pokud není poloměr zadán, určí se blízkost vůči všem budovám v síti.

Pro hodnotu přitažlivosti sítě G , daný poloměr r a budovu i platí

$$Closeness[i]^r = \frac{1}{\sum_{j \in G - \{i\}, d[i,j] \leq r} (d[i,j] \cdot W[j])}$$

kde $W[j]$ jsou váhy cílových budov.

Přímost

Index přímosti popisuje, jak nejkratší možná cesta z dané budovy do jiné budovy kopíruje euklidovskou vzdálenost mezi těmito budovami. Přímost zachycuje pozitivní odchylky v cestovní vzdálenosti, které vyplývají z geometrických omezení uliční sítě.

Pro hodnotu přitažlivosti sítě G , daný poloměr r a budovu i platí

$$Straightness[i]^r = \sum_{j \in G - \{i\}, d(i,j) \leq r} \frac{\delta[i,j]}{d[i,j]} \cdot W[j],$$

kde $\delta[i,j]$ je euklidovská vzdálenost mezi budovami i a j .

3.2.3 Využití grafů v dopravě

Využití teorie grafů v dopravních úlohách je velmi široké. Většina aplikací teorie grafů je zařaditelná do oblasti síťových analýz. Základní přehled o síťových analýzách lze nalézt v publikacích Longley (2005) či Tuček (1998). V oblasti geoinformatiky jsou využívány nejrůznější typy síťových analýz:

- Optimální trasy (Delavar a kol., 2004; Chang a kol., 1997; Jerby a Ceder, 2006; Cheng a Chang, 2001; Derekenaris a kol., 2001; Rybansky, 2014)
- Spádové oblasti (Cinnamon a kol., 2008; Kimplel a kol., 2007; Upchurch a kol., 2004)
- Analýzy dopravní dostupnosti (Horák a kol., 2014; Jindra, 2016; Burian a kol., 2016; Liu a Zhu, 2004; Taylor a kol., 2006)
- Location – allocation analýzy (Yeh a Chow, 1996; Caddy a Carocci, 1999)
- Geokódování (Lee, 2009; Cayo a Talbot, 2003; Bonner a kol., 2003; Ratcliffe, 2004)
- Problém obchodního cestujícího (Mosheiov, 1994; Gavish a Graves, 1978; Laporte a kol., 1996; Reinelt, 1991; Laporte a Martello, 1990; Dorigo a Gambardella, 1997)
- Problém čínského poštěáka (Granot a Hamers, 2004; Malandraki a Daskin, 1993; Nobert a Picard, 1996; Minieka, 1979)

Mapové portály jako Mapy.cz, Google maps, Open Street Map, Bing maps či HERE maps běžně nabízejí možnost vyhledání optimální trasy. Tento proces využívá standardně grafových algoritmů blíže popsanych v kapitole 3.2, mezi nimi nejčastěji Dijkstrův algoritmus.

3.2.4 Další využití teorie grafů v geoinformatice

Kromě výše uvedených využití teorie grafů je možné najít aplikace i v dalších oblastech týkajících se geoinformatiky. V lodní dopravě využil síťové analýzy Coffin (2009), při vyhodnocování real-time dat (big data) Rathore a kol. (2015), pro konstrukci evakuačních plánů Dhamala (2014). Dopravní problémy v sociálních studiích řešil Roberts (1978). Při zkoumání vztahů mezi kvalitou krajiny a populací rybářských kolonií využil teorii grafů Garroway a kol. (2008). Ve 3D analýzách lze nalézt využití geometrických datových modelů v Lee (2004).

Široké uplatnění má teorie grafů také v kartografii. Například Wolff (2013), Mackaness a Beard (1993b), Neumann (1994) či (Wilson, 2002).

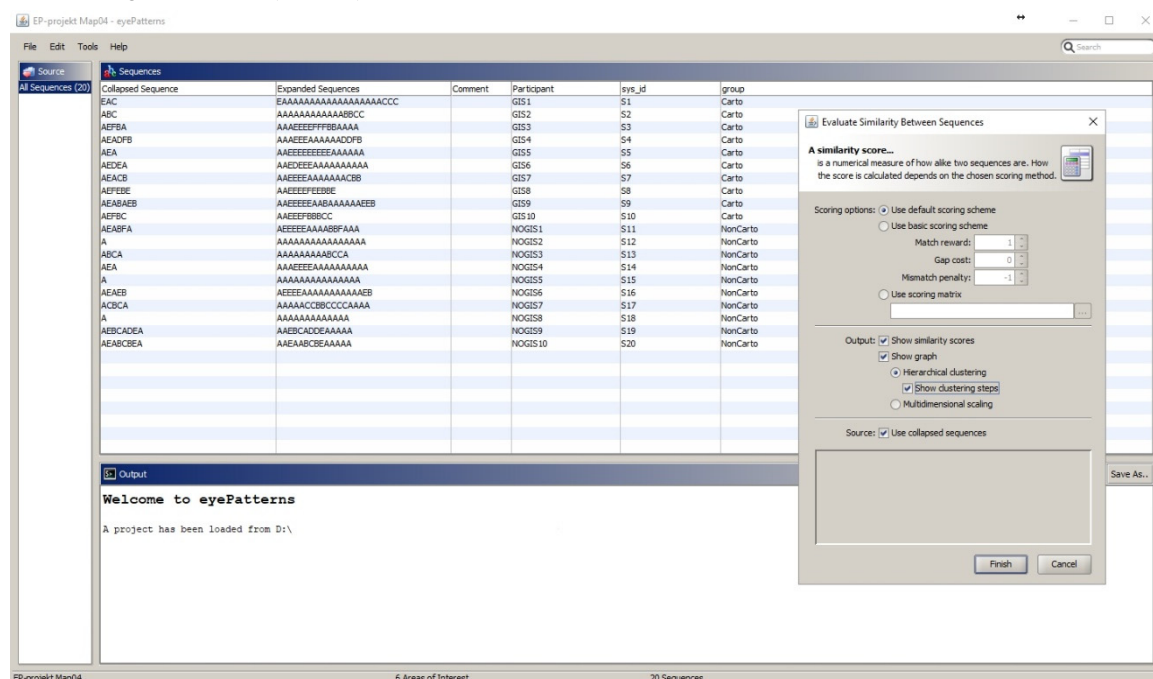
Dalším využitím teorie grafů v geografii a geoinformatice se zabývali například Arlinghaus a kol. (2002), Pitts (1965) či Zetterberg a kol. (2010).

4 DC1 – SCANGRAPH - TEORIE GRAFŮ V ANALÝZE EYE-TRACKING DAT

Výstupem prvního dílčího cíle je nově navržená metoda pro hledání podobnosti v eye-tracking datech a její aplikace v nástroji *ScanGraph* dostupném na adrese eyetracking.upol.cz/scangraph.

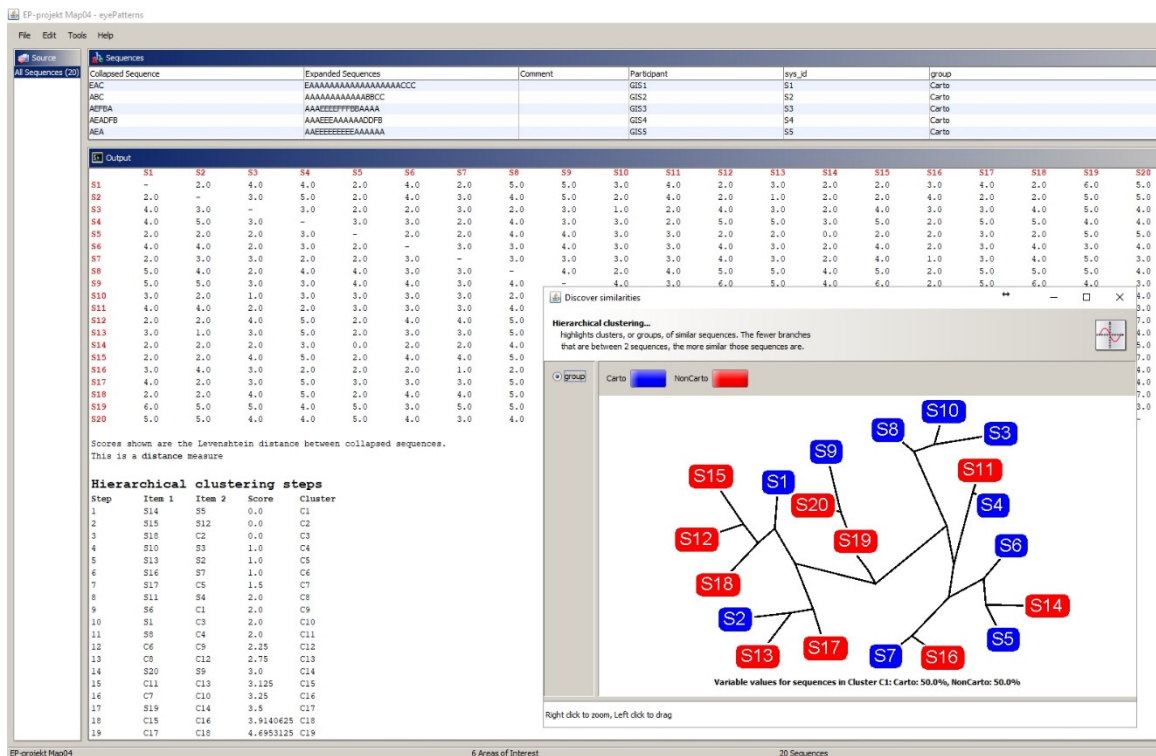
4.1 Důvody vytvoření vlastního nástroje

Při vyhodnocování eye-tracking experimentu, jehož se účastnilo n respondentů, je v rámci vyhodnocení každému respondentovi přiřazena sekvence nebo též řetězec písmen určující trajektorii pohybu jejich očí při průchodu navštívených oblastí zájmu. Pro reprezentaci daného experimentu pomocí grafu představují jeho vrcholy jednotlivé respondenty a hrany mezi vrcholy určují jisté vazby (podobnosti) mezi nimi. S tímto modelem pracuje nástroj *eyePatterns* (obr. 9).



Obr. 9 Prostředí nástroje *eyePatterns*

Pro výpočet podobnosti mezi jednotlivými sekvencemi je možno zvolit si výpočet pomocí Levenshteinovy vzdálenosti a Needleman-Wunschova algoritmu, kdy je mezi každými dvěma respondenty vypočtena hodnota určující podobnost mezi nimi. Získává se tak tabulka kladných celočíselných hodnot, kterou je možné chápat jako matici vzdáleností požadovaného grafu. Vytvoření grafu pak probíhá na základě hierarchical clustering algoritmu (obr. 10).



Obr. 10 Konstrukce grafu na základě hierarchical clustering algoritmu

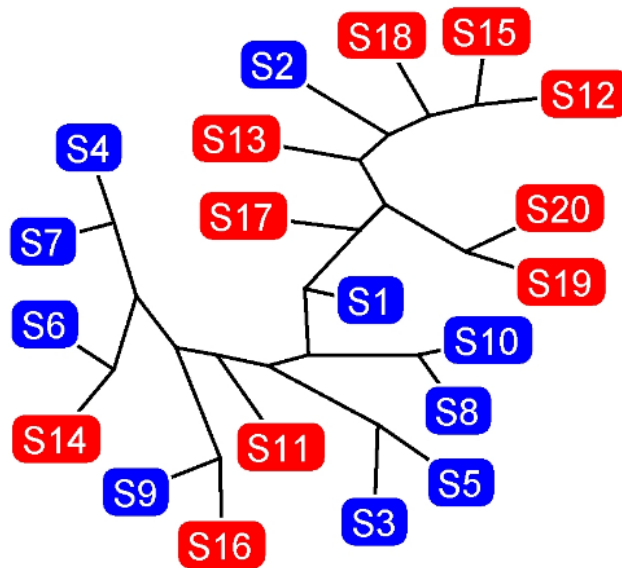
Graf, který je takto získán, je stromem, v němž listy představují respondenty daného experimentu a ostatní uzly představují jednotlivé skupiny, které vznikaly při iteracích hierarchical clustering algoritmu. Hrany grafu představují připojování skupin. Podle manuálu softwaru *eyePatterns* je tento graf vyhodnocován tak, že čím menší je hranová vzdálenost mezi listy (počet hran nutných k projití od jednoho listu ke druhému), tím větší je podobnost mezi respondenty reprezentovanými danými listy grafu.

Jak uvádí West (2006), software *eyePatterns* používá hierarchical clustering algoritmus pro výpočet podobností mezi sekvencemi. Klastrování rozděluje data do disjunktních podmnožin prvků, které sdílejí podobné znaky. Aglomerativní hierarchical clustering algoritmus postupuje od spojení nejbližších si sekvencí a končí připojením nejvzdálenější sekvence či klastru. Přičemž význam slova *nejbližší* a *nejvzdálenější* je v tomto kontextu chápáno jako nejvíce či nejméně podobné. Výstupem je pak graf, který má vizualizovat nejvíce a nejméně podobné sekvence (West a kol., 2006).

Nakreslení grafu vychází z konstrukce dendrogramu, který je výstupem hierarchical clustering algoritmu. Daný graf je tedy stromem – listy grafu představují jednotlivé analyzované sekvence. Avšak původní charakter dendrogramu již není z tohoto nakreslení patrný a tedy není vizuálně určitelný první krok hierarchical clustering algoritmu, tedy které dvě sekvence si dle výpočtů byly nejbližší a které byly naopak připojeny v dalších krocích. Tuto skutečnost lze zjistit pouze po zobrazení jednotlivých kroků algoritmu (show clustering steps). Navíc nakreslení grafu je možné tahem myši měnit. Informace v manuálu nástroje „*The fewer branches that are between two sequences, the more similar those sequences are.*“ tak nekoresponduje s touto vizualizací. Tento nesoulad je ilustrován na příkladu z případové studie.

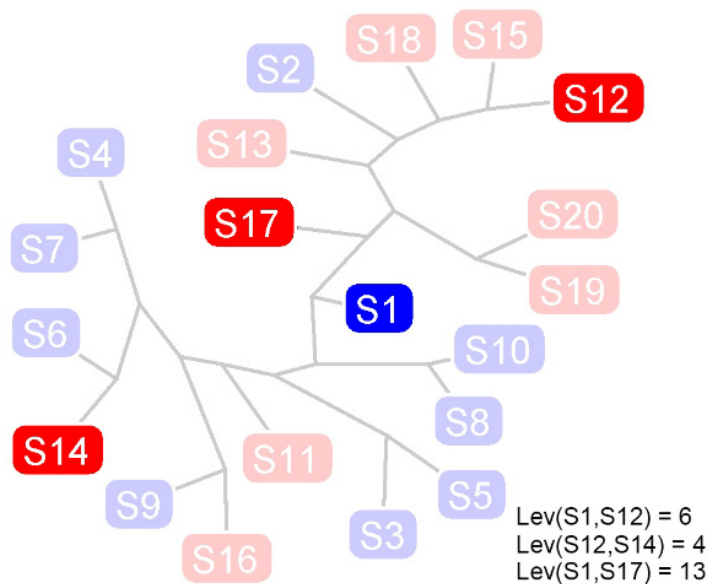
Případová studie zaměřená na vnímání mapové kompozice, která byla vytvořena v rámci výuky na Katedře geoinformatiky UP, obsahovala dvacet sekvencí (označených S1 až S20). Strom na obrázku 11 představuje výstup z nástroje *eyePatterns* při použití Levenshteinovy vzdálenosti (v nástroji označována jako defaultní výpočetní metoda).

Modré listy S1 až S10 představují sekvence respondentů kartografů a červené listy S11 až S20 respondenty laiky.

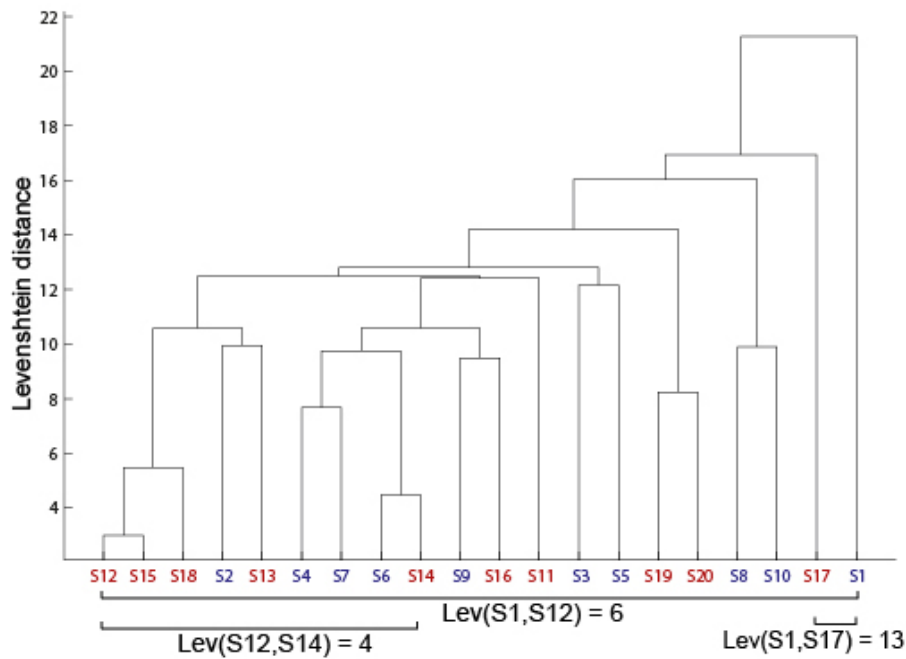


Obr. 11 Výstup z nástroje *eyePatterns*
– strom sestrožený na základě hierarchical clustering algoritmu

Na obrázku 12 jsou pak vyznačeny některé listy, na kterých je rozpor patrný. Zatímco mezi listy S1 a S17 existuje cesta délky tři, Levenshteinova vzdálenost mezi nimi je rovna 13. Délka cesty dle manuálu *eyePatterns* naznačuje velkou podobnost mezi těmito sekvencemi, avšak hodnota Levenshteinovy vzdálenosti oproti ostatním dvojicím sekvencí nevyovídá to samé. Zatímco mezi listy S12 a S14 existuje délka cesty 14, je hodnota jejich Levenshteinovy vzdálenosti rovna pouze 4, což znamená, že pouze čtyři změny (vlození, smazání a záměna znaku) byly nutné k přetvoření jedné sekvence v druhou a je tedy očekávána větší podobnost. Přesto délka cesty mezi těmito listy by nasvědčovala opačnému případu. Tuto situaci je možné pozorovat i na dendrogramu (obr. 13), jenž zobrazuje stejná data.



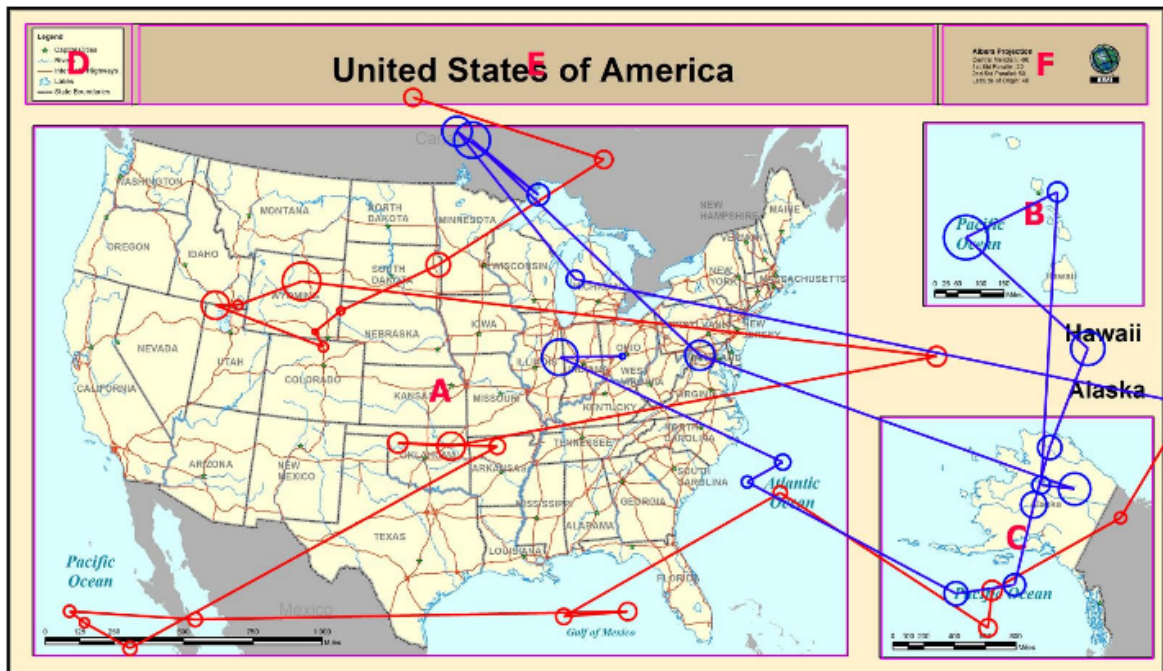
Obr. 12 Výstup z *eyePatterns* s vyznačenými nesrovnalostmi



Obr. 13 Výstup z eyePatterns ve formě dendrogramu

Scanpath trajektorie sekvencí S1 a S17 byly zobrazeny ve Scanpath modulu programu OGAMA (obr. 14). Je evidentní, že tyto trajektorie jsou velice odlišné, jak je také patrné z příslušných sekvencí (S1=EAAAAA AAAAAAAAAAAAAACCC, S17=AAAAACCBCCCC AAAA).

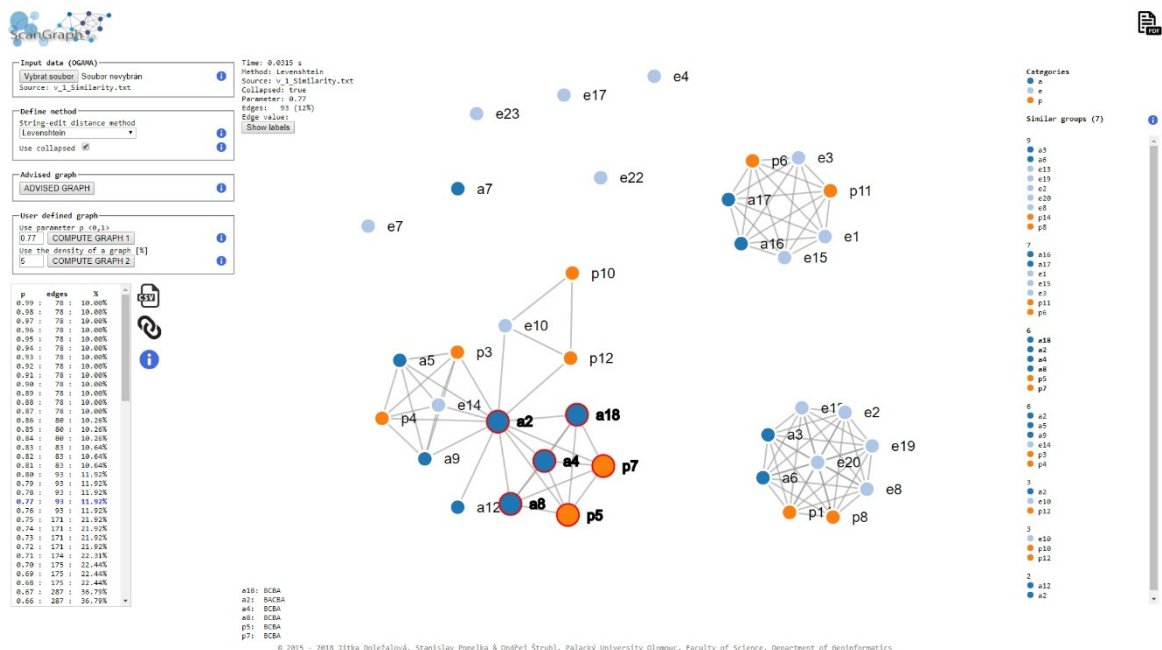
U respondenta S1 (modrá barva) bylo zaznamenáno méně fixací v mapovém poli. Prohlížel také oblast zájmu E (název mapy) a oblast zájmu B (mapa Aljašky), zatímco fixace respondenta S17 (červená barva) zde nebyly vůbec zaznamenány.



Obr. 14 Srovnání trajektorií S1 a S17, které nástroj eyePatterns označil jako velmi podobné

Při určování vzdálenosti (podobnosti) mezi jednotlivými sekvencemi používá nástroj *eyePatterns* pouze základní metriky popsané v kapitole 3.2.1. Nezohledňuje tak možnou délku sekvencí.

Vzhledem k nepřesnostem, které mohou použitím nástroje *eyePatterns* vzniknout, vznikla motivace vytvořit nový nástroj *ScanGraph* pro hledání podobností mezi sekvencemi eye-tracking dat. Oproti nástroji *eyePatterns*, kde všechny sekvence tvoří listy vizualizovaného stromu, vyzdvihuje *ScanGraph* pouze ty, které jsou si podobné při požadované míře podobnosti. Výstupem je tedy obyčejný graf, v němž podobné skupiny tvoří kliky (obr. 15).



Obr. 15 Prostředí nástroje ScanGraph

4.2 Popis nástroje ScanGraph

Nástroj byl vytvořen v programovacích jazycích PHP, HTML a C#, grafická vizualizace byla zajištěna knihovnamy D3.js a jQuery. Nástroj je umístěn na serverovém úložišti UP. Bližší popis úložiště lze nalézt v kapitole 2.6.

4.3 Výpočty ve ScanGraph

4.3.1 Modifikace algoritmů a popis jejich výpočtu

Vzdálenost a podobnost jsou veličiny vyjadřující vztah dvou bodů v prostoru, kdy se bodem rozumí vektor, objekt či v případě eye-tracking dat sekvenci navštívených AOI. Vzdálenost je možné chápat jako „míru nepodobnosti“. Čím je vzdálenost mezi dvěma objekty větší, tím jsou si méně podobné. Podobnost je tedy duální mírou ke vzdálenosti.

Pro korektní výpočet vzdálenosti a podobnosti mezi jednotlivými sekvencemi zavedeme pojem metriky a metrického prostoru.

Definice: Metrikou D na množině X nazýváme funkci $D: X \times X \rightarrow \mathbb{R}$, kde X je n -rozměrný prostor a splňující následující předpoklady:

$$0 \leq D(x, y) < \infty, \forall x, y \in X;$$

$$D(x, x) = 0, \forall x \in X$$

a má následující vlastnosti:

$$\begin{aligned}
D(x, y) &= D(y, x), \forall x, y \in X; \\
D(x, y) &= 0 \Leftrightarrow x = y; \\
D(x, z) &\leq D(x, y) + D(y, z), \forall x, y, z \in X.
\end{aligned}$$

Prostor X , ve kterém je metrika definována, nazýváme metrickým prostorem. Vzdálenost dvou objektů x a y je pak hodnota $D(x, y)$.

Dva objekty mají nulovou vzdálenost právě tehdy, jsou-li totožné.

Dále budou metriky použité v této práci splňovat vlastnost

$$\max_{x, y} (D(x, y)) = 1.$$

Tedy $D(x, y) \in (0, 1), \forall x, y \in X$ a dále budeme uvažovat pouze metriky s touto vlastností.

Definice: Metrikou podobnosti S na množině X je taková funkce $S: X \times X \rightarrow \mathbb{R}$, kde X je n -rozměrný prostor, splňující následující předpoklady:

$$\begin{aligned}
\exists S_0 \in \mathbb{R}: 0 \leq S(x, y) \leq S_0 < \infty, \forall x, y \in X; \\
S(x, x) &= S_0, \forall x \in X
\end{aligned}$$

a má následující vlastnosti:

$$\begin{aligned}
S(x, y) &= S(y, x), \forall x, y \in X; \\
S(x, y) &= S_0 \Leftrightarrow x = y.
\end{aligned}$$

V této práci budeme konstantu S_0 uvažovat rovnu 1. Tedy dva objekty jsou si nanejvýše podobné právě tehdy, jsou-li totožné.

Dualitu metriky vzdálenosti a metriky podobnosti vyjádříme vztahem:

$$S(x, y) = 1 - D(x, y), \forall x, y \in X.$$

Odtud vyjádříme trojúhelníkovou nerovnost pro metriku podobnosti:

$$S(x, z) \geq S(x, y) + S(y, z) - 1.$$

S ohledem na požadované vlastnosti byly metriky pro výpočet podobnosti mezi sekvencemi a, b navštívených oblastí zájmu vytvořeny následovně:

$$\begin{aligned}
Lev'(a, b) &= 1 - \frac{Lev(a, b)}{\max\{|a|, |b|\}}, \\
DL'(a, b) &= 1 - \frac{DL(a, b)}{\max\{|a|, |b|\}}, \\
NW'(a, b) &= \frac{NW(a, b)}{\max\{|a|, |b|\}}.
\end{aligned}$$

Pro všechny výše uvedené metriky podobnosti S platí:

$$\begin{aligned}
S(a, b) &\in (0, 1), \\
S(a, b) &= 0 \Leftrightarrow a = b, \\
S(a, b) &= 1 \Leftrightarrow a \cap b = \emptyset, \forall a, b \in X.
\end{aligned}$$

Poslední vlastnost interpretujeme tak, že sekvence a a b nemají žádný společný znak, přičemž délka daných sekvencí může být různá.

4.3.2 Doporučení pro volbu algoritmu

Před vyhodnocováním experimentu musí uživatel vybrat metodu výpočtu míry podobnosti. Nástroj nabízí tři možnosti: *Levenshtein distance*, *Damerau-Levenshtein distance* a *Needleman-Wunsch algoritmus*. Tyto metody jsou blíže popsány v kapitolách 3.2.1 a 4.3.1 a jejich limity jsou popsány v kapitole 4.5. Zatímco Needleman-Wunsch algoritmus vyhledává souhlasné znaky v sekvencích, Levenshteinova vzdálenost zjišťuje

počet nutných kroků vložení, smazání a substituce (v případě Damerau-Levenshteinovy vzdálenosti ještě i transpozice) pro transformaci jedné sekvence ve druhou.

Jelikož Damerau-Levenshteinova vzdálenost umožňuje transpozici znaků jako jeden krok, je více vhodná pro experimenty, kde větší důraz je kladen spíše na to, zda jednotlivé oblasti zájmu byly respondentem vůbec navštíveny, než na jejich pořadí. Jestliže cílem experimentu je například hodnotit, zda se respondent více zaměřuje například na pravý horní roh stimulu než na levý dolní roh, je vhodnější použít Damerau-Levenshteinovu vzdálenost. Obdobně, je-li cílem experimentu zjistit, zda uživatel například použil legendu mapy, bez většího významu, ve které fázi prohlížení se na ni zaměřil, pak je doporučena Damerau-Levenshteinova vzdálenost. Na druhou stranu, jestliže stimul má podobu cyklu či vývojového diagramu a je tedy důležité, v jakém pořadí uživatel jednotlivé prvky, zastoupené oblastmi zájmu, navštívil, pak je doporučena Levenshteinova vzdálenost. Stejně doporučení platí, pokud je pro uživatele významné, zda si respondent prohlédl nejprve legendu a až po té navštívil mapové pole nebo zda nejprve řešil zadaný úkol experimentu v mapovém poli a až po té řešení zkontroloval v legendě.

Na níže uvedeném příkladu je patrné, že dvě sekvence, které prochází stejnými oblastmi zájmu, ale v jiném pořadí, jsou metrikou založenou na Levenshteinově vzdálenosti vyhodnoceny jako zcela odlišené. Zatímco pro metriku založenou na Damerau-Levenshteinově vzdálenosti, kde pro transpozici prvků postačuje jeden krok, je hodnota již 0,25.

$$Lev'(ABCD, DCBA) = 0$$

$$DL'(ABCD, DCBA) = 0,25$$

Obecně je možné Levenshteinovu a Damerau-Levenshteinovu vzdálenost nazvat jako „přísnější“ metriky. Levenshteinova vzdálenost sice dosahuje velice blízkých hodnot (ve velké míře totožných) jako Needleman-Wunsch algoritmus, avšak může nabývat hodnot o malý rozdíl nižší. Rozdíl je možné najít opět u sekvencí, kde jsou navštívené stejné oblasti zájmu v odlišném pořadí, jak ilustruje níže uvedený příklad.

$$Lev(ABCD, DABC) = 0,25$$

$$NW(ABCD, DABC) = 0,5$$

$$Lev(ABABAB, BABABA) = 0,67$$

$$NW(ABABAB, BABABA) = 0,8$$

Souhrnně lze tedy říci, že hlavním kritériem pro výběr metody výpočtu podobnosti je významnost pořadí navštívených oblastí zájmu, přičemž nejvíce benevolentní je metrika založená na Needleman-Wunschově algoritmu a naopak nejstriktnější je metrika založená na Levenshteinově vzdálenosti.

4.3.3 Určení míry podobnosti

Při hledání vzájemně si podobných sekvencí vypočítá nástroj *ScanGraph* míru podobnosti pro všechny dvojice analyzovaných sekvencí. Při výpočtu podobnosti všech sekvencí jsou tak postupně vytvořeny tři symetrické čtvercové matice řádu n , kde n je počet analyzovaných sekvencí.

První matice s názvem *Original matrix* ($O = (o_{ij})_{i,j=1}^n$) je maticí hodnot zvoleného algoritmu (Levenshteinova vzdálenost, Damerau-Levenshteinova vzdálenost, Needleman-Wunsch algoritmus).

Jestliže sekvence a odpovídá i -tému řádku a sekvence b odpovídá j -tému sloupci, pak

$$o_{ij} = \begin{cases} Lev(a, b) \\ DL(a, b), \\ NW(a, b) \end{cases}$$

$\forall a, b \in X$. Speciálně pro $a = b$ platí

$$o_{ii} = \begin{cases} 0, Lev(a, a), DL(a, a) \\ |a|, NW(a, a) \end{cases}$$

Druhou maticí je pak *Modified matrix* ($M = (m)_{i,j=1}^n$), jejíž prvky jsou hodnoty modifikovaných algoritmů popsanych v kapitole 4.3.1. Tedy

$$m_{ij} = \begin{cases} Lev'(a, b) \\ DL'(a, b), \\ NW'(a, b) \end{cases}$$

$\forall a, b \in X$. Speciálně pro $a = b$ platí $m_{ii} = 1$.

Matice *Modified matrix* může být označována jako matice míry podobnosti. Prvky této matice mohou nabývat hodnot od 0 do 1, jak bylo popsáno v kapitole 4.3.1.

Určení poslední matice *Adjacency matrix* závisí na hodnotě parametru p , který je zadán uživatelem, a to buď přímou volbou, nebo pomocí hustoty grafu. Podrobný popis zadávání této hodnoty je obsahem kapitoly 4.4.2. Hodnota parametru p je mírou hraniční podobnosti požadované uživatelem.

Matice *Adjacency matrix* ($A = (a_{ij})_{i,j=1}^n$), neboli matice sousednosti, je výchozí maticí pro vyhodnocení výsledků. Její hodnoty jsou určeny z *Modified matrix* následovně:

$$a_{ij} = \begin{cases} 1, m_{ij} \geq p \\ 0, m_{ij} < p \end{cases}$$

Obecné vlastnosti matice sousednosti jsou popsány v kapitole 3.1.

Hodnota 1 je tedy přiřazena těm dvojicím sekvencí a a b , jejichž podobnost je rovna nebo větší než daná míra p . Hodnota 0 pak náleží těm dvojicím sekvencí, jejichž podobnost je menší než zadaná míra p .

4.3.4 Vizualizace výsledků

Na základě matice sousednosti je sestaven obyčejný graf o n vrcholech. Vrcholy grafu reprezentují jednotlivé sekvence a hranami jsou pak spojeny ty vrcholy, jejichž míra podobnosti je rovna nebo větší než daná míra p . V případě, že sekvence daného experimentu jsou uživatelem rozděleny do různých sledovaných kategorií (muži, ženy, experti, laici apod.), jsou vrcholy grafu barevně odlišeny podle těchto kategorií.

Skupiny vzájemně si podobných sekvencí pak představují kliky v tomto grafu. Jejich prohlížení je pak popsáno v následující kapitole 4.4.2. Vyhledávání klik je realizováno pomocí Bron-Kerbosch algoritmu, který je popsán v kapitole 3.1.

4.4 Použití ScanGraph (uživatelem)

Použití nástroje *ScanGraph* je přímočaré. Uživatel zobrazí naměřená eye-tracking data v open-source aplikaci OGAMA nebo SMI BeGaze, kde vytvoří jednotlivá AOI a vyexportuje řetězce fixací. Soubor vyexportovaný z prostředí OGAMA může být přímo nahrán do nástroje *ScanGraph*. V případě použití SMI, data mohou být převedena za použití vytvořeného nástroje *smi2scangraph*.

4.4.1 Import dat

OGAMA

Nástroj *ScanGraph* byl vyvinut pro práci s daty exportovanými z open-source prostředí OGAMA (OpenGazeAndMouseAnalyzer), což je software pro analýzu pohybů očí a počítačové myši pro prezentační účely (Voškůhler a kol., 2008). OGAMA obsahuje nástroj s názvem *Výpočet Levenshteinovy vzdálenosti*, který umožňuje výpočet Levenshteinovy vzdálenosti mezi trajektoriemi jednotlivých respondentů. Vzdálenosti mezi sekvencemi mohou být počítány na základě pravidelné gridové sítě nebo uživatelem definovaných oblastí zájmu.

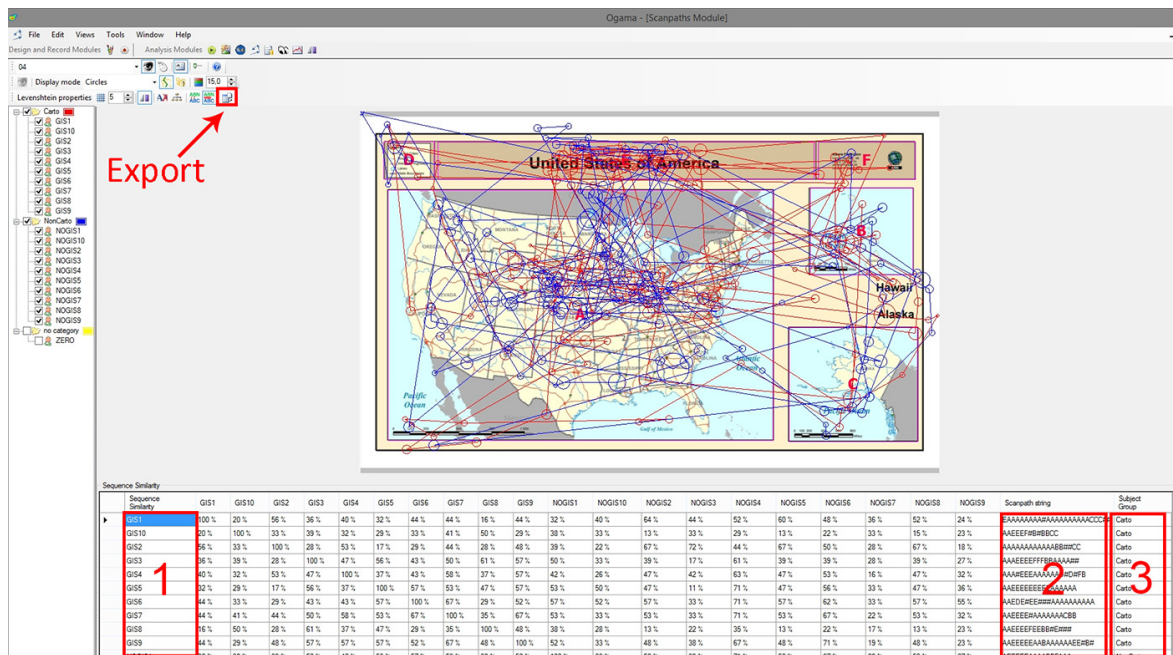
Výstupem tohoto nástroje je matice, jež podle manuálu *eyePatterns* má vyjadřovat míru podobnosti mezi sekvencemi (v procentech) podle Levenshteinovy vzdálenosti. Avšak při kontrole těchto hodnot je zjevné, že se jedná o hodnoty vycházející z Needleman-Wunschova algoritmu. Dále je výstupem také seznam řetězců fixací pro každého respondenta. Pro import do nástroje *ScanGraph* jsou důležité pouze řetězce fixací. Ovšem hodnoty podobnosti vypočítané v OGAMA nejsou využity, jelikož nástroj *ScanGraph* používá algoritmus modifikovaný (kapitola 4.3.1) tak, aby zohledňoval délku jednotlivých sekvencí.

První krok analýzy nástrojem *ScanGraph* je vytvoření oblastí zájmu v analyzovaném stimulu. Například na stimulu na obrázku 16 je vyznačeno šest oblastí zájmu.



Obr. 16 Oblasti zájmu vyznačené na stimulu s mapovou kompozicí # 1

Dalším krokem je zobrazení trajektorií respondentů a řetězců fixací. Textový dokument s řetězci fixací může být exportován z prostředí OGAMA a přímo použit jako vstupní data do nástroje *ScanGraph* (obr. 17).



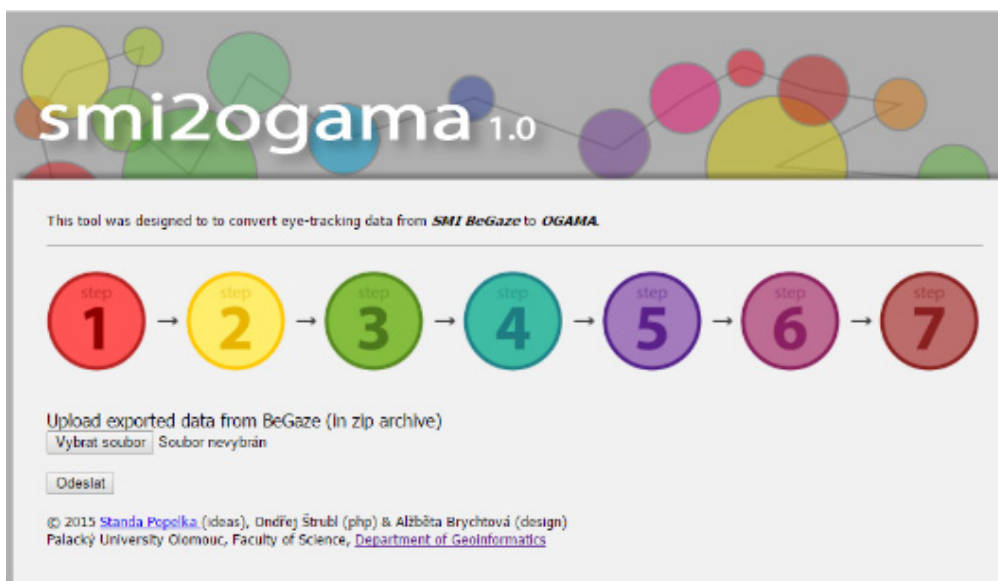
Obr. 17 Prostedí OGAMA Scanpath modulu. Jsou vypočítány hodnoty Levenshteinovy vzdálenosti mezi vybranými respondenty a mohou být exportovány jako textový dokument. Pouze pole Subject name (1), Scanpath string (2) a přiřazení ke kategori (3) jsou použity v nástroji ScanGraph.

Konverze dat z SMI BeGaze

Jak bylo popsáno výše, *ScanGraph* pracuje s přímým výstupem z prostředí OGAMA. Dokument obsahuje identifikaci respondentů, řetězce navštívených oblastí zájmu a název kategorie, do které jsou jednotliví respondenti zahrnuti (muži, ženy, experti, laici apod.). Jelikož v eye-tracking laboratoři na Katedře geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci je často využíván SMI eye-tracker, byl z praktických důvodů vytvořen nástroj na konverzi surových dat z SMI BeGaze do formátu kompatibilního s prostředím OGAMA. Bohužel OGAMA však nepodporuje import dynamických stimulů jako video nebo záznam na monitoru.

Z tohoto důvodu byl vyvinut jednoduchý nástroj s názvem *smi2scangraph*. Při použití toho nástroje může uživatel přímo exportovat údaje o oblastech zájmu z SMI BeGaze. Po konverzi exportovaného dokumentu přes *smi2scangraph* může být výstup přímo nahrán do nástroje *ScanGraph* bez přechodu přes prostředí OGAMA. Tato možnost je obzvláště přínosná v případech dynamických stimulů a dynamických oblastí zájmu.

Pro konverzi dat z SMI BeGaze do prostředí OGAMA byl vyvinut nástroj SMI2OGAMA. Jeho použití je také velmi jednoduché. Uživatel exportuje souřadnice pohledu jako surová (raw) data z SMI BeGaze jako samostatné soubory pro každého respondenta. ZIP soubor s těmito surovými daty je nahrán do SMI2OGAMA a konverze do OGAMA formátu pak probíhá na serveru. Po několika sekundách je ZIP soubor stažen. V další fázi importuje uživatel po jednom tyto dokumenty do nového projektu aplikace OGAMA. Nakonec obrázky stimulů jsou zkopírovány do *SlideResources* složky projektu OGAMA. V rozhraní SMI2OGAMA je možné nalézt více informací o tomto procesu (obr. 18).



Obr. 18 Interface webového nástroje SMI2OGAMA

4.4.2 Použití ScanGraph

Nástroj *ScanGraph* umožňuje analyzovat jeden soubor či „multifile“ – ZIP soubor obsahující data z několikastimulového experimentu.

Při prvním spuštění webového nástroje *ScanGraph* se zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadovaný způsob výpočtu. Nahrání souboru provede uživatel stiskem tlačítka *Choose file*.

V obou případech vybere dále uživatel v poli *Define method* algoritmus výpočtu pro *String-Edit-Distance*. K dispozici jsou tři algoritmy – *Levenshtein Distance*, *Needleman-Wunsch algoritmus* a *Damerau-Levenshtein Distance* popsané v kapitole 3.2.1. Konkrétní metodu volí uživatel na základě charakteru dat svého experimentu a zaměřením cíle experimentu. Jistá doporučení lze nalézt v kapitole 4.3.2.

Ve stejném poli *Define method* uživatel dále specifikuje zaškrtnutím políčkem *Use collapsed*, zda požaduje, aby řetězec navštívených AOI neobsahoval násobné znaky.

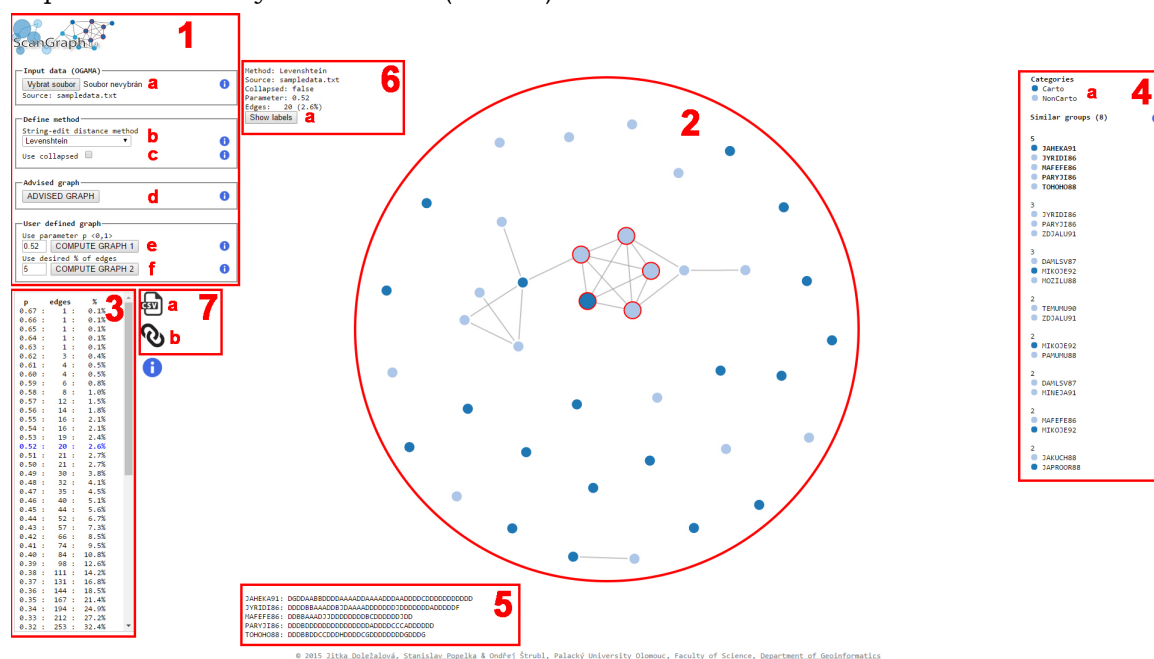
Tuto možnost uživatel opět využívá na základě očekávaného charakteru dat a stanoveného cíle experimentu. Při nezaškrtnutí políčka *Use collapsed* zůstanou analyzované sekvence v původním tvaru, tedy tak, jak je uživatel nahrál. Pokud naopak tuto možnost uživatel zaškrtně, budou všechny násobné znaky nahrazeny jedním znakem - AAABCCCCDD → ABCD. Tuto možnost volí uživatel v případě, že pro analýzu experimentu není podstatný počet fixací v jednotlivých oblastech zájmu, ale pouze pořadí jejich navštívení.

V případě analýzy multifile souboru zvolí uživatel navíc metodu zpracování těchto souborů. Na výběr jsou dvě možnosti, metoda *Mean* (průměr) a metoda *Occurance* (výskyt). Metody jsou blíže popsány v závěru této kapitoly.

Ke spuštění výpočtu může dojít na základě zadaného parametru p podobnosti nebo hustoty hran grafu. Parametr p určuje požadovanou míru podobnosti řetězců a jeho výpočet je blíže popsán v kapitole 4.3.3 **Chyba! Nenalezen zdroj odkazů.** Parametr p nabývá hodnot z intervalu $\langle 0,1 \rangle$, přičemž hodnota 0 představuje nulovou (žádnou) podobnost mezi řetězci a hodnota 1 absolutní podobnost, tedy totožnost řetězců. Čím vyšší je hodnota tohoto parametru, tím vyšší nárok je kladen na vzájemnou podobnost

mezi řetězci. Stiskem tlačítka *Compute graph 1* se spustí výpočet, který určí skupiny řetězců vykazující zadanou míru podobnosti a vyšší. Alternativou je zadání hustoty grafu. Tato hodnota je chápána jako procento zastoupení hran výsledného grafu vůči úplnému grafu při daném počtu vrcholů a tedy nabývá hodnot z intervalu $(0,100)$. Úplný graf s $n \geq 2$ vrcholy obsahuje $\binom{n}{2}$ hran a hustota hran grafu (v procentech) je pak hodnota $\frac{|E|}{\binom{n}{2}} \cdot 100$, kde $|E|$ je počet hran grafu a n je počet jeho vrcholů. Stiskem tlačítka *Compute graph 2* uživatel spustí výpočet skupin podobných řetězců tak, aby výsledný graf reprezentující tyto skupiny měl právě zadanou hustotu hran nebo nejbližší možnou nižší. Pro uživatele, kteří nemají předchozí zkušenost s nástrojem, je vhodné využít tlačítka *Advised graph*, které spustí výpočet podobných skupin reprezentovaných grafem o hustotě hran 5 % nebo nejbližší možné nižší. Tato hodnota byla určena na základě předchozích zkušeností s vyhodnocováním eye-trackingových dat. Graf s pětiprocentní hustotou hran má zvykle poměrně dobrou vypovídající hodnotu. Tento graf může sloužit pouze pro spuštění výpočtu a uživatel může dále s výstupem pracovat a nastavovat jeho požadované parametry v poli *User define graph*. Graf s požadovanou mírou podobnosti má jednoznačně určenou hustotu hran. Nelze tedy tyto dva parametry nastavovat současně a vždy se zobrazí pouze graf odpovídající volbě posledního stisknutého tlačítka *Compute graph 1* a *Compute graph 2* či *Advised graph*.

Nyní se zobrazí zbylá pole nástroje. V centru prohlížeče je zobrazen samotný graf, který odpovídá nastaveným hodnotám (obr. 19).



Obr. 19 Jednotlivé prvky nástroje ScanGraph

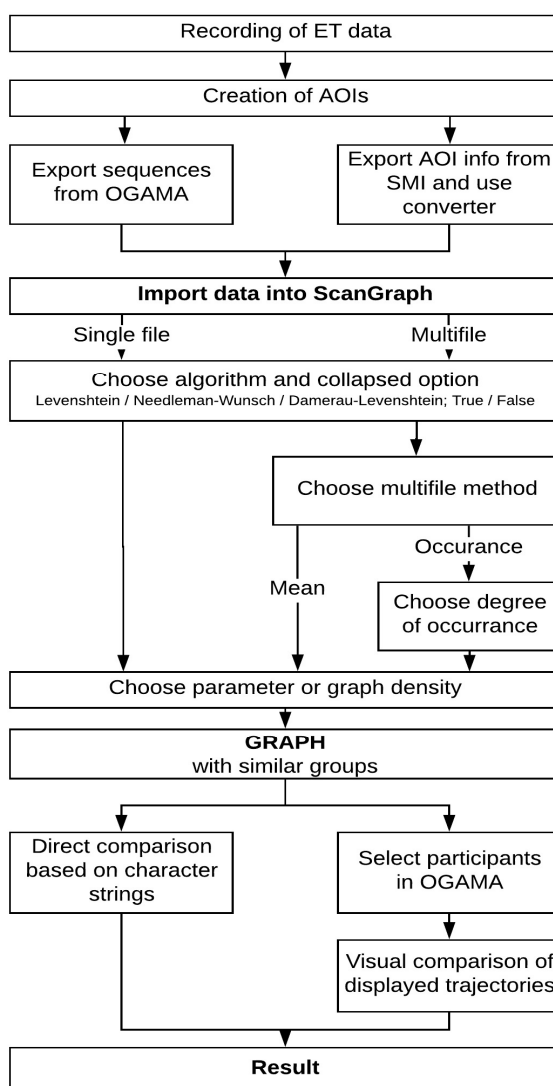
Vrcholy zobrazeného grafu reprezentují jednotlivé sekvence vyhodnocovaného experimentu. Rozdílné barvy určují příslušnost sekvencí k vyšetřovaným kategoriím (např. muž/žena, odborník/laik apod.). Ty sekvence, které jsou si podobné s mírou podobnosti větší či rovnou zadané míře, jsou spojeny hranou. Skupiny vzájemně si podobných sekvencí tak vytváří kliky v příslušném grafu. Množina netriviálních klik grafu tvoří jeho pokrytí. Jejich seznam je možné si prohlédnout v poli v pravé části. Po kliknutí na seznam vrcholů kliky se červeně podbarví příslušné vrcholy v grafu a současně v části se zobrazí příslušné řetězce. Kliky jsou popsány seznamem svých vrcholů a jsou seřazeny sestupně podle velikosti.

V levém horním rohu pole grafu nalezne uživatel celkový přehled svého experimentu a pomocí tlačítka *Show labels* přidá popisky k vrcholům grafu. Tento přehled obsahuje

údaje *Time* (doba trvání výpočtu), *Method* (zvolená metoda pro výpočet String-Edit-Method), *Source* (název nahraného analyzovaného souboru), *Collapsed* (zda uživatel zvolil originální řetězce nebo řetězce bez násobnosti fixací), *Parameter* (zvolená hodnota parametru při volbě *Compute graph 1* či přiřazená hodnota parametru při volbě *Compute graph 2*), *Edges* (počet hran výsledného grafu včetně hustoty hran) a *Edge value* (ohodnocení hrany mírou podobnosti).

V levé dolní části se zobrazí tabulka, která uvádí všechny možné varianty grafu, který může data daného experimentu reprezentovat. Varianty jsou seřazeny podle hodnoty parametru (míry podobnosti) od nejvyšší možné po nejnižší nulovou. V prvním sloupci tabulky je uvedena hodnota míry podobnosti, k níž je jednoznačně dán počet hran zobrazených ve druhém sloupci a tedy i hustota hran zobrazena ve třetím sloupci. Kliknutím na libovolný řádek této tabulky zobrazí uživatel graf příslušný tomuto řádku a tedy příslušné míře podobnosti.

Kliknutím na ikonu CSV (7a) uživatel stáhne soubor se všemi výpočty, seznamem podobných skupin a jejich příslušných řetězců, vstupními daty a přehledem nastavení. Ikona 7b umožní vytvoření permanentního odkazu pro zobrazení grafu na jiném zařízení.



Obr. 20: Schéma použití nástroje ScanGraph

Multifile

Kromě zpracování dat z jednotlivých stimulů umí nástroj *ScanGraph* zpracovávat data také ze ZIP souborů obsahující TXT soubory pro více stimulů z jednoho experimentu. V takovém případě vypočítaná podobnost a výsledný graf budou představovat podobnosti napříč všemi stimuly obsaženými v ZIP souboru.

Jakmile uživatel nahraje ZIP soubor, rozhraní nástroje *ScanGraph* se automaticky změní a bude zobrazovat možnosti výběru multifile metod. Nabízeny jsou dvě možnosti: *Mean* a *Occurrence*.

V případě volby *Mean* vypočítá *ScanGraph* Modified matice M_1, \dots, M_k pro všech k stimulů obsažených v ZIP souboru a následně vypočítá jejich aritmetický průměr a vytvoří tak novou Modified matici $\bar{M} = (\bar{m}_{ij})_{i,j=1}^n$, kde

$$\bar{M} = \frac{1}{k} \sum_{l=1}^k M_l.$$

S hodnotami v matici \bar{M} se dále pracuje stejně jako v případě Modified matice u single stimulů.

V případě volby metody *Occurrence* vychází výpočet nové Modified matice $\tilde{M} = (\tilde{m}_{ij})_{i,j=1}^n$ z jednotlivých Adjacency matic všech k stimulů ZIP souboru. Uživatel v tomto případě kromě míry podobnosti p musí zadat navíc hodnotu Degree of Occurrence c , tedy hodnotu jež představuje požadované procentuální zastoupení výskytu podobnosti mezi jednotlivými dvojicemi respondentů napříč stimuly. Tato hodnota pak slouží pro určení nové Adjacency matice $\tilde{A} = (\tilde{a}_{ij})_{i,j=1}^n$ stejně jako míra podobnosti p v případě single stimulů. Tedy

$$\tilde{M} = \frac{1}{k} \sum_{l=1}^k A_l,$$

$$\tilde{a}_{ij} = \begin{cases} 1, & \tilde{m}_{ij} \geq c \\ 0, & \tilde{m}_{ij} < c \end{cases}$$

Například, jestliže ZIP soubor obsahuje deset stimulů a zadaná hodnota *degree of occurrence* bude 0,8, budou spojeni hranou právě ti respondenti, kteří si byli podobní při dané míře podobnosti v 8 a více stimulech.

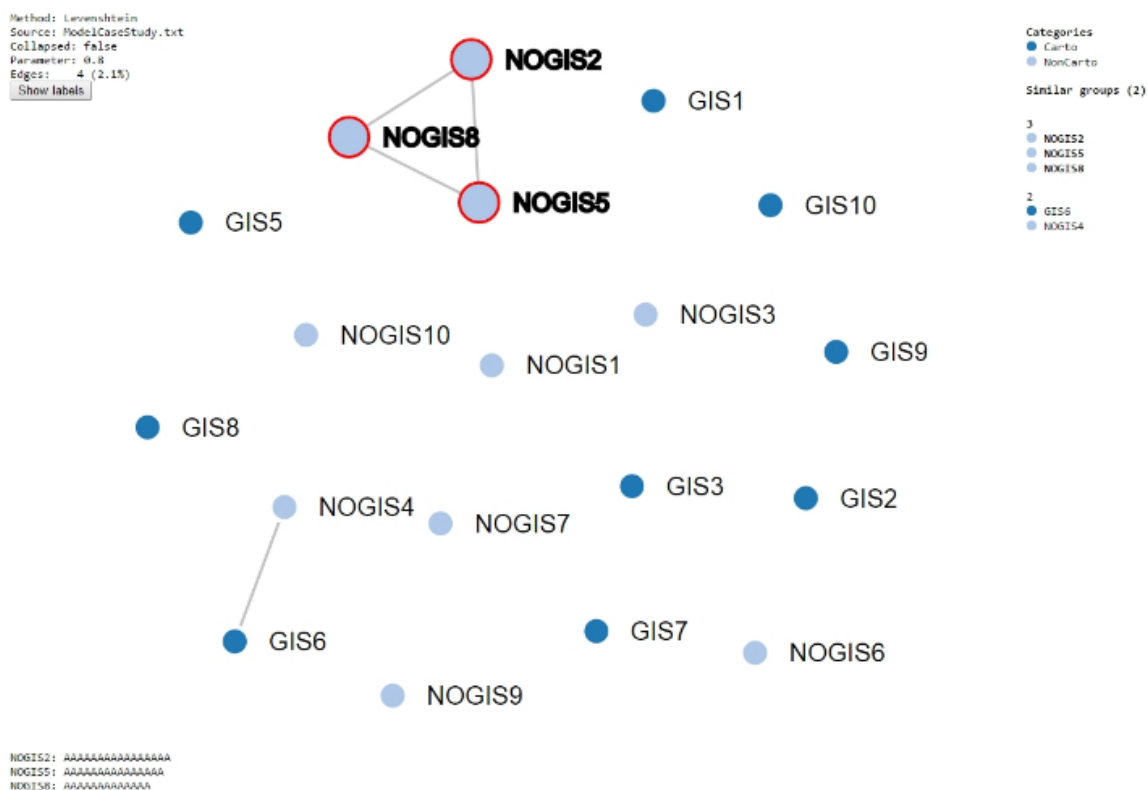
Volba *Mean* je doporučena v případě, že rozptyl mezi jednotlivými hodnotami na souhlasných pozicích v Modified maticích není příliš velký. Tedy když tyto hodnoty jsou srovnatelné a není očekávána velká směrodatná odchylka. V opačném případě hodnota aritmetického průměru může být zavádějící a je tedy vhodnější použít metodu *Occurrence*. Obzvláště pokud jsou očekávány extrémní odchylky vlivem chyb a problémů při sběru dat.

4.4.3 Ukázka výsledků a jejich interpretace

Funkcionalita nástroje *ScanGraph* bude demonstrována na jednoduché případové studii zaměřené na percepci map s různou kompozicí. Data pro tuto studii byla naměřena v rámci předmětu KGI/KAPRO (Kartografický projekt). Autory experimentu vytvořeného v roce 2011 jsou Marek Adamec, Martin Klícha, Jan Procházka a Roman Zbranek.

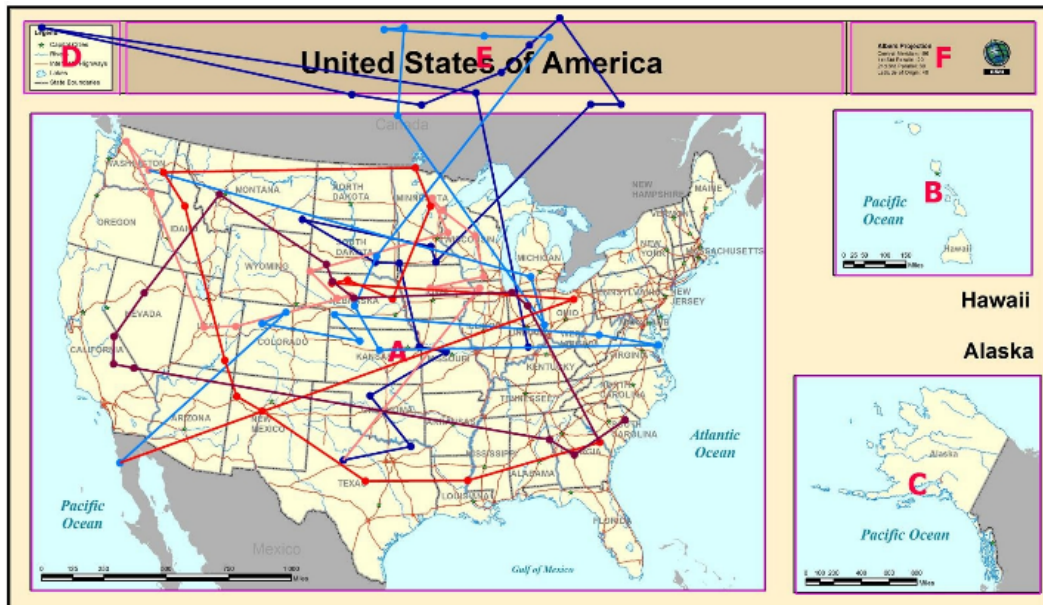
Vyšetřováno bylo jak pořadí navštívených oblastí zájmu, tak počet fixací. Pro analýzu byla nastavena varianta non-collapsed a vybrána varianta Levenshteinovy vzdálenosti. Při

míře podobnosti $p = 0,8$ byly nalezeny dvě netriviální kliky – jedna dvouprvková a jedna tříprvková (obr. 21).



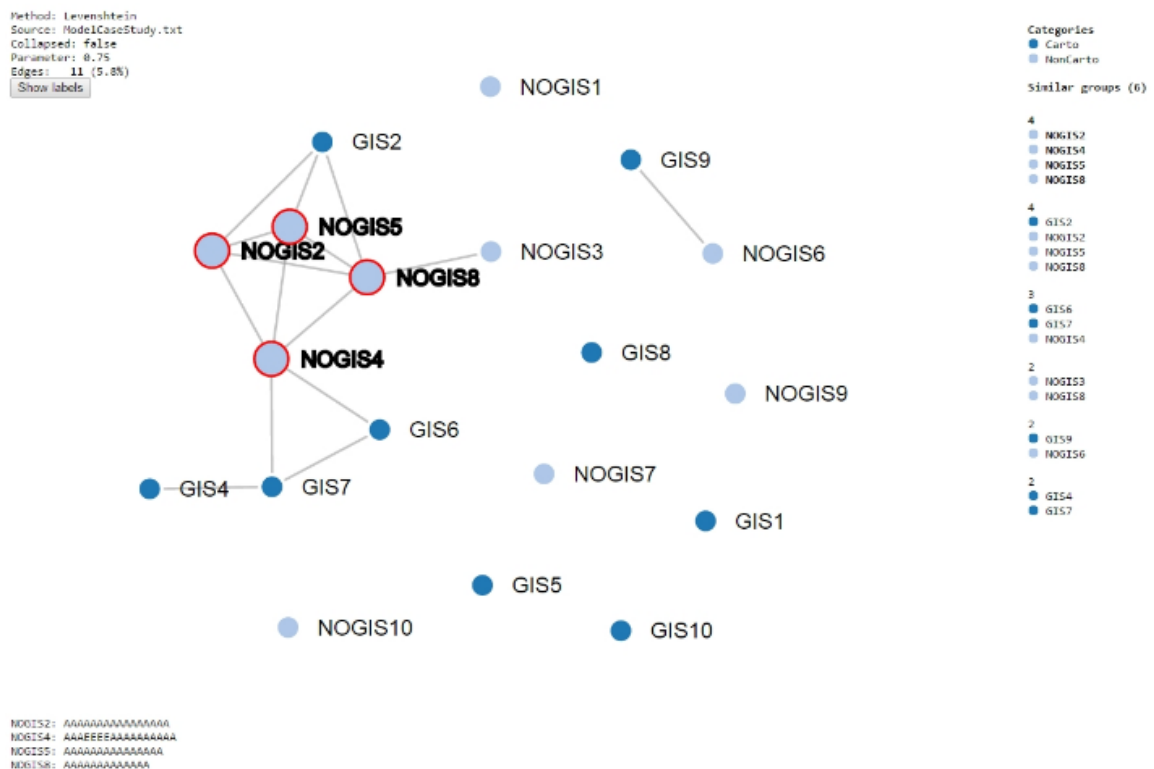
Obr. 21 Výstup nástroje ScanGraph zobrazující dvě netriviální kliky pro hodnotu míry podobnosti $p = 0,8$

Obrázek 22 zobrazuje trajektorie těchto respondentů. Tříprvková skupina je sestavena ze tří respondentů nekartografů (jako „nekartograf“ jsou v této práci označováni respondenti, kteří uvedli, že nemají kartografické vzdělání). Všichni tři strávili celý čas prohlížením oblasti zájmu A, která obsahovala mapové pole. Jejich trajektorie jsou znázorněny v odstínech červené. Druhá klika obsahovala sekvence dvou respondentů a jejich trajektorie jsou znázorněny v odstínech modré. U obou těchto respondentů byl stejný počet fixací (17), většina z nich taktéž v mapovém poli (AOI A) a ostatní na titulu mapy (AOI E). U respondenta GIS6 (tmavě modrá trajektorie) byla zaznamenána jedna fixace v legendě mapy.



Obr. 22 Trajektorie pěti respondentů, kteří vytváří dvě kliky (při míře podobnosti $p = 0,8$). Tři respondenti, zobrazení v odstínech červené, strávili veškerý čas v mapovém poli. Respondenti GIS6 a NOGIS4 (odstíny modré) navštívili také mapové pole.

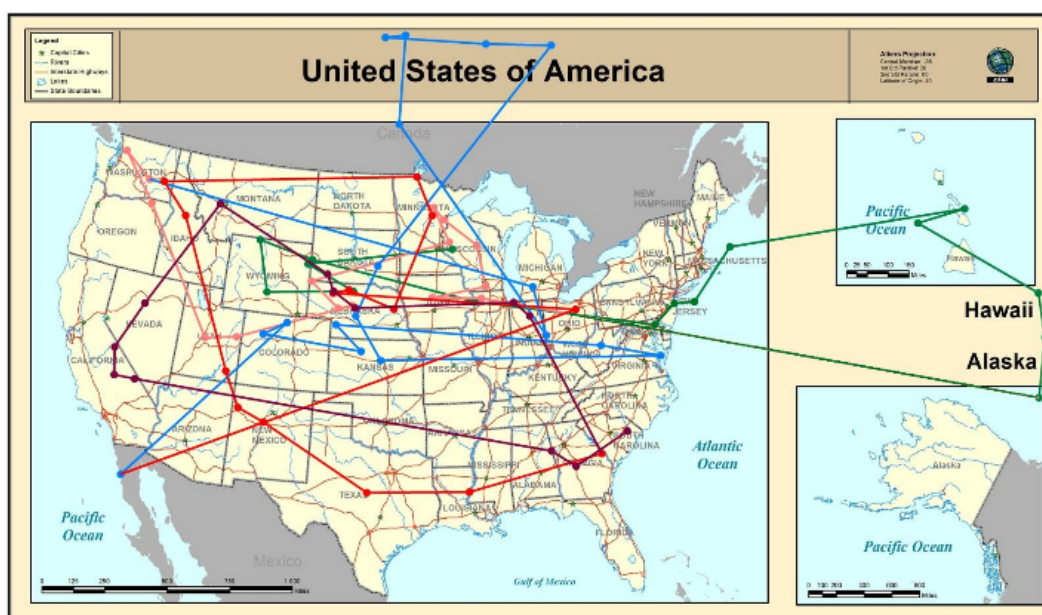
Při snížení míry podobnosti na hodnotu $p = 0,75$ bylo nalezeno šest netriviálních klik (obr. 23). Dvě čtyřprvkové kliky obsahovaly původní tříprvkovou kliku sestavenou z nekartografů jako svoji podmnožinu. V jednom případě byla rozšířena o respondenta NOGIS4 a ve druhém o respondenta GIS2. Což znamená, že respondenti NOGIS4 a GIS2 si byli určitým způsobem podobní s původními respondenty (NOGIS2, NOGIS5 a NOGIS8) při míře podobnosti 0,75, ale nebyli si podobní při této míře navzájem.



Obr. 23 Výstup nástroje ScanGraph zobrazující kliky při míře podobnosti $p = 0,75$

Trajektorie respondentů z obrázku 23 jsou zobrazeny na obrázku 24. Trajektorie respondentů NOGIS2, NOGIS5 a NOGIS8 jsou opět znázorněny v odstínech červené.

Modrou barvou je znázorněn respondent NOGIS4 a zelená barva náleží respondentu GIS2. Je patrné, že lze nalézt podobnost modré a zelené trajektorie vůči červeným trajektoriím, ale podobnost mezi nimi navzájem není znatelná.



Obr. 24 Trajektorie pěti respondentů vytvářejících vytvářejících dvě kliky při míře podobnosti $p = 0,75$. Respondenti NOGIS2, NOGIS5 a NOGIS8 jsou zobrazeny v odstínech červené. Respondenti NOGIS4 (modrá trajektorie) a GIS2 (zelená linie) jsou oba podobní červeným trajektoriím, nikoliv však sobě navzájem.

4.5 Limity nástroje ScanGraph

V současném stavu vývoje nástroje *ScanGraph*, tedy při aktuálně používaných algoritmech, je *ScanGraph* schopen počítat podobnosti pro přinejmenším 320 respondentů nebo sekvence o maximální délce 320 znaků. Náročnost výpočtu je ovlivněna volbou *String-Edit-Distance* algoritmu. Pro zjištění výpočetních limitů nástroje *ScanGraph* byl proveden test. Z tohoto důvodu byly vytvořeny náhodné sekvence generované s použitím webové aplikace¹, která vytváří sekvence dané délky. Pro tento test bylo koncipováno několik scénářů. Byly vygenerovány soubory obsahující 20, 40, 80, 160 a 320 respondentů se stejnými hodnotami délek sekvencí a byl změřen výpočetní čas. Všechny tři algoritmy (*Needleman-Wunsch*, *Levenshtein*, *Damerau-Levenshtein*) proběhly při hustotě hran grafu 5 %. Výsledek tohoto testování je shrnut v tabulce 1.

Jako nejrychlejší algoritmus se projevila Levenshteinova vzdálenost. Nástroj *ScanGraph* pracuje velice rychle pro sekvence s počtem znaků nejvýše 250. Při větších délkách sekvencí časová náročnost prudce roste. Pro délku sekvence o 160 znacích pracuje *ScanGraph* spolehlivě až pro 320 respondentů. Výpočetní čas byl v takovém případě naměřen 11,38 s. Výpočet pomocí *Needleman-Wunsch* algoritmu a *Damerau-Levenshteinovy* vzdálenosti zabírá poměrně více času a *ScanGraph* tak při vyšším počtu respondentů dosahuje svých limitů. Kliky je schopen nalézt při těchto dvou algoritmech pro sekvence délek 80 znaků a pro 80 respondentů. Pro sekvence s větší délkou nebyl *ScanGraph* schopen nalézt řešení (N/A v tabulce 1).

¹ <http://www.converter.cz/passgen/pswdgen.php>

Tab. 1 Souhrn ukázkového testu pro 20, 40, 80, 160 a 320 znaků a respondentů. Hodnoty v tabulce představují délku výpočtu pro graf s 5% hranovou hustotou v sekundách. Pro kombinace s hodnotou N/A nebyl schopen ScanGraph výpočet provést.

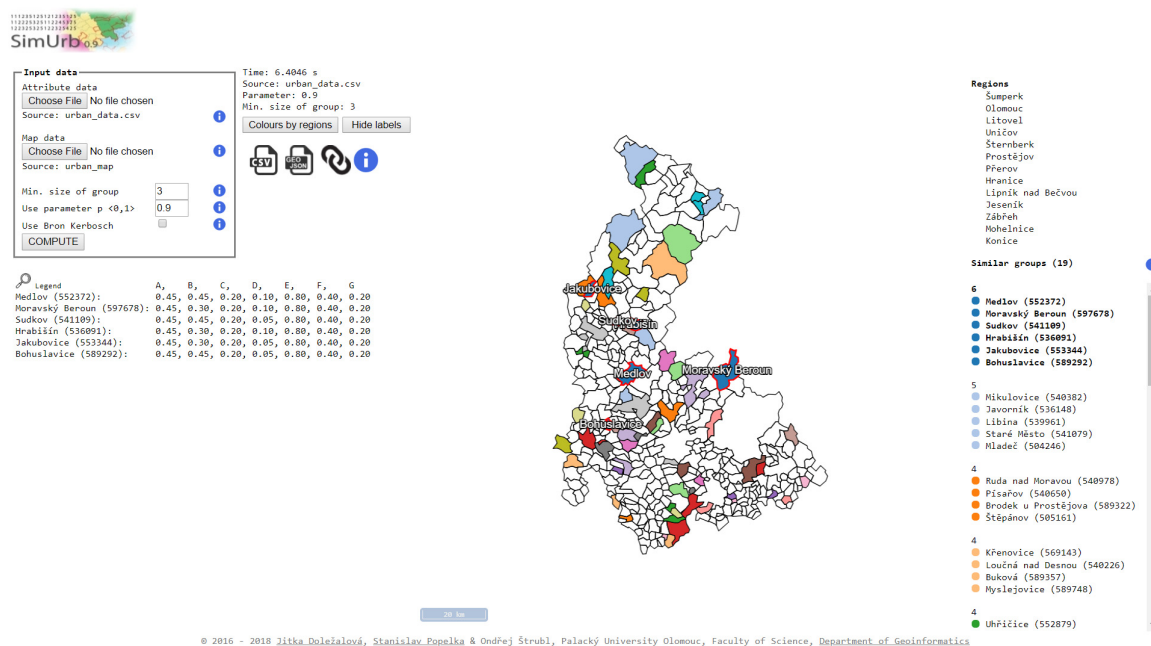
			Number of participants				
			20	40	80	160	320
Number of characters	20	LEV	0,02	0,03	0,09	0,30	1,32
		N-W	2,01	2,04	7,94	32,21	135,80
		D-L	2,64	2,60	9,82	42,44	159,07
	40	LEV	0,02	0,03	0,10	0,37	2,03
		N-W	2,00	7,63	30,10	126,43	596,40
		D-L	2,78	10,32	43,47	203,32	766,15
	80	LEV	0,02	0,05	0,19	0,85	3,72
		N-W	9,73	37,79	162,21	N/A	N/A
		D-L	13,34	50,64	202,74	N/A	N/A
	160	LEV	0,06	0,16	0,74	3,40	11,38
		N-W	63,90	259,14	N/A	N/A	N/A
		D-L	63,04	220,49	N/A	N/A	N/A
	320	LEV	37,32	151,77	657,62	N/A	N/A
		N-W	542,09	N/A	N/A	N/A	N/A
		D-L	254,17	N/A	N/A	N/A	N/A

5 DC2 – SIMURB – TEORIE GRAFŮ V PROSTOROVÉM PLÁNOVÁNÍ

Výstupem druhého dílčího cíle je navržená metoda pro hledání podobnosti aplikovatelná v prostorovém plánování a dostupná v nástroji *SimUrb* na adrese eyetracking.upol.cz/simurb.

5.1 Důvody vytvoření vlastního nástroje

Shlukovací metody, jež jsou běžně dostupné například díky softwaru ArcGIS Pro, zařazují všechny vyhodnocované objekty do skupin a tvoří tak rozklad celé množiny na třídy podobnosti. Tyto přístupy jsou v praxi běžně používány. Jejich nevýhodou může však být to, že vyskytnou-li se objekty s velkou odchylkou hodnot koeficientů, mohou být přiřazeny ke skupině, se kterou nevykazují žádnou či malou podobnost. Z tohoto důvodu byla navržená nová metoda vyhledávání podobných objektů, jež se vyznačuje dvěma odlišnými vlastnostmi od v praxi běžně používaných metod. Uživateli je umožněno nastavit si výši míry podobnosti dle jeho předpokladů o vyhodnocovaných datech a velikost skupin, jež jsou pro jeho analýzu zajímavé. Výsledné skupiny podobných objektů pak jsou tvořeny pouze těmi objekty, jež danou míru podobnosti vykazují. Ostatní objekty nejsou zařazeny do žádné skupiny.



Obr. 25 Prostředí nástroje *SimUrb*

5.2 Popis nástroje *SimUrb*

Nástroj byl vytvořen v programovacích jazycích PHP, HTML a C#, grafická vizualizace byla zajištěna knihovnami D3.js a jQuery. Nástroj je umístěn na serverovém úložišti UP. Bližší popis úložiště lze nalézt v kapitole 2.6.

5.3 Výpočty v SimUrb

5.3.1 Metrika

Cílem nástroje *SimUrb* je vyhledání vzájemně si podobných skupin objektů na základě předem definovaných atributů. Každému z n zkoumaných objektů je tak přiřazena uspořádaná m -tice koeficientů, které pocházejí z předem definovaných intervalů.

Označme $X = \{x_1, x_2, \dots, x_n\}$ množinu zkoumaných objektů, $K = \{K_1, K_2, \dots, K_m\}$ množinu atributů, $\underline{x}_i = (k_{i1}, k_{i2}, \dots, k_{im})$ uspořádanou m -tici koeficientů objektu x_i , kde $k_i \in \langle \min k_i; \max k_i \rangle, \forall i \in 1, \dots, m$, a $w_i = (w_1, w_2, \dots, w_m)$ uspořádaná m -tice vah, kde w_i je vahou atributu $K_i, w_i \in \mathbb{R}, \forall i = 1, \dots, m$.

Jednotlivé objekty s jejich přiřazenou m -ticí koeficientů lze chápat jako body v m -rozměrném prostoru.

Pro určení míry podobnosti byla navržena metrika, jež vychází z euklidovské vzdálenosti. Nechť x_i a x_j jsou dva objekty z množiny X . Pak pro míru jejich podobnosti $d(x_1, x_2)$ platí:

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^m \frac{w_j (k_{1j} - k_{2j})^2}{w_j (\max K_j - \min K_j)^2}}.$$

Pro hodnoty míry podobnosti $d(x_1, x_2)$ platí $d(x_1, x_2) \in \langle 0, 1 \rangle$, přičemž $d(x_1, x_2) = 1 \Leftrightarrow \underline{x}_1 = \underline{x}_2$ a $d(x_1, x_2) = 0 \Leftrightarrow \underline{x}_1 = (\min K_1, \min K_2, \dots, \min K_m) \wedge \underline{x}_2 = (\max K_1, \max K_2, \dots, \max K_m)$. Je zřejmé, že totéž bude platit, jestliže minimálních hodnot intervalů atributů bude nabývat objekt x_2 a maximálních hodnot objekt x_1 . Hodnota $d(x_1, x_2)$ je tedy rostoucí s rostoucí podobností mezi objekty.

5.3.2 Určení míry podobnosti

Při hledání vzájemně si podobných objektů vypočítá nástroj *SimUrb* míru podobnosti pro všechny dvojice analyzovaných objektů. Při výpočtu podobnosti všech objektů jsou tak postupně vytvořeny tři symetrické čtvercové matice řádu n , kde n je počet analyzovaných objektů.

První matice s názvem *Original matrix* ($O = (o_{ij})_{i,j=1}^n$) je maticí hodnot euklidovské vzdálenosti pro všechny dvojice x_i, x_j z množiny všech objektů X .

Druhou maticí je pak *Modified matrix* ($M = (m_{ij})_{i,j=1}^n$), jejíž prvky jsou hodnoty $d(x_i, x_j), \forall x_i, x_j \in X$.

Určení poslední matice *Adjacency matrix* závisí na hodnotě parametru p , který je zadán uživatelem. Hodnota parametru p je mírou hraniční podobnosti požadované uživatelem a nabývá hodnot z intervalu $\langle 0, 1 \rangle$, kdy $p = 1$ indikuje totožné sekvence a $p = 0$ zcela odlišné (jedna sekvence je tvořena minimálními možnými hodnotami a druhá maximálními).

Matice *Adjacency matrix* ($A = (a_{ij})_{i,j=1}^n$), nebo-li matice sousednosti, je výchozí maticí pro vyhodnocení výsledků. Její hodnoty jsou určeny z Modified matrix následovně:

$$a_{ij} = \begin{cases} 1, & m_{ij} \geq p \\ 0, & m_{ij} < p \end{cases}$$

Obecné vlastnosti matice sousednosti jsou popsány v kapitole 3.1.

Hodnota 1 je tedy přiřazena těm dvojicím objektů x_i a x_j , jejichž podobnost je rovna nebo větší než daná míra p . Hodnota 0 pak náleží těm dvojicím sekvencí, jejichž podobnost je menší než zadaná míra p .

5.3.3 Vizualizace výsledků

Vzájemně si podobné skupiny tvoří opět kliky v obyčejném grafu. Vzhledem k tomu, že však většina analýz, k nimž je nástroj *SimUrb* uzpůsoben, pracuje s desítkami až stovkami řetězců, by vizualizace formou grafu byla nepřehledná a matoucí. Navíc protože nástroj umožňuje vložit i mapový podklad pro sekvence s geografickou lokací, je vhodnější metodou vizualizace právě v mapě.

Cílem uživatele nástroje *SimUrb* bývá obvykle vyhledat disjunktní shluky vzájemně si podobných objektů. Metody shlukování jsou blíže popsány v kapitole 3.2.2. Nedisjunktní skupiny často ztrácí přehlednost a především význam pro jejich analýzu. Z toho důvodu pracuje nástroj *SimUrb* odlišným způsobem než popsaný nástroj *ScanGraph* (kapitola 4.3.4). Algoritmus výpočtu je popsán v kapitole 5.3.

5.4 Použití SimUrb (uživatelé)

5.4.1 Import dat

Obdobně jako v případě nástroje *ScanGraph* je prvním krokem použití *SimUrbu* import dat. Na rozdíl od *ScanGraph*, kde byl využit přímo export z programu OGAMA, případně z konvertoru SMI2SCANGRAPH, je u tohoto nástroje nutné data předem připravit do požadované struktury. Nástroj je navržen tak, aby četl soubory ve formátu CSV. Tento soubor musí obsahovat tři povinná pole, a to ID, NAME a REGION. V poli ID je obsažen unikátní identifikátor konkrétní obce. V poli NAME je uveden její název a v poli REGION pak název územně správního celku, do kterého tato obec patří. Typicky se může jednat o obec s rozšířenou působností (ORP), okres či kraj. Následuje libovolný počet polí, ve kterých jsou uvedeny hodnoty jednotlivých indikátorů. Názvy těchto polí budou po výpočtu podobností zobrazeny po zobrazení legendy.

Kromě této dané struktury souboru je nutné, aby byly přesně definovány také první tři řádky tabulky. Ty jsou označeny jako *weight*, *min* a *max*. V poli *weight* je uvedena hodnota váhy jednotlivých indikátorů (jako desetinné číslo oddělené čárkou), v polích *min* a *max* pak minimální a maximální hodnoty škály, ve které se indikátory pohybují. Toto řešení bylo zvoleno proto, aby se předešlo problémům s různými intervaly hodnot jednotlivých indikátorů. Struktura souboru pro import je znázorněna v tabulce 2. Pokud byly hodnoty indikátoru „PRIR_PR_14“ vybírány z intervalu 1 až 5, bude v řádku *min* hodnota 1 a v poli *max* hodnota 5.

Tab. 2 Vzorová struktura souboru CSV pro import do nástroje *SimUrb*.

ID	NAME	REGION	PRIR_PR_14	MIGR_SA_14	DOK_BYT_14	M_URBAN_16	IND_STA_14	IND_VZD_11	TR_CEPR_14
<i>weight</i>			0,15	0,15	0,2	0,05	0,2	0,2	0,05
<i>min</i>			1	1	1	1	1	1	1
<i>max</i>			5	5	5	5	5	5	5
500020	Petrov nad Desnou	Šumperk	3	3	2	1	3	2	4
500496	Olomouc	Olomouc	3	3	4	2	3	5	4
500526	Bělkovice-Lašťany	Olomouc	3	4	4	3	5	3	4

Protože nástroj *SimUrb* zobrazuje nalezené podobnosti vizuálně v mapě, je nutné do aplikace nahrát rovněž soubor formátu JSON obsahující polygony analyzovaných obcí. Tento soubor má rovněž jasně danou strukturu. Musí obsahovat tři pole. Jedná se o ID,

NAME a REGION_NAME. Obsah těchto polí je stejný jako v případě CSV souboru. Vzorová struktura je znázorněna v tabulce 3.

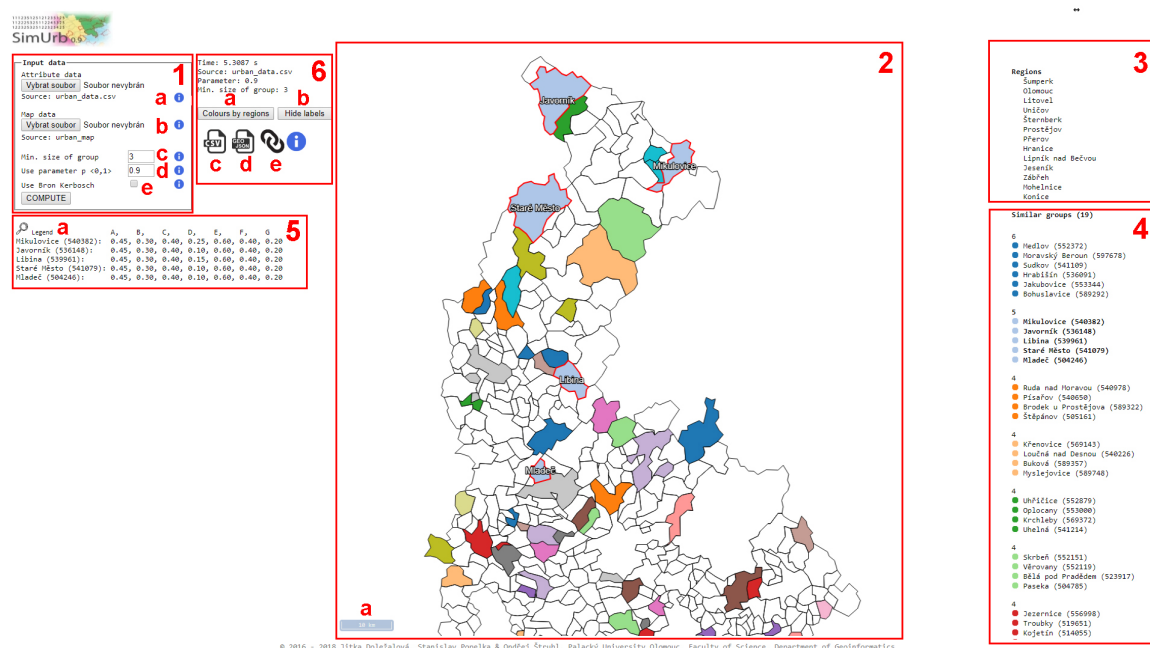
Tab. 3 Vzorová struktura souboru JSON pro import do nástroje SimUrb.

	ID	NAME	REGION_NAME
0	500020	Petrov nad Desnou	Šumperk
1	500496	Olomouc	Olomouc
2	500526	Bělkovice-Lašťany	Olomouc

Data musí mít projekci WGS84 (EPSG: 4326). Je rovněž vhodné, aby nebyl soubor příliš datově objemný, což by zvýšilo čas pro načtení nástroje. Toho lze dosáhnout zjednodušením hranic jednotlivých obcí (simplify geometry).

V případě, že by uživatel do nástroje soubor s mapou nenahrál, výsledné kliky reprezentující podobné obce by byly zobrazeny pouze v pravé části nástroje a nikoliv v mapě.

5.4.2 Použití SimUrb



Obr. 26 Jednotlivé prvky nástroje SimUrb

Při prvním spuštění webového nástroje *SimUrb* (obr. 26) se zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadovaný způsob výpočtu.

Nahrání souboru s daty provede uživatel volbou tlačítka *Choose file* v poli *Attribute data*. Dále může uživatel nahrát mapový podklad, opět volbou tlačítka *Choose file*, ovšem v poli *Map data*. Tento výběr není povinný. Metoda výpočtu podobnosti je v nástroji *SimUrb* pouze jediná a tedy se volba uživateli nenabízí. Metoda je blíže popsána v kapitole 5.3

Dalším krokem je nastavení minimálního počtu prvků ve skupině a hodnota míry podobnosti. Jelikož často uživatele při analýze zajímají pouze skupiny s větším

množstvím prvků, byla do nástroje přidána právě volba nastavení minimálního počtu prvků ve skupině. Algoritmus výpočtu se pak zastaví u takto velkých skupin a dále již nepokračuje, čímž se celková doba výpočtu zkrátí. Míra podobnosti p může nabývat hodnot z intervalu $(0,1)$, přičemž hodnota rovna jedné značí zcela totožné sekvence a hodnota rovna nule značí sekvence s největší možnou odlišností. Podrobnější popis je v kapitole 5.3.1.

Mezi volby nástroje bylo přidáno také políčko pro možnost použití Bron-Kerbosch algoritmu pro vyhledávání všech klik mezi sekvencemi. Tato volba je doporučena pouze pro malé soubory, kde cílem analýzy není vytvoření disjunktních shluků a charakter dat umožňuje pro správné vyhodnocení překryv mezi skupinami. Volbou tlačítka *Compute* je spuštěn výpočet.

Poté se zobrazí v pravé části seznam všech nalezených klik. Kliknutím na kteroukoliv skupinu se v levé části zobrazí jejich seznam společně s příslušnými sekvencemi koeficientů. Najetím myši na ikonu lupy se pak uživateli zobrazí legenda s popisem názvu koeficientů a jejich vahami.

V levé horní části je uvedena délka trvání výpočtu, název zdrojových dat, zvolená míra podobnosti a minimální velikost skupiny. K dispozici jsou stejně jako v nástroji *ScanGraph* možnosti stáhnout CSV soubor s analyzovanými daty a vypočtenými hodnotami v *Original*, *Modified* a *Adjacency matici* a možnost permanentního odkazu.

Jestliže uživatel vložil spolu s daty i mapový podklad, jsou výsledné skupiny vizualizovány i v mapě. Primárně jsou charakterizovány areálovým znakem, barevně rozlišené podle příslušných nalezených skupin, přičemž jestliže uživatel klikl na některou skupinu v jejich seznamu, jsou v mapě hranice příslušných regionů zvýrazněny liniovým znakem červené barvy.

Volbou tlačítka *Colours by groups* se pak přepne vizualizace podle větších administrativních celků, jsou-li v datech zadány. Volbou tlačítka *Show labels* nebo *Hide labels* pak uživatel může zobrazit či skrýt názvy objektů v mapě. Zobrazit detaily jednotlivých skupin lze pak i kliknutím na regiony v mapovém podkladu.

5.4.3 Interpretace výsledků

Jednotlivé objekty množiny X mohou být reprezentovány vrcholy grafu. Sestrojení grafu je pak dáno maticí sousednosti dle daného parametru p . Skupiny vzájemně si podobných objektů pak tvoří kliky v tomto grafu. Jelikož však běžné případové studie prostorového plánování obsahují desítky až stovky objektů, není vyhledávání všech klik v daném grafu vhodné. Ačkoliv by takovéto řešení bylo výpočetně správné, ztrácí s rostoucím počtem objektů interpretační hodnotu. Cílem uživatele je nalézt velké skupiny objektů vykazujících podobnost, které se ve svých prvcích nepřekrývají. Současně malé skupiny (jednoprvkové, dvouprvkové, tříprvkové) nemají pro vyhodnocení studie velký význam. Z tohoto důvodu rozklad na třídy podobnosti probíhá následujícím způsobem:

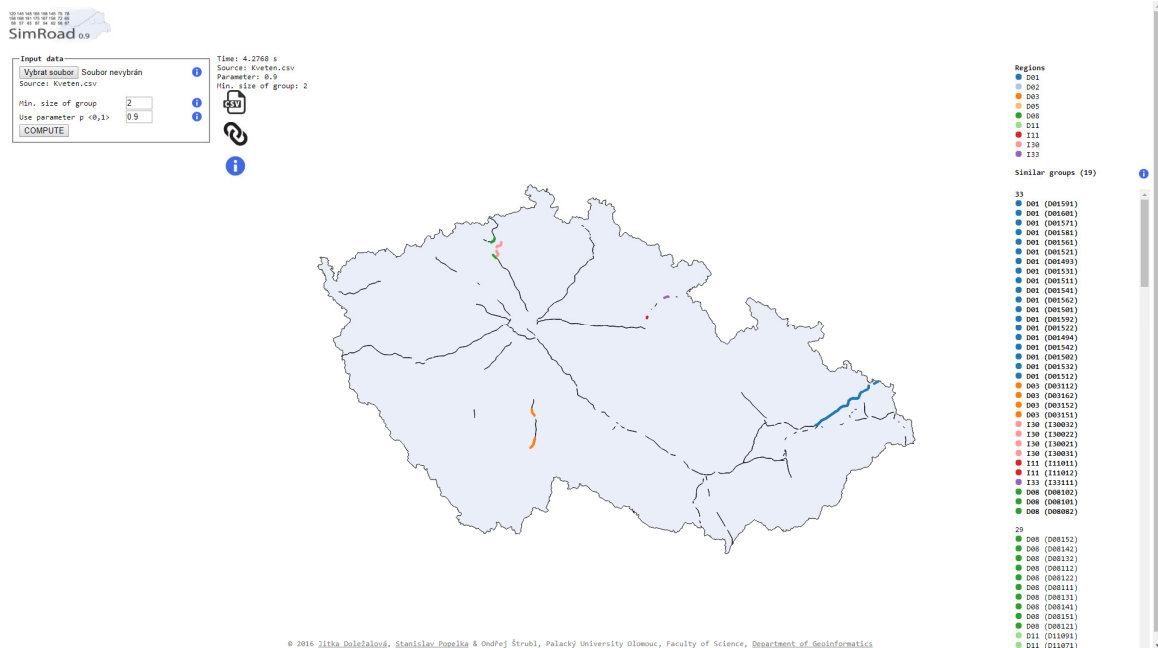
1. Nástroj nalezne největší kliku grafu (maximum clique problem). V případě více největších klik zvolí první nalezenou.
2. Objekty patřící do nalezené kliky vyřadí z množiny X .
3. Opakuje kroky 1 a 2, dokud množina X není prázdná.

Výsledkem jsou pak skupiny vzájemně si podobných objektů při dané míře podobnosti, jež lze nazvat shluky.

6 DC3 – SIMROAD – TEORIE GRAFŮ V DOPRAVĚ

6.1 Důvody vytvoření vlastního nástroje

V úlohách prostorového plánování je možné se často setkat s typizací na základě podobnosti mezi územními celky, přičemž lze využít mnoho různých nástrojů. Pro obdobné vyhodnocování liniových úseků však není žádný nástroj k dispozici, ačkoliv vytváření typů i u takovýchto objektů má svůj význam. Tato skutečnost byla tedy motivací pro vytvoření nástroje *SimRoad*.



Obr. 27 Prostředí nástroje *SimRoad*

6.2 O nástroji *SimRoad* jak byl vytvořen

Nástroj byl vytvořen v programovacích jazycích PHP, HTML a C#, grafická vizualizace byla zajištěna knihovnamy D3.js a jQuery. Nástroj je umístěn na serverovém úložišti UP. Blíže popis úložiště lze nalézt v kapitole 2.6.

6.3 Výpočty v *SimRoad*

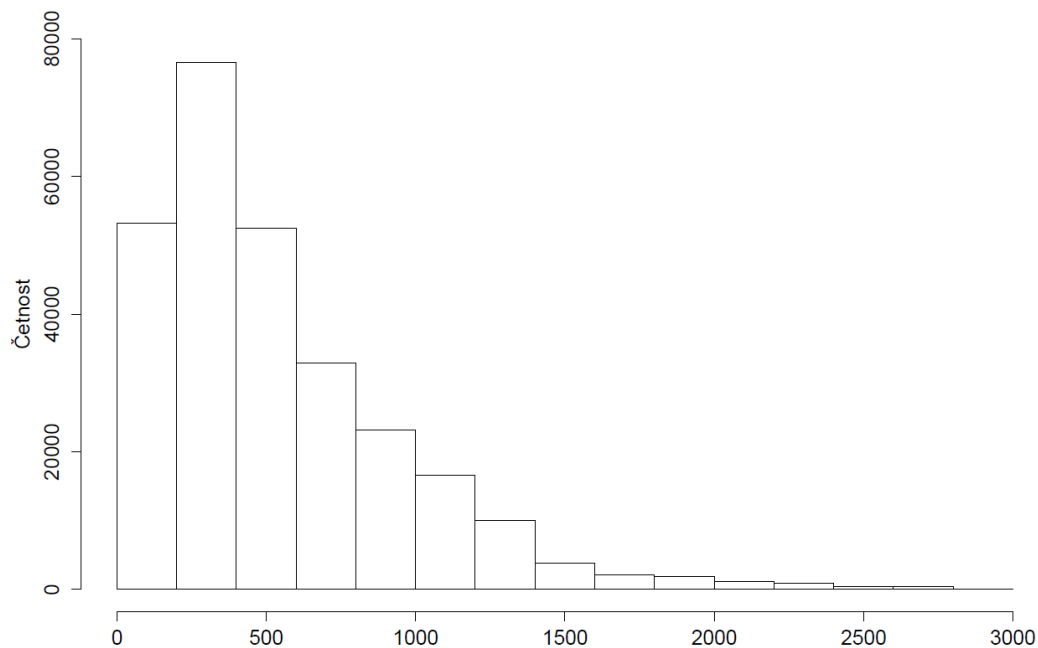
6.3.1 Metrika

Metrika používaná nástrojem *SimRoad* je velice blízká metrice používané nástrojem *SimUrb*. Obecný popis je dostupný v kapitole 5.3.1.

6.3.2 Určení míry podobnosti

Prvotní určení míry podobnosti v nástroji *SimRoad* probíhá obdobně jako v nástroji *SimUrb*. Matice *Modified matrix* a *Original matrix* jsou určeny stejným výpočtem, který je blíže popsán v kapitole 5.3.2. Avšak jelikož očekávané rozdělení hodnot míry podobnosti pro všechny dvojice sekvencí, které nástroj zpracovává, má vysokou šikmost (1,569 – obr. 28), je parametr p , jenž je zadáván uživatelem, stanoven jako hodnota kvantilové funkce.

Rozdělení hodnot original matrix



Obr. 28 Histogram rozdělení četnosti hodnot „Original matrix“

Nechť $F(x)$ je distribuční funkce náhodné veličiny X . Funkce

$$F^{-1}(u) = \inf\{x \in \mathbb{R}; F(x) \geq u\}, \quad 0 < u < 1,$$

se nazývá kvantilová funkce náhodné veličiny X . Pro $0 < p < 1$ se hodnota $F^{-1}(p)$ nazývá p -kvantil. (Kunderová, 1997)

Nadále tedy pro matici sousednosti $A = (a_{ij})_{i,j=1}^n$ (Adjacency matrix) platí

$$a_{ij} = \begin{cases} 1, & m_{ij} \geq p \\ 0, & m_{ij} < p \end{cases}$$

6.3.3 Vizualizace výsledků

Stejně jako u nástroje *SimUrb* by vytváření nedisjunktních skupin vzájemně si podobných objektů nástrojem *SimRoad* bylo nepřehledné a bezúčelné. Skupiny vzájemně si podobných objektů jsou tedy opět vyhledávány jako maximální kliky v grafu sestrojeném podle matice sousednosti, kdy v každém kroku je vybrána největší klika a její vrcholy jsou vyřazeny z dalšího výpočtu, viz kapitola 6.3.

6.4 Použití SimRoad (uživatel)

6.4.1 Import dat

Nástroj *SimRoad* byl vytvořen pro porovnání podobnosti dálničních úseků na základě dat o počtu průjezdů z mýtných bran. Vrstva dálničních úseků a hranic České republiky je v nástroji pevně nastavena. Uživatel může nahrát vlastní tabulku ve formátu CSV, která obsahuje tři povinné sloupce – *usek*, *dalnice* a *popis*. Pole *usek* obsahuje jedinečný identifikátor úseku vymezeného dvěma mýtnými bránami. V poli *dalnice* je uvedeno, na které dálnici se daný úsek nachází, a v poli *popis* je pro lepší orientaci uživatele úsek popsán slovně. V tabulce následuje libovolný počet sloupců obsahující počty průjezdů

v daném úseku dálnice. Každý sloupec reprezentuje data pro jeden den. Část vzorové tabulky je znázorněna v tabulce 4.

Tab. 4 Vzorová struktura souboru CSV pro import do nástroje *SimRoad*.

usek	dálnice	popis	04.05.2015	05.05.2015	06.05.2015	07.05.2015	08.05.2015	09.05.2015	10.05.2015	11.05.2015
D01011	D01	Praha-Spořilov – Praha-Chodov	372	333	341	419	301	306	417	366
D01012	D01	Praha-Chodov – Praha-Spořilov	357	357	340	416	344	284	353	396
D01021	D01	Praha-Chodov – Průhonice	553	508	517	597	370	365	482	552
D01022	D01	Průhonice – Praha-Chodov	542	534	526	600	422	351	439	567
D01031	D01	Průhonice – Modletice	471	435	447	521	322	321	432	476

6.4.2 Použití *SimRoad*

Nástroj *SimRoad* umožňuje analyzovat najednou pouze jeden soubor s daty.

Při prvním spuštění webového nástroje *SimRoad* se stejně jako v nástroji *SimUrb* zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadované parametry.

Nahrání souboru s daty provede uživatel volbou tlačítka *Choose file* v poli *Attribute data*. Metoda výpočtu míry podobnosti je v nástroji *SimRoad* pouze jediná a tedy se volba uživateli nenabízí. Metoda je blíže popsána v kapitolách 5.3 a 6.3.

Dalším krokem je nastavení minimálního počtu prvků ve skupině a hodnota míry podobnosti. Míra podobnosti p může nabývat hodnot z intervalu $\langle 0,1 \rangle$, přičemž hodnota rovna jedné značí zcela totožné sekvence a hodnota rovna nule značí sekvence s největší možnou odlišností. Podrobnější popis je v kapitole 5.3.1 a 6.3.2.

Poté se zobrazí v pravé části seznam všech nalezených klik. Kliknutím na kteroukoliv skupinu se tato skupina barevně zvýrazní v mapě. Vzhledem k rozsáhlé délce sekvencí, jenž určují jednotlivé objekty analyzované nástrojem *SimRoad*, nejsou tyto v základním prostředí nástroje zobrazovány. Uživatel si je však může stáhnout a zobrazit díky CSV souboru.

V levé horní části je uvedena délka trvání výpočtu, název zdrojových dat, zvolená míra podobnosti a minimální velikost skupiny. K dispozici jsou stejně jako v nástroji *ScanGraph* možnosti stáhnout CSV soubor s analyzovanými daty a vypočtenými hodnotami v *Original*, *Modified* a *Adjacency matici* a možnost permanentního odkazu.

Skupiny, jež jsou výstupem nástroje *SimRoad*, jsou vizualizovány v mapě. Jelikož se jedná o dálnice, silnice I. třídy a rychlostní silnice, jsou vizualizovány liniovými znaky a barevně rozlišeny podle příslušných kategorií.

6.4.3 Interpretace výsledků

Jednotlivé silniční úseky mohou být opět reprezentovány vrcholy grafu. Sestrojení grafu je dáno maticí sousednosti dle daného parametru p . Skupiny vzájemně si podobných silničních úseků pak tvoří kliky v tomto grafu. Stejně však jako u nástroje *SimUrb* by určování všech klik v takovémto množství dat postrádalo přehlednost. Z tohoto důvodu jsou skupiny vzájemně si podobných silničních úseků vytvářeny následujícím způsobem:

1. Nástroj nalezne největší kliku grafu (maximum clique problem). V případě více největších klik zvolí první nalezenou.
2. Silniční úseky patřící do nalezené kliky vyřadí z prohledávané množiny silničních úseků.
3. Opakuje kroky 1 a 2, dokud množina úseků není prázdná.

Výsledkem jsou pak skupiny vzájemně si podobných silničních úseků při dané míře podobnosti.

7 DC4 – PŘÍPADOVÉ STUDIE VYUŽÍVAJÍCÍ VYTVOŘENÉ NÁSTROJE

Všechny nástroje a metody, které tyto nástroje využívají, byly navrženy tak, aby byly použitelné v praxi. Pro uživatele jsou k dispozici informační ikony s nápovědou, k čemu jednotlivé prvky nástrojů slouží.

V rámci disertační práce byly vypracovány tři případové studie, na nichž byla demonstrována funkcionality jednotlivých nástrojů a ověřena správnost navržených metod. Tyto studie jsou podrobně popsány v kapitolách 7.1.1, 7.2.1 a 7.3.1.

Analýza zaznamenaných eye-tracking dat může být zpracována nástrojem *ScanGraph* v jakémkoliv případě, kdy je vhodné porovnávat rozdílné skupiny respondentů. Ačkoliv je *ScanGraph* poměrně novým nástrojem, byl již použit pro několik studií. Studie odlišných strategií prohlížení různých mapových kompozic mezi kartografy a nekartografy byla vypracována v rámci výuky na Katedře geoinformatiky UP. Rozdílné prvky mezi strategiemi mužů a žen při hledání určeného symbolu v mapě byly studovány v (Dolezalova a Popelka, 2016). Snopková (2016) analyzovala odlišnosti ve čtení map mezi respondenty s normálním viděním a respondenty s barvoslepostí. *ScanGraph* tedy může být použit i v jiných oblastech výzkumu, nikoliv pouze v kartografii. Pulkrtová (2016) použila nástroj ve své psychologické práci zabývající se rozdílnou percepcí červené barvy mezi muži a ženami. Hájková (2016) využila nástroj *ScanGraph* při studiu na ústavu fyzioterapie Fakulty zdravotnických věd UP s pacienty po mozkové příhodě. Ve své práci porovnávala kontrolní skupinu se dvěma skupinami pacientů po rozdílném typu mozkové příhody.

Nástroj *SimUrb* byl navržen pro vyhledávání podobných skupin libovolných objektů, jež je možné charakterizovat pomocí uspořádané n -tice kvantitativních koeficientů. Jako takový lze tedy využít nejen pro úlohy prostorového plánování. Obdobně jako nástroj *ScanGraph* byl tedy již využit v dalších studiích, které jsou uvedeny dále v kapitolách 7.2.2 a 7.2.3 a další možná využití jsou aktuálně připravována.

Nástroj *SimRoad*, jakožto nejmladší z těchto nástrojů, byl prozatím využit pouze v jedné případové studii, již je věnována kapitola 7.3.1.

7.1 Případové studie *Scanpath Comparison*

7.1.1 Vizualizace nejistoty geografických dat

Případová studie z oblasti eye-trackingu navazovala na diplomovou práci Kučery (2016) a byla zaměřena na nejistotu vizualizace v mapách. Nejistota je považována za nejasnost, náhodnost podmínek nebo výsledek určitých procesů a jevů. Koncept nejistoty je také poměrně často používán k popisu malé jistoty o konkrétním jevu v mapách (Brus a kol., 2013). Mnoho přístupů a metod vizualizace nejistoty bylo vyvinuto na příklad na Bertinově teorii grafických proměnných a kombinování statických i dynamických prvků vizualizace (Kubiček, Šašinka a Stachoň, 2012).

Data pro tuto případovou studii byla použita z diplomové práce Kučery. V této závěrečné práci byl vytvořen soubor kartografických znaků pro vizualizaci nejistoty pomocí bodových, liniových i plošných znaků. Bodové znaky byly vytvořeny podle studie MacEachrena a kol. (2012). Tyto znaky byly umístěny do map, jež byly použity jako stimuly pro online dotazníkové šetření a následný eye-tracking experiment. Cílem diplomové práce bylo zhodnotit, které vizualizace jsou nejvíce srozumitelné pro čtenáře mapy. Experiment zahrnoval 40 respondentů. Dvacet z nich byli studenti kartografie

a dvacet respondenti bez kartografického vzdělání. Eye-tracking metriky (*Trial Duration*, *Scanpath Length*) a přesnost odpovědí byly použity při analýze všech stimulů. Celkem bylo v experimentu použito 27 map s bodovými znaky. Třináct z nich zobrazovalo jeden jev, osm představovalo kombinaci více jevů. Šest map zobrazovalo časoprostorovou nejistotu.

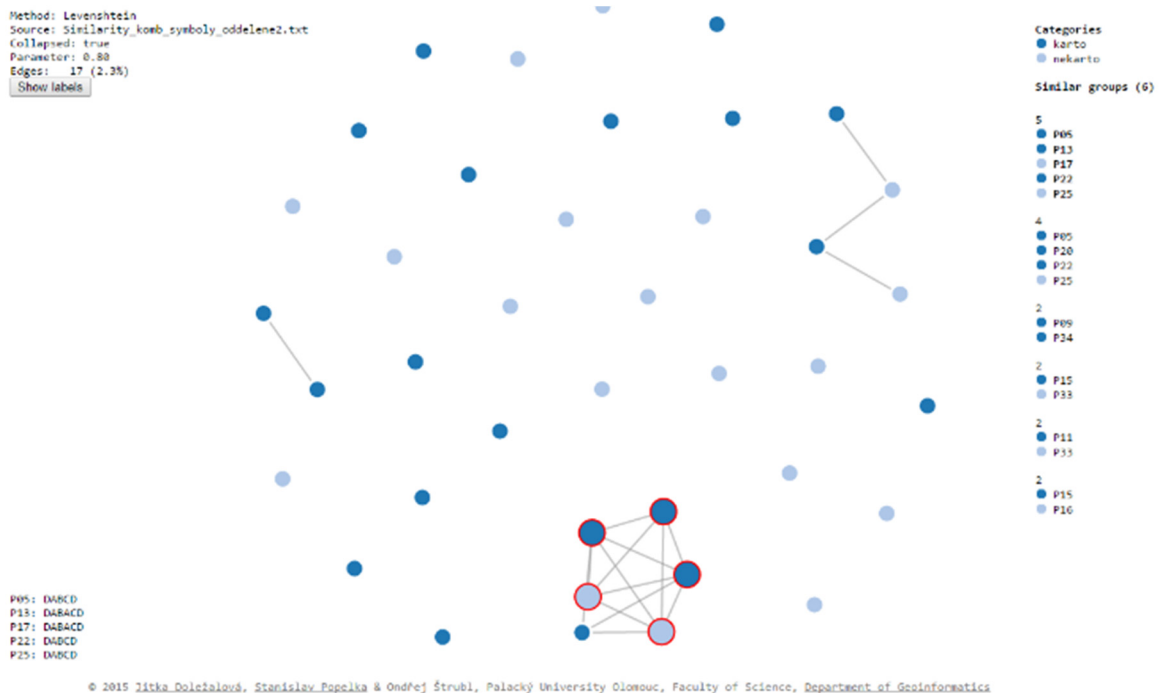
Nejprve byly analyzovány strategie prohlížení stimulu BK07. V tomto případě 16 bodových znaků reprezentovalo možný výskyt tří živočišných druhů (prase divoké, zajíc polní a liška obecná) s rozdílným stupněm nejistoty. Úkolem respondentů bylo určit lokalitu s největší pravděpodobností výskytu každého z druhů. Legenda se všemi třemi druhy byla umístěna na pravé straně stimulu. Levá část stimulu pak obsahovala ortofotomapsu s bodovými znaky. Oblasti zájmu byly vyznačeny okolo mapového pole a okolo každé části legendy (obr. 29).



Obr. 29 Oblasti zájmu vyznačené ve stimulu BK07

Trajektorie pohybu očí byly konvertovány do sekvencí znaků podle pozice fixací ve vyznačených oblastech zájmu v prostředí OGAMA. Ze scanpath (v levé části obrázku) byly vygenerovány řetězce fixací. Pro analýzu byla zvolena varianta collapsed sekvencí (v pravé části obrázku) a dále pro výpočet podobností mezi sekvencemi byla zvolena modifikovaná metoda Levenshteinovy vzdálenosti.

Míra podobnosti byla nastavena na hodnotu $p = 0,8$. Ve výsledném grafu bylo nalezeno šest netriviálních klik (obr. 30). Největší klika zahrnovala pět sekvencí. Tři z nich patřily kartografům a dvě nekartografům.



Obr. 30 Výstup z nástroje ScanGraph pro stimul BK07 a míru podobnosti $p = 0,8$

Strategie těchto respondentů může být charakterizována jako ideální. Všichni začali prohlížet stimul od středu obrazovky (oblast zájmu D). Poté přesunuli pohled do všech částí legendy (oblasti zájmu A, B a C) a nakonec se vrátili zpět do mapového pole, kde vyhledávali správnou odpověď. U respondentů P13 a P17 se navíc vyskytla fixace v oblasti zájmu A po prohlížení oblasti zájmu B.

Ve čtyřprvkové klice byla situace velice podobná. V tomto případě všechny sekvence měly pořadí navštívených oblastí zájmu „DABCD“. Jedinou výjimkou byl respondent P20, u kterého byla navíc nalezena fixace v oblasti zájmu B na konci prohlížení stimulu.

Zbylé vizualizované netriviální kliky obsahovaly pouze dva respondenty. Všichni tito respondenti při svém prohlížení stimulu vynechali oblast zájmu C. Oblast zájmu C byla vyznačena okolo poslední části legendy (reprezentující možný výskyt lišky obecné). Jelikož všechny části legendy mají stejnou vypovídající hodnotu, nepotřebovali respondenti prohlížet její poslední část.

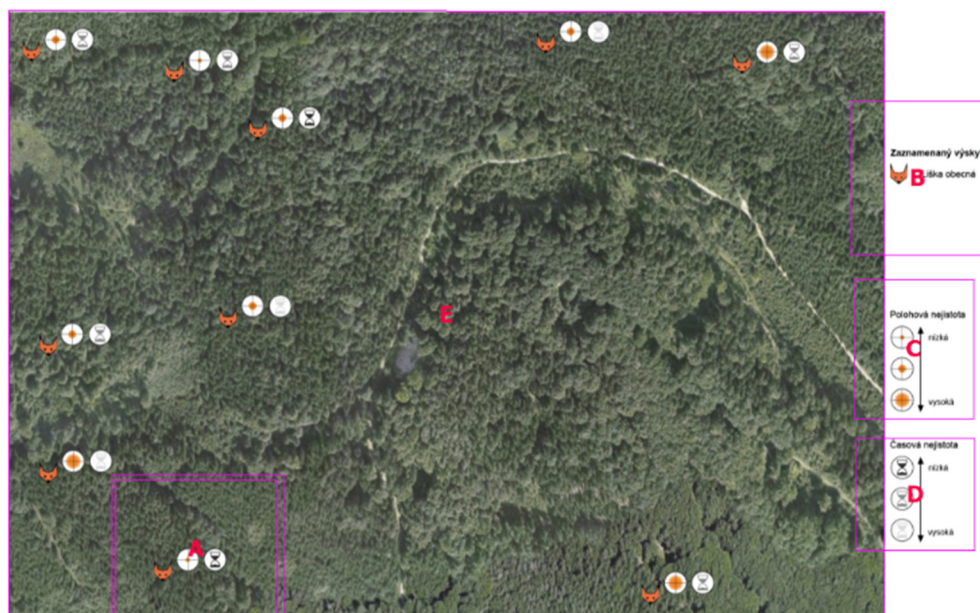
Zbylé vrcholy grafu, tedy zbylé sekvence, tvořily jednoprvkové kliky neboli izolované vrcholy. Což znamená, že jejich sekvence navštívených oblastí zájmu nebyla podobná žádné jiné sekvenci (při dané míře podobnosti $p = 0,8$). Příkladem takovéto sekvence může být respondent P29 se svou sekvencí „DABADADADADABDBCBDAD“, respondent P30 s velice krátkou sekvencí „DAD“ nebo respondent P43, jehož fixace byly všechny v mapovém poli, tedy v oblasti zájmu D. Všichni tito tři respondenti patřili do kategorie nekartografů.

Za použití nástroje ScanGraph jsme byli schopni rychle nalézt skupinu respondentů, kteří prohlíželi daný stimul podobným způsobem. Po vyšetření jejich sekvencí bylo zjištěno, že tato sekvence se dá klasifikovat jako „ideální“.

Druhou analyzovanou mapou z experimentu byl stimulus s označením C03. Tato mapa zobrazovala možný výskyt lišky obecné. Na rozdíl od předchozího stimulu, časoprostorová nejistota byla zobrazena dvěma různými znaky. Úkolem respondentů bylo nalézt místo, kde je největší pravděpodobnost výskytu (polohová i časová) lišky obecné. Oblasti zájmu byly opět vyznačeny (obr. 31).

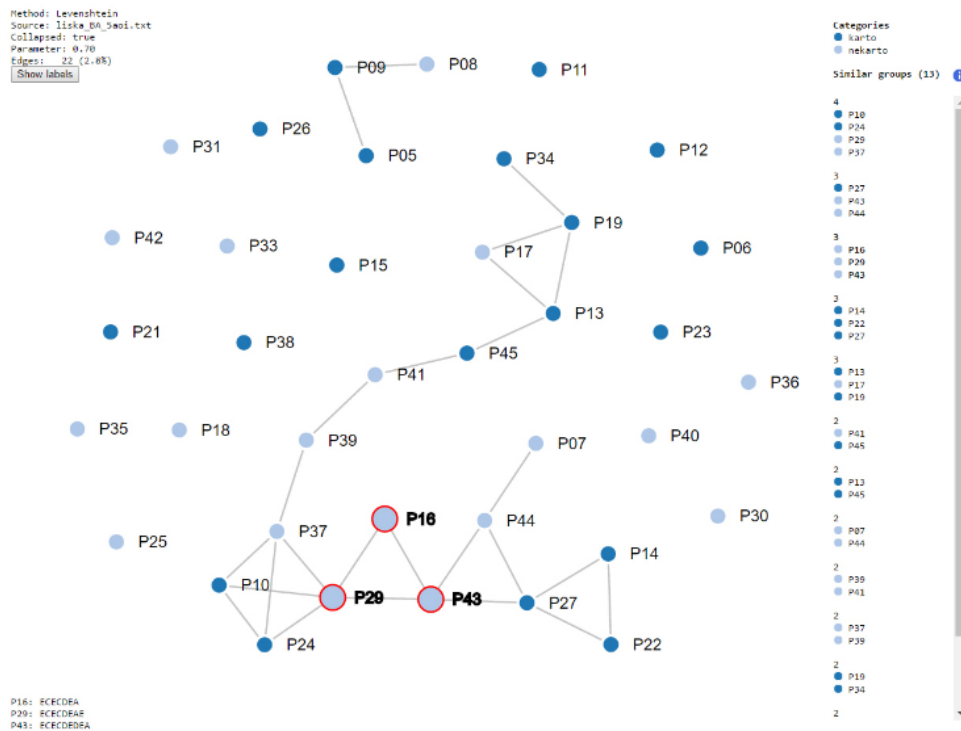
Správná odpověď se nacházela v oblasti zájmu A. Oblast zájmu B byla vyznačena okolo znaku pro lišku obecnou a zahrnovala také název mapy (Zaznamenaný výskyt lišky

obecné). Další dvě oblasti zájmu byly vyznačeny okolo dvou částí legendy (polohová nejistota AOI C a časová nejistota AOI D). Poslední oblast zájmu (AOI E) byla vyznačena okolo mapového pole (kromě lokace správné odpovědi).



Obr. 31 Oblasti zájmu vyznačené ve stimulu C03

Pro nástroj *ScanGraph* bylo použito stejné nastavení jako v předchozím případě, pouze míra podobnosti byla nastavena na $p = 0,7$. Tento stimul obsahoval více oblastí zájmu, tedy byla očekávána menší míra podobnosti mezi sekvencemi. Výsledný graf je zobrazen na obrázku 32. Celkem bylo nalezeno 13 netriviálních klik. Zajímavé je srovnání mezi dvěma tříprvkovými klikami. Prvních z nich je klika respondentů P14, P22 a P27. Všichni tito respondenti patří do kategorie kartografů. Druhou zajímavou klikou je klika tvořená třemi respondenty z kategorie nekartografů. Rozdíl mezi těmito dvěma klikami je ten, že žádný z nekartografů neprohlížel oblast zájmu B vyznačenou okolo názvu mapy. Studenti kartografie vědí, že názvu mapy musí také věnovat pozornost.



Obr. 32 Výstup nástroje ScanGraph pro stimul C03 a míru podobnosti $p = 0,7$

7.1.2 Hodnocení nefotorealistické 3D vizualizace měst

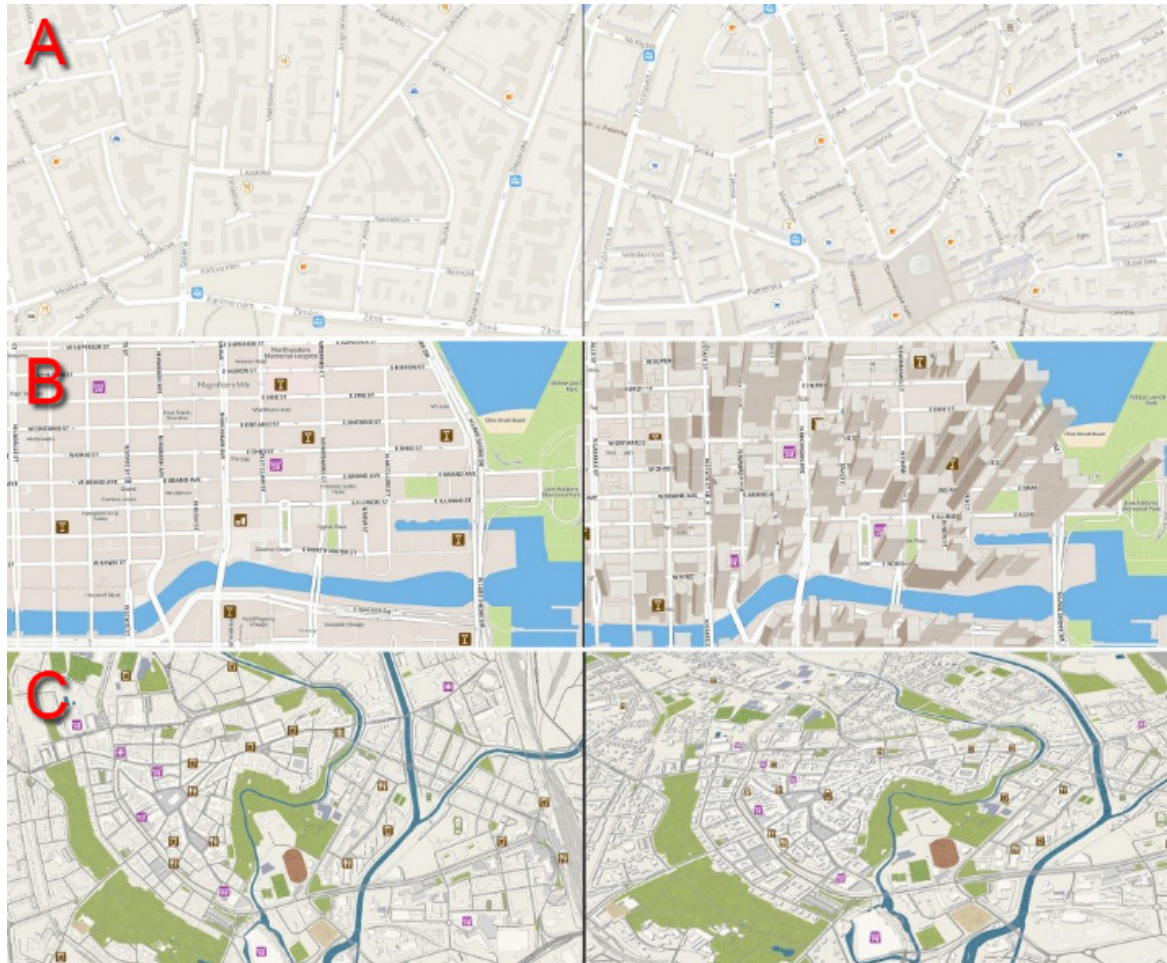
Pro hodnocení nefotorealistické 3D vizualizace byl vytvořen eye-tracking experiment s cílem zjistit, zda 3D vizualizace ovlivňuje uživatelskou percepci při používání městských map při hledání konkrétního bodu v mapě.

Experimentu se zúčastnilo celkem 40 respondentů (24 žen a 16 mužů). Mezi nimi byli kartografové i nekartografové. Většina respondentů byla ve věku 20–25 let.

Cílem testování bylo ověřit, zda hledání určitého cíle na mapě bude těžší či snazší na stimulu s 3D variantou mapy.

Experiment obsahoval screenshoty různých internetových mapových portálů (obr. 33). Mapy byly doplněny bodovými znaky a byly použity dva typy map. První z nich byla standardní mapa s budovami reprezentovanými polygony, druhá obsahovala 3D vizualizace těchto budov.

Data byla zaznamenávána se vzorkovací frekvencí 120 Hz. Pro statistickou analýzu naměřených dat bylo použito několik metod – *Time to Answer* (klik myši), *Fixation Count*, *Fixation Duration Median* a *Scanpath Length*.



Obr. 33 Dvojice 2D (vlevo) a 3D (vpravo) stimulů. A: Google Maps (source: <http://maps.google.com>), B: OSMbuildings (source: <http://osmbuildings.org>), C: F4map (source: <http://map.f4-group.com>)

Vyšší hodnoty pro *Fixation Count* ve 2D variantě byly pozorovány pro tři páry map. U zbývajících map byly vyšší hodnoty pozorovány u 3D varianty. Podobná situace nastala pro všechny ostatní použité metody.

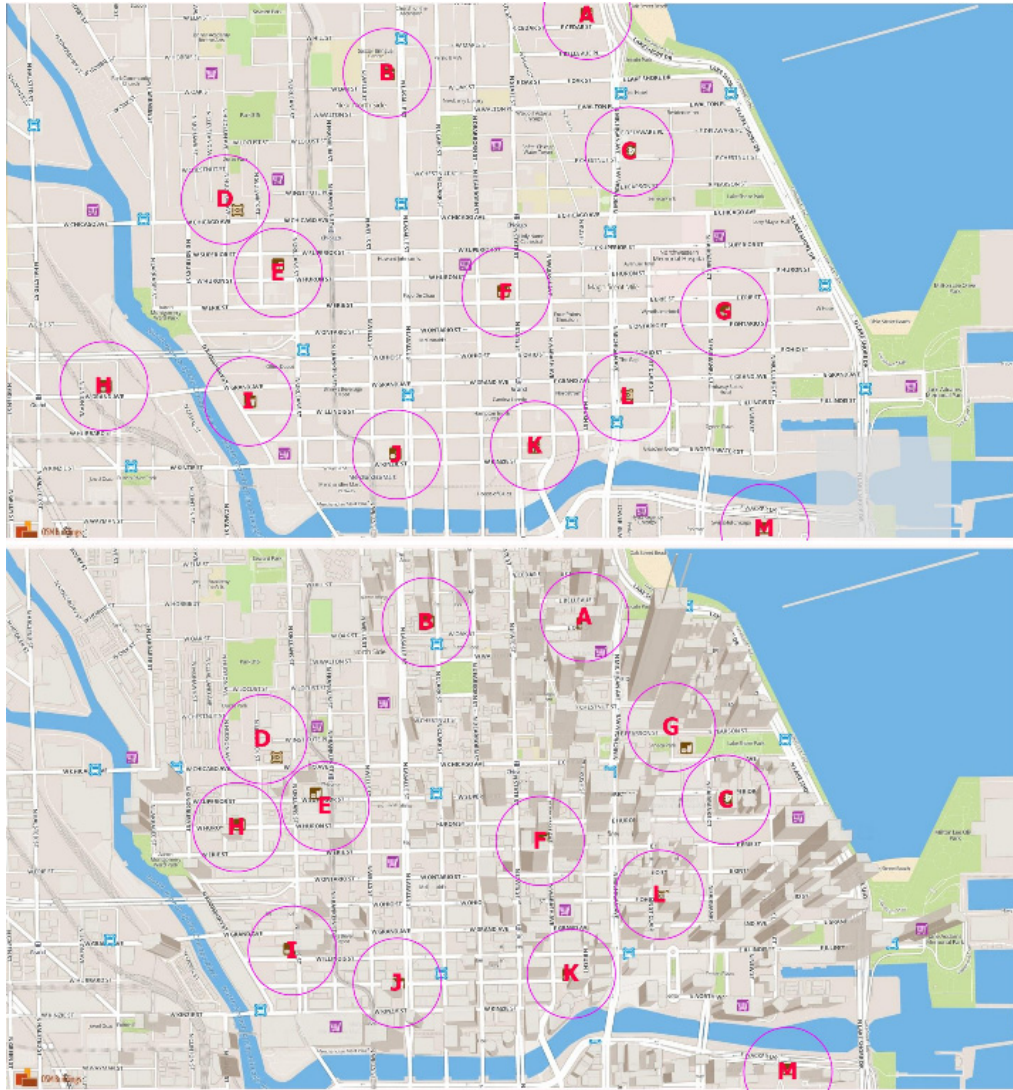
Poté bylo provedeno srovnání rozdílů 2D a 3D variant pro všechny stimuly. Statisticky významného rozdílu bylo dosaženo pro metody *Time to Answer* a *Fixation Duration*. Vyšší hodnoty u obou metod byly pozorovány u 3D variant map. Největší rozdíl mezi 2D a 3D variantou mapy byl pozorován pro pár stimulu číslo 9.

Podrobnější informace ke statistické analýze a vizuální analýze za použití *FlowMap* metody (Adrienko a kol., 2012) jsou popsány v Popelka a Doležalová (2015).

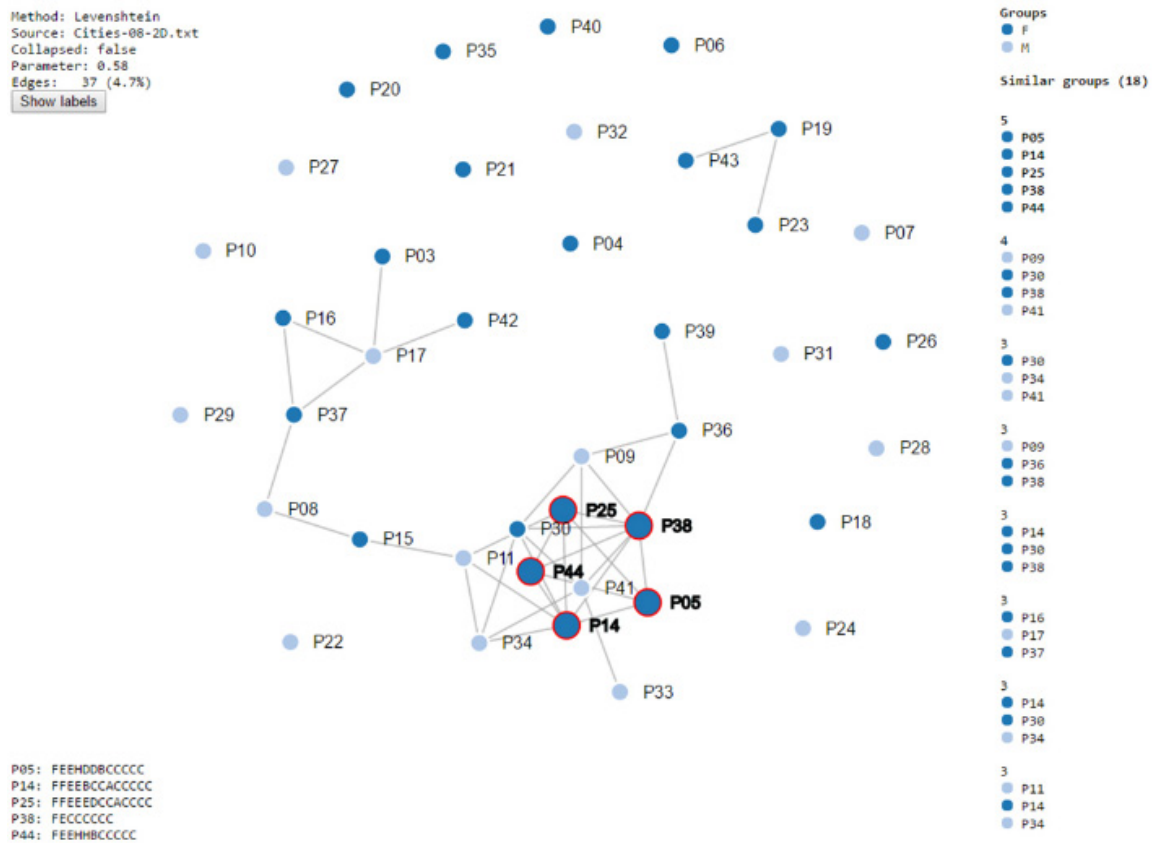
Pro analýzu těchto stimulů byl použit i nástroj *ScanGraph*. Jeho aplikace je ilustrována na jednom páru stimulů.

Oblasti zájmu byly vyznačeny okolo všech přidanych bodových znaků. Správná odpověď, v tomto případě, se nacházela v oblasti zájmu C (obr. 34).

Pro analýzu pomocí nástroje *ScanGraph* bylo zvoleno nastavení Levenshteinovy vzdálenosti a non-collapsed varianty. Pro 2D variantu byl sestaven graf s 5% hustotou hran a tedy obsahoval 37 hran. Míra podobnosti v tomto případě byla rovna $p = 0,58$ a bylo nalezeno 18 netriviálních klik. Největší z nich obsahovala 5 respondentů (sekvenci). Těmito respondenty byly ženy. Jejich trajektorie započaly v oblasti zájmu F a pokračovaly do oblasti zájmu E. Poté respondentky navštívily různé oblasti zájmu, ale všechny skončily na správné odpovědi, tedy oblasti zájmu C, a u všech zde bylo zaznamenáno 5 fixací.

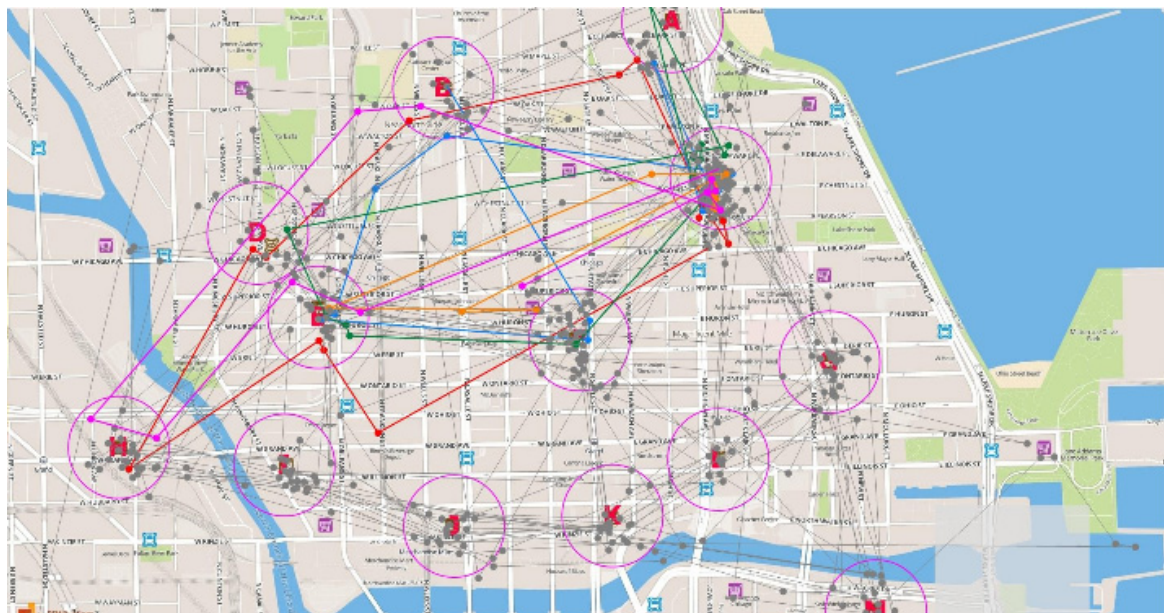


Obr. 34 Oblasti zájmu vyznačené okolo doplňujících bodů ve 2D (nahore) a 3D (dole) variantě stimulu 08



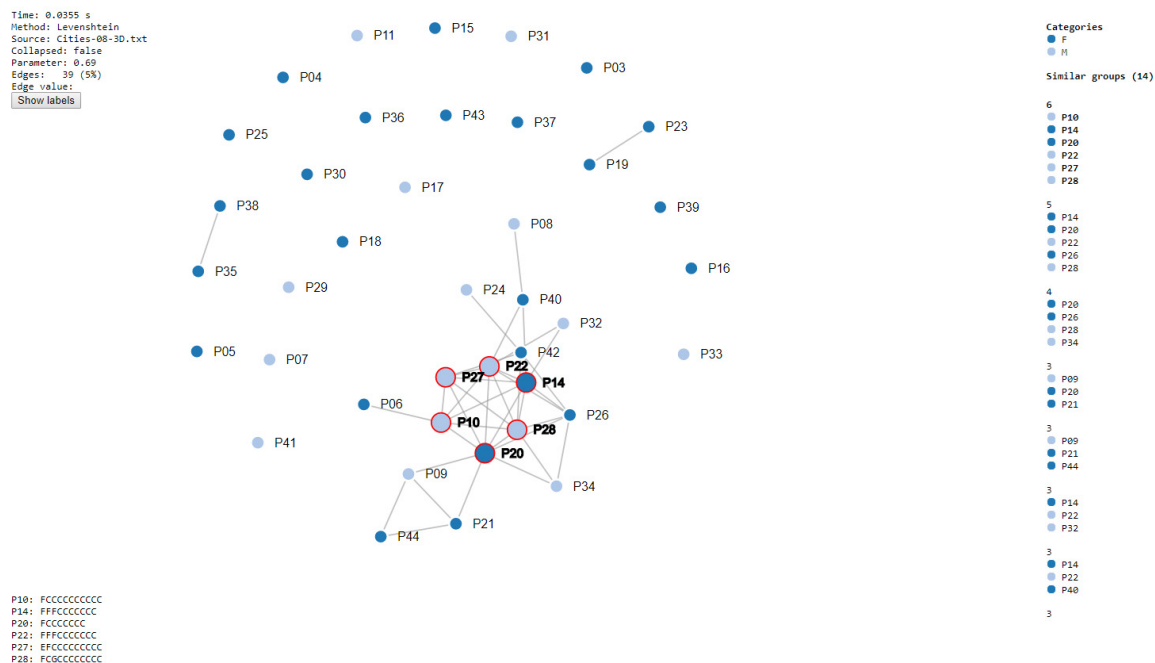
Obr. 35 Výstup nástroje ScanGraph pro 2D variantu stimulu

Trajektorie pěti z respondentů z největší kliky byly zobrazeny v prostředí OGAMA (obr. 35). Je zřejmé, že tyto respondenti prohlíželi pouze malou část stimulu, většinou okolo oblastí zájmu F, E a C. Odlišnost od zbytku respondentů (znázornění šedou barvou) je viditelná.



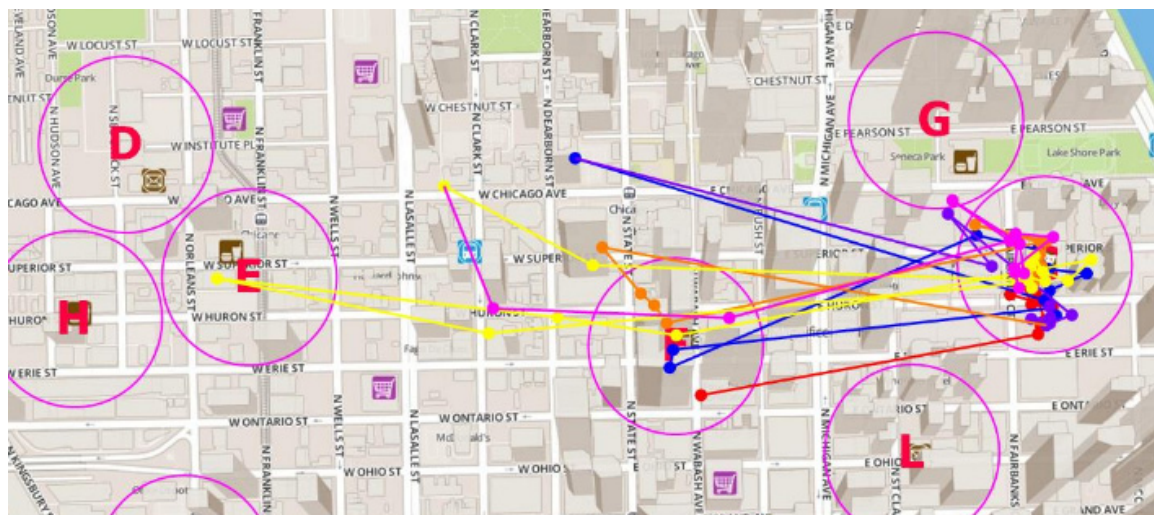
Obr. 36 Trajektorie pěti respondentů tvořících největší kliku (barevné trajektorie) a zbytek respondentů (šedé trajektorie)

Zopakování stejného procesu pro 3D variantu stimulu přineslo graf se 38 hranami a mírou podobnosti $p = 0,69$. V případě 3D varianty byla tedy podobnost mezi respondenty větší. Největší nalezená klika obsahovala šest respondentů, z toho čtyři muži a dvě ženy (obr. 37).



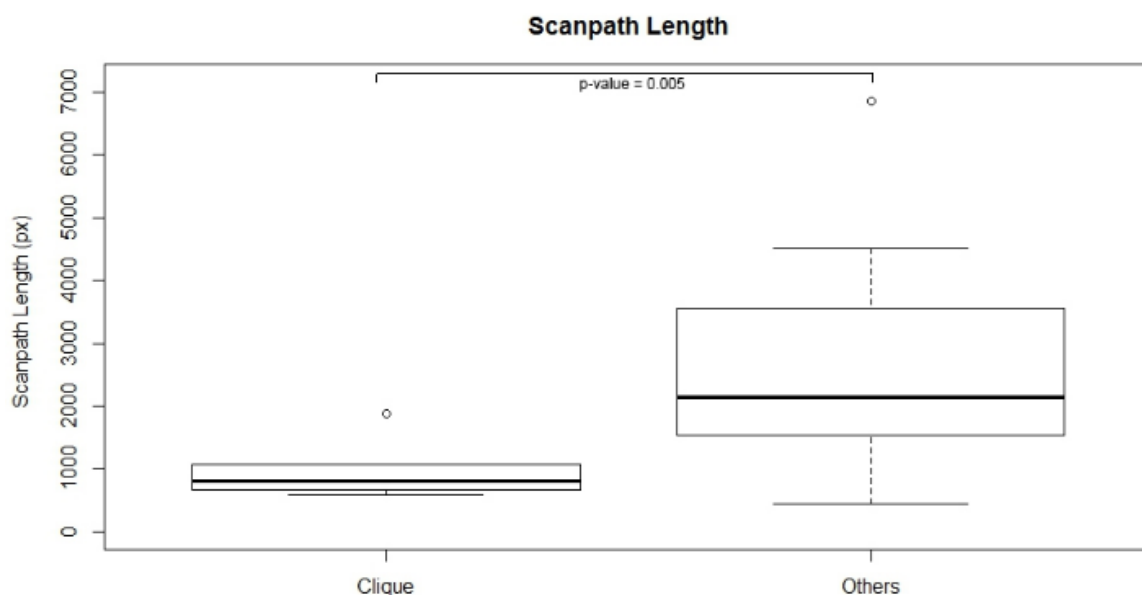
Obr. 37 Výstup nástroje ScanGraph pro 3D variantu stimulu

Většina sekvencí z této kliky začínala opět v oblasti zájmu F, ale po té pokračovala přímo do oblasti zájmu C, kde se nacházela správná odpověď. Obrázek 38 představuje výřez z 3D varianty zobrazující tuto situaci.



Obr. 38 Trajektorie šesti respondentů tvořících největší kliku

Bylo provedeno statistické hodnocení rozdílu skupiny těchto šesti respondentů a zbytku respondentů. Statisticky významný rozdíl byl nalezen u eye-tracking metody *ScanPath Length*. Hodnota *ScanPath Length* byla u skupiny šesti respondentů nalezené nástrojem *ScanGraph* znatelně nižší (obr. 39). Podle Goldberga a kol. (2002), kratší scanpath naznačuje účinnější prohledávání.

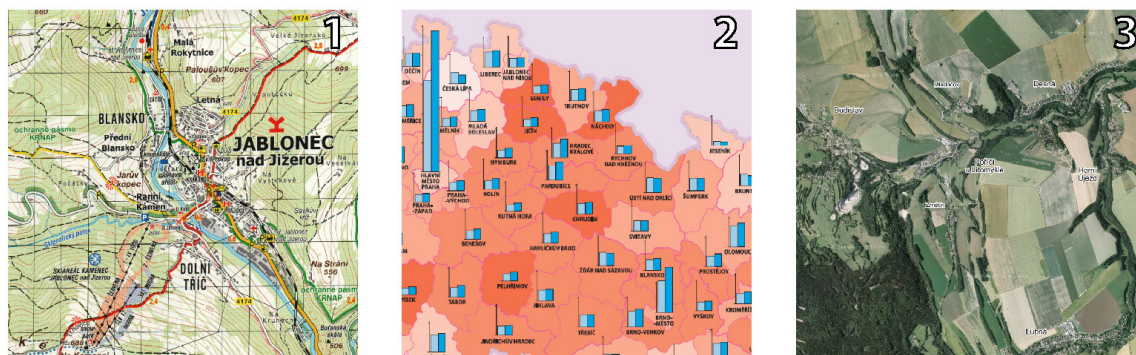


Obr. 39 Boxploty zobrazující rozdíl ve Scanpath Length metrice mezi skupinou šesti respondentů označených nástrojem ScanGraph jako vzájemně si podobní a zbytkem respondentů

7.1.3 Hodnocení kartografické gramotnosti

Případová studie (Beitlová, 2017) se zaměřila na hodnocení kartografické gramotnosti 55 respondentů z vybraných uživatelských skupin: kartografové (17), právníci (11) a administrativní pracovníci (12). Eye-tracking experiment obsahoval 30 stimulů zobrazujících nejpopulárnější typy map (turistické mapy, automapy, ortofotomapy, kartogramy a kartodiagramy). Ke všem stimulům byly přiřazeny úkoly (identifikace, výběr, odhad vzdálenosti). Data byla zaznamenávána pomocí low-cost zařízení EyeTribe s frekvencí 60 Hz. Více informací o této případové studii je dostupných v magisterské práci Beitlové (2017).

Cílem experimentu bylo zjistit stupeň kartografické gramotnosti u různých uživatelů map. Odlišnost mezi skupinami byla zjišťována za použití tří eye-tracking metod. Podobnost mezi čtením map jednotlivých respondentů byla zkoumána za pomoci nástroje *ScanGraph*. Experiment je ilustrován na stimulu obsahujícím výřezy tří různých druhů map (Obr. 40). Stimul je pokryt trajektoriemi všech respondentů. Barva trajektorie reprezentuje příslušnost respondentů k jednotlivým skupinám. Je zřejmé, že nalezení podobností pouhým vizuálním porovnáním trajektorií by bylo poměrně obtížné.

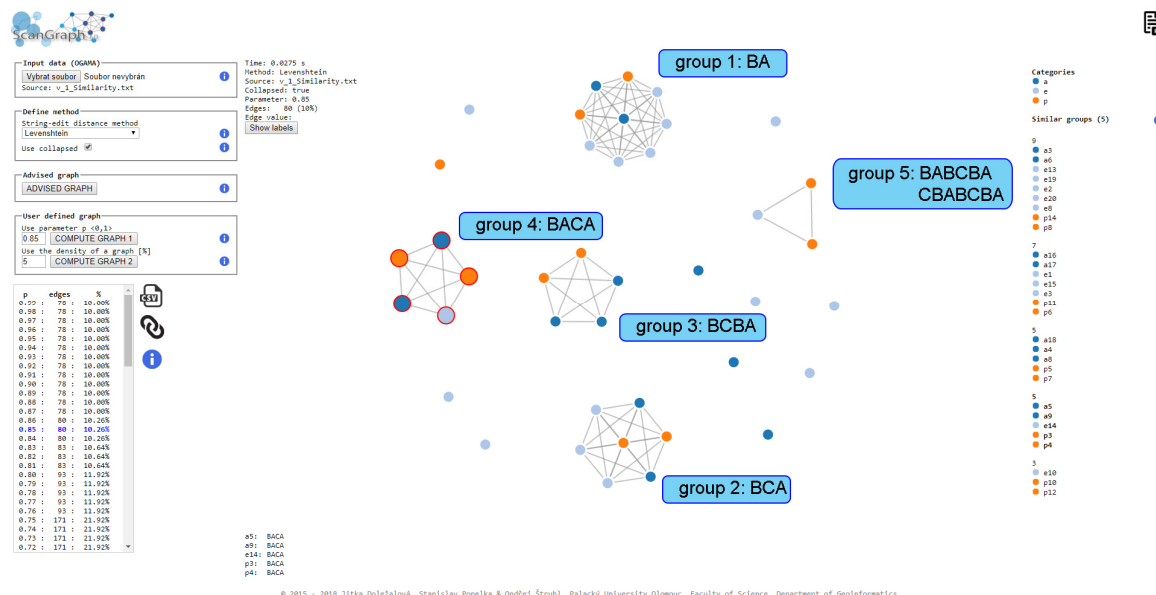


Obr. 40 Ukázka jednoho ze stimulů ze studie (Beitlová, 2017)

Respondenti v tomto případě měli rozhodnout, která mapa bude nejlepší pro plánování turistického výletu. Správnou odpovědí byla mapa A (vlevo). Zadání úkolu bylo

jednoznačné a výběr turistické mapy očekávaný u všech respondentů. Avšak strategie výběru mapy se mohla napříč uživateli, potažmo kategoriemi lišit.

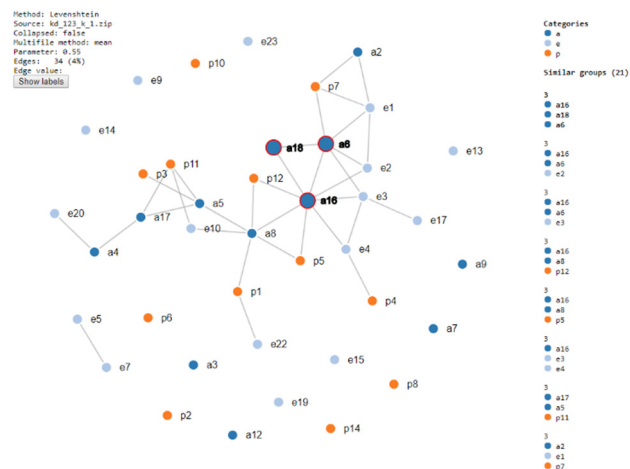
Pro určení odlišných strategií při prohlížení stimulu byl použit nástroj *ScanGraph*. Byla nastavena míra podobnosti $p = 0.85$ a varianta collapsed, jak je zobrazeno na obrázku 41.



Obr. 41 Výstup nástroje *ScanGraph* při míře podobnosti $p = 0,75$, collapsed varianta. Barva vrcholů reprezentuje příslušnost ke kategoriím.

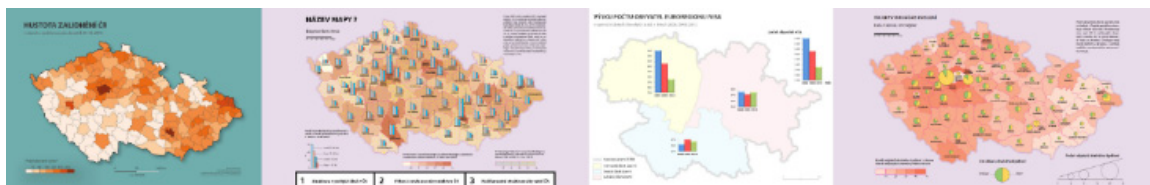
Celkem bylo nalezeno pět netriviálních skupin. Na obrázku 41 jsou sekvence navštívených oblastí zájmu (a tedy prohlížených map) zobrazeny v modrých obdélníčkách. Je zřejmé, že většina respondentů započala prohlížení v oblasti zájmu B, která byla vymezena okolo mapy vprostřed.

Světle modré vrcholy představovaly respondenty kartografy. Většina kartografů (5) je obsažena ve skupině 1. Všichni tyto respondenti zaměřili svůj pohled ihned po oblasti zájmu B do oblasti zájmu A, kde se nacházela správná odpověď. To může naznačovat, že mohli použít periferní vidění ke zjištění, že ortofotomapa na pravé straně není správnou odpovědí.



Obr. 42 Multifile analýza pro čtyři stimuly při míře podobnosti $p = 0,55$, metoda mean

Při experimentu byla využita i funkce pro analýzu multifile souboru. Pro analýzu byly vybrány čtyři stimuly obsahující kartogramy a kartodiagramy (obr. 43). Protože je vyšetřována podobnost napříč několika stimuly, je očekávána nižší míra podobnosti. Byla zvolena hodnota $p = 0,55$ a vybrána metoda Mean. Nástroj našel 21 netriviálních kliků o nanejvýše třech vrcholech.

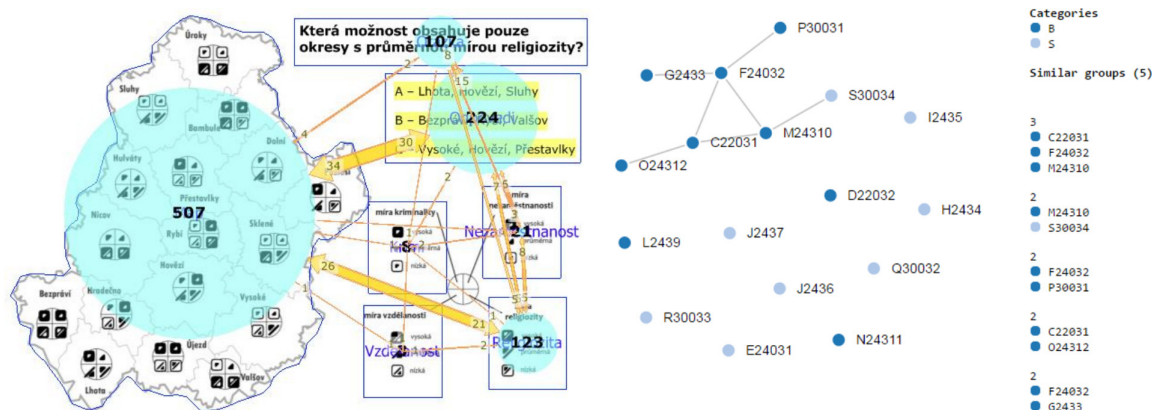


Obr. 43 Ukázka čtyř stimulů použitých pro multifile analýzu.

7.1.4 Další studie využívající ScanGraph

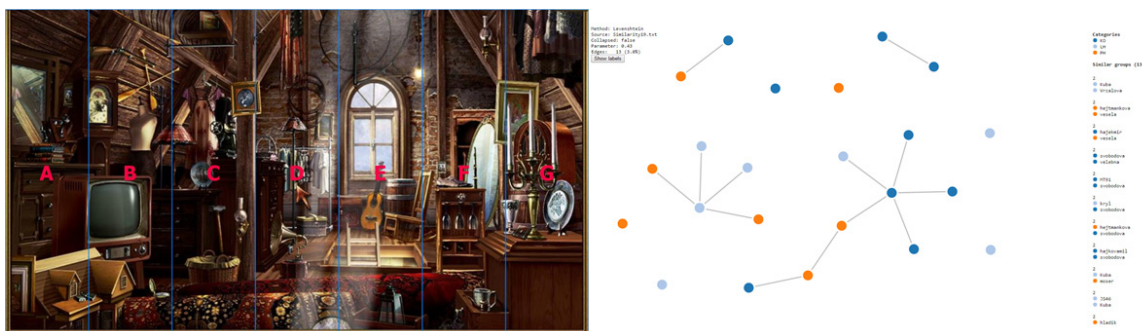
Kromě výše zmíněných studií proběhlých na Katedře geoinformatiky Univerzity Palackého v Olomouci byl nástroj ScanGraph využit i ve studiích řešených na jiných pracovištích.

V oblasti kartografie byl ScanGraph využit na Geografickém ústavu Masarykovy univerzity v Brně, kdy Snopková (2016) hodnotila čtení map respondenty se sníženou schopností rozpoznání barev (barvosleposti). Test obsahoval dvě varianty map, jednu barevnou a druhou, ve které byly barvy nahrazeny symboly (obr. 44). Výsledky ukázaly, že existují statisticky významné rozdíly mezi skupinou osob s poruchou barvocitů a zdravou populací u počtu zaznamenaných fixací a délce trajektorie pohybu oka.



Obr. 44 Ukázka stimulu ze studie (Snopková, 2016) s vytvořenými oblastmi zájmu a výsledek nástroje ScanGraph.

Příkladem může být diplomová práce Hájkové (2016) z ústavu fyzioterapie Fakulty zdravotnických studií Univerzity Palackého v Olomouci. Cílem této práce bylo zjistit, zda se u pacientů po prodělané cévní mozkové příhodě vyskytují poruchy vizuospaciálního vnímání, a dochází tak k určité stranové prostorové predilekci. Měření probíhalo pomocí eye-trackeru EyeTribe. Měřena byla skupina 20 pacientů s cévní mozkovou příhodou (10 s pravostrannou hemiparézou, 10 s levostrannou hemiparézou), ve srovnání s 10 zdravými probandy. Vyšetření pomocí eye-trackingu bylo prováděno pomocí 19 úkolů zaměřených na prostorové vnímání nad stimuly ve formě obrázků. Nad těmito obrázky bylo vytvořeno sedm oblastí zájmu, a bylo zjišťováno pořadí, v jakém docházelo k vizuální inspekci těchto obrázků (obr. 45).



Obr. 45 Ukázka stimulu ze studie (Hájková, 2016) s vytvořenými oblastmi zájmu a výsledek nástroje ScanGraph.

ScanGraph byl rovněž využit na Psychologickém ústavu Masarykovy univerzity v Brně, kde Pulkrtová (2016) zkoumala vliv červené barvy v porovnání s modrou na vnímání atraktivitu žen v závislosti na pohlaví hodnotitele. Důležitým aspektem práce bylo zjistit, zda červená barva upoutá více pozornosti než barva kontrolní, a tím podpořit významnost barvy v hodnoticím procesu. Vliv salience červené barvy u mužského pohlaví byl prokázán u celkového počtu fixací a celkové délky trvání fixací na červenou část oděvu. U ženské populace se potvrdil vyšší počet fixací na rušivé pozadí v případě podnětových materiálů kontrolní barvy (obr. 46).



Obr. 46 Ukázka stimulu ze studie (Pulkrtová, 2016) a výsledky nástroje ScanGraph.

To, že ScanGraph najde uplatnění i mimo studie využívající eye-tracker prokázal Herman (2018), který se zabývá analýzou chování respondenta v 3D prostředí. Pomocí vytvořeného nástroje zaznamenává uživatelské interakce (rotace, posun, zoom) a hledá respondenty s podobnou sekvencí interakcí. V ukázce na obrázku 47 je znázorněna vizualizace posloupnosti jednotlivých způsobů interakce pro 40 respondentů. Hledání podobností vizuálně by bylo velmi náročné. Z toho důvodu Herman převedl data do formátu vhodného pro použití v nástroji ScanGraph a následně našel skupiny respondentů, jejichž posloupnost interakcí byla ze 70 % podobná. V dolní části obrázku 47 jsou sekvence znaků reprezentujících typy interakcí doplněné o barevnou vizualizaci.

Jednotlivé indikátory nabývají pouze hodnot z množiny {1,2,3,4,5}, přičemž každému indikátoru je navíc přiřazena jeho váha. Výstup realizovaný zpracovatelem byl vytvořen následujícím způsobem: Pro jednotlivé obce byl vypočítán vážený průměr příslušných indikátorů samostatně pro jednotlivé pilíře (Územní podmínky pro soudržnost společenství obyvatel území – sociální pilíř, Územní podmínky pro hospodářský rozvoj – hospodářský pilíř, Územní podmínky pro příznivé životní prostředí – environmentální pilíř) a tyto hodnoty byly rozčleněny do šesti kategorií. Tři z těchto kategorií byly hodnoceny jako pozitivní a tři jako negativní. Celkové hodnocení rozboru udržitelného rozvoje území je pak tvořeno osmi kategoriemi ($V'_3(2) = 2^3$), dle toho, zda daná obec byla v jednotlivých pilířích hodnocena jako pozitivní či negativní.

Pro srovnání s výstupem nástroje *SimUrb* bylo zvoleno takové nastavení, které vygeneruje také právě šest skupin vzájemně si podobných obcí. Pro sociální pilíř bylo zvoleno nastavení míry podobnosti 0,73 a skupiny o minimálním počtu 15 objektů, pro hospodářský pilíř byla nastavena míra podobnosti 0,75 a skupiny o minimálním počtu 16 objektů a pro environmentální pilíř byla zvolena míra podobnosti 0,71 a opět skupiny o minimálním počtu 15 objektů. Nástroj *SimUrb* však nezbytně nezařadí všechny obce do kategorií. Ty obce, jež nevyhovují zadaným parametrům, zůstanou mimo vytvořené skupiny podobných obcí. Nástroj *SimUrb* tak může více odpovídat realitě, kdy skutečně obec (či obecně jakýkoliv vyhodnocovaný objekt) si není podobná s žádnou či minimální požadovanou skupinou jiných obcí. Na okraji tak zůstanou ty obce, které mohou vykazovat jistou extrémnost v jednom či více indikátorech.

Shoda s kategoriemi vymezenými v rámci ÚAP je uvedena v tabulkách.

Hodnoty v tabulce 5 představují procentuální zastoupení obcí z podobných skupin nástroje *SimUrb* v kategoriích vymezených v rámci ÚAP pro sociální pilíř, v tabulce 6 pak srovnání pro hospodářský pilíř, v tabulce 7 pro environmentální pilíř a v tabulce 11 srovnání celkového hodnocení dle metodiky ÚAP s výstupem nástroje *SimUrb*.

Tab. 5 Srovnání hodnocení sociálního pilíře s výstupem nástroje *SimUrb*

	Počet obcí	Procentuální podíl						Střední hodnota	Směrodatná odchylka
		Velmi nevyhovující (2,25 a méně)	Nevyhovující (2,26 – 2,60)	Spíše nevyhovující (2,61 – 2,85)	Spíše vyhovující (2,86 – 3,15)	Vyhovující (3,16 – 3,50)	Velmi vyhovující (3,51 a více)		
A	47		2,13	36,17	25,53	34,04	2,13	3,03	0,241
B	24		37,5	41,67	20,83			2,71	0,238
C	21					28,57	71,43	3,69	0,229
D	21	85,71	14,29					2,05	0,215
E	19			26,32	36,84	36,84		3,06	0,247
F	18			27,78	38,89	27,78	5,55	2,77	0,226

Tab. 6 Srovnání hodnocení hospodářského pilíře s výstupem nástroje SimUrb

	Počet obcí	Procentuální podíl						Střední hodnota	Směrodatná odchylka
		Velmi nevyhovující (2,45 a méně)	Nevyhovující (2,46 – 2,80)	Spíše nevyhovující (2,81 – 3,05)	Spíše vyhovující (3,06 – 3,30)	Vyhovující (3,31 – 3,60)	Velmi vyhovující (3,61 a více)		
A	53		1,89	28,3	43,39	26,42		3,2	0,191
B	33				6,06	42,42	51,52	3,62	0,2
C	23	4,35	52,17	39,13	4,35			2,78	0,204
D	22				13,63	54,55	31,82	3,55	0,177
E	16	56,25	43,75					2,46	0,146
F	16					6,25	93,75	3,88	0,208

Tab. 7 Srovnání hodnocení environmentálního pilíře s výstupem nástroje SimUrb

	Počet obcí	Procentuální podíl						Střední hodnota	Směrodatná odchylka
		Velmi nevyhovující (2,00 a méně)	Nevyhovující (2,01 – 2,35)	Spíše nevyhovující (2,36 – 2,55)	Spíše vyhovující (2,56 – 2,85)	Vyhovující (2,86 – 3,15)	Velmi vyhovující (3,16 a více)		
A	39		12,82	38,46	46,16	2,56		2,55	0,184
B	32	37,5	53,13	9,37				2,08	0,164
C	22						100	3,77	0,245
D	19		47,37	36,84	15,79			2,42	0,179
E	17		23,53	29,41	47,06			2,54	0,23
F	15				20	53,33	26,67	3,07	0,209

V jednotlivých pilířích je možné pozorovat odlišnosti od původního vyhodnocení, jež bylo provedeno pomocí pouhého průměru hodnot. V níže uvedených tabulkách je možné na příkladech srovnat obce se stejným průměrem, které byly zařazeny nástrojem *SimUrb* do odlišných skupin.

V sociálním pilíři (tab. 8) byla obec Rovensko zařazena nástrojem *SimUrb* do skupiny B, Chromeč do skupiny E a Lipník nad Bečvou do skupiny F, kdežto při původním vyhodnocování byly zařazeny všechny do stejné kategorie, díky stejnému průměru. Avšak je možné pozorovat, že obec Rovensko má nižší hodnoty v počtu dokončených bytů, míře dostupné urbanizace a v trendu hrubé míry celkového přírůstku, ale výrazně vyšší v indexu stáří. Dále obec Chromeč se od Lipníku nad Bečvou odlišuje nižšími hodnotami v indexu stáří a indexu vzdělanosti, ale zato vyššími hodnotami v hrubé míře přirozeného přírůstku, počtu dokončených bytů a míře dostupné urbanizace. Obdobně lze srovnat i obec Rovensko s Lipníkem nad Bečvou. Je tedy zřejmé, že ačkoliv obce vykazují stejnou průměrnou hodnotu indikátorů, nejsou si svými hodnotami indikátorů příliš podobné.

Tab. 8 Ukázka obcí se stejným průměrem v sociálním pilíři zařazených nástrojem SimUrb do různých kategorií.

Obec	Skupina Simurb	Hrubá míra přirozeného přírůstku	Hrubá míra migračního salda	Počet dokončených bytů	Míra dostupné urbanizace	Index stáří	Index vzdělanosti	Trend hrubé míry celkového přírůstku	
Rovensko	B	0,6	0,3	0,4	0,05	0,8	0,6	0,1	2,85
Chromeč	F	0,6	0,3	0,6	0,15	0,4	0,6	0,2	2,85
Lipník nad Bečvou	F	0,45	0,3	0,4	0,1	0,6	0,8	0,2	2,85

Jak je patrné z příkladu uvedeného v tabulce 9, obec Vlkoš má výrazně nižší hodnotu u míry podnikatelské aktivity oproti ostatním obcím a naopak vyšší u indexu dostupnosti dálniční sítě.

Tab. 9 Ukázka obcí se stejným průměrem v hospodářském pilíři zařazených nástrojem SimUrb do různých kategorií

Obec	Skupina Simurb	Podíl osob zaměstnaných v terciéru	Míra podnikatelské aktivity	Index dostupnosti dálniční sítě	Podíl cizích zdrojů k celkovým aktivitám obce	Hrubá míra salda dojížďky	Podíl nezaměstnaných osob	Podíl dlouhodobě nezaměstnaných osob	Trend podílu nezaměstnaných osob	
Štarnov	A	0,4	0,6	0,6	0,15	0,3	0,75	0,5	0,1	3,4
Vlkoš	B	0,3	0,4	0,8	0,2	0,3	0,75	0,5	0,15	3,4
Ochoz	D	0,3	0,8	0,4	0,2	0,3	0,75	0,5	0,15	3,4

V tabulce 10 je možné pozorovat například obec Přerov, která vyniká vysokou hodnotou podílu pásem ochrany vod, ovšem u dalších několika indexů vykazuje hodnoty nižší. Obce z tohoto příkladu jsou opět zařazeny do stejné kategorie metodou ÚAP, ovšem nástrojem SimUrb jsou rozřazeny do různých skupin.

Tab. 10 Ukázka obcí se stejným průměrem v environmentálním pilíři zařazených nástrojem SimUrb do různých kategorií

Obec	Skupina Simurb	Podíl pásem ochrany vod	Podíl ploch ochrany přírody	Koncentrace prachových částic PM10	Ekol. fragmentace nezastavěného území	Koeficient ekologické stability	Podíl lesa	Podíl půd 1. a 2. třídy ochrany	Stupeň realizace pozemkových úprav	
Přerov	A	0,6	0,3	0,15	0,3	0,1	0,2	0,6	0,2	2,45
Kožušany-Tážaly	B	0,3	0,3	0,3	0,4	0,1	0,1	0,75	0,2	2,45
Hrubčice	D	0,15	0,15	0,3	0,4	0,1	0,1	0,75	0,5	2,45

Dle metodiky ÚAP byly jednotlivé kategorie rozlišeny na negativní (Velmi nevyhovující, Nevyhovující a Spíše nevyhovující) a pozitivní (Spíše vyhovující, Vyhovující a Velmi vyhovující) a byly jim tak přiděleny symboly „+“ a „-“. Jednotlivé obce tak byly uspořádány do nových shluků dle přidělené uspořádané trojice těchto symbolů. Nástrojem SimUrb byla provedena analýza pomocí funkcionality *Multifile* pro ZIP soubory, kdy se vyhledaly podobné obce v rámci všech tří pilířů. Procentuální shodné zastoupení je opět znázorněno v tabulce 11.

Tab. 11 Srovnání celkového hodnocení dle metodiky ÚAP s výstupem nástroje SimUrb

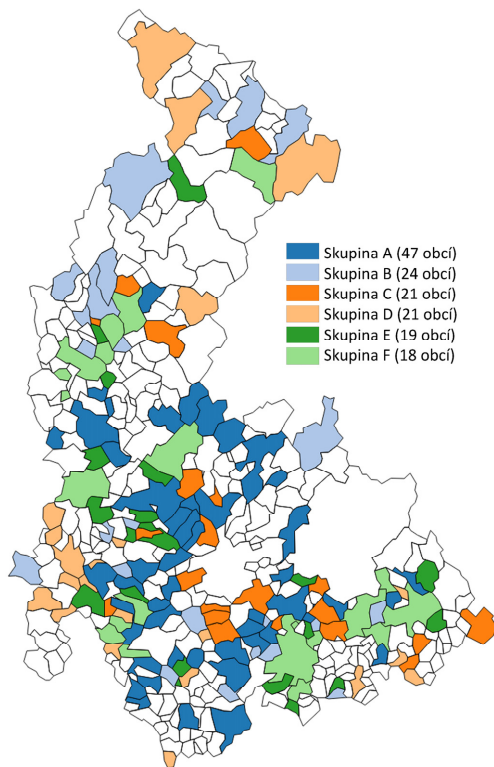
		Procentuální podíl			
		---	+-	-+	+++
			+-	++	
			++	++	
A	30		3,33	33,34	33,33
			3,33		
			3,33	23,34	
B	19	26,31	36,85		5,27
			15,78		
			5,26	10,53	
C	18			16,67	5,55
			27,78		
				50	
D	16	6,25		6,25	18,75
				12,5	
			56,25		
E	15		6,67		40
				6,67	
				46,66	
F	13	38,46	7,69	7,69	7,69
			38,47		
G	12			33,34	
				66,66	
H	11		9,1		
				90,9	

Analyzovaná data byla pro srovnání vyhodnocena také nástroji dostupnými v softwaru ArcGIS Pro, které umožňují vytvářet shluky obcí na základě jejich atributů (nástroje *Grouping Analysis* a *Multivariate Clustering*). Nástroj *Grouping Analysis* nabízí možnost výpočtu s prostorovým kontextem i bez něj. V případě, že je do výpočtu zahrnuto i prostorové rozložení objektů, využívá nástroj minimální kostry grafu pro nalezení přirozených shluků. Jestliže prostorová složka objektů nevstupuje do určení shluků, použije nástroj *K Mean algoritmus*, jenž se řadí mezi NP-těžké úlohy. Při jeho inicializaci je náhodně zvoleno tolik objektů, jaký je požadovaný počet skupin, a to tak, aby byly od sebe co nejvíce vzdáleny v rámci svých přiřazených hodnot. Jelikož počáteční objekty, ke kterým se přiřazují další objekty pro vytvoření shluků, jsou voleny náhodně, je při každém spuštění nástroje výsledná varianta výstupu jiná. Při spuštění *Multivariate Clustering* při zadaném počtu požadovaných shluků je cílem nalézt co nejvíce homogenní shluky navzájem co nejvíce heterogenní. Shluky jsou opět tvořeny pomocí *K Mean algoritmu*.

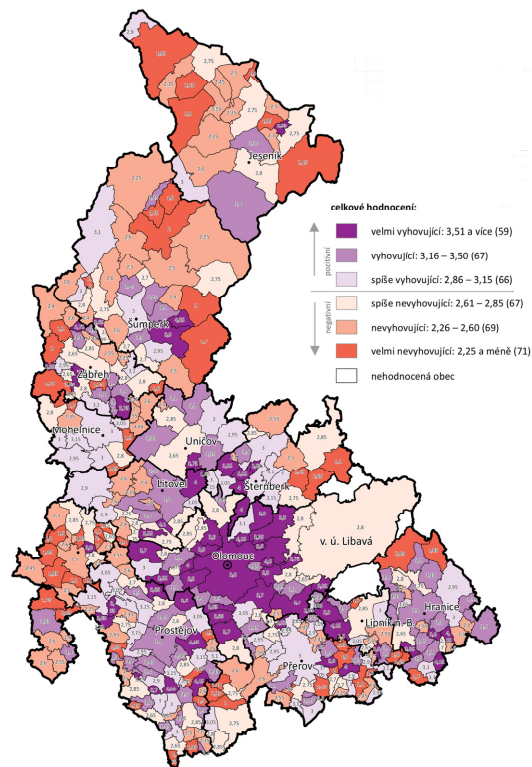
Na níže uvedených obrázcích 48 až 51 jsou vizuálně porovnané výstupy jednotlivých nástrojů. Lze pozorovat, že zařazení dle metodiky ÚAP a metody *Grouping Analysis* a *Multivariate Clustering* vytváří srovnatelné výstupy, na rozdíl od nástroje *SimUrb*, který přináší nový přístup k vyhledávání podobných skupin. Zatímco nástroj *SimUrb* zařadí do skupin pouze některé z vyhodnocovaných obcí, ostatní nástroje či metody zařadí vždy všechny územní celky. Ačkoliv lze pozorovat mezi výstupy nástrojů *Grouping Analysis* a *Multivariate Clustering* drobné odchylky, je možno výsledné skupiny téměř zaměnit. Například pro sociální pilíř je skupina 3 nástroje *Grouping Analysis* (dále jen GA) zaměnitelná se skupinou 4 nástroje *Multivariate Clustering* (dále jen MC) nebo skupina 6 nástroje GA se skupinou 3 nástroje MC. Přiřadit je možné i skupiny vytvořené metodikou ÚAP, kdy například skupina velmi nevyhovující se překrývá se skupinou 4 nástroje MC. U hospodářského pilíře lze srovnat například velmi vyhovující obce se skupinou 4 nástroje

GA a skupinou 5 nástroje MC. U environmentálního pilíře lze pozorovat například shody u skupiny 1 nástroje GA se skupinou 6 nástroje MC a skupin 3 u GA i MC. Výstupy na obrázcích 48 až 51 jsou v plném rozlišení součástí digitální přílohy na CD.

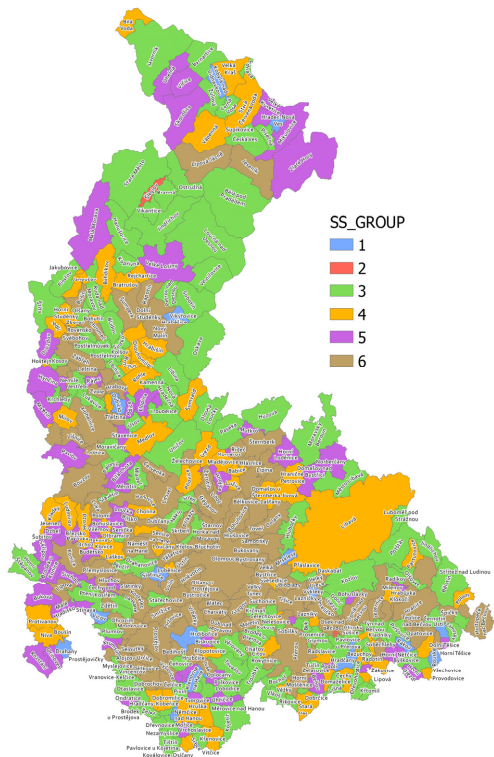
ÚZEMNÍ PODMÍNKY PRO SOUDRŽNOST SPOLEČENSTVÍ ÚZEMÍ



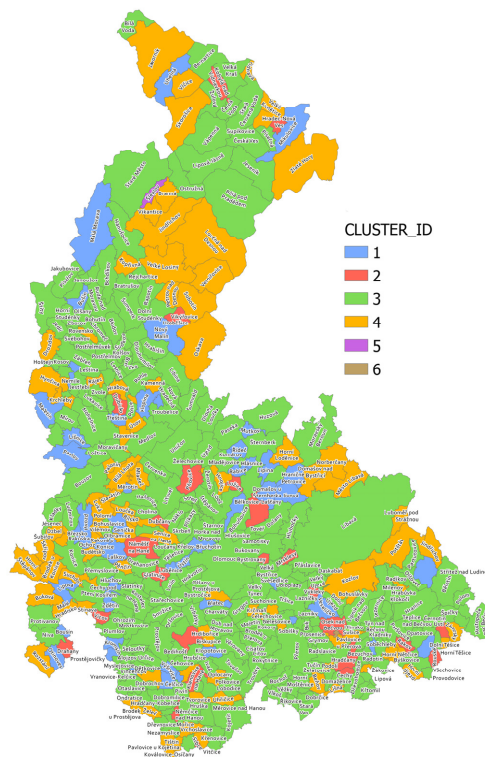
A - Výstup nástroje SimUrb



B - Výstup dle metodiky RURÚ



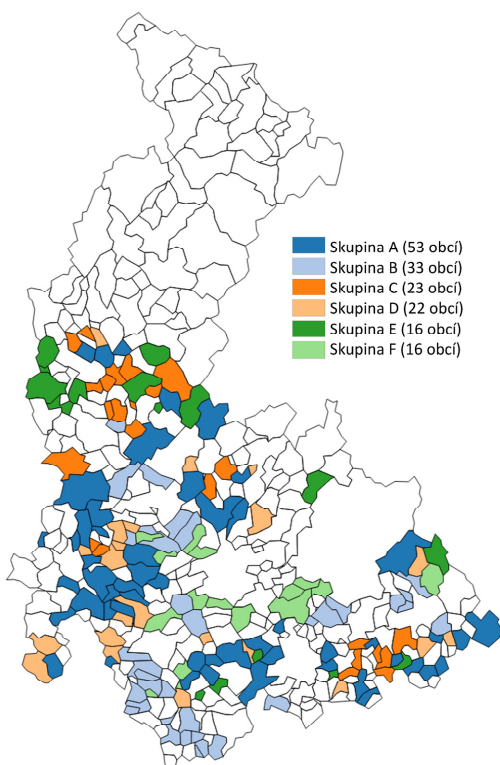
C - Grouping Analysis (ArcGIS)



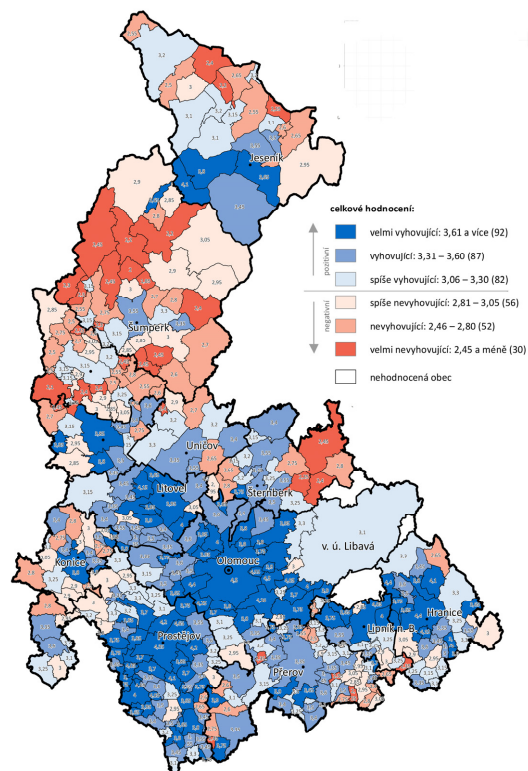
D - Multivariate Clustering (ArcGIS)

Obr. 48 Srovnání výstupů nástrojů A – SimUrb, B – Urban Planner, C – Grouping Analysis, D – Multivariate Clustering pro sociální pílř

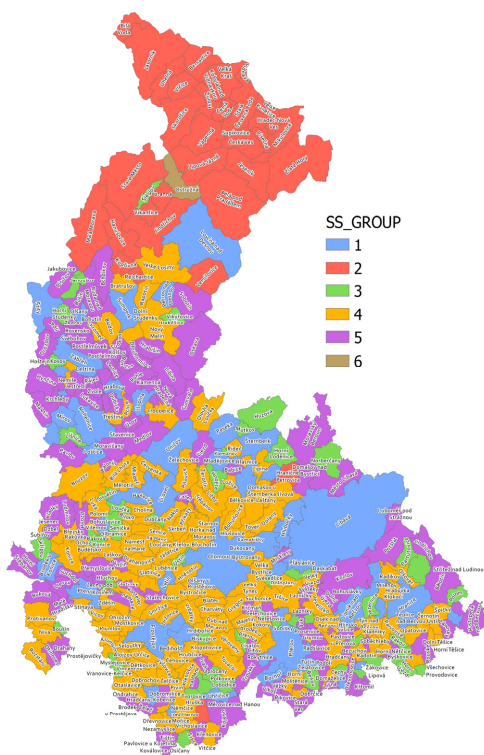
ÚZEMNÍ PODMÍNKY PRO HOSPODÁŘSKÝ ROZVOJ



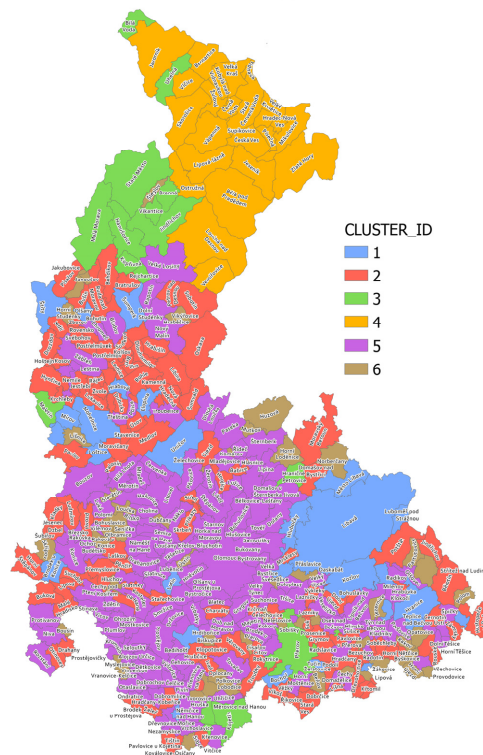
A - Výstup nástroje SimUrb



B - Výstup dle metodiky RURÚ



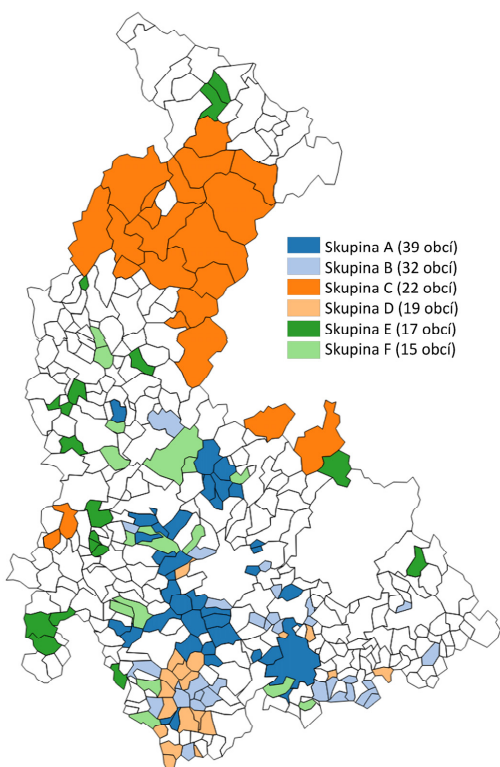
C - Grouping Analysis (ArcGIS)



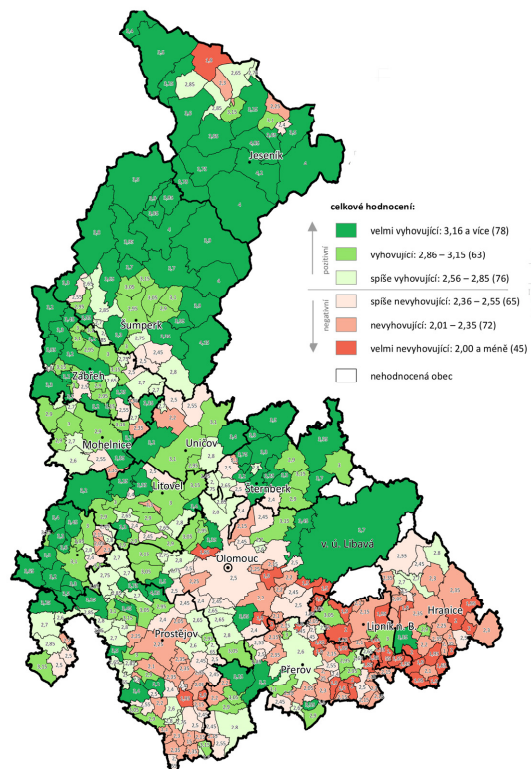
D - Multivariate Clustering (ArcGIS)

Obr. 49 Srovnání výstupů nástrojů A – SimUrb, B – Urban Planner, C – Grouping Analysis, D – Multivariate Clustering pro hospodářský pilíř

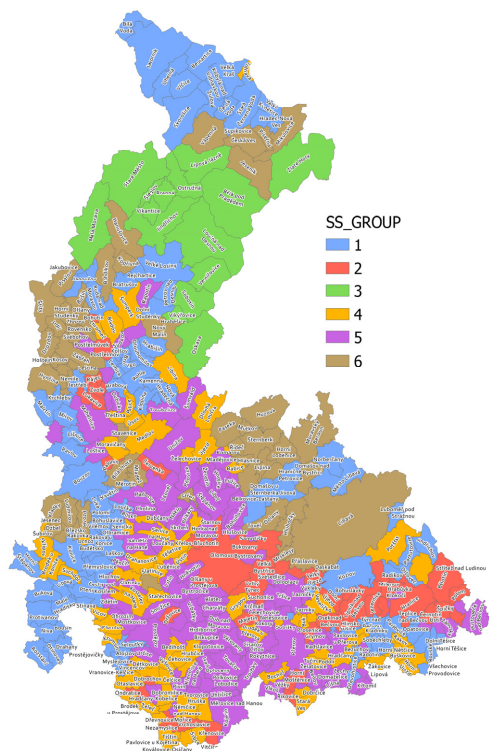
ÚZEMNÍ PODMÍNKY PRO PŘÍZNVÉ ŽIVOTNÍ PROSTŘEDÍ



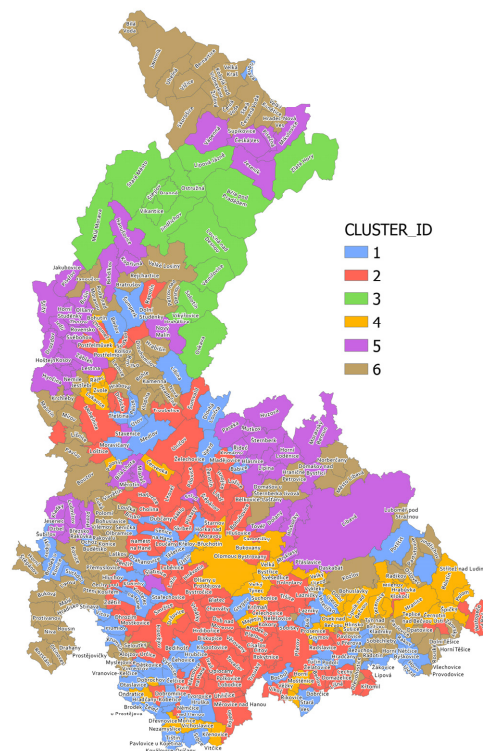
A - Výstup nástroje SimUrb



B - Výstup dle metodiky RURÚ



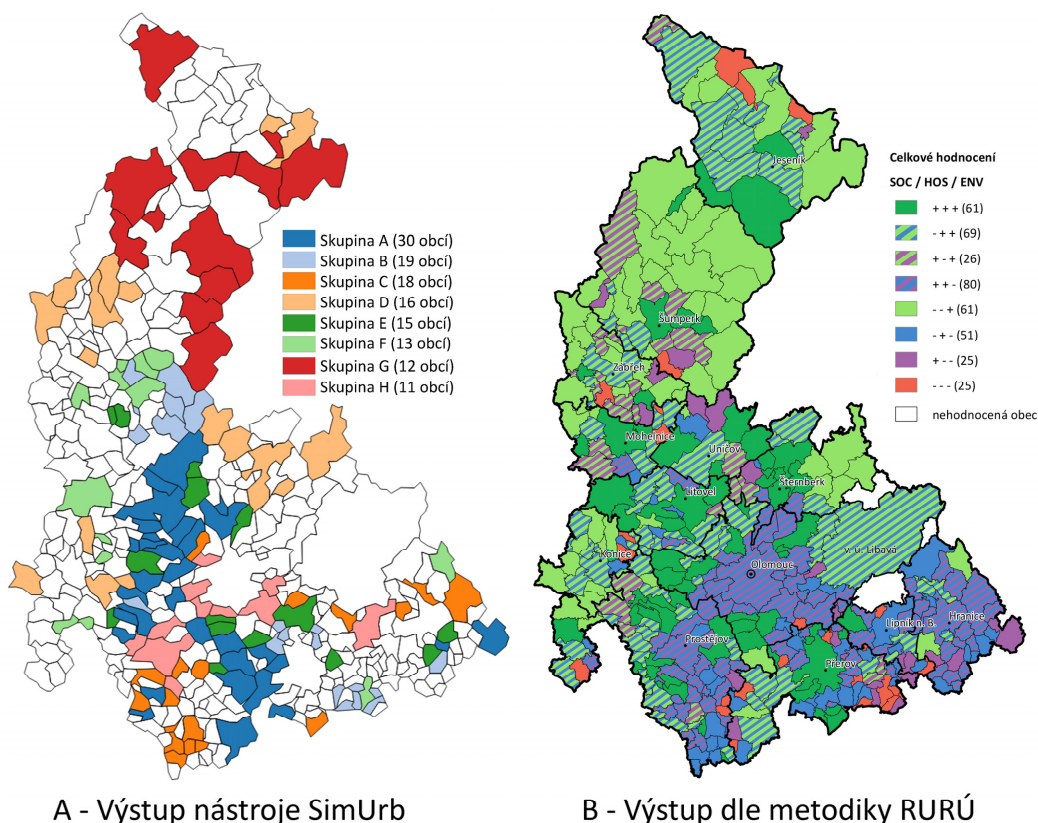
C - Grouping Analysis (ArcGIS)



D - Multivariate Clustering (ArcGIS)

Obr. 50 Srovnání výstupů nástrojů A – SimUrb, B – Urban Planner, C – Grouping Analysis, D – Multivariate Clustering pro environmentální pílť

CELKOVÉ HODNOCENÍ ROZBORU UDRŽITELNÉHO ROZVOJE ÚZEMÍ



Obr. 51 Srovnání výstupů nástrojů A – SimUrb, B – Urban Planner pro celkové hodnocení rozboru udržitelného rozvoje území

7.2.2 Časoprostorové analýzy typu a stavu zemědělských plodin

Bakalářská práce Dratvy (2018) se zabývá časoprostorovou analýzou zemědělských plodin pomocí dat ze systému Sentinel-2 z let 2016 a 2017. Cílem práce analyzovat data za účelem zjištění podobnosti plodin pomocí čtyř vegetačních indexů. Pro analýzu byly vybrány vegetační indexy NDVI, CVI, SAVI a GLI, které byly určeny pro každé zájmové pole. Kromě korelace hodnot mezi sledovanými lety byl pro analýzu podobnosti využit nástroj *SimUrb*, jenž vyhledal vzájemně si podobná pole v jednotlivých letech.

Analýzováno bylo 34 polí se sedmi plodiny v roce 2016 a devíti plodinami v roce 2017, pro něž byly dostupné hodnoty jednotlivých indexů pro 11 dní v roce v obou letech. Data měření tvořila párové hodnoty pro rok 2016 a 2017, s maximální odchylkou pěti dní.

Analýza nástrojem *SimUrb* byla provedena z pohledu subjektu (obec), jehož cílem je vyhledat všechna pole, na kterých se v daném roce pěstoval daný druh plodiny. Předpokladem je, že subjekt zná spektrální projev dané plodiny během roku, nebo ví alespoň o jednom poli, na kterém požadovaná plodina rostla. Díky tomu by mohl být schopen na základě shluků vyhledat všechna pole, na kterých byla zkoumaná plodina pěstována a tím urychlit zjišťování typu plodiny.

Vstupními daty pro nástroj *SimUrb* byly tedy sekvence 11 hodnot pro každý ze čtyř indexů každého z 34 polí. Porovnávány byly sekvence příslušných indexů pro jednotlivé roky. Pro každý index bylo nalezeno ideální nastavení míry podobnosti, váha všech indexů byla nastavena rovna jedné.

Výsledné skupiny podobných polí odpovídaly skutečnosti v převážné většině případů (tab. 12 - 15). Největších odklonů od reálného stavu bylo pozorováno u kukuřice, která byla často slučována se sójou či cukrovou řepou.

Tab. 12 Skupiny vytvořené nástrojem SimUrb pro index NDVI v roce 2017 s hodnotou míry podobnosti $p = 0,9$

Shluk	ID Pole	Plodina 2017
1	19	tráva
1	25	tráva
1	30	tráva
2	28	tráva
2	29	tráva
2	31	tráva
3	15	ječmen jarní
3	17	ječmen jarní
3	24	ječmen jarní
4	18	tráva
4	23	tráva
4	32	tráva
5	7	cukrovka
5	8	cukrovka
5	9	cukrovka
6	22	pšenice jarní
6	26	sója
7	33	tráva
7	34	vojtěška
8	12	tráva
8	13	tráva
9	2	kukuřice
9	6	kukuřice

Tab. 13 Skupiny vytvořené nástrojem SimUrb pro index CVI v roce 2017 s hodnotou míry podobnosti $p = 0,9$

Shluk	ID Pole	Plodina 2017
1	3	tráva
1	32	tráva
1	29	tráva
1	11	tráva
1	19	tráva
1	18	tráva
1	13	tráva
1	16	tráva
1	23	tráva
1	25	tráva
2	26	sója
2	22	pšenice jarní
2	9	cukrovka
2	8	cukrovka
2	7	cukrovka
2	4	cukrovka
3	17	ječmen jarní
3	24	ječmen jarní
3	15	ječmen jarní
4	30	tráva
4	28	tráva
4	12	tráva
5	20	jetel
5	21	jetel
6	31	tráva
6	33	tráva
7	2	kukuřice
7	10	sója
8	5	kukuřice
8	6	kukuřice

Tab. 14 Skupiny vytvořené nástrojem SimUrb pro index GLI v roce 2017 s hodnotou míry podobnosti $p = 0,9$

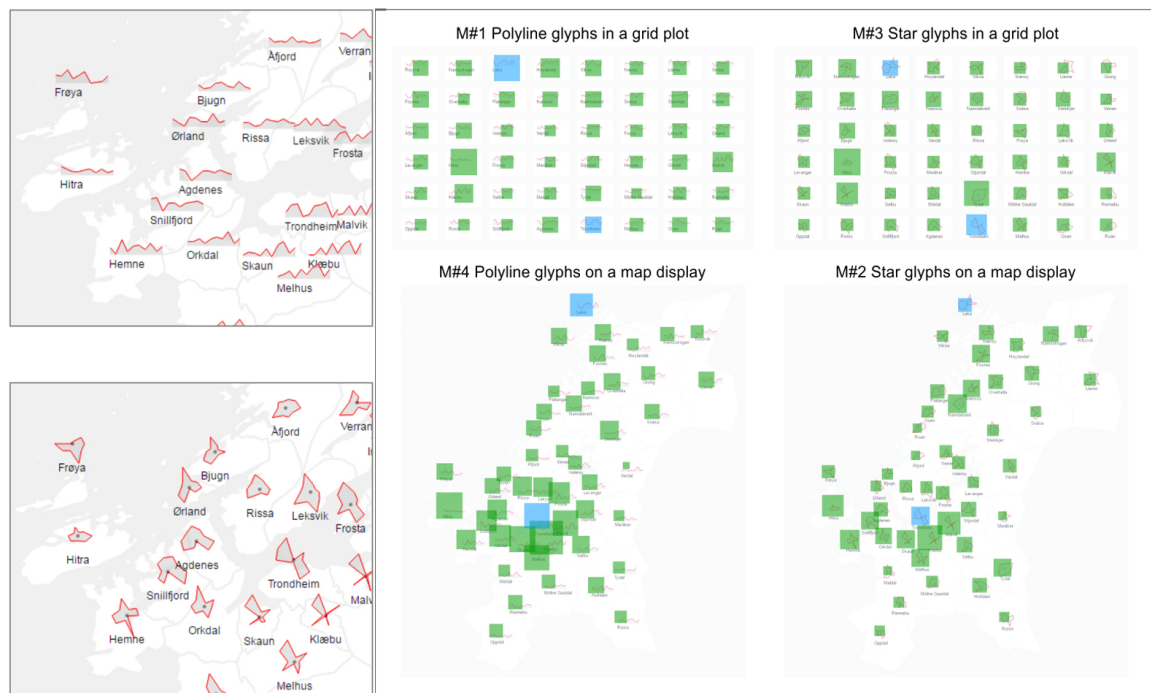
Shluk	ID Pole	Plodina 2017
1	7	cukrovka
1	4	cukrovka
1	8	cukrovka
1	9	cukrovka
1	10	sója
1	29	tráva
1	30	tráva
1	31	tráva
1	23	tráva
1	12	tráva
1	18	tráva
1	25	tráva
1	19	tráva
1	16	tráva
1	28	tráva
1	5	kukuřice
1	2	kukuřice
1	6	kukuřice
1	27	kukuřice
2	22	pšenice jarní
2	26	sója
2	15	ječmen jarní
2	24	ječmen jarní
2	17	ječmen jarní
3	13	tráva
3	3	tráva
3	32	tráva
3	11	tráva
4	33	tráva
4	34	vojtěška
4	1	pšenice ozimá
4	14	pšenice ozimá
5	20	jetel
5	21	jetel

Tab. 15 Skupiny vytvořené nástrojem SimUrb pro index SAVI v roce 2017 s hodnotou míry podobnosti $p = 0,9$

Shluk	ID Pole	Plodina 2017
1	29	tráva
1	31	tráva
1	28	tráva
1	18	tráva
1	25	tráva
1	23	tráva
2	4	cukrovka
2	7	cukrovka
2	9	cukrovka
2	8	cukrovka
3	16	tráva
3	30	tráva
3	19	tráva
4	1	pšenice ozimá
4	15	ječmen jarní
4	17	ječmen jarní
5	26	sója
5	22	pšenice jarní
6	27	kukuřice
6	2	kukuřice
7	33	tráva
7	34	vojtěška
8	13	tráva
8	12	tráva
9	5	kukuřice
9	6	kukuřice

7.2.3 Porovnání polyline a star glyphů

Nástroj *SimUrb* byl využit také ve studii Opacha a kol. (2017), kde byly analyzovány rozdíly ve vnímání tzv. *Polyline* a *Star glyphů* (obr. 52 vlevo). Pomocí těchto glyphů je možné zobrazit vícedimenzionální data. Ve studii bylo zjišťováno, který typ je vhodnější pro různé úkoly. Pomocí obou typů glyphů byly vizualizovány hodnoty deseti indikátorů pro obce v části Norska. Respondenti řešili úkoly jako nalezení konkrétního glyphu, odhad hodnoty indikátoru, ale také hledání podobných glyphů. Právě pro tento typ úkolu byl využit výsledek nástroje *SimUrb*, díky kterému mohla být číselně vyjádřena podobnost jednotlivých obcí a tím pádem kvantifikována přesnost odpovědí respondentů. V úkolu číslo 4 měli respondenti za úkol nalézt dva nejpodobnější glyphy. V pravé části obrázku 52 jsou znázorněny počty fixací zaznamenaných nad jednotlivými glyphy. Nejpodobnější dvojice glyphů (rovněž zjištěna pomocí *SimUrb*) je znázorněna pomocí modré barvy.



Obr. 52 Polyline a Star glyphy (vlevo), využitě ve studii Opacha a kol. (2017). V pravé části obrázku je znázorněn počet fixací při hledání dvojice nejpodobnějších glyphů.

7.3 Případová studie *Road Similarity*

7.3.1 Analýza toků veřejné dopravy na základě dat z mýtných bran

Případová studie využívající nástroj *SimRoad* navazuje na bakalářskou práci Roubalíka (Roubalík, 2017). Snahou zde není porovnat výsledky této práce s výstupy z nástroje *SimRoad*, ale přinést nové výsledky, které jsou dostupné pouze díky tomuto nástroji.

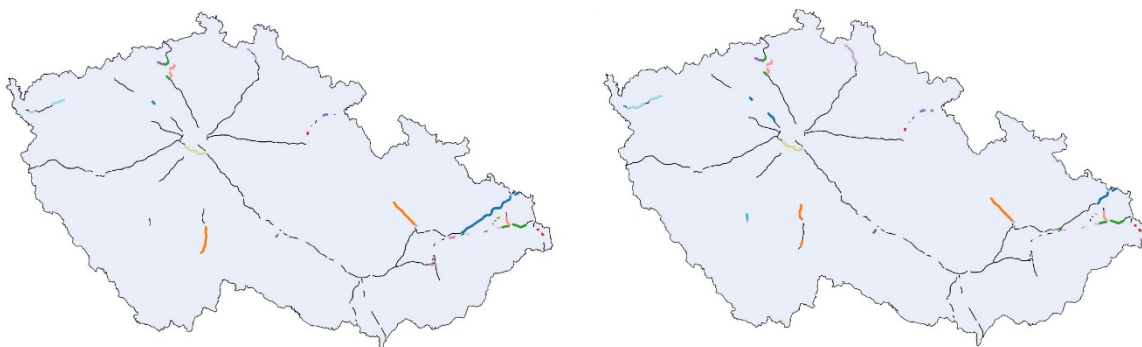
Bakalářská práce Štěpána Roubalíka se zabývá analýzou toků veřejné dopravy na základě dat z mýtných bran. V práci jsou použita data poskytnutá Ředitelstvím silnic a dálnic v průběhu roku 2015. V praktické části se práce zaměřuje na statistické vyhodnocení a grafickou vizualizaci dat o průjezdech autobusových dopravců mýtnou zónou. Dále popisuje časové změny intenzity dopravy v průběhu roku 2015 z celkového hlediska i z hlediska vybraných dopravců a také srovnává intenzitu dopravy s datovou sadou z Celostátního sčítání dopravy z roku 2010. Je popsána časová i prostorová variabilita dopravního toku a rozdíly mezi individuálními dopravci.

Bakalářská práce Roubalíka se však nezabývá vyhledáváním podobných silničních úseků z hlediska průjezdnosti. Tato analýza byla provedena nástrojem *SimRoad* nad rámec cílů této bakalářské práce.

Pro případovou studii byla vybrána data z května 2015. Studie analyzovala data dvěma přístupy. Prvním z nich bylo porovnání dat průjezdů na všech typech silnic během pracovních dnů a během víkendů.

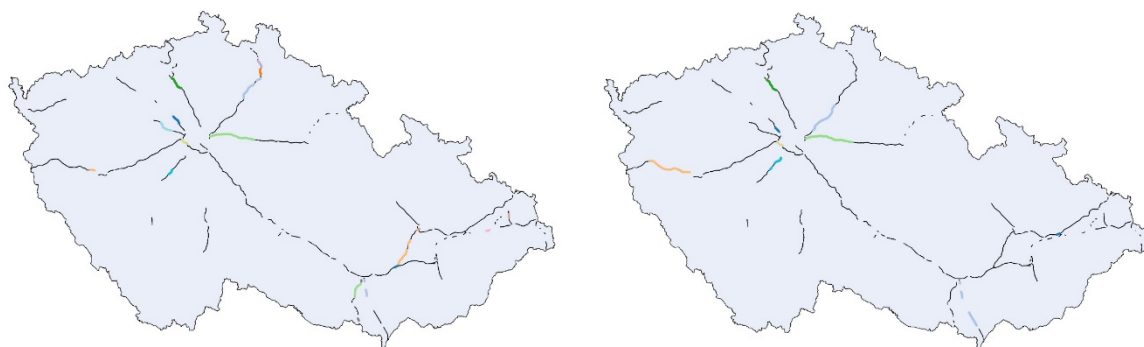
Pro oba výstupy byla zvolena hodnota míry podobnosti $p = 0,8$ a vyhodnocovány byly pouze skupiny o deseti a více úsecích. V obou případech tak bylo nalezeno 12 skupin. Zmíněno je několik pozorování, která mohou být dále studována například v bakalářských pracích.

Největší skupina v pracovních dnech byla tvořena 101 úseky, o víkendech 113 úseky (obr. 53). Seznamem svých úseků se příliš neliší. V obou případech je do této skupiny zahrnuta rychlostní silnice R35 na úseku Křelov–Mohelnice, rychlostní silnice R06 a rychlostní silnice R01 (jižní část Pražského okruhu). Zatímco v pracovních dnech je do této skupiny zahrnuta dálnice D01 na delším úseku Butovice–Věrnovice (PL), o víkendech se jedná pouze o část při hranicích s Polskem Bohumín–Věrnovice (PL). Dálnice D03 je v pracovních dnech zahrnuta svou jižní částí na úseku Tábor–Veselí nad Lužnicí, o víkendech je to zejména severní část na úseku Čekanice–Mezno. Průměrná průjezdnost v pracovních dnech na těchto úsecích byla 57,3 vozidel autobusových dopravců (dále jen vozidel) na pracovní den a o víkendech 61,8 vozidel na víkendový den.



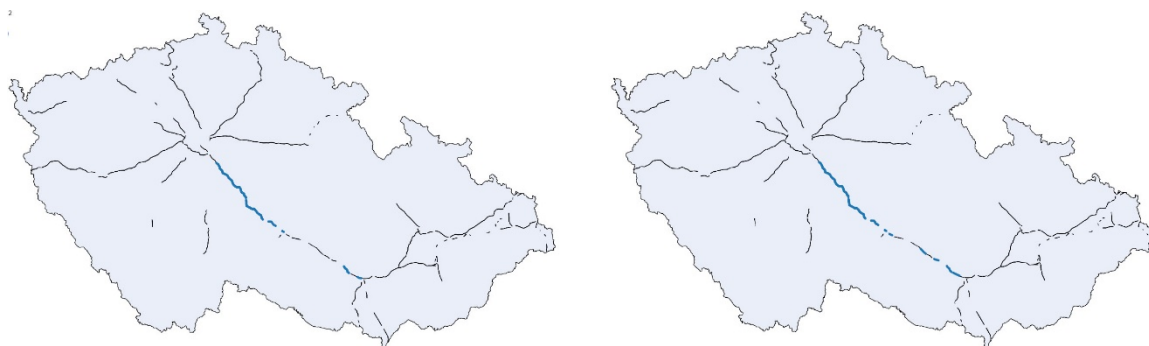
Obr. 53 Největší skupiny pro všechny komunikace nalezené nástrojem *SimRoad*. Vlevo pracovní dny, vpravo víkendy.

V obou případech našel nástroj skupinu silnic s podobnou průjezdností v okolí Prahy a Brna (obr. 54). V případě pracovních dnů se jednalo o 54 úseků, o víkendech o 52 úseků. Jednalo se zejména v okolí Prahy o úseky dálnice D11 Praha–Počernice–Poděbrady, část dálnice D08, rychlostní silnice R01, R10, a v okolí Brna rychlostní silnice R46 a R52. O víkendy k pražským silnicím přibýly úseky dálnice D05, zatímco v okolí Brna nebyly R46 a R52 zařazeny vůbec. Místo nich však vykazuje podobnou průjezdnost dálnice D02. Průměrná průjezdnost v pracovních dnech na těchto úsecích byla 125,9 vozidel na den a o víkendech 150,5 vozidel na den.



Obr. 54 Okolí Prahy a Brna pro všechny komunikace. Vlevo pracovní dny, vpravo víkendy.

Dálnice D01 mezi Prahou a Brnem v obou případech vykazovala podobnost pouze v rámci svých vlastních úseků (obr. 55). V pracovních dnech byla nalezena skupina 26 úseků Modletice–Jihlava a Devět křížů–Brno-centrum a o víkendech 28 úseků, kdy k úsekům z pracovních dnů přibyl úsek Velká Bíteš–Devět křížů a Jihlava–Velký Beranov. Průměrná průjezdnost v obou případech převyšovala úseky všech ostatních komunikací. V případě pracovních dnů se jednalo o 257,8 vozidel na den, v případě víkendů o 311,5 vozidel na den.



Obr. 55 Podobné úseky dálnice D mezi Prahou a Brnem, větší část. Vlevo pracovní dny, vpravo víkendy.

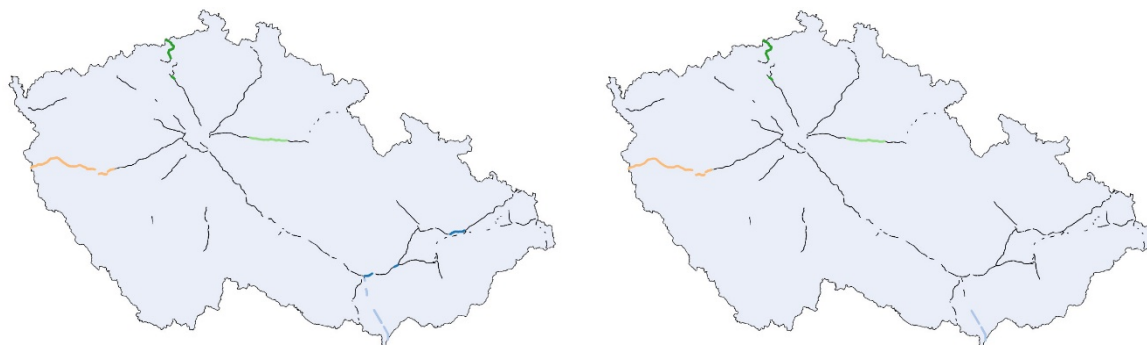
Zbylé úseky dálnice D01 mezi Prahou a Brnem byly zařazeny v obou případech opět do samostatné skupiny, pouze u pracovních dnů k nim byl přiřazen ještě jeden úsek dálnice D08 Praha–Březiněves–Zdiby. V pracovních dnech jejich průměrná průjezdnost činila 216,7 vozidel na den a o víkendech 280,7 vozidel na den.



Obr. 56 Podobné úseky dálnice D01 mezi Prahou a Brnem, menší část. Vlevo pracovní dny, vpravo víkendy.

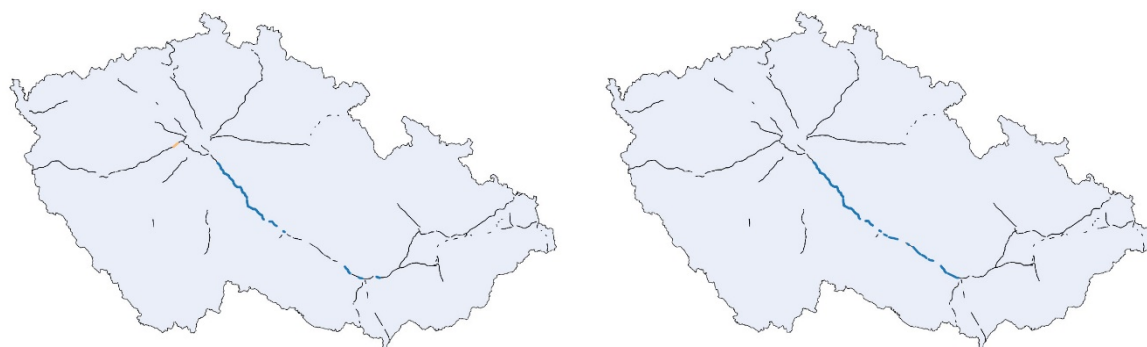
Druhým přístupem bylo srovnávání průjezdnosti v pracovních dnech a víkendech pouze u dálnic. Opět byla nastavena hodnota míry podobnosti na $p = 0,8$. V obou případech tak bylo nalezeno 7 skupin o více než 10 dálničních úsecích.

Největší skupina o víkendech byla zastoupena 40 úseky, zatímco v pracovních dnech to bylo 54 úseků (obr. 57). Jednalo se v obou případech úseky dálnice D05, D02, D08 a D11, přičemž v pracovních dnech přibyly ještě úseky dálnice D01 Brno-jih-Brno-východ, Vyškov-západ-Vyškov-východ a Lipník nad Bečvou-Hranice. Průměrná průjezdnost v pracovních dnech činila 104,4 vozidel na den a o víkendech 124,1 vozidel na den.



Obr. 57 Největší skupina nalezená pro dálniční úseky. Vlevo pracovní dny, vpravo víkendy.

Dálnice D01 mezi Prahou a Brnem byla o víkendech zařazena celá do jedné skupiny s průměrnou průjezdností 307,3 vozidel na den. V pracovních dnech však byla opět rozdělena do dvou skupin, a to na úseky Modletice-Jihlava a Devět křížů-Brno-centrum s průměrnou průjezdností 258,1 vozidla na den a zbylé úseky s průměrnou průjezdností 216,8 vozidel na den (obr. 58).



Obr. 58 Dálnice D01 mezi Prahou a Brnem. Vlevo pracovní dny, vpravo víkendy.

Dálnice D01 v příhraniční oblasti s Polskem na úseku Bělotín-západ-Věrnovice (PL) vykazovala podobnou průjezdnost s dálnicí D03, částí dálnice D11 na úseku Chlumec-Sedlice a částí dálnice D08 na úseku Lovosice-Doksany a Řehlovice-Trmice (obr. 59). Průměrná průjezdnost v případě pracovních dnů byla 61,7 vozidel na den a 81,5 vozidel na den.



Obr. 59 Dálniční úseky podobné s příhraničními úseky dálnice D01. Vlevo pracovní dny, vpravo víkendy.

8 VÝSLEDKY

Disertační práce *Teorie grafů při vyhodnocování časoprostorových dat* je zaměřena na implementaci grafů do tří tematických okruhů, a to v analýze eye-tracking dat, prostorovém plánování a dopravě. Ve všech třech oblastech se práce zabývala vyhledáváním podobných skupin v časoprostorových datech. Výsledkem každého ze tří studovaných okruhů byla navržena metoda pro adekvátní vyhledávání těchto skupin a nástroj, jenž tuto metodu využívá. Funkcionalita nástrojů byla ověřena na třech případových studiích. Nástroje byly využity také v několika dalších případových studiích jiných autorů.

8.1 DC1 – Využití grafů v analýze eye-tracking dat

V rámci prvního dílčího cíle byly popsány dosavadní přístupy k vyhodnocování podobnosti mezi daty eye-tracking experimentů, jimiž jsou sekvence navštívených oblastí zájmu. Byly charakterizovány algoritmy, které jsou běžně využívány při výpočtu *String Edit Distance* (*Levenshteinova vzdálenost*, *Damerau-Levenshteinova vzdálenost* a *Needleman-Wunsch algoritmus*). Krátce byly popsány nástroje, které s těmito algoritmy pracují a které jsou běžně používány pro vyhodnocování podobnosti eye-tracking dat (*eyePatterns*, *ClustalG* a *TramineR*). Byly zhodnoceny možnosti a limity jejich využití. Na základě toho byla navržena nová metoda využívající grafových algoritmů pro vyhledávání podobných skupin mezi eye-tracking daty, která pracuje již bez omezení předchozích nástrojů. Největším přínosem této metody je skutečnost, že při určení míry podobnosti mezi sekvencemi zahrnuje také délku jednotlivých sekvencí a nedochází tak k výpočetním nesrovnalostem při porovnávání množiny sekvencí s výrazně odlišným počtem znaků. Výsledkem této metody je míra podobnosti označovaná jako p , jež nabývá hodnot z intervalu $(0,1)$, přičemž $p = 0$ v případě zcela odlišných sekvencí (neshodují se v jediném znaku) a $p = 1$ pro totožné sekvence. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných sekvencí.

Tuto metodu využívá navržený online nástroj s názvem *ScanGraph* volně dostupný na adrese eyetracking.upol.cz/scangraph.

Tento nástroj umožňuje analyzovat podobnost mezi sekvencemi s vyznačenými oblastmi zájmu jednoho či více stimulů. V případě více stimulů je zapotřebí, aby k jednotlivým stimulům byla vázána naměřená data od stejných respondentů. Pro analýzu si uživatel může vybrat mezi třemi metodami, jež vycházejí z algoritmů *Levenshteinova vzdálenost*, *Damerau-Levenshteinova vzdálenost* a *Needleman-Wunsch algoritmus*. Volba metody záleží na charakteru analyzovaných dat a na cíli, jenž si uživatel stanovil. Doporučení pro volbu metody je blíže popsáno v kapitole 4.3.2.

Při analýze více stimulů najednou má uživatel navíc na výběr ze dvou metod – *Mean* a *Occurrence*.

Kromě určení podobných skupin sekvencí při zadané míře podobnosti p je možnost spuštění výpočtu i pomocí stanovení hustoty hran grafu. Grafem je míněn obyčejný graf, jenž je zobrazen na výstupu nástroje a jehož vrcholy reprezentují jednotlivé analyzované sekvence a hrany pak reprezentují relaci podobnosti mezi nimi. Podobné skupiny sekvencí jsou tak reprezentovány klikami v tomto grafu.

Omezení tohoto nástroje bylo testováno na náhodně vygenerovaných sekvencích s předem danými parametry (délka, počet). Výsledky tohoto testování jsou popsány v kapitole 4.5.

Nástroj *ScanGraph* byl použit v několika případových studiích. Ve studiích autorky *Vizualizace nejistoty geografických dat* (kapitola 7.1.1), *Hodnocení nefotorealistické 3D*

vizualizace měst (kapitola 7.1.2) či ve studiích jiných autorů, jako Beitlová (2017) *Hodnocení kartografické gramotnosti* (kapitola 7.1.3) či v dalších studiích (kapitola 7.1.4). Případové studie jsou náplní čtvrtého dílčího cíle.

Navržená metoda a použití nástroje *ScanGraph* již bylo publikováno:

Doležalová, J., Popelka, S. (2016) *Evaluation of user strategy on 2D and 3D city maps based on novel scanpath comparison method and graph visualization* (Proceedings of ISPRS)

Popelka, S., **Dolezalova, J.** (2017) *Scanpath comparison with the use of web applications SMI2OGAMA and ScanGraph*. Eye Movements in Programming: Spring Academy 2017, TR-B-17-02, Berlin

Dolezalova, J., Popelka, S. (2016) *ScanGraph: A novel scanpath comparison method using graph cliques visualization*. Journal of Eye Movement Research

Dolezalova, J., Popelka, S. (2016) *The use of simple graphs and cliques for analysis of cartographic eye-tracking data*. Proceedings of ITAT

8.2 DC 2 – Využití teorie grafů v prostorovém plánování

Druhý dílčí cíl byl zaměřen do oblasti prostorového plánování. Byly popsány běžně používané přístupy k vytváření podobných skupin mezi daty prostorového plánování, tedy základní shlukovací metody, a byly také zmíněny metody síťové analýzy, jež využívají grafových algoritmů. Nástroj, který s těmito metodami pracuje – Urban Analyst Network – byl také krátce charakterizován.

Objekty, jež jsou podrobovány analýze podobnosti, musí být jednoznačně charakterizovány uspořádanou n -ticí koeficientů, vyjadřujících hodnoty zkoumaných vlastností. Byla navržena metoda pro vyhledávání podobných objektů mezi daty územního plánování, jež využívá grafových algoritmů pro určování podobných skupin. Od ostatních běžně využívaných metod shlukové analýzy se odlišuje tím, že na jejím výstupu nemusí být zařazeny všechny vyhodnocované objekty do skupin. Uživatel si zvolí míru podobnosti p , jež vyžaduje, aby analyzovaná data vykazovala. Analyzované objekty jsou chápány jako body v n -rozměrném euklidovském prostoru, přičemž jejich souřadnice jsou určeny n -ticí koeficientů, které je charakterizují. Míra podobnosti mezi těmito objekty je pak určena za pomoci euklidovské vzdálenosti mezi nimi. Míra podobnosti p nabývá hodnot z intervalu $(0,1)$, přičemž $p = 0$ pro objekty s největší možnou vzdáleností a $p = 1$ pro totožné objekty. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných objektů.

Tuto metodu využívá navržený online nástroj s názvem *SimUrb* volně dostupný na adrese eyetracking.upol.cz/simurb.

Tento nástroj umožňuje analyzovat podobnost mezi objekty z jednoho či více souborů. V případě více souborů je zapotřebí, aby k nim byla vázána data od stejných subjektů.

Nástroj *SimUrb* byl použit v několika případových studiích. Ve studiích autorky *Hodnocení udržitelného rozvoje území* (kapitola 7.2.1) nebo Opacha a kol. (2017), či ve studiích jiných autorů, jako *Dratva (2018) Časoprostorové analýzy typu a stavu zemědělských plodin*. Zmíněné případové studie jsou náplní čtvrtého dílčího cíle.

SimUrb byl již použit při přípravě publikace:

Opach, T., Popelka, S., **Dolezalova, J.**, Rod, J. K. (2017) *Star and Polyline Glyphs in a Grid Plot and on a Map Display: Which Perform Better?*. Cartography and Geographic Information Science.

8.3 DC 3 – Využití grafů v dopravě

Třetí dílčí cíl byl zaměřen do oblasti dopravy. V práci byly zmíněny nejčastější přístupy k dopravním datům, tedy metody síťové analýzy a grafové algoritmy, jenž jsou v těchto analýzách aplikovány.

Práce se zaměřuje speciálně na data, která jsou charakterizována uspořádanou n -tici koeficientů, vyjadřujících hodnoty zkoumaných vlastností. V této práci jsou jimi hodnoty průjezdů autobusů mýtnými bránami. Byla navržena metoda pro vyhledávání podobných silničních úseků mezi těmito daty, využívající teorie grafů. Navržená metoda pro výpočet míry podobnosti vychází z metody používané nástrojem *SimRoad*. Jelikož však data z mýtných bran vykazují velkou hodnotu koeficientu *šikmost*, byla pro úpravu výpočtu použita kvantilová funkce. Míra podobnosti p ale stále nabývá hodnot z intervalu $(0,1)$, přičemž $p = 0$ pro objekty s největší možnou vzdáleností a $p = 1$ pro totožné objekty, v tomto případě pro silniční úseky. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných objektů.

Tuto metodu využívá navržený online nástroj s názvem *SimRoad* volně dostupný na adrese eyetracking.upol.cz/simroad.

Nástroj *SimRoad* byl použit v jedné případové studii a to pro analýzu toků veřejné dopravy na základě dat z mýtných bran. Případová studie je obsažena ve čtvrtém dílčím cíli.

8.4 DC 4 – Případové studie využívající vytvořené nástroje

Ve třech případových studiích, provedených v rámci disertační práce, byla aplikována teorie grafů do úloh z oblasti eye-trackingu, prostorového plánování a dopravy. V těchto studiích byly využívány navržené metody pro určení míry podobnosti a vyhledávání podobných skupin mezi objekty časoprostorových dat různého charakteru.

Výsledkem případové studie z oblasti eye-trackingu byly nalezené skupiny vzájemně si podobných strategií respondentů prohlížejících stimuly s různými vizualizacemi nejistoty v mapách. Experiment byl zaměřen na studium vizualizace nejistoty geografických dat. Byla vyhodnocována naměřená data 40 respondentů (20 studentů kartografie a 20 respondentů bez kartografického vzdělání) ze dvou stimulů. V prvním případě, kdy byl vyhodnocován stimulus zobrazující nejistotu možného výskytu třech živočišných druhů, byl nástroj *ScanGraph* nastaven na míru podobnosti $p = 0,8$, metodu Levenshtein distance a variantu collapsed. Nalezl tak 6 netriviálních klik, přičemž největší z nich obsahovala sekvence pěti respondentů. V druhém případě, kdy byl vyhodnocován stimulus nejistoty možného výskytu pouze jednoho živočišného druhu, byl nástroj *ScanGraph* nastaven na míru podobnosti $p = 0,7$ a opět metodu Levenshtein distance a variantu *Collapsed*. V tomto případě pak našel 13 netriviálních klik, s největší z nich o čtyřech prvcích.

Případová studie z oblasti prostorového plánování srovnávala výstupy z navrženého nástroje *SimUrb* s existujícími výsledky metodiky Olomouckého kraje hodnocení udržitelného rozvoje území. Vyhodnocována byla podobnost mezi obcemi olomouckého kraje na základě atributů rozdělených do tří kategorií (sociální pilíř, hospodářský pilíř a environmentální pilíř). Při použití metodiky Olomouckého kraje byly obce rozčleněny v rámci každého pilíře do šesti skupin (tři s pozitivní charakteristikou a tři s negativní) na základě hodnoty váženého průměru příslušných atributů. Jako celkové hodnocení byly obce rozčleněny ještě do osmi kategorií dle toho, zda byly v jednotlivých pilířích hodnoceny jako pozitivní či negativní. Pro adekvátní srovnání s nástrojem *SimUrb* bylo zvoleno nastavení tak, aby bylo vytvořeno také šest skupin v rámci jednotlivých pilířů a osm skupin v rámci celkového hodnocení. Pro sociální pilíř byl nástroj *SimUrb* na míru

podobnosti $p = 0,73$ a minimální velikost skupiny 15, pro hospodářský pilíř $p = 0,75$ a minimální velikost skupiny 16 a pro environmentální pilíř $p = 0,71$ a minimální velikost skupiny opět 15. Pro celkové hodnocení byla v nástroji *SimUrb* použita metoda pro analýzu více souborů najednou, tedy *Multifile* a výstupem bylo osm skupin. Při použití nástroje *SimUrb* nemusí být do skupin nutně zařazeny všechny vyhodnocované objekty. Pro srovnání tak byla vypočteno procentuální zastoupení obcí ze skupin nalezených nástrojem *SimUrb* v kategoriích určených metodikou ÚAP. Vypočtená byla také střední hodnota a směrodatná odchylka u koeficientů obcí dle jejich zařazení do skupin nástrojem *SimUrb*. Vizuální srovnání proběhlo také s výstupy dvou nástrojů dostupných skrze software ArcGIS Pro, a to *Grouping Analysis* a *Multivariate Clustering*. Byly sledovány případy, kdy obce byly metodikou ÚAP zařazeny do stejné kategorie, avšak nástroj *SimUrb* je rozčlenil do různých skupin. V takovýchto případech lze pozorovat, že ačkoliv průměr vážených hodnot atributů je u těchto obcí stejný, v jednotlivých hodnotách se od sebe výrazně odlišují.

Ve třetí případové studii, zaměřené na dopravu, byla vyhledávána podobnost mezi silničními úseky dálnic, silnic I. třídy a rychlostních silnic na základě dat z mytných bran. Ve studii byl využit nástroj *SimRoad* a navržená metoda, kterou nástroj používá. Ve studii byly vyhledávány silniční úseky s podobnou průjezdností autobusových dopravců během jednoho měsíce, přičemž byly srovnávány hodnoty průjezdů v pracovních dnech a o víkendech. Zvoleny byly dva přístupy, a to srovnání přes všechny úseky a srovnání pouze pro dálniční úseky. Nástroj *SimRoad* byl pokaždé nastaven na míru podobnosti $p = 0,8$. Výsledkem pak bylo sedm pozorování a to pro největší skupiny nalezené nástrojem *SimRoad* pro každý z přístupů, pozorování dálnice D01 mezi Prahou a Brnem a pozorování pro silniční úseky v okolí Prahy a Brna. Jednotlivé výstupy pro pracovní dny a víkendy se z hlediska zastoupení silničních úseků ve skupinách vytvořených nástrojem ve většině případů odlišovali jen v menším rozsahu. Z hlediska hodnoty průměrné průjezdnosti na den již odchylky byly pozorovatelné.

Kromě těchto tří případových studií byly metody a nástroje využity také v dalších případových studiích jiných autorů, které jsou v práci krátce představeny.

9 DISKUZE

V disertační práci byly navrženy tři metody pro vyhledávání podobných skupin různých objektů časoprostorových dat s využitím teorie grafů a tři nástroje, jež tyto metody používají. Dále byly vypracovány tři případové studie, na nichž byla funkcionality nástrojů prezentována.

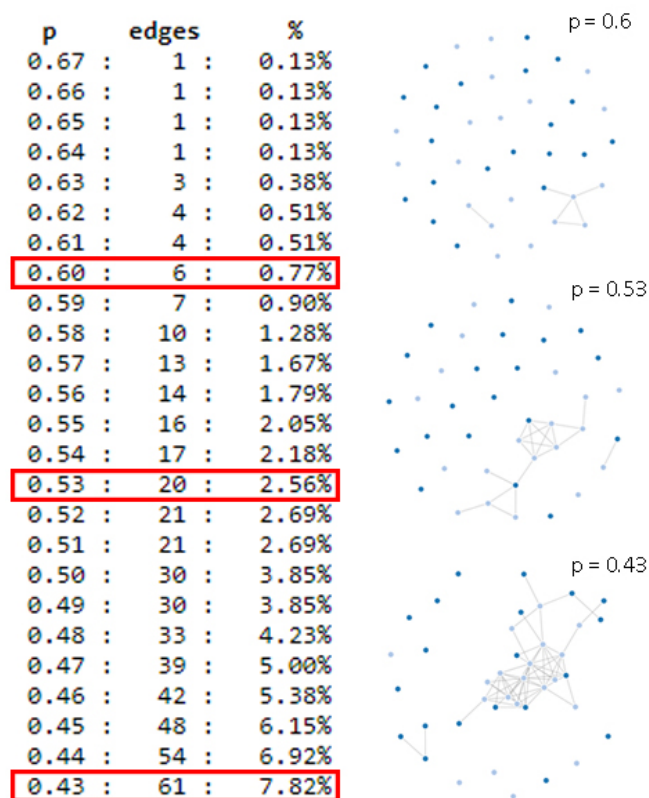
Vytvořené nástroje (*ScanGraph*, *SimUrb* a *SimRoad*) přistupují k problematice určování a vyhledávání podobných skupin obdobným způsobem. Přesto však každý z nich je uzpůsoben na vyhodnocování dat požadovaného charakteru jinými vlastnostmi.

Při zpracovávání disertační práce bylo zvažováno, zda všechny tři přístupy nezahrnout do jednoho nástroje, kde by pak uživateli bylo umožněno zvolit si příslušný charakter. Jelikož však jednotlivé nástroje mohou být používány pro zcela odlišné experimenty, mohlo by být sjednocení pod jeden nástroj pro uživatele spíše matoucí.

Jako první byla navržena metoda pro vyhodnocování podobnosti sekvencí navštívených oblastí zájmu v eye-tracking experimentech. Z toho důvodu také nástroj *ScanGraph*, jež tuto metodu používá, disponuje nejvíce funkcionalitami. Od svého vzniku byl aplikován v mnoha případových studiích, a tedy byl několikrát díky zpětné vazbě uživatelů rozšiřován.

Použití nástroje *ScanGraph* je přímočaré. Uživatel vytvoří oblasti zájmu, nahraje data do nástroje *ScanGraph* a zvolí míru podobnosti či hustotu hran grafu. Výstupem je pak seznam klik reprezentujících podobné strategie respondentů. Je zřejmé, že výsledek analýzy závisí na vytvořených oblastech zájmu a na zvolené míře podobnosti.

Je vždy výhodnější vyznačit oblasti zájmu okolo významnějších částí stimulu. V případě kartografie jsou to kompoziční prvky mapy. Nalezené podobnosti mezi sekvencemi pak budou mít vyšší vypovídající hodnotu než v případě pravidelné mřížky bez ohledu na obsah mapy.



Obr. 60 Vliv míry podobnosti p na výsledný graf. Uživatel může zobrazit počet hran pro každou možnou hodnotu míry podobnosti a na základě této informace zobrazit příslušný graf

Protože podobnost mezi sekvencemi respondentů závisí na pohybech očí a na vyznačených oblastech zájmu, nemůže být jednoznačně řečeno, které nastavení míry podobnosti může být označeno jako ideální. Mezi experimenty se dosažené míry podobnosti podstatně liší. Příhodnou cestou jak najít vhodnou míru podobnosti pro svůj experiment je nejprve zvolením tlačítka *Advised graph*. Tato funkce určí graf s 5% hustotou hran a jeho příslušnou míru podobnosti. Dle zkušeností s vyhodnocováním eye-tracking experimentů (Popelka, 2018) má takto zkonstruovaný graf vysokou vypovídající hodnotu. Po vykreslení grafu se zobrazí tabulka pro všechny možné míry podobnosti, jichž může podobnost nabývat. Nejsou uvedeny ty míry, pro které by množina hran grafu byla prázdná. Z této tabulky může uživatel vybrat a zobrazit další variantu grafu pro větší či menší míru podobnosti, dle potřeb jeho experimentu.

Cílem výzkumníka bývá nalézt skupiny s co největší mírou podobnosti, avšak mnohdy naměřená data takovou podobnost nevykazují a tedy ani není možné nalézt skupiny respondentů s velkou podobností. Je pak tedy vhodné zvolit tak vysokou míru podobnosti, jakou data umožňují, což uživatel může přehledně vybrat. Vliv míry podobnosti p na konstrukci grafu je znázorněn na obrázku 60.

Jak bylo zmíněno v kapitole 4.5, nástroj *ScanGraph* má své výpočetní limity. Testováním bylo zjištěno, že je možné určit podobnosti mezi 320 sekvencemi o délce 160 znaků s použitím Levenshteinovy vzdálenosti. Pro metody *Needleman-Wunsch algoritmus* a *Damerau-Levenshteinova vzdálenost* je mezní hodnotou pro funkční výpočet 80 sekvencí o délce 80 znaků. Většina eye-tracking experimentů však skýtá méně než 80 respondentů, tedy toto omezení nástroje není fatálně limitující. Komplikace však může způsobovat délka sekvencí, která při mnohých experimentech 80 znaků překročí. Tento problém může být částečně vyřešen volbou *collapsed varianty*. Nebudou se tak vyskytovat násobné znaky a sekvence budou kratší.

Snahou je optimalizovat výpočetní algoritmy tak, aby se tyto limitní hodnoty zvýšily. Dalším budoucím přínosem bude přidání funkcionality, která nalezne podobnosti sekvencí z experimentu k jiné uživatelem dané sekvenci. Praktickým použitím je selekce participantů, kteří k prohlížení stimulu zvolili správnou či naopak chybnou cestu.

Nástroj *SimUrb* neobsahuje tolik funkcionalit jako nástroj *ScanGraph*. Částečně je to dáno stářím nástroje (mladší než nástroj *ScanGraph*) a částečně jednodušším určením metriky (není potřeba volit si mezi metrikami či volbu *collapsed*). Pro použití nástroje *SimUrb* je nutné, aby jednotlivé objekty určené k vyhodnocování, byly charakterizovány uspořádanou n -ticí kvantitativních koeficientů. Uživatel nahraje takováto data do nástroje *SimUrb* a zvolí míru podobnosti a minimální velikost skupiny, která je pro něj zajímavá. Výstupem je pak seznam disjunktních skupin objektů, které vykazují stejnou či vyšší míru podobnosti požadovanou uživatelem. Výsledek analýzy tak závisí na volbě hodnoty míry podobnosti. Pokud má uživatel tu možnost, nahraje i mapový podklad, ve kterém se pak jednotlivé skupiny budou zobrazovat. Není možné stanovit pro všechny analýzy, které nastavení míry podobnosti má být voleno jako optimální. U každého jednotlivého vyhodnocování záleží na rozdělení dat a na cíli, jenž si uživatel stanovil.

Snahou uživatele bývá nalézt skupiny s co největší mírou podobnosti. Mnohdy však vyhodnocované objekty takovou podobnost nevykazují a tedy ani není možné nalézt skupiny s velkou podobností. Uživatel pak tedy musí explicitně vyzkoušet, které nastavení je pro jeho analýzu nejvhodnější. Čím nižší je hodnota míry podobnosti, tím větší „benevolence“ je dána objektům zařadit se do skupin.

Nástroj *SimUrb* vyhledává skupiny podobných objektů tak, aby byly disjunktní. Může se tedy stát, že množinu objektů, jež by také vykazovala požadovanou podobnost, přičemž by její prvky náležely již do jiných skupin, by zanedbal. Jelikož však nástroj při určování skupin dává přednost skupinám o větším počtu objektů, je tedy pravděpodobné, že menší skupiny, jež by byly zanedbány, by pro uživatele neměly přínosnou hodnotu.

Nástroj *SimRoad*, jakožto nejmladší z nástrojů, byl prověřen zatím pouze na jedné případové studii. Obdobně jako u nástroje *SimUrb* je u nástroje *SimRoad* nutné, aby jednotlivé objekty určené k vyhodnocování byly charakterizovány uspořádanou n -tici kvantitativních koeficientů. Avšak na rozdíl od nástroje *SimUrb* je metrika používaná nástrojem *SimRoad* uzpůsobena pro data, jejichž rozdělení má velkou hodnotu šikmosti. To znamená taková data, v nichž výrazně převládají objekty s velkou nebo malou vzdáleností.

Uživatel nahraje data v požadovaném formátu a má-li k dispozici, pak i mapový podklad, v němž pak výsledné skupiny vzájemně si podobných objektů budou vizualizovány.

Stejně jako u nástroje *SimUrb* se může stát, že nástroj zanedbá skupinu silničních úseků, jež sice vykazují podobnost zadanou uživatelem, avšak velikost jejich skupiny je menší, než u jiných nalezených skupin. Obdobně však lze říci, že tyto skupiny jsou pro uživatele méně zajímavé.

V nástroji *SimRoad* je největší možnost dalšího rozvoje. Vlivem toho, že nástroj byl vyzkoušen prozatím na jedné případové studii, postrádá jeho vývoj zpětnou vazbu uživatelů.

10 ZÁVĚR

Cílem disertační práce *Teorie grafů při vyhodnocování časoprostorových dat* bylo navrhnout metody pro vyhodnocování podobnosti mezi časoprostorovými daty a vyhledávání skupin, jež podobnost vykazují, a tyto metody aplikovat v případových studiích. Speciálně se disertační práce zaměřila na tři oblasti geoinformatiky. Jednalo se o zpracování eye-tracking dat, úlohy prostorového plánování a dopravu.

Pro vypracování disertační práce bylo nutné seznámit se s dosavadními metodami a programy používanými pro vyhodnocování podobnosti mezi daty tohoto charakteru a nalézt současné aplikace teorie grafů a grafových algoritmů při analýzách těchto dat.

Výstupem prvního dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů navržená speciálně pro eye-tracking data, tedy pro sekvence navštívených oblastí zájmu. Současně byl navržen a vytvořen nástroj *ScanGraph*, jenž tuto metodu používá. Nástroj je volně dostupný na adrese eyetracking.upol.cz/scangraph. Nástroj umožňuje analýzu dat nejen z jednoho stimulu, ale i z několika stimulů tvořících celý eye-tracking experiment. Nástroj byl využit v několika případových studiích a díky zpětné vazbě jeho uživatelů byl při svém vývoji obohacen o nové funkcionality.

Výstupem druhého dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů navržená speciálně pro data definovaná uspořádanou n -ticí kvantitativních koeficientů. Současně byl navržen a vytvořen nástroj *SimUrb*, jenž tuto metodu používá. Nástroj je volně dostupný na adrese eyetracking.upol.cz/simurb, který umožňuje kromě samotné analýzy dat vizualizovat nalezené skupiny v mapě. Nástroj je tedy aplikovatelný v úlohách prostorového plánování.

Výstupem třetího dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů dat z oblasti dopravy. Metoda je úzce vázána na metodu druhého dílčího cíle. Jelikož analyzovaná data z oblasti dopravy mají velkou hodnotu koeficientu šikmosti, byla metoda uzpůsobena tomuto charakteru pro názornější interpretovatelnost. Současně byl navržen nástroj *SimRoad* tuto metodu využívající. Nástroj je volně dostupný na adrese eyetracking.upol.cz/simroad a umožňuje výsledné skupiny vizualizovat také v mapě.

Náplní čtvrtého dílčího cíle bylo vypracování případových studií, v nichž se navržené metody aplikují. Z oblasti eye-trackingu byla vypracována studie na hodnocení podobnosti mezi strategiemi respondentů prohlížejícími stimuly s vizualizací nejistoty geografických dat. V oblasti prostorového plánování se studie zaměřila na srovnání s výstupem hodnocení udržitelného rozvoje území vypracovaného za pomoci metodiky ÚAP. Třetí případová studie se věnovala vyhledávání podobných silničních úseků z hlediska průjezdnosti autobusových dopravců mýtnými bránami na dálnicích, silních I. třídy a rychlostních silnicích.

Kromě případových studií autorky obsahuje disertační práce i ukázky ze studií jiných autorů, v nichž byly navržené metody a nástroje aplikovány. Všechny vymezené cíle práce tak byly naplněny.

Metody a nástroje navržené v této disertační práci jsou aplikovatelné v praxi a v průběhu psaní disertační práce již byly využity řadou uživatelů. Autorka věří, že si své příznivce budou nadále získávat a díky jejich referencím se budou nadále vyvíjet.

11 POUŽITÉ ZDROJE

- Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, 47(4), 1377–1392. <https://doi.org/10.3758/s13428-014-0550-3>
- Appel, K., & Haken, W. (1976). EVERY PLANAR MAP IS FOUR COLORABLE 1. *BULLETIN OF THE AMERICAN MATHEMATICAL SOCIETY*, 82(5).
- Arlinghaus, S. L., Arlinghaus, W. C., Harary, F., & Arlinghaus, S. L. (2002). *Graph Theory and Geography: An Interactive View* (Ebook).
- Bahill, A. T., & Stark, L. (1979). The Trajectories of Saccadic Eye Movements. *Scientific American*. Scientific American, a division of Nature America, Inc. <https://doi.org/10.2307/24965071>
- Beitlová, M. (2017). *Analýza kartografické gramotnosti vybraných skupin uživatelů map*. Olomouc. Univerzita Palackého v Olomouci.
- Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., & Freudenheim, J. L. (2003). Positional Accuracy of Geocoded Addresses in Epidemiologic Research. *Epidemiology*, 14(4), 408–412. <https://doi.org/10.1097/01.EDE.0000073121.63254.c5>
- Bron, C., & Kerbosch, J. (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9), 575–577. <https://doi.org/10.1145/362342.362367>
- Brus, J., Pechanec, V., & Kilianová, H. (2013). UNCERTAINTY VS. SPATIAL DATA QUALITY VISUALISATIONS: A CASE STUDY ON ECOTONES.
- Buhl, J., Gautrais, J., Reeves, N., Solé, R. V., Valverde, S., Kuntz, P., & Theraulaz, G. (2006). Topological patterns in street networks of self-organized urban settlements. *The European Physical Journal B*, 49(4), 513–522. <https://doi.org/10.1140/epjb/e2006-00085-1>
- Burian, J., Zajíčková, L., & Ivan, I. (2016). *Analýza dopravního chování obyvatel Olomouce a Ostravy*.
- Caddy, J., & Carocci, F. (1999). The spatial allocation of fishing intensity by port-based inshore fleets: a GIS application. *ICES Journal of Marine Science*, 56(3), 388–403. <https://doi.org/10.1006/jmsc.1999.0477>
- Cardillo, A., Scellato, S., Latora, V., & Porta, S. (2006). Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73(6), 066107. <https://doi.org/10.1103/PhysRevE.73.066107>
- Cayo, M. R., & Talbot, T. O. (2003). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics*, 2(1), 10. <https://doi.org/10.1186/1476-072X-2-10>
- Chang, N.-B., Lu, H. Y., & Wei, Y. L. (1997). GIS Technology for Vehicle Routing and Scheduling in Solid Waste Collection Systems. *Journal of Environmental Engineering*, 123(9), 901–910. [https://doi.org/10.1061/\(ASCE\)0733-9372\(1997\)123:9\(901\)](https://doi.org/10.1061/(ASCE)0733-9372(1997)123:9(901))
- Cheng, M.-Y., & Chang, G.-L. (2001). Automating utility route design and planning through GIS. *Automation in Construction*, 10(4), 507–516. [https://doi.org/10.1016/S0926-5805\(00\)00103-5](https://doi.org/10.1016/S0926-5805(00)00103-5)
- Cinnamon, J., Schuurman, N., & Crooks, V. A. (2008). A method to determine spatial access to specialized palliative care services using GIS. *BMC Health Services Research*, 8(1), 140. <https://doi.org/10.1186/1472-6963-8-140>
- Coffin, A. W. (2009). ROAD NETWORK DEVELOPMENT AND LANDSCAPE DYNAMICS IN THE SANTA FE RIVER WATERSHED, NORTH-CENTRAL FLORIDA, 1975 TO 2005.
- Crainic, T. G., & Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 97(3), 409–438. [https://doi.org/10.1016/S0377-2217\(96\)00298-6](https://doi.org/10.1016/S0377-2217(96)00298-6)
- Damerau, F. J., & J., F. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171–176. <https://doi.org/10.1145/363958.363994>

- Delavar, M. R., Samadzadegan, F., & Pahlavani, P. (2004). A GIS-Assisted Optimal Urban Route Finding Approach Based On Genetic Algorithms.
- Demel, J. (2002). *Grafy a jejich aplikace*. Academia.
- Derekenaris, G., Garofalakis, J., Makris, C., Prentzas, J., Sioutas, S., & Tsakalidis, A. (2001). Integrating GIS, GPS and GSM technologies for the effective management of ambulances. *Computers, Environment and Urban Systems*, 25(3), 267–278. [https://doi.org/10.1016/S0198-9715\(00\)00025-9](https://doi.org/10.1016/S0198-9715(00)00025-9)
- Dhamala, T. N. (2014). A survey on models and algorithms for discrete evacuation planning network problems. *Journal of Industrial and Management Optimization*, 11(1), 265–289. <https://doi.org/10.3934/jimo.2015.11.265>
- Dolezalova, J., & Popelka, S. (2016). Evaluation of the user strategy on 2D and 3D city maps based on novel scanpath comparison method and graph visualization. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*.
- Dorigo, M., & Gambardella, L. M. (1997). Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1), 53–66. <https://doi.org/10.1109/4235.585892>
- Dratva, P. (2018). ČASOPROSTOROVÉ ANALÝZY TYPU A STAVU ZEMĚDĚLSKÝCH PLODIN Z DAT SENTINEL-2. Olomouc. Univerzita Palackého v Olomouci.
- Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). Scanpath comparison revisited. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10* (p. 219). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1743666.1743719>
- Fan, D., & Shi, P. (2010). Improvement of Dijkstra's algorithm and its application in route planning. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 1901–1904). IEEE. <https://doi.org/10.1109/FSKD.2010.5569452>
- Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach.
- Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, 1(4), 620–630. <https://doi.org/10.1111/j.1752-4571.2008.00047.x>
- Gavish, B., & Graves, S. C. (1978). The Travelling Salesman Problem and Related Problems.
- Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '02* (p. 51). New York, New York, USA: ACM Press. <https://doi.org/10.1145/507072.507082>
- Granot, D., & Hamers, H. (2004). On the equivalence between some local and global Chinese postman and traveling salesman graphs. *Discrete Applied Mathematics*, 134(1–3), 67–76. [https://doi.org/10.1016/S0166-218X\(03\)00225-7](https://doi.org/10.1016/S0166-218X(03)00225-7)
- Hájková, M. (2016). Eye tracking vyšetření predilekce očních pohybů u pacientů po cévní mozkové příhodě. Olomouc. Univerzita Palackého v Olomouci.
- Handy, S. L., & Niemeier, D. A. (1997). Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A*, 29(7), 1175–1194. <https://doi.org/10.1068/a291175>
- Herman, L., Řezník, T., Stachoň, Z., & Rusznák, J. (2018). Experimental Tool for Usability Testing of Interactive 3D Maps: Requirements, Design, Implementation, and Evaluation. *Cartographic Perspectives, in print*.
- Horák, J., Ivan, I., Fojtík, D., Inspektor, T., Zajíčková, L., & Voženílek, V. (2014). DOSTUPNOST VEŘEJNOU LINKOVOU DOPRAVOU V ČR.
- Jerby, S., & Ceder, A. (2006). Optimal Routing Design for Shuttle Bus Service. *Transportation Research Record: Journal of the Transportation Research Board*, 1971, 14–22. <https://doi.org/10.3141/1971-04>

- Jindra, M. (2016). Analýza dopravní dostupnosti veřejných institucí v Moravských krajích.
- Khodr, H. M., Melian, J. A., Quiroz, A. J., Picado, D. C., Yusta, J. M., & Urdaneta, A. J. (2003). A probabilistic methodology for distribution substation location. *IEEE Transactions on Power Systems*, 18(1), 388–393. <https://doi.org/10.1109/TPWRS.2002.807088>
- Kim, H., Lee, J., YS Jeong - US Patent 6, 564,145, & 2003, undefined. (n.d.). Method for finding shortest path to destination in traffic network using Dijkstra algorithm or Floyd-warshall algorithm. *Google Patents*.
- Kimplel, T. J., Dueker, K. J., & El-Geneidy, A. M. (2007). Using GIS to Measure the Effect of Overlapping Service Areas on Passenger Boardings at Bus Stops. *Journal of the Urban and Regional Information Systems Association*, 19(1).
- Kong, F., Yin, H., Nakagoshi, N., & Zong, Y. (2010). Urban green space network development for biodiversity conservation: Identification based on graph theory and gravity modeling. *Landscape and Urban Planning*, 95(1–2), 16–27. <https://doi.org/10.1016/J.LANDURBPLAN.2009.11.001>
- Kubiček, P., Šašinka, Č., & Stachoň, Z. (2012). Uncertainty visualization testing. Sofia, Bulgaria.
- Kučera, M. (2016). *Uživatelské testování a optimalizace vizualizací nejistoty prostorových dat*. Olomouc. Univerzita Palackého v Olomouci.
- Kunderová, P. (1997). Úvod do teorie pravděpodobnosti a matematické statistiky. Olomouc. Univerzita Palackého v Olomouci.
- Laporte, G., Asef-Vaziri, A., & Sriskandarajah, C. (1996). Some Applications of the Generalized Travelling Salesman Problem. *Journal of the Operational Research Society*, 47(12), 1461–1467. <https://doi.org/10.1057/jors.1996.190>
- Laporte, G., & Martello, S. (1990). The selective travelling salesman problem. *Discrete Applied Mathematics*, 26(2–3), 193–207. [https://doi.org/10.1016/0166-218X\(90\)90100-Q](https://doi.org/10.1016/0166-218X(90)90100-Q)
- Lee, J. (2004). A Spatial Access-Oriented Implementation of a 3-D GIS Topological Data Model for Urban Entities. *GeoInformatica*, 8(3), 237–264. <https://doi.org/10.1023/B:GEIN.0000034820.93914.d0>
- Lee, J. (2009). GIS-based geocoding methods for area-based addresses and 3D addresses in urban areas. *Environment and Planning B: Planning and Design*, 36(1), 86–106. <https://doi.org/10.1068/b31169>
- Levenshtein, V. I., & I., V. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady, Vol. 10, p.707, 10, 707*.
- Liu, L., & Meng, L. (2009). Algorithms of Multi-Modal Route Planning Based on the Concept of Switch Point. *Photogrammetrie - Fernerkundung - Geoinformation*, 2009(5), 431–444. <https://doi.org/10.1127/1432-8364/2009/0031>
- Liu, S., & Zhu, X. (2004). Accessibility Analyst: An Integrated GIS Tool for Accessibility Analysis in Urban Transportation Planning. *Environment and Planning B: Planning and Design*, 31(1), 105–124. <https://doi.org/10.1068/b305>
- Longley, P. (2005). *Geographical information systems and science*. Wiley.
- MacEachren, A., Roth, R., ... J. O.-... on V. and, & 2012, U. (2012). Visual semiotics & uncertainty visualization: An empirical study. *Ieeexplore.Ieee.Org*.
- Mackanness, W. A., & Beard, K. M. (1993a). Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Systems*, 20(4), 210–221. <https://doi.org/10.1559/152304093782637479>
- Mackanness, W. A., & Beard, K. M. (1993b). Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Systems*, 20(4), 210–221. <https://doi.org/10.1559/152304093782637479>
- Malandraki, C., & Daskin, M. S. (1993). The maximum benefit Chinese postman problem and the maximum benefit traveling salesman problem. *European Journal of Operational Research*, 65(2), 218–234. [https://doi.org/10.1016/0377-2217\(93\)90335-K](https://doi.org/10.1016/0377-2217(93)90335-K)

- Minieka, E. (1979). The Chinese Postman Problem for Mixed Networks. *Management Science*, 25(7), 643–648. <https://doi.org/10.1287/mnsc.25.7.643>
- Mosheiov, G. (1994). The Travelling Salesman Problem with pick-up and delivery. *European Journal of Operational Research*, 79(2), 299–310. [https://doi.org/10.1016/0377-2217\(94\)90360-3](https://doi.org/10.1016/0377-2217(94)90360-3)
- Nara, K. (1997). A new algorithm for distribution feeder expansion planning for urban area. In *APSCOM-97. International Conference on Advances in Power System Control, Operation and Management* (Vol. 1997, pp. 192–197). IEE. <https://doi.org/10.1049/cp:19971829>
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453. [https://doi.org/10.1016/0022-2836\(70\)90057-4](https://doi.org/10.1016/0022-2836(70)90057-4)
- Neumann, J. (1994). The Topological Information Content of a Map / An Attempt At A Rehabilitation Of Information Theory In Cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 31(1), 26–34. <https://doi.org/10.3138/U626-551H-64K4-9687>
- Nobert, Y., & Picard, J.-C. (1996). An optimal algorithm for the mixed Chinese postman problem. *Networks*, 27(2), 97–108. [https://doi.org/10.1002/\(SICI\)1097-0037\(199603\)27:2<97::AID-NET1>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-0037(199603)27:2<97::AID-NET1>3.0.CO;2-8)
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11(9), 929–IN8. [https://doi.org/10.1016/0042-6989\(71\)90213-6](https://doi.org/10.1016/0042-6989(71)90213-6)
- Opach, T., Popelka, S., Dolezalova, J., & Rød, J. K. (2017). Star and polyline glyphs in a grid plot and on a map display: which perform better? *Cartography and Geographic Information Science*, 1–20. <https://doi.org/10.1080/15230406.2017.1364169>
- Pitts, F. R. (1965). A GRAPH THEORETIC APPROACH TO HISTORICAL GEOGRAPHY. *The Professional Geographer*, 17(5), 15–20. https://doi.org/10.1111/j.0033-0124.1965.015_m.x
- Popelka, S. (2018). *Eye-tracking (nejen) v kognitivní kartografii – praktický průvodce tvorbou a vyhodnocením experimentu*. Olomouc. Univerzita Palackého v Olomouci.
- Pradhan, A., & Mahinthakumar, G. (Kumar). (2013). Finding All-Pairs Shortest Path for a Large-Scale Transportation Network Using Parallel Floyd-Warshall and Parallel Dijkstra Algorithms. *Journal of Computing in Civil Engineering*, 27(3), 263–273. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000220](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000220)
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9), 970–982. <https://doi.org/10.1109/34.877520>
- Pulkrtová, T. (2016). *Vliv červené barvy na vnímání atraktivity žen*. Brno. Masarykova Univerzita.
- Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science*, 18(1), 61–72. <https://doi.org/10.1080/13658810310001596076>
- Rathore, M. M., Ahmad, A., Paul, A., & Jeon, G. (2015). Efficient Graph-Oriented Smart Transportation Using Internet of Things Generated Big Data. In *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 512–519). IEEE. <https://doi.org/10.1109/SITIS.2015.121>
- Reinelt, G. (1991). TSPLIB—A Traveling Salesman Problem Library. *ORSA Journal on Computing*, 3(4), 376–384. <https://doi.org/10.1287/ijoc.3.4.376>
- Roberts, F. S. (1978). *Graph Theory and Its Applications to Problems of Society* - Fred S. Roberts - Knihy Google.
- Roubalík, Š. (2017). Analýza toků veřejné dopravy na základě dat z mýtných bran.
- Rybansky, M. (2014). Modelling of the optimal vehicle route in terrain in emergency situations using GIS data. *IOP Conference Series: Earth and Environmental Science*, 18(1), 012131. <https://doi.org/10.1088/1755-1315/18/1/012131>
- Saalfeld, A. (1995). Map generalization as a graph drawing problem (pp. 444–451). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-58950-3_398

- Sathyaraj, B. M., Jain, L. C., Finn, A., & Drake, S. (2008). Multiple UAVs path planning algorithms: a comparative study. *Fuzzy Optimization and Decision Making*, 7(3), 257–267. <https://doi.org/10.1007/s10700-008-9035-0>
- Sevtsuk, A., & Mekonnen, M. (2012). Urban network analysis A new toolbox for ArcGIS, 287–305.
- Snopková, D. (2016). *Tvorba a užití map osobami se sníženou schopností rozpoznání barev*. Brno. Masarykova Univerzita.
- Soltani, A. R., Tawfik, H., Goulermas, J. Y., & Fernando, T. (2002). Path planning in construction sites: performance evaluation of the Dijkstra, A*, and GA search algorithms. *Advanced Engineering Informatics*, 16(4), 291–303. [https://doi.org/10.1016/S1474-0346\(03\)00018-1](https://doi.org/10.1016/S1474-0346(03)00018-1)
- Taylor, M. A. P., Sekhar, S. V. C., & D'Este, G. M. (2006). Application of Accessibility Based Methods for Vulnerability Analysis of Strategic Road Networks. *Networks and Spatial Economics*, 6(3–4), 267–291. <https://doi.org/10.1007/s11067-006-9284-9>
- Treml, E. A., Halpin, P. N., Urban, D. L., & Pratson, L. F. (2008). Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology*, 23(S1), 19–36. <https://doi.org/10.1007/s10980-007-9138-y>
- Tuček, J., & Ján. (1998). *GIS - Geografické informační systémy: principy a praxe*. Computer Press.
- Upchurch, C., Kuby, M., Zoldak, M., & Barranda, A. (2004). Using GIS to generate mutually exclusive service areas linking travel on and off a network. *Journal of Transport Geography*, 12(1), 23–33. <https://doi.org/10.1016/J.JTRANGE0.2003.10.001>
- Vrkoč, J. (2014). *Testování využitelnosti nástroje Urban Network Analyst v případových studiích*. Olomouc. Univerzita Palackého v Olomouci.
- Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., & González, M. C. (2012). Understanding road usage patterns in urban areas. *Scientific Reports*, 2. <https://doi.org/10.1038/srep01001>
- West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). eyePatterns. In *Proceedings of the 2006 symposium on Eye tracking research & applications - ETRA '06* (p. 149). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1117309.1117360>
- Wilson, R. (2002). Four Colors Suffice: How the Map Problem Was Solved - Robin Wilson - Knihy Google.
- Wolff, A. (2013). Graph Drawing and Cartography. *Handbook of Graph Drawing and Visualization*, 697–736.
- Yeh, A. G.-O., & Chow, M. H. (1996). An integrated GIS and location-allocation approach to public facilities planning—An example of open space planning. *Computers, Environment and Urban Systems*, 20(4–5), 339–350. [https://doi.org/10.1016/S0198-9715\(97\)00010-0](https://doi.org/10.1016/S0198-9715(97)00010-0)
- Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010a). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, 95(4), 181–191. <https://doi.org/10.1016/J.LANDURBPLAN.2010.01.002>
- Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010b). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, 95(4), 181–191. <https://doi.org/10.1016/J.LANDURBPLAN.2010.01.002>

ANNOTATION

The dissertation thesis is focused on the implementation of graph theory in geoinformatics problems. The main aim of the thesis is to propose tools and methods for finding similarity classes using graph theory. The methods are designed to fit specific data types. The tools are then used in case studies developed by the author, focusing on the evaluation of cartographic works by eye-tracking, analyzing the data of territorial analytical data, and analyzing the road sections. There are also given cases when the tools were used in case studies of other authors.

The first part of the thesis evaluates the relation between simple graphs and spatio-temporal data, introduces basic concepts from graph theory and describes the approaches to evaluation of time-spatial data from eye-tracking data, spatial planning and transport.

The first aim is focused on the analysis of time-space data from eye-tracking. The result is the creation of a new method for searching similar eye-tracking data sequences and the tool that uses this method. The uniqueness of the method consists in taking into account the length of the individual sequences to determine the degree of similarity between them.

The focus of the second aim is directed to spatial planning issues. As a result, it suggests a method for searching similar municipalities based on predefined properties using graph theory, which, however, does not necessarily include all objects in comparison to standard clustering methods. The advantage of this approach is to remove objects (communities) with extreme properties.

The third aim focuses on the analysis of traffic data. The result is a proposal for a method to search for similar road sections and a tool that uses this method. The similarity of road sections is evaluated on the basis of data on the passage of vehicles obtained from toll gates.

The purpose of the fourth aim is to produce three case studies with real data in which the above methods are applied. The first study, which is the application of graph theory in eye-tracking data, touches the visualization of uncertainty of geographic data. The second case study, as a graphical application of spatial planning, deals with the assessment of sustainable land development. The third study applies simple graphs to traffic data and evaluates data from toll gates for road sections for buses. Designed methods and created tools are also applied to case studies of other authors.

Keywords:

Graph Theory, Cliques in a Graph, Similarity, Cluster Analysis

SUMMARY

The aim of the dissertation Graph theory in analyses of spatio-temporal data was to propose methods for evaluating the similarity between spatio-temporal data and the search of similarity groups and applying these methods in case studies. In particular, the dissertation focused on three areas of geoinformatics. These were eye-tracking data processing, spatial planning and transport tasks.

For the dissertation thesis it was necessary to get acquainted with the existing methods and programs used for evaluation of similarity between data of this character and to find current application of the theory of graphs and graph algorithms in analyzes of these data.

The output of the first aim dissertation thesis was a method for determining the degree of similarity and searching for groups of similar objects using graph theory designed specifically for eye-tracking data, ie for sequences of visited areas of interest. At the same time, ScanGraph was designed and created using this method. The tool is freely available at eyetracking.upol.com/scangraph. The tool allows data analysis not only from one stimulus but also from several stimuli that make up the entire eye-tracking experiment. The tool has been used in several case studies and thanks to feedback from its users it has been enriched in its development with new functionality.

The output of the second partial goal of the dissertation was a method for determining the degree of similarity and searching for groups of similar objects using graph theory designed specifically for the data defined by an ordered set of quantitative coefficients. At the same time, SimUrb was designed and created using this method. The tool is freely available at eyetracking.upol.com/simurb, which allows you to visualize the groups found on the map in addition to the data analysis itself. The tool is applicable in spatial planning tasks.

The output of the third aim dissertation thesis was the method for determining the degree of similarity and searching for groups of similar objects using the theory of transport data graphs. The method is closely linked to the second aim method. However, since the traffic data analyzed have a high value of slope coefficient, the method has been adapted to this character for more readable interpretation. At the same time, SimRoad was designed to use this method. The tool is freely available at eyetracking.upol.com/simroad, which also allows you to visualize the resulting group on the map.

The purpose of the fourth aim was to develop case studies in which the proposed methods are applied. From the field of eye-tracking, a study was conducted to assess the similarity between respondent strategies by viewing the incentives with the visualization of the uncertainty of geographic data. In the sphere of spatial planning, the study focused on the comparison with the outcome of the Sustainable Development Assessment of the Territory. The third case study focused on the search for similar road sections in terms of the accessibility of bus carriers through toll gates on motorways and highways.

Apart from the case studies of the author, the dissertation includes examples from studies of other authors in which the proposed methods and tools were applied.

The methods and tools proposed in this dissertation are applicable in practice and have already been used by a number of users during the dissertation. The author believes that they will continue to earn their fans and, thanks to their references, will continue to develop.

PŘÍLOHY

SEZNAM PŘÍLOH

Volné přílohy:

1× CD

Struktura CD:

Data_ScanGraph	- Data z případové studie
Data_SimRoad	- Data z případové studie
Data_SimUrb	- Data z případové studie
Nastroje	- Zdrojové soubory vytvořených nástrojů
Dolezalova.pdf	- Text disertační práce
Priloha1.jpg	- Územní podmínky pro soudržnost území
Priloha2.jpg	- Územní podmínky pro hospodářský rozvoj
Priloha3.jpg	- Územní podmínky pro příznivé životní prostředí
Priloha4.jpg	- Celkové hodnocení rozboru udržitelného rozvoje území



KATEDRA GEOINFORMATIKY

Univerzita Palackého v Olomouci | Přírodovědecká fakulta

TEORIE GRAFŮ PŘI VYHODNOCOVÁNÍ ČASOPROSTOROVÝCH DAT

AUTOREFERÁT DISERTAČNÍ PRÁCE

Studijní program: P1314 Geografie

Obor studia: 1302V011 Geoinformatika a kartografie

Školitel: prof. Mgr. Radomír Halaš, Dr.

Mgr. Jitka Doležalová

GRAPH THEORY IN ANALYSES OF SPATIO-TEMPORAL DATA

Ph.D. THESIS SUMMARY

Study Programme: Geography

Specialization: Geoinformatics and Cartography

Supervisor: prof. Radomír Halaš, Ph.D.

**Department of Geoinformatics
Faculty of Science, Palacký University Olomouc**

Olomouc 2018

Disertační práce byla vypracována v prezenční formě doktorského studia na Katedře geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci.

Dissertation thesis was compiled within Ph.D. study at the Department of Geoinformatics, Faculty of Science, Palacký University Olomouc.

Předkladatel / Submitter:

Mgr. Jitka Doležalová

Školitel / Supervisor:

prof. Mgr. Radomír Halaš, Dr.

Katedra algebry a geometrie

Přírodovědecká fakulta Univerzity Palackého v Olomouci

17. listopadu 12

771 46 Olomouc

Oponenti / Opponents:

doc. Ing. Jitka Komárková, PhD. (Univerzita Pardubice)

doc. Ing. Peter Fabián, PhD. (Žilinská univerzita)

doc. Ing. František Staněk, CSc. (VŠB-Technická Univerzita Ostrava)

Autoreferát byl rozeslán dne 30. června 2018 / Summary was posted on June 30, 2018.

Obhajoba disertační práce se koná dne 6. září 2018 před komisí pro obhajoby disertačních prací doktorského studia v oboru P1314 Geografie, studijním oboru 1302V011 Geoinformatika a kartografie, v prostorách Katedry geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci, 17. listopadu 50, 771 46 Olomouc.

The defence of the dissertation thesis will be held on September 6, 2018 at the commission for the defence of dissertation thesis of Ph.D. degree in study programme P1314 Geography, specialization Geoinformatics and cartography, in the premises of the Department of Geoinformatics, Faculty of Science, Palacký University Olomouc, 17. listopadu 50, 771 46 Olomouc.

S disertační prací je možno se seznámit na studijním oddělení Přírodovědecké fakulty Univerzity Palackého v Olomouci, 17. listopadu 12, 77 46 Olomouc.

The dissertation thesis is available at the Study Department, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 771 46 Olomouc.

© Jitka Doležalová, 2018

© Univerzita Palackého v Olomouci, 2018

ISSN 1805-7500

ISBN 978-80-244-5330-9

Obsah

1. Abstrakt.....	4
2. Cíle práce a postup řešení	5
3. Metody a postup zpracování	8
4. Současný stav problematiky.....	11
5. SCANGRAPH – teorie grafů v analýze eye-tracking dat.....	12
6. SIMURB – teorie grafů v prostorovém plánování	17
7. SIMROAD – teorie grafů v dopravě	19
8. Případové studie využívající vytvořené nástroje.....	22
9. Výsledky	30
10. Diskuze	32
11. Závěr	35
12. Použité zdroje.....	36
Odborný životopis autora	38
Seznam vybraných publikací autora souvisejících s disertační prací	39
13. Annotation.....	41
14. Summary	42

1. ABSTRAKT

Disertační práce je zaměřena na implementaci teorie grafů do problematiky geoinformatických úloh. Hlavním cílem práce je navržení nástrojů a metod pro hledání tříd podobnosti využívající teorii grafů. Metody jsou vytvořeny tak, aby odpovídaly specifickým charakterům dat. Jednotlivé nástroje jsou pak použity v případových studiích vypracovaných autorkou, zaměřených na hodnocení kartografických děl pomocí eye-trackingu, na analýzu dat územně analytických podkladů a na analýzu průjezdnosti silničních úseků. Uvedeny jsou i příklady, kdy byly nástroje využity v případových studiích jiných autorů.

V první části práce je zhodnocen vztah obyčejných grafů a časoprostorových dat, jsou zavedeny základní pojmy z teorie grafů a jsou popsány dosavadní přístupy k vyhodnocování časoprostorových dat z oblasti eye-tracking dat, územního plánování a dopravy.

První dílčí cíl je zaměřen na analýzu časoprostorových dat z oblasti eye-trackingu. Jeho výsledkem je vytvoření nové metody pro vyhledávání podobných sekvencí eye-tracking dat a nástroje, jež tuto metodu využívá. Jedinečnost metody spočívá zejména v zohledňování délky jednotlivých sekvencí pro určení míry podobnosti mezi nimi.

Zaměření druhého dílčího cíle je směřováno do problematiky prostorového plánování. Jeho výsledkem je navržení metody pro vyhledávání podobných obcí na základě předem definovaných vlastností využívající teorie grafů, která však oproti standardním shlukovacím metodám nezařazuje nutně do podobných skupin všechny objekty. Výhodou tohoto přístupu je vyřazení objektů (obcí) s extrémními vlastnostmi.

Třetí dílčí cíl se zaměřuje na analýzu dat dopravních úloh. Jeho výsledkem je návrh metody pro vyhledávání podobných silničních úseků a nástroj tuto metodu využívající. Podobnost silničních úseků je vyhodnocována na základě dat o průjezdech vozidel získaných z mýtných bran.

Náplní čtvrtého dílčího cíle je vypracování tří případových studií s reálnými daty, v nichž se výše uvedené metody aplikují. První studie, jež je aplikací teorie grafů v eye-tracking datech, se dotýká vizualizace nejistoty geografických dat. Druhá případová studie, jakožto aplikace grafů v prostorovém plánování, se zabývá hodnocením udržitelného rozvoje území. Třetí studie aplikuje grafy do dopravních dat a vyhodnocuje data z mýtných bran silničních úseků pro autobusy. Navržené metody a vytvořené nástroje jsou následně aplikovány i v případových studiích jiných autorů.

Klíčová slova:

Teorie grafů, kliky v grafu, podobnost, shluková analýza.

2. CÍLE PRÁCE A POSTUP ŘEŠENÍ

Hlavním cílem práce je navrhnout metody pro hledání tříd podobnosti využívající obyčejné grafy, vytvořit nástroje využívající tyto metody a ověřit jejich funkčnost při aplikaci v časoprostorových úlohách v dopravě, urbanismu (prostorovém plánování) a hodnocení kartografických děl pomocí eye-trackingu. V dalším textu práce je používán více pojem prostorové plánování než urbanismus, jelikož blíže vystihuje zaměření výpočetních metod.

V první části práce bude zhodnocen vztah obyčejných grafů a časoprostorových dat a budou zavedeny základní pojmy z teorie grafů pro potřeby tří vytvořených nástrojů. Tyto nástroje budou aplikovány ve výzkumných případových studiích.

Následně budou vytvořeny webové nástroje umožňující hledání skupin vzájemně si podobných sekvencí v časoprostorových datových sadách. V první části bude nástroj využit pro analýzu dat o pohybu očí při čtení (nejen) kartografických děl a uzpůsoben pro co nejjednodušší používání.

V další části řešení disertační práce bude vytvořený nástroj transformován pro využití k řešení časoprostorových úloh v prostorovém plánování, kde bude možné nalézt skupiny podobných obcí na základě předem definovaných vlastností, a také v dopravě, kde budou hledány úseky dopravní sítě s podobnou průjezdností.

Výstupem disertační práce budou tři vytvořené nástroje pro hledání tříd podobnosti. Práce přinese relevantní výsledky případových studií, ve kterých budou nástroje aplikovány na konkrétní data ze tří tematických oblastí (hodnocení kartografických děl (pomocí eye-trackingu), prostorové plánování, doprava). Výsledky těchto studií budou obsahovat okomentovaný popis vztahů mezi sledovanými entitami, statistická hodnocení a mapy.

Cíle disertační práce jsou rozděleny na čtyři dílčí části:

- DC1** – Využití grafů v analýze eye-tracking dat
- DC2** – Využití grafů v prostorovém plánování
- DC3** – Využití grafů v dopravě
- DC4** – Případové studie využívající vytvořené nástroje

Náplní **prvního dílčího cíle** je návrh metody Scanpath Comparison využívající teorii grafů pro analýzu podobnosti v eye-tracking datech. Tato metoda je následně implementována při tvorbě webového nástroje pro časoprostorovou analýzu eye-tracking dat. Tento nástroj, s názvem ScanGraph, vyhledává podobné uživatelské strategie čtení mapy na základě porovnávání řetězců navštívených oblastí zájmu.

Požadavky na funkcionalitu nástroje:

- import dat přímo z prostředí programu eye-tracking softwaru
- stanovení míry podobnosti uživatelem nástroje

- nalezení všech skupin vzájemně si podobných řetězců oblastí zájmu při dané míře podobnosti
- možnost exportu souboru s výpočty a původními daty
- možnost permanentního odkazu na výsledek analýzy
- možnost analýzy více stimulů najednou.

Náplní **druhého dílčího cíle** je návrh metody využívající teorii grafů pro hledání podobných obcí na základě jejich kvantitativních indikátorů. Tato metoda je následně implementována při tvorbě webového nástroje SimUrb pro analýzu časoprostorových dat v urbánních úlohách, která budou ve formátu uspořádané n -tice ohodnocených indikátorů přiřazených jednotlivým sídelním jednotkám. Nástroj vyhledává disjunktní skupiny sídelních jednotek, které vykazují míru podobnosti danou uživatelem.

Požadavky na funkcionalitu nástroje:

- import dat jako CSV
- určení disjunktních skupin vzájemně si podobných sídelních jednotek při dané míře podobnosti
- stanovení míry podobnosti uživatelem
- rozlišení sídelních jednotek podle nadregionů (např. ORP, okresy, atd.)
- stanovení nejmenší velikosti skupiny, která je pro uživatele zajímavá
- možnost exportu souboru s výpočty a původními daty
- možnost permanentního odkazu na výsledek analýzy
- interaktivní vyznačení podobné skupiny v mapě

Náplní **třetího dílčího cíle** je návrh metody využívající teorii grafů pro hledání disjunktních skupin silničních úseků vykazujících danou míru podobnosti z hlediska počtu průjezdů vozidel. Následně byl vytvořen webový nástroj SimRoad, využívající tuto nově vytvořenou metodu. Nástroj je použitelný zejména pro analýzu průjezdnosti dálničních úseků a rychlostních silnic na základě dat z mýtných bran.

Požadavky na funkcionalitu nástroje:

- import dat jako CSV
- určení disjunktních skupin vzájemně si podobných silničních úseků při dané míře podobnosti
- stanovení míry podobnosti uživatelem
- rozlišení silničních úseků podle označení (např. D1, D46 atd.)
- stanovení nejmenší velikosti skupiny, která je pro uživatele zajímavá
- možnost exportu souboru s výpočty a původními daty
- možnost permanentního odkazu na výsledek analýzy
- interaktivní vyznačení podobných úseků v mapě

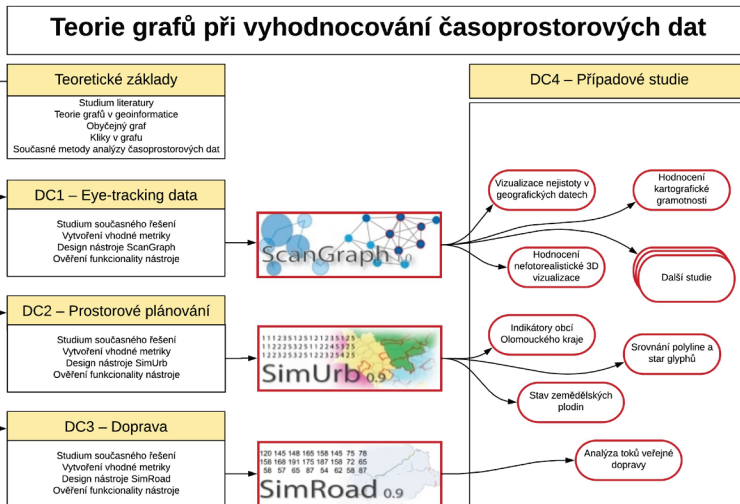
V předchozích třech dílčích cílech byly navrženy metody pro časoprostorovou analýzu dat ze tří oblastí a následně vytvořeny webové nástroje tyto metody využívající. **Čtvrtým dílčím cílem** je aplikace těchto nově vytvořených metod a nástrojů ve třech případových studiích.

- nástroj ScanGraph je použit v případové studii zaměřené vizualizaci nejistoty v geografických datech
- nástroj SimUrb je použit v případové studii zaměřené na hodnocení podobnosti obcí v Olomouckém kraji na základě tří tematických skupin indikátorů a porovnání s dosavadním způsobem hodnocení podobnosti obcí
- nástroj SimRoad bude použit v případové studii zaměřené na časoprostorovou analýzu podobnosti dálničních úseků na základě porovnání počtu průjezdů hromadné dopravy (autobusů) mýtnými branami

Navržené metody a nástroje byly využity i v dalších případových studiích jiných autorů. Nástroj ScanGraph byl využit například při hodnocení nefotorealistické 3D vizualizace měst či v hodnocení kartografické gramotnosti. Nástroj SimUrb byl využit například při časoprostorové analýze typu a stavu zemědělských plodin či při porovnání polyline a star glyphů. Nástroj SimRoad, jakožto nejmladší z těchto nástrojů, nebyl zatím aplikován jinými autory.

3. METODY A POSTUP ZPRACOVÁNÍ

Pro vývoj tří nástrojů vymezených v cílech disertační práce byla použita programová prostředí PHP, C# a D3.js. Každá z metod vyžaduje jiné přístupy a využívá jiné algoritmy.



Obr. 1 Schéma postupu práce

DC1 – Využití grafů v analýze eye-tracking dat

Řešení prvního dílčího cíle bylo zaměřeno na analýzu dat z eye-tracking experimentů. Pro hledání podobných sekvencí v eye-tracking datech byla vytvořena metoda ze skupiny String Edit Distance. Principem této metody je hledání podobností v sekvencích pohledem navštívených oblastí zájmu. Teoreticky by bylo možné hodnotit přímo podobnost vlastních sekvencí fixací a sakád, avšak výsledky takového srovnání by byly v praxi jen těžko použitelné. Proto je výhodnější agregovat eye-tracking data k oblastem zájmu odpovídajícím logickým částem zobrazovaného stimulu (v případě mapy například kompozičním prvkům).

Analýzovaná eye-tracking data byla naměřena pomocí zařízení SMI RED 250 a následně pomocí nástroje SMI2OGAMA převedena do prostředí open-source aplikace OGAMA, která slouží pro záznam a analýzu eye-tracking a mouse-tracking dat. Eye-tracking data by bylo možné zaznamenávat přímo v prostředí OGAMA s využitím low-cost zařízení EyeTribe. Jak bylo popsáno v (Popelka a kol., 2016), přesnost tohoto zařízení je pro kartografické účely dostatečná. V aplikaci je následně nutné označit oblasti zájmu. Pro korektní výsledek analýzy je vhodné, aby tyto oblasti zájmu odpovídaly logickým celkům analyzovaného stimulu.

Jednou z výhod aplikace OGAMA je funkce Levenshtein Distance Calculation, která umožňuje spočítat vzdálenost mezi sekvencemi navštívených AOI. Tento výpočet není dostačující pro srovnávání prvků množiny více než dvou řetězců z důvodu jejich rozdílné délky, proto jej nelze přímo použít. Důležité ovšem je to, že je při využití této funkce generována sekvence, kterou lze jednoduše vyexportovat. Vytvořený nástroj ScanGraph pracuje přímo s tímto exportem, aby nebylo nutné data nijak upravovat.

Po nahrání dat do prostředí aplikace ScanGraph si uživatel zvolí jeden ze tří dostupných algoritmů a pomocí parametru p nebo hustoty grafu zvolí požadovanou míru podobnosti. Výstupem aplikace je obyčejný graf, ve kterém jednotlivé uzly reprezentují uživatele (sekvence). Pomocí Bron-Kerbosch algoritmu jsou v tomto grafu nalezeny všechny netriviální kliky. Vrcholy nalezených klik jsou v prostředí ScanGraph vypsány v uspořádaném seznamu a po kliknutí lze zobrazit sekvence jednotlivých uživatelů obsažených v klikách.

DC2 – Využití grafů v prostorovém plánování

Pro hledání skupin podobných sídelních jednotek, vykazujících míru podobnosti danou uživatelem, byla nejprve připravena data, na základě kterých byla podobnost určována. Jednalo se o hodnoty indexů z oblasti sociální, hospodářské a environmentální, které jsou každoročně počítány v rámci aktualizací Územně analytických podkladů Olomouckého kraje. Každý index může mít přidělenou určitou váhu. Do výpočtu podobnosti pomocí nástroje SimUrb vstupuje CSV (Comma Separated Value) tabulka obsahující hodnoty zvolených indexů ve sloupcích a její řádky reprezentují jednotlivé obce. Obce jsou definovány podle kódu Základní Územní Jednotky (ZUJ). Pro potřeby vizualizace výsledků je v tabulce rovněž obsažen název a příslušnost k vyšší správní jednotce (např. ORP).

Pro vlastní vizualizaci výsledků v mapě je nutné na server dopředu nahrát předem vytvořené vrstvy obcí a ORP ve formátu geojson, se kterými jsou následně data propojena a pomocí nichž jsou výsledky zobrazované v mapě.

Uživatel si v nástroji SimUrb volí hodnotu míry podobnosti p a minimální velikost skupiny, která jej zajímá.

Nástroj SimUrb je navržen tak, aby mohl pracovat se stovkami záznamů. Vyhledávání všech netriviálních klik by v takovém množství dat bylo obtížné až nemožné a pro danou studii bezúčelné. Proto podobné skupiny se v nástroji SimUrb vyhledávají jako disjunktní kliky příslušného grafu. Algoritmus pracuje tak, že v prvním kroku vyhledá největší kliku a její vrcholy následně vyřadí ze seznamu. Tento krok opakuje, dokud nezůstanou pouze izolované vrcholy či kliky o menším počtu vrcholů, než bylo požadováno uživatelem. Mohou tedy existovat i kliky, které nebudou výstupem, avšak pro vyhodnocení nejsou významné.

Výsledné netriviální kliky jsou uživateli zobrazeny v pravé části interface programu SimUrb obdobně jako v případě ScanGraph, ale místo zobrazení pomocí grafu jsou podobné obce zobrazené přímo v mapě.

DC3 – Využití grafů v dopravě

Metody využití pro tvorbu třetího nástroje s názvem SimRoad jsou velmi podobné těm, pomocí kterých byly hledány podobné obce v nástroji SimUrb. Data zpracovávaná nástrojem SimRoad mají však odlišné rozdělení než data nástroje SimUrb. Pro názornější vyhodnocování je proto určení míry upraveno za pomoci percentilů.

Příprava dat pro výpočet podobnosti úseků silnic na základě dat naměřených mytými branami je rovněž obdobná jako v případě SimUrb. Do výpočtu tentokrát vstupuje CSV tabulka, jejíž řádky reprezentují určité úseky silnic a ve sloupcích jsou uvedeny počty průjezdů vozidel v jednotlivých časových úsecích, např. dny.

DC4 – Případové studie využívající vytvořené nástroje

Vypracované případové studie se dotýkaly témat z oblasti eye-trackingu, urbanismu a dopravy. Případová studie z oblasti eye-trackingu se zabývala analýzou podobnosti z experimentu vizualizace nejistoty geografických dat. Pro analýzu byl použit nástroj ScanGraph využívající navrženou metodu pro vyhledávání podobných sekvencí navštívených oblastí zájmu. Případová studie navazovala na diplomovou práci Kučery (2016), v níž byl vytvořen soubor kartografických znaků pro vizualizaci nejistoty v bodovém, líniovém i plošném znaku. Stimuly z tohoto experimentu představovaly mapy s vyznačenými bodovými znaky. Při analýze stimulů v práci Kučery bylo využito eye-tracking metrik jako Trial Duration a Scanpath Length. V případové studii autorky byly analyzovány dva stimuly v podobě ortofoto snímků, a to v prvním případě s možným výskytem prasete divokého, zajíce polního a lišky obecné a ve druhém případě s možným výskytem pouze lišky obecné.

Pro případovou studii z oblasti urbanismu byly vyhodnocovány územně analytické podklady a srovnávány s již existujícími výstupy metodiky používané Olomouckým krajem. Studie byla provedena s použitím nástroje SimUrb a navržené metody vyhodnocování podobnosti pro data tohoto charakteru. Podobnost byla hledána mezi obcemi Olomouckého kraje, jež byly charakterizovány systémem atributů sociálního, hospodářského a environmentálního pilíře. Současně byl výstup nástroje SimUrb porovnáván i s výstupy nástrojů dostupných v softwaru ArcGIS (Grouping Analysis a Multivariate Clustering), které také vytvářejí shluky objektů na základě předem definovaných atributů.

Ve třetí případové studii, jež se týkala oblasti dopravy, byl aplikován nástroj SimRoad s navrženou metodou. Vyhodnocována byla data z mytých bran s údaji o průjezdech autobusů. Studie se věnovala srovnání průjezdů v pracovních dnech a o víkendech v rámci jednoho měsíce a to jak pro všechny úseky dálnic, silnic I. třídy a rychlostních komunikací, tak i pouze u dálnic. Z výstupů nástroje byly vybrány ty výstupy, které by mohly být dále studovány v rámci diplomových prací.

4. SOUČASNÝ STAV PROBLEMATIKY

Teorie grafů nachází své aplikace v mnoha vědních oborech. Geoinformatika tak není žádnou výjimkou. Odkazovat se můžeme již na prvopočátky teorie grafů, kdy Leonhard Euler (1736) řešil úlohu Kaliningradských mostů jak projít sedm mostů přes řeku Pregolu, a to každý z nich právě jednou, a vrátit se do výchozího místa. Tomuto problému odpovídá v teorii grafů eulerovský graf.

Uplatnění teorie grafů nalezneme také v kartografii. Snad nejznámější úlohou je problém obarvení mapy tak, aby žádná dvě sousední území nebyla obarvena stejnou barvou a současně bylo využito minimálního potřebného množství barev. Tato úloha je aplikací barvení vrcholů rovinného grafu a určení chromatického čísla grafu. Všeobecně známým se stal tzv. problém čtyř barev, tedy zda na obarvení libovolné politické mapy stačí čtyři barvy tak, aby žádné dva sousední státy nebyly vykresleny stejnou barvou. Tento problém byl vyřešen až v roce 1976, kdy Appel a Haken (1976) využili pro důkaz počítačový model všech možných konfigurací. Při tomto modelu jsou politické jednotky reprezentovány vrcholem a hrana mezi vrcholy vyjadřuje sdílení společné hranice.

Při generalizaci mapy je možno liniové prvky na mapě reprezentovat orientovanými ohodnocenými hranami grafu. Počátky, rozdvojení či křížení liniových prvků jsou pak reprezentovány vrcholy tohoto grafu. Generalizační výběr je pak simulován úlohami o cestách v grafu. Typické zpracování nalezneme u Saalfeld (1995) či Mackaness & Beard (1993).

Teorii grafů lze pro analýzu eye-tracking dat využít zejména při hodnocení podobnosti v průběhu trajektorií pohybu očí (Scanpath Comparison). Tento typ analýzy je velmi vhodný zejména v případech, kdy chceme z eye-tracking dat zjistit odlišnost chování (čtení stimulu) různými skupinami uživatelů. Nejčastěji se jedná o experty a laiky, muže a ženy atd. Historie studia trajektorie pohybu očí se dá vysledovat od Noton & Stark (1971), kteří vyhodnocovali podobnost v prohlížení kreseb respondenty při různých příležitostech. Jednou z nejběžněji používaných metod Scanpath Comparison je takzvaná String Edit Distance (Anderson a kol., 2015).

Využití teorie grafů v dopravních úlohách je velmi široké. Většina aplikací teorie grafů je zařaditelná do oblasti síťových analýz. Základní přehled o síťových analýz lze nalézt v Longley (2005) či Tuček (1998).

Kromě výše uvedených využití teorie grafů je možné najít aplikace i v dalších oblastech týkajících se geoinformatiky. V lodní dopravě využil síťové analýzy Coffin (2009), při vyhodnocování real-time dat (Big Data) Rathore a kol. (2015), pro konstrukci evakuačních plánů Dhamala (2014). Dopravní problémy v sociálních studiích řešil Roberts (1978). Při zkoumání vztahů mezi kvalitou krajiny a populací rybářských kolonií využil teorii grafů Garroway a kol. (2008). V 3D analýzách lze nalézt užití geometrických datových modelů v Lee (2004).

Dalším využitím teorie grafů v geografii a geoinformaticce se zabývali například Arlinghaus a kol. (2002), Pitts (1965) či Zetterberg a kol. (2010).

5. SCANGRAPH – TEORIE GRAFŮ V ANALÝZE EYE-TRACKING DAT

Výstupem prvního dílčího cíle je nově navržená metoda pro hledání podobnosti v eye-tracking datech a její aplikace v nástroji ScanGraph dostupném na adrese eyetracking.upol.cz/scangraph.

Vzdálenost a podobnost jsou veličiny vyjadřující vztah dvou bodů v prostoru, kdy bodem rozumíme vektor, objekt či v případě eye-tracking dat sekvenci navštívených AOI. vzdálenost můžeme chápat jako „míru nepodobnosti“. Čím je vzdálenost mezi dvěma objekty větší, tím jsou si méně podobné. Podobnost je tedy duální mírou ke vzdálenosti.

Pro korektní výpočet vzdálenosti a podobnosti mezi jednotlivými sekvencemi zavedeme pojem metriky a metrického prostoru.

Definice: Metrikou vzdálenosti D na množině X nazýváme funkci $D: X \times X \rightarrow \mathbb{R}$, kde X je n -rozměrný prostor splňující následující předpoklady:

$$0 \leq D(x, y) < \infty, \forall x, y \in X;$$

$$D(x, x) = 0, \forall x \in X$$

a má následující vlastnosti:

$$D(x, y) = D(y, x), \forall x, y \in X;$$

$$D(x, y) = 0 \Leftrightarrow x = y;$$

$$D(x, z) \leq D(x, y) + D(y, z), \forall x, y, z \in X.$$

Prostor X , ve kterém je metrika definována, nazýváme metrickým prostorem. Vzdálenost dvou objektů x a y je pak hodnota $D(x, y)$.

Metriky použité v této práci splňovat vlastnost

$$\max_{x,y} (D(x, y)) = 1.$$

Tedy $D(x, y) \in \langle 0, 1 \rangle, \forall x, y \in X$ a dále budeme uvažovat pouze metriky s touto vlastností.

Definice: Metrikou podobnosti S na množině X je taková funkce $S: X \times X \rightarrow \mathbb{R}$, kde X je n -rozměrný prostor, splňující následující předpoklady:

$$0 \leq S(x, y) \leq S_0 < \infty, \forall x, y \in X;$$

$$S(x, x) = S_0, \forall x \in X$$

a má následující vlastnosti:

$$S(x, y) = S(y, x), \forall x, y \in X;$$

$$S(x, y) = S_0 \Leftrightarrow x = y.$$

V této práci budeme konstantu S_0 uvažovat rovnu 1. Tedy dva objekty jsou si nanejvýše podobné právě tehdy, jsou-li totožné.

Dualitu metriky vzdálenosti a metriky podobnosti vyjádříme vztahem:

$$S(x, y) = 1 - D(x, y), \forall x, y \in X.$$

Odtud vyjádříme trojúhelníkovou nerovnost pro metriku podobnosti:

$$S(x, z) \geq S(x, y) + S(y, z) - 1.$$

S ohledem na požadované vlastnosti byly metriky pro výpočet podobnosti mezi sekvencí a , b navštívených oblastí zájmu navrženy následovně:

$$Lev'(a, b) = 1 - \frac{Lev(a, b)}{\max\{|a|, |b|\}},$$

$$DL'(a, b) = 1 - \frac{DL(a, b)}{\max\{|a|, |b|\}},$$

$$NW'(a, b) = \frac{NW(a, b)}{\max\{|a|, |b|\}}.$$

Pro všechny výše uvedené metriky podobnosti S platí:

$$S(a, b) \in \langle 0, 1 \rangle,$$

$$S(a, b) = 1 \Leftrightarrow a = b,$$

$$S(a, b) = 0 \Leftrightarrow a \cap b = \emptyset, \forall a, b \in X.$$

Poslední vlastnost interpretujeme tak, že sekvence a a b nemají žádný společný znak, přičemž délka daných sekvencí může být různá.

Při hledání vzájemně si podobných sekvencí vypočítá nástroj ScanGraph míru podobnosti pro všechny dvojice analyzovaných sekvencí. Při výpočtu podobnosti všech sekvencí jsou tak postupně vytvořeny tři symetrické čtvercové matice řádu n , kde n je počet analyzovaných sekvencí.

První matice s názvem Original matrix ($O = (o_{ij})_{i,j=1}^n$) je maticí hodnot zvoleného algoritmu (Levenshteinova vzdálenost, Damerau-Levenshteinova vzdálenost, Needleman-Wunsch algoritmus).

Jestliže sekvence a odpovídá i -tému řádku a sekvence b odpovídá j -tému sloupci, pak

$$o_{ij} = \begin{cases} Lev(a, b) \\ DL(a, b), \\ NW(a, b) \end{cases}$$

$\forall a, b \in X$. Speciálně pro $a = b$ platí

$$o_{ii} = \begin{cases} 0, Lev(a, a), DL(a, a) \\ |a|, NW(a, a) \end{cases}$$

Druhou maticí je pak Modified matrix ($M = (m)_{i,j=1}^n$), jejíž prvky jsou hodnoty modifikovaných algoritmů. Tedy

$$m_{ij} = \begin{cases} Lev'(a, b) \\ DL'(a, b), \\ NW'(a, b) \end{cases}$$

$\forall a, b \in X$. Speciálně pro $a = b$ platí $m_{ii} = 1$.

Matice Modified matrix může být označována jako matice míry podobnosti. Prvky této matice mohou nabývat hodnot od 0 do 1.

Určení poslední matice – Adjacency matrix, závisí na hodnotě míry podobnosti p , která je zadána uživatelem, a to buď přímou volbou nebo pomocí hustoty grafu. Hodnota míry podobnosti p je mírou hraniční podobnosti požadované uživatelem.

Matice Adjacency matrix ($A = (a_{ij})_{i,j=1}^n$), nebo-li matice sousednosti, je výchozí maticí pro vyhodnocení výsledků. Její hodnoty jsou určeny z Modified matrix následovně:

$$a_{ij} = \begin{cases} 1, m_{ij} \geq p \\ 0, m_{ij} < p \end{cases}$$

Hodnota 1 je přiřazena těm dvojicím sekvencí a a b , jejichž podobnost je rovna nebo větší než daná míra p . Hodnota 0 pak náleží těm dvojicím sekvencí, jejichž podobnost je menší než zadaná míra p .

Na základě matice sousednosti je sestrojen obyčejný graf o n vrcholech. Vrcholy grafu reprezentují jednotlivé sekvence a hranami jsou pak spojeny ty vrcholy, jejichž míra podobnosti je rovna nebo větší než daná míra p . V případě, že sekvence daného experimentu jsou uživatelem rozděleny do různých sledovaných kategorií (muži, ženy, experti, laici,...), jsou vrcholy grafu barevně odlišeny podle těchto kategorií.

Skupiny vzájemně si podobných sekvencí pak představují kliky v tomto grafu. Vyhledávání klik je realizováno pomocí Bron-Kerbosch algoritmu.

Použití nástroje ScanGraph je přímočaré. Nástroj ScanGraph umožňuje analyzovat jeden soubor či „multifile“ – ZIP soubor obsahující data z několikastimulového experimentu.

Při prvním spuštění webového nástroje ScanGraph se zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadovaný způsob výpočtu.

Nahrání souboru provede uživatel stiskem tlačítka Choose file.

V obou případech vybere dále uživatel v poli Define method algoritmus výpočtu pro String-Edit-Distance. K dispozici jsou tři algoritmy – Levenshtein Distance, Needleman-Wunsch algoritmus a Damerau-Levenshtein Distance. Konkrétní metodu volí uživatel na základě charakteru dat svého experimentu a zaměřením cíle experimentu.

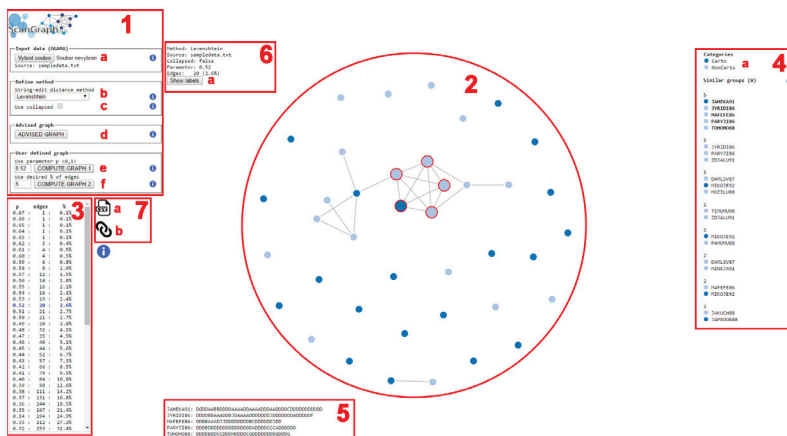
Ve stejném poli Define method uživatel dále specifikuje zaškrtávacím políčkem Use collapsed, zda požaduje, aby řetězec navštívených AOI neobsahoval násobné znaky.

Tuto možnost uživatel opět využívá na základě očekávaného charakteru dat a stanoveného cíle experimentu. Při nezaškrtnutí políčka Use collapsed zůstanou analyzované sekvence v původním tvaru, tedy tak, jak je uživatel nahrál. Pokud naopak tuto možnost uživatel zaškrtně, budou všechny násobné znaky nahrazeny jedním znakem - AAABBCCCCDD → ABCD. Tuto možnost volí uživatel v případě, že pro analýzu experimentu není podstatný počet fixací v jednotlivých oblastech zájmu, ale pouze pořadí jejich navštívení.

V případě analýzy multifile souboru zvolí uživatel navíc metodu zpracování těchto souborů. Na výběr jsou dvě možnosti, metoda Mean (průměr) a metoda Occurance (výskyt). Metody jsou blíže popsány v závěru této kapitoly.

Ke spuštění výpočtu může dojít na základě zadané míry podobnosti p nebo hustoty hran grafu. Míra podobnosti p nabývá hodnot z intervalu $(0,1)$, přičemž hodnota 0 představuje nulovou (žádnou) podobnost mezi řetězci a hodnota 1 absolutní podobnost, tedy totožnost řetězců. Čím vyšší je hodnota této míry podobnosti, tím vyšší nárok je kladen na vzájemnou podobnost mezi řetězci. Stiskem tlačítka Compute graph 1 se spustí výpočet, který určí skupiny řetězců vykazující zadanou míru podobnosti a vyšší. Alternativou je zadání hustoty grafu. Tato hodnota je chápána jako procento zastoupení hran výsledného grafu vůči úplnému grafu při daném počtu vrcholů a tedy nabývá hodnot z intervalu $(0,100)$. Úplný graf s $n \geq 2$ vrcholy obsahuje $\binom{n}{2}$ hran a hustota hran grafu (v procentech) je pak hodnota $\frac{|E|}{\binom{n}{2}} \cdot 100$, kde $|E|$ je počet hran grafu a n je počet jeho vrcholů. Stiskem tlačítka Compute graph 2 uživatel spustí výpočet skupin podobných řetězců tak, aby výsledný graf reprezentující tyto skupiny měl právě zadanou hustotu hran nebo nejbližší možnou nižší. Pro uživatele, kteří nemají předchozí zkušenost s nástrojem je vhodné využít tlačítka Advised graph, které spustí výpočet podobných skupin reprezentovaných grafem o hustotě hran 5% nebo nejbližší možné nižší. Tato hodnota byla určena na základě předchozích zkušeností s vyhodnocováním eye-trackingových dat. Graf s pětiprocentní hustotou hran má zvykle poměrně dobrou vypovídající hodnotu (Popelka, 2018). Tento graf může sloužit pouze pro spuštění výpočtu a uživatel může dále s výstupem pracovat a nastavovat jeho požadované parametry v poli User define graph. Graf s požadovanou mírou podobnosti má jednoznačně určenou hustotu hran. Nelze tedy tyto dva parametry nastavovat současně a vždy se zobrazí pouze graf odpovídající volbě posledního stisknutého tlačítka Compute graph 1 a Compute graph 2 či Advised graph.

Nyní se zobrazí zbylá pole nástroje. V centru prohlížeče je zobrazen samotný graf, který odpovídá nastaveným hodnotám.



Obr. 2 Jednotlivé prvky nástroje ScanGraph

Vrcholy zobrazeného grafu reprezentují jednotlivé sekvence vyhodnocovaného experimentu. Rozdílné barvy určují příslušnost sekvencí k vyšetřovaným kategoriím (např. muž/žena, odborník/laik apod.). Ty sekvence, které jsou si podobné s mírou podobnosti větší či rovnou zadané míře, jsou spojeny hranou. Skupiny vzájemně si podobných sekvencí tak vytváří kliky v příslušném grafu. Množina netriviálních klik grafu tvoří jeho pokrytí. Jejich seznam je možné si prohlédnout v poli v pravé části. Po kliknutí na seznam vrcholů kliky se červeně podbarví příslušné vrcholy v grafu a současně v levé části se zobrazí příslušné řetězce. Kliky jsou popsány seznamem svých vrcholů a jsou seřazeny sestupně podle velikosti.

V levém horním rohu pole grafu nalezne uživatel celkový přehled svého experimentu a pomocí tlačítka Show labels přidá popisky k vrcholům grafu. Tento přehled obsahuje údaje Time (doba trvání výpočtu), Method (zvolená metoda pro výpočet String-Edit-Method), Source (Název nahraného analyzovaného souboru), Collapsed (zda uživatel zvolil originální řetězce nebo řetězce bez násobnosti fixací), Parameter (zvolená hodnota parametru při volbě Compute graph 1 či přiřazená hodnota parametru při volbě Compute graph 2), Edges (počet hran výsledného grafu včetně hustoty hran) a Edge value (ohodnocení hrany mírou podobnosti).

V levé dolní části se zobrazí tabulka, která uvádí všechny možné varianty grafu, který může data daného experimentu reprezentovat. Varianty jsou seřazeny podle hodnoty parametru (míry podobnosti) od nejvyšší možné po nejnižší nulovou. V prvním sloupci tabulky je uvedena hodnota míry podobnosti, k níž jednoznačně je dán počet hran zobrazených ve druhém sloupci a tedy i hustota hran zobrazena ve třetím sloupci. Kliknutím na libovolný řádek této tabulky zobrazí uživatel graf příslušný tomuto řádku a tedy příslušné míře podobnosti.

6. SIMURB – TEORIE GRAFŮ V PROSTOROVÉM PLÁNOVÁNÍ

Cílem nástroje SimUrb je vyhledání vzájemně si podobných skupin objektů na základě předem definovaných atributů. Každému z n zkoumaných objektů je tak přiřazena uspořádaná m -tice koeficientů, které pocházejí z předem definovaných intervalů.

Označme $X = \{x_1, x_2, \dots, x_n\}$ množinu zkoumaných objektů, $K = \{K_1, K_2, \dots, K_m\}$ množinu atributů, $\underline{x}_i = (k_1, k_2, \dots, k_m)$ uspořádanou m -tici koeficientů objektu x_i , kde $k_i \in \langle \min k_i; \max k_i \rangle, \forall i \in 1, \dots, m$, a $w_i = (w_1, w_2, \dots, w_m)$ uspořádaná m -tice vah, kde w_i je vahou atributu $K_i, w_i \in \mathbb{R}, \forall i = 1, \dots, m$.

Jednotlivé objekty s jejich přiřazenou m -ticí koeficientů můžeme chápat jako body v m -rozměrném prostoru.

Pro určení míry podobnosti byla navržena metrika, jenž vychází z euklidovské vzdálenosti.

Nechť x_i a x_j jsou dva objekty z množiny X . Pak pro míru jejich podobnosti $d(x_1, x_2)$ platí:

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^m \frac{w_j(k_{1j} - k_{2j})^2}{w_j(\max K_j - \min K_j)^2}}.$$

Pro hodnoty míry podobnosti $d(x_1, x_2)$ platí $d(x_1, x_2) \in \langle 0, 1 \rangle$, přičemž $d(x_1, x_2) = 1 \Leftrightarrow \underline{x}_1 = \underline{x}_2$ a

$d(x_1, x_2) = 0 \Leftrightarrow \underline{x}_1 = (\min K_1, \min K_2, \dots, \min K_m) \wedge \underline{x}_2 = (\max K_1, \max K_2, \dots, \max K_m)$. Je zřejmé, že totéž bude platit, jestliže minimálních hodnot intervalů atributů bude nabývat objekt x_2 a maximálních hodnot objekt x_1 . Hodnota $d(x_1, x_2)$ je tedy rostoucí s rostoucí podobností mezi objekty.

Při hledání vzájemně si podobných objektů vypočítá nástroj SimUrb míru podobnosti pro všechny dvojice analyzovaných objektů. Při výpočtu podobnosti všech objektů jsou tak postupně vytvořeny tři symetrické čtvercové matice řádu n , kde n je počet analyzovaných objektů.

První matice s názvem Original matrix ($O = (o_{ij})_{i,j=1}^n$) je maticí hodnot euklidovské vzdálenosti pro všechny dvojice x_i, x_j z množiny všech objektů X .

Druhou maticí je pak Modified matrix ($M = (m)_{i,j=1}^n$), jejíž prvky jsou hodnoty $d(x_i, x_j), \forall x_i, x_j \in X$.

Určení poslední matice – Adjacency matrix, závisí na hodnotě parametru p , který je zadán uživatelem. Hodnota parametru p je mírou hraniční podobnosti požadované uživatelem a nabývá hodnot z intervalu $\langle 0, 1 \rangle$, kdy $p = 1$ indikuje

totožné sekvence a $p = 0$ zcela odlišné (jedna sekvence je tvořena minimálními možnými hodnotami a druhá maximálními).

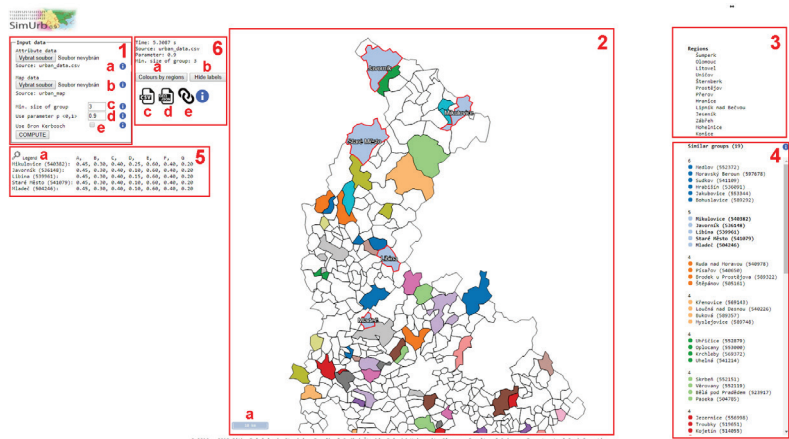
Matice Adjacency matrix ($A = (a_{ij})_{i,j=1}^n$), nebo-li matice sousednosti, je výchozí maticí pro vyhodnocení výsledků. Její hodnoty jsou určeny z Modified matrix následovně:

$$a_{ij} = \begin{cases} 1, & m_{ij} \geq p \\ 0, & m_{ij} < p \end{cases}$$

Hodnota 1 je tedy přiřazena těm dvojicím objektů x_i a x_j , jejichž podobnost je rovna nebo větší než daná míra p . Hodnota 0 pak náleží těm dvojicím sekvencí, jejichž podobnost je menší než zadaná míra p .

Vzájemně si podobné skupiny tvoří opět kliky v obyčejném grafu. Vzhledem k tomu, že většina analýz, k nimž je nástroj SimUrb uzpůsoben, pracuje s desítkami až stovkami řetězců, by vizualizace formou grafu byla nepřehledná a matoucí. Navíc je u sekvencí s geografickou lokalizací vhodnější vizualizovat výsledky přímo v mapě.

Cílem uživatele nástroje SimUrb bývá obvykle vyhledat disjunktní shluky vzájemně si podobných objektů. Nedisjunktní skupiny často ztrácí přehlednost a především význam pro jejich analýzu. Z toho důvodu pracuje nástroj SimUrb odlišným způsobem než výše popsany nástroj ScanGraph.



Obr. 3 Jednotlivé prvky nástroje SimUrb

Při prvním spuštění webového nástroje SimUrb se zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadovaný způsob výpočtu.

Nahrání souboru s daty provede uživatel volbou tlačítka Choose file v poli Attribute Data. Dále může uživatel nahrát mapový podklad, opět volbou tlačítka

Choose file, ovšem v poli Map data. Tento výběr není povinný. Metoda výpočtu podobnosti je v nástroji SimUrb pouze jediná a tedy se volba uživateli nenabízí.

Dalším krokem je nastavení minimálního počtu prvků ve skupině a hodnota míry podobnosti. Jelikož často uživatele při analýze zajímají pouze skupiny s větším množstvím prvků, byla do nástroje přidána právě volba nastavení minimálního počtu prvků ve skupině. Algoritmus výpočtu se pak zastaví u takto velkých skupin a dále již nepokračuje, čímž se celková doba výpočtu zkrátí. Míra podobnosti p může nabývat hodnot z intervalu $(0,1)$, přičemž hodnota rovna jedné značí zcela totožné koeficienty a hodnota rovna nule značí koeficienty s největší možnou odlišností.

U voleb nástroje se nachází také možnost použití Bron-Kerbosch algoritmu pro vyhledávání všech klik mezi sekvencemi. Tato volba je doporučena pouze pro malé soubory, kde cílem analýzy není vytvoření disjunktních shluků a charakter dat umožňuje pro správné vyhodnocení překryv mezi skupinami. Volbou tlačítka Compute je spuštěn výpočet.

Poté se zobrazí v pravé části seznam všech nalezených klik. Kliknutím na kteroukoliv skupinu se v levé části zobrazí jejich seznam společně s příslušnými sekvencemi koeficientů. Najetím myši na ikonu lupy se pak uživateli zobrazí legenda s popisem názvu koeficientů a jejich vahami.

V levé horní části je uvedena délka trvání výpočtu, název zdrojových dat, zvolená míra podobnosti a minimální velikost skupiny. K dispozici jsou stejně jako v nástroji ScanGraph možnosti stáhnout CSV soubor s analyzovanými daty a vypočtenými hodnotami v Original, Modified a Adjacency matici a možnost permanentního odkazu.

Jestliže uživatel vložil spolu s daty i mapový podklad, jsou výsledné skupiny vizualizovány i v mapě. Primárně jsou charakterizovány areálovým znakem a barevně rozlišené podle příslušných nalezených skupin. Jestliže uživatel klikl na některou skupinu v seznamu, jsou v mapě hranice příslušných regionů zvýrazněny pomocí červené barvy.

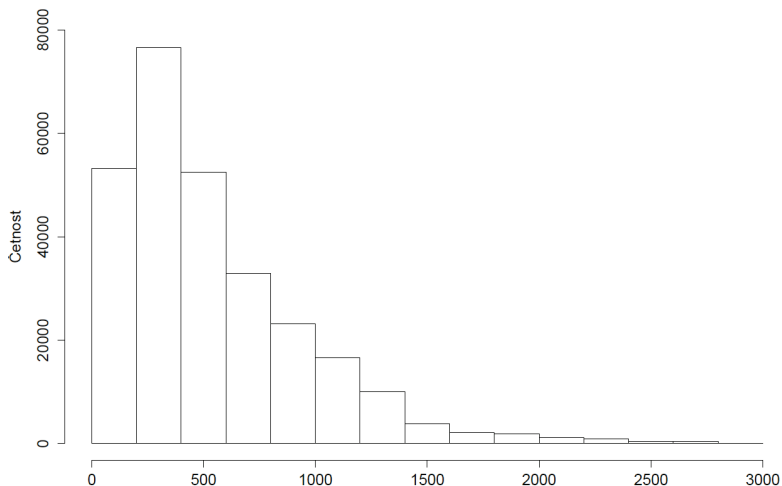
Volbou tlačítka Colours by groups se pak přepne vizualizace podle nadregionů, jsou-li v datech zadány. Volbou tlačítka Show labels / Hide labels pak uživatel může zobrazit či skrýt názvy objektů v mapě. Zobrazit detaily jednotlivých skupin lze pak i kliknutím na regiony v mapovém podkladu.

7. SIMROAD – TEORIE GRAFŮ V DOPRAVĚ

Metrika používaná nástrojem SimRoad je velice blízká metrice používané nástrojem SimUrb.

Prvotní určení míry podobnosti v nástroji SimRoad probíhá obdobně jako v nástroji SimUrb. Matice Modified matrix a Original matrix jsou určeny stejným výpočtem jako v případě nástroje ScanGraph, avšak jelikož očekávané rozdělení hodnot míry podobnosti pro všechny dvojice sekvencí, které nástroj zpracovává, má vysokou šikmost (obr. 4), je parametr p , jenž je zadáván uživatelem, stanoven jako hodnota kvantilové funkce.

Rozdělení hodnot original matrix



Obr. 4 Histogram rozdělení četnosti hodnot Original matrix

Nechť $F(x)$ je distribuční funkce náhodné veličiny X . Funkce

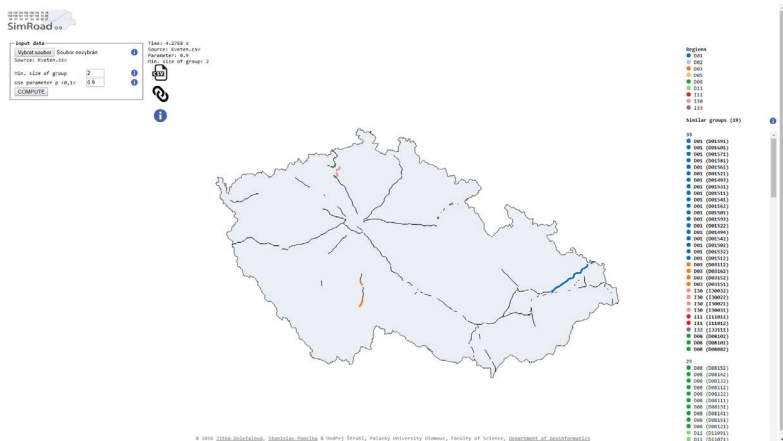
$$F^{-1}(u) = \inf\{x \in \mathbb{R}; F(x) \geq u\}, \quad 0 < u < 1,$$

se nazývá kvantilová funkce náhodné veličiny X . Pro $0 < p < 1$ se hodnota $F^{-1}(p)$ nazývá p -kvantil. (Kunderová, 1997)

Nadále tedy pro matici sousednosti $A = (a_{ij})_{i,j=1}^n$ (Adjacency matrix) platí

$$a_{ij} = \begin{cases} 1, & m_{ij} \geq p \\ 0, & m_{ij} < p \end{cases}$$

Stejně jako u nástroje SimUrb by vytváření nedisjunktních skupin vzájemně si podobných objektů nástrojem SimRoad bylo nepřehledné a bezúčelné. Skupiny vzájemně si podobných objektů jsou tedy opět vyhledávány jako maximální kliky v grafu sestaveném dle matice sousednosti, kdy v každém kroku je vybrána největší klika a její vrcholy jsou vyřazeny z dalšího výpočtu.



Obr. 5 Prostředí nástroje SimRoad

Nástroj SimRoad umožňuje analyzovat najednou pouze jeden soubor s daty.

Při prvním spuštění webového nástroje SimRoad se stejně jako v nástroji SimUrb zobrazí pouze několik základních polí nástroje, ve kterých si uživatel nahraje soubor a nastaví požadované parametry.

Nahrání souboru s daty provede uživatel volbou tlačítka Choose file v poli Attribute Data. Metoda výpočtu míry podobnosti je v nástroji SimRoad pouze jediná a tedy se volba uživateli nenabízí.

Dalším krokem je nastavení minimálního počtu prvků ve skupině a hodnota míry podobnosti. Míra podobnosti p může nabývat hodnot z intervalu $(0,1)$, přičemž hodnota rovna jedné značí zcela totožné sekvence a hodnota rovna nule značí sekvence s největší možnou odlišností.

Poté se zobrazí v pravé části seznam všech nalezených klik. Kliknutím na kteroukoliv skupinu se tato skupina barevně zvýrazní v mapě. Vzhledem k rozsáhlé délce sekvencí, jenž určují jednotlivé objekty analyzované nástrojem SimRoad, nejsou tyto v základním prostředí nástroje zobrazovány. Uživatel si je však může stáhnout a zobrazit v CSV souboru.

V levé horní části je uvedena délka trvání výpočtu, název zdrojových dat, zvolená míra podobnosti a minimální velikost skupiny. K dispozici jsou stejně jako v nástroji ScanGraph možnosti stáhnout CSV soubor s analyzovanými daty a vypočtenými hodnotami v Original, Modified a Adjacency matici a možnost permanentního odkazu.

Skupiny, jenž jsou výstupem nástroje SimRoad jsou vizualizovány v mapě. Jelikož se jedná o dálnice a silnice I. třídy, jsou vizualizovány liniovými znaky a barevně rozlišeny podle příslušných kategorií.

8. PŘÍPADOVÉ STUDIE VYUŽÍVAJÍCÍ VYTVOŘENÉ NÁSTROJE

Vytvořené nástroje byly použity ve třech případových studiích.

Případová studie z oblasti eye-trackingu navazovala na diplomovou práci Kučery (2016).

Kartografická případová studie byla zaměřena na nejistotu vizualizace v mapách. Nejistota je považována za nejasnost, náhodnost podmínek nebo výsledek určitých procesů a jevů. Koncept nejistoty je také poměrně často používán k popisu malé jistoty o konkrétním jevu v mapách (Brus a kol., 2013). Mnoho přístupů a metod vizualizace nejistoty bylo vyvinuto na příklad na Bertinově teorii grafických proměnných a kombinování statických i dynamických prvků vizualizace (Kubíček, Šašínska, & Stachoň, 2012).

V závěrečné práci Kučery (2016) byl vytvořen soubor kartografických znaků pro vizualizaci nejistoty pomocí bodových, liniových i plošných znaků. Bodové znaky byly vytvořeny podle studie MacEachrena a kol. (2012). Tyto znaky byly umístěny do map, jež byly použity jako stimuly pro online dotazníkové šetření a následný eye-tracking experiment. Cílem diplomové práce bylo zhodnotit, které vizualizace jsou pro čtenáře mapy nejvíce srozumitelné. Experiment zahrnoval 40 respondentů. Dvacet z nich bylo studenti kartografie a dvacet respondenti bez kartografického vzdělání. Eye-tracking metriky (Trial Duration, Scanpath Length) a přesnost odpovědí byly použity při analýze všech stimulů. Celkem bylo v experimentu použito 27 map s bodovými znaky. Třináct z nich zobrazovalo jeden jev, osm představovalo kombinaci více jevů. Šest map zobrazovalo časoprostorovou nejistotu.

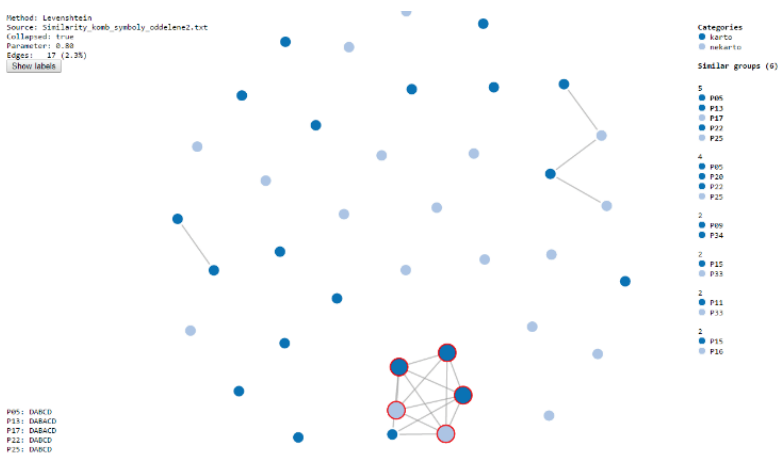
Nejprve byly analyzovány strategie prohlížení stimulu s označením BK07 (obr. 6). V tomto případě 16 bodových znaků reprezentovalo možný výskyt tří živočišných druhů (prase divoké, zajíc polní a liška obecná) s rozdílným stupněm nejistoty. Úkolem respondentů bylo určit lokalitu s největší pravděpodobností výskytu každého z druhů. Legenda se všemi třemi druhy byla umístěna na pravé straně stimulu. Levá část stimulu pak obsahovala ortofoto mapu s bodovými znaky. Oblasti zájmu byly vyznačeny okolo mapového pole a okolo každé části legendy.

Trajektorie pohybu očí byly konvertovány do sekvencí znaků podle pozice fixací ve vyznačených oblastech zájmu v prostředí OGAMA. Z trajektorie pohybu očí byly vygenerovány řetězce fixací. Pro analýzu byla zvolena varianta collapsed sekvencí a dále pro výpočet podobností mezi sekvencemi byla zvolena modifikovaná metoda Levenshteinovy vzdálenosti.



Obr. 6 Oblasti zájmu vyznačené ve stimulu BK07

Míra podobnosti byla nastavena na hodnotu $p = 0,8$. Ve výsledném grafu bylo nalezeno šest netriviálních klik (obr. 7). Největší klika zahrnovala pět sekvencí. Tři z nich patřily kartografům a dvě nekartografům.



© 2015 Jiřka Dolnáková, Stanislav Popelka & Ondřej Štrubel, Palacký University Olomouc, Faculty of Science, Department of Geoinformatics

Obr. 7 Výstup z nástroje ScanGraph pro stimul BK07 a míru podobnosti $p = 0,8$

Strategie těchto respondentů může být charakterizována jako ideální. Všichni začali prohlížet stimulus od středu obrazovky (oblast zájmu D). Poté přesunuli pohled do všech částí legendy (oblasti zájmu A, B a C) a nakonec se vrátili zpět

do mapového pole, kde vyhledávali správnou odpověď. U respondentů P13 a P17 se navíc vyskytla fixace v oblasti zájmu A po prohlížení oblasti zájmu B.

Ve čtyřprvkové klíče byla situace velice podobná. V tomto případě všechny sekvence měly pořadí navštívených oblastí zájmu „DABCD“. Jedinou výjimkou byl respondent P20, u kterého byla navíc nalezena fixace v oblasti zájmu B na konci prohlížení stimulu.

Zbylé vizualizované netriviální kliky obsahovaly pouze dva respondenty. Všichni tyto respondenti při svém prohlížení stimulu vynechali oblast zájmu C. Oblast zájmu C byla vyznačena okolo poslední části legendy (reprezentující možný výskyt lišky obecné). Jelikož všechny části legendy mají stejnou vypovídající hodnotu, nepotřebovali respondenti prohlížet její poslední část.

Zbylé vrcholy grafu, tedy zbylé sekvence, tvořily jednoprvkové kliky neboli izolované vrcholy. Což znamená, že jejich sekvence navštívených oblastí zájmu nebyla podobná žádné jiné sekvenci (při dané míře podobnosti $p = 0,8$). Příkladem takovéto sekvence může být respondent P29 se svou sekvencí „DABADADADABDBCBDAD“, respondent P30 s velice krátkou sekvencí „DAD“ nebo respondent P43, jehož fixace byly všechny v mapovém poli, tedy v oblasti zájmu D. Všichni tyto tři respondenti patřili do kategorie nekartografů.

Za použití nástroje ScanGraph bylo možné rychle nalézt skupinu respondentů, kteří prohlíželi daný stimul podobným způsobem. Po vyšetření jejich sekvencí bylo zjištěno, že tato sekvence se dá klasifikovat jako „ideální“.

Nástroj SimUrb byl použit pro vyhodnocení územně analytických podkladů a srovnání výstupů s existujícími výstupy používané metodiky (metodika používaná Olomouckým krajem pro zpracování rozboru udržitelného rozvoje území v rámci územně analytických podkladů). Data byla sestavena celkem z 23 indikátorů (sociální pilíř 7 indikátorů, environmentální pilíř 8 indikátorů a hospodářský pilíř 8 indikátorů), které jsou součástí územně analytických podkladů Olomouckého kraje. Data jsou platná pro rok 2016 a byla vypočtena zpracovatelem ÚAP Olomouckého kraje (Urban Planner s.r.o.) z několika vstupních datových sad (datový sklad ÚAP Olomouckého kraje, Ministerstvo financí ČR, ČSÚ a AOPK). Data byla poskytnuta ve formátu XLS a SHP za všechny obce Olomouckého kraje.

Jednotlivé indikátory nabývají pouze hodnot z množiny {1,2,3,4,5}, přičemž každému indikátoru je navíc přiřazena jeho váha. Výstup realizovaný zpracovatelem byl vytvořen následujícím způsobem: Pro jednotlivé obce byl vytvořen vážený průměr příslušných indikátorů samostatně pro jednotlivé pilíře (Územní podmínky pro soudržnost společenství obyvatel území, Územní podmínky pro hospodářský rozvoj, Územní podmínky pro příznivé životní prostředí) a tyto hodnoty byly rozčleněny do šesti kategorií. Tři z těchto kategorií byly hodnoceny jako pozitivní a tři jako negativní. Celkové hodnocení rozboru udržitelného rozvoje území je pak tvořeno osmi kategoriemi (2³), dle toho, zda daná obec byla v jednotlivých pilířích hodnocena jako pozitivní či negativní.

Pro srovnání s výstupem nástroje SimUrb bylo zvoleno takové nastavení, které vygeneruje také právě šest skupin vzájemně si podobných obcí. Pro sociální pilíř bylo zvoleno nastavení míry podobnosti $p = 0,73$ a skupiny o minimálním počtu 15 objektů, pro hospodářský pilíř byla nastavena míra podobnosti $p = 0,75$ a skupiny o minimálním počtu 16 objektů a pro environmentální pilíř byla zvolena míra podobnosti $p = 0,71$ a opět skupiny o minimálním počtu 15 objektů. Nástroj SimUrb však nezbytně nezařadí všechny obce do kategorií. Ty obce, jež nevyhovují zadaným parametrům, zůstanou mimo vytvořené skupiny podobných obcí. Nástroj SimUrb tak může více odpovídat realitě, kdy skutečně obec (či obecně jakýkoliv vyhodnocovaný objekt) si není podobná s žádnou či minimální požadovanou skupinou jiných obcí. Na okraji tak zůstanou ty objekty, jež mohou vykazovat jistou extrémnost v jednom či více indikátorech.

Shoda s kategoriemi vymezenými v rámci ÚAP je uvedena v tabulkách 1 a 2.

Hodnoty uvedené tabulce 1 představují procentuální zastoupení obcí z podobných skupin nástroje SimUrb v kategoriích vymezených v rámci ÚAP pro sociální pilíř.

Tab. 1 Srovnání hodnocení sociálního pilíře s výstupem nástroje SimUrb

	Počet obcí	Procentuální podíl						Střední hodnota	Směrodatná odchylka
		Velmi nevyhovující (2,25 a méně)	Nevyhovující (2,26 – 2,60)	Spíše nevyhovující (2,61 – 2,85)	Spíše vyhovující (2,86 – 3,15)	Vyhovující (3,16 – 3,50)	Velmi vyhovující (3,51 a více)		
A	47		2,13	36,17	25,53	34,04	2,13	3,03	0,241
B	24		37,5	41,67	20,83			2,71	0,238
C	21					28,57	71,43	3,69	0,229
D	21	85,71	14,29					2,05	0,215
E	19			26,32	36,84	36,84		3,06	0,247
F	18			27,78	38,89	27,78	5,55	2,77	0,226

V jednotlivých pilířích můžeme pozorovat odlišnosti od původního vyhodnocení pomocí pouhého průměru hodnot. Na níže uvedených obrázcích můžeme na příkladech srovnat obce se stejným průměrem, které byly zařazeny nástrojem SimUrb do odlišných skupin.

V sociálním pilíři (tab. 2) byla obec Rovensko zařazena nástrojem SimUrb do skupiny B, Chromeč do skupiny E a Lipník nad Bečvou do skupiny F, kdežto při původních vyhodnocování byly zařazeny všechny do stejné kategorie, a to díky stejnému průměru. Avšak můžeme pozorovat, že obec Rovensko má nižší hodnoty ve třetím, čtvrtém a sedmém indikátoru, ale výrazně vyšší v pátém indikátoru. Dále obec Chromeč se od Lipníku nad Bečvou odlišuje nižšími hodnotami v pátém a šestém indikátoru, ale za to vyššími hodnotami v prvním, třetím a čtvrtém. Obdobně můžeme srovnat i obec Rovensko s Lipníkem nad Bečvou. Je tedy zřejmé, že ačkoliv obce vykazují stejnou průměrnou hodnotu indikátorů, nejsou si svými hodnotami indikátorů příliš podobné.

Tab. 2 Ukázka obcí se stejným průměrem zařazených nástrojů SimUrb do různých kategorií.

Obec	Skupina Simurb	Hrubá míra přirozeného přírůstu	Hrubá míra migračního salda	Počet dekoncentrací bytů	Míra dostupné urbanizace	Index stáří	Index vzdělanosti	Trend hrubé míry celkového přírůstu	
Rovensko	B	0,6	0,3	0,4	0,05	0,8	0,6	0,1	2,85
Chroměč	E	0,6	0,3	0,6	0,15	0,4	0,6	0,2	2,85
Lipník nad Bečvou	F	0,45	0,3	0,4	0,1	0,6	0,8	0,2	2,85

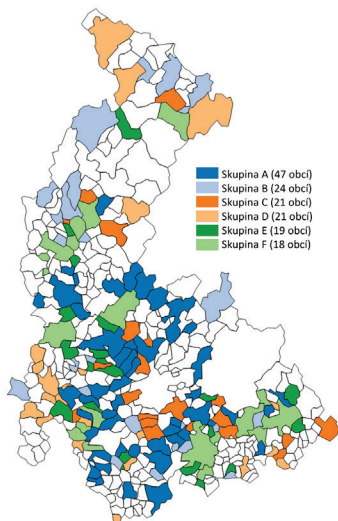
Analýzovaná data byla pro srovnání vyhodnocena také nástroji dostupnými v softwaru ArcGIS, které umožňují vytvářet shluky obcí na základě jejich atributů (nástroje Grouping Analysis a Multivariate Clustering). Nástroj Grouping Analysis nabízí možnost výpočtu s prostorovým kontextem i bez něj. V případě, že je do výpočtu zahrnuto i prostorové rozložení objektů, využívá nástroj minimální kostry grafu pro nalezení přirozených shluků. Jestliže prostorová složka objektů nevstupuje do určení shluků, použije nástroj K Mean algoritmus, jenž se řadí mezi NP-těžké úlohy. Při jeho inicializaci je náhodně zvoleno tolik objektů, jaký je požadovaný počet skupin, a to tak, aby byly od sebe co nejvíce vzdáleny v rámci svých přiřazených hodnot. Jelikož počáteční objekty, ke kterým se přiřazují další objekty pro vytvoření shluků, jsou voleny náhodně, je při každém spuštění nástroje výsledná varianta výstupu rozdílná. Při spuštění Multivariate Clustering při zadaném počtu požadovaných shluků je cílem nalézt co nejvíce homogenní shluky navzájem co nejvíce heterogenní. Shluky jsou opět tvořeny pomocí K Mean algoritmu.

Na obrázku 8 jsou vizuálně porovnané výstupy jednotlivých nástrojů.

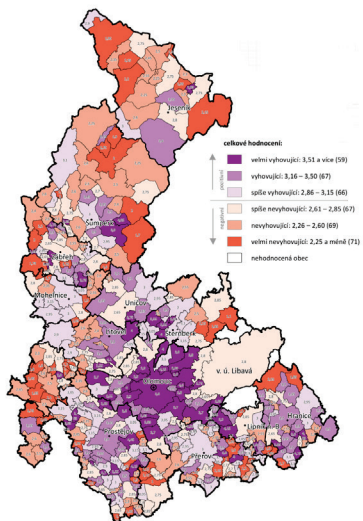
Případová studie využívající nástroj SimRoad navazuje na bakalářskou práci Roubalíka (Roubalík, 2017). Její snahou není porovnat výsledky této práce s výstupy z nástroje SimRoad, ale přinést nové výsledky, jenž jsou dostupné pouze díky tomuto nástroji.

Bakalářská práce Roubalíka se zabývá analýzou toků veřejné dopravy na základě dat z mýtných bran. V práci jsou použita data poskytnutá Ředitelstvím silnic a dálnic v průběhu roku 2015. V praktické části se práce zaměřuje na statistické vyhodnocení a grafickou vizualizaci dat o průjezdech autobusových dopravců mýtnou zónou. Dále popisuje časové změny intenzity dopravy v průběhu roku 2015 z celkového hlediska i z hlediska vybraných dopravců a také srovnává intenzitu dopravy s datovou sadou z Celostátního sčítání dopravy z roku 2010. Je popsána časová i prostorová variabilita dopravního toku a rozdíl mezi individuálními dopravci.

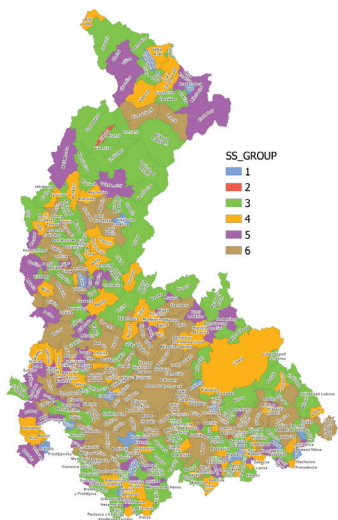
ÚZEMNÍ PODMÍNKY PRO SOUDRŽNOST SPOLEČENSTVÍ ÚZEMÍ



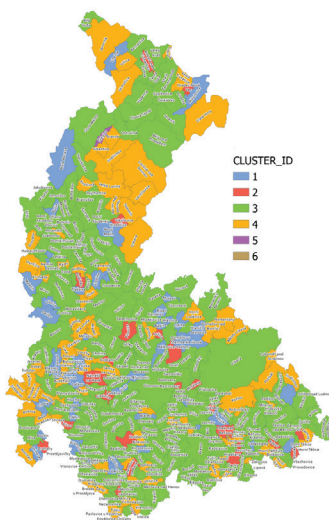
A - Výstup nástroje SimUrb



B - Výstup dle metodiky RURÚ



C - Grouping Analysis (ArcGIS)



D - Multivariate Clustering (ArcGIS)

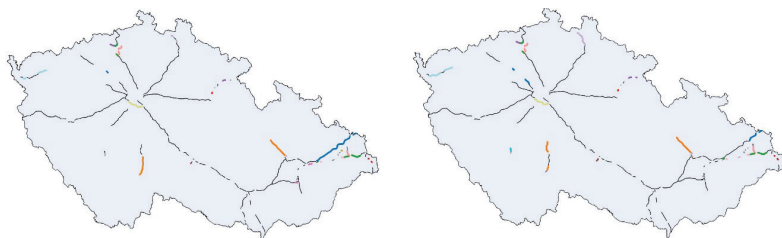
Obr. 8 Srovnání výsledků nástroje SimUrb s dalšími přístupy na příkladu územních podmínek pro soudržnost společenství území.

Tato práce se však nezabývá vyhledáváním podobných silničních úseků z hlediska průjezdnosti. Tato analýza byla provedena nástrojem SimRoad nad rámec cílů této bakalářské práce.

Pro případovou studii byla vybrána data z května 2015. Studie analyzovala data dvěma přístupy. Prvním z nich bylo porovnání dat průjezdů na všech typech silnic během pracovních dnů a během víkendů.

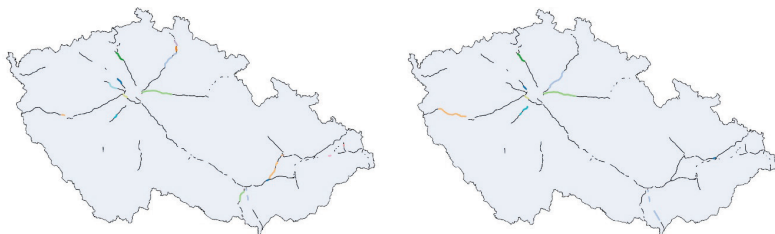
Pro oba výstupy byla zvolena hodnota míry podobnosti $p = 0,8$ a vyhodnocovány byly pouze skupiny o 10 a více úsecích. V obou případech tak bylo nalezeno 12 skupin. Zmíněno je několik pozorování, která mohou být dále studována například v bakalářských pracech.

Největší skupina v pracovních dnech byla tvořena 101 úsecy, o víkendech 113 úsecy (obr. 9). Seznamem svých úseků se příliš neliší. V obou případech je do této skupiny zahrnuta rychlostní silnice R35 na úseku Křelov – Mohelnice, rychlostní silnice R06 a rychlostní silnice R01 (jižní část pražského okruhu). Zatímco v pracovních dnech je do této skupiny zahrnuta dálnice D01 na delším úseku Butovice – Věřňovice (PL), o víkendech se jedná pouze o část při hranicích s Polskem Bohumín – Věřňovice (PL). Dálnice D03 je v pracovních dnech zahrnuta svou jižní částí na úseku Tábor – Veselí nad Lužnicí, o víkendech je to zejména severní část na úseku Čekanice – Mezno. Průměrná průjezdnost v pracovních dnech na těchto úsecích byla 57,3 vozidel autobusových dopravců (dále jen vozidla) na pracovní den a o víkendech 61,8 vozidla na víkendový den.



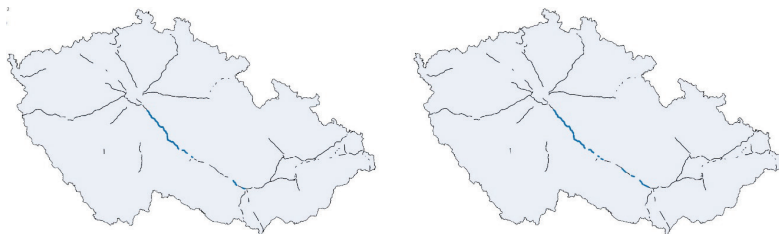
Obr. 9 Největší skupiny pro všechny komunikace nalezené nástrojem SimRoad. Vlevo pracovní dny, vpravo víkendy.

V obou případech našel nástroj skupinu silnic s podobnou průjezdností v okolí Prahy a Brna (obr. 10). V případě pracovních dnů se jednalo o 54 úseků, o víkendech o 52 úseků. Jednalo se zejména v okolí Prahy o úseky dálnice D11 Praha-Počernice – Poděbrady, část dálnice D08, rychlostní silnice R01, R10 a v okolí Brna rychlostní silnice R46 a R52. O víkendy k pražským silnicím přibýly úseky dálnice D05, zatímco v okolí Brna nebyly R46 a R52 zařazeny vůbec, místo nich však vykazuje podobnou průjezdnost dálnice D02. Průměrná průjezdnost v pracovních dnech na těchto úsecích byla 125,9 vozidla na den a o víkendech 150,5 vozidla na den.



Obr. 10 Okolí Prahy a Brna pro všechny komunikace. Vlevo pracovní dny, vpravo víkendy.

Dálnice D01 mezi Prahou a Brnem v obou případech vykazovala podobnost pouze v rámci svých vlastních úseků (obr. 11). V pracovních dnech byla nalezena skupina 26 úseků Modletice – Jihlava a Devět křížů – Brno-centrum a o víkendech 28 úseků, kdy k úsekům z pracovních dnů přibyl úsek Velká Bíteš – Devět křížů a Jihlava – Velký Beranov. Průměrná průjezdnost v obou případech převyšovala úseky všech ostatních komunikací. V případě pracovních dnů se jednalo o 257,8 vozidel na den, v případě víkendů o 311,5 vozidel na den.



Obr. 11 Podobné úseky dálnice D01 mezi Prahou a Brnem, větší část. Vlevo pracovní dny, vpravo víkendy.

Zbýlé úseky dálnice D01 mezi Prahou a Brnem byly zařazeny v obou případech opět do samostatné skupiny, pouze u pracovních dnů k nim byl přiřazen ještě jeden úsek dálnice D08 Praha-Břežiněves – Zdíby. V pracovních dnech jejich průměrná průjezdnost činila 216,7 vozidel na den a o víkendech 280,7 vozidel na den.



Obr. 12 Podobné úseky dálnice D01 mezi Prahou a Brnem, menší část. Vlevo pracovní dny, vpravo víkendy.

9. VÝSLEDKY

Disertační práce Teorie grafů při vyhodnocování časoprostorových dat byla zaměřena na implementaci grafů do tří tematických okruhů, a to v analýze eye-tracking dat, prostorovém plánování a dopravě. Ve všech třech oblastech se práce zabývala vyhledáváním podobných skupin v časoprostorových datech. Výsledkem každého ze tří studovaných okruhů byla navržena metoda pro adekvátní vyhledávání těchto skupin a nástroj, jenž tuto metodu využívá. Funkcionalita nástrojů byla ověřena na třech případových studiích. Nástroje byly využity také v několika dalších případových studiích jiných autorů.

V rámci prvního dílčího cíle byly popsány dosavadní přístupy k vyhodnocování podobnosti mezi daty eye-tracking experimentů, jimiž jsou sekvence navštívených oblastí zájmu. Byly charakterizovány algoritmy, jenž jsou běžně využívány při výpočtu String Edit Distance (Levenshteinova vzdálenost, Damerau-Levenshteinova vzdálenost a Needleman-Wunsch algoritmus). Byly krátce popsány nástroje, jenž s těmito algoritmy pracují a jsou běžně používány pro vyhodnocování podobnosti eye-tracking dat (eyePatterns, ClustalG a TramireR). Byly zhodnoceny možnosti a limity jejich využití. Na základě toho byla navržena nová metoda využívající grafových algoritmů pro vyhledávání podobných skupin mezi eye-tracking daty, která pracuje již bez omezení předchozích nástrojů. Největším přínosem této metody je skutečnost, že při určení míry podobnosti mezi sekvencemi zahrnuje také délku jednotlivých sekvencí a nedochází tak k výpočetním nesrovnalostem při porovnávání množiny sekvencí s výrazně odlišným počtem znaků. Výsledkem této metody je míra podobnosti označovaná jako p , jenž nabývá hodnot z intervalu $(0,1)$, přičemž $p = 0$ v případě zcela odlišných sekvencí (neshodují se v jediném znaku) a $p = 1$ pro totožné sekvence. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných sekvencí. Tuto metodu využívá navržený online nástroj s názvem ScanGraph, který je dostupný na adrese eyetracking.upol.cz/scangraph.

Tento nástroj umožňuje analyzovat podobnost mezi sekvencemi s vyznačenými oblastmi zájmu jednoho či více stimulů. V případě více stimulů je zapotřebí, aby k jednotlivým stimulům byla vázána naměřená data od stejných respondentů. Pro analýzu si uživatel může vybrat mezi třemi metodami, jenž vycházejí z algoritmů Levenshteinova vzdálenost, Damerau-Levenshteinova vzdálenost a Needleman-Wunsch algoritmus. Volba metody záleží na charakteru analyzovaných dat a na cíli, jenž si uživatel stanovil. Kromě samostatných stimulů lze analyzovat i více stimulů najednou a to pomocí dvou metod – Mean a Occurrence.

Druhý dílčí cíl byl zaměřen do oblasti prostorového plánování. Byly popsány běžně používané přístupy k vytváření podobných skupin mezi daty prostorového plánování, tedy základní shlukovací metody a byly také zmíněny metody síťové analýzy, jenž využívají grafových algoritmů. Nástroj, který s těmito metodami pracuje – Urban Analyst Network – byl také krátce charakterizován.

Objekty, jež jsou podrobovány analýze podobnosti, musí být jednoznačně charakterizovány uspořádanou n -tící koeficientů, vyjadřujících hodnoty zkoumaných vlastností. Byla navržena metoda pro vyhledávání podobných objektů mezi daty územního plánování, jež využívá grafových algoritmů pro určování podobných skupin. Od ostatních běžně využívaných metod shlukové analýzy se odlišuje tím, že na jejím výstupu nemusí být zařazeny všechny vyhodnocované objekty do skupin. Uživatel si zvolí míru podobnosti p , jež vyžaduje, aby analyzovaná data vykazovala. Analyzované objekty jsou chápány jako body v n -rozměrném euklidovském prostoru, přičemž jejich souřadnice jsou určeny n -tící koeficientů, jež je charakterizují. Míra podobnosti mezi těmito objekty je pak určena za pomoci euklidovské vzdálenosti mezi nimi. Míra podobnosti p nabývá hodnot z intervalu $\langle 0,1 \rangle$, přičemž $p = 0$ pro objekty s největší možnou vzdáleností a $p = 1$ pro totožné objekty. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných objektů.

Tuto metodu využívá navržený online nástroj s názvem SimUrb, který je dostupný na adrese eyetracking.upol.cz/simurb.

Tento nástroj umožňuje analyzovat podobnost mezi objekty z jednoho či více souborů. V případě více souborů je zapotřebí, aby k nim byla vázána data od stejných subjektů.

Třetí dílčí cíl byl zaměřen do oblasti dopravy. V práci byly zmíněny nejčastější přístupy k dopravním datům, tedy metody síťové analýzy a grafové algoritmy, jež jsou v těchto analýzách aplikovány.

Práce se zaměřuje speciálně na data, kterou jsou charakterizována uspořádanou n -tící koeficientů, vyjadřujících hodnoty zkoumaných vlastností. V této práci jsou jimi hodnoty průjezdů autobusových dopravců mýtnými bránami. Byla navržena metoda pro vyhledávání podobných silničních úseků mezi těmito daty, využívající teorie grafů. Navržená metoda pro výpočet míry podobnosti vychází z metody používané nástrojem SimRoad. Jelikož však data z mýtných bran vykazují velkou hodnotu koeficientu šikmost, byla pro úpravu výpočtu použita kvantilová funkce. Míra podobnosti p ale stále nabývá hodnot z intervalu $\langle 0,1 \rangle$, přičemž $p = 0$ pro objekty s největší možnou vzdáleností a $p = 1$ pro totožné objekty, v tomto případě pro silniční úseky. Současně hodnota této míry podobnosti p roste s rostoucí podobností daných objektů.

Tuto metodu využívá navržený online nástroj s názvem SimRoad, který je dostupný na adrese eyetracking.upol.cz/simroad.

10. DISKUZE

V disertační práci byly navrženy tři metody pro vyhledávání podobných skupin různých objektů časoprostorových dat s využitím teorie grafů a tři nástroje, jež tyto metody používají. Dále byly vypracovány tři případové studie, na nichž byla funkcionalita nástrojů prezentována.

Vytvořené nástroje (ScanGraph, SimUrb a SimRoad) přistupují k problematice určování a vyhledávání podobných skupin obdobným způsobem. Přesto však každý z nich je uzpůsoben na vyhodnocování dat požadovaného charakteru jinými vlastnostmi.

Při zpracovávání disertační práce bylo zvažováno, zda všechny tři přístupy nezahrnout do jednoho nástroje, kde by pak uživatelé bylo umožněno zvolit si příslušný charakter. Jelikož však jednotlivé nástroje mohou být používány pro zcela odlišné experimenty, mohlo by být sjednocení pod jeden nástroj pro uživatele spíše matoucí.

Jako první byla navržena metoda pro vyhodnocování podobnosti sekvencí navštívených oblastí zájmu v eye-tracking experimentech. Z toho důvodu také nástroj ScanGraph, jež tuto metodu používá, disponuje nejširší funkcionalitou. Od svého vzniku byl aplikován v mnoha případových studiích a tedy byl několikrát díky zpětné vazbě uživatelů rozšiřován.

Výstupem nástroje je seznam klik reprezentující respondenty s podobnou strategií. Je zřejmé, že výsledek analýzy závisí na vytvořených oblastech zájmu a na zvolené míře podobnosti.

Je vždy výhodnější vyznačit oblasti zájmu okolo významnějších částí stimulu. V případě kartografie jsou to kompoziční prvky mapy. Nalezené podobnosti mezi sekvencemi pak budou mít vyšší vypovídající hodnotu než v případě pravidelné mřížky bez ohledu na obsah mapy.

Protože podobnost mezi sekvencemi respondentů závisí na pohybech očí a na vyznačených oblastech zájmu, nemůže být jednoznačně řečeno, které nastavení míry podobnosti může být označeno jako ideální. Mezi experimenty se dosažené míry podobnosti podstatně liší. Příhodnou cestou jak najít vhodnou míru podobnosti pro svůj experiment je nejprve zvolením tlačítka Advised graph. Tato funkce určí graf s 5 % hustotou hran a jeho příslušnou míru podobnosti. Dle zkušeností s vyhodnocováním eye-tracking experimentů (Popelka, 2018) má takto zkonstruovaný graf vysokou vypovídající hodnotu. Po vykreslení grafu se zobrazí tabulka pro všechny možné míry podobnosti, jichž může podobnost nabývat. Nejsou uvedeny ty míry, pro které by množina hran grafu byla prázdná. Z této tabulky může uživatel vybrat a zobrazit další variantu grafu pro větší či menší míru podobnosti, dle potřeb jeho experimentu.

Cílem výzkumníka bývá nalézt skupiny s co největší mírou podobnosti, avšak mnohdy naměřená data takovou podobnost nevykazují a tedy ani není možné nalézt skupiny respondentů s velkou podobností. Je pak tedy vhodné zvolit tak vysokou míru podobnosti, jakou data umožňují, což uživatel může přehledně vybrat.

Nástroj ScanGraph má své výpočetní limity. Testováním bylo zjištěno, že je možné určit podobnosti mezi 320 sekvencemi o délce 160 znaků s použitím Levenshteinovy vzdálenosti. Pro metody Needleman-Wunsch algoritmus a Damerau-Levenshteinova vzdálenosti je mezní hodnotou pro funkční výpočet 80 sekvencí o délce 80 znaků. Většina eye-tracking experimentů však skýtá méně než 80 respondentů, tedy toto omezení nástroje není fatálně limitující. Komplikace však může způsobovat délka sekvencí, která při mnohých experimentech 80 znaků překročí. Tento problém může být částečně vyřešen volbou collapsed varianty. Nebudou se tak vyskytovat násobné znaky a sekvence budou kratší.

Snahou je optimalizovat výpočetní algoritmy tak, aby se tyto limitní hodnoty zvýšily. Dalším budoucím přínosem bude přidání funkcionality, která nalezne podobnosti sekvencí z experimentu k jiné uživatelem dané sekvenci. Praktickým použitím je selekce participantů, kteří k prohlížení stimulu zvolili správnou či naopak chybnou cestu.

Nástroj SimUrb neobsahuje tolik funkcionalit jako nástroj ScanGraph. Částečně je to dáno stářím nástroje (mladší než nástroj ScanGraph) a částečně jednodušším určením metriky (není potřeba volit si mezi metrikami či volbu collapsed). Pro použití nástroje SimUrb je nutné, aby jednotlivé objekty určené k vyhodnocování, byly charakterizovány uspořádanou n -ticí kvantitativních koeficientů. Uživatel nahraje takováto data do nástroje SimUrb a zvolí míru podobnosti a minimální velikost skupiny, která je pro něj zajímavá. Výstupem je pak seznam disjunktčních skupin objektů, které vykazují stejnou či vyšší míru podobnosti požadovanou uživatelem. Výsledek analýzy tak závisí na volbě hodnoty míry podobnosti. Pokud má uživatel tu možnost, nahraje i mapový podklad, ve kterém se pak jednotlivé skupiny budou zobrazovat. Není možné stanovit pro všechny analýzy, které nastavení míry podobnosti má být voleno jako optimální. U každého jednotlivého vyhodnocování záleží na rozdělení dat a na cíli, jenž si uživatel stanovil.

Snahou uživatele bývá nalézt skupiny s co největší mírou podobnosti. Mnohdy však vyhodnocované objekty takovou podobnost nevykazují a tedy ani není možné nalézt skupiny s velkou podobností. Uživatel pak tedy musí explicitně vyzkoušet, které nastavení je pro jeho analýzu nevhodnější. Čím nižší je hodnota míry podobnosti, tím větší „benevolence“ je dána objektům zařadit se do skupin.

Nástroj SimUrb vyhledává skupiny podobných objektů tak, aby byly disjunktí. Může se tedy stát, že množinu objektů, jež by také vykazovala požadovanou podobnost, přičemž by její prvky náležely již do jiných skupin, by zanedbal. Jelikož však nástroj při určování skupin dává přednost skupinám o větším počtu objektů, je tedy pravděpodobné, že menší skupiny, jenž by byly zanedbány, by pro uživatele neměly přínosnou hodnotu.

Nástroj SimRoad, jakožto nejmladší z nástrojů, byl prověřen zatím pouze na jedné případové studii. Obdobně jako u nástroje SimUrb je u nástroje SimRoad nutné, aby jednotlivé objekty určené k vyhodnocování byly charakterizovány uspořádanou n -ticí kvantitativních koeficientů. Avšak na rozdíl od nástroje

SimUrb je metrika používaná nástrojem SimRoad uzpůsobena pro data, jejichž rozdělení má velkou hodnotu šikmosti. To znamená taková data, v nichž výrazně převládají objekty s velkou nebo malou vzdáleností.

Uživatel nahraje data v požadovaném formátu a má-li k dispozici, pak i mapový podklad, v němž pak výsledné skupiny vzájemně si podobných objektů budou vizualizovány.

Stejně jako u nástroje SimUrb se může stát, že nástroj zanedbá skupinu silničních úseků, jež sice vykazují podobnost zadanou uživatelem, avšak velikost jejich skupiny je menší, než u jiných nalezených skupin, a tedy je nástroj zanedbá. Obdobně však můžeme říct, že tyto skupiny jsou pro uživatele méně zajímavé.

V nástroji SimRoad je největší možnost dalšího rozvoje. Vlivem toho, že nástroj byl vyzkoušen prozatím na jedné případové studii, postrádá jeho vývoj zpětnou vazbu uživatelů.

11. ZÁVĚR

Cílem disertační práce *Teorie grafů při vyhodnocování časoprostorových dat* bylo navrhnout metody pro vyhodnocování podobnosti mezi časoprostorovými daty a vyhledávání skupin, jež podobnost vykazují a tyto metody aplikovat v případových studiích. Speciálně se disertační práce zaměřila na tři oblasti geoinformatiky. Jednalo se o zpracování eye-tracking dat, úlohy prostorového plánování a dopravu.

Pro vypracování disertační práce bylo nutné seznámit se s dosavadními metodami a programy používanými pro vyhodnocování podobnosti mezi daty tohoto charakteru a nalézt současné aplikace teorie grafů a grafových algoritmů při analýzách těchto dat.

Výstupem prvního dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů navržená speciálně pro eye-tracking data, tedy pro sekvence navštívených oblastí zájmu. Současně byl navržen a vytvořen nástroj ScanGraph, jež tuto metodu používá. Nástroj je volně dostupný na adrese eyetracking.upol.cz/scangraph. Nástroj umožňuje analýzu dat nejen z jednoho stimulu, ale i z několika stimulů tvořících celý eye-tracking experiment. Nástroj byl využit v několika případových studiích a díky zpětné vazbě jeho uživatelů byl při svém vývoji obohacen o nové funkcionality.

Výstupem druhého dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů navržená speciálně pro data definovaná uspořádanou n -tici kvantitativních koeficientů. Současně byl navržen a vytvořen nástroj SimUrb, jež tuto metodu používá. Nástroj je volně dostupný na adrese eyetracking.upol.cz/simurb, který umožňuje kromě samotné analýzy dat vizualizovat nalezené skupiny v mapě. Nástroj je tedy aplikovatelný v úlohách prostorového plánování.

Výstupem třetího dílčího cíle disertační práce byla metoda pro určení míry podobnosti a vyhledávání skupin vzájemně si podobných objektů využívající teorie grafů dat z oblasti dopravy. Metoda je úzce vázána na metodu druhého dílčího cíle. Jelikož však analyzovaná data z oblasti dopravy mají velkou hodnotu koeficientu šikmosti, byla metoda uzpůsobena tomuto charakteru pro názornější interpretovatelnost. Současně byl navržen nástroj SimRoad tuto metodu využívající. Nástroj je volně dostupný na adrese eyetracking.upol.cz/simroad, který umožňuje výsledné skupiny vizualizovat také v mapě.

Náplní čtvrtého dílčího cíle bylo vypracování případových studií, v nichž se navržené metody aplikují. Z oblasti eye-trackingu byla vypracována studie na hodnocení podobnosti mezi strategiemi respondentů prohlížejícími stimuly s vizualizací nejistoty geografických dat. V oblasti prostorového plánování se studie zaměřila na srovnání s výstupem hodnocení udržitelného rozvoje území vypracovaného za pomoci metody ÚAP. Třetí případová studie se věnovala

vyhledávání podobných silničních úseků z hlediska průjezdnosti autobusových dopravců mýtnými bránami na dálnicích, silních I. třídy a rychlostních silnicích.

Kromě případových studií autorky obsahuje disertační práce i ukázky ze studií jiných autorů, v nichž byly navrženy metody a nástroje aplikovány.

Metody a nástroje navržené v této disertační práci jsou aplikovatelné v praxi a v průběhu psaní disertační práce již byly využity řadou uživatelů. Autorka věří, že si své příznivce budou nadále získávat a díky jejich referencím se budou nadále vyvíjet.

12. POUŽITÉ ZDROJE

- Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, 47(4), 1377–1392.
- Appel, K., & Haken, W. (1976). Every Planar Map is Four Colorable 1. *Bulletin of the American Mathematical Society*, 82(5).
- Arlinghaus, S. L., Arlinghaus, W. C., Harary, F., & Arlinghaus, S. L. (2002). *Graph Theory and Geography: An Interactive View* (Ebook).
- Brus, J., Pechanec, V., & Kilianová, H. (2013). Uncertainty vs. Spatial Data Quality Visualisations: A Case Study on Ecotones.
- Coffin, A. W. (2009). Road Network Development and Landscape Dynamics in The Santa Fe River Watershed, North-Central Florida, 1975 to 2005.
- Dhamala, T. N. (2014). A survey on models and algorithms for discrete evacuation planning network problems. *Journal of Industrial and Management Optimization*, 11(1), 265–289.
- Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, 1(4), 620–630.
- Kunderová, P. (1997). Úvod do teorie pravděpodobnosti a matematické statistiky. Olomouc. Univerzita Palackého v Olomouci.
- Lee, J. (2004). A Spatial Access-Oriented Implementation of a 3-D GIS Topological Data Model for Urban Entities. *GeoInformatica*, 8(3), 237–264.
- Longley, P. (2005). *Geographical information systems and science*. Wiley.
- Mackaness, W. A., & Beard, K. M. (1993). Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Systems*, 20(4), 210–221.
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11(9), 929–942.

- Pitts, F. R. (1965). A Graph Theoretic Approach to Historical Geography. *The Professional Geographer*, 17(5), 15–20.
- Popelka, S. (2018). Eye-tracking (nejen) v kognitivní kartografii – praktický průvodce tvorbou a vyhodnocením experimentu, Olomouc. Univerzita Palackého v Olomouci.
- Rathore, M. M., Ahmad, A., Paul, A., & Jeon, G. (2015). Efficient Graph-Oriented Smart Transportation Using Internet of Things Generated Big Data. In *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 512–519.
- Roberts, F. S. (1978). *Graph Theory and Its Applications to Problems of Society*. Rutgers University.
- Roubalík, Š. (2017). Analýza toků veřejné dopravy na základě dat z mýtných bran. Olomouc. Univerzita Palackého v Olomouci.
- Saalfeld, A. (1995). Map generalization as a graph drawing problem (pp. 444–451). Springer, Berlin, Heidelberg.
- Tuček, J., & Ján. (1998). *GIS - Geografické informační systémy: principy a praxe*. Computer Press.
- Zetterberg, A., Mörtberg, U. M., & Balfors, B. (2010). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, 95(4), 181–191.

ODBORNÝ ŽIVOTOPIS AUTORKY / CURRICULUM VITAE

OSOBNÍ ÚDAJE / PERSONAL INFORMATION

Jméno / Name Jitka Doležalová
Bydliště / Address Fibichova 5, 787 01 Šumperk
E-mail jitka.dolezalova@upol.cz
Narozena / Birth 19. 3. 1986, Šumperk



VZDĚLÁNÍ / EDUCATION

- 2013–dosud Univerzita Palackého v Olomouci, **doktorské studium**, obor Geoinformatika a kartografie / *PhD study: Geoinformatics and Cartography*
- 2008–2010 Univerzita Palackého v Olomouci, navazující **Mgr. studium**, obor Diskrétní matematika / *Master's degree: specialization Discrete Mathematics*
- 2005–2008 Univerzita Palackého v Olomouci, **Bc. studium**, obor *Diskrétní matematika* / *Bachelor's degree: Discrete Mathematics*

PRAXE / EXPERIENCE

- 2012–dosud Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra geoinformatiky (technicko-hospodářský pracovník / *clerical staff*)
- 2016–dosud Univerzita Palackého v Olomouci, Centrum popularizace, Pevnost poznání, (koordinátor / *coordinator*)
- 2012–2013 Tutor, s.r.o. (Lektor přípravných kurzů matematiky / *Lecturer of preparatory mathematical courses*)
- 2010–2011 Janta.cz (osobní asistent jednatele společnosti / *Personal Assistant of Managing Director*)
- 2007–2010 Nemovitosti.cz (osobní asistent generálního ředitele / *Personal Assistant of CEO*)

VÝUKOVÉ AKTIVITY / TEACHING

- 2011–dosud individuální kurzy matematiky / *individual math courses*
- 2010–dosud cvičení Matematika, Statistika a Geostatistika / *seminars of: Mathematics, Statistics and Geostatistics*

STÁŽE / MOBILITY

- 2016 Univerzita Komenského v Bratislavě, Slovensko
- 2017 Slovenská technická univerzita v Bratislavě, Slovensko

Seznam vybraných publikací autorky souvisejících s disertační prací / Author's selected publications related to the dissertation

Publikace v odborném časopise v databázi Web of Science (J_{imp}) / Publications in a scientific journal indexed on the Web of Science (J_{imp})

Opach, T., Popelka, S., **Dolezalova, J.**, Rod, J.K. (2017) Star and Polyline Glyphs in a Grid Plot and on a Map Display: Which Perform Better?, Cartography and Geographic Information Science, DOI: 10.1080/15230406.2017.1364169.

Dolezalova, J. & Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. Journal of Eye Movement Research, 9(4), 13p. doi:10.16910/jemr.9.4.5.

Popelka, S., Stachon, Z., Sasinka, C., & **Dolezalova, J.** (2016). EyeTribe Tracker Data Accuracy Evaluation and Its Interconnection with Hypothesis Software for Cartographic Purposes. Computational Intelligence and Neuroscience, 2016, 14. doi: 10.1155/2016/9172506.

Publikace v recenzovaném časopise ve světově uznávané databázi (Scopus) (J_{neimp}) / Publications in a scientific journal indexed on other world database (Scopus) (J_{neimp})

Dolezalova, J., Popelka, S. (2016). The use of simple graphs and cliques for analysis of cartographic eye-tracking data. Paper presented at the ITAT 2016, pp. 206-211. <http://ceur-ws.org/Vol-1649>.

Kapitola v odborné knize (C) / Chapter in a scientific book (C)

Popelka, S., **Dolezalova, J.** (2015): Non-photorealistic 3D Visualization in City Maps: An Eye-Tracking Study. Modern Trends in Cartography Springer International Publishing, pp. 357-367.

Publikace v konferenčním sborníku v databázi ISI Proceedings (D) / Papers in conference proceedings on the database ISI Proceedings (D)

Popelka, S., **Dolezalova, J.**, Beitlova, M. (2018). New Features of ScanGraph - a Tool for Revealing Participants' Strategy from Eye-movement Data. In ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland. ACM, New York, NY, USA, 2 p.

Doležalová, J., Popelka, S. (2016) Evaluation of user strategy on 2D and 3D city maps based on novel scanpath comparison method and graph visualization, ISPRS-International Archives of the Photogrammetry, Remote

Sensing and Spatial Information Sciences, pp. 637-640, doi:10.5194/isprs-archives-XLI-B2-637-2016.

Popelka, S., **Dolezalova, J.** (2016) Differences between 2D map and virtual globe containing point symbols – an eye-tracking study, Geoconference on Informatics, Geoinformatics and Remote Sensing, Vol III, pp. 175-182, doi:10.5593/sgem2016B23.

Miřijovský J., Brus, J., **Doležalová J.**, Místecký, S. (2015): Evaluation of the development agricultural plants for precision farming based on remote sensing methods, SGEM 2015.

Publikace v konferenčním sborníku domácí (ostatní) / Papers in conference proceedings – Czech (other)

Doležalová, J., Popelka, S. (2016) Hledání podobností ve strategii čtení 3D map, Sborník abstraktů příspěvků, ISAF Geomatics in projects; Plan4All conference, Tribun EU, 1 p.

Brychtová, A., **Doležalová, J.** (2015) Sequential Color Schemes Generator 1.0: nástroj pro tvorbu barevných stupnic, Sborník abstraktů příspěvků ze semináře Geomatika v projektech 2015 Tribun EU, 2 p. 978-80-263-0988-8.

Publikace v konferenčním sborníku zahraniční (ostatní) / Papers in conference proceedings – foreign (other)

Popelka, S., **Dolezalova, J.** (2017) Scanpath comparison with the use of web applications SMI2OGAMA and ScanGraph, Eye Movements in Programming: Spring Academy 2017, TR-B-17-02, Berlin, 1 p.

Brychtova, A., **Dolezalova, J.** (2015) Designing Usable Sequential Color Schemes for Geovisualizations, Proceedings of the 1st ICA European Symposium on Cartography, 2 p.

Doležalová, J. (2014): Optimizing colors for effective visualization of geographic data, Third InDOG doctoral conference, 4 p.

13. ANNOTATION

The dissertation thesis is focused on the implementation of graph theory in geoinformatics problems. The main aim of the thesis is to propose tools and methods for finding similarity classes using graph theory. The methods are designed to fit specific data types. The tools are then used in case studies developed by the author, focusing on the evaluation of cartographic works by eye-tracking, analyzing the data of territorial analytical data, and analyzing the road sections. There are also given cases when the tools were used in case studies of other authors.

The first part of the thesis evaluates the relation between simple graphs and spatio-temporal data, introduces basic concepts from graph theory and describes the approaches to evaluation of time-spatial data from eye-tracking data, spatial planning and transport.

The first aim is focused on the analysis of time-space data from eye-tracking. The result is the creation of a new method for searching similar eye-tracking data sequences and the tool that uses this method. The uniqueness of the method consists in taking into account the length of the individual sequences to determine the degree of similarity between them.

The focus of the second aim is directed to spatial planning issues. As a result, it suggests a method for searching similar municipalities based on predefined properties using graph theory, which, however, does not necessarily include all objects in comparison to standard clustering methods. The advantage of this approach is to remove objects (communities) with extreme properties.

The third aim focuses on the analysis of traffic data. The result is a proposal for a method to search for similar road sections and a tool that uses this method. The similarity of road sections is evaluated on the basis of data on the passage of vehicles obtained from toll gates.

The purpose of the fourth aim is to produce three case studies with real data in which the above methods are applied. The first study, which is the application of graph theory in eye-tracking data, touches the visualization of uncertainty of geographic data. The second case study, as a graphical application of spatial planning, deals with the assessment of sustainable land development. The third study applies simple graphs to traffic data and evaluates data from toll gates for road sections for buses. Designed methods and created tools are also applied to case studies of other authors.

Keywords:

Graph Theory, Cliques in a Graph, Similarity, Cluster Analysis

14. SUMMARY

The aim of the dissertation Graph theory in analyses of spatio-temporal data was to propose methods for evaluating the similarity between spatio-temporal data and the search of similarity groups and applying these methods in case studies. In particular, the dissertation focused on three areas of geoinformatics. These were eye-tracking data processing, spatial planning and transport tasks.

For the dissertation thesis it was necessary to get acquainted with the existing methods and programs used for evaluation of similarity between data of this character and to find current application of the theory of graphs and graph algorithms in analyses of these data.

The output of the first aim dissertation thesis was a method for determining the degree of similarity and searching for groups of similar objects using graph theory designed specifically for eye-tracking data, ie for sequences of visited areas of interest. At the same time, ScanGraph was designed and created using this method. The tool is freely available at eyetracking.upol.com/scangraph. The tool allows data analysis not only from one stimulus but also from several stimuli that make up the entire eye-tracking experiment. The tool has been used in several case studies and thanks to feedback from its users it has been enriched in its development with new functionality.

The output of the second partial goal of the dissertation was a method for determining the degree of similarity and searching for groups of similar objects using graph theory designed specifically for the data defined by an ordered set of quantitative coefficients. At the same time, SimUrb was designed and created using this method. The tool is freely available at eyetracking.upol.com/simurb, which allows you to visualize the groups found on the map in addition to the data analysis itself. The tool is applicable in spatial planning tasks.

The output of the third aim dissertation thesis was the method for determining the degree of similarity and searching for groups of similar objects using the theory of transport data graphs. The method is closely linked to the second aim method. However, since the traffic data analyzed have a high value of slope coefficient, the method has been adapted to this character for more readable interpretation. At the same time, SimRoad was designed to use this method. The tool is freely available at eyetracking.upol.com/simroad, which also allows you to visualize the resulting group on the map.

The purpose of the fourth aim was to develop case studies in which the proposed methods are applied. From the field of eye-tracking, a study was conducted to assess the similarity between respondent strategies by viewing the incentives with the visualization of the uncertainty of geographic data. In the sphere of spatial planning, the study focused on the comparison with the outcome of the Sustainable Development Assessment of the Territory. The third case study focused on the search for similar road sections in terms of the accessibility of bus carriers through toll gates on motorways and highways.

Apart from the case studies of the author, the dissertation includes examples from studies of other authors in which the proposed methods and tools were applied.

The methods and tools proposed in this dissertation are applicable in practice and have already been used by a number of users during the dissertation. The author believes that they will continue to earn their fans and, thanks to their references, will continue to develop.

Mgr. Jitka Doležalová

TEORIE GRAFŮ PŘI VYHODNOCOVÁNÍ ČASOPROSTOROVÝCH DAT
GRAPH THEORY IN ANALYSES OF SPATIO-TEMPORAL DATA

Určeno pro studenty, partnerská akademická pracoviště a veřejnost.

Výkonný redaktor prof. RNDr. Karel Lemr, Ph.D.
Odpovědná redaktorka Mgr. Lucie Loutocká
Technická redakce Mgr. Jitka Doležalová

Publikace neprošla redakční jazykovou úpravou.

Vydala a vytiskla Univerzita Palackého v Olomouci
Křížkovského 8, 771 47 Olomouc
www.vydavatelstvi.upol.cz
www.e-shop.upol.cz
vup@upol.cz

1. vydání
Olomouc 2018
Edice GEOINFO-CARTO-THESIS, svazek XIV.
ISSN 1805-7500
ISBN 978-80-244-5330-9

Neprodejná publikace