2019 Michigan IT Symposium

Building a Computable Patient History

Using Natural Language Processing and Graph Databases to Turn Words into Ideas

Kate Weber

Sr. Data Analyst

kgweber@umich.edu

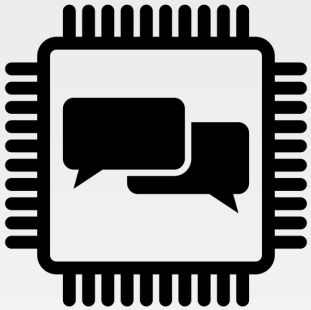
SCHOOL OF DENTISTRY

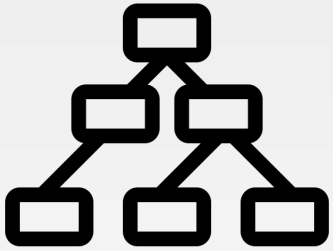
DENTAL INFORMATICS

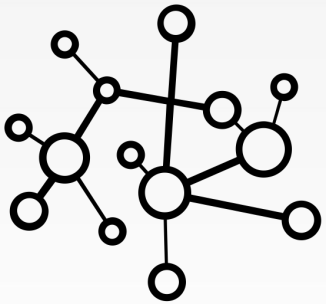
UNIVERSITY OF MICHIGAN

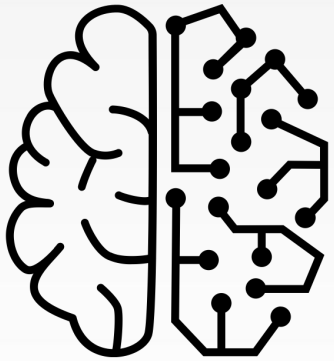
Today's Talk

- Introduction to the Problem
- Components of the solution
  - Natural Language Processing
  - Standard Medical Ontologies
  - Network Databases and Algorithms
- Analysis Example
- Applications
- Ways to Improve
- Technical Components
- Q&A



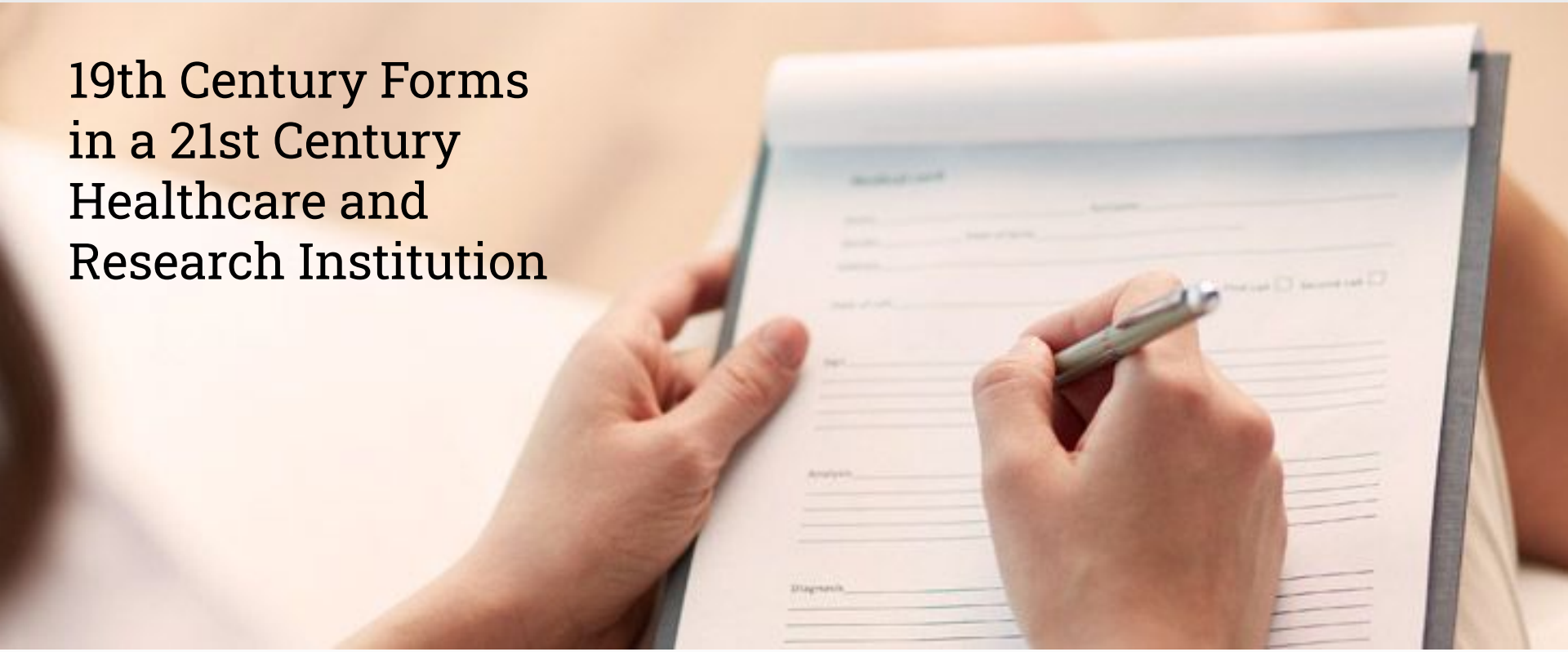




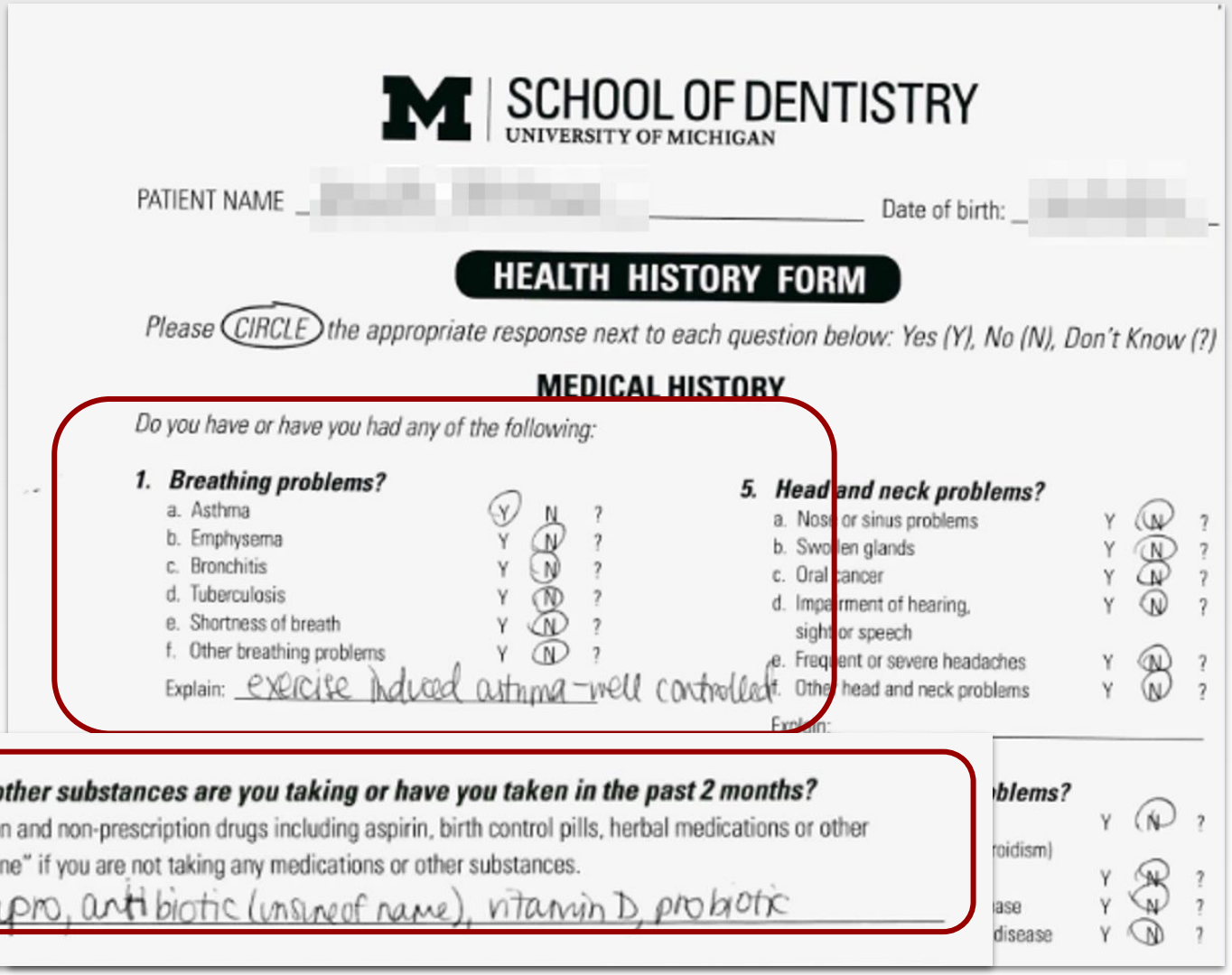


The Problem

19th Century Forms in a 21st Century Healthcare and Research Institution



We Still Collect Data On Paper



SCHOOL OF DENTISTRY
UNIVERSITY OF MICHIGAN

PATIENT NAME _____ Date of birth: _____

HEALTH HISTORY FORM

Please CIRCLE the appropriate response next to each question below: Yes (Y), No (N), Don't Know (?)

MEDICAL HISTORY

Do you have or have you had any of the following:

1. Breathing problems?

a. Asthma	(Y)	N	?
b. Emphysema	Y	(N)	?
c. Bronchitis	Y	(N)	?
d. Tuberculosis	Y	(N)	?
e. Shortness of breath	Y	(N)	?
f. Other breathing problems	Y	(N)	?

Explain: exercise induced asthma - well controlled

5. Head and neck problems?

a. Nose or sinus problems	Y	(N)	?
b. Swollen glands	Y	(N)	?
c. Oral cancer	Y	(N)	?
d. Impairment of hearing, sight or speech	Y	(N)	?
e. Frequent or severe headaches	Y	(N)	?
f. Other head and neck problems	Y	(N)	?

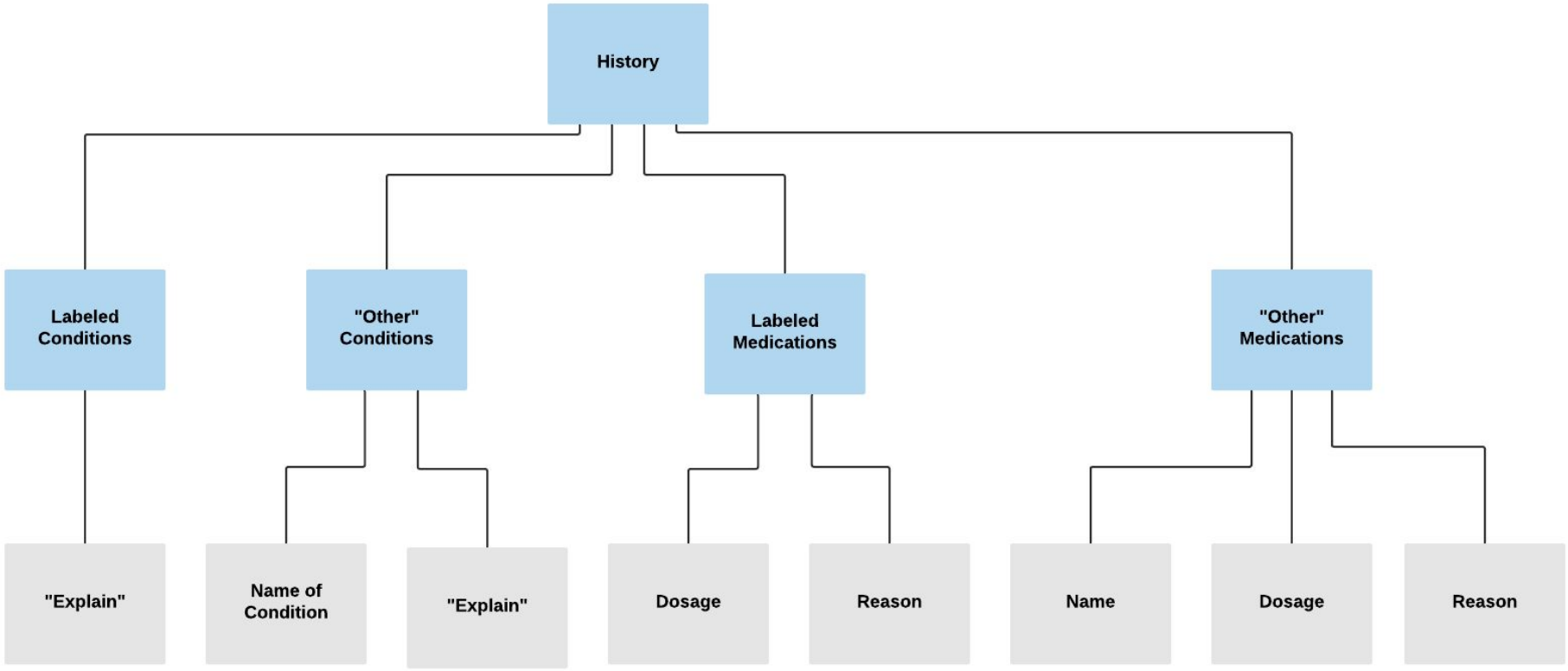
Explain:

11. What medications or other substances are you taking or have you taken in the past 2 months?

a. Please list all prescription and non-prescription drugs including aspirin, birth control pills, herbal medications or other supplements. Write "none" if you are not taking any medications or other substances.

Singulair, Lexapro, antibiotic (unsure of name), vitamin D, probiotic

...and translate a fraction into computable form



Things We Can't Do

- Extracting Facts
- Linking Concepts
- Connecting the Same Information from Different Forms
- Answering Questions
- Identifying Entities
- Literature Processing
- Phenotyping
- Epidemiology and Public Health
- Linking to other Health Systems

To: DENT Data Analytics
From: diligent_researcher@umich.edu
Subject: HELP!!!

My research is on "Relationship between periimplantitis and cardiovascular disease".

I want to recruit patients who have at least one dental implant and cardiovascular disease with the specific time sequence "implant placement first, CVD second".

Patient Services has been really helpful cross-matching the patient list from the Frankel cardiovascular center and the Dental school, and we tried to cross-match the medication for CVD as well.

But it hasn't been really fruitful. There's no way to perform keyword searches for "cardiovascular disease" or "stroke" or "cardiac stent" etc.

What can we do?!

Natural Language Processing in Action

The patient is a 63-year-old female with a three-year history of bilateral hand numbness and occasional weakness.

Within the past year, these symptoms have progressively gotten worse, to encompass also her feet.

She had a workup by her neurologist and an MRI revealed a C5-6 disc herniation with cord compression and a T2 signal change at that level.

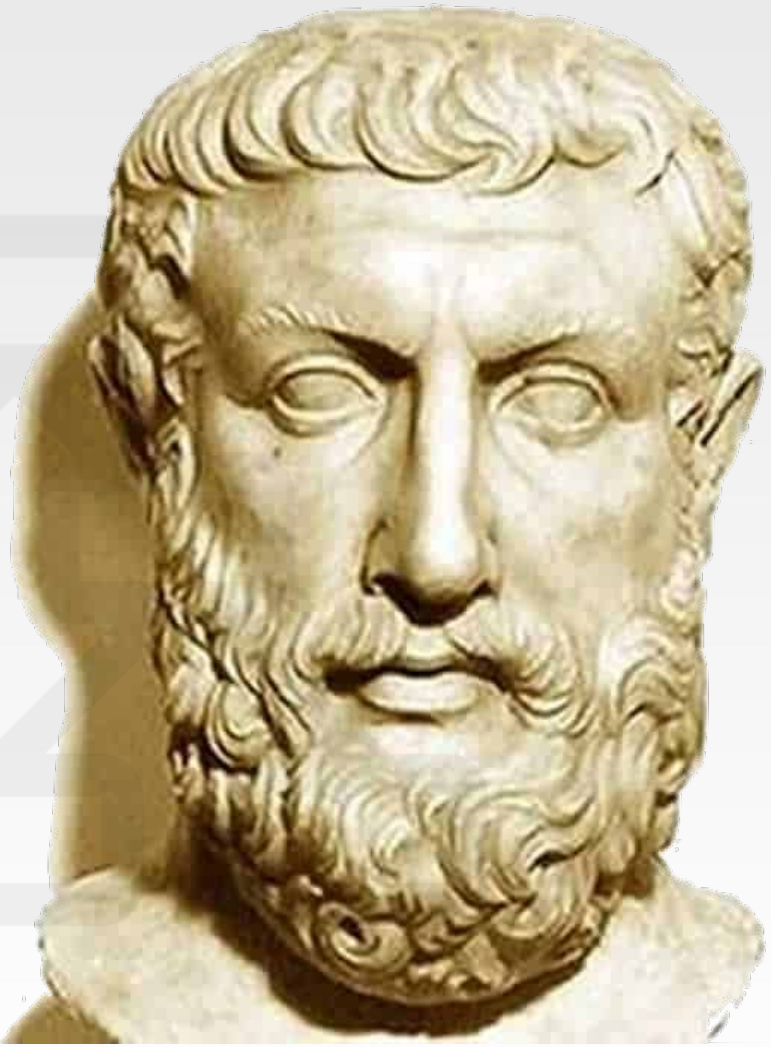
Challenges

- **Synonymy:** Same idea, different words
  - “Diabetes” “Type 1 Diabetes” “Juvenile Diabetes” “Sugar Disease” “Dbaites”
- **Polysemy**: Same word, different ideas
  - “CT” = Computed Tomography; Cardiothoracic; Carpal Tunnel; Cognitive therapy…
- **Negation:** About half of clinical concepts are negated
  - “Patient does not show signs of diabetes.”
- **Uncertainty**:
  - “Tentative diagnosis of tuberculosis.”
- **Messy Record Structure**: partially structured data, copy/paste templates
  - “CARDIOVASCULAR: [ ] Angina [ ] MI [x] HTN [ ] CHF [ ] Denies CV Problems”
- **Temporality** - keeping track of history vs current
- **Specialized Dialects** - Each specialty or subculture can use language differently

Connections to Knowledge

"**Ontology** is the philosophical study of being. More broadly, it studies concepts that directly relate to being, in particular becoming, existence, reality, as well as the basic categories of being and their relations."

Wikipedia

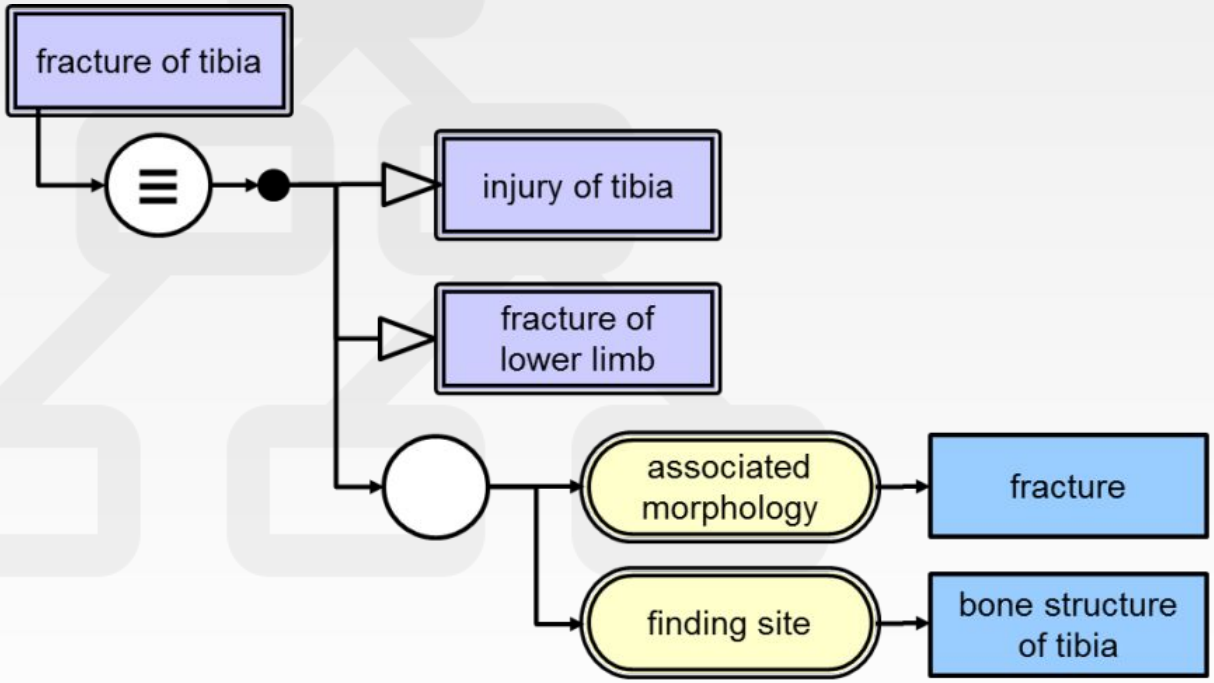


The Leg Bone's Connected to the Foot Bone

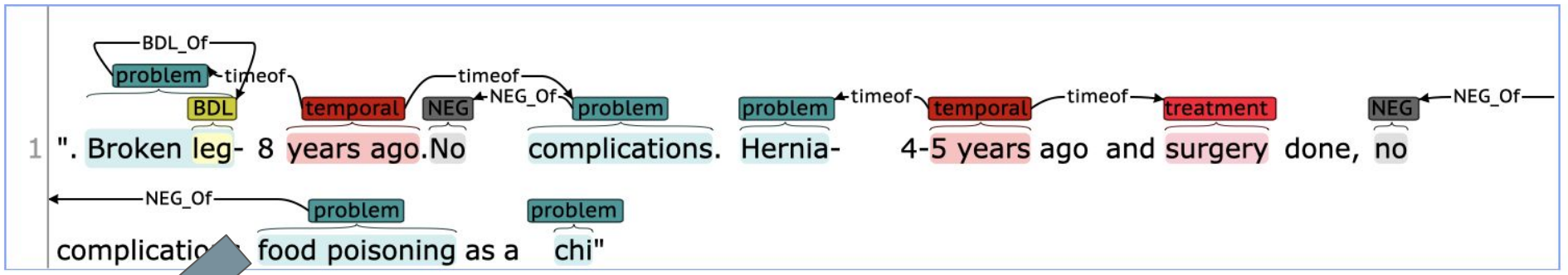
"**Medical ontologies** are built by humans but are intended to be used computationally and thus are formally structured. For example, a medical ontology can express that a disease whose clinical course is sudden or short is equivalent to an acute disease."

Lexigram.io

- **OMOP**
- **UMLS**
- **SNOMED-CT**
- **RXNORM**
- **ICD-9/10**

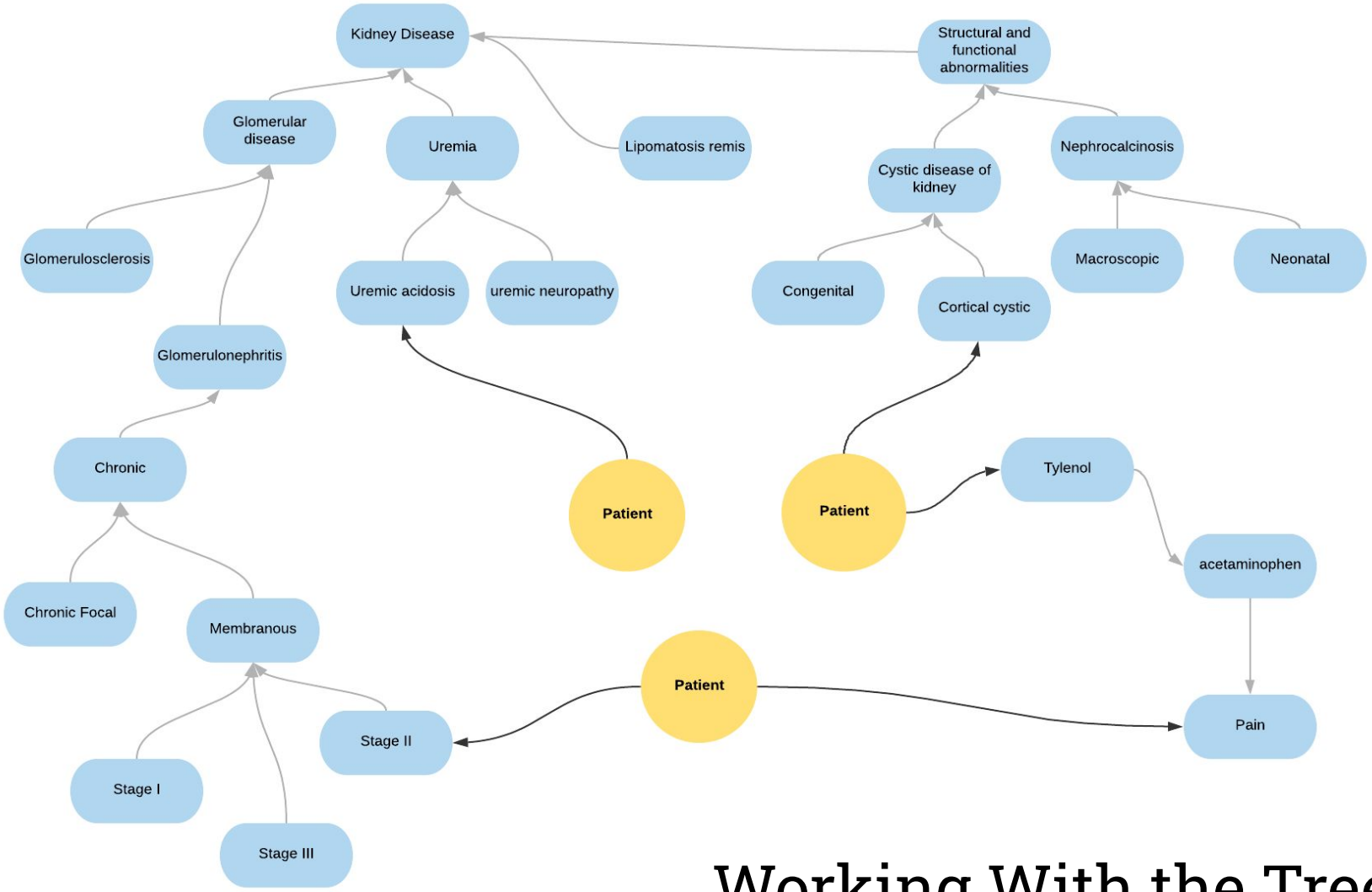


Tying It Together



problem

Attribute	Value
semanticTag	problem
startPos	105
endPos	119
text	food poisoning
assertion	absent
cui	C0016479
cuiDesc	Food Poisoning



Working With the Tree

Adapted from El-Sappagh S, Elmogy M. An encoding methodology for medical knowledge using SNOMED CT ontology. Journal of King Saud University-Computer and Information Sciences. 2016 Jul 1;28(3):311-29.

Tools

I'm way too cool for fancy-pants logos

Python Integration is awesome

MedEx

MetaMap

CLAMP

Clinical Language Annotation, Modeling, and Processing Toolkit

Output

```
00000000|MMI|5.18|Practice Experience|C0237607|[menp]|["experience"-tx-1-"experience"-verb-0]|TX|174
00000000|MMI|5.18|State|C1442792|[ftcn]|["State"-tx-1-"states"-verb-0]|TX|97/6|
00000000|MMI|5.18|Substance|C0439861|[sbst]|["Substances"-tx-1-"substances"-noun-0]|TX|48/10|
00000000|MMI|5.18|US State|C3148680|[geoa]|["State"-tx-1-"states"-verb-0]|TX|97/6|
00000000|MMI|3.68|Physical therapy (field)|C0699718|[bmod]|["PT"-tx-1-"Pt"-noun-0]|TX|94/2|
00000000|MMI|3.60|Weakness|C3714552|[sosy]|["WEAKNESS"-tx-1-"weakness"-noun-0]|TX|33/8|
00000000|MMI|3.59|allergic|C0700624|[ftcn]|["Allergic"-tx-1-"Allergic"-adj-0]|TX|59/8|
00000000|MMI|3.55|Day 5|C3842674|[fndg]|["Day 5"-tx-1-"5 day"-noun-1]|TX|322/1,357/3|
00000000|MMI|3.48|mg/mg|C1319635|[qnco]|["mg/mg"-tx-1-"mg mg"-noun-0]|TX|151/2,158/2|
00000000|MMI|3.46|100%|C3817553|[qnco]|["100%"-tx-1-"100"-integer-0]|TX|147/3|
00000000|MMI|3.46|Medications|C4284232|[inpr]|["MEDS"-tx-1-"meds"-noun-0]|TX|142/4|
00000000|MMI|3.45|atorvastatin|C0286651|[orch,phsu]|["ATORVASTATIN"-tx-1-"Atorvastatin"-noun-0]|TX|7
00000000|MMI|3.42|Does stand up|C0600132|[fndg]|["stand up"-tx-1-"stand up"-adj-0]|TX|217/8|
00000000|MMI|3.42|STAND-UP (product)|C0310197|[phsu]|["STAND-UP"-tx-1-"stand up"-adj-0]|TX|217/8|
00000000|MMI|3.42|Standing up|C0444796|[spco]|["stand up"-tx-1-"stand up"-adj-0]|TX|217/8|
00000000|MMI|3.42|CMTR2 gene|C2681537|[gngm]|["AFT"-tx-1-"aft"-adj-0]|TX|226/3|
00000000|MMI|3.42|CMTR2 wt Allele|C4319972|[gngm]|["AFT"-tx-1-"aft"-adj-0]|TX|226/3|
00000000|MMI|3.42|Drop (unit of presentation)|C4318619|[qnco]|["Drop"-tx-1-"drop"-noun-0]|TX|349/4|
00000000|MMI|3.42|Drop - Unit of Volume|C4048603|[qnco]|["DROP"-tx-1-"drop"-noun-0]|TX|349/4|
00000000|MMI|3.42|Drop British|C2348338|[qnco]|["Drop"-tx-1-"drop"-noun-0]|TX|349/4|
```

Cool, but now what?

Where to store it?

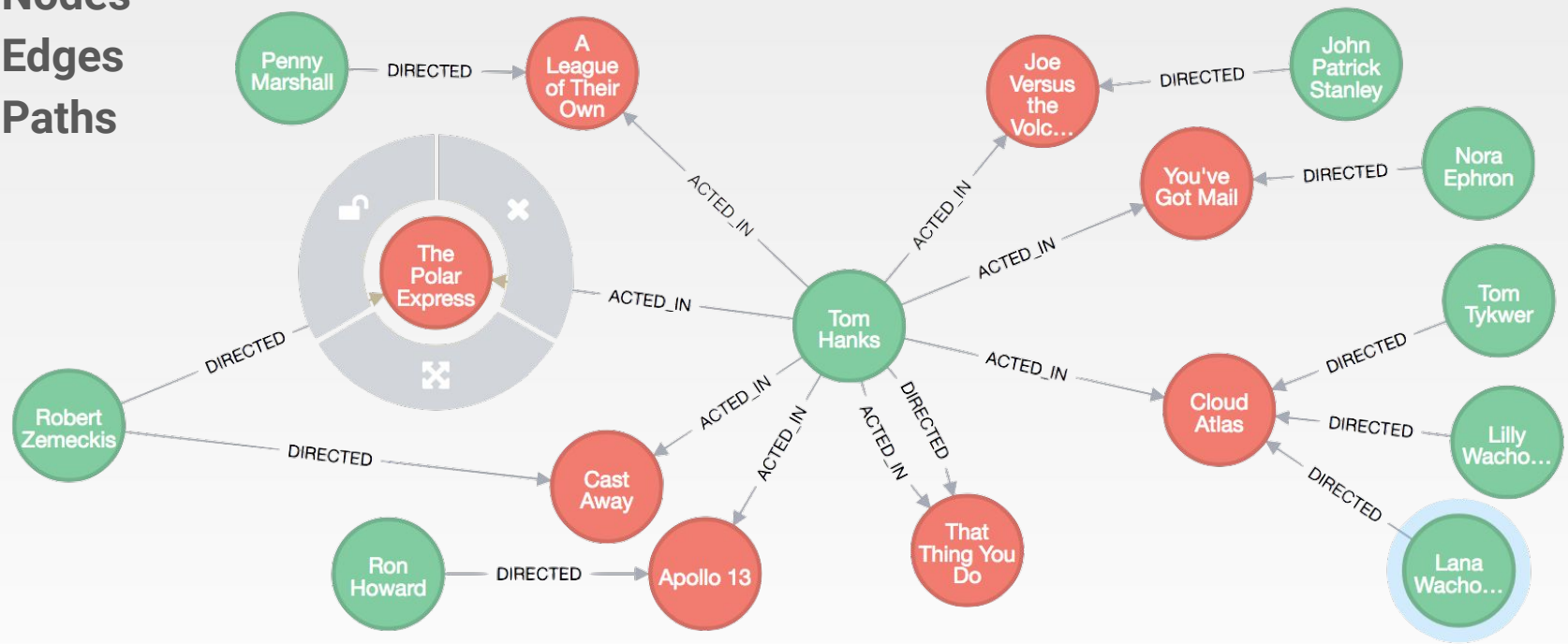
How to process it?

Is it good for anything?

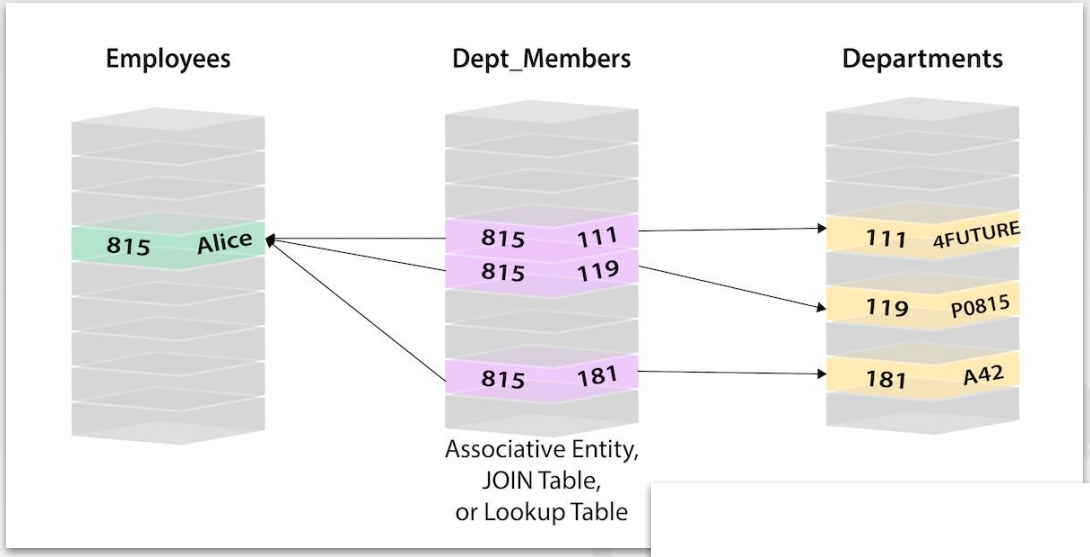
Graph Databases

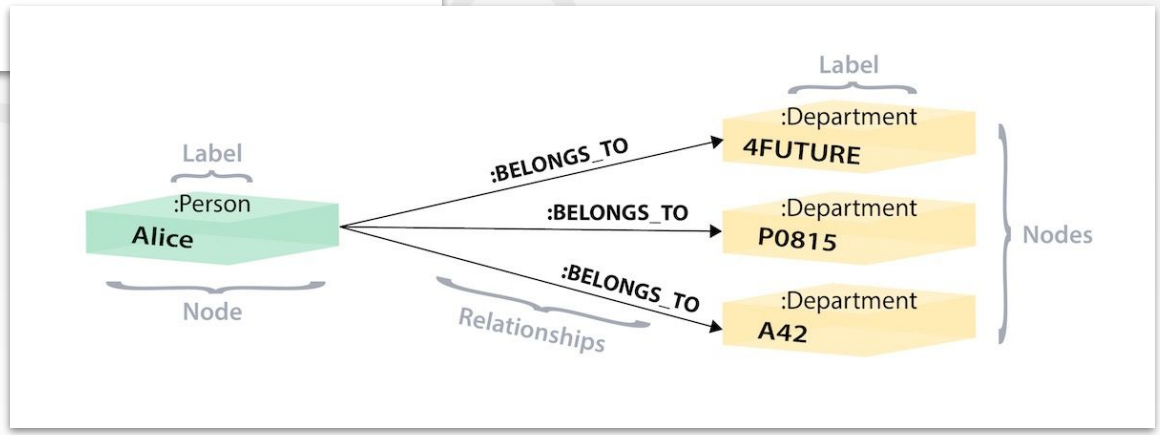
All about the connections

- Nodes
- Edges
- Paths

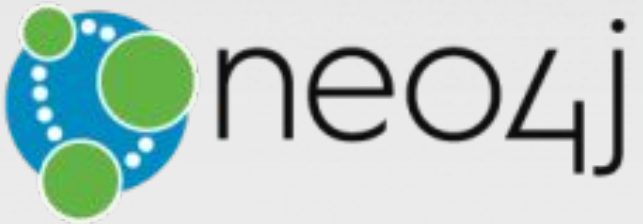


Graph vs RDBMS





Demo




```
MATCH (tom {name: "Tom Hanks"}) RETURN tom;

MATCH (tom:Person {name: "Tom Hanks"})-[:ACTED_IN]->(tomHanksMovies)
RETURN tom,tomHanksMovies;

MATCH (TomH:Person{name:'Tom Hanks'})
MATCH (TomH)-[:ACTED_IN]->(m)<-[:DIRECTED]-(d)
RETURN TomH,m,d LIMIT 10;

MATCH (bacon:Person {name:"Kevin Bacon"})-[*1..4]-(hollywood)
RETURN DISTINCT hollywood;

MATCH p=shortestPath(
 (bacon:Person {name:"Kevin Bacon"})-[*]-(meg:Person {name:"Meg Ryan"}))
RETURN p;

MATCH (tom:Person {name:"Tom Hanks"})-[:ACTED_IN]->(m)<-[:ACTED_IN]-(coActors),
    (coActors)-[:ACTED_IN]->(m2)<-[:ACTED_IN]-(cruise:Person {name:"Tom Cruise"})
RETURN tom, m, coActors, m2, cruise;
```

Graph Algorithms: Centrality

Who is best connected to other nodes?

What is the most important node?

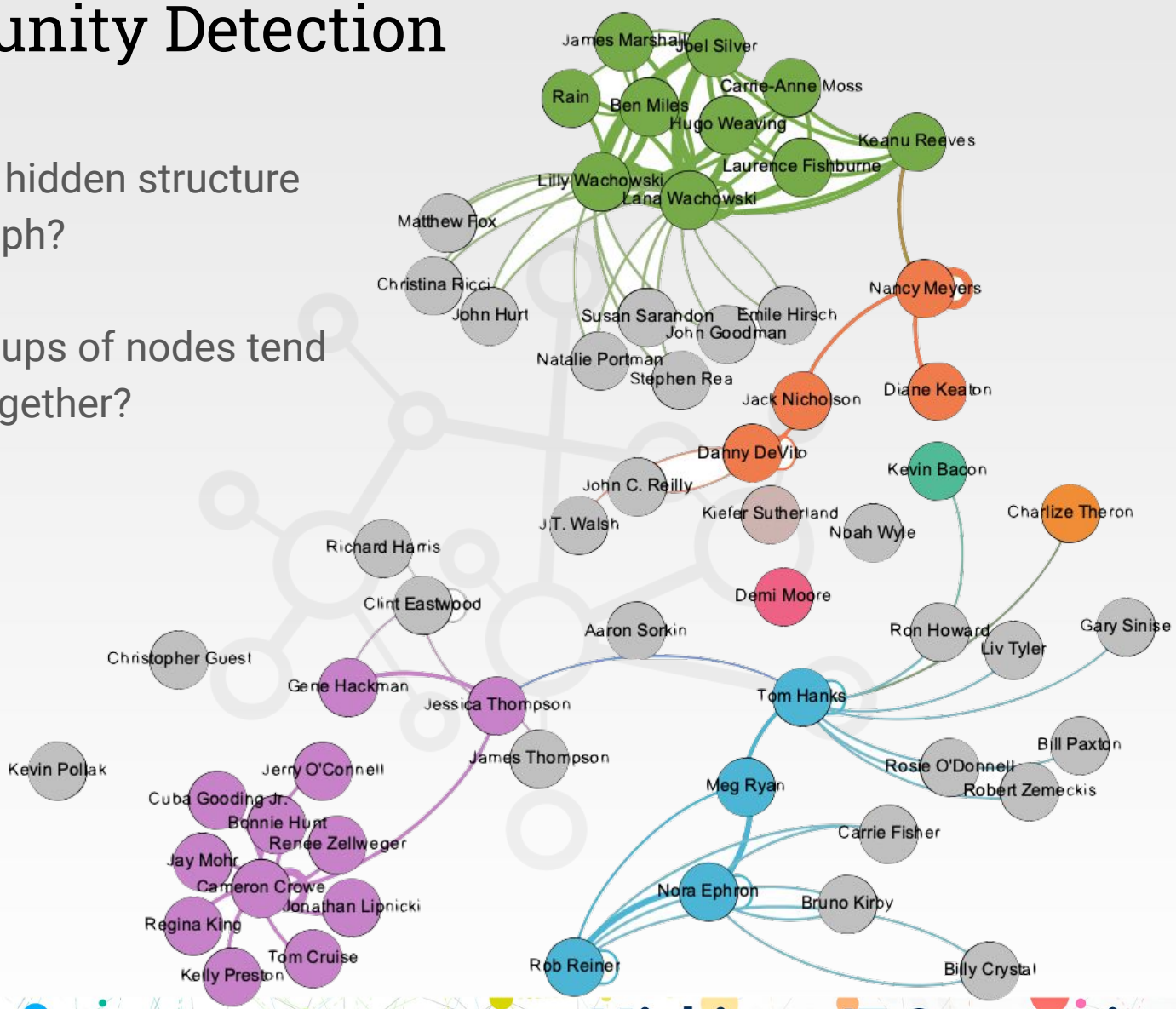
```
call algo.pageRank.stream(
  'MATCH (p:Person) return id(p) as id',
  'MATCH (p1:Person)-->(:Movie)<--(p2:Person)
   WITH p1,p2,count(*) as commonMovies
   WHERE commonMovies > 2
   RETURN id(p1) as source, id(p2) as target',
{graph:'cypher'})
YIELD nodeId, score
WITH nodeId, score
ORDER BY score desc limit 10

RETURN algo.asNode(nodeId).name as person, score
```

"person"	"score"
"Cameron Crowe"	5.0992637382820245
"Lana Wachowski"	1.7025186547078193
"Lilly Wachowski"	1.7025186547078193
"Nancy Meyers"	1.3125946707092224
"Keanu Reeves"	1.2911624179221688
"Nora Ephron"	1.254318251507357
"Meg Ryan"	1.254318251507357
"Joel Silver"	1.1893135151360181
"Jessica Thompson"	1.0404261087300257
"Laurence Fishburne"	1.0249650554265828

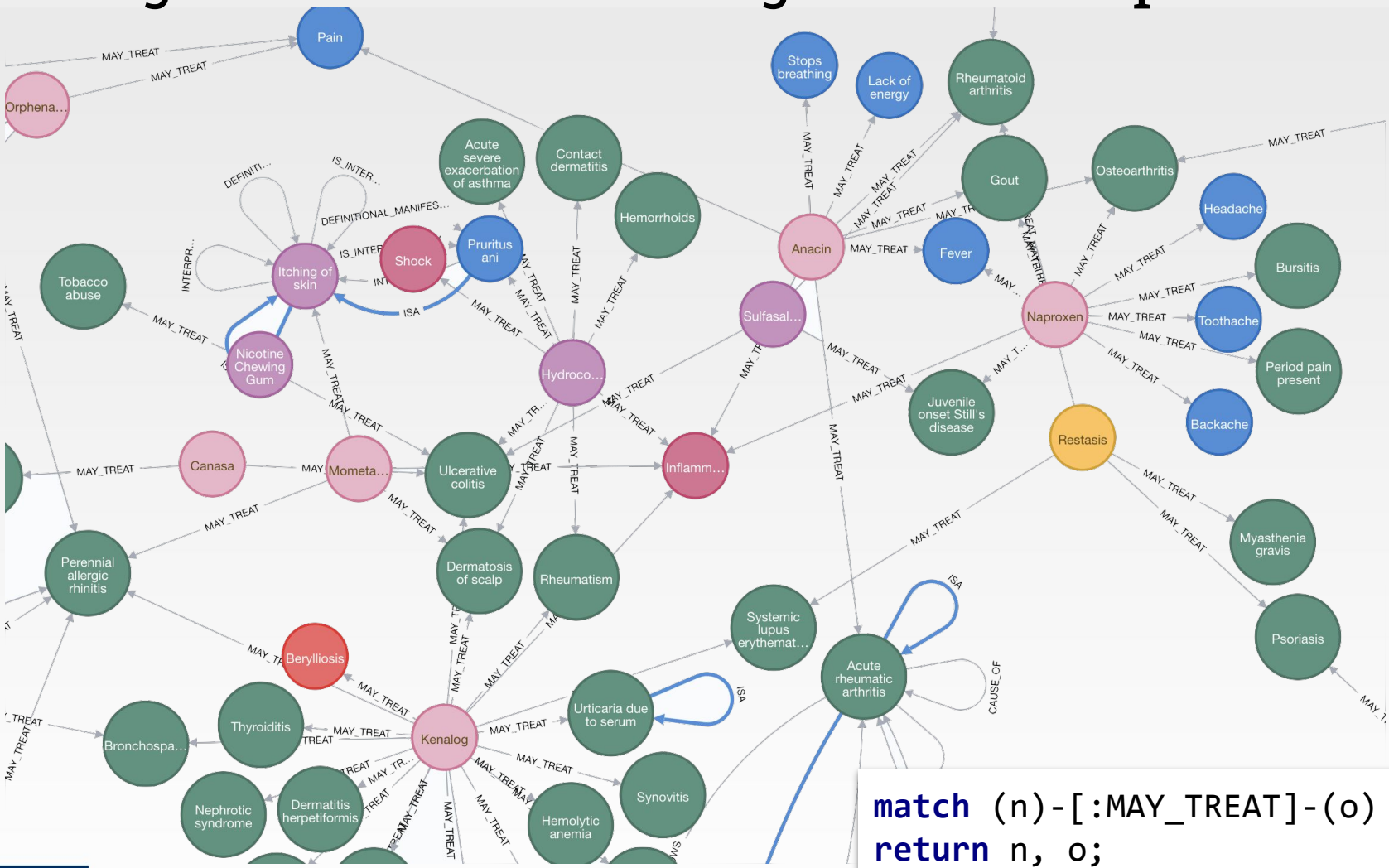
Community Detection

What is the hidden structure within a graph?

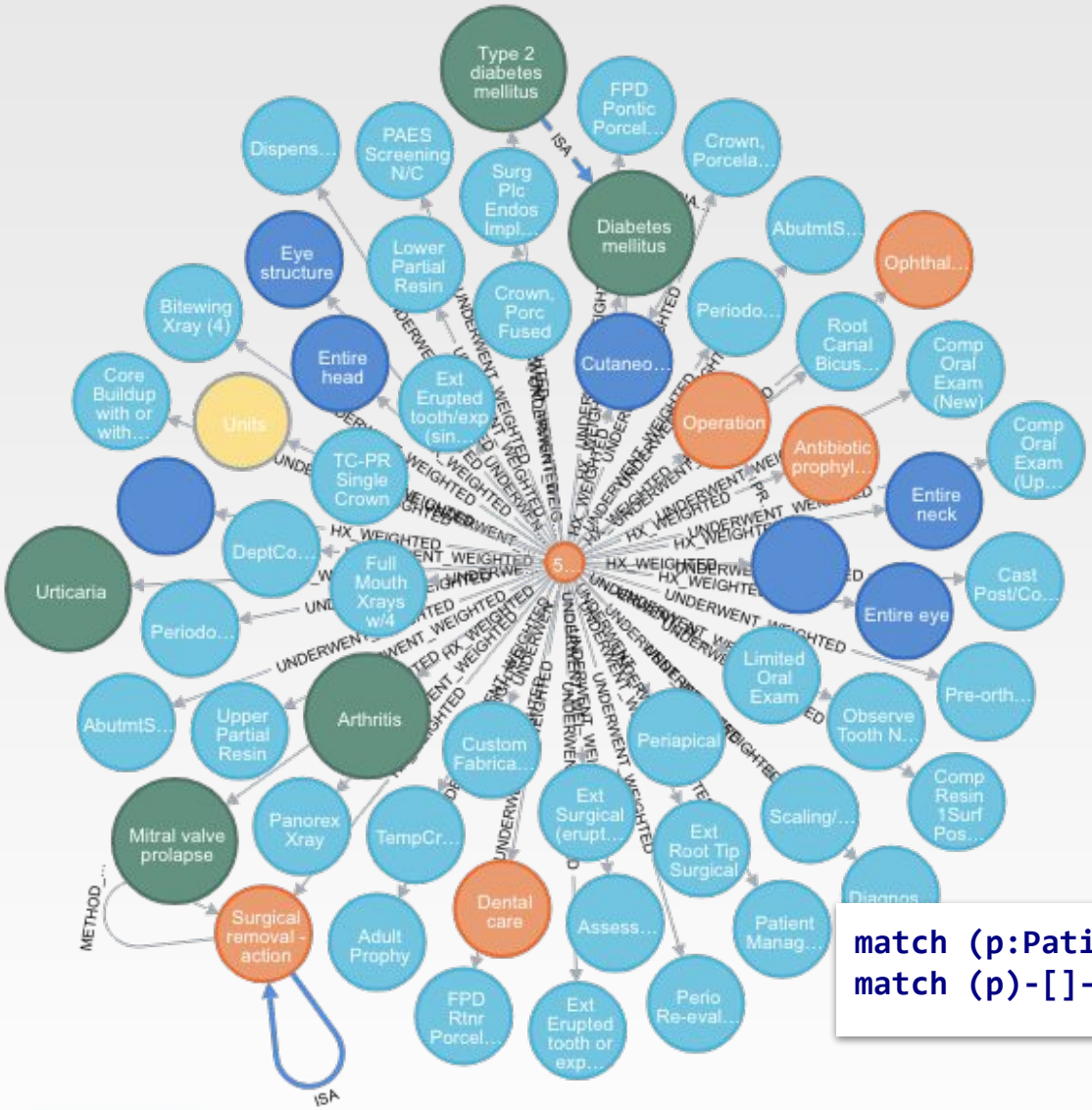
How do groups of nodes tend to clump together?



The Light Dawneth - Ontologies ARE Graphs



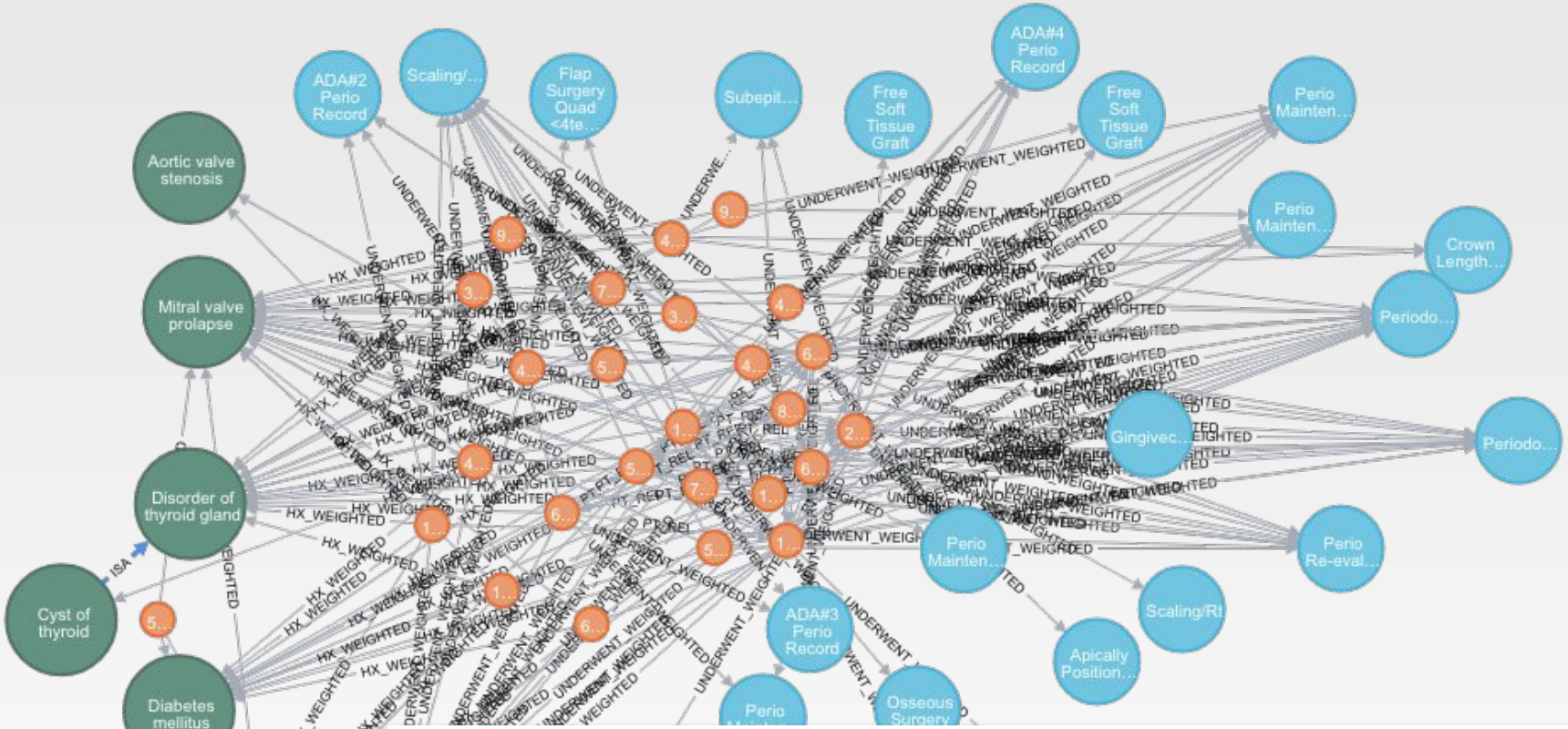

```
match (n)-[:MAY_TREAT]-(o)
return n, o;
```



A Patient's Relationships to:

- Conditions
- Medications
- Body Systems
- Procedures

```
match (p:Patient{patient:'5551212'})-[]-(c:Concept)
match (p)-[]-(pr:Procedure)return p, c, pr
```



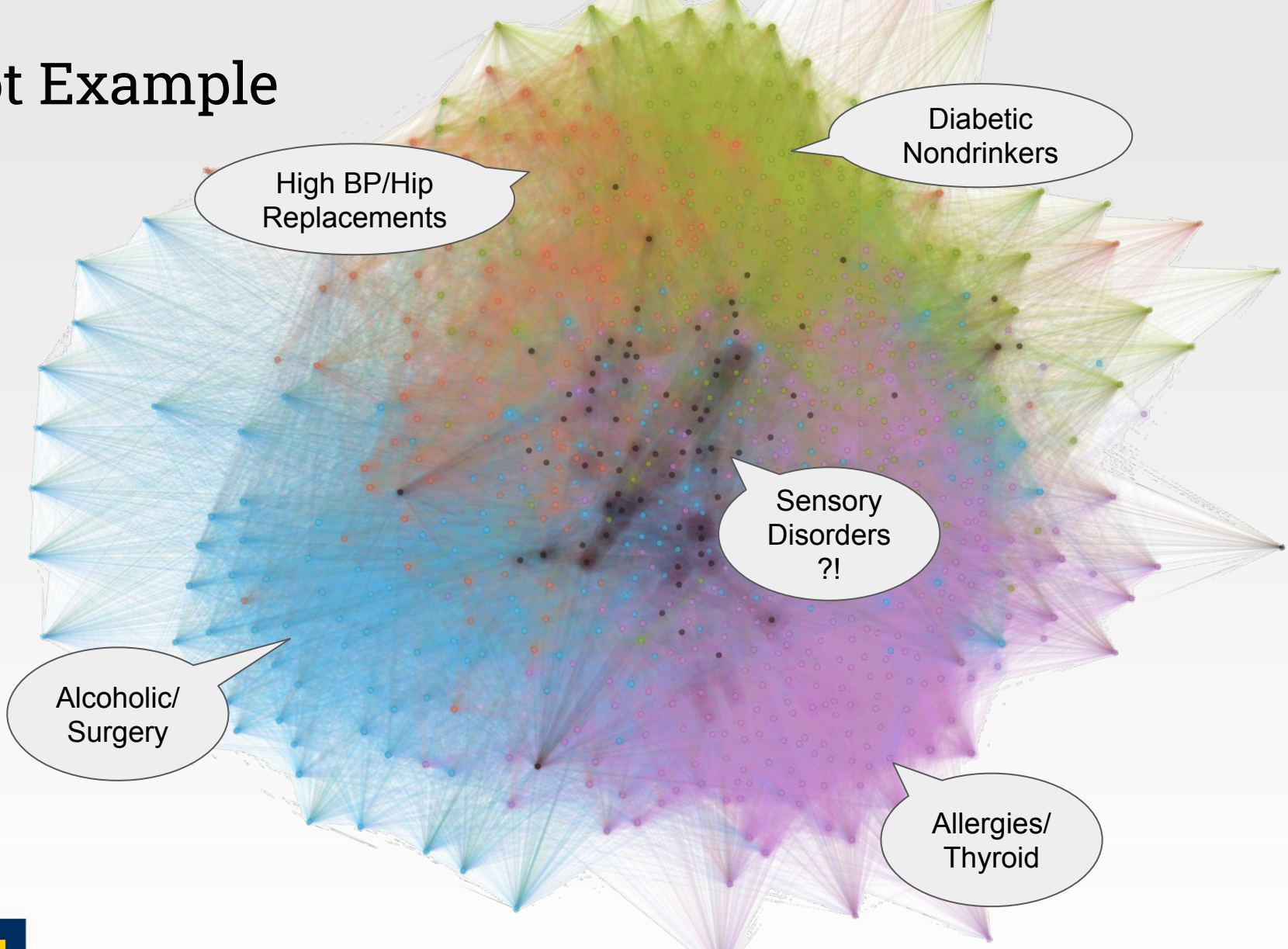

```
match path = (c1:Concept)<-[:ISA*0..2]-(c2)
where tolower(c1.str) contains 'heart valve disorder'
match path2 = (c3:Concept)<-[:ISA*0..2]-(c4)
where tolower(c3.str) contains 'disorder of endocrine system'
with c2, c4
match (pr:Procedure)-[:UNDERWENT_WEIGHTED]-(p:Patient)-[:HX_WEIGHTED]-(c2)
where pr.procedure starts with 'D4'
match (p)-[:HX_WEIGHTED]-(c4)
return p, pr, c2, c4;
```

Well, so what?

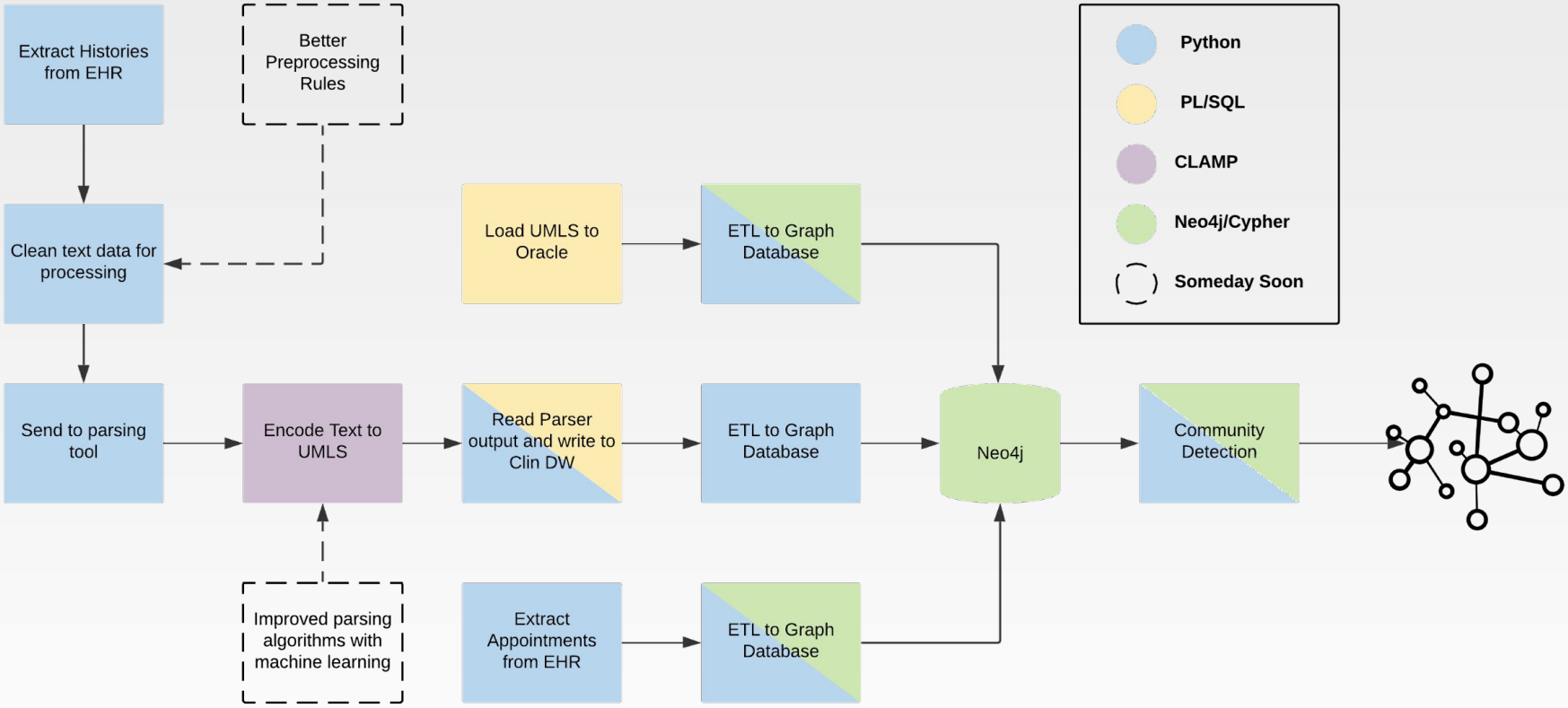
Now we can:

- Identify groups of patients by category of disorder
- Identify patterns of appointment attendance with health status in mind
- Explore relationships between oral and overall health
- Look at patient flow through clinics
- Model communities of patients within the population
- Reconcile patient medications with Michigan Medicine
- Analyze patterns of condition recurrence
- Try to better understand patient's health with sparse data
- Identify special needs patients based on consistent criteria

Pilot Example



Overall Processing Flow



Ways to Improve

- **Collaborate** with clinicians to **narrow the concept set**
- **Improve Parsing:** Build machine learning algorithms to help find common misspellings and drop unhelpful concepts
- **Structure Data for Sharing** with Michigan Medicine
- Align with **standards-based observational data sets** (OHDSI)
- **Quantify Data Quality** for researchers
- **Integrate** NLP-discovered concepts in our clinical **data mart**
- **Build self-serve** tools
- Formalize **security and data access rules**
- Build **de-identified** data set
- Find or build cloud-based **HIPAA-Safe Graph Database**

Tools

NLP

- Regular Expressions - http://regex101.com
- UMLS (also MetaMap)- https://www.nlm.nih.gov/research/umls/index.html
- CLAMP - https://clamp.uth.edu
- MedEx - https://sbmi.uth.edu/ccb/resources/medex.htm

Python

- pandas - https://pandas.pydata.org/
- scikit-learn - https://scikit-learn.org/
- Natural Language Toolkit (nltk) - https://www.nltk.org/
- pymetamap - https://github.com/AnthonyMRios/pymetamap
- igraph - https://igraph.org/python/
- NetworkX - https://networkx.github.io/
- Py2neo - https://py2neo.org/

Neo4j - https://neo4j.com/

Gephi - https://gephi.org/

Questions?

Kate Weber

kgweber@umich.edu

Thanks To:

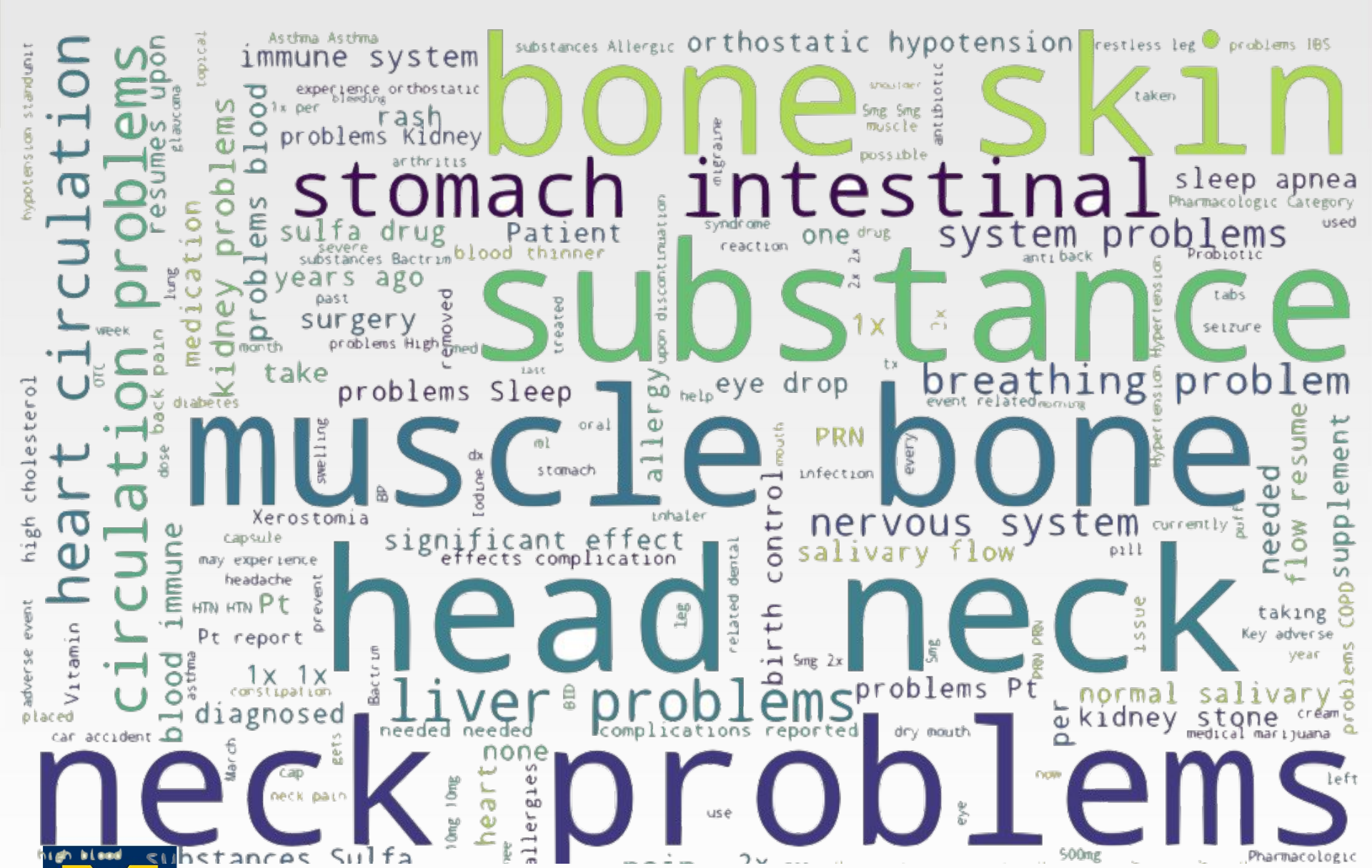
V. G. Vinod Vydiswaran, Assistant Professor of Learning Health Sciences, Medical School and Assistant Professor of Information, School of Information vgvinodv@umich.edu

UM School of Dentistry Learning Health Systems Task Force

https://tbgraph.wordpress.com/2017/11/27/neo4j-graph-algorithms-projecting-a-virtual-graph/

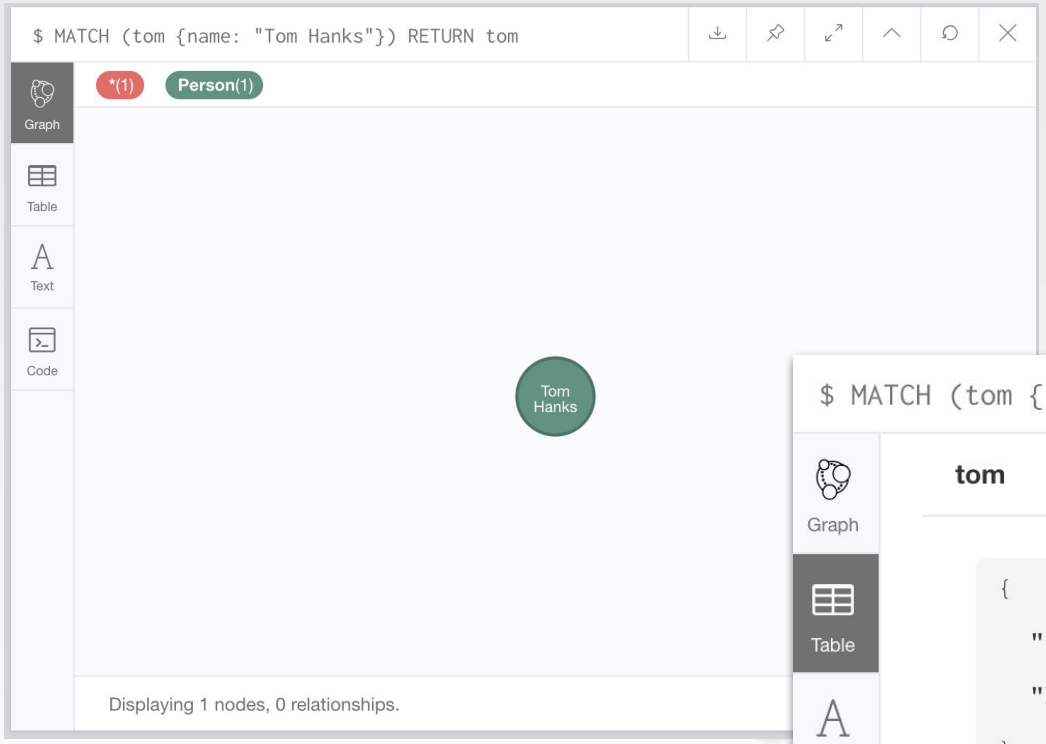
Icons

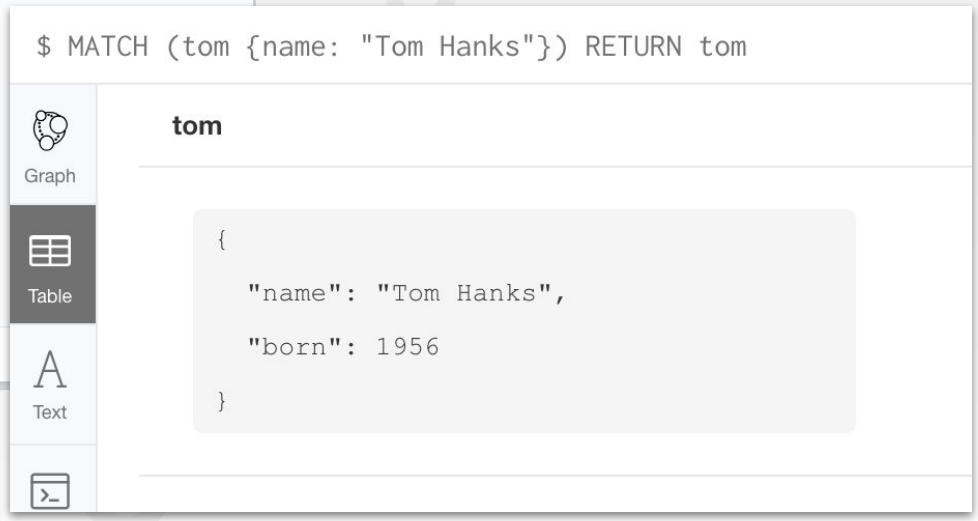
- Network by Meaghan Hendricks from the Noun Project
- Machine Learning by Angela from the Noun Project
- Visual Hierarchy by Eynav Raphael from the Noun Project
- Natural Language Process by Trevor Dsouza from the Noun Project


bone skin
substance
muscle bone
head neck
neck problems
stomach intestinal
heart circulation problems
circulation problems
immune system
orthostatic hypotension
sleep apnea
system problems
breathing problem
nervous system
liver problems
diagnosed
kidney stone
normal salivary
significant effect
salivary flow
birth control
allergy
surgery
rash
problems Kidney
sulfa drug
Patient
years ago
take
problems Sleep
eye drop
PRN
Xerostomia
blood immune
kidney problems
medication
resumes upon
problems blood
high cholesterol
Vitamin
Pt report
1x 1x
needed needed
complications reported
dry mouth
heart none
Substances Sulfa
Michigan IT Symposium
SPARKING CONNECTIONS & IGNITING IDEAS ACROSS U-M
UNIVERSITY OF MICHIGAN

Demo Backup

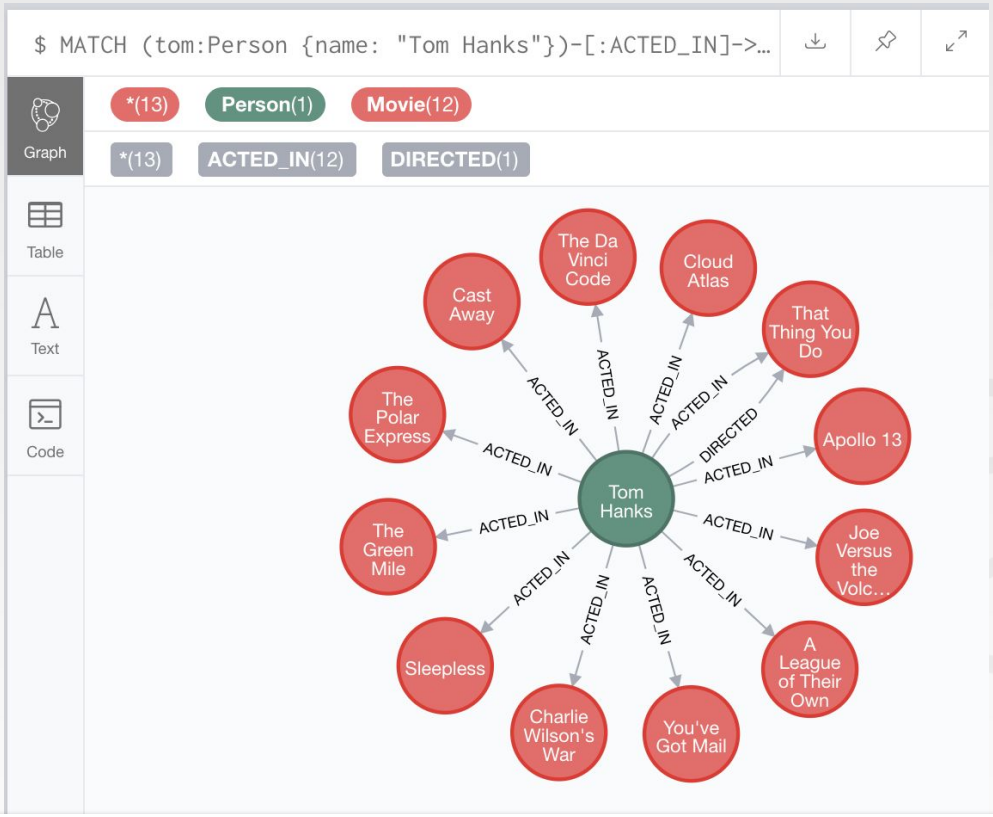
Demo





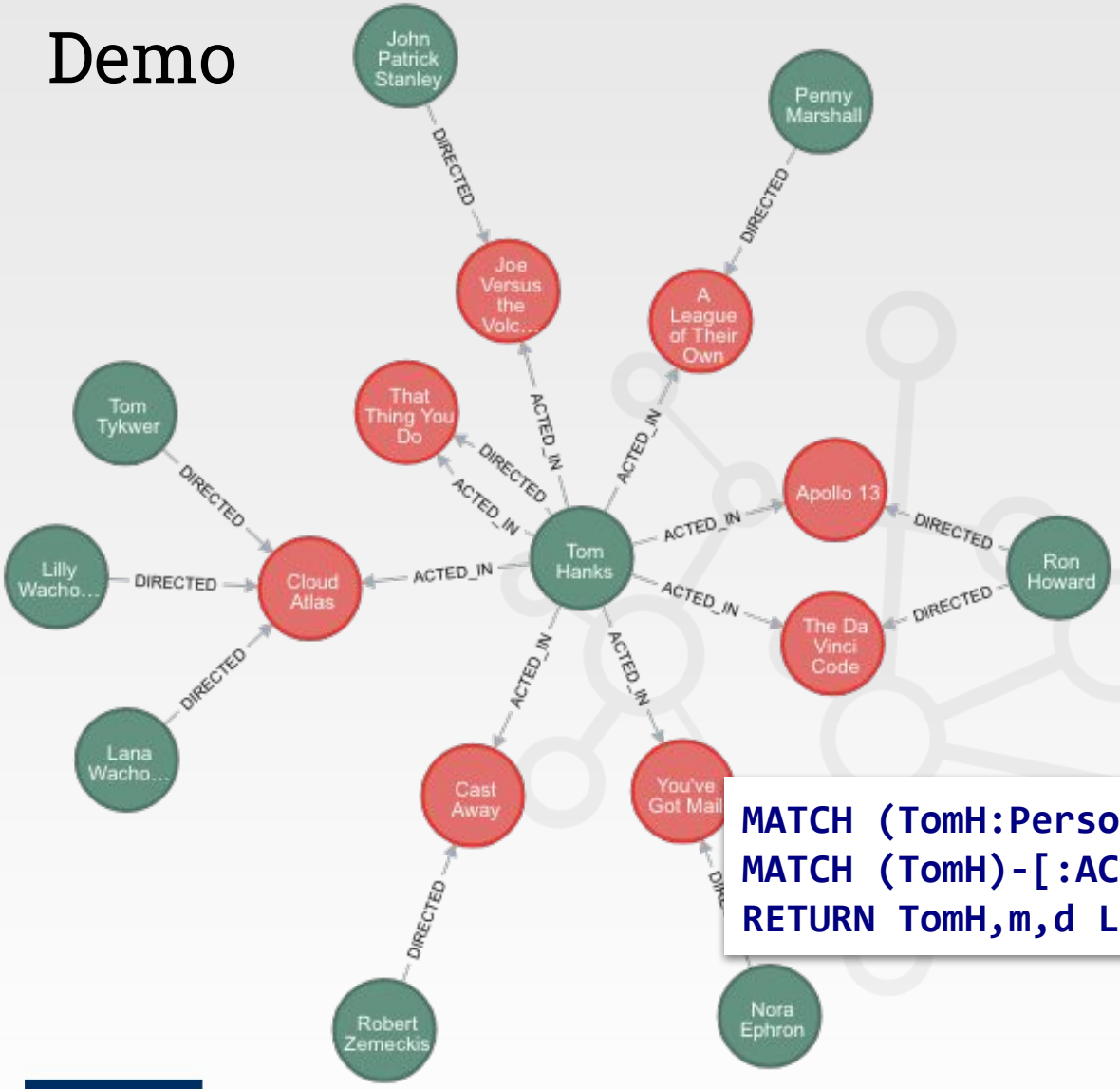

```
MATCH (tom {name: “Tom Hanks”})
RETURN tom
```

Demo



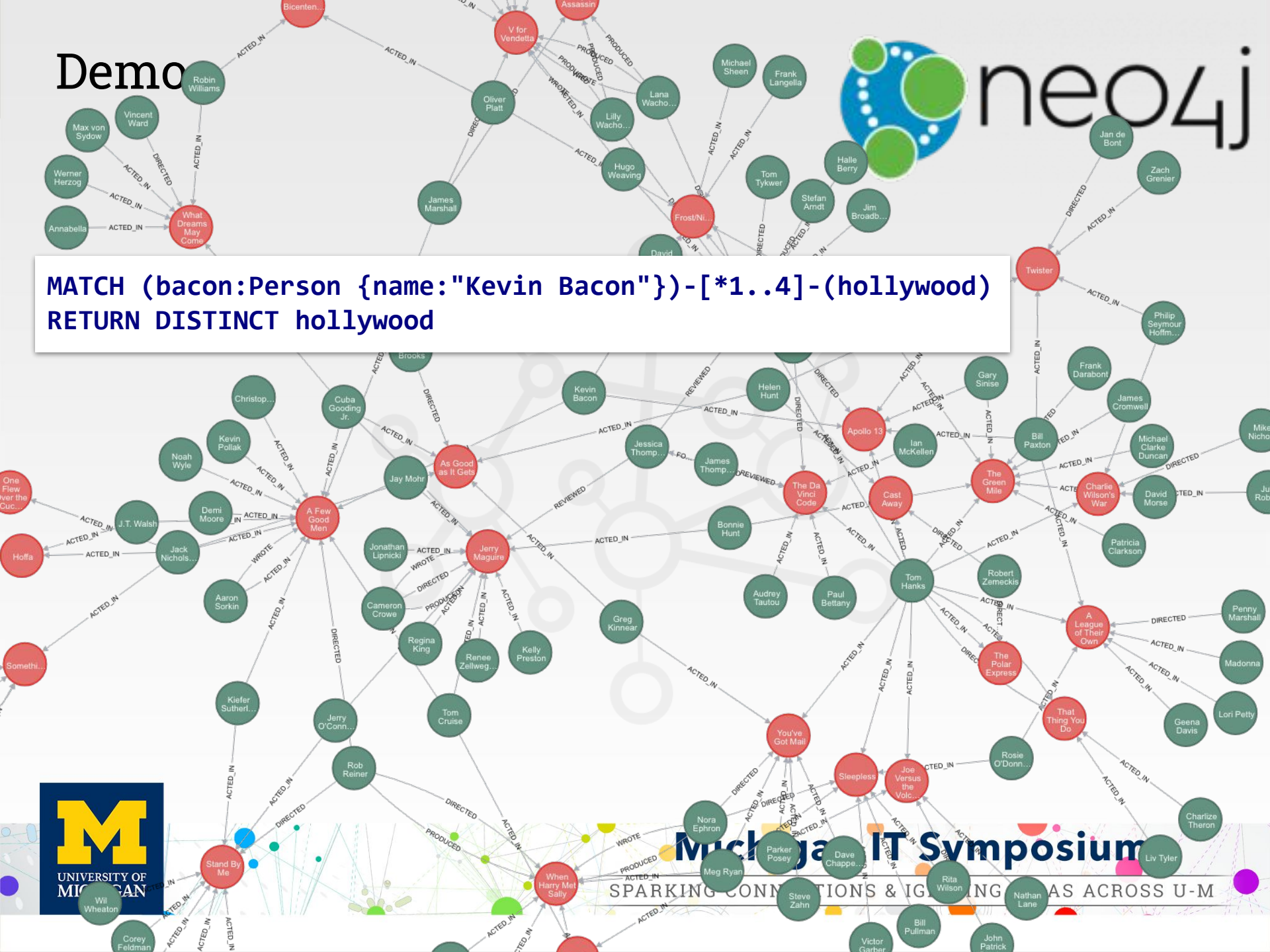

```
MATCH (tom:Person {name: "Tom Hanks"})-[:ACTED_IN]->(tomHanksMovies)
RETURN tom,tomHanksMovies
```

Demo



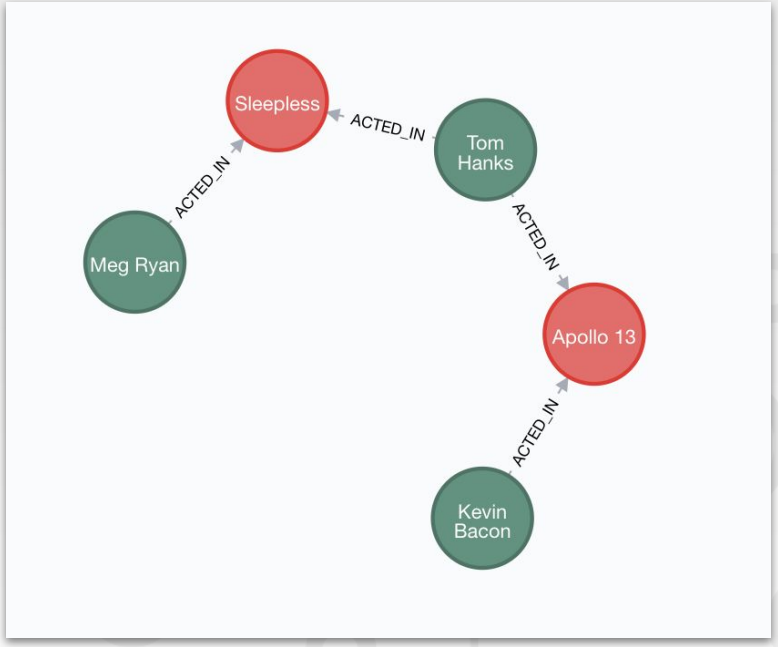

```
MATCH (TomH:Person{name:'Tom Hanks'})
MATCH (TomH)-[:ACTED_IN]->(m)<-[:DIRECTED]-(d)
RETURN TomH,m,d LIMIT 10
```

Demo



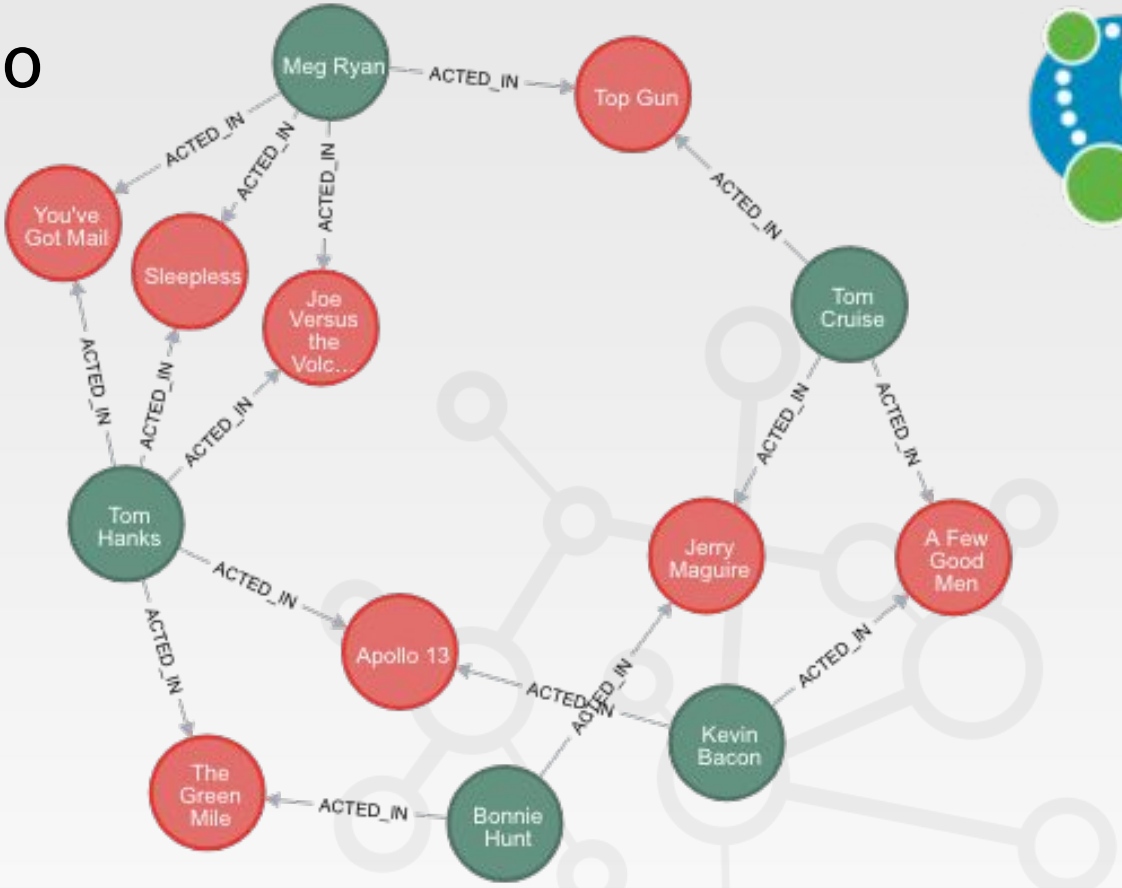
```
MATCH (bacon:Person {name:"Kevin Bacon"})-[*1..4]-(hollywood)
RETURN DISTINCT hollywood
```

Demo




```
MATCH p=shortestPath(
  (bacon:Person {name:"Kevin Bacon"})-[*]-(meg:Person {name:"Meg Ryan"}))
RETURN p
```

Demo




```
MATCH
 (tom:Person {name:"Tom Hanks"})-[:ACTED_IN]->(m)<-[:ACTED_IN]-(coActors),
 (coActors)-[:ACTED_IN]->(m2)<-[:ACTED_IN]-(cruise:Person {name:"Tom Cruise"})
RETURN tom, m, coActors, m2, cruise
```