THE UNIVERSITY OF ALABAMA IN HUNTSVILLE | COLLEGE OF SCIENCE

Research and Creative Experience for Undergraduates (RCEU) Program 2024

Using Bayesian Networks to Predict Disease Comorbidities

Nolan Stuessy, Mentor: Dr. Satyaki Roy,

Department of Mathematical Sciences

Overview

- **Motivation.** Undetected or late-detected comorbidities lead to worse patient outcomes.

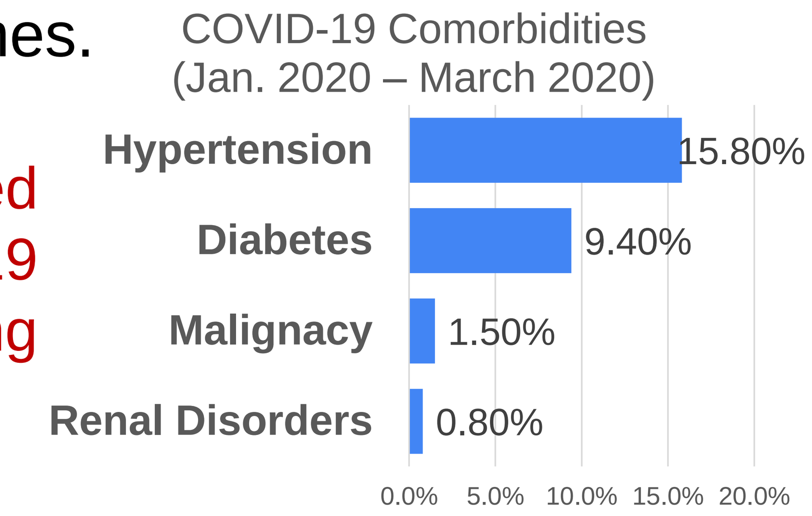

Figure. There is increased prevalence of COVID-19 among patients with underlying conditions or comorbidities[1].

- **Research gap.** Current models focus on phenotypic links, struggling to predict disease subtype relationships.
- **Aim.** Use graph-based relationships to better predict missing and future comorbidities at the patient level.

Methodology

- Data was obtained from *MIMIC-III*[2], a critical care database, and formatted to compare diseases, indexed by their *ICD-9 codes*, across multiple patients.

Diseases:	A	B	C
Patient 1	1	1	0
Patient 2	1	1	0
Patient 3	0	0	1

- Disease types that co-occur were represented graphically using *Bayesian networks* and disease prevalence was measured by conditional probabilities.

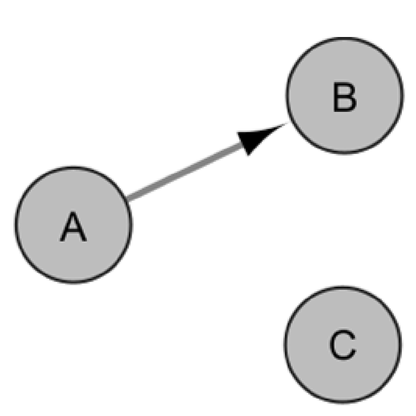

Figure. The nodes represent disease codes and directed link (u, v) represents the influence of node u on the occurrence of v.

- An *Independent Cascade Model*[3] was used to simulate the effect of diseases affecting each other and compare their marginal utilities.

$$\Delta\sigma(A,v) = \sigma(A \cup \{v\}) - \sigma(A)$$

- Seed nodes that affected the most other diseases were identified with a greedy algorithm that searched through and scored different nodes to maximize the spread of influence.

Proposed approach. Find seed nodes in BayesNet[4], that maximize a *cost function* based on 2 criteria:

1. Reward the activation of disease nodes of interest.
1. Penalize the activation of off-target nodes.

Key Findings

- We were able to analyze the influence spread of disease subtypes within a given category.

Respiratory Sinus Influence Maximization

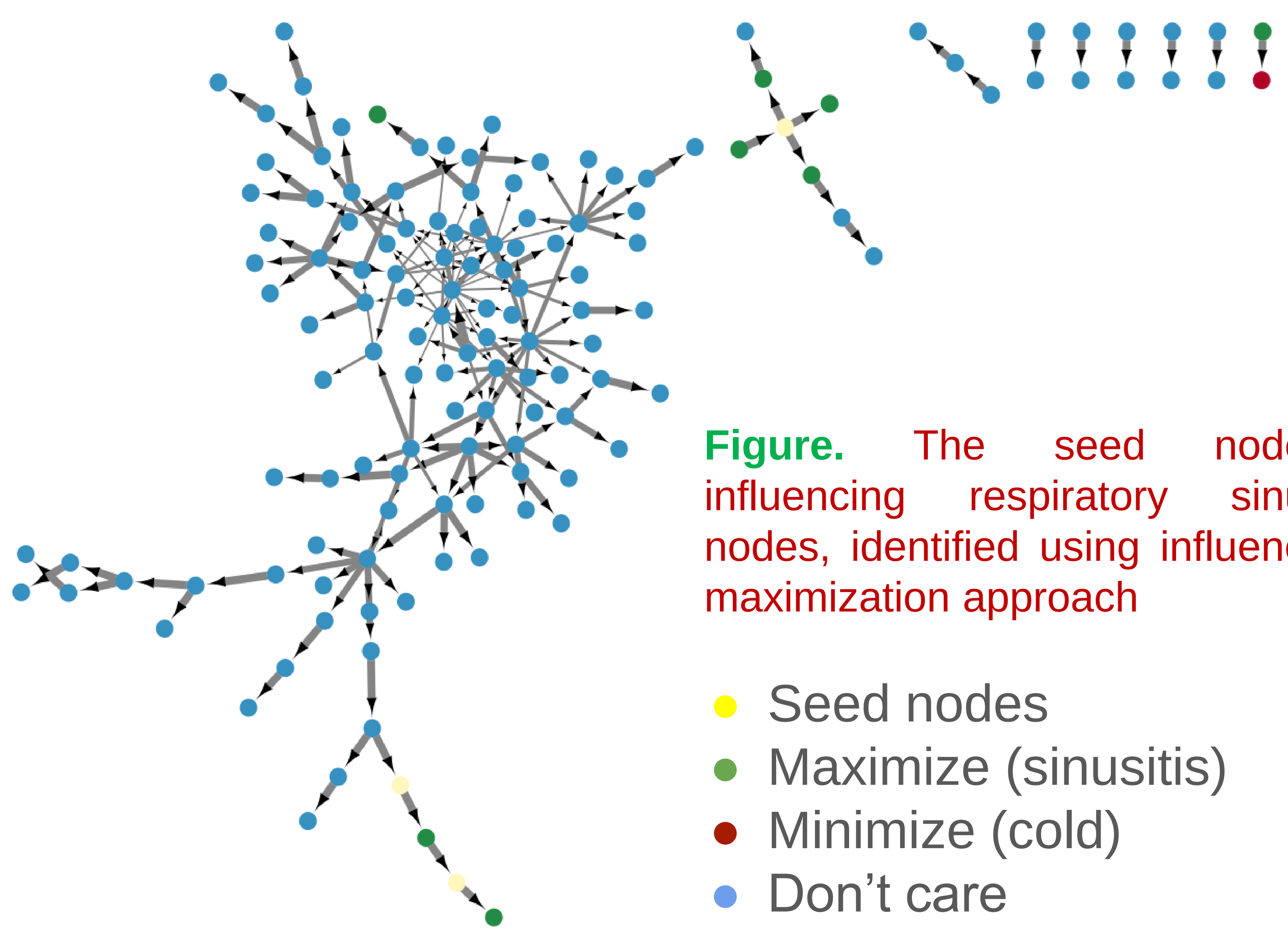

Figure. The seed nodes influencing respiratory sinus nodes, identified using influence maximization approach

- We compared the clustered subgroups of diseases with the overall dataset to find places where observations vary from the predefined categories.

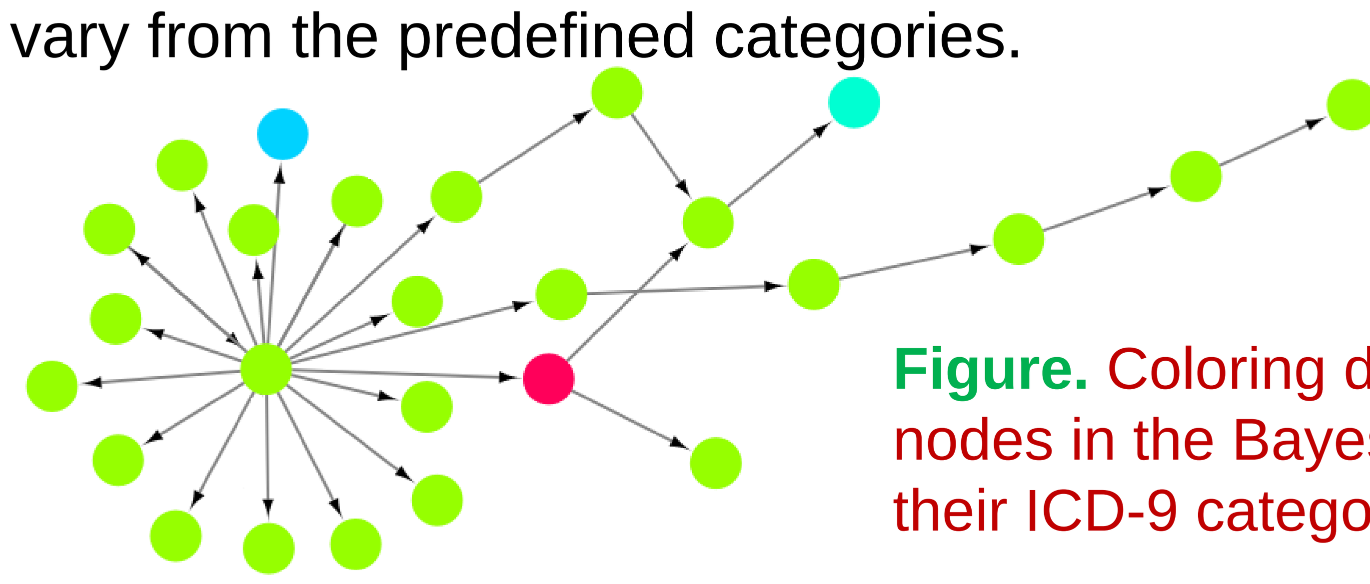

Figure. Coloring disease nodes in the BayesNet by their ICD-9 categories.

- **Ongoing efforts.** Use *diversity indices*[5] to test the alignment of the ICD-9 category with co-occurrence based on the BayesNet.

Conclusions

- Visualizing and predicting missing comorbidities enhances risk assessment and personalizes treatment.
- Future work may explore diverse node clusters and the clinical predictiveness of the BayesNet.

References

1.A. Sanyaolu, et al. "Comorbidity and its impact on patients with COVID-19." SN comprehensive clinical medicine 2 (2020): 1069-1076.
2. A. Johnson, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3.1 (2016): 1-9.
3. D. Kempe, J. Kleinberg, and É. Tardos. "Maximizing the spread of influence through a social network." Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003.
4. I. Kim and J. Yong-Gyu. "Using Bayesian networks to analyze medical data." International workshop on machine learning and data mining in pattern recognition. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.
5. P. Sen. "Gini diversity index, Hamming distance, and curse of dimensionality." Metron-International Journal of Statistics 63.3 (2005): 329-349.

Acknowledgements

All RCEU projects were sponsored in part by the Alabama Space Grant Consortium, the UAH Office of the President, Office of the Provost, Office of the Vice President for Research and Economic Development, the College of Science, the College of Arts, Humanities, and Social Sciences, and the College of Education.

HONORS COLLEGE
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE