

# Vizualizácia vedeckej komunikácie pomocou open source softvéru na analýzu grafov a sietí

**PhDr. Matúš Formanek, PhD.**

matus.formanek@fhv.uniza.sk

**Marcela Kajanová**

marcela.kajanova@st.fhv.uniza.sk

Katedra mediamatiky a kultúrneho dedičstva

Fakulta humanitných vied

Žilinská univerzita v Žiline

## Abstrakt

Príspevok sa venuje konkrétnym spôsobom a metódam, pomocou ktorých je možné vizuálnou formou zachytiť a analyzovať vzťahy medzi vzájomne pôsobiacimi entitami. V kontexte vedeckej komunikácie v novoveku sa týmito entitami stávajú vybraní predstavitelia európskej inteligencie. Vzťahy v pomyselných grafoch predstavuje korešpondencia v podobe orientovaných hrán. Použitý voľne dostupný softvér Gephi, ako open source nástroj, umožňuje nielen vizualizáciu tejto komplexnej siete, ale i jej následnú analýzu z viacerých možných uhlov pohľadu.

Kľúčové slová : vizualizácia dát, open source, teória grafov, vedecká korešpondencia

## Úvod

Predložený príspevok sa viaže k projektu APVV s názvom *Intelektuálne dedičstvo a vedecká komunikácia 1500-1800 so slovenskými vzťahmi ako súčasť európskej histórie a identity*, ktorý taktiež vystupuje pod akronymom INDED. Spomínaný projekt nadobúda národný, ako aj európsky charakter, pričom očakávame, že výsledky projektu budú následne zapájané do väčších medzinárodných projektov či akcií, ako napríklad *COST IS1310 - Reassembling the Republic of Letters, 1500-1800 A digital framework for multi-lateral collaboration on Europe`s intellectual history*. V projekte INDED ide o aplikovaný výskum vo vedeckej oblasti

knižnično-informačných štúdií pod záštitou Žilinskej univerzity v Žiline. Výskumná časť projektu je tematicky orientovaná na komunikáciu a intelektuálne dedičstvo. Výskumnú problematiku predstavuje dokumentácia konkrétnych prejavov vedeckej korešpondencie, ako aj výmena poznatkov a informácií medzi vzdelancami z obdobia novoveku po materiálovej stránke. Špecifickým výskumným problémom projektu INDED sa stáva digitalizovanie výstupov, ktoré budú inovatívne sprístupňovať poznatky o vedeckej komunikácii a intelektuálnom dedičstve z historického obdobia novoveku ako vedeckej komunite, tak aj širokej verejnosti. Medzi konkrétne ciele patrí napríklad:

- pasportizácia pertinentných informačných zdrojov a dokumentov,
- analýza stavu unikátnych rukopisných dokumentov,
- návrh postupu na ich ochranu,
- reštaurovanie ako aj odhad životnosti týchto dokumentov,
- experimenty, ktorých náplňou je zvýšenie kvality digitalizácie a najmä schopnosť rozlišovať rukopisné znaky,
- následne sprístupnenie dokumentov a poznatkov z projektu v publikáciách do európskeho projektu digitalizácie a vizualizácie dokumentov a dát s názvom COST IS1310.
- V neposlednom rade popularizácia poznatkov a výsledkov projektu prostredníctvom vedeckých podujatí webového sídla (Fakulta humanitných vied 2017).

V prvej kapitole príspevku definujeme komunikáciu ako informačný proces a tiež rozoberáme jej širší význam. Obsahom podkapitoly viažucej sa ku komunikácii sú charakteristiky jednotlivých prvkov komunikácie nevyhnutných pre jej uskutočnenie a efektívnosť. Druhú kapitolu venujeme základným teoretickým východiskám súvisiacim s vizualizáciou, na ktoré nadväzujeme v podkapitolách, kde sa prechádzame do praktickej roviny. V podkapitole 2.1. stanovujeme ciele, pričom stručne opisujeme metodickú rovinu, podľa ktorej bola realizovaná praktická časť. V nasledujúcej podkapitole 2.2 pojednávame o teórii siete, ako o prepojení viacerých osôb, pričom ide o vzdelancov z obdobia novoveku. Zároveň sa zaoberáme pojmom *graf*, v zmysle grafického znázornenia vzťahov v sieti. Časť 2.3 venujeme open source softvéru s názvom Gephi, cez ktorý bola uskutočnená vizualizácia dát. Veľmi stručne opisujeme vznik a vývoj tohto nástroja. Ďalej tu popisujeme spôsob úpravy dát v súbore CSV a aplikáciu konkrétneho algoritmu na usporiadanie grafu.

# 1. Komunikácia

Komunikácia je nevyhnutnou súčasťou civilizácie ako aj samotnej ľudskej histórie. Pochádza z latinského slova *communicare*, čo značí *robiť spoločným, medzi sebou*. Bežní ľudia, rovnako ako aj historické osobnosti zdieľali svoje myšlienky a znalosti. Pokiaľ medzi sebou komunikujú vzdelanci svoje poznatky, jedná sa o vedeckú komunikáciu, ktorá sa stáva podmnožinou samotnej komunikácie a prejavuje sa rovnakým spôsobom. Písomná forma komunikácie umožňuje prechovávať tieto zdieľané myšlienky a znalosti, ako aj intelektuálne dedičstvo, ktoré tak vytvárajú obrazy minulosti.

Komunikácia nesie charakter informačného procesu, pričom informačný proces sa prejavuje ako tok dát, ktoré sú výsledkom ľudského poznania. Tieto dáta taktiež obsahujú obraz, alebo opis minulej, súčasnej či budúcej reality v závislosti od obsahu samotnej komunikovanej správy (Harausová 2013).

Komunikácia je tiež „*prenos vzájomného porozumenia pomocou symbolov*“. Takto ju definoval Donnely (2004), pričom pod symbolmi môžeme rozumieť jazyk, v ktorom je vedená komunikácia, avšak taktiež môže predstavovať binárny kód, v ktorom medzi sebou komunikujú stroje.

## 1.1.Prvky komunikácie

Podľa Donnelyho (2004) komunikáciu tvoria prvky, bez ktorých by komunikácia strácala význam. Sú to komunikátor, komunikant, komunikát, komunikačný kanál, kódovanie, dekódovanie a spätná väzba. Súčasťou komunikácie je však tiež šum ako škodlivý a nežiadúci prvok.

**Komunikátor** je v roli pôvodcu alebo zdroja správy takzvaný odosielateľ. Odosielateľ pri verbálnej komunikácii vysiela svoju myšlienku či informáciu pričom ju transformuje do slov, kedy volí dĺžku, štruktúru, jazyk aj formu podania informácie. Popri formulovaní správy by mal brať komunikátor ohľad na dekódovacie schopnosti komunikanta.

**Komunikant** ako prijímateľ správy, ktorý je v roli, kedy informácie prijíma, musí správu dekódovať a pochopiť. Správne pochopenie informácie môže ovplyvniť úroveň vzdelania a vedomostí komunikanta, komunikačné zručnosti, postoj prijímateľa správy ku odosielateľovi a aj prostredie, v ktorom sa nachádza komunikant.

**Komunikát** je zakódovaný obsah, informácie určené pre komunikanta. Zároveň je to súbor znakov alebo symbolov. Pri verbálnej komunikácii má formu písanú alebo vyslovenú. Pokiaľ ide o neverbálnu komunikáciu, komunikát dokáže byť obsiahnutý v geste či mimike komunikátora.

**Kódovanie** definujeme ako proces kedy komunikátor prevádza informácie určené pre komunikanta do rady symbolov, ktoré vyjadrujú úmysel komunikátora. Posledná fáza kódovania informácie je nazývaná komunikát, teda správa v podobe v akej bude odoslaná.

**Komunikačný kanál** inak nazývaný aj médium je spôsob prenosu komunikátu ku prijímateľovi správy. Práve komunikačný kanál ovplyvňuje formu komunikácie, a to ústnou alebo písomnou formou. Príklad média môže byť telefónne spojenie, počítačová sieť alebo list.

**Dekódovanie** alebo interpretácia komunikátu. Ide o proces kedy sa prijímateľ snaží vysvetliť si obsah správy od komunikátora, kedy výsledkom je buď porozumenie alebo neporozumenie komunikátu. Práve dekodovanie komunikátu sa najčastejšie stáva kritickým bodom komunikácie medzi komunikátorom a komunikantom. Aby bola komunikácia čo najefektívnejšia, je žiadúce aby komunikát bol čo najviac zhodný s dekodovanými informáciami.

**Spätná väzba** predstavuje moment, kedy si komunikátor a komunikant navzájom vymenia role. Vďaka spätnej väzbe sa pôvodný komunikátor uisťuje, že obsah jeho správy bol dekodovaný správne, a zároveň je to reakcia pôvodného komunikanta na komunikát, čím sa zvyšuje efektivita komunikácie. Pri niektorých typoch komunikácie spätná väzba nie je uskutočniteľná.

**Komunikačný šum** svojim pôsobením skresľuje komunikát a narúša mieru efektivity komunikácie. Jeho pôsobenie sa môže prejaviť pri dekodovaní správy. Za šum môžeme považovať akýkoľvek rušivý element, ktorý narúša proces komunikácie a môže pôsobiť na ktorýkoľvek z prvkov komunikácie. V ideálnej komunikácii, kedy nedochádza k žiadnym rušivým vplyvom a komunikant plne porozumie správe od komunikátora, komunikačný šum neexistuje avšak o ideálnej komunikácii môžeme uvažovať iba v teoretickej rovine (Zoralová 2012).

## 2. Vizualizácia korešpondencie a vzájomných vzťahov

Známa Aristotelova myšlienka vypovedala o dôležitosti ľudskej predstavivosti a obrazov v mysliach ľudí. Človek vnímajúci obrazy dokáže lepšie porozumieť danej problematike, pretože vizualizované údaje sú mu svojou povahou omnoho bližšie a rýchlejšie ich pochopí.

Porozumenie rozsiahlym sieťam prostredníctvom vizualizácie množstva údajov sa uplatňovalo už v minulosti pri riešení množstva úspešných projektov<sup>1</sup>. Vizualizácia napomáha ľuďom nájsť skryté vlastnosti v danej štruktúre, hoci ide stále o proces relatívne náročný a vyžadujúci si voľbu správnej exploratívnej stratégie (Bastian et al. 2009).

### 2.1. Ciele a stručný popis metodiky

Cieľ predloženého príspevku spočíva v návrhu metód možnej vizualizácie vzájomných vzťahov medzi predstaviteľmi európskej inteligencie, ktorí spolu komunikovali medzi rokmi 1500–1800. Vzdelanci žijúci v tomto období si vymieňali množstvo korešpondencie, o čom svedčia mnohé záznamy v príslušnej odbornej literatúre. Pri vizualizácii teda vychádzame priamo z už zozbieraných údajov z praktického výskumu. Aby bolo možné realizovať ďalšie skúmanie v rámci stanovených cieľov projektu, považujeme za nutnosť tieto vzťahy vhodnými metódami zachytiť vizuálnou formou. Tá umožní bližší a detailnejší pohľad na vedecké komunity tej doby s prihliadnutím i na iné kvalitatívne ukazovatele, ktoré budeme schopní sledovať a zachytiť.

Metodologická stránka príspevku vychádza z teoretických východísk procesov vzájomnej komunikácie pričom sa primárne ďalej opiera o princípy prevzaté z matematickej teórie grafov. Vzťahy medzi osobnosťami zaznamenávame a následne reprezentujeme pomocou vhodného aparátu, čím dostávame explicitne vyjadrené údaje v textovej forme spracovateľnej pomocou počítača – v elektronických tabuľkách. Z tabuliek sa konverziou vytvárajú jednoduché textové súbory CSV<sup>2</sup> obsahujúce vlastné údaje v požadovanom formáte. Tieto počítačové súbory následne slúžia ako vstupy pre softvér, ktorý s nimi umožňuje ďalej pracovať. K dispozícii tak máme, vďaka softvéru, široké možnosti, ako výslednú grafickú vizualizáciu modifikovať, korigovať či usporiadať do vyhovujúcej podoby. Softvér je v tomto prípade len prostredníkom

---

<sup>1</sup> Adar, E. 2006. Guess: a language and interface for graph exploration. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, 791–800.

Batagelj, M. 1998. Pajek - program for large network analysis. In *Connections* 21:47–57.

Shannon, M. et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. In *Genome Res* 13(11):2498–2504.

<sup>2</sup> Skratka z ang. comma selected values - ide o textový súbor s hlavičkou údajov, pričom sú jednotlivé stĺpce údajov navzájom oddelené stanoveným oddeľovacím znakom, napríklad čiarkou, bodkočiarkou a pod.

ponúkajúci možnosti následnej vedeckej práce pri hlbšom a detailnejšom skúmaní. Náš príspevok pritom predstavuje len úvod do rozoberanej problematiky písomnej vedeckej komunikácie v historickom kontexte.

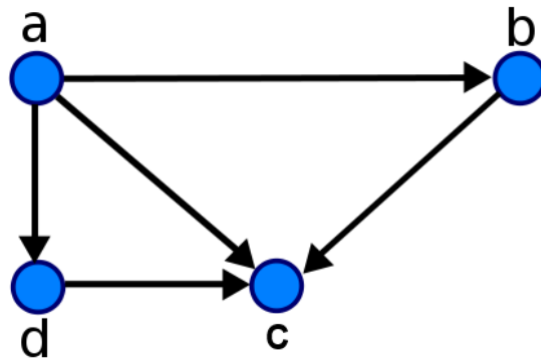
## 2.2.Sieť a graf

K definícii pojmu sieť môžeme pristúpiť z viacerých hľadísk. My sa však zameriame predovšetkým na sociologické hľadisko. Online slovníky definujú pojem *sieť* ako určitý systém prepojených vecí alebo ľudí (Oxford University Press 2018). V informatike ide zjednodušene o množinu technických zariadení prepojených navzájom pomocou rozličných komunikačných kanálov. V sociologickom rozmere, ktorý je najviac relevantný k téme tohto príspevku, budeme sieť vnímať ako vzájomne prepojenú skupinu ľudí, ktorí si vymieňajú korešpondenciu obsahujúcu rozličné informácie s dôrazom na profesionálny, vedecký alebo iný druh komunikácie. Napĺňa sa tak primárny cieľ siete – umožnenie výmeny informácií cez komunikačné kanály, resp. linky (Businessdictionary n.d.).

K explicitnému zachyteniu štruktúry siete môžeme využiť útvar nazývaný graf. Pravdou však ostáva, že v širšom význame slova sa grafom môže rozumieť akákoľvek grafická reprezentácia ľubovoľných údajov. Istým spôsobom teda môže ísť aj o určitý diagram. Pre naše účely však využijeme grafy v ponímaní, prostredníctvom ktorého k nim pristupuje matematická teória grafov. Ponúknuté prostriedky využijeme pri reprezentácii siete tvorenej navzájom komunikujúcimi osobnosťami.

Podľa teórie grafov je graf definovaný ako abstraktný objekt pozostávajúci z vrcholov a hrán: „Nech  $V$  je konečná množina. Nech  $E$  je systém dvojprvkových podmnožín množiny  $V$ . Pojmom graf označujeme usporiadanú dvojicu  $(V, E)$ . Prvky množiny  $V$  nazývame vrcholy grafu a prvky množiny  $E$  nazývame hrany grafu“ (Pokorný 2013). Graf teda tvoria dve usporiadané dvojice množín – množina vrcholov grafu a množina hrán. Vrcholy alebo uzly reprezentujú v našom konkrétnom prípade osobnosti, kým hrany vyjadrujú smerovanie korešpondencie. V kontexte tejto práce budeme považovať označenia *vrchol* a *uzol* za ekvivalentné.

Vďaka usporiadanosti dvojíc tvoriacich hranu, hovoríme o tzv. orientovanom grafe (vzťah vyjadrený šípkou). V každej dvojici je na prvý pohľad zrejмый odosielateľ i príjemca, resp. smer zaslanej korešpondencie – vid' obrázok 1.



**Obrázok 1 Príklad grafu vyjadrujúceho korešpondenciu medzi osobnosťami**

Vychádzajúc z obrázka 1, ak vrcholy a, b, c, d predstavujú v grafe jednotlivé osobnosti, ich korešpondenciu môžeme vyjadriť v zmysle definície týmito usporiadanými dvojicami: (a,b), (a,c), (a,d), (b,c), (d,c). Ide teda o orientované hrany daného grafu. Vidíme, že najviac aktívnym prispievateľom je v našom prípade z obrázka 1 osoba „a“, pretože stupeň vrcholu smerom von  $\text{deg}(a)=3$ . Taktiež stupeň vrcholu „c“ smerom dnu  $\text{deg}(c)=3$ . Osoba „c“ je tak najčastejším príjemcom. Stupeň definuje Pokorný (2013) ako „počet hrán grafu, ktoré sú s daným vrcholom incidentné... Okolím vrcholu v grafe nazývame množinu všetkých vrcholov grafu, ktorú sú susedné s týmto vrcholom.“ Na základe stupňa vieme určiť najaktívnejších odosielateľov a najvýznamnejších adresátov.

### 2.3. Gephi

Gephi patrí medzi pokročilé, zdarma dostupné open source nástroje určené na vedeckú analýzu výskumných dát, konkrétne grafov a sietí. Umožňuje hlbší prienik do problematiky a preskúmanie vzájomných vzťahov medzi entitami vyjadrenými prostredníctvom grafu. Pri zobrazovaní rozsiahlych sietí Gephi využíva vizuálne 3D renderovanie prebiehajúce v reálnom čase s využitím vlastností moderných grafických kariet a viacjadrových procesorov. Flexibilná architektúra využívajúca paralelné spracovanie dát umožňuje softvéru prehľadne spracovávať i nesmierne komplexné siete s niekoľkými desiatkami tisíc uzlov (Bastian et al. 2009).

Získané vizualizácie často pomôžu odhaliť na prvý pohľad skryté súvislosti. Pomocou zmeny základných nastavení v programe dokážeme dynamicky meniť výsledné rozloženie grafu, nastaviť filtrovanie alebo zhlukovanie vrcholov na základe stanovených kritérií. Tvar, veľkosť, farba i menovky vrcholov grafu môžu odrážať rozličné sledované vlastnosti, napríklad incidenciu uzla, či inú používateľom definovanú vlastnosť. Podobné charakteristiky možno skúmať aj pri hranách grafu. Obyčajnou zmenou farby, usporiadaním či modifikovaním grafu

je možné intuitívnym spôsobom objavovať v grafe skryté vzory, vlastnosti, zvláštnosti, prípadne i chyby, ktoré sme schopní neskôr využiť, napríklad, pri stanovovaní hypotéz, ďalšom výskume atď. Gephi však zároveň disponuje i matematickým pohľadom na graf a explicitné logické vyjadrenie vzťahov medzi zapojenými uzlami, vrátane kalkulácie rôznych vlastností a dynamických štatistík v reálnom čase (Gephi 2017; Bastian et al. 2009).

### 2.3.1. Príprava, formátovanie a import dát

V rámci výskumu na projekte INDED, sme začali vizualizovať doposiaľ spracovanú korešpondenciu Mateja Bela – dovedna 378 listov. Dáta budeme postupne dopĺňať aj o ďalšie osobnosti. Pôvodné zozbierané výskumné údaje o korešpondencii Mateja Bela sme získali v tabuľke formátu MS Excel. Dáta vzišli z viacerých terénnych výskumov a pracovných ciest, ktoré realizoval iný odborný tím pracovníkov Katedry mediamatiky a kultúrneho dedičstva. Čo sa týka povahy vlastných údajov, principiálne šlo v základe o dva samostatné oddelené stĺpce s menami reprezentujúcimi odosielateľov a príjemcov. V prípade, že odosielateľ zaslal adresátovi viacero listov, vyjadřila sa túto skutočnosť odpovedajúcim viacnásobným opakovaním toho istého riadku v predmetnej tabuľke. V terminológii teórie grafov sa jedná o vyjadrenie hrán toho istého grafu.

Keďže Gephi nepodporuje priamy import z prostredia MS Excel, bolo nutné získané dáta pred vložením náležite upraviť a naformátovať. Bezprostredne z Excelu sme tabuľku exportovali ako CSV súbor, pričom selekčným znakom medzi stĺpcami sa stala bodkočiarka. V open source softvéri Notepad++<sup>3</sup> sme čistý text uložený v CSV súbore prekonvertovali do kódovacej znakovej tabuľky UTF-8, ktorú Gephi natívne očakáva pri importe dát z textového súboru. Doplnili sme taktiež hlavičky údajov – *Source* a *Target* – vid' obrázok 2. Samozrejme, v súbore je možné stanoviť aj ďalšie doplnkové, vlastné hlavičky údajov.

---

<sup>3</sup> Viac informácií o softvéri je možné nájsť na <https://notepad-plus-plus.org/>

1	Source;Target
2	Bel, Matej;Francke, August Hermann
3	Bel, Matej;Francke, August Hermann
4	Bel, Matej;Francke, August Hermann
5	Bel, Matej;Francke, August Hermann
6	Bel, Matej;Francke, August Gotthilf
7	Bel, Matej;Bohuš, Juraj
8	Bel, Matej;Bohuš, Juraj
9	Bel, Matej;Bohuš, Juraj
10	Bel, Matej;Bohuš, Juraj
11	Bel, Matej;Bohuš, Juraj

Zdroj

Cieľ

**Obrázok 2** Štruktúra CSV súboru spoločne s vloženými hlavičkami údajov

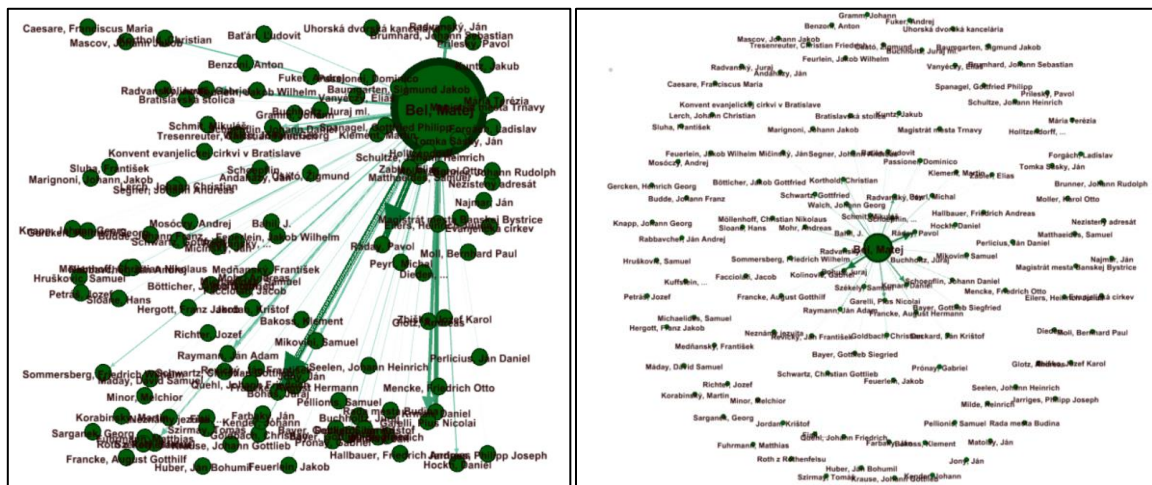
Správne naformátovaný CSV súbor obsahujúci výskumné dáta sme importovali do dátového laboratória softvéru Gephi ako orientované hrany (voľba : *Directed*) budúceho grafu. Ak sa nejaký riadok v CSV súbore opakoval viackrát (viac korešpondencie medzi tými istými osobami v rovnakom smere), softvér Gephi kumuloval ich počet a vyjadril túto hodnotu pomocou tzv. váhy hrany (*Weight*) – v osobitnom stĺpci charakterizujúcom vlastnosť hrany. Ak teda osoba *a* poslala osobe *b* 37 listov, hrana v tvare (*a,b*) bude vytvorená softvérom len raz, no s odpovedajúcou váhou 37. Váhu hrany môžeme neskôr využiť tak, že hodnotu premietneme do výsledného grafu zvoleným vizuálnym spôsobom.

Softvér sa ďalej postaral o automatické vytvorenie príslušných vrcholov (mená osobností), pričom im pridelil identifikátor (*ID*) na základe menného pomenovania, ktorým je v našom prípade meno osobnosti. Unikátne pomenovanie (tzv. *Label* alebo menovka) je nutné manuálne dopísať v osobitnom stĺpci, pretože vo výslednom grafe/vizualizácii sa zobrazujú vždy len menovky uzlov, nie ich interné identifikátory. Gephi umožňuje kliknutím hromadne prekopírovať hodnotu zo stĺpca *ID* do stĺpca *Label*. Tým pádom sa stali obe hodnoty identickými. Pri ďalšom importe nových výskumných dát o korešpondencii konkrétnej osoby softvér daný vrchol nebude duplicitne vytvárať, keďže bol už raz vytvorený niekedy v minulosti.

### 2.3.2. Rozloženie grafu – aplikácia algoritmu Force Atlas 2

Rozloženie grafu predstavuje vzájomné usporiadanie jeho vrcholov. Predvolené rozloženie *Random layout* nepovažujeme za vhodné. Na obrázku 3 vidíme, že vizualizácia umiestnená vľavo je neprehľadná, hoci sa tu váha hrany premietla do hrúbky orientovanej spojnice medzi vrcholmi grafu (na obrázku zelené šípky). Po mnohých pokusoch sme zistili, že z ponúkaných

možností rozloženia grafu našim potrebám najviac vyhovovalo rozloženie zabezpečené algoritmom nazvaným *Force Atlas 2* (na obrázku 3 vpravo).



**Obrázok 3 Pôvodné *Random* rozloženie vs. rozloženie *Force Atlas 2***

Algoritmus *Force Atlas 2* je jedným z mnohých algoritmov, ktorými umožňuje softvér Gephi dynamicky modifikovať aktuálne rozloženie grafu, teda vzájomnú polohu vrcholov. Algoritmus pritom využíva techniku prevzatú zo simulácie správania sa prírodných fyzikálnych systémov, v ktorých vystupujú sily ako gravitácia, príťažlivosť, či odpudivosť. Vďaka pôsobiacim „silám“ dokážeme našu sieť, resp. graf rozmiestniť v priestore v tzv. rovnovážnom stave (Jacomy et al. 2014).

Algoritmus *Force Atlas 2* vytvoril Mathieu Jacomy, zamestnanec vedeckých laboratórií v Paríži a zakladajúci člen konzorcia Gephi, ako nasledovníka pôvodnej, prvej verzie algoritmu. Algoritmus je primárne určený na analýzu kvalitatívnych aspektov v menších až stredne veľkých grafoch obsahujúcich do 10 000 vrcholov. Druhá verzia prvotného algoritmu prináša viaceré nové možnosti, ako aj inovatívnu optimalizáciu s názvom *Barnes Hut*<sup>4</sup>, vďaka ktorej sa stáva *Force Atlas 2* ešte rýchlejšim (Ralinovski 2014). Vďaka spomínanej optimalizácii nedochádza k tak dramatickému poklesu rýchlosti pri prudkom zvyšovaní počtu zapojených uzlov, ako pri prvej verzii. Samotný tvar výsledného grafu môžeme zaradiť niekam medzi Früchtermanov–Rheingoldovo<sup>5</sup> a Noack LinLogovo<sup>6</sup> rozloženie. Aj vďaka tejto skutočnosti algoritmus vyniká rýchlosťou adaptívnej konvergencie grafu (Jacomy 2011).

<sup>4</sup> Viac informácií sa čitateľ dozvie v publikácii Ralinovski (2014)

<sup>5</sup> Viac informácií sa čitateľ dozvie na web stránke : <https://github.com/gephi/gephi/wiki/Fruchterman-Reingold>

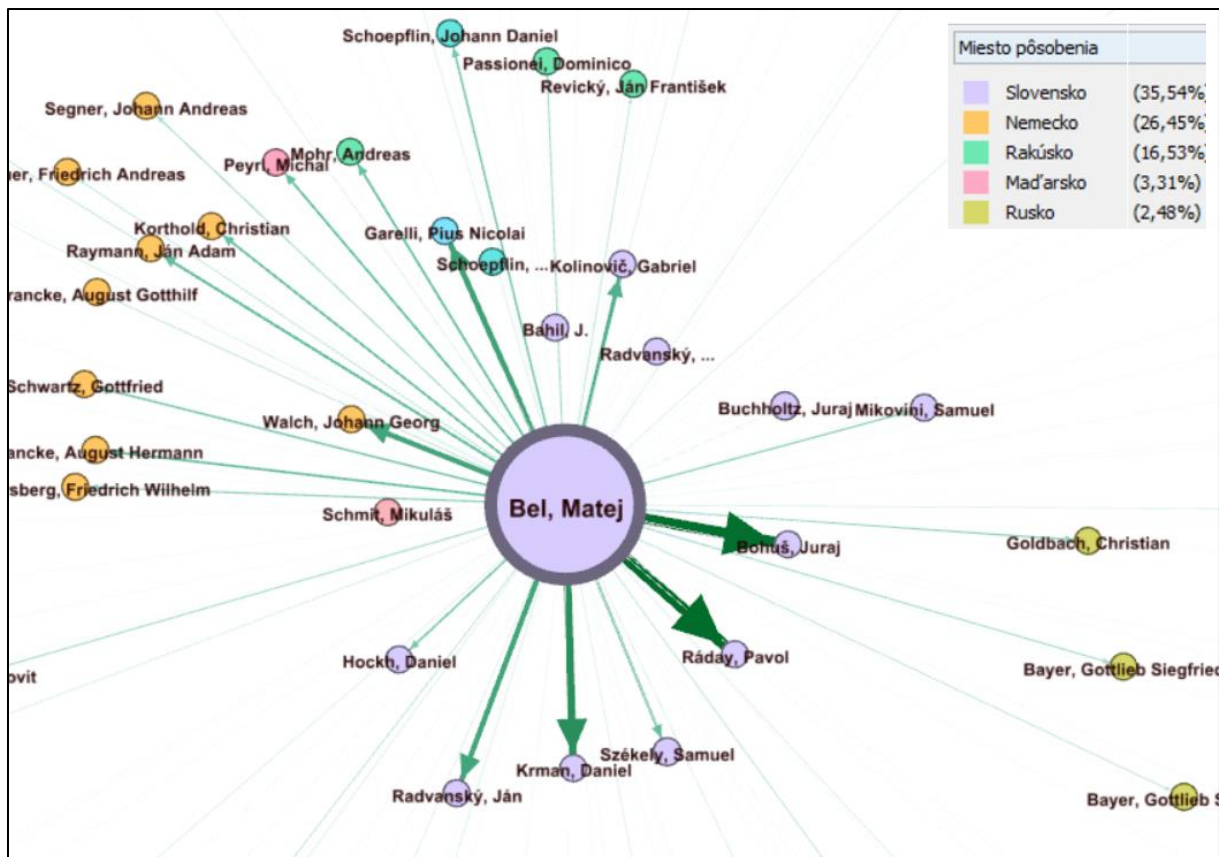
<sup>6</sup> Viac informácií sa čitateľ dozvie v publikácii Gibson et al. (2012)

Počas procesu renderovania umožňuje Gephi manipulovať s nastaveniami algoritmu, čo sa vo výsledku premietne aj do výslednej podoby zobrazeného grafu. Ako vyplýva z podstaty algoritmu Force Atlas 2, dôraz sa kladie na vzdialenosť medzi vrcholmi – na príťažlivosť a odpudivosť. „Ťažšie“ uzly sú príťahované viac do stredu. Odpudzovanie „ľahších“ smerom von je možné škálovať. Pozíciu vrcholu tým pádom nemôžeme posudzovať absolútne, ale vždy len relatívne, t. j. vzhľadom na ostatné vrcholy. Vzájomne pôsobiace sily vytvárajú počas behu algoritmu dynamický pohyb uzlov, ktoré následne smerujú k rovnovážnemu stavu. Podľa slov autora, existuje predpoklad, že takto utvorená konfigurácia uzlov pomôže pri interpretácii údajov (Jacomy et al. 2014). Vrcholom grafu nemusíme povoliť prekrývanie, čo opäť spôsobí výrazne sprehládnenie výsledkov – najmä ak používame menovky vrcholov. My sme túto možnosť využili (Jacomy 2011).

### **2.3.3. Farbenie vrcholov grafu na základe ich vlastností**

K jednotlivým hranám i vrcholom grafu umožňuje Gephi asociovať ľubovoľné atribúty, na základe ktorých vieme do vizualizácie zakomponovať ďalšie sekundárne polia údajov a na ich základe aplikovať rôzne filtre na množiny vrcholov a hrán.

Vychádzajúc z pôvodných dát v tabuľkách MS Excel sme z biografických údajov osobností vydedukovali približnú geografickú oblasť ich aktívneho pôsobenia v kultúrno-spoločenskom kontexte tej doby. Vtedajšie historické pomenovania území (Sasko, Sedmohradsko,...a pod.) sme sa snažili čo možno najvernejšie mapovať k súčasným oblastným názvom, resp. názvom odpovedajúcich štátov. Každý osobnosti bol týmto spôsobom priradený práve jeden nový atribút v stĺpci *štát*. Ak osobnosť pôsobila na viacerých miestach, vybrali sme pre ňu najcharakteristickejšie územie. V Gephi sme následne vybrali atribút miesta pôsobenia ako sledovaný parameter a nechali na vrcholy grafu automaticky aplikovať farebnú škálu – vid' výrez z vizualizácie na obrázku 4.



**Obrázok 4 Detailný pohľad graf farbený na základe vybranej vlastnosti**

Z údajov, ktoré sme mali v čase prípravy príspevku k dispozícii, sme pripravili vyššie uvedenú vizualizáciu, resp. graf. Z obrázka 4 je na prvý pohľad zrejmé, že Matej Bel posielal najviac korešpondencie Jurajovi Bohušovi a Pavlovi Rádayovi, čo boli osobnosti pôsobiace predovšetkým na území terajšej Slovenskej republiky. Významná časť Belovej korešpondencie ale smerovala taktiež do dnešného Nemecka. Na základe farebného pokrytie sme teraz schopní identifikovať osobnosti pôsobiace na rovnakom geografickom území.

Hrúbka šípok (orientovaných hrán) v grafe odzrkadľuje ich váhu, pričom konkrétne hodnoty ponúka k nahliadnutiu tzv. laboratórium dát, kde sú údaje jednotlivých atribútov vyjadrené v odpovedajúcich stĺpcoch. Pomerne jednoduchým spôsobom dokážeme z údajov vyčítať tie osobnosti, ktoré sa na komunikácii podieľali najväčším podielom. Nie je pritom problémom sledovať aj miesta pôsobnosti týchto ľudí, lokalizovať ich na mape Európy, čím ich zaradíme do užšieho či širšieho kontextu,...atď.. Táto vlastnosť môže, spoločne s inými, neskôr priniesť nesmierne zaujímavé zistenia – po spracovaní korešpondencie niekoľkých desiatok až stoviek ďalších osobností.

## Záver

Proces vizualizácie vedeckej komunikácie v rokoch 1500–1800 sa aktuálne nachádza len vo svojej prvotnej, inicializačnej fáze, no predpovedáme jej úspešné napredovanie. Hoci sme zatiaľ analyzovali len korešpondenciu Mateja Bela, plánujeme neskôr prístupit' k analýze listov aj ostatných významných učencov v období rokov 1500-1800. Sme presvedčení, že ďalším spracovaním a skúmaním dát získame omnoho zložitejšie a zaujímavejšie (grafické) výstupy, ktoré nám pomôžu poodhaliť na prvý pohľad menej zrejme skutočnosti.

Moderné postupy a metódy vizualizovania vzájomných vzťahov sme realizovali s využitím moderného nástroja – softvéru Gephi, ktorý, ako sme prakticky preukázali v tomto príspevku, predstavuje vhodný prostriedok napomáhajúci pri naplňaní stanovených cieľov projektu INDED. Prioritne sa ďalšiemu výskumu bude vo svojej záverečnej bakalárskej práci venovať študentka študijného programu Mediamatika a kultúrne dedičstvo – Marcela Kajanová. Získané výsledky plánujeme opäť publikovať.

Ďalšie možnosti vizualizácie dát sa môžu týkať odlišenia korešpondencie na základe dátumu, alebo časového obdobia, kedy si vzdelanci odosielali korešpondenciu. V takomto prípade by bolo možné ponechať názvy odosielateľov a prijímateľov a podľa časového rozpätia odoslania správy meniť farbu jednotlivých uzlov. Taktiež sa ponúka možnosť pripraviť webové sídlo, kde by mohla byť umiestnená vizualizácia slúžiaca verejnosti.

## Zoznam bibliografických odkazov

BASTIAN, Mathieu. 2009. *Gephi: An Open Source Software for Exploring and Manipulating Networks* [online]. [cit. 2018-10-20]. Dostupné na:

<https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>

BUSINESSDICTIONARY. N.D. *Network* [online]. [cit. 2018-10-10]. Dostupné na:

<http://www.businessdictionary.com/definition/network.html>

DONNELLY J., H. Management. Praha: Grada Publishing, 2004. 824 s. ISBN 80-7169-422-3.

FAKULTA HUMANITNÝCH VIED. 2017. *Intelektuálne dedičstvo a vedecká komunikácia 1500-1800 so slovenskými vzťahmi ako súčasť európskej histórie a identity*.

GEPHI. 2017. *Features* [online]. [cit. 2018-10-26]. Dostupné na: <https://gephi.org/features/>

GIBSON, Helen et al. 2012. *A survey of two-dimensional graph layout techniques for information visualisation* [online]. [cit. 2018-09-19]. DOI: 10.1177/1473871612455749. Dostupné na : <http://www.leonidzhukov.net/hse/2015/sna/papers/gibson2013>

HARAUŠOVÁ, Helena. 2013. *Komunikácia v organizácii* [online]. [cit. 2018-10-22] dostupné na: <https://www.upjs.sk/public/media/5596/Komunikacia-v-organizacii.pdf>

JACOMY, Mathieu et al. 2014. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. In: *PLoS ONE*. Vol. 9, iss 6. DOI: <https://doi.org/10.1371/journal.pone.0098679> . Dostupné na: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679#s2>

JACOMY, Mathieu. 2011. *Force Atlas* [online]. 2011-06-06 [cit. 2018-10-25]. Dostupné na <https://gephi.wordpress.com/tag/force-atlas/>

OXFORD UNIVERSITY PRESS. 2018. *Network* [online]. [cit. 2018-10-10]. Dostupné na: <https://en.oxforddictionaries.com/definition/network>

POKORNÝ, Milan. 2013. *Diskrétna matematika* [on-line]. Trnava : Pedagogická fakulta Trnavskej univerzity v Trnave, 2013. 112 s. [5,6 AH]. ISBN 978--80--8082--569--0. Dostupné na : <http://pdf.truni.sk/e-ucebnice/diskretna-matematika/>

RALINOVSKI, Kiril. 2014. Improving the firefly algorithm through the Barnes-Hut tree code [online]. In : *ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges [Belgium], 23–25 April 2014* [cit. 2018-09-29]. ISBN 978–287419095–7. Dostupné na: <http://www.i6doc.com/fr/livre/?GCOI=28001100432440>

ZORALOVÁ, Martina. 2012. *Systém komunikace na Obecním úřadě v Bělé pod Pradědem* [online]. [cit. 2018-10-21]. Dostupné z: [https://theses.cz/id/8ln19p/BP\\_Zaoralov.pdf](https://theses.cz/id/8ln19p/BP_Zaoralov.pdf)