

## Popis softwarového výsledku

# ANALYZA – Vizualizační komponenta – Visilant

## vytvořeného v rámci řešení projektu

### *Komplexní analýza a vizualizace heterogenních dat velkého rozsahu („ANALYZA“)*

*Období řešení projektu:* 1. ledna 2017 – 31. prosince 2020

*Výzkumný program:* Program bezpečnostního výzkumu České republiky 2015-2020

*Hlavní řešitel:* RNDr. Tomáš Rebok, Ph.D.

#### *Řešitelský tým:*

RNDr. Tomáš Rebok, Ph.D. (2017-2020)	RNDr. Katarína Furmanová, Ph.D. (2019)
RNDr. Michal Batko, Ph.D. (2017-2020)	Bc. Denis Kasanič (2019-2020)
RNDr. Milan Čermák (2017-2020)	Bc. Marko Řeháček (2019-2020)
RNDr. Martin Drašar, Ph.D. (2017-2020)	Bc. Denisa Šrámková (2019-2020)
Mgr. Miroslava Jarešová (2017)	Matej Babej (2019-2020)
doc. RNDr. Barbora Kozlíková, Ph.D. (2017-2020)	Bc. Dávid Brilla (2019-2020)
RNDr. Vladimír Míč, Ph.D. (2017-2018)	Bc. Martin Kažimír (2019-2020)
RNDr. Filip Nálepa, Ph.D. (2017-2018)	Erik Hrcišák (2019-2020)
RNDr. David Novák, Ph.D. (2017-2018)	Bc. Jozef Bátorňa (2019-2020)
RNDr. Daniel Tovarňák, Ph.D. (2017-2018)	Daniel Plakinger (2020)
Mgr. Kristína Zákopčanová (2017-2020)	Vladimír Lazárik (2020)
prof. Ing. Pavel Zezula, CSc. (2017-2020)	Kristián Gutič (2020)
Mgr. Martin Macák (2018)	Ing. Helena Mojdlová (2017)
Mgr. Matúš Guoth (2018)	Ing. Jitka Ročková (2018-2020)
RNDr. Jakub Peschel (2019-2020)	

V Brně, dne 30. ledna 2021

#### **Masarykova univerzita**

Žerotínovo nám. 617/9, 601 77 Brno, Česká republika

T: +420 549 49 1111, E: [info@muni.cz](mailto:info@muni.cz), [www.muni.cz](http://www.muni.cz)

Bankovní spojení: KB Brno-město, ČÚ: 85636621/0100, IČ: 00216224, DIČ: CZ00216224

V odpovědi prosím uvádějte naše číslo jednací.

## Obsah

1	Kontext SW „Vizualizační komponenta“ ve vyvinutém systému .....	3
2	SW „Vizualizační komponenta“ .....	6
2.1	Klasifikace požadavků na vizualizační komponentu .....	6
2.2	Návrh komponenty .....	8
2.3	Vizualizační dokumenty .....	8
2.4	Pokročilé techniky uživatelské interakce .....	9
2.5	Vizuální rozhraní pro komplexní podporu analytických procesů .....	18
3	Návod pro nasazení „Vizualizační komponenty“ .....	28
4	Uživatelská dokumentace .....	29
4.1	Přehledový dashboard .....	29
4.2	Vztahová vizualizace .....	32
5	Programátorská dokumentace .....	36
5.1	Moduly .....	36
5.2	Komponenty .....	38
5.3	GraphLink .....	40
5.4	Služby .....	41
5.5	Formát Dat .....	43
6	Poděkování .....	44

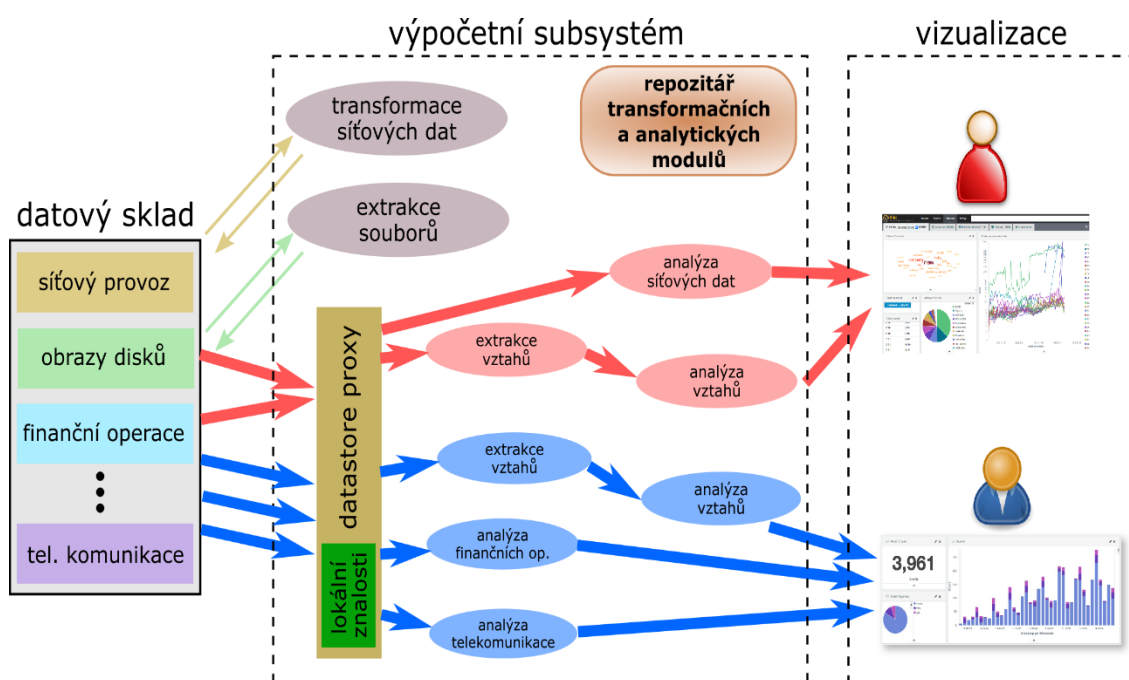
## 1 Kontext SW „Vizualizační komponenta“ ve vyvinutém systému

Hlavním cílem námi řešeného projektu byl návrh a realizace distribuovaného systému umožňujícího komplexní analýzu heterogenních dat velkého rozsahu. Navržený systém je koncipován jako nástroj umožňující jednotné uložení netriviálně velkých datových sad (např. záchytů síťového provozu, obrazů disků, komunikačních informací či znalostí z terénu), nad nimiž jsou vyvolávány nejrůznější mezidoménové analýzy (např. identifikace skupin, které jsou v kontaktu nezávisle na tom, jaký komunikační prostředek volí, případně identifikace po telekomunikační síti komunikujících stran, které si současně vyměňují netriviální finanční prostředky atp.). Takto navržená koncepce má za cíl prozkoumat možnosti a limity podobných mezidoménových analýz, jejichž ambicemi je omezit izolované analýzy různých datových domén a nezbytnost budování mezidoménového kontextu jen v režii analytika. Navržená koncepce vypomůže jak s budováním a udržováním tohoto kontextu přímo nad analyzovanými datovými sadami, tak i s odhalováním nových nebo složitě zaznamenaných datových souvislostí a kontextů, které by jinak byly buď stěží nebo zcela vůbec odhalitelné.

Od tohoto cíle se odvíjely základní požadavky na vyvinutý systém, z nichž ty nejzásadnější lze v krátkosti sumarizovat následovně:

- *Škálovatelnost* – základní požadavek jakéhokoli systému, jehož ambicemi je práce s velkými daty či náročnými výpočty. Nejinak je tomu i v případě námi vyvinutého systému, kdy jak návrh celé architektury, tak i návrh a implementace dílčích komponent podporuje principy distribuovaného či paralelizovaného zpracování a práce s velkými daty.
- *Flexibilita* – neméně zásadní požadavek na systém, jehož ambicemi je práce s různými datovými doménami a různorodými analýzami. Námi vyvinutý systém je vyvinut s předpokladem, že nelze specifikovat konkrétní datové sady a informace, které má systém udržovat, ale naopak musí být připraven v budoucnu jednoduše integrovat různé, i nyní neznámé datové sady. Podobně i všechny analytické funkce systému nebylo možno specifikovat v době jeho vývoje, ale poskytnutím obecného frameworku s dostatečnou flexibilitou je umožněno jejich budoucí doplňování a adaptace novým potřebám.
- *Interaktivita* – požadavek úzce související se škálovatelností systému, jehož ambicí je poskytnout co nejinteraktivnější zpracování a reakce na požadavky uživatele – v závislosti na množství analyzovaných dat – v co nejkratším čase, což je nezbytně důležité zejména pro účely tzv. explorativních analýz (postupného odhalování znalostí).
- *Spolehlivost a důvěryhodnost* – neméně důležité požadavky, posilující důvěru uživatelů v odhalené skutečnosti a schopnosti systému.

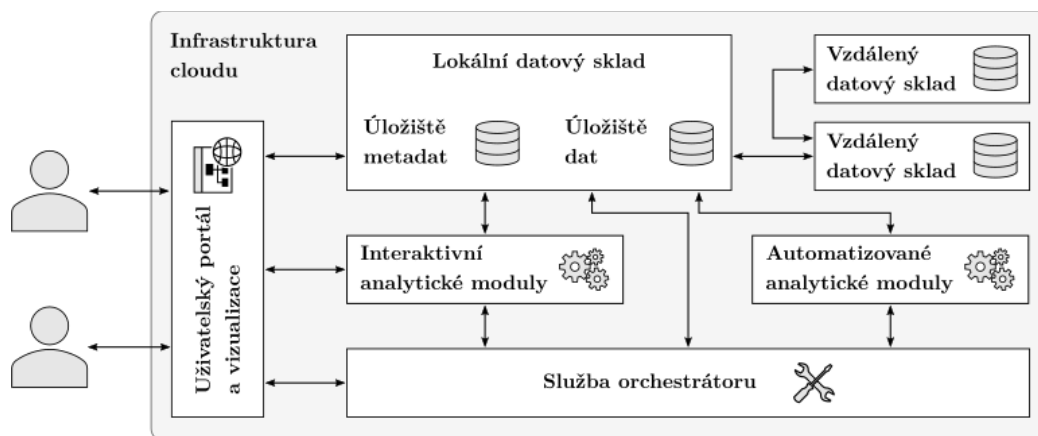
- *Bezpečnost* – jelikož tato netriviální oblast byla nad časový i personální rámec aktivit tohoto projektu, systém jsme vyvinuli tak, aby byl připraven pro dodatečnou implementaci rozsáhlých bezpečnostních mechanismů, včetně práce principů AAI (autentizace, autorizace, identity).



Obrázek 1: Ilustrace myšlenky sjednocených datových analýz s využitím analytických operací sestavených do datového workflow

Pro demonstraci této myšlenky sjednocených analýz (Obrázek 1) jsme navrhli a implementovali architekturu systému, jehož základní komponenty jsou ilustrovány na následujícím obrázku (Obrázek 2). Těmito základními komponentami jsou:

- **Datový sklad** v hierarchické sestavě distribuovaných datových úložišť udržujících veškerá analyzovaná data,
- **Výpočetní subsystém**, poskytující sadu transformačních a analytických modulů a **Orchestrační službu**, která tyto moduly vhodně skládá do datových workflow realizujících transformace dat v Datovém skladu či vlastní datové analýzy,
- **Vizualizační komponentu** poskytující analytické možnosti a prezentující výsledky jednotlivých analýz uživateli (z technického, programátorského a uživatelského pohledu popsaná v následujících kapitolách).



Obrázek 2: Architektura vyvinutého distribuovaného systému pro analýzu rozsáhlých heterogenních dat

Jádro celého implementovaného systému tvoří *analytické moduly*, které vhodně sestaveny do datového workflow (tzv. *analytické operace*) reprezentují analytické možnosti celého modulárního systému. Veškerá vstupní data – včetně poznatků, pozorování a výsledků provedených analýz – udržuje komponenta Datového skladu, ze které jsou potřebná data vyzvedávána analytickými moduly (a následně tamtéž ukládány zpracované mezivýsledky). Analytické moduly jsou ovládány Orchestrátorem, který zajišťuje jejich správné spuštění, včetně předání parametrů specifických pro daný modul. Současně zajišťuje řetězení jednotlivých modulů do analytických operací tak, aby se analytik mohl soustředit pouze na cílový analytický modul, který mu zpřístupní požadovaná data bez toho, aby je musel zpracovávat a připravovat odděleně. Různé analytické operace a moduly jsou spouštěny skrze Vizualizační komponentu, která umožňuje zachycení jednotlivých poznatků a případně výsledků analýz ve formě vztahového diagramu. Na základě výběru jednotlivých uzlů a hran daného grafu pak může analytik vybrat data, se kterými chce zacházet, a předat je Orchestrátoru a jednotlivým analytickým modulům. Jednotlivé komponenty analytického systému tvoří efektivně propojený celek tak, aby se plně využil potenciál pokročilých analýz a nových nástrojů pro zpracování dat spojených s policejním vyšetřováním.

Analytické operace jsou typicky složeny ze specializovaných automatizovaných a interaktivních analytických modulů. Automatizované moduly poskytují funkce transformace daných dat a různé formy detekce na základě předaných parametrů. Výsledky těchto modulů jsou uloženy v rámci Datového skladu, kde je možné s nimi dále pracovat ať už v navazujících analytických modulech nebo ve vztahové vizualizaci. Interaktivní analytické moduly zpřístupňují analytikovi daná data v rámci uživatelského rozhraní, kde je možné je analyzovat pomocí pokročilých funkcí daného analytického nástroje. S využitím pokročilých funkcí Orchestrátoru je možné analytické moduly spouštět samostatně nebo jako klastr distribuovaných výpočtů, pokud to daný modul podporuje.

## 2 SW „Vizualizační komponenta“

Vizualizační komponenta je nedílnou, a z pohledu uživatele také nejdůležitější komponentou celého vyvinutého systému, výrazně ovlivňující jak jeho analytické schopnosti, tak uživatelský komfort. Proto jsme – na základě vstupů z mnoha realizovaných setkání se zástupci policejních vyšetřovatelů – věnovali její analýze významné množství času. Takto identifikované požadavky se navíc dále rozvíjely na dalších realizovaných setkáních, kdy návrh a námi prezentované dílčí řešení policejní analytiky inspirovalo k dalším a dalším nápadům, které jsme se v návrhu a implementaci vyvinutého prototypového analytického systému snažili zohlednit.

Při prvotní analýze a návrhu Vizualizační komponenty jsme vycházeli z dostupného řešení, a to konkrétně z práce s vizualizacemi v rámci nástroje *IBM i2 Analyst's Notebook*, který nabízí řešení vztahové vizualizace dat a je čteně využíván právě v procesu policejních datových analýz. Jako hlavní nedostatky nástroje *IBM i2 Analyst's Notebook* znemožňující efektivní vizuální analýzu jsme zaznamenali:

- omezené možnosti zpracování velkého množství dat,
- omezené možnosti interakce,
- nekvalitní zobrazení velkého množství dat způsobené zejména špatnou čitelností zobrazených informací a četnými překryvy datových položek.

Takto identifikované nedokonalosti jsme se spolu s praktickými vstupy policejních analytiků snažili zohlednit v požadavcích, jejichž klasifikovaný výčet je uveden v následující sekci.

### 2.1 Klasifikace požadavků na vizualizační komponentu

Jak již bylo zmíněno, na základě v úvodu projektu realizovaných a průběžně zpřesňovaných/rozvíjených analýz jsme si stanovili klasifikaci požadavků na Vizualizační komponentu, která se skládá z jednoho primárního požadavku na celou Vizualizační komponentu a desíti dalších požadavků, které jsme rozdělili do dvou kategorií: vizualizační požadavky a interakční požadavky.

#### Požadavky na vizualizační komponentu

##### 1. *Množství různých pohledů na data velkého rozsahu nabízející rozdílné úrovně detailu*

Návrh vizualizační komponenty poskytne množství různých pohledů na data s několika úrovněmi detailu prezentovaných informací. Volba těchto pohledů a úrovní bude závislá na množství zobrazovaných dat a současně preferencích analytika. Jednotlivé pohledy pokryjí rozsah od vysokoúrovňových, abstraktních vizualizací tisíců datových záznamů, až po nízkoúrovňové vizualizace prezentující informace o jednotlivých záznamech.

## Vizualizační požadavky

### 1. Vztahová vizualizace

Primární vizualizací bude vztahová vizualizace, která bude prezentovat jednotlivé entity a jejich vzájemné vztahy. Bude tak představovat základní pohled na data v Datovém skladu.

### 2. Specializované vizualizace

Množina interaktivních specializovaných vizualizací poskytne uživateli odpovědi na specifické dotazy a nabídne možnost zobrazit specifické datové typy vhodným způsobem (např. finanční, telekomunikační, či časoprostorová data). Tyto vizualizace budou moci být propojeny se vztahovou vizualizací, respektive s dalšími specializovanými vizualizacemi, pokud to bude žádoucí.

### 3. Časoprostorové vizualizace

Vizualizace zaměřené na analýzu časoprostorových dat se zaměřením na intuitivní reprezentaci dat a interakční techniky. Tyto vizualizace umožní různými způsoby zkoumat časovou souslednost událostí, například pomocí animace nebo manuální interakcí s časovou osou.

### 4. Export dat ve vizuální podobě

Možnost exportovat analytické výsledky, které budou prezentované ve vhodném kontextu s možností výběru úrovně detailu, která bude závislá na cílovém čtenáři, tedy například soudním znalci. Export bude probíhat jak do klasické textové podoby, tak do vizuální reprezentace (například jako sada prezentačních slides).

### 5. Intuitivní a pochopitelná reprezentace dat

Jednotlivé vizualizace je třeba navrhnout tak, aby jejich reprezentace dat byla intuitivní a pochopitelná pro uživatele z kriminální oblasti. To znamená, že zvolená reprezentace dat bude vycházet jak z poznatků o lidském vnímání, tak z konvencí používaných v oblasti kriminálního vyšetřování.

## Interakční požadavky

### 1. Real-time prozkoumávání dat

Vizualizační komponenta umožní prozkoumávání dat v reálném čase se zaměřením na filtrování, pokročilé interakční techniky a dostupnost specializovaných vizualizací.

### 2. Pokročilé interakční techniky

Kombinace navigačních a interakčních technik, které usnadní orientaci ve vizualizacích a zároveň podpoří vizuální analýzu.

### 3. Lokální modifikace datových položek

Analytik bude moci do vizualizací přidat doplňující informace buď vytvořením nových datových záznamů či úpravou existujících záznamů. Tyto změny bude následně možné pozorovat ve vizualizacích, aniž by byly ovlivněny data v centrálním datovém skladě.

### 4. Nastavení prostředí a preferencí

Analytik si bude moci nastavit preference pro vizualizační komponentu, jako například preference mapování vizuální reprezentace, preference nastavení jednotlivých vizualizací, vytvoření šablon atd.

### 5. Vizualizační dokumenty (workspaces)

Každá vizuální analýza bude probíhat ve vizualizačním dokumentu, ve kterém bude uložen aktuální stav analýzy a všechna aktuální nastavení vizualizační komponenty (například nastavení úrovně detailu zobrazení jednotlivých uzlů nebo skupin uzlů).

## 2.2 Návrh komponenty

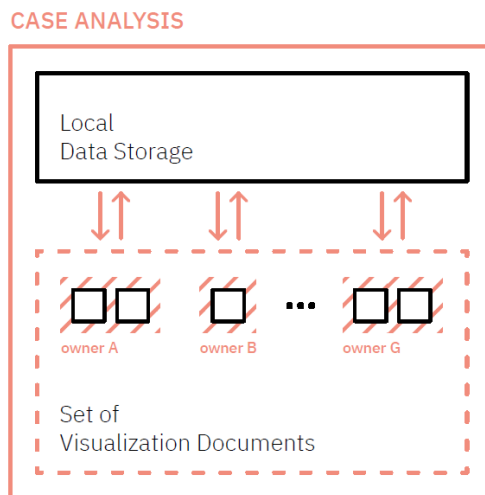
Návrh vyvinuté vizualizační komponenty je úzce spjat s pracovním postupem kriminálních vyšetřovatelů. Na základě několika schůzek se partnery projektu a diskusích o hlavních požadavcích na vizualizační komponentu, které jsou uvedeny výše, bylo rozhodnuto, že klíčovou vizualizací v celé komponentě bude *vztahová vizualizace*. Tato reprezentace přehledným způsobem zobrazuje zvolenou podmnožinu vstupních dat (vybranou pomocí dotazů na centrální datový sklad). Za podpory vysoké míry uživatelské interakce, vztahová vizualizace umožňuje uživateli získat přehled o klíčových objektech pro vyšetřování a jejich vzájemných vztazích. Na tuto vizualizaci budou následně napojeny doplňující specializované vizualizace. Mezi ně bude patřit například jednoduchá přehledová vizualizace, která uživateli předá informaci o množství a typu vybrané podmnožiny dat v kontextu celého obsahu datového skladu týkajícího se daného případu. Mezi další vizualizace podporující práci analytika bude patřit například vizualizace finančních transakcí mezi vybranými subjekty (jejich výběr může být proveden interakcí se vztahovou vizualizací) či časoprostorová vizualizace pohybu subjektů (např. sledování pohybu vozidla).

## 2.3 Vizualizační dokumenty

Každá vizuální analýza (viz Obrázek 3) probíhá v tzv. *vizualizačním dokumentu*. Vizualizační dokument umožňuje zobrazit data pomocí vztahové vizualizace, která může být následně využívána pro vyvolávání doplňkových specializovaných vizualizací. Vizualizační dokument ukládá současně aktuální stav a nastavení související s vizualizací vybrané množiny dat, tzn. změny, které vyvolal uživatel, jako například přemístění uzlu vztahové vizualizace, agregace uzlů do skupin a jiné operace ovlivňující zobrazení vizualizace. Vizualizační dokument si můžeme představit



jako analogii textového dokumentu Word, pouze s výjimkou toho, že nebude prezentovat text, ale vizuálně znázorní jednotlivé datové položky z datového skladu.



Obrázek 3: Diagram analytického projektu zobrazující organizaci vizualizačních dokumentů

Vizualizační dokument může být v libovolný moment uložen a opět načten později. Tento způsob práce a uložení aktuálního stavu současně poskytuje možnost pro sdílení práce mezi analytiky. Analytik současně může v rámci analytického projektu vytvořit libovolné množství vizualizačních dokumentů. Tyto dokumenty mohou být potom využity pro vzájemné porovnání.

## 2.4 Pokročilé techniky uživatelské interakce

Vizuální reprezentace dat nabízí výrazné usnadnění čitelnosti a pochopení velkého množství dat, nicméně teprve až ve spojení s technikami interakce představuje přínosný nástroj pro vizuální analýzu. Tento nástroj pak umožňuje nejen snadnější orientaci ve vizualizaci, ale především analytickou navigaci, prozkoumávání dat v reálném čase a zpřístupňování informací na různých úrovních detailu.

Do návrhu a prototypové implementace vztahové vizualizace jsme v rámci vyvinutého systému proto zahrnuli následující interakce:

- *lokální prozkoumávání dat* – zpřístupnění dostupné informace o uzlech a vazbách;
- *technika focus+context* – rychlá orientace ve vizualizaci pomocí zvýrazňování aktuálně důležitých objektů (uzlů a vazeb) za současného vizuálního potlačení ostatních objektů;
- *výběry* – efektivní výběr množiny uzlů a vazeb;

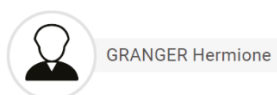
- *lokální modifikace dat* – zavádění nových informací do vizualizace pomocí doplnění nových uzlů či vazeb a jejich atributů;
- *agregace uzlů* – snadná organizace uzlů do skupin;
- *anotace* – doplnění nové znalost, pozorování, či otázky, na které je třeba odpovědět, v podobě tagů, zvýrazňování objektů či poznámek;
- *záložky* – ukládání a následné zpřístupnění vybraných pohledů na data, včetně filtrů, seřazení dat, agregací a dalších uživatelských interakcí;
- *rychlé operace* – zpřístupnění častých a důležitých operací, které je možné nad objekty vyvolat.

## Lokální prozkoumávání dat

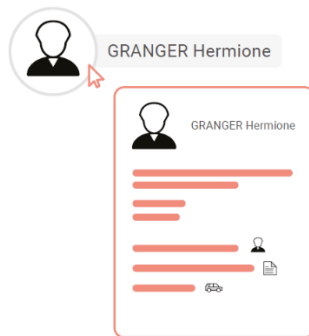
Cílem lokálního prozkoumávání dat je umožnit analytikovi zpřístupnit dostupné informace o jednotlivých uzlech či vazbách mezi nimi. Tato technika je zásadní vzhledem k tomu, že se ve vizualizacích bude pracovat s velkým množstvím dat, které není možné zobrazit najednou a současně je třeba zachovat čitelnost vizualizace.

S ohledem na rozdílnou důležitost a potřebu jednotlivých informací, lokální prozkoumávání dat umožňuje na data o uzlech a vazbách nahlížet ve třech úrovních detailu. Tyto úrovně detailu utvářejí jednoduchou hierarchii, od úrovně zobrazující pouze zásadní informace umožňující uzel či vazbu jasně identifikovat, po úroveň zpřístupňující veškeré informace, které jsou o uzlu či vazbě dostupné. Jednotlivé úrovně jsou následující:

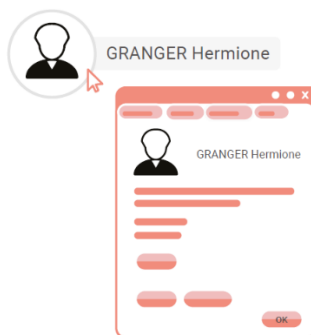
1. *okamžitá informace* (viz Obrázek 4) poskytující pouze klíčové informace pro identifikaci objektu, které jsou zobrazeny ve štítku objektu;
2. *přehledová informace* (viz Obrázek 5) zobrazující pomocí tooltipu ty informace, které analytik zvolil jako důležité;
3. *detailní informace* (viz Obrázek 6) zpřístupňující veškeré informace, které jsou o objektu dostupné v popup okně.



Obrázek 4: Ukázka zobrazení okamžité informace objektu



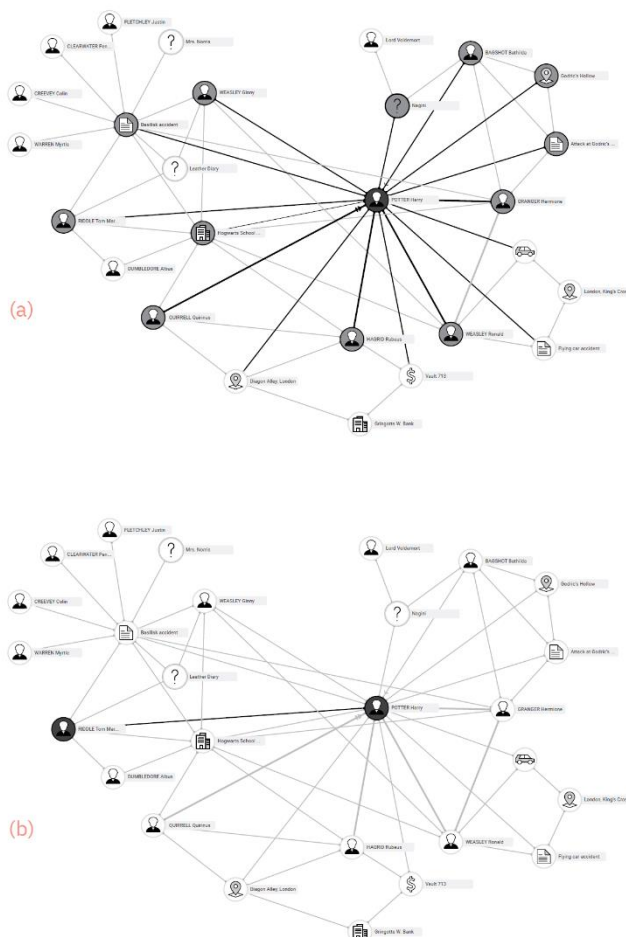
Obrázek 5: Ukázka zobrazení přehledové informace objektu



Obrázek 6: Ukázka zobrazení detailní informace objektu

## Technika focus+context

Vztahová vizualizace může obsahovat velké množství uzlů, které jsou vzájemně propojeny a nemusí být snadné se v takové vizualizaci zorientovat či nalézt propojené uzly. Technika focus+context (viz Obrázek 7) je navržena s cílem tyto překážky překonat. Poskytuje jednak rychlý přehled vazeb jednotlivých datových položek a současně usnadňuje nalezení propojení mezi uzly a vazbami a tím současně nalezení sousedících uzlů ve vizualizaci.

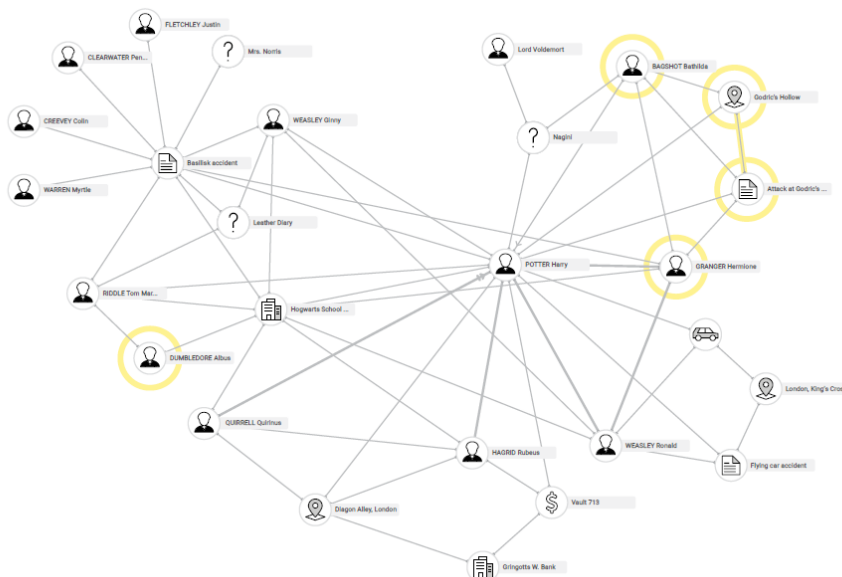


Obrázek 7: Znárodnění techniky focus+context. (a) ukázka zvýraznění uzlu, jeho vazeb a sousedů. (b) ukázka zvýraznění vazby a uzlů, které daná vazba propojuje

## Výběry

Nedílnou součástí interakčních technik vizuální analýzy je možnost vytváření výběrů (viz Obrázek 8), které umožňují definovat podmnožinu uzlů a vazeb, která typicky sdílí nějakou vlastnost, ať už na základě dat či úvah analytika. Nad vybranou množinou uzlů a vazeb mohou být následně vyvolány další operace, například zvýraznění, tagování, či seskupení uzlů.

Analytik může definovat množinu uzlů a vazeb buď manuálně, tedy výběrem jednotlivých objektů nebo využitím filtrování. Množina vybraných uzlů a spojnic může být následně pojmenována a uložena jako pojmenovaný výběr ve vizualizačním dokumentu.



Obrázek 8: Ukázka reprezentace výběru množiny uzlů

## Lokální modifikace dat

Důležitou součástí nástroje pro vizuální analýzu je umožnit zavádění nových informací do vizualizace a následné pozorování, jak tyto změny ovlivňují samotnou vizualizaci. Tyto informace je možné zavést buď v podobě nových uzlů a vazeb nebo jako nové atributy již existujících datových záznamů. Každý nově doplněný záznam je uložen současně s identifikací uživatele, který záznam vytvořil, a časovou značkou vytvoření.

Je důležité, aby takto nově zavedené informace, které mohou být v mnoha případech pouze domněnky či nepotvrzené informace, byly pro analytika odlišitelné od informací, které jsou součástí centrálního datového skladu. Z tohoto důvodu mají nové informace odlišnou vizuální podobu oproti původním datům (viz Obrázek 9). Zatímco uzly a vazby z centrálního datového skladu jsou znázorňovány pomocí plného obrysu, nově přidané uzly a vazby jsou znázorněny pomocí šrafovaného obrysu.

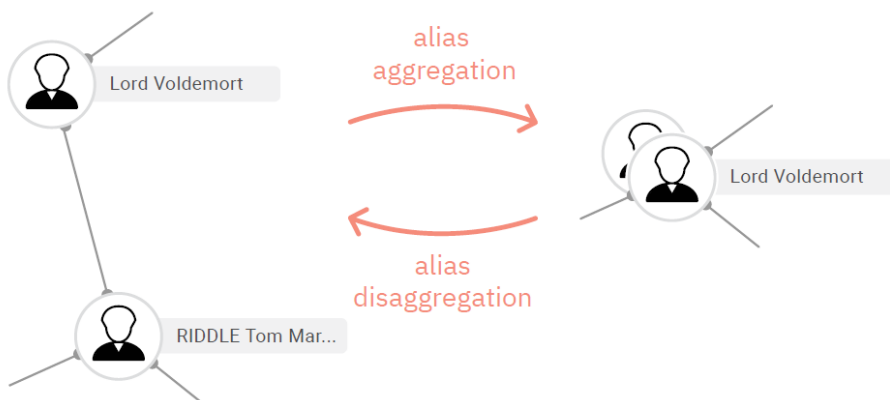


Obrázek 9: Ilustrace zobrazující rozdílnou vizuální reprezentaci pro objekty obsažené v centrálním datovém skladě a objekty doplněné analytikem

## Agregace uzlů

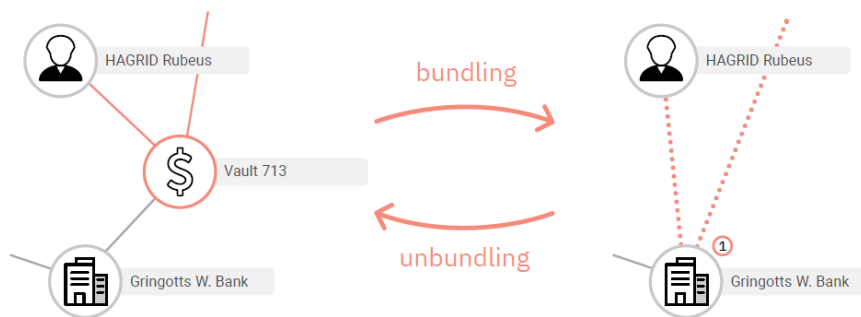
V průběhu explorativní analýzy dat je agregace uzlů velmi užitečná a důležitá operace, která umožňuje měnit strukturu dat zaváděním logických vztahů mezi uzly a často vede ke zjednodušení pohledu na data. Pro doménu kriminálního vyšetřování jsme navrhli následující tři možné způsoby agregace, které mohou podpořit vizuální analýzu – seskupení aliasů, bundling, obecná skupina. Každý typ agregace má speciální vizuální reprezentaci.

- *Seskupení aliasů* (viz Obrázek 10) je typ agregace navržený pro situaci, kdy jsou dva či více objektů podezřelé z reprezentování téhož objektu v reálném světě. Typickým příkladem je situace, kdy máme dva objekty reprezentující tutéž osobu. Tento typ agregace následně umožňuje pozorovat, jak sloučení objektů ovlivní vazby a vztahy vůči ostatním objektům a může umožnit zpozorovat nové poznatky.



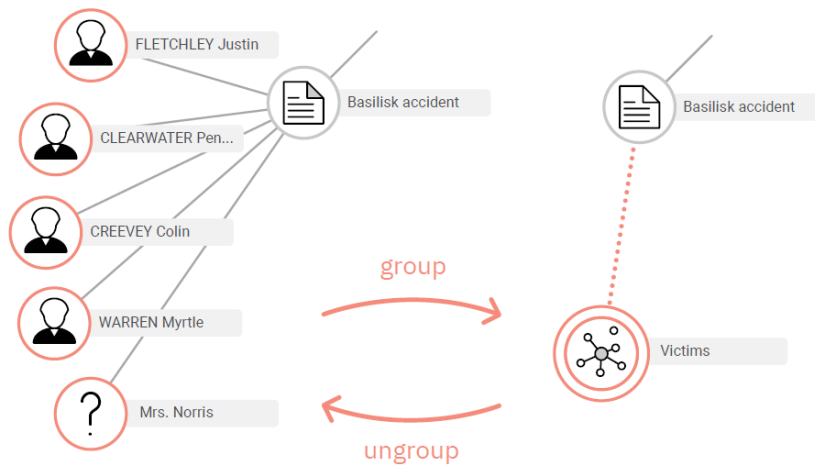
Obrázek 10: Ukázka seskupení a rozdělení aliasů

- *Bundling* neboli zabalení (viz Obrázek 11) je dalším typem agregace, který umožňuje skrýt (zabalit) aktuálně nedůležité uzly pod sousedící uzly, na které má logickou či hierarchickou vazbu. Například v případě, že osoba vlastní auto, které není důležité pro aktuální analýzu. Uzel reprezentující auto se může pomocí operace zabalení skrýt pod uzel reprezentující osobu, která auto vlastní. Současně se vedle uzlu, pod kterým jsou skryté další uzly, zobrazí ukazatel (bublínka) s číslem reprezentujícím počet uzlů, které jsou pod ním skryté.



Obrázek 11: Ukázka zabalení a rozbalení uzlů

- *Obecná skupina* (viz Obrázek 12 **Chyba! Nenalezen zdroj odkazů.**) je poslední typ agregace, který umožňuje seskupit libovolnou množinu uzlů, pojmenovat ji a reprezentovat ji nadále jen jedním uzlem.



Obrázek 12: Ukázka vytvoření a zrušení skupiny

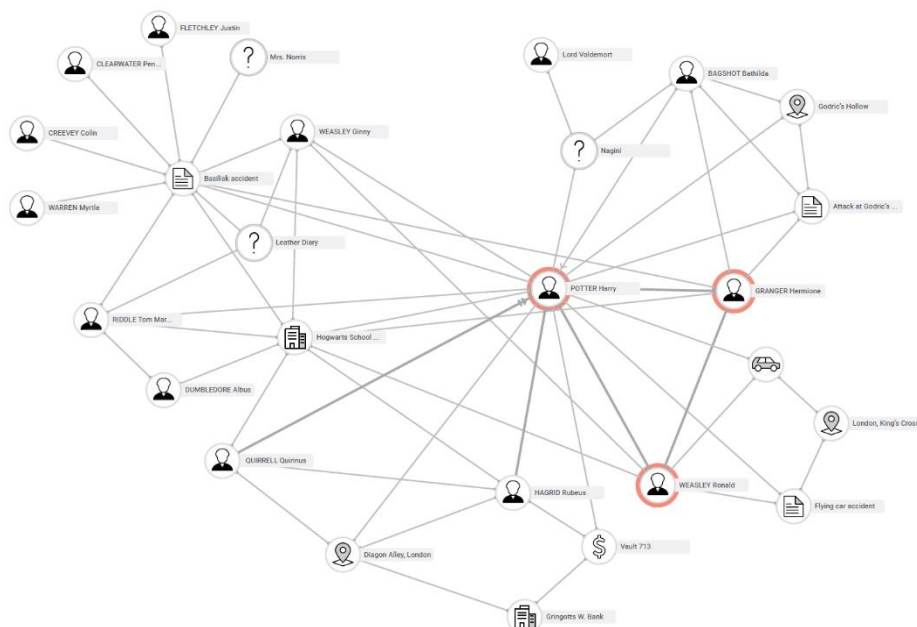
## Anotace

Operace anotace nabízí analytikovi možnost zahrnout do vizualizačního dokumentu vlastní pozorování, otázky, úkoly či jen obecné pracovní poznámky. Anotace podporují proces vizuální analýzy v základní verzi nabízí tři typy anotačních technik: focus, tagy a poznámky.

*Focus* (viz Obrázek 13) je anotační technika, která umožňuje vizuálně zdůraznit ty objekty, které jsou v rámci analýzy považovány za klíčové. Takovéto vizuální zvýraznění umožňuje objekt vždy snadno zpozorovat v rámci celé vizualizace a zároveň vytváří jakýsi výchozí bod pro další analýzu dat, který usnadňuje orientaci ve vizualizaci.







Obrázek 13: Ukázka anotační techniky focus ve vztahovém diagramu

Tagy nabízí další možnost vizuálního zvýraznění, které poskytuje obecný způsob, jak zvýraznit vybrané zajímavé vlastnosti objektů, které jsou relevantní pro aktuální analytický projekt. Tag může být vytvořen pro reprezentování informace, jako například “oběti útoku” či “odsouzení v minulosti”, nebo jen pro analytickovy pracovní poznámky, např. “nekompletní informace” či “podezřelý”.



Každý tag se skládá z pojmenování a barvy a je následně přiřazen jednotlivým objektům. Jednotlivé tagy se zobrazují v okolí objektu jako barevné body, tudíž v následném prozkoumávání jednotlivých objektů analytik dostává rychle podstatnou informaci, kterou si v minulosti zvýraznil. V rohu vizualizace se současně nachází interaktivní legenda tagu, která zobrazuje všechny aktuální tagy. Po njetí na jednotlivý tag v legendě se zvýrazní ty objekty ve vizualizaci, které mají přiřazený daný tag.

Poznámky umožňují analytikovi vložit jeho myšlenky, pozorování či obecné pracovní poznámky do vizualizace v podobě textových poznámek. Tyto se mohou nacházet na třech úrovních:

1. jednotlivé uzly či vazby ve vztahové vizualizaci;
2. vizualizační dokument;
3. analytický projekt.

S každou poznámkou je zároveň uložena identifikace analytika, který ji vytvořil, současně s časovou známkou vytvoření. Tyto poznámky mohou být také zpřístupněny ostatním analytikům v rámci analytického projektu.

## Záložky

Analytikovy interakce, manipulace s daty, změny vizualizačních nastavení, filtrování či výběry mohou měnit celkové rozložení a zobrazení dat ve vizualizaci. Je tedy důležité, aby tyto změny mohly být zachovány pro budoucí práci, sdílení s kolegy či aby se k nim dalo zpětně vrátit. Záložky umožňují, aby tyto interakce byly uloženy jako vizualizační záložka, která může být znovu zpřístupněna v jakýkoliv moment analýzy.

## Rychlé operace

V průběhu analýzy je běžné, že některé operace bývají vyvolávány častěji než jiné a je tedy žádoucí, aby byly i rychleji přístupné analytikovi. Za tímto účelem jsme vytvořili panel rychlých operací (viz obrázek napravo) pro jednotlivé uzly či vazby, které zahrnují důležité a často využívané operace, jako je například tagování, focus či skrytí nedůležitých uzlů. Panel rychlých operací se zobrazí po kliknutí myši na uzel či vazbu a jeho obsah může být uzpůsoben analytiky dle jejich potřeb.



## 2.5 Vizualní rozhraní pro komplexní podporu analytických procesů

Během testování a diskutování Vizualizační komponenty s kriminalisty na konci roku 2019 jsme odhalili významnou výzvu pro vylepšení dosavadní podpory analytického procesu vyšetřování, které je často velmi komplexní a představuje netriviální kognitivní zátěž pro vyšetřovatele. Velký výzkumný i aplikační potenciál této výzvy nás inspiroval k vytvoření návrhu a prototypu vhodných vizuálních reprezentací, které by tento proces podpořily. Tento netriviální analytický proces totiž může být významně podpořen vhodnými vizualizačními technikami a přístupy, proto jsme návrh a implementaci Vizualizační komponenty prototypově rozšířili (byť bez hlubší integrace se zbylými komponentami vyvinutého systému) o nové vizuální rozhraní pro interaktivní uchování a prozkoumávání jednotlivých analytických stavů, jejich sdílení, synchronizaci a export.

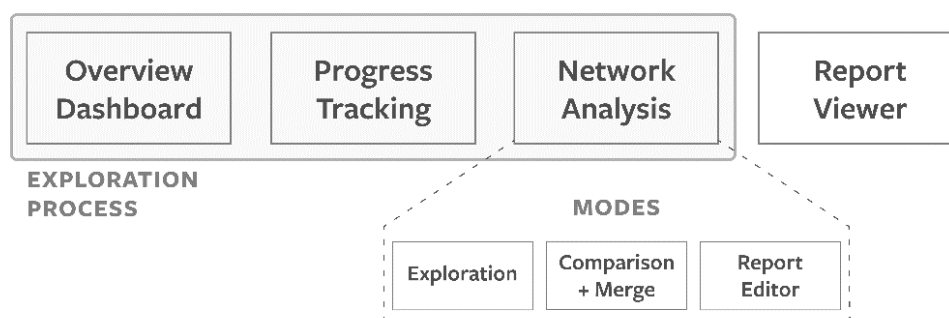
Navržené rozšíření analytických procesů jsme rovněž popsali v publikaci<sup>1</sup> představující tento navržený vizuální nástroj, která byla publikována na prestižní mezinárodní konferenci IEEE VIS 2020

---

<sup>1</sup> Zákopčanová, K., Řeháček, M., Bátor, J., Plakinger, D., Stoppel, S., Kozlíková, B. Visilant: Visual Support for the Exploration and Analytical Process Tracking in Criminal Investigations. *IEEE Transactions on Visualization and Computer Graphics*. IEEE, 2021, 10 s. ISSN 1077-2626. doi:10.1109/TVCG.2020.3030356.

a v roce 2021 bude rovněž uveřejněna v impaktovaném časopise *IEEE Transactions on Visualization and Computer Graphics*. Součástí bylo vytvoření prototypové verze rozhraní, které bylo následně testováno kriminalisty. Nicméně plná integrace tohoto rozhraní již byla nad rámec časových možností řešení projektu v daném roce a tvoří tedy jednu z plánovaných aktivit po skončení projektu.

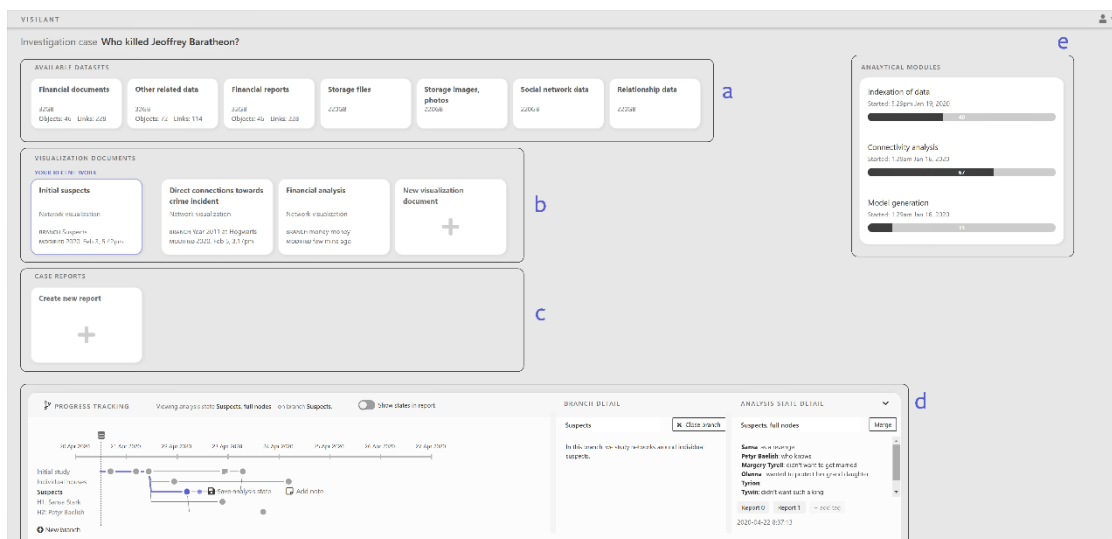
Navržené vizuální rozhraní představuje koncepční řešení návrhu vizuální reprezentace procesu externalizace, sledování vývoje vyšetřování a spolupráce mezi kriminalisty. Dále se zaměřuje na podporu nelineárního způsobu prozkoumávání dat, který odráží způsob práce kriminalistů, ověřujících řadu vstupních hypotéz. To v důsledku vede k rozsáhlé síti potenciálních vyšetřovacích větví, které je nutné prozkoumat, aby bylo možné dané hypotézy potvrdit či vyvrátit. Pro tyto účely jsme navrhli nástroj, který se skládá z následujících částí (viz Obrázek 14):



Obrázek 14: Schéma navrženého nástroje a jeho částí

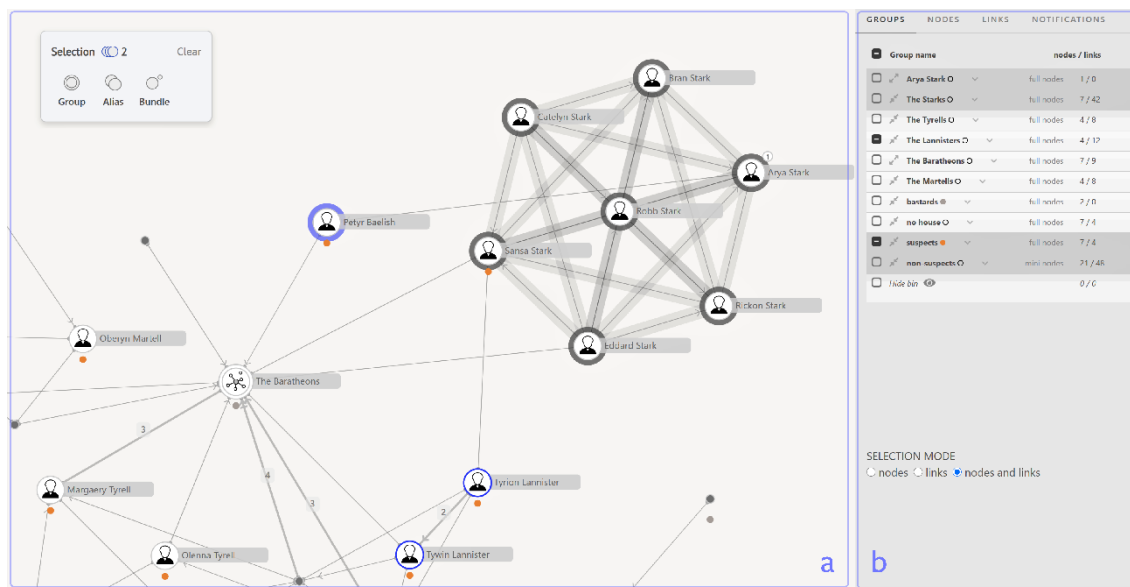
První komponenta představuje vstupní bránu do procesu vyšetřování, kde uživatel získává přehled o aktuálně dostupných datech (viz Obrázek 15, část a), která jsou uložena v centrálním Datovém skladu. Zde si uživatel následně zvolí podmnožinu dat, kterou chce aktuálně prozkoumat a zahrnout do vyšetřování, čímž nadefinuje a vytvoří takzvaný *vizualizační dokument*. Tímto termínem označujeme schéma celého vyšetřovacího procesu, od počátečního výběru dat, přes vytvoření vhodných vizuálních reprezentací a náhledů na data, po potvrzení či vyvrácení hypotéz a výsledné uzavření případu, s případným předáním důkazních podkladů pro další zpracování. v rámci *Overview Dashboard* panelu tak uživatel zároveň získává přehled o již vytvořených vizualizačních dokumentech a může je dále prozkoumat (viz Obrázek 15, část b). V další části tohoto panelu může vyšetřovatel vygenerovat report z výsledků vyšetřování pro státního zástupce (viz Obrázek 15, část c, detaily o reportu dále), a může sledovat aktuálně spuštěné analytické procesy a jejich stav (viz Obrázek 15, část e). Dolní část panelu obsahuje tzv. *Progress Tracking* diagram (viz Obrázek 15, část d), což je nově navržená vizuální reprezentace, která je přítomna skrze celý

analytický proces a tvoří přehledové rozhraní pro nelineární exploraci procesu vyšetřování (detaily v samostatné sekci níže).



*Obrázek 15: Overview Dashboard a jeho součásti: a) seznam dat dostupných v centrálním datovém skladu, b) seznam již vytvořených vizualizačních dokumentů a možnost vytvořit nový, c) seznam vytvořených výsledných reportů o případu a možnost vytvořit nový, d) Progress Tracking Diagram pro podporu nelineárního prozkoumávání hypotéz a trasování vyšetřovacích větví, e) seznam aktuálně běžících analytických procesů*

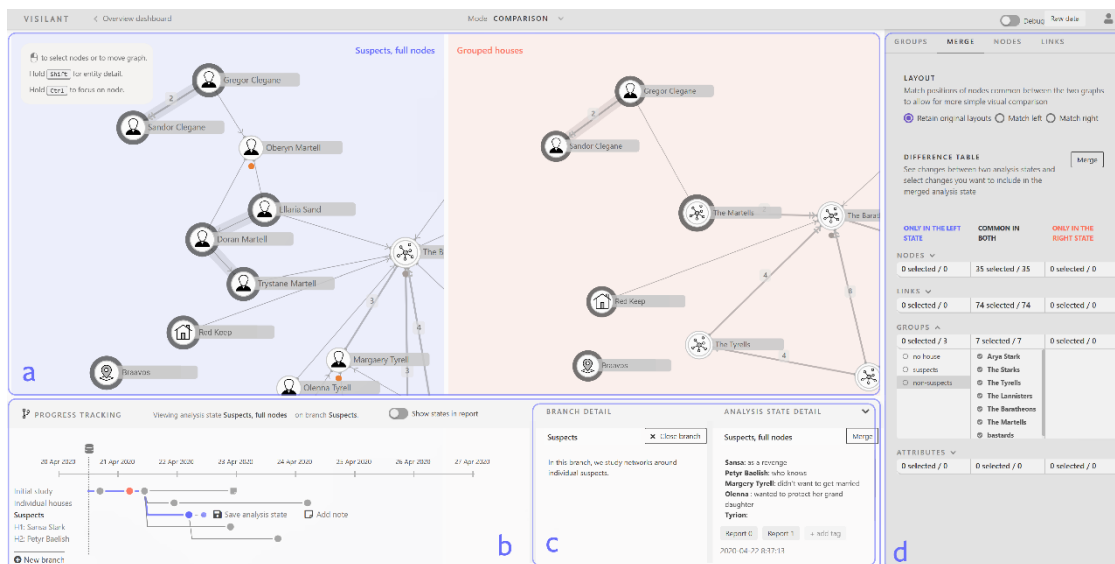
V rámci jednotlivých fází vyšetřování, kdy je nutné prozkoumat vztahy mezi zkoumanými entitami, nástroj využívá *Network Analysis* panel, jehož ústřední částí je vztahová vizualizace, která byla navržena a detailně popsána již v předchozích zprávách projektu (viz Obrázek 16).



Obrázek 16: Network Analysis panel, jehož centrální část tvoří vztahová vizualizace

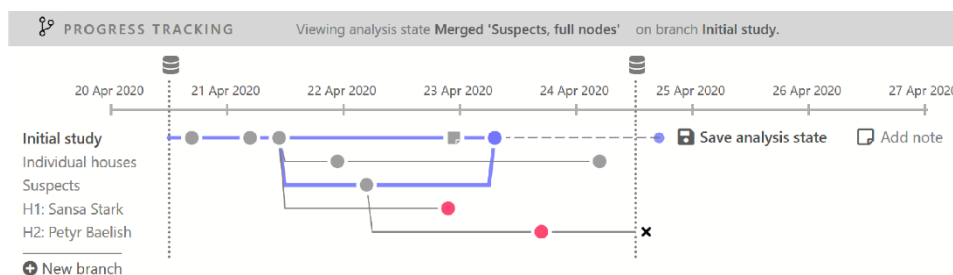
Tento panel operuje ve třech základních módech:

1. **Exploration mode** – v tomto módu vyšetřovatel prozkoumává jednu instanci vztahové vizualizace.
2. **Comparison mode** – tento mód slouží k porovnání aktuálních stavů dvou vztahových vizualizací (viz Obrázek 17, část a). Ty mohou pocházet buď od různých vyšetřovatelů nebo z různých časových období daného vyšetřování (více v části o *Progress Tracking*). V tomto módu je možné dané vztahové vizualizace spojit do jedné – uživatel může ve vedlejším panelu (viz Obrázek 17, část d) definovat uzly, které mají dané vizualizace společné a rovněž které si má spojená vizualizace převzít z levé a pravé vizualizace. Tímto způsobem je možné intuitivně spojovat například stavy vyšetřování téhož případu, prováděné různými vyšetřovateli.
3. **Report mode** – zde má vyšetřovatel možnost vytvořit vizuální reprezentaci vztahových vizualizací pro generování výsledného reportu z vyšetřování. Uživatel zde může doplnit textové popisky pro jednotlivé stavy a objekty a definovat sekvenci snímků vztahové vizualizace, která postihuje vývoj vztahů v čase.



Obrázek 17: Network Analysis panel v Comparison módu, kdy jsou porovnávány vztahové vizualizace dvou stavů vyšetřování (a). Součástí nástroje je možnost tyto vztahové vizualizace spojit do jedné, na základě předem definovaných uživatelských preferencí. Ty je možné definovat v panelu (d). V dolní části panelu se opět nachází Progress Tracking diagram (b) a část obsahující details o vybrané větvi (c).

Progress Tracking diagram (viz Obrázek 18) je jednou z nejvýznamnějších částí celého návrhu. Umožňuje zaznamenávat, sledovat a zpětně trasovat průběh celého vyšetřování. Obsahuje sadu uložených stavů, které jsou generovány buď automaticky (pro předem definované operace) nebo na základě uživatelského požadavku. Jednotlivé vyšetřovací větve odpovídají větvení diagramu. Každá větev obsahuje sadu uložených stavů, po jejich aktivaci se uživateli zpřístupní vztahová vizualizace odpovídající danému stavu. Při spojení dvou stavů z různých větví pomocí Comparison módu (viz Obrázek 17) dojde ke spojení větví v tomto diagramu. Jednotlivé stavy, resp. jejich vztahové vizualizace, mohou být využity v Report módu panelu Network Analysis, který generuje finální report.

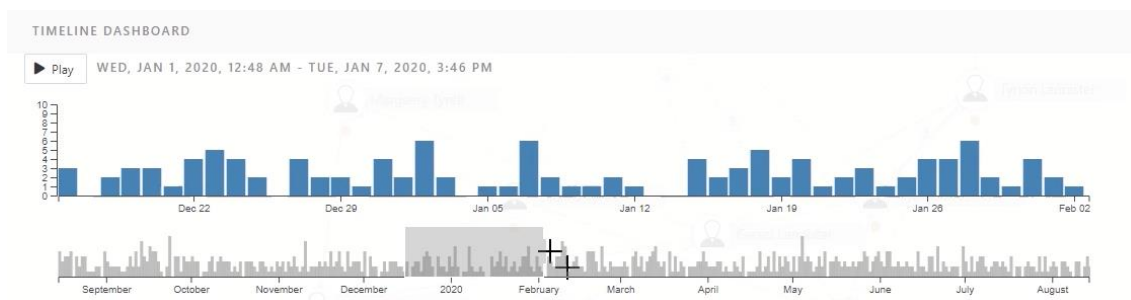


Obrázek 18: Progress Tracking diagram

Pro zhodnocení vhodnosti a využitelnosti navrženého přístupu nástroj postoupil dvoufázové testování s experty, jehož průběh je popsán v další kapitole věnující se testování komponent implementovaného systému.

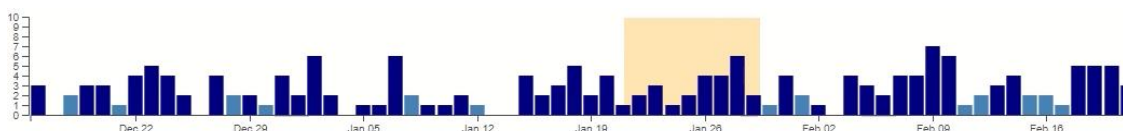
## 2.5.1 Vizualní prozkoumávání časových a prostorových závislostí mezi objekty

V rámci této části jsme se nejdříve zaměřili na návrh a implementaci způsobu zacházení s časovými známkami v datech, kdy jednotlivé objekty v sobě mohou nést informaci o době svého trvání. Návaznost jednotlivých časových úseků pak může vést k detekci důležitých návazností událostí. Zkoumání časových závislostí objektů je úzce spjato se vztahovou vizualizací, nově navržený panel (viz Obrázek 19) je tedy se vztahovou vizualizací úzce propojen.



Obrázek 19: Panel pro prozkoumávání časových závislostí v datech

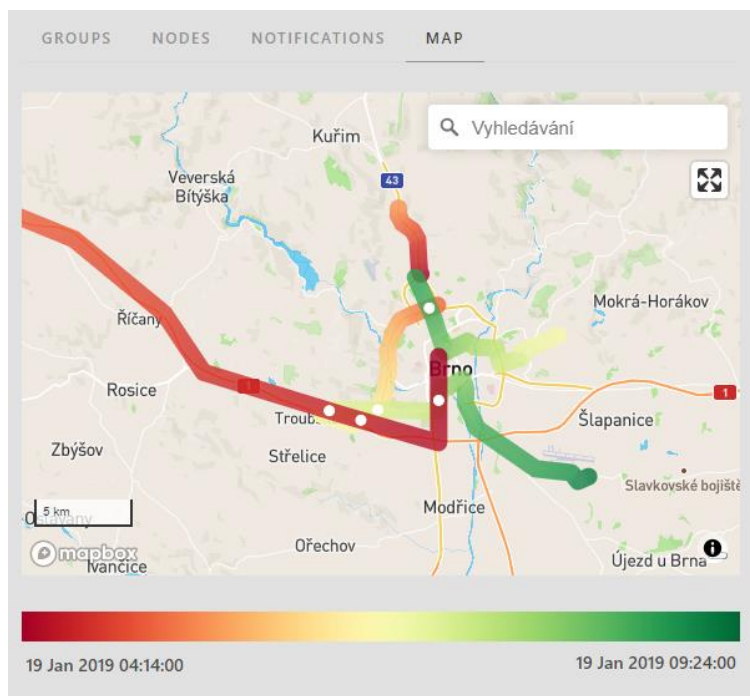
Panel obsahuje dva sloupcové diagramy. Spodní diagram slouží jako přehledový a navigační prvek, kde je udržován přehled počtu objektů s časovou známkou v daném časovém rozsahu (počet takových objektů je projektován do výšky sloupce v daném čase) a uživatel zde může vybrat časový rozsah zájmu, který je pak zobrazen ve větším detailu v horním sloupcovém grafu (modrý). Stisknutím tlačítka *Play* je spuštěna animace vztahové vizualizace, kdy objekty, které v daném časovém kroku mají platnou časovou známku (tj. hrají roli v daném čase) jsou vizuálně zvýrazněny pomocí selekce. Pro aktuální moment (zvýrazněný žlutým obdélníkem) je navíc v grafu zvýrazněno (tmavě modrou barvou), ve kterých časových okamžicích mají objekty důležité v tomto okamžiku rovněž časovou známku (tj. jsou potenciálně důležité). V rámci přehrávání si uživatel navíc může nastavit časové okno pro sledování, které se v rámci přehrávání iterativně posouvá (viz Obrázek 20).



Obrázek 20: Nastavení delšího časového okna (žlutě) pro přehrávání



V jistých případech hraje významnou roli nejen časová složka, ale rovněž informace o lokaci, ve které k dané události došlo. Pro tyto případy jsme navrhli a implementovali dodatečný panel (viz Obrázek 21), jenž přehledně zobrazuje časovou souslednost událostí na mapovém podkladu. Jednotlivé čáry na mapě představují trajektorie zvoleného objektu či objektů (uzlů ve vztahové vizualizaci) a obarvení těchto trajektorií informuje uživatele o časové návaznosti těchto trajektorií (události reprezentované červenými trajektoriemi se udály dříve než ty zelené). Jednotlivé trajektorie jsou rovněž zobrazeny barevnou škálou, je tedy možné jednoduše určit, ve kterém místě událost začala a kde skončila.

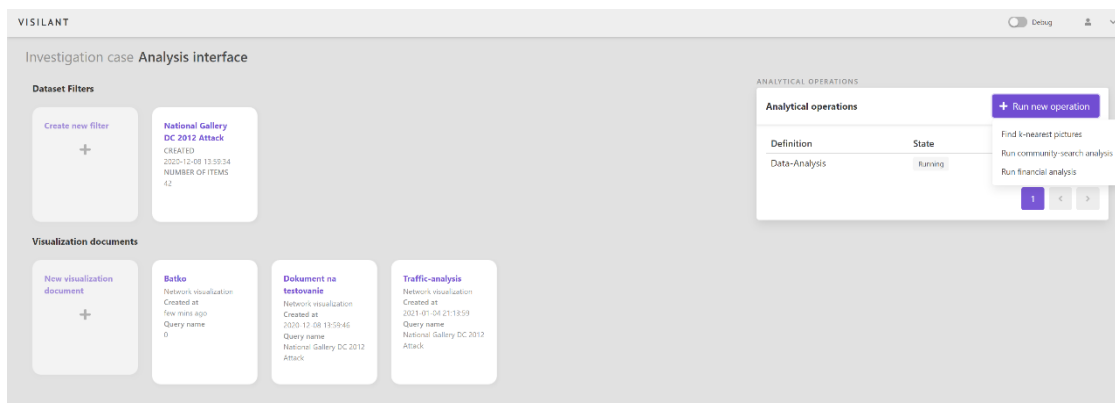


Obrázek 21: Panel pro zobrazení prostorové lokalizace událostí v čase

## 2.5.2 Integrace Vizualizační komponenty

V rámci posledního roku řešení projektu jsme se významně soustředili na integraci Vizualizační komponenty do celého vyvíjeného systému a jednotlivé demonstrační případy užití. Centrálním prvkem systému je dashboard rozhraní (viz Obrázek 22), které bylo navrženo v rámci návrhu vizuálního rozhraní pro podporu analytických procesů, detailně popsáno výše.





Obrázek 22: Implementované dashboard rozhraní vyvinutého systému. Rozhraní umožňuje vytvářet a spravovat výběry dat pro analýzu z centrálního Datového skladu, jednotlivé vizualizační dokumenty a analytické operace nad daty.

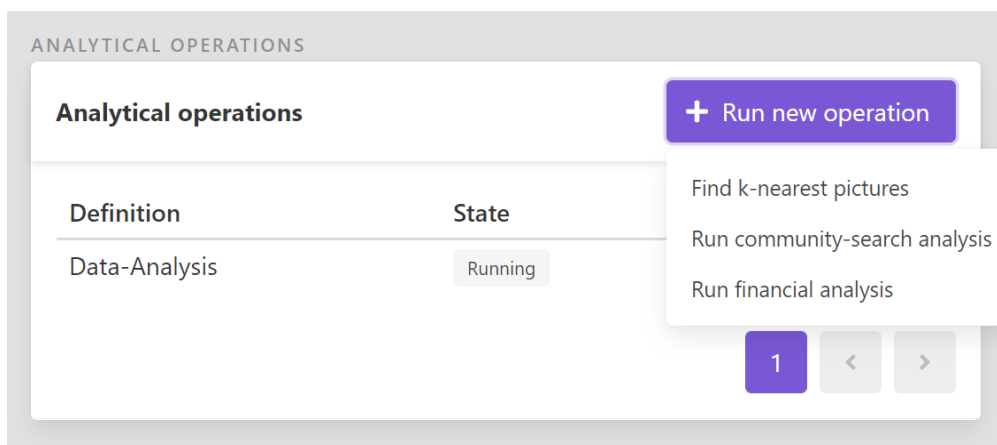
Nejdůležitějším prvkem rozhraní je vytváření jednotlivých analytických operací, které jsou úzce navázány na implementované demonstrační případy užití, detailně popsané v předcházející roční zprávě a následující kapitole. Ty můžeme z hlediska ovládání rozdělit na dvě kategorie:

4. analytické operace, které jsou spouštěny z hlavního dashboard rozhraní,
5. analytické operace, které jsou spouštěny pro vybraný uzel vztahové vizualizace.

Do první kategorie řadíme následující implementované případy užití:

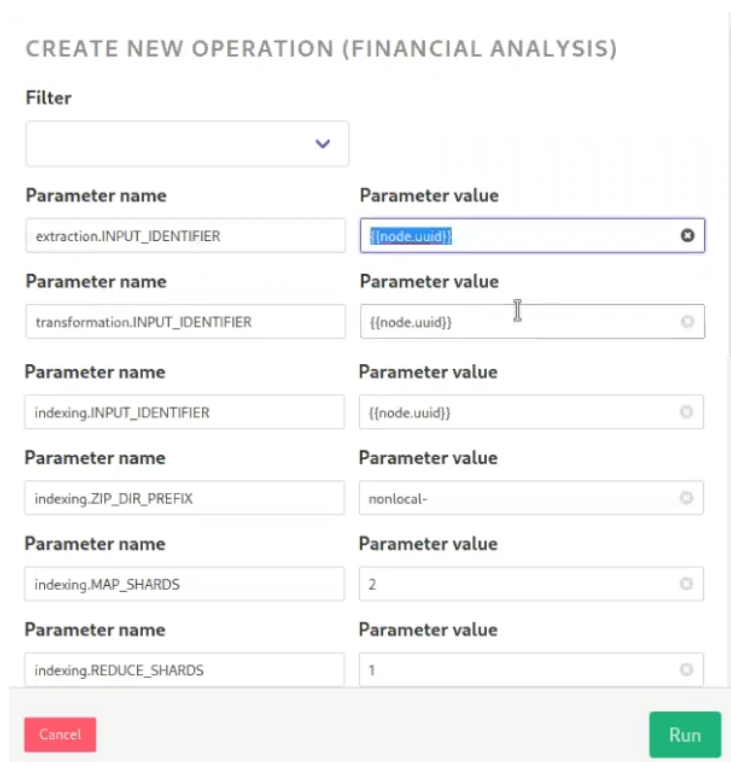
- vyhledávání vizuálně podobných obrázků,
- identifikace komunit,
- identifikace podezřelé činnosti s fixními procesy.

Na následujícím obrázku (Obrázek 23) lze vidět menu v rozhraní, kde uživatel volí mezi těmito možnostmi.



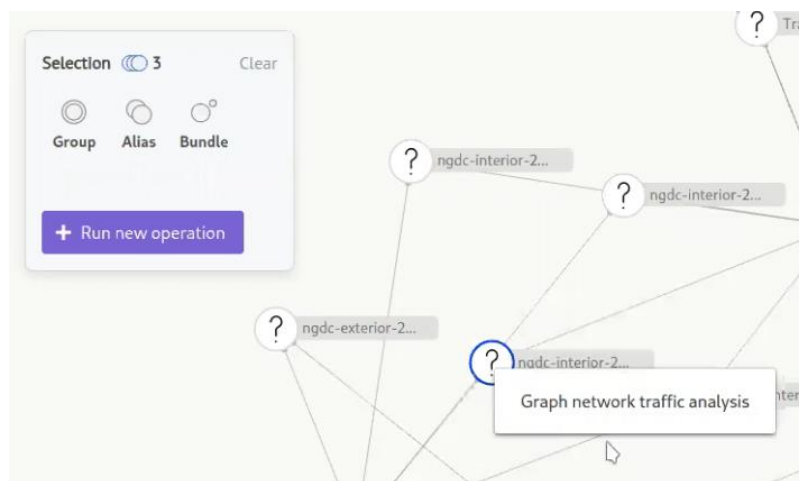
Obrázek 23: Vytváření jednotlivých analytických operací v dashboard panelu

Při vytváření dané analytické operace je uživateli zobrazeno okno pro specifikaci jednotlivých parametrů, které je nutné pro spuštění operace pro daný případ užití zadat. Obrázek 24 zobrazuje ukázkou takového okna pro případ identifikace podezřelé činnosti s fixními procesy.



Obrázek 24: Okno pro vytvoření nové analytické operace

Specifickým případem užití je případ, kdy je pro spuštění analytické operace nutné, aby uživatel zvolil vybraný objekt zájmu, tj. uzel ve vztahové vizualizaci (kategorie 2). Tuto operaci je tedy možné spustit až v situaci, kdy máme vybraná data zobrazena pomocí síťové vizualizace (mám již vytvořený odpovídající vizualizační dokument). Obrázek 25 ilustruje tento případ, kdy je možné po kliknutí na vybraný uzel ve vztahové vizualizaci spustit analytickou operaci pro vztahovou analýzu komplexních síťových událostí.



Obrázek 25: Spuštění síťové analýzy pro daný uzel ve vztahové vizualizaci

Výše uvedený popis integrace Vizualizační komponenty a její integrace do celého systému je pouze základním přehledem podporované funkcionality. Detaily lze nalézt v dokumentaci vytvořené komponenty.

## 3 Návod pro nasazení „Vizualizační komponenty“

Vizualizační komponentu je možné spustit ve dvou režimech. První využívá vývojové prostředí jazyka NodeJS (verze 14+ včetně) spolu s nástrojem pro správu balíčků yarn. Tento způsob je vhodný, pokud je zapotřebí komponentu dále modifikovat, protože dochází k vygenerování potřebných souborů a jejímu následnému zobrazení. V případě produkčního spuštění je druhým způsobem vybudování Image za pomoci Dockeru (verze 19+ včetně). Oba tyto způsoby jsou popsány níže v relevantních sekcích, společným krokem je rozbalení přiloženého zip souboru obsahujícího zdrojové kódy a následná konfigurace.

### Konfigurace

Vzhledem k použití renderování na straně klienta je nutno tyto konfigurace nastavit předem ještě před samotným spuštěním. Pro fungování komponenty je potřebné správné napojení na ostatní součásti systému jako jsou Datový sklad, Orchestrátor a interní rozhraní vizualizační komponenty, sloužící k ukládání dat uživatele. Obě tyto komponenty mají své API rozhraní, pomocí kterého je navázáno spojení pomocí protokolu http předáním URL adresy odkazující na toto rozhraní. Konfigurační soubor se nachází ve složce public uložený pod názvem *config.js*. V tomto souboru je potřeba změnit odkazy na tyto služby ponecháním klíčových hodnot a nahrazením jim odpovídajícím hodnotám.

### Spuštění z vývojového prostředí

- Zadáním příkazu `yarn` pro stáhnutí balíčků
- Následným příkazem `yarn serve` se zobrazí na výstupu terminálu webová adresa, z níž je dostupná vizualizační komponenta.

### Spuštění pomocí vybudování Docker Image

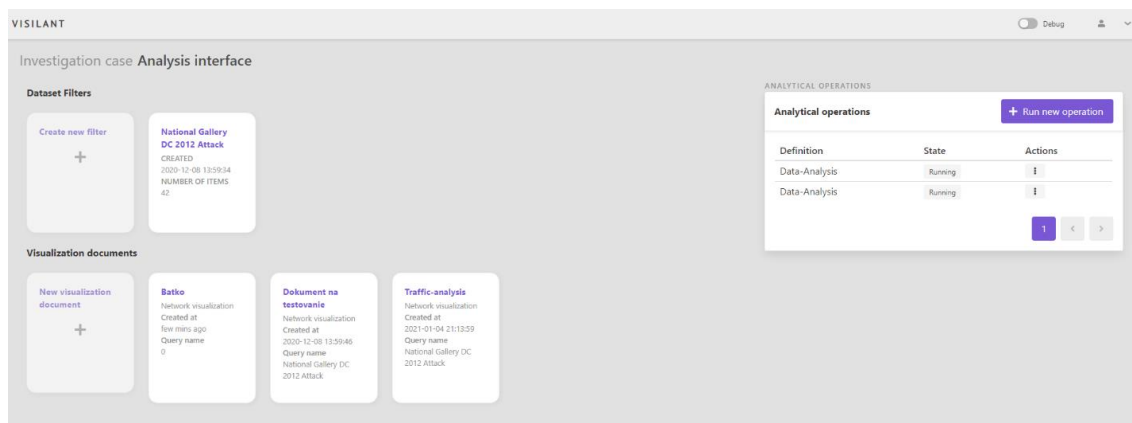
- I v tomto případě je nutno nainstalovat balíčky pomocí příkazu `yarn`
- Dále spuštěním příkazu `yarn build` vygenerujeme zobrazení pro webového klienta
- Příkazem `docker build -t visualization` spustíme v hlavním adresáři vytvoření Docker Image
- Následným příkazem `docker run -it -p 80:80 visualization` se spustí kontejner běžící na portu 80
- Vizualizační komponentu si můžeme zobrazit na webové stránce `http://localhost`

## 4 Uživatelská dokumentace

Tato sekce obsahuje stručný uživatelský návod na použití vizualizační komponenty výsledného nástroje.

### 4.1 Přehledový dashboard

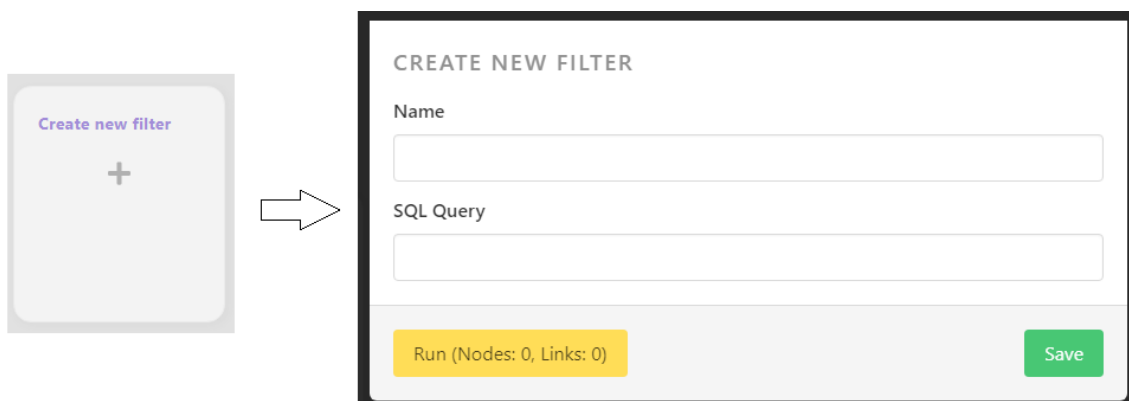
Dashboard tvoří základní vstupní rozhraní celé aplikace. Obsahuje základní prvky pro vytvoření nové analytické operace, vytvoření filtru nad daty a vytvoření nového vizualizačního dokumentu, který operuje nad vyfiltrovanými daty. Náhled vstupního dashboardu zobrazuje Obrázek 26.



Obrázek 26: Přehledový dashboard, který tvoří vstupní prvek do celé aplikace

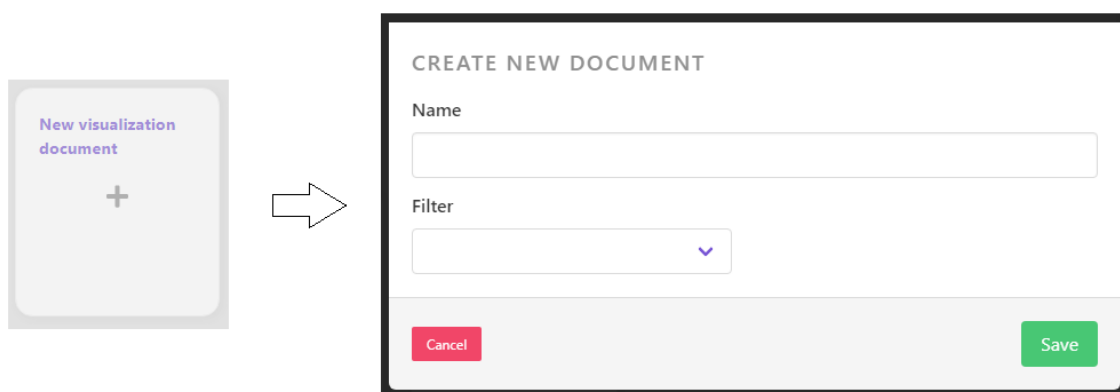
V rámci vstupního dashboardu uživatelé získávají přehled o vytvořených filtrech, existujících vizualizačních dokumentech (které mohou být opětovně navštíveny) a aktuálně běžících operacích.

Výběr dat pro nově vytvářenou analýzu je prováděn následujícím způsobem. Uživatel vybere možnost „Create new filter“ a následně se otevře dialogové okno (Obrázek 27), kde uživatel zadá unikátní jméno nového filtru a SQL dotaz do centrálního datového skladu, na jehož základě dojde k výběru odpovídajících dat a jejich zpřístupnění uživateli pro další prozkoumávání.



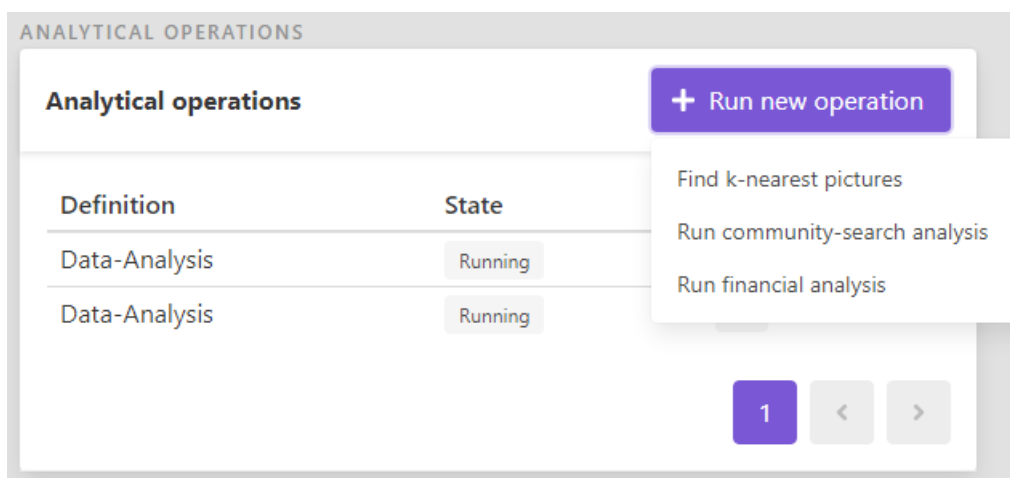
Obrázek 27: Vytvoření nového filtru pro výběr dat z centrálního datového skladu

Při vytváření nového vizualizačního dokumentu, který reprezentuje jednotlivé prováděné analýzy nad daty, je postupováno obdobným způsobem. Uživatel nejdříve vybere možnost „New visualization document“, která otevře nové dialogové okno (Obrázek 28). V něm uživatel zadá identifikační název nového dokumentu a v menu „Filter“ vybere data, která byla získána pomocí filtru popsaného výše. V tomto menu se objeví seznam všech již vytvořených datových filtrů. Po vytvoření vizualizačního dokumentu se tento objeví, podobně jako datové filtry, v přehledovém dashboardu.



Obrázek 28: Vytvoření nového vizualizačního dokumentu a nastavení dat, se kterými má operovat

V rámci vstupního dashboardu je dále možné vytvářet jednotlivé analytické operace nad zvolenými daty. Panel pro tyto operace (Obrázek 29) obsahuje jak přehled jednotlivých spuštěných operací a jejich stav, tak i tlačítko pro vytváření nových operací („Run new operation“). Pod touto volbou se nachází dodatečné menu, jež slouží k výběru typu operace, odpovídající podporovaným případem užití.



Obrázek 29: Panel pro vytváření jednotlivých analytických operací a jejich přehled

Jednotlivé, na tomto okně podporované, případy užití jsou:

- nalezení k-obrázků ve zvolené datové sadě, které jsou nejpodobnější zadanému obrázku (např. fotografii),
- vyhledávání komunit ve zvolené datové sadě dle integrovaného algoritmu,
- spuštění analýzy finančního toků nad danou datovou sadou.

Každá z těchto možností vyvolá nové dialogové okno, ve kterém musí uživatel zadat dodatečně požadované parametry pro analýzu. Například, dialogové okno pro spuštění operace nalezení k-nejpodobnějších obrázků ve zvolené datové sadě (Obrázek 30). Nejprve musí uživatel vybrat existující datovou sadu (předem vytvořenou z hlavního dashboardu) a poté nastavit další parametry. Uživatel může vyhledávání blíže specifikovat díky možnosti přidávat libovolné množství vlastních parametrů. Po stisknutí tlačítka „Run“ se spustí požadovaná operace a její stav lze vidět v přehledové tabulce v dashboardu.

The dialog box is titled "CREATE NEW OPERATION (K-NEAREST PICTURES)". It features a "Filter" dropdown menu. Below it are two rows of parameter configuration. The first row has "extraction.filter" as the parameter name and "{{selectedNodes}}" as the parameter value. The second row has "Group.ParameterName" as the parameter name and an empty field for the parameter value. There is a blue "Add parameter" button below the second row. At the bottom of the dialog are a red "Cancel" button and a green "Run" button.

Obrázek 30: Dialogové okno pro vytvoření analytické operace vyhledání k-nejpodobnějších obrázků v dané datové sadě

## 4.2 Vztahová vizualizace

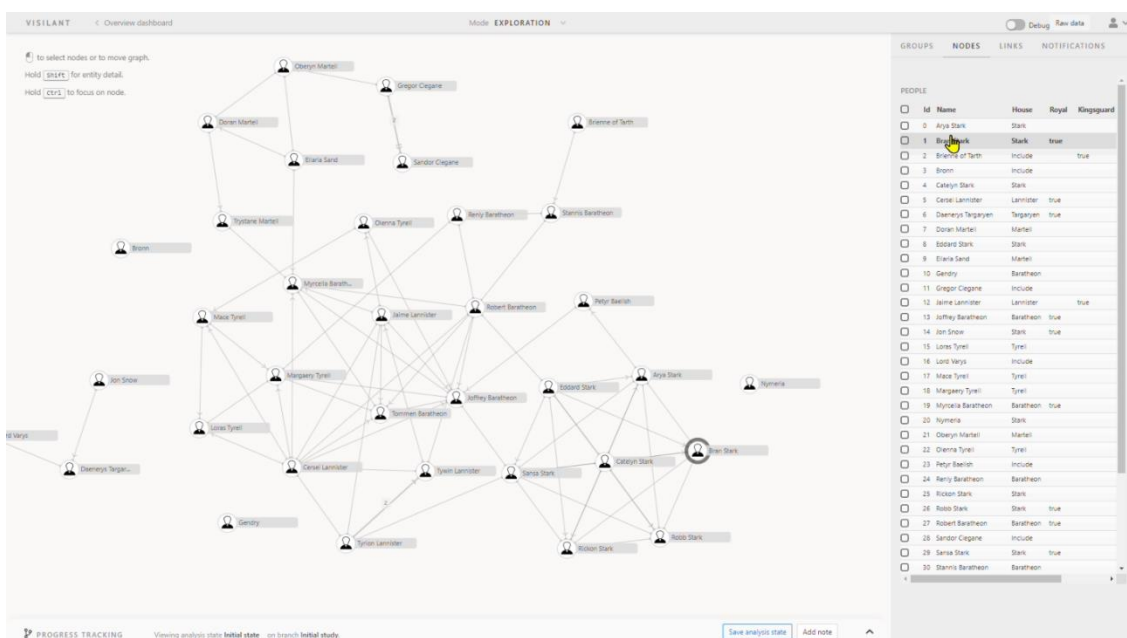
Každý vizualizační dokument a dokončenou analytickou operaci je možné vizuálně prozkoumat pomocí vztahové vizualizace, která je dostupná po kliknutí na daný objekt (např. vizualizační dokument v přehledovém dashboardu). Vztahová vizualizace operuje nad daty pocházejícími z vytvořeného filtru a pokud byla nad daty spuštěna nějaká analytická operace, zobrazuje rovněž její výsledky. Základem vztahové vizualizace je přehledný diagram (Obrázek 31), kde je každý objekt ze vstupní sady reprezentován uzlem a vztah mezi objekty je určen vazbou mezi nimi.

Mezi hlavní podporované funkce patří:

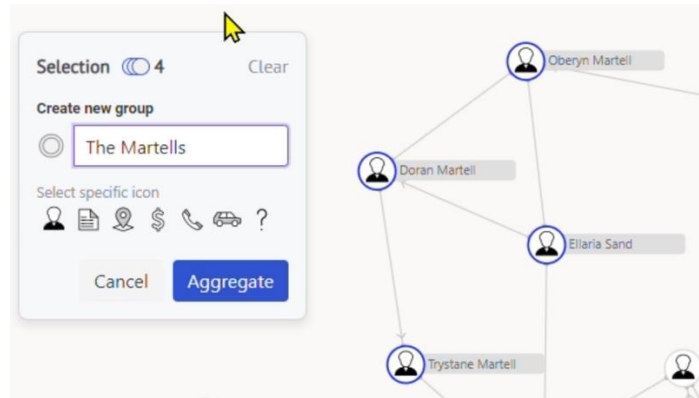
- Možnost vytvářet pojmenované skupiny uzlů (Obrázek 32) a po výběru sady uzlů uživatel může zvolit, zda chce dané uzly spojit do jednoho („Group“), vytvořit tzv. *Alias* uzel (pro případy, kdy je podezření, že dva uzly reprezentují tentýž objekt, např. osobu), či vytvořit tzv. *Bundle* (zabalení aktuálně nedůležitých uzlů pod sousedící uzel, na který mají tyto uzly logickou nebo hierarchickou vazbu). Kontextové menu pro tyto volby je dostupné pro vytvořenou selekci a ilustruje jej Obrázek 33.



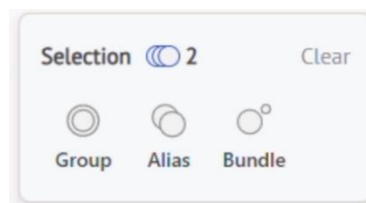
- Pro vybrané uzly přiřazovat doplňkové „Tagy“, čímž vizuálně zvýrazňujeme zajímavé vlastnosti odpovídajících objektů. Toto je realizováno pomocí barevné značky vedle uzlu, jež slouží pro rychlou vizuální navigaci skrz takto zajímavé uzly. Přiřazení těchto tagů je možné realizovat v pravém panelu obsahujícím seznam jednotlivých objektů (Obrázek 34).
- Rychlý přehled detailních informací o zvoleném uzlu při najetí myši nad daný uzel (Obrázek 35). Každý uvedený atribut daného uzlu navíc obsahuje „semafor“ informující uživatele o důvěryhodnosti dané informace.



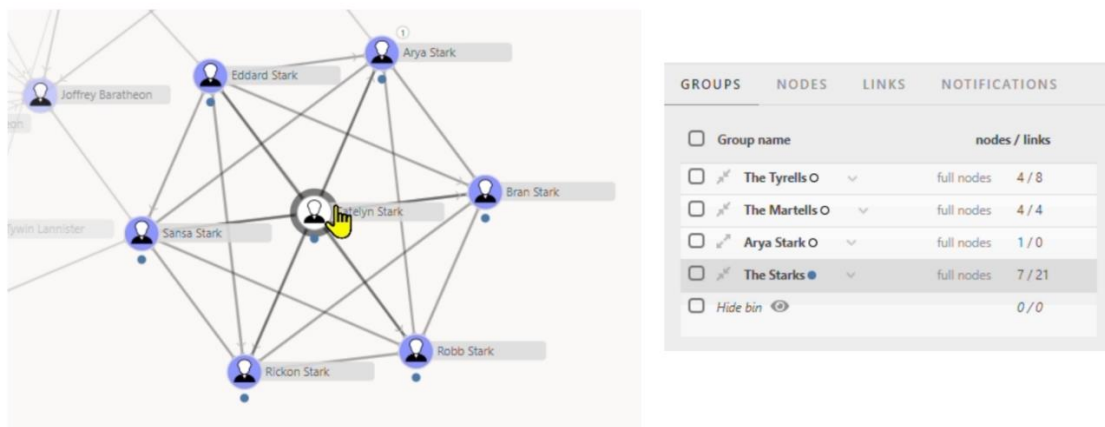
Obrázek 31: Vztahová vizualizace nad vybranými daty, doplněná o seznam objektů a jejich vybraných parametrů



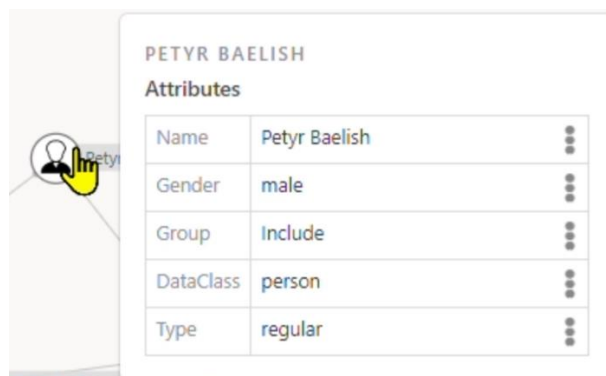
Obrázek 32: Vytváření pojmenované skupiny objektů, jež lze různými způsoby agregovat (viz Obrázek 33)



Obrázek 33: Kontextové menu pro volby typu agregace vybraných uzlů



Obrázek 34: Vytváření barevných tagů pro jednotlivé uzly. Vlevo zobrazení tagu vedle jednotlivých uzlů, vpravo přiřazení tagu v přehledové tabulce



Obrázek 35: Zobrazení detailních atributů vybraného uzlu, pro každý atribut lze vidět důvěryhodnost dané informace pomocí jednoduchého semaforu

## 5 Programátorská dokumentace

Vizualizační komponenta má k splnění svých případů užití vytvořen vlastní agregační algoritmus, komunikační rozhraní, které se podílí na transformaci dat pro Datový sklad. Pro splnění těchto případů užití byl vybrán framework VueJS ve verzi 2.6, společně s knihovnou Vuex pro správu modulů.

### 5.1 Moduly

Moduly jsou umístěny ve složce `src/modules`. Jsou použity pro správu globálního stavu aplikace napříč uživatelským prostředím, ukládají se do nich informace jako pozice jednotlivých uzlů, jejich propojení a skupiny, do kterých patří. Datový modul umožňuje práci s různými verzemi dat pomocí uložení jednotlivých dat do příslušných slotů. Každý slot je následně identifikován svým ID.

#### 5.1.1 `src/store/modules/data.store.js`

##### **INIT(pl)**

Inicializace počátečního stavu, ve kterém probíhá ověření vstupních údajů společně s transformací přijatých dat z Datového skladu.

##### **ADD\_NODE(pl)**

Přidá nový uzel do existujícího grafu, na základě obsahu objektu *pl*.

##### **UPDATE\_NODES(pl)**

Upraví jeden nebo více uzlů na základě dodaných atributů v objektu *pl*.

##### **Parametry**

Název	Typ	Popis
<b>pl</b>	Object	payload
<b>pl.slotId</b>	number	dataset slot
<b>pl.diff</b>	Array.<Object>	
<b>pl.diff.uuid</b>	Number	Každý objekt nebo rozdílové pole potřebují ID uzlu a atributy pro náhradu [...{ uuid: nodeId, ...updateAttr: newVal}]

## AGGREGATE\_NODES(state, pl)

Agreguje uzly do skupin, aliasů nebo bundles. Provádí operaci *Push* nebo *Update* nové skupiny uzlů/spojnic.

Agregace typu ALIAS vytváří klasický uzel reprezentující jednu entitu, nikoliv skupinu. Případný *merge* spojnic nebo dalších atributů musí být řešen uživatelem.

BUNDLING schová sadu uzlů pod vybraný uzel; uzly jsou schovány, spojnice jsou přesunuty (jsou připojeny na *bundle* node).

GENERAL GROUP uzly jsou schovány, spojnice jsou přesunuty.

### Parametry

Název	Typ	Popis
<b>pl</b>	Object	payload
<b>pl.slotId</b>	number	dataset slot
<b>pl.action</b>	group alias bundle	Typ agregace
<b>pl.nodeIds</b>	Array.<Number>	co agregovat
<b>pl.disableTags</b>	boolean	Vypnutí funkcionality tagování
<b>pl.targetGroupNode</b>	Node	Nová skupina uzlů s ID, která má být uložena ve stavu
<b>pl.removedLinkIds</b>	Array.<Number>	OUT: bude obsahovat vztahy skryté během agregace, řazeno od nejmenších
<b>pl.newVisibleLinks</b>	Array.<Object>	OUT: bude obsahovat nové vztahy, řazené podle ID od nejmenšího

## DISAGGREGATE\_NODE(state, pl)

Opačná funkce k funkci agregace pro skupinu, alias nebo bundle.

Nová data pro graf jsou umístěna do objektu payload.

## Parametry

Název	Typ	Popis
<b>pl</b>	Object	
<b>pl.slotId</b>	number	dataset slot
<b>pl.nodeId</b>	Number	skupinový uzel pro rozdělení skupin
<b>pl.revertAction</b>	group bundle	
<b>pl.removeNode</b>	boolean	OUT: bude obsahovat informaci, zda se má odstranit objekt mimo skupinu
<b>pl.removedLinkIds</b>	Array.<Number>	OUT: bude obsahovat odstraněné vztahy, řazeny od nejmenšího
<b>pl.newVisibleLinks</b>	Array.<Object>	OUT: bude obsahovat nahrazující vztahy, řazeny podle ID od nejmenšího
<b>pl.newVisibleNodes</b>	Array.<Object>	OUT: bude obsahovat nahrazující uzly, řazeny podle ID od nejmenšího

## 5.2 Komponenty

### 5.2.1 GraphNode

Uzel zobrazený v grafu.

#### Vlastnosti

Název	Typ	Popis
<b>node required</b>	Object	Data k uzlu získaná z datového skladu
<b>tags</b>	Array	Seznam tagů patřících uzlu
<b>visuals required</b>	Object	Objekt vizualizačních atributů, jako pozice v grafu.
<b>move-callback</b>	Function	Funkce, která se spustí po přesunutí uzlu
<b>is-focused</b>	Boolean	Příznak označení uzlu
<b>is-permanently-focused</b>	Boolean	Příznak stálého označení uzlu
<b>is-globally-highlighted</b>	Boolean	Příznak globálního označení uzlu
<b>is-detailed</b>	Boolean	Příznak označení detailního náhledu uzlu
<b>is-selected</b>	Boolean	Příznak výběru uzlu
<b>is-hovered</b>	Boolean	Příznak najetí myši
<b>is-selection-blocked</b>	Boolean	Příznak blokování označení selekce uzlů
<b>distance-from-focused-node</b>	Number	Příznak vzdálenosti od označeného uzlu
<b>focus-visuals</b>	Object	Objekt uchovávací označení v pozici grafu
<b>is-highlighted-in-diff</b>	Boolean	Příznak označení v zobrazení rozdílných uzlů

## Počítané vlastnosti

Jméno	Typ	Popis
<b>bundleSize</b>	unknown	Velikost skupiny uzlů
<b>hasMiniatureStyle</b>	boolean	Příznak miniatury
<b>hasUserDefinedStyle</b>	boolean	Příznak uživatelsky definovaných stylů
<b>visCircleRadius</b>	object	Objekt obsahující velikost označeného kruhu okolo uzlu
<b>visBundleCircleRadius</b>	unknown	Objekt obsahující velikost označeného kruhu okolo skupiny
<b>visBundleCircleDistance</b>	object	Vzdálenost kruhu od označené skupiny
<b>visibleTags</b>	unknown	Pole viditelných značek
<b>strokeColor</b>	boolean	Barva označení
<b>mainCircleFillColor</b>	string	Hlavní barva výplně uzlu
<b>icon</b>	unknown	Ikona uzlu

## Události

Název	Popis
<b>node-dragstart</b>	Spuštění chycení uzlu myší
<b>node-dragend</b>	Ukončení chycení uzlu myší
<b>node-click</b>	Kliknutí na uzel
<b>node-doubleclick</b>	Dvojklik na uzel
<b>node-mouseover</b>	Najetí myší na uzel
<b>node-mouseout</b>	Ukončení najetí myší na uzel
<b>context-menu</b>	Požadavek na zobrazení kontextového menu

## 5.3 GraphLink

Spojení mezi dvěma uzly.

### Vlastnosti

Název	Typ	Popis
<i>link required</i>	Object	Napojený link
<i>source required</i>	Object	Zdrojový uzel
<i>target required</i>	Object	Cílový uzel
<i>is-highlighted required</i>	Boolean	Příznak vyznačení
<i>is-detailed required</i>	Boolean	Příznak zobrazení
<i>is-selected required</i>	Boolean	Příznak označení
<i>is-hovered required</i>	Boolean	Příznak najetí myší
<i>is-focused required</i>	Boolean	Příznak výběru
<i>is-globally-highlighted</i>	Boolean	Příznak globálního výběru
<i>source-node-type required</i>	String	Typ zdrojového uzlu
<i>target-node-type required</i>	String	Typ cílového uzlu
<i>is-source-node-miniature</i>	Boolean	Je zdrojový uzel miniaturou
<i>is-target-node-miniature</i>	Boolean	Je cílový uzel miniaturou
<i>distance-from-focused-node</i>	Number	Vzdálenost od označeného uzlu
<i>is-highlighted-in-diff</i>	Boolean	Příznak označení v zobrazení rozdílných uzlů

### Vlastní atributy

Název	Typ	Popis
<i>hasUserDefinedStyle</i>	boolean	Příznak uživatelsky definovaného stylu
<i>showLabel</i>	object	Název linku
<i>isTransferred</i>	unknown	Příznak přenesení
<i>isMultilink</i>	boolean	Příznak existence v jiných spojnicích
<i>multilinkSize</i>	unknown	Velikost v jiných spojnicích, kam patří

### Události

Název	Popis
<i>link-mouseover</i>	Při přejetí myší nad uzlem
<i>link-highlighted</i>	Při označení linku
<i>link-clicked</i>	Při kliknutí na link



## 5.4 Služby

O komunikaci se zdroji dat se starají služby. V komponentě se nacházejí tři takové služby:

- *orchestrator.service.js* – správa a spuštění operací nad Orchestrátorem,
- *storage.service.js* – správa dat z Datového skladu,
- *visilant.service.js* – interní uložení komponenty.

### 5.4.1 orchestrator.service.js

#### **getOperations()**

Získá právě běžící operace v Orchestrátoru, které jsou zobrazeny v Dashboard panelu. Návratovou hodnotou je seznam operací. Volání HTTP: GET /operation

#### **runOperation(parameters)**

Spustí operaci na Datovém skladě s požadovanými parametry, na vyplnění těchto parametrů jsou použity templates:

```
{
  "name": "Data-Analysis",
  "parameters": {
    "extraction": {
      "INPUT_IDENTIFIER": "{{node.uuid}}"
    },
    "transformation": {
      "INPUT_IDENTIFIER": "{{node.uuid}}",
      "RDF_FILE_NAME": "zeek-dgraph",
      "SCHEMA_FILE_NAME": "zeek-dgraph",
      "CASE_ID": "test case national gallery"
    },
    "indexing": {
      "INPUT_IDENTIFIER": "{{node.uuid}}",
      "ZIP_DIR_PREFIX": "nonlocal-",
      "MAP_SHARDS": "2",
      "REDUCE_SHARDS": "1",
      "CUSTOM_TOKENIZERS": "cldr-plugin.so",
      "BULK_ARGS": ""
    },
    "handling-alpha": {
      "INPUT_IDENTIFIER": "{{node.uuid}}",
      "LRU_MN": "2048",
      "ALPHA_0": "",
      "ALPHA_ARGS": "",
      "CUSTOM_TOKENIZERS": "cldr-plugin.so"
    }
  }
}
```

Funkce doplní na místo parametru příslušnou hodnotu, která vytvoří operaci. Návrátovou hodnotou je ID nově vytvořené operace. Volání HTTP: POST /operation

### **removeOperation(id)**

Odstranění operace, která je aktuálně aktivní, prostřednictvím její ID.

Volání HTTP: DELETE /operation/{id}

### **getOperation(id)**

Získá informace o aktuální operaci prostřednictvím její ID.

Volání HTTP: GET /operation/{id}

## **5.4.2 visilant.service.js**

### **getWorkspace(id)**

Získá uživatelský pracovní prostor, který uchovává informaci o aktuálně definovaných filtrech a vizualizačních dokumentech. Volání HTTP: GET /workspace/{id}

### **updateWorkspace(id, data)**

Aktualizuje informace o pracovním prostoru, například po vytvoření nového filtru nebo vizualizačního dokumentu. Volání HTTP: PUT /workspace/{id}

## **5.4.3 storage.service.js**

### **getStorage(query)**

Získá data o uzlech a spojnicích z Datového skladu, netransformuje je. Volání HTTP: GET /visualisation?query={query}

### **saveStorage(nodes, links)**

Uloží nové uzly a spojnice. Volání HTTP: POST /visualisation

### **runFilter(query)**

Spustí požadovaný filtr nad daty uloženými ve skladě, pro ověření získaných uzlů a spojnic. Volání HTTP: PUT /generic-objects/filter

## 5.5 Formát Dat

V následujících příkladech je zobrazena ukázka transformací dat, které Vizualizační komponenta přijímá z Datového skladu a Orchestrátoru, a následně je upravuje pro zobrazení ve vizualizaci.

### Úprava pro Datový sklad

Data přijatá pod vlastností s názvem obsahujícím # jsou interpretována jako pole hodnot.

```
"properties": [  
  {  
    "propertyKey": "dataClass",  
    "valueString": "regular"  
  },  
  {  
    "propertyKey": "color",  
    "valueString": "purple"  
  },  
  {  
    "propertyKey": "#comunities",  
    "valueString": "purple,red"  
  }  
]
```

## 6 Poděkování

Vytvořeno v rámci projektu „*Komplexní analýza a vizualizace heterogenních dat velkého rozsahu (ANALYZA)*“ (kód projektu VI20172020096, období řešení 1.1.2017 – 31.12.2020) v rámci Programu bezpečnostního výzkumu České republiky v letech 2015-2020 Ministerstva vnitra ČR.