

Rare Disease Diagnostic Support Through Network Analysis

Angel Nugroho, Dac-Trung Nguyen, Dr. Eric Sid



NIH
National Center
for Advancing
Translational Sciences

The Problem Space

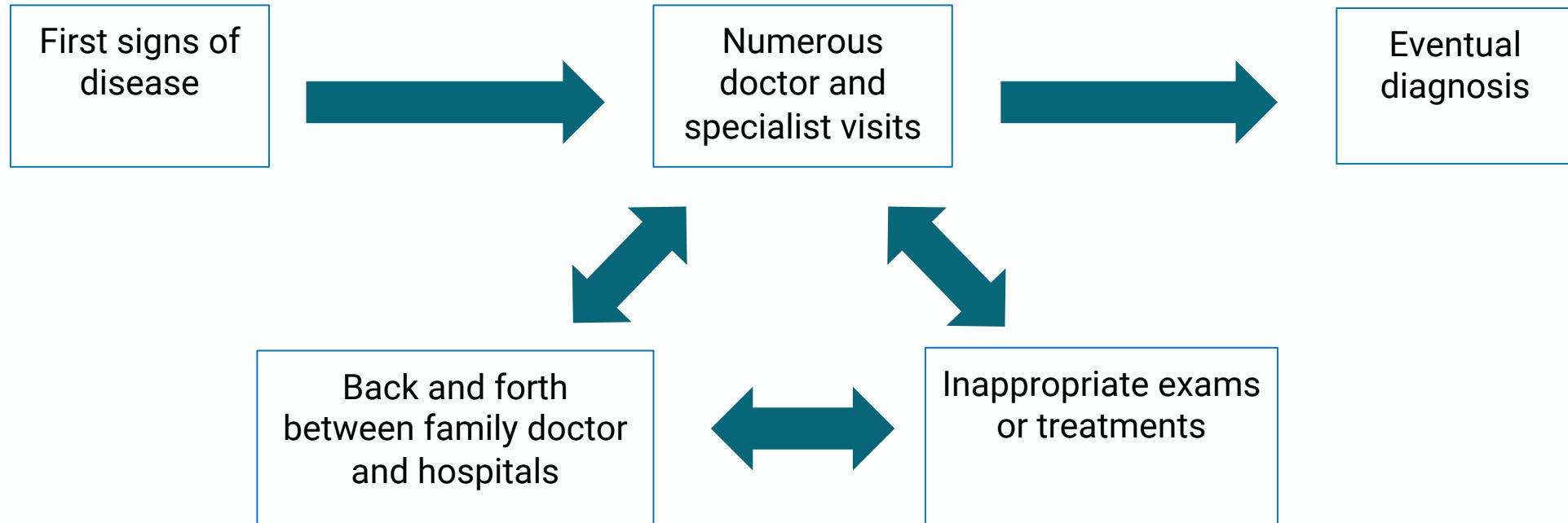
- Rare diseases are defined as conditions that affect fewer than 200,000 people in the United States
- Around ~7000 rare diseases
- Affecting 30 million people
- Extremely limited data



National Center
for Advancing
Translational Sciences

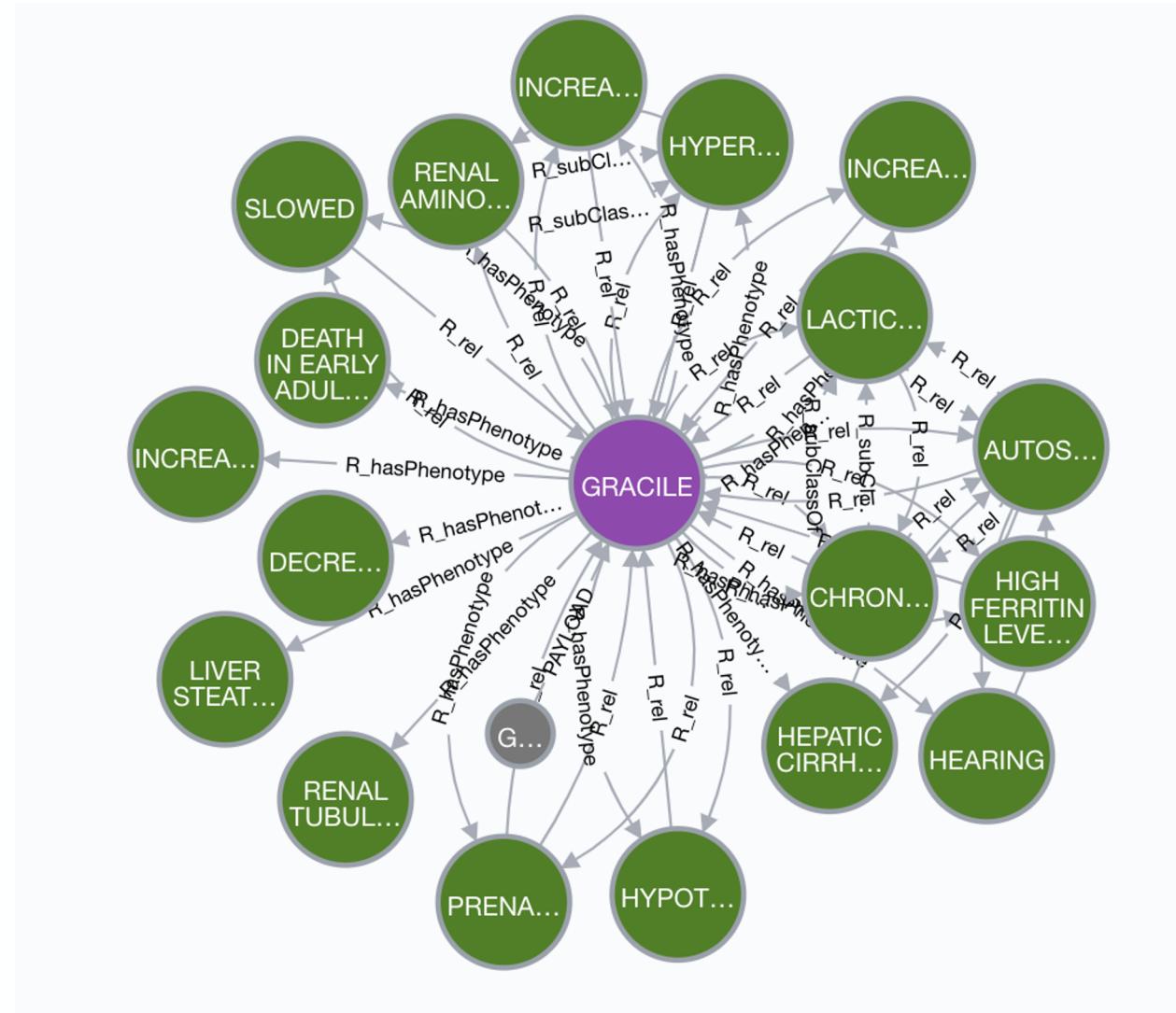
The Diagnostic Odyssey

“When you hear hoofbeats behind you, don't expect to see a zebra.”



Neo4j Graph Database

- Rare diseases and associations represented in a network
- Uses GARD, HP, Orphanet and other ontologies



Creating a Disease Ranking System

Process

- Based on phenotypes given, output matching diseases
- Assign values to disease and phenotype nodes
 - Page Rank algorithm
- Values of matching phenotypes help rank diseases over each other
 - TF-IDF

Implementation

- Neo4j
 - Cypher queries
- Jupyter notebooks
 - pandas
 - networkx
 - numpy
- Django



NIH
National Center
for Advancing
Translational Sciences

Website Demo: Zebra Rank



National Center
for Advancing
Translational Sciences

What comes next?

- More than data: What can a potential rare disease patient do with this?
 - Providing actionable steps
- Future network analysis:
 - Group related diseases
 - Other nodes, like affected gene
 - Creating a knowledge graph



NIH
National Center
for Advancing
Translational Sciences