Master thesis

Jannik Gut
12. may

Table of content

- SubGNN tryouts
  - Also cluster update
- SPOKE test
- Other papers
  - Safe medical recommendations, close to ours
  - EHR Coding with Multi-scale Feature Attention and Structured Knowledge Graph Propagation
  - Other
- Discussion
  - Dataset/knowledge base
  - Next steps

SubGNN tryouts

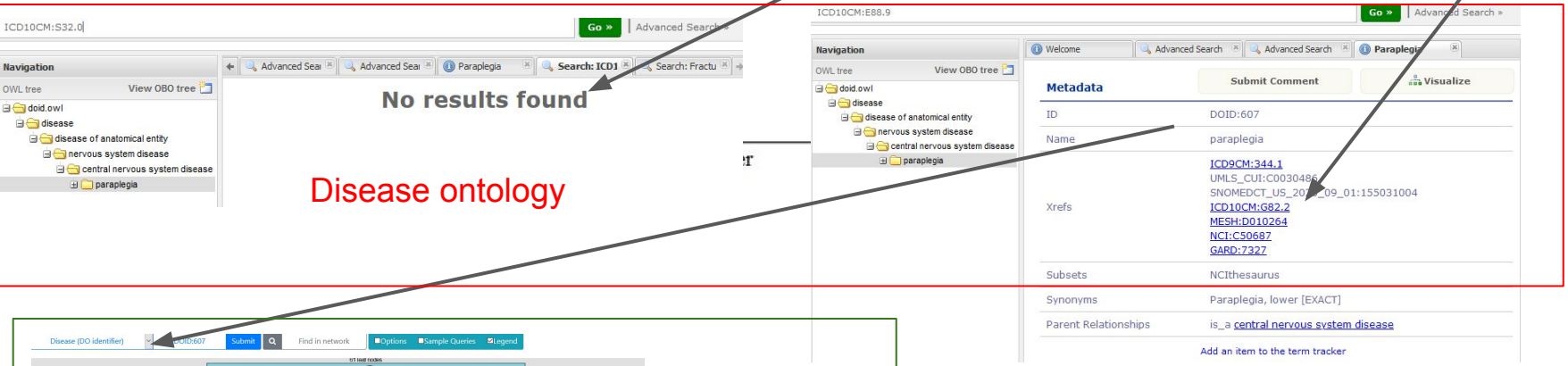
- Google Colab was too problematic considering real data can't be used
- Brother's computer only has Windows and SubGNN needs Linux libraries
- My computer barely has a GPU and can not work with the server libraries
- Cluster
  - Spin up a Linux distro
    - PuTTy and OpenSSH have different key file layouts
    - Now have Markdown with LaTeX
  - Can connect to cluster
  - But gogo scripts do not work → in contact with Kieran
    - No news from dataset
  - Did not look into previous code yet

SPOKE

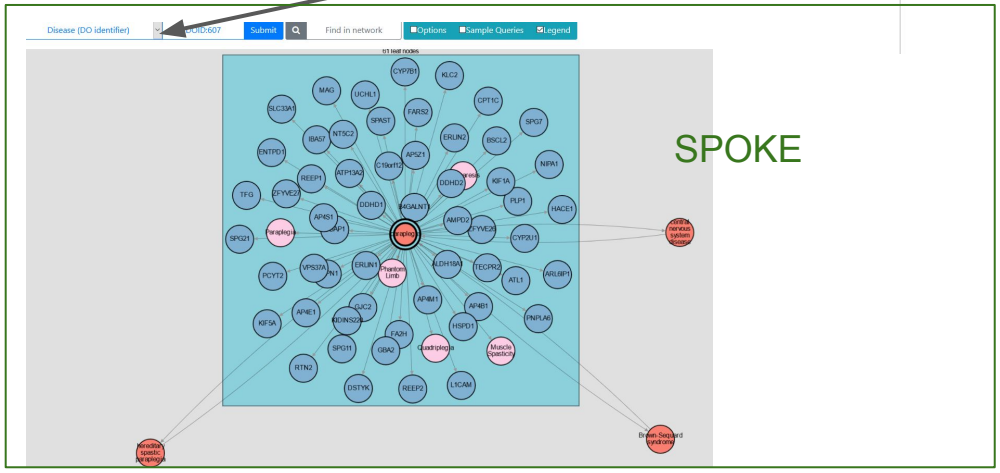
Has the Disease Ontology integrated, which can use ICD codes.

BFS EHR

Hauptdiagnose	Lumbale Wirbelfraktur	S32.0
Zusatz zu Hauptdiagnose	Sturz von Gerüst	W12.6
1. Nebendiagnose	Paraplegie ab L2	G82.2
2. Nebendiagnose	Blasen- und Sexualfunktionsstörung	G83.4



Disease ontology



SPOKE

Hauptdiagnose	Lumbale Wirbelfraktur	S32.0
Zusatz zu Hauptdiagnose	Sturz von Gerüst	W12.6
1. Nebendiagnose	Paraplegie ab L2	G82.2
2. Nebendiagnose	Blasen- und Sexualfunktionsstörung	G83.4


ICD10CM:S32.0
Go »
Advanced Search »
Navigation
OWL tree
View OBO tree
doid.owl
disease
disease of anatomical entity
nervous system disease
central nervous system disease
paraplegia
Advanced Sear
Advanced Sear
Paraplegia
Search: ICD1
Search: Fractu
No results found

ICD10CM:E88.9

Go » | Advanced Search »

Navigation

OWL tree View OBO tree

- doid.owl
  - disease
    - disease of anatomical entity
      - nervous system disease
        - central nervous system disease
          - paraplegia

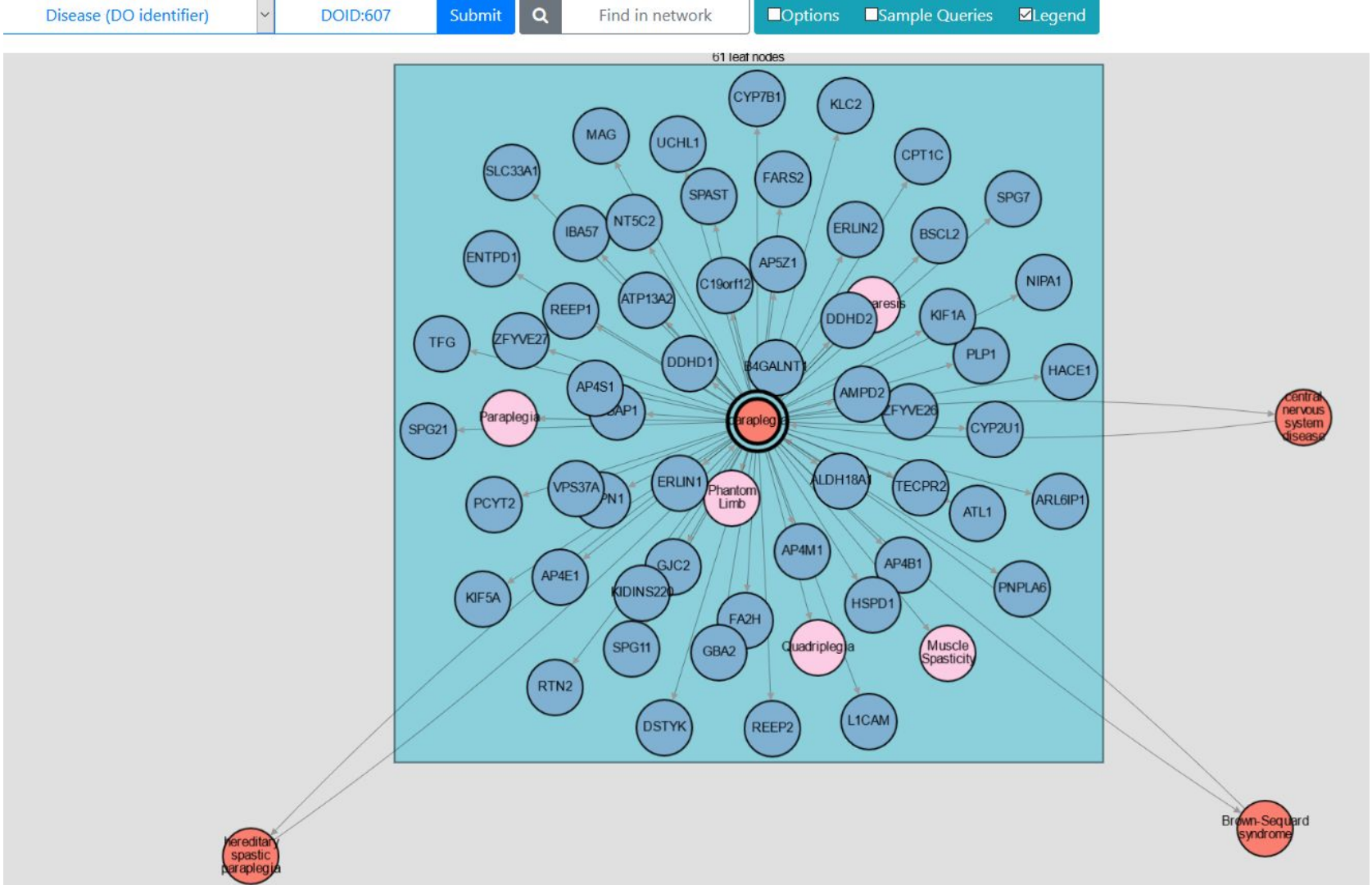
Welcome | Advanced Search | Advanced Search | **Paraplegia**

Metadata

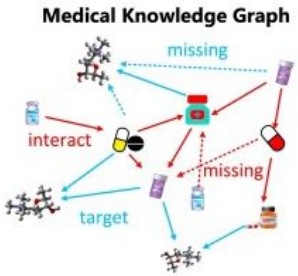
Submit Comment | Visualize

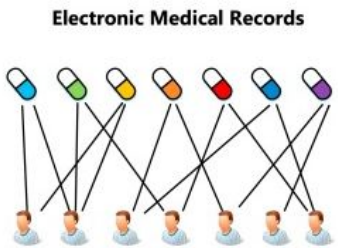
ID	DOID:607
Name	paraplegia
Xrefs	ICD9CM:344.1 UMLS_CUI:C0030486 SNOMEDCT_US_2020_09_01:155031004 ICD10CM:G82.2 MESH:D010264 NCI:C50687 GARD:7327
Subsets	NCIthesaurus
Synonyms	Paraplegia, lower [EXACT]
Parent Relationships	is_a central nervous system disease

Add an item to the term tracker


Disease (DO identifier)
DOID:607
Submit
Find in network
Options
Sample Queries
Legend
61 leaf nodes
CYP7B1
KLC2
MAG
UCHL1
CPT1C
SLC33A1
FARS2
SPG7
SPAST
NT5C2
IBA57
ERLIN2
BSCL2
ENTPD1
AP5Z1
NIPA1
C19orf12
ATP13A2
REEP1
DDHD2
KIF1A
TFG
ZFYVE27
PLP1
DDHD1
B4GALNT1
HACE1
AP4S1
AMPD2
SPG21
Paraplegia
ZFYVE26
CYP2U1
ERLIN1
Phantom Limb
ALDH18A1
TECPR2
PCYT2
VPS37A
ARL6IP1
ATL1
AP4M1
AP4B1
GJC2
AP4E1
PNPLA6
KIF5A
HSPD1
FA2H
SPG11
GBA2
Quadriplegia
Muscle Spasticity
RTN2
DSTYK
REEP2
L1CAM
central nervous system disease
Hereditary spastic paraplegia
Brown-Sequard syndrome

Safe medical recommendations





- Embed MIMIC-III in DrugBank and ICD-9
  - No better dataset with e.g. “cures”
- $$S(q, m_N) = p^T m_n - \sum_{o=1}^{n-1} ||m_n + r_{interaction} - m_o||_{L1/L2}$$
- 3 graphs jointly embedded with TransR (triplet translation)
  - Knowledge base (disease → disease, medicine → medicine)
  - Patient → medicine
  - Patient → disease
- Patient is a sum of diseases, which have to be covered by medicine
-
- Problem of cold start of never seen therapy
-
- Demographics not used, combinations not possible, no flow between graphs
- **Can we model the graph without the patients as intermediate?**

EHR Coding with Multi-scale Feature Attention and Structured Knowledge Graph Propagation

- Assign ICD code based on description
-
- EHR NLP data through this *NER* model
- NLP description of ICD as node representation and run GNN 2-3 times to get leaf representations
- Sigmoid of dot product as metric for probability
-
- Two MIMIC-III modes, one with all labels and the other with only the top 50
    - somewhat popular

Other

- Many papers that use NLP data to form a knowledge graph
  - Gives good results, even with rudimentary techniques
- No knowledge base between disease and cures
  - If there was, others would have used it
- Tasks
  - Medication rank, considering drug interactions and symptoms
  - Disease prediction, given symptoms
  - Finding new interactions/symptoms

Discussion

- Dataset
  - Only use ontologies as knowledge base
  - Only take the ones in Disease Ontology
  - Create one by ourselves
  - Other idea?
- Next steps
  - Move forward with cluster if possible
    - Dataset exploration
    - SubGNN testing
  - ???
    - Papers also drying up
  - Else Markdown + LaTeX