



Váňovaná slovná sieť pre zdravých ľudí a ľudí s problémom reči

vedúci práce : doc. RNDr. Mária Markošová, PhD.

katedra: FMFI.KAI - Katedra aplikovanej informatiky

riešiteľ: Bc. Lenka Čakurdová

FMFI 2024/2025

Cieľ práce

- ❑ cieľom diplomovej práce je spracovať ľudskú reč/text pomocou váhovaných slovných sietí
- ❑ skúmať reč zdravých ľudí a tiež ľudí s poškodením mozgu a afáziou – poruchou reči
- ❑ skúmať rôzne vlastnosti slovných sietí pomocou metód teórie grafov
- ❑ zistiť signifikantné rozdiely v daných slovných sieťach pre potenciálne diagnostikovanie afázie

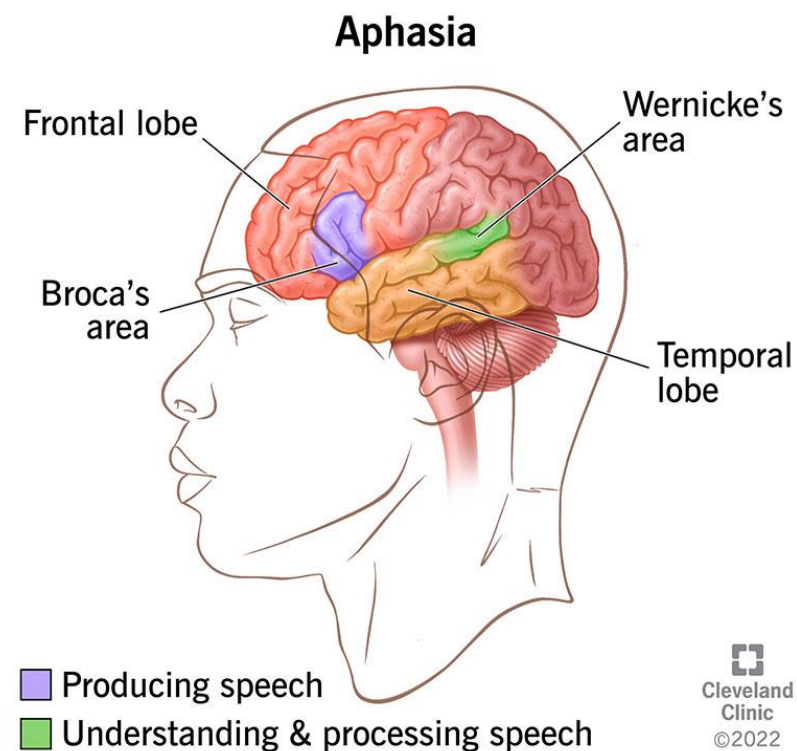
Afázia

- ❑ je porucha reči spôsobená poškodením mozgovej kôry v oblastiach spracovania reči
- ❑ poškodenia sú rôznej závažnosti a môžu ovplyvniť sluchové vnímanie reči, tvorbu reči, čítanie a/alebo písanie
- ❑ môže byť dôsledkom takmer akéhokoľvek typu poškodenia mozgu, vrátane traumatického poškodenia mozgu, nádoru, chirurgického odstránenia mozgového tkaniva alebo infekcie, tiež môže byť aj dôsledkom neurodegeneratívnych ochorení, ako je frontotemporálna demencia



Afázia

- ❑ poškodenia v rôznych častiach mozgu sa pri afázií prejavujú rozdielne
- ❑ Wernickeho oblasť : rýchla plynulá reč, žiadne motorické problémy, no neschopnosť porozumieť reči, tvorenie nezmyselných viet, náhodný slovosled, opakovanie slov
- ❑ Brockova oblasť : plne rozumejú reči, no majú problém sa vyjadriť, vynechávajú slová (agramatizmus) a rozprávajú veľmi pomaly
- ❑ vážnejšie typy afázie môžu predstavovať úplnú neschopnosť porozumieť reči

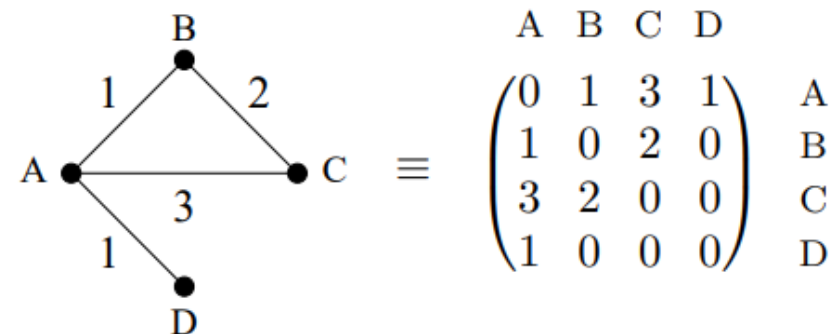


Graf

- ❑ množina vrcholov (uzlov), ktorých dvojice sú prepojené hranou
- ❑ mnoho grafov/sietí, ktoré boli skúmané vo fyzikálnej literatúre bolo binárneho charakteru, t. j., hrany medzi vrcholmi sú buď prítomné, alebo nie
- ❑ takéto siete môžu byť reprezentované (0, 1) alebo binárnymi maticami
- ❑ graf s n vrcholmi je reprezentovaný maticou A , $n \times n$, pre ktorú platí
$$A_{ij} = 1, \text{ ak vrcholy } i, j \text{ sú prepojené hranou}$$
$$0, \text{ ak nie sú prepojené}$$
- ❑ pojmy : slučka, podgraf, stupeň vrcholu, najmenšia priemerná vzdialenosť, klasterizačný koeficient

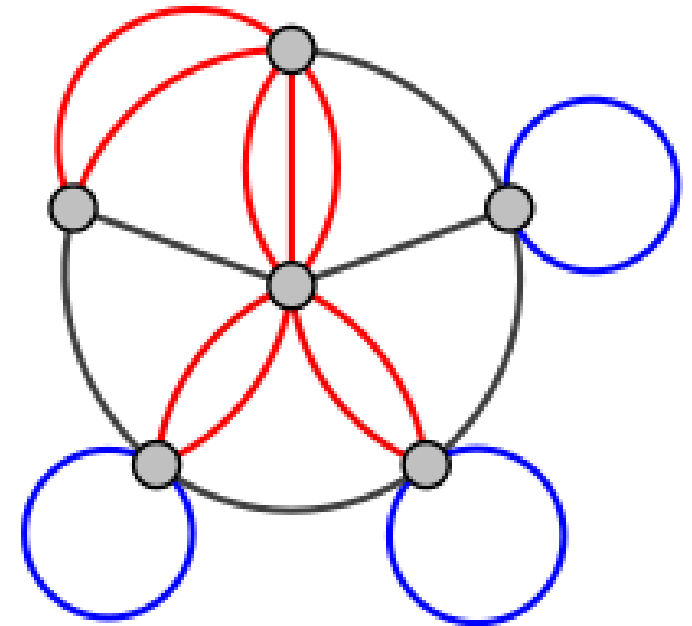
Váňovaný graf

- ❑ množina vrcholov (uzlov), ktorých dvojice sú prepojené hranou, každá hrana má priradenú váhu
- ❑ váha predstavuje „silu“ prepojenia uzlov, nesie informáciu o pridanej hodnote/vlastnosti prepojenia (vzdialenosť, počet prepravených pasažierov, počet strávených hodín ...)
- ❑ rovnako môže byť reprezentovaný maticou, no v prípade zaznačenia prepojenia danej dvojice uzlov je na príslušnom mieste hodnota váhy prepojenia
- ❑ do úvahy sa berie konektivita a rovnako aj váha prepojenia uzlov



Multigraf

- ❑ váhované grafy sú veľmi komplexné na spracovanie a skúmanie
- ❑ možnosť premapovať váhovaný graf na multigraf
- ❑ pôvodný graf má hranu medzi vrcholmi i, j označenú váhou w , následne premapovaný multigraf bude mať medzi danými vrcholmi w hrán označených váhou 1
- ❑ štruktúra grafu ostáva zachovaná
- ❑ jednoduchší prístup a spracovanie



Slovné siete

Ľudský jazyk je môžeme vnímať ako komplexnú sieť slov. No nie je možné ju definovať staticky, pretože jazyk sa časom vyvíja, mení, slová pribúdajú a niektoré sa vytrácajú. Rovnako sa môže meniť koncept a význam slov. Je teda nevyhnutné túto sieť vnímať ako neustále sa dynamicky meniacu a vyvíjajúcu.

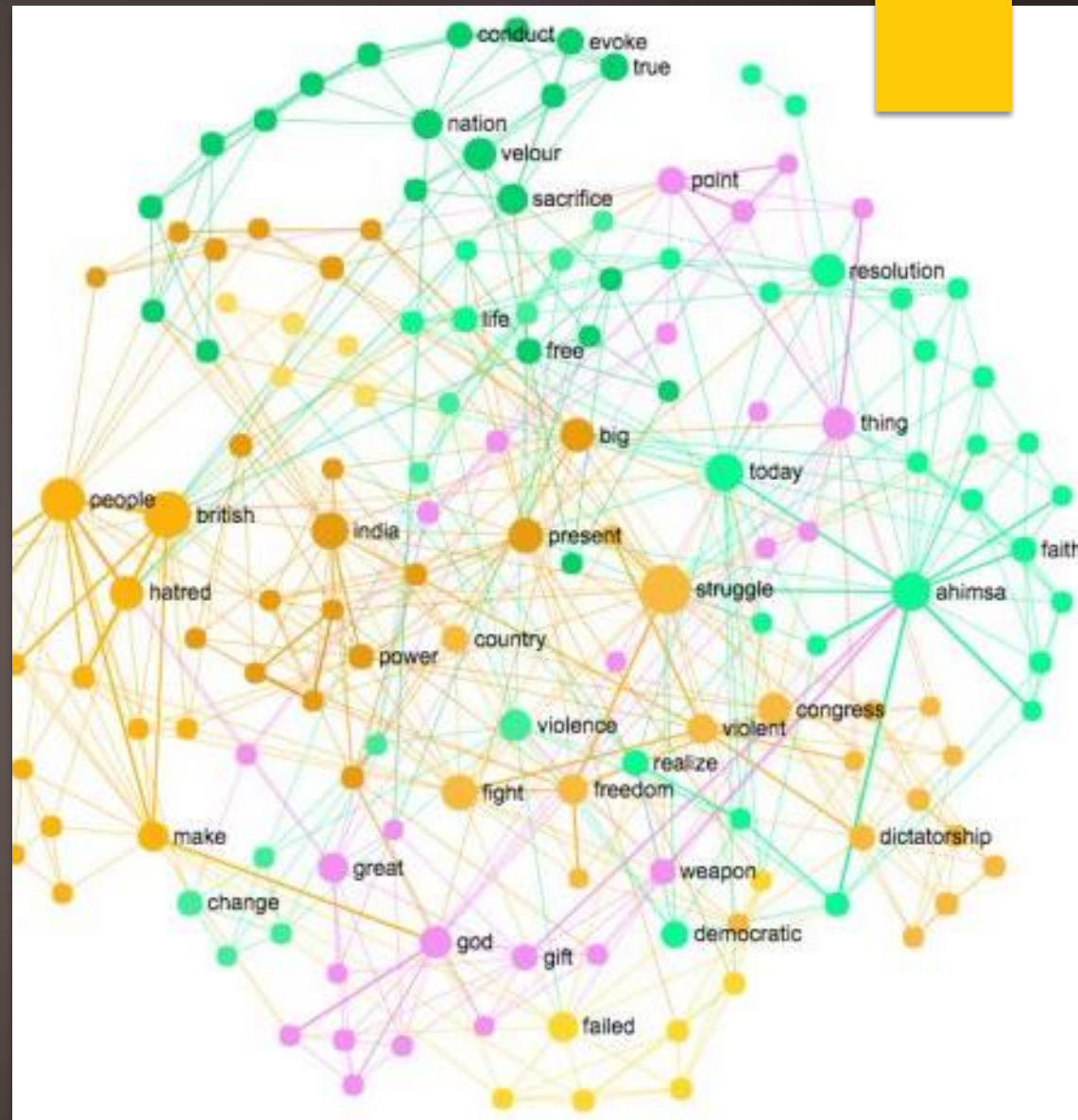
Nové slová pribúdajú a vytvárajú nové prepojenia/hrany, no rovnako, po čase môžu aj už existujúce spojenia medzi slovami vymiznúť, či vzniknúť úplne nové spojenia medzi staršími slovami.

Pretrhnutie veľmi dôležitých prepojení v tejto sieti môže viesť k identifikácií rôznych rečových porúch.

Slovná sieť predstavuje veľmi komplexnú architektúru podobnú vlastnostiam sietí World wide web a Internet.

Slovné siete

- graf, ktorý je tvorený z vrcholov (uzlov), ktoré predstavujú jednotlivé slová a z hrán grafu, ktoré predstavujú interakcie slov medzi sebou
- prepojenia slov môžu byť určené napríklad na základe susednosti vo vetách
- slová vo vetách môžu tiež interagovať vďaka syntaktickým väzbám, či zaužívaným výrazom. (základná škola, New York ...)



Skúmané vlastnosti slovných sietí

- ❑ pri spracovaní a porovnávaní slovných sietí zdravých ľudí a ľudí s afáziou sa budú skúmať vlastnosti sietí
- ❑ počet hrán a uzlov v sieti
- ❑ sila prepojenia uzlov
- ❑ stupeň uzlov a priemerný stupeň uzlov
- ❑ klasterizačný koeficient, priemerná najkratšia vzdialenosť
- ❑ unikátne slová, opakovanie slov, vynechávanie slov
- ❑ následne sa pomocou signifikantných rozdielov v hodnotách získaných zo slovných sietí určia ukazovatele potenciálneho poškodenia mozgu a afázie

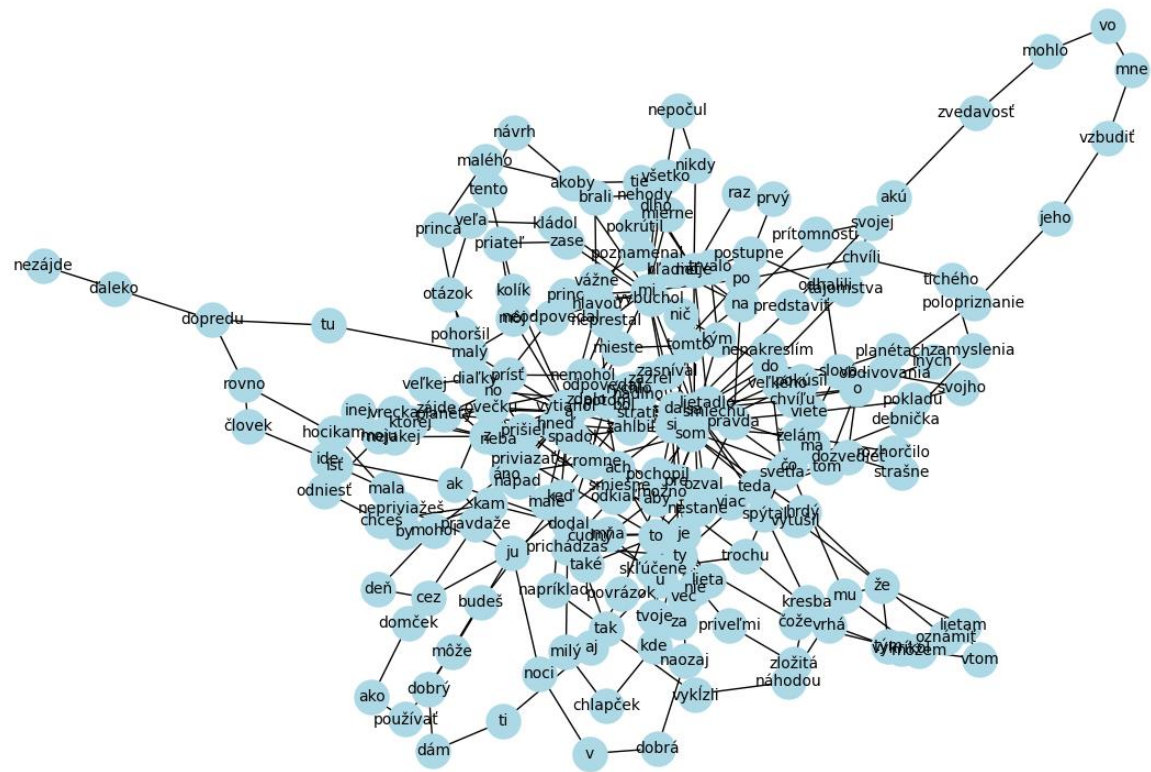
Plán práce

- ❑ získanie a spracovanie textov zdravých ľudí a ľudí s afáziou
- ❑ vybudovanie váhovanej slovnej siete z príslušných vstupov
- ❑ získavanie štatistických údajov a porovnávanie hodnôt u zdravých ľudí a ľudí s afáziou
- ❑ porovnávanie štatistických údajov získaných z váhovanej a neváhovanej slovnej siete
- ❑ vizualizácia slovnej siete
- ❑ písanie diplomovej práce

vlastnosť	neváňovaná sieť	váňovaná sieť
počet uzlov	209	209
počet hrán	307	313
maximálny stupeň uzla	26	13
minimálny stupeň uzla	1	1
priemerný stupeň uzla	2.937799043062201	1.6057692307692308
klasterizačný koeficient	0.04249291878337754	0.0268
priemerná najkratšia cesta	4.512881854987119	7.2799

Tabuľka 5.1: Analýza slovnej siete - Malý princ

Výskum



Výskum

Výskum

vlastnosť	<i>The Rat man of Paris</i>	<i>The shadow factory</i>
počet uzlov	8392	6557
počet hrán	36805	26510
maximálny stupeň uzla	2969	2045
minimálny stupeň uzla	1	1
priemerný stupeň uzla	7.033484270734032	6.347414976361141
klasterizačný koeficient	0.2671	0.2559
priemerná najkratšia cesta	2.92499	2.89893

Tabuľka 5.2: Analýza váhovanej slovnej siete textu *The Rat man of Paris* a *The shadow factory*

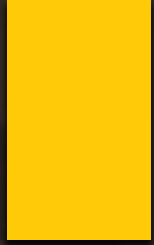
Pokračovanie práce

- ❑ GUI pre jednoduché ovládanie a zobrazovanie výsledkov
- ❑ analýza distribúcie interpunkčných znamienok
- ❑ analýza podgrafov
- ❑ vyhodnotenie výsledkov a určenie ukazovateľov potenciálnej diagnostiky afázie
- ❑ písanie diplomovej práce

Literatúra



- Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Modeling the evolution of weighted networks. *Physical Review E*, 70(6), December 2004.
- Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications*, 346(1–2):34–43, February 2005.
- S. N. Dorogovtsev and J. F. F. Mendes. *Language as an evolving word web*, 2001.
- Ramon Ferrer-i Cancho and Ricard Sole. The small world of human language. *Proceedings. Biological sciences / The Royal Society*, 268:2261–5, 12 2001.
- Ilona Henseler, Frank Regenbrecht, and Hellmuth Obrig. Lesion correlates of pathological profiles in chronic aphasia: Comparisons of syndrome-, modality- and symptom-level assessment. *Brain : a journal of neurology*, 137, 02 2014
- M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5), November 2004.



Ďakujem za
pozornosť