

# Using UMLS Metathesaurus Relations for Managing Biomedical Knowledge from MEDLINE Citations

Liqin Wang

August 9, 2012

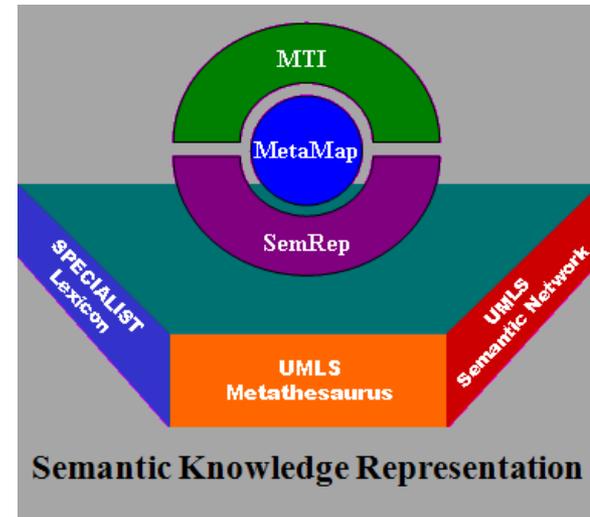


The University of Utah  
School of Medicine

# Knowledge Resources and Tools



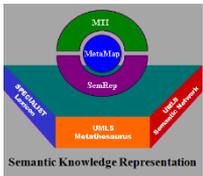
**Unified Medical  
Language System**®





# Metathesaurus

- 169 source terminologies (UMLS 2012AA)
  - SNOMED CT, ICD, RxNorm, Gene Ontology, etc.
  - Over 2.7 million concepts
- Definitional Knowledge
  - Congestive heart failure ISA Disorder of cardiac ventricle
- Essentially contains hierarchical relations
  - Over 8.7 million relations counted by using CUI in both directions
  - Indicated by: Parent (PAR), Children (CHD), Narrower Than (RN), Broader Than (RB)



# SemRep

- MetaMap
- Semantic Network

antioxidant and neuroprotective capabilities and thus, may contribute to the overt activity of its parent compound, rasagiline. This paper will review the earlier and present studies in the development of rasagiline for treatment of PD and discuss its pharmacology and applicable mechanism of action.

- Output is a semantic predication
  - **rasagiline TREATS parkinson disease**

# Semantic MEDLINE DB

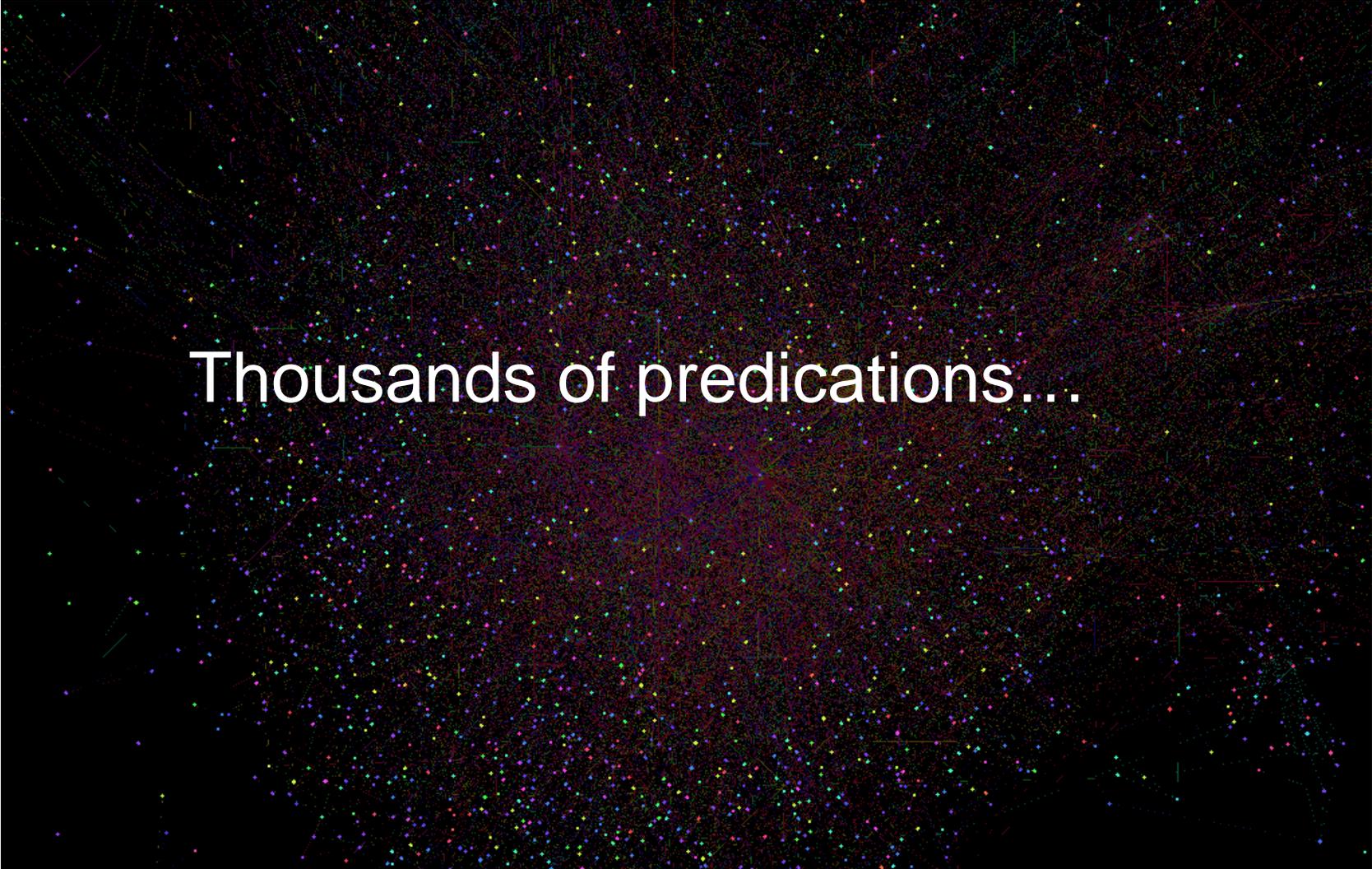
---

- A database containing the output of SemRep after parsing the MEDLINE citations including titles and abstracts
- 57 million Predications
  - Associative predications up to 54 million (93.67%)
- Essentially contains assertional knowledge
  - Enalapril treats Congestive heart failure



# For a simple query...

---



Thousands of predications...

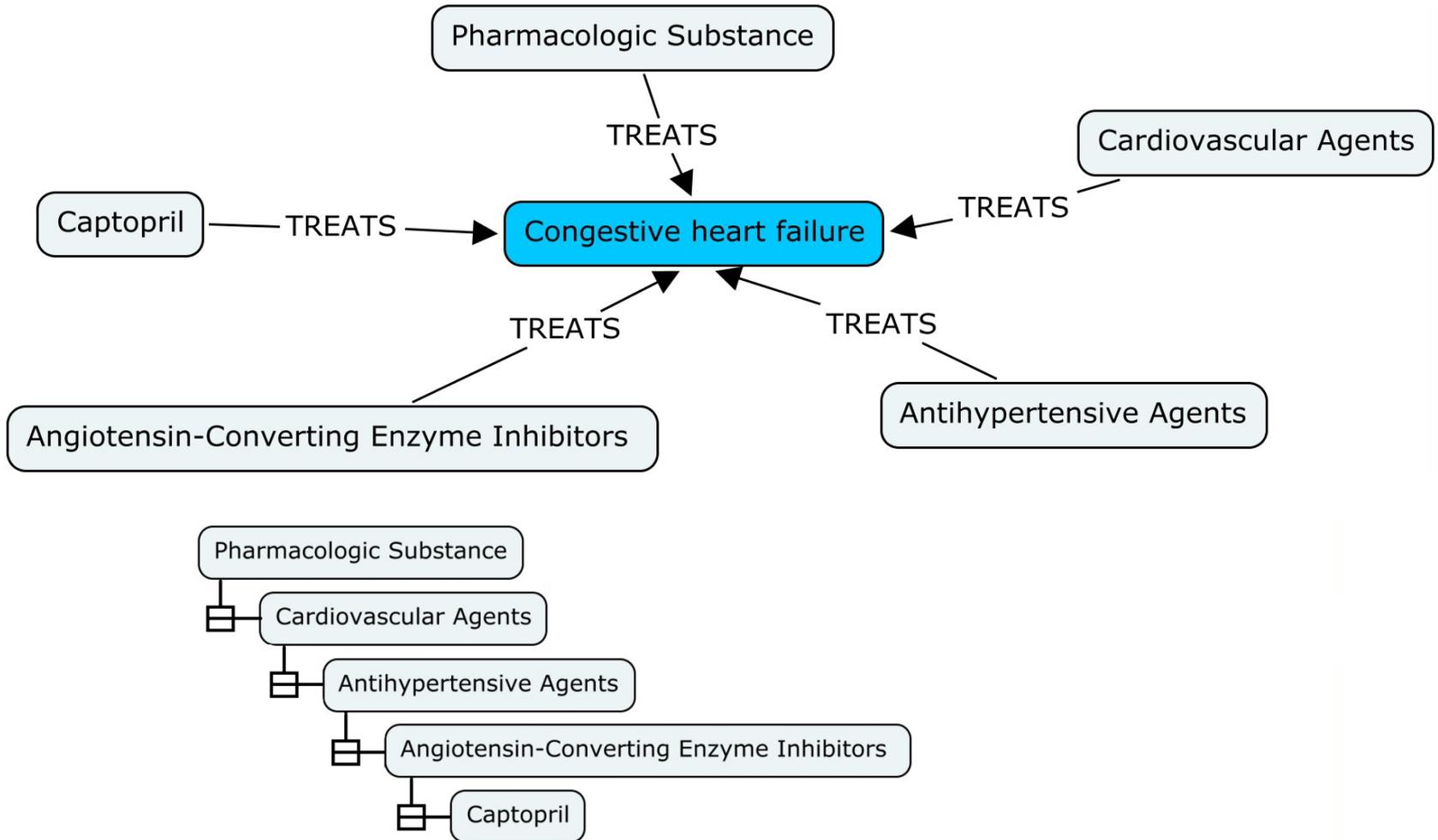


The information needs to be effectively organized and analyzed in order to be useful!



**LEVEL OF GRANULARITY**

# An Example



Understand the level of granularity could support an effectively management of biomedical knowledge!

# Related work

---

- Abstraction summarization
  - Relevance, connectivity, novelty, and saliency
  - Manipulate the data only from semantic MEDLINE
  
- Marcelo Fiszman, Thomas C. Rindflesch, Halil Kilicoglu. Abstraction Summarization for Managing the Biomedical Research Literature. *Proceedings of the Workshop on Computational Lexical Semantics*. 2004:76–83.
- Han Zhang, Marcelo Fiszman, Dongwook Shin, Christopher M Miller, Graciela Rosembat, Thomas C Rindflesch. Degree centrality for semantic abstraction summarization of therapeutic studies. *J Biomed Inform*; 2011: 830-838.
- Fiszman M, Demner-Fushman D, Kilicoglu H, Rindflesch TC. Automatic summarization of MEDLINE citations for evidence-based medical treatment: a topic-oriented evaluation. *J Biomed Inform*. 2009 Oct;42(5):801-13. Epub 2008 Nov 5.

# Objectives

---

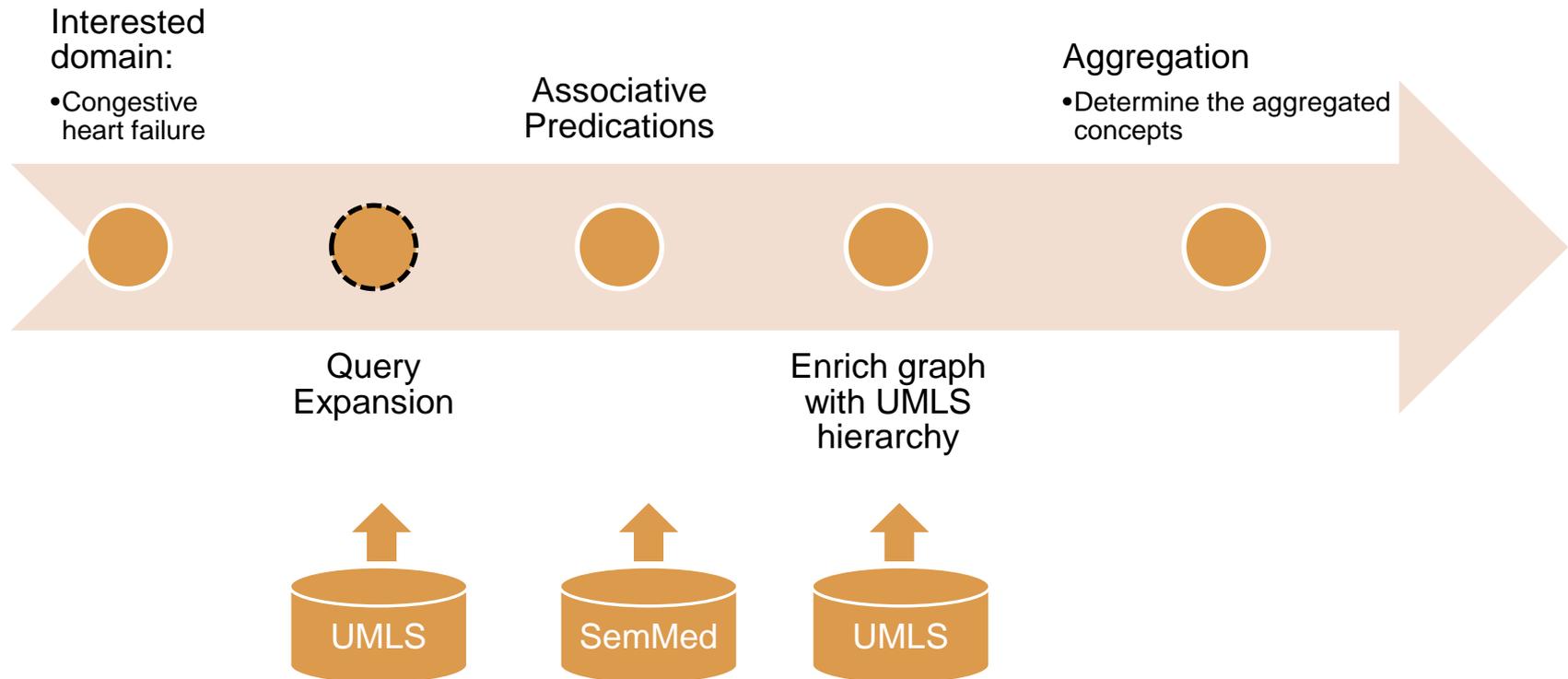
- Use UMLS hierarchical relations to enrich large and unorganized graphs built from semantic predications
- Aggregate information according to the hierarchy for biomedical knowledge management prospects
  - Navigation
  - Knowledge extraction
  - Summarization
  - Pattern recognition

# Congestive heart failure

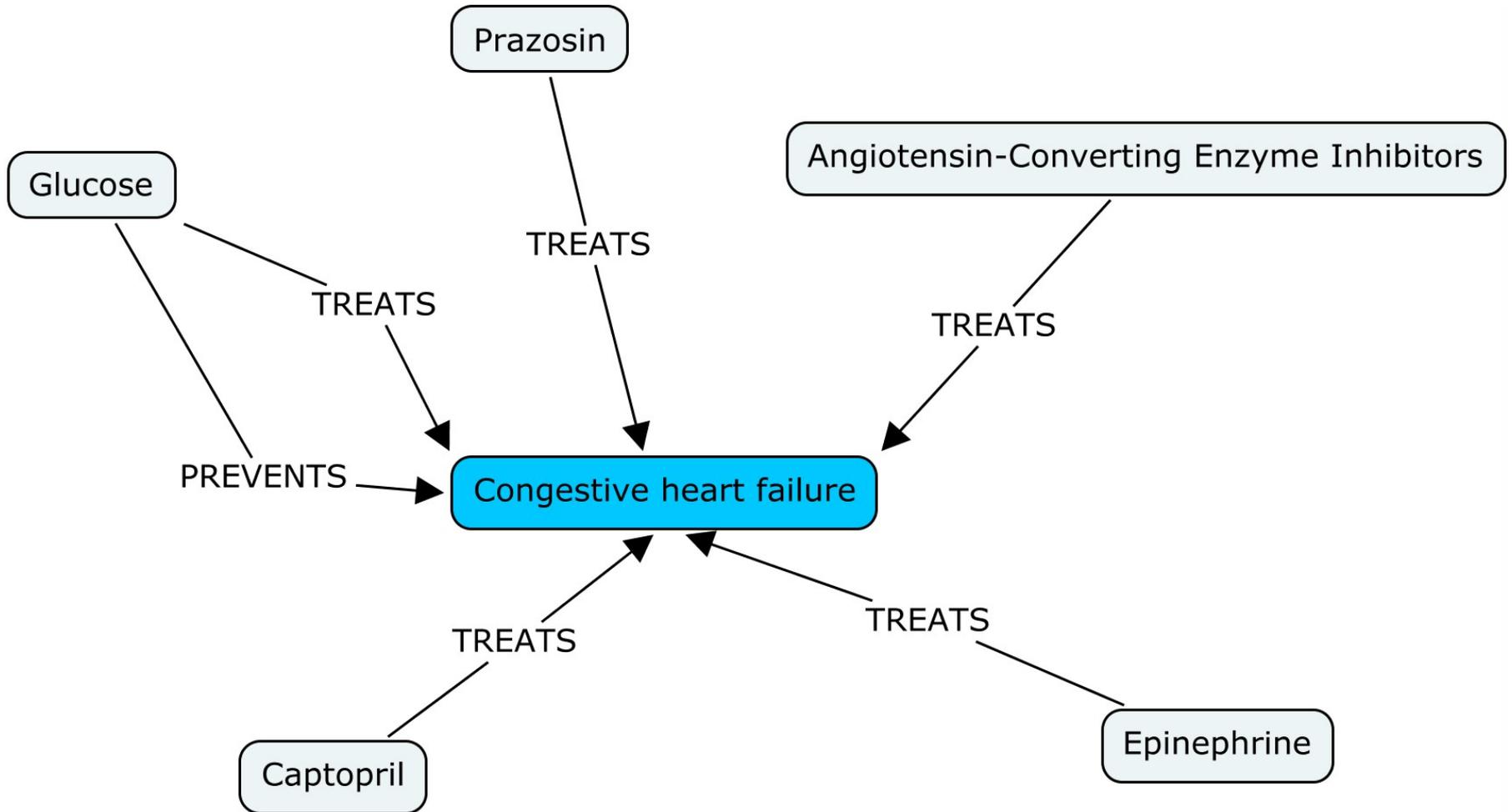
---

- A condition in which the heart can no longer pump enough blood to the rest of the body
- An example syndrome in this study

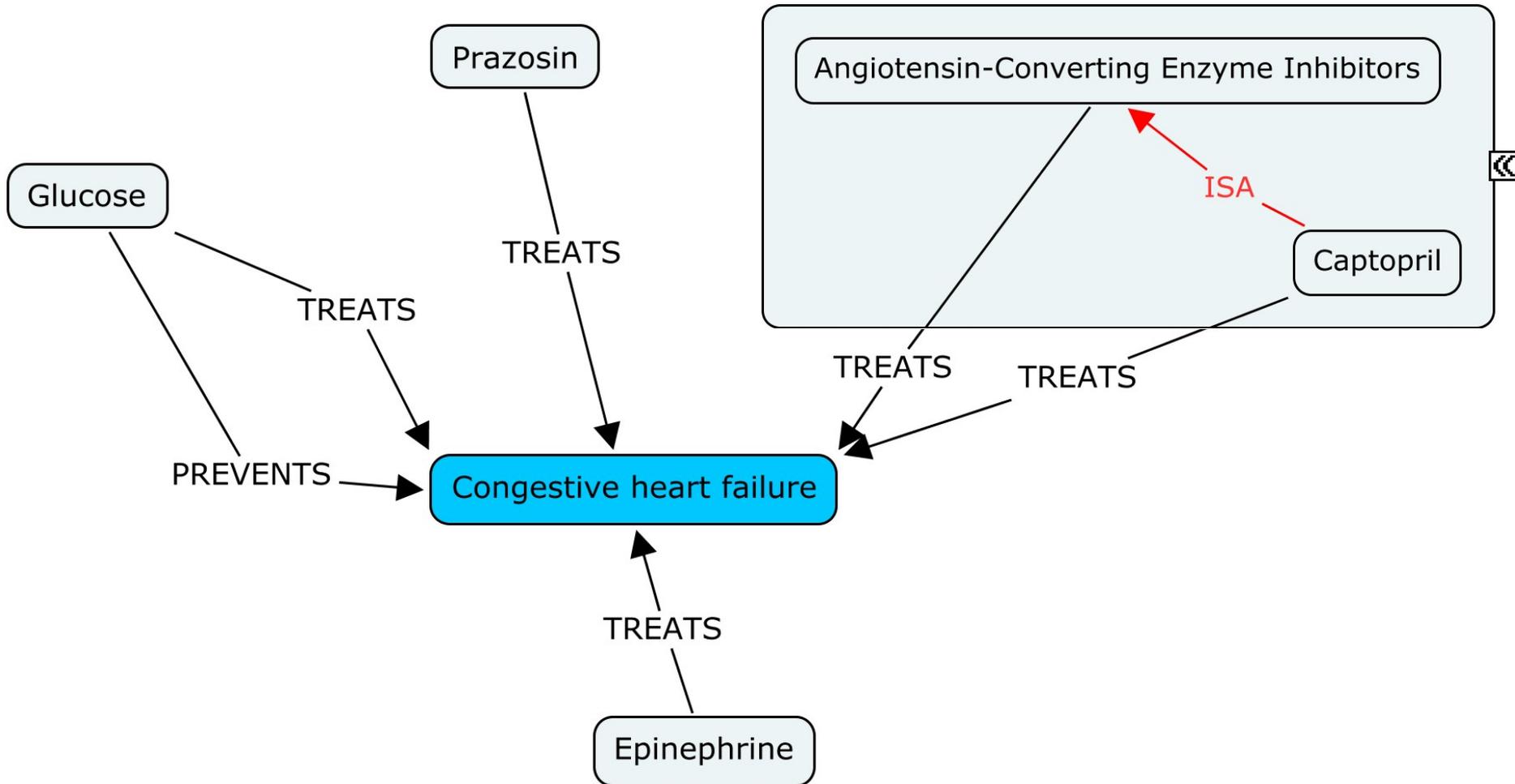
# Experiment Process



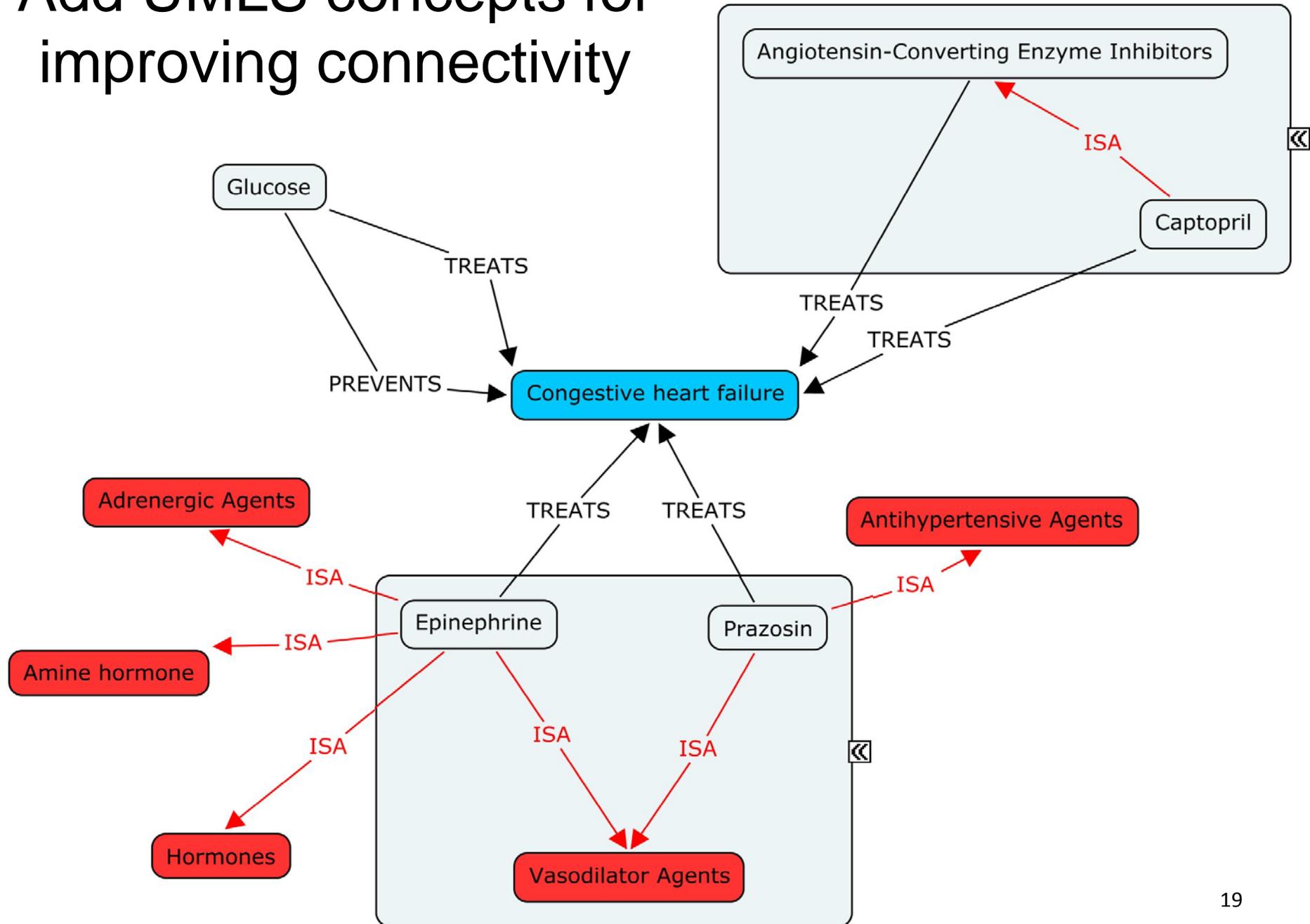
# Original graph



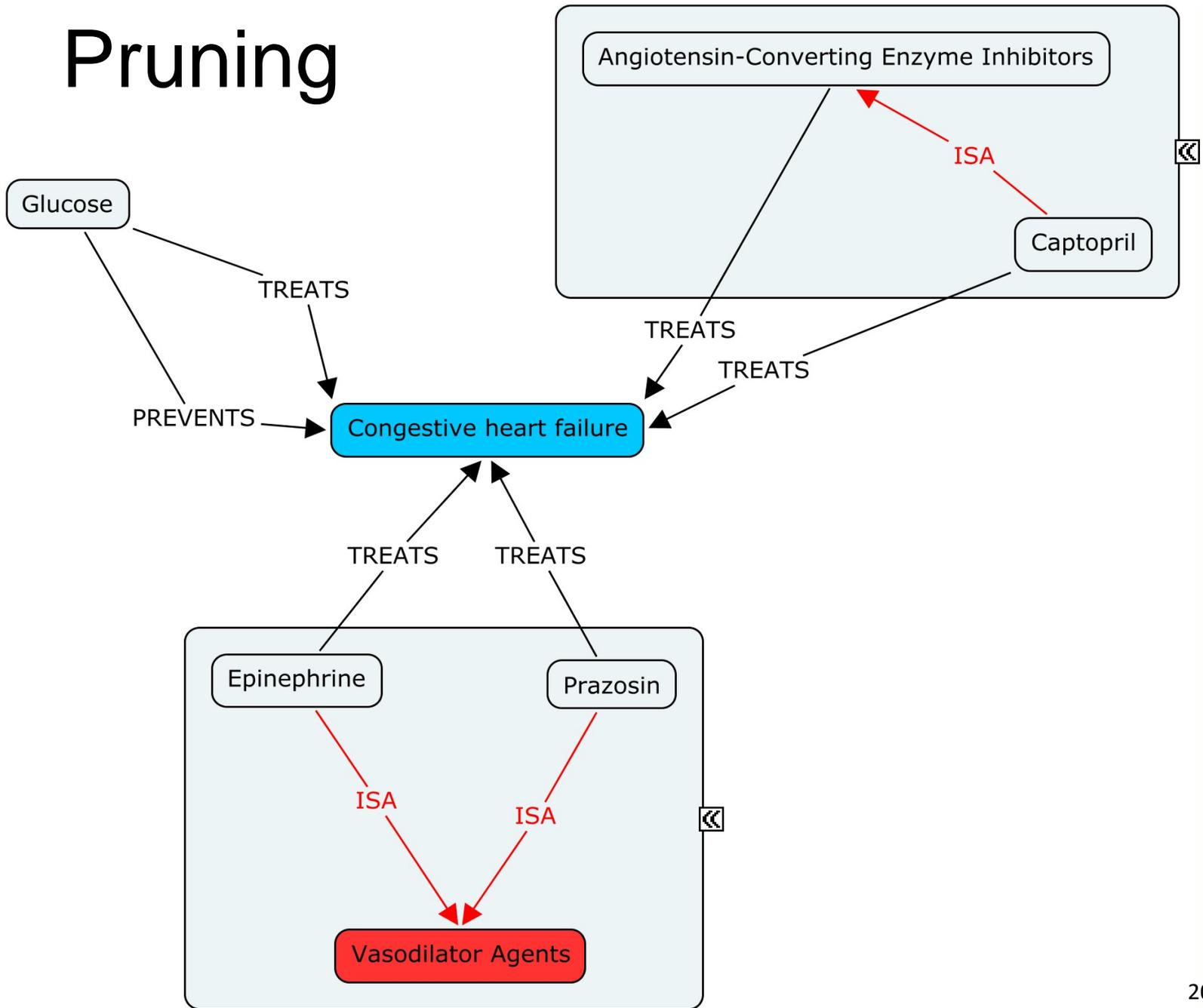
# Adding hierarchy between any two concepts in the graph

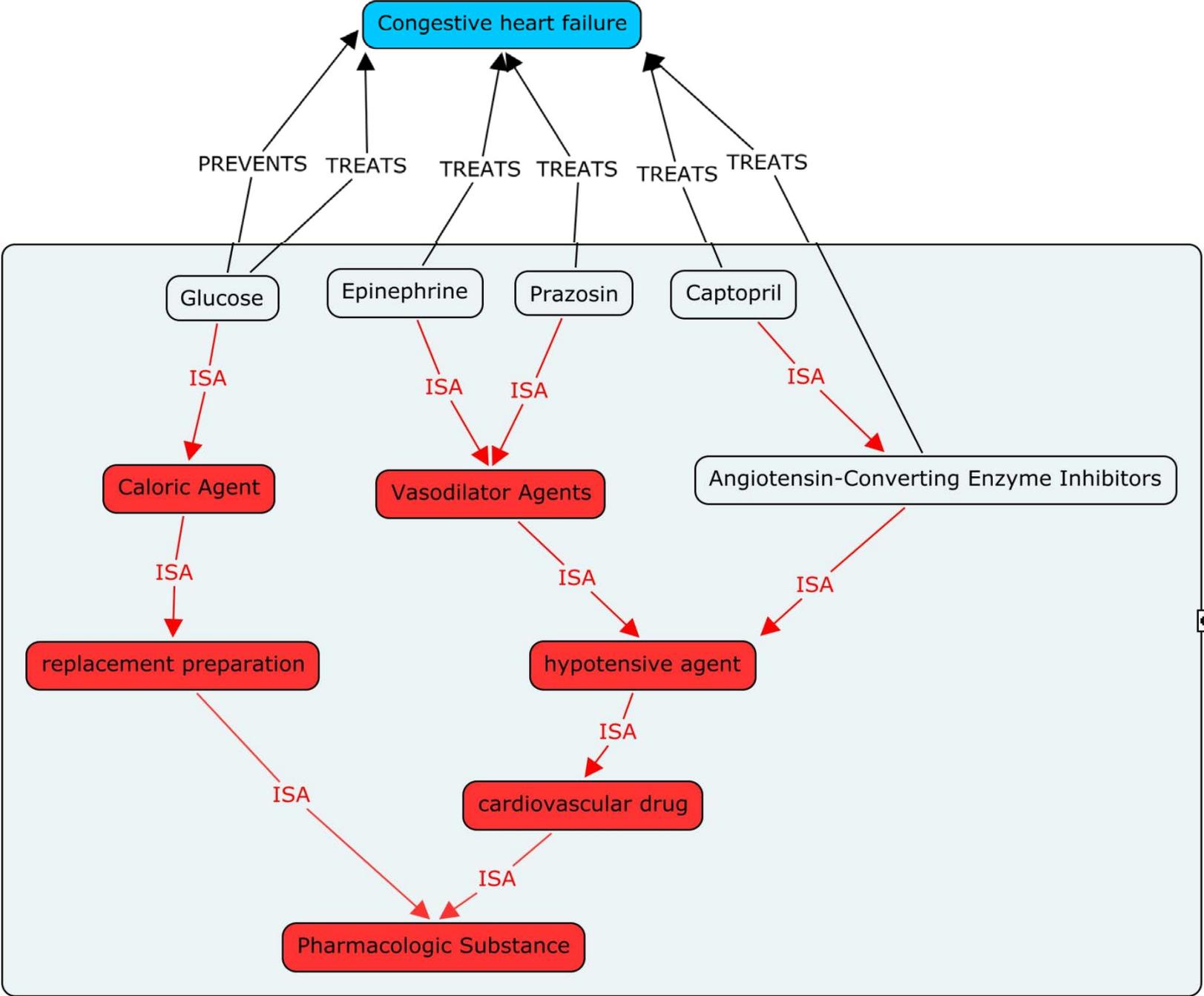


# Add UMLS concepts for improving connectivity

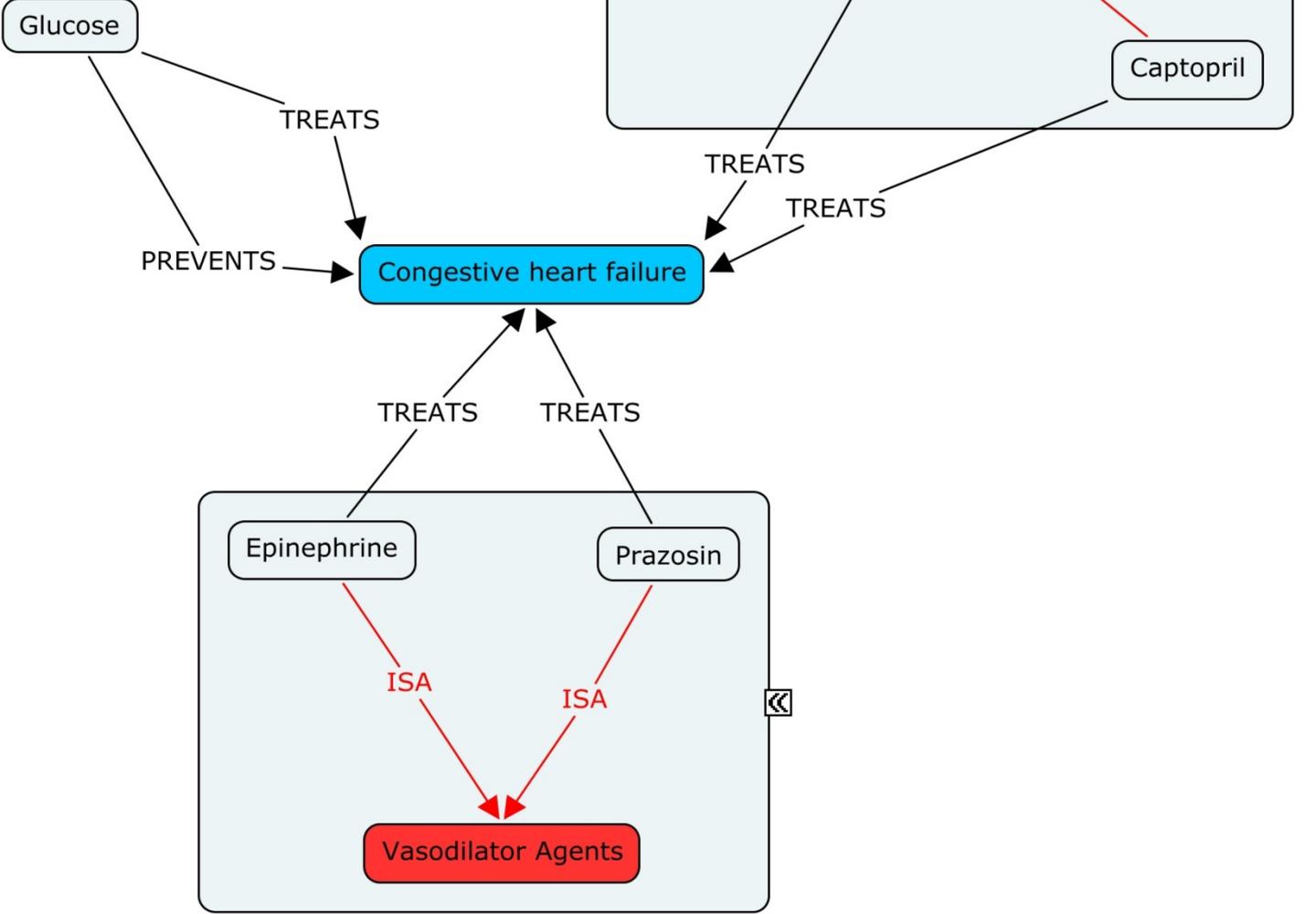


# Pruning

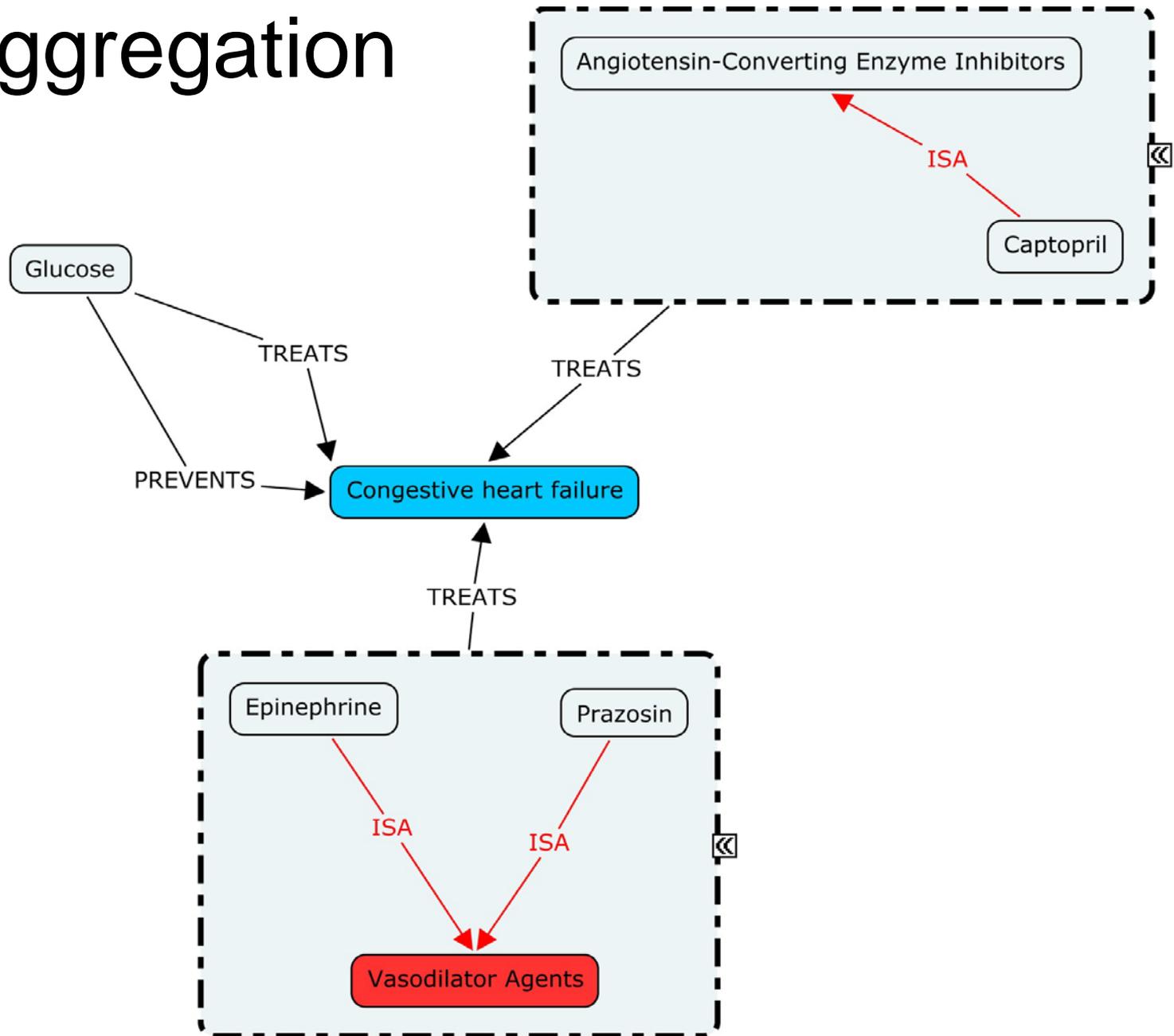




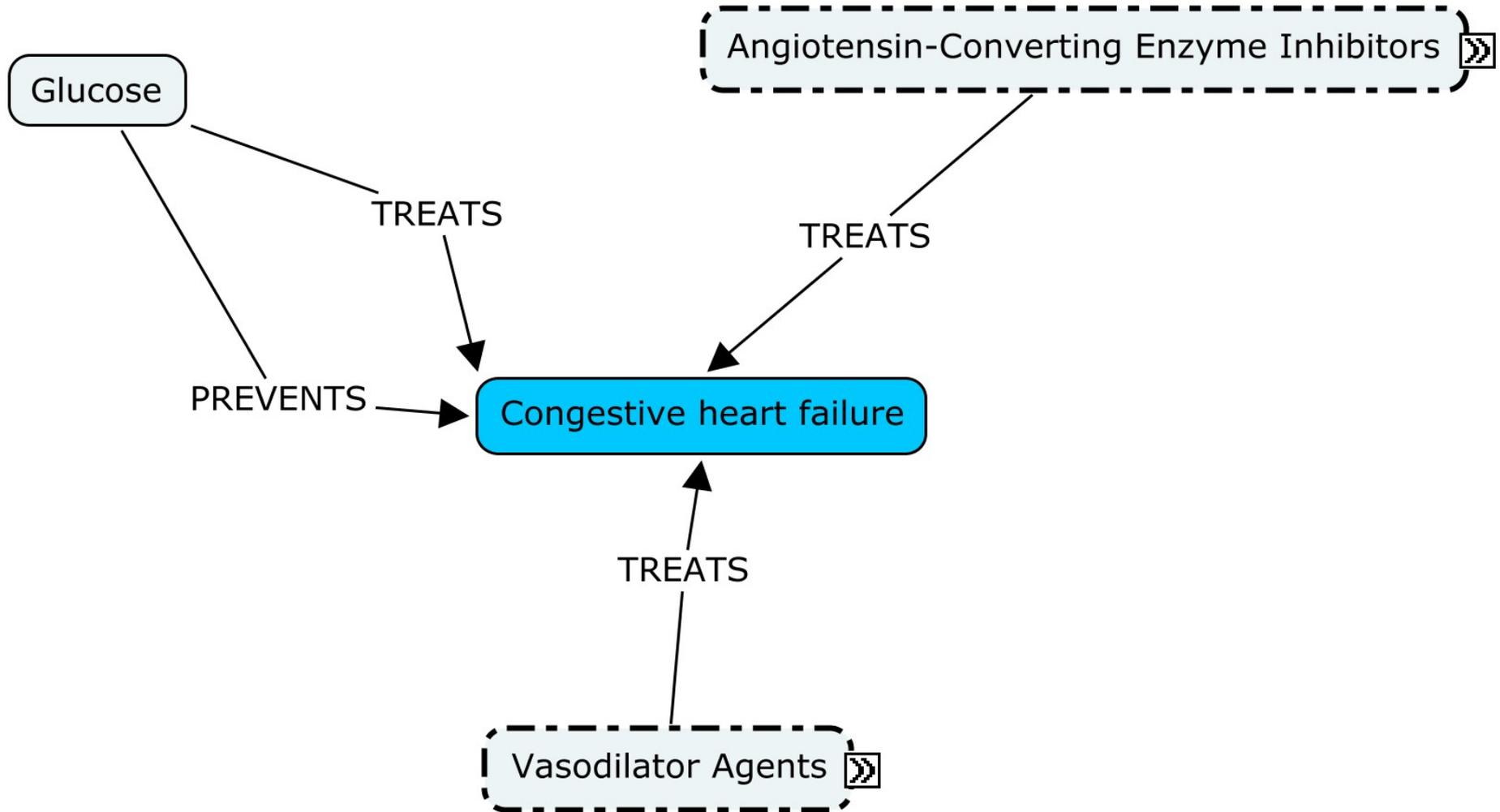
# Break up large cluster



# Aggregation



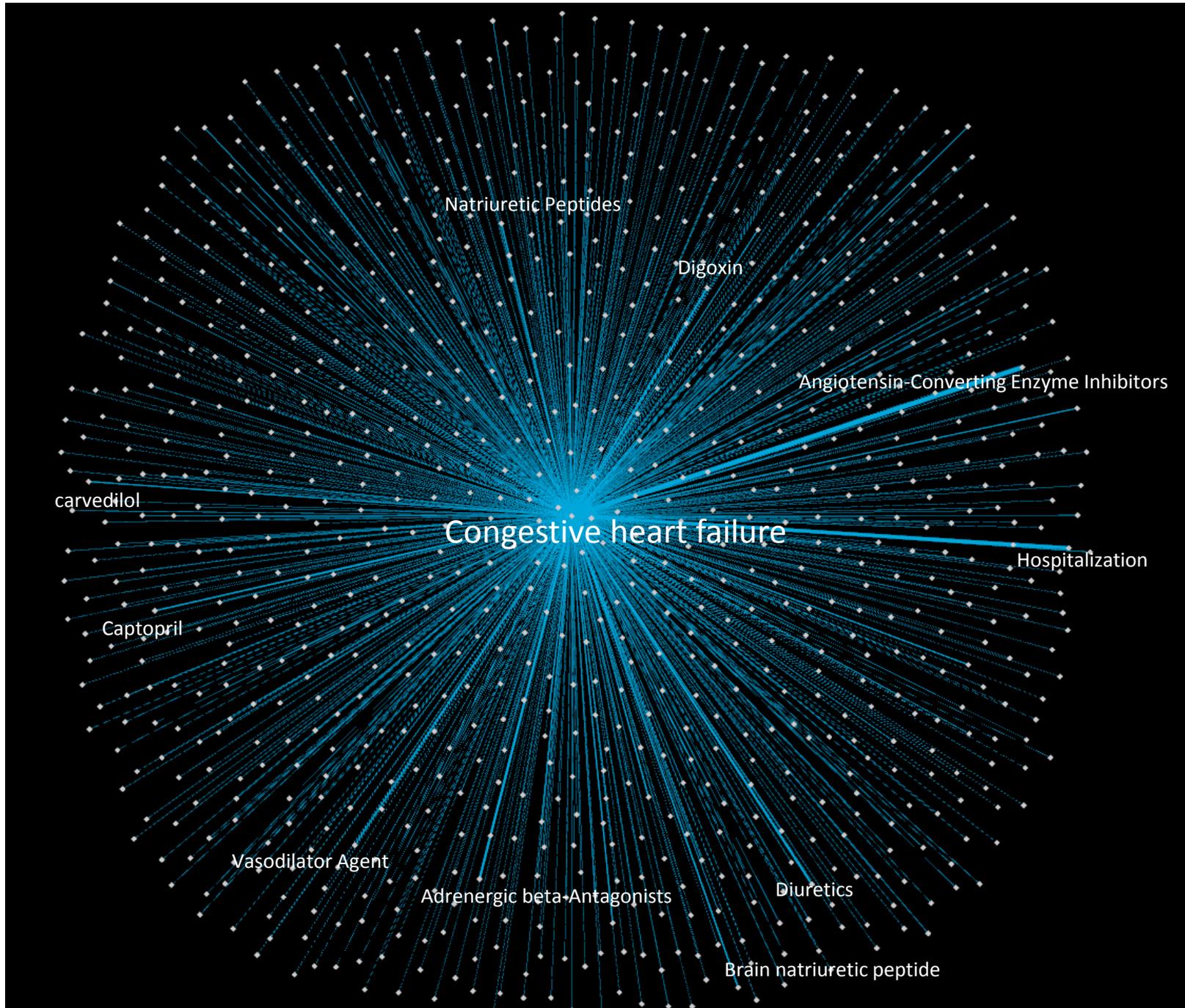
# After aggregation



# Predication Query

---

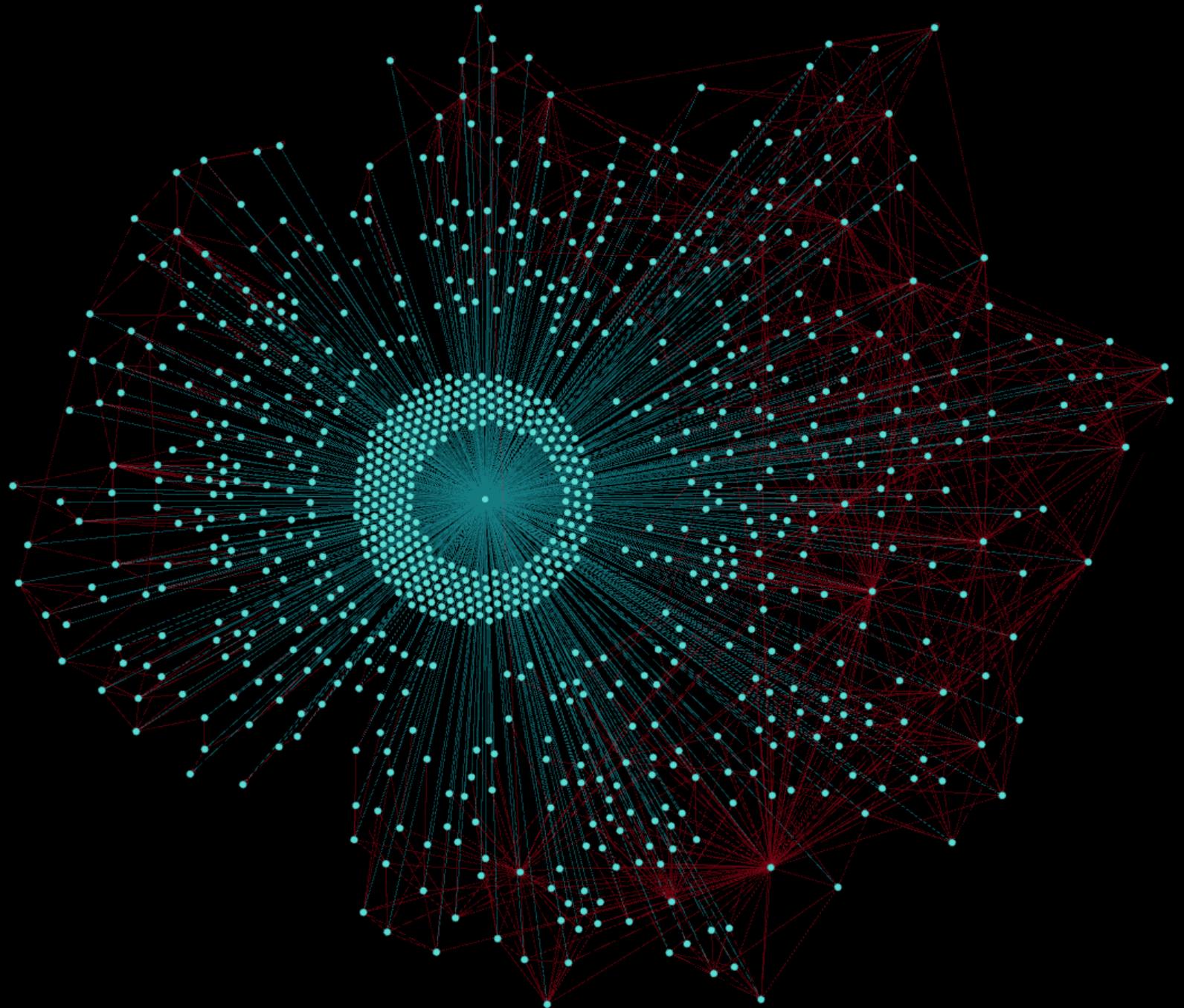
- Given a concept: Congestive Heart Failure (CHF), retrieving associative predications related to CHF
  - Limit the predicate to “TREATS” and “PREVENTS”
  - Remove less informative predications
    - Pharmaceutical Preparations TREATS Congestive heart failure
  - Obtain the frequency of the predication
- Results
  - 971 predications
    - 95.78% are “TREATS” predications
  - 924 concepts directly associated with CHF



# Graph Enrichment

---

- Enrich the graph with UMLS hierarchical relations between any two concepts in the graph
- Results
  - 1060 directly hierarchical relations among 617 (66.7%) concepts
  - 303 (33.33%) concepts only connect to CHF



# Graph Enrichment

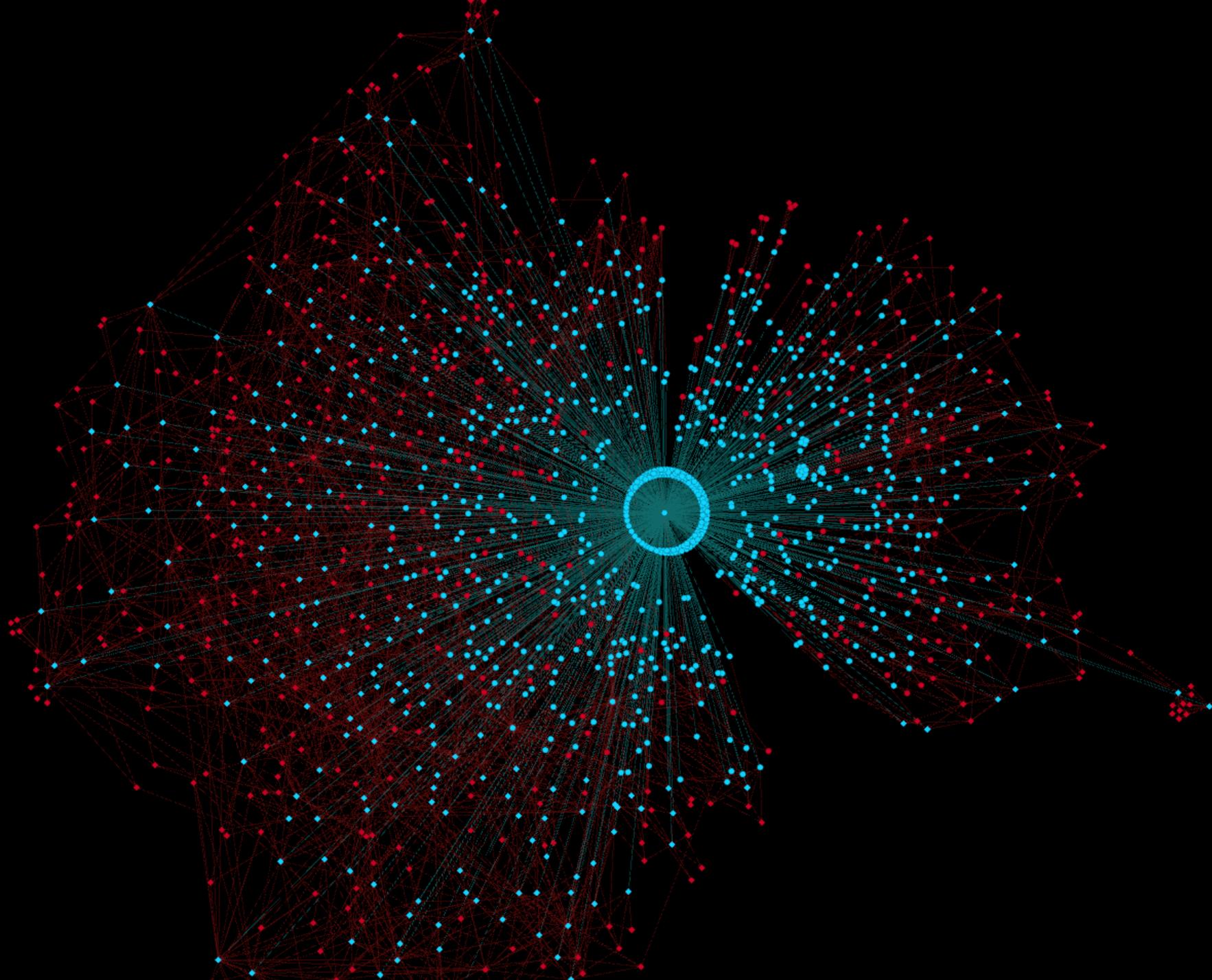
---

- Enrich the graph by adding directly upper level hierarchical relations
  - add all one-hop direct hierarchical relations for each concepts in the graph
  
- Results
  - 1652 new concepts added to the graph

# Graph Enrichment

---

- Pruning the graph
  - Remove any added UMLS concepts that are connected to others
- Results
  - 527 UMLS concepts remaining to the graph
  - 150 (16.22%) concepts of **original** graph are not connected to others except the CHF
    - Using Indirect hierarchy...

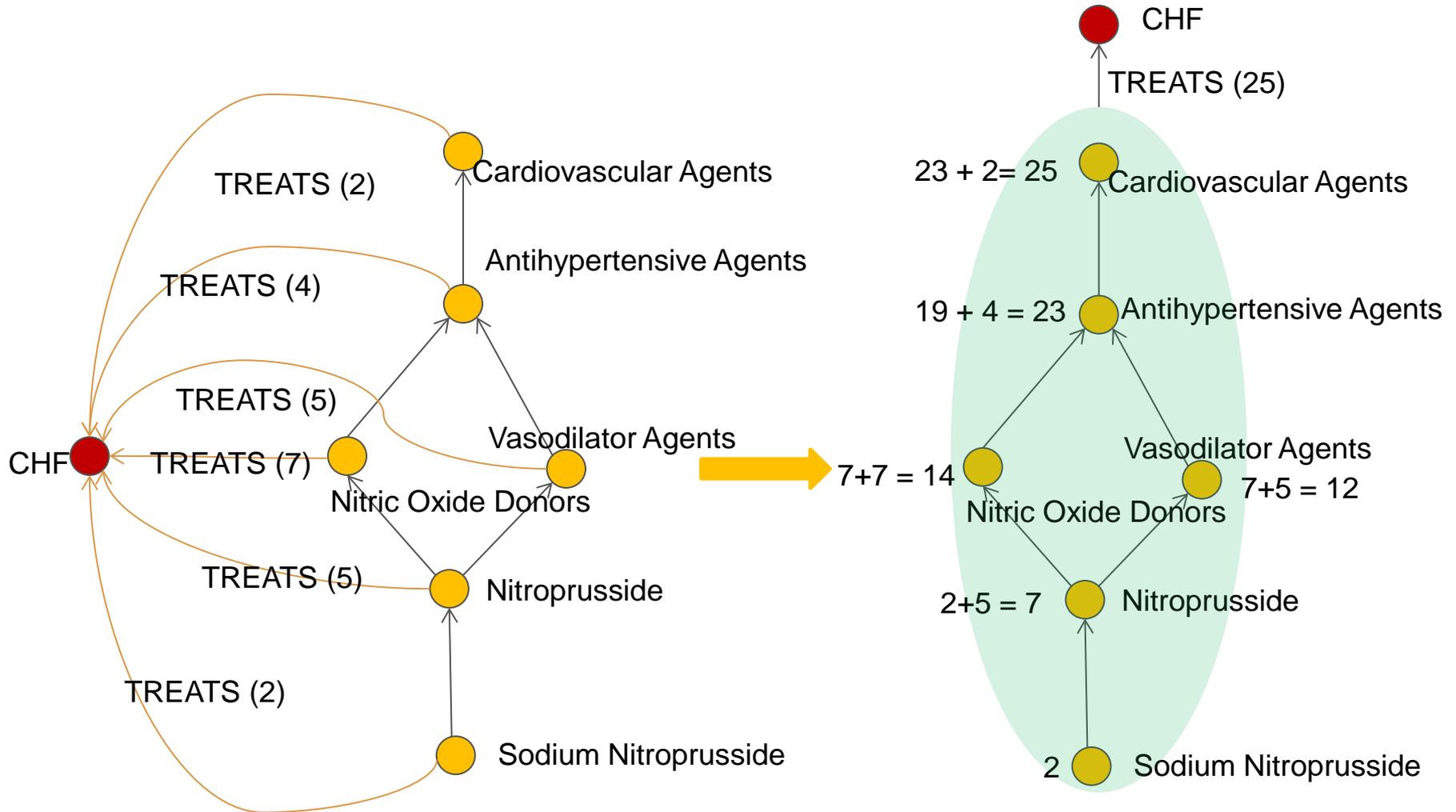


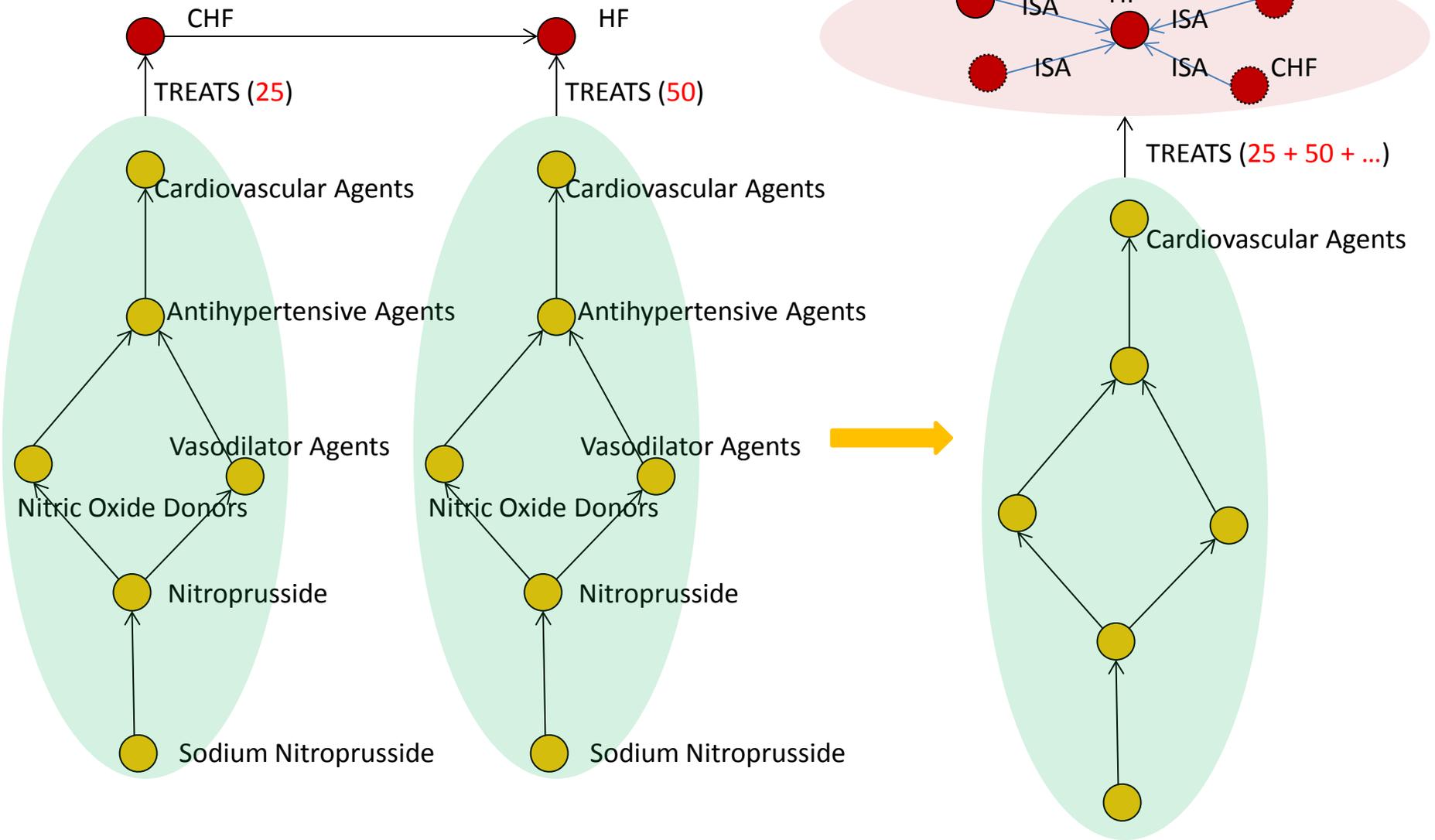
# Information Aggregation

---

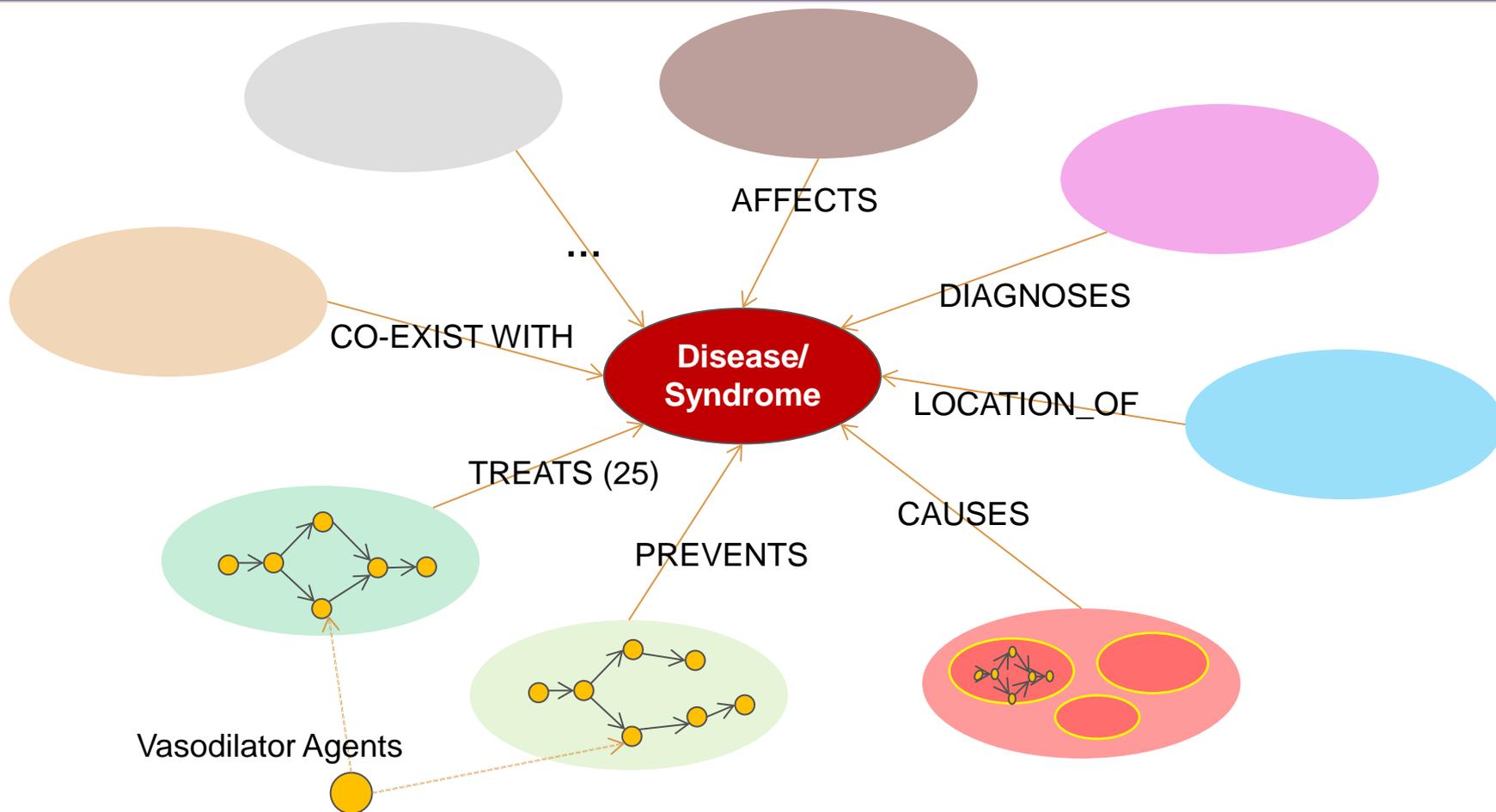
- Aggregate individual concepts into collections could condense large graphs
- Propagate the frequency information to the representative concepts
- Challenges:
  - Determine the level of granularity
  - Determine the most representative concepts in each cluster

# Process of Aggregation





# Clustering by predicate types



One concept could have membership to multiple clusters

# Future Work

---

- Measure the interestingness of concepts in a specific hierarchy
  - Degree centrality
  - Information content
- Develop use cases for demonstrating the usefulness of information aggregation for navigation, knowledge extraction, summarization, etc.

# Acknowledgments

---

- Olivier Bodenreider (Mentor)
- Bastien Rance
- Thomas Rindfleisch
- Marcelo Fiszman
- Dongwook Shin
- Caleb Goodwin
- Rui Zhang
- Liz Workman
- May Cheh
- National Library of Medicine

**Thank you!**