Network-Based Exploration of Nut Allergy Prevalence Across Demographic and Clinical Cohorts

In this study, I used the Childhood Allergies: Prevalence, Diagnosis, and Treatment Outcomes dataset, which contains data aimed to understand the prevalence and treatment of childhood allergies. It includes retrospective data from healthcare providers, detailing the number of individuals affected by allergic disorders ranging from asthma to food allergies and contains demographic information such as gender, race, and ethnicity. In this report, I will dive into my methodology, results, and conclusions.

In this project, I analyzed a dataset on childhood allergies using graph theory to understand the connections between individuals and their specific nut allergy statuses. The program reads data from a CSV file, containing detailed information about individuals' demographic characteristics and their allergy history. It then constructs a directed graph, where nodes represent either individuals or various types of nut allergies. The relationships between individuals and their allergies are established in this graph. The program calculates the centrality measures for each node, providing insights into the prevalence and demographic distribution of these allergies. This approach allows an analysis of the relationship between different demographic factors and allergy types, giving a unique perspective on allergy prevalence and patterns.

To create the nodes and edges, I defined two main types of nodes in the graph using the NodeType enum: Individual and NutAllergyStatus. Individual nodes represent patients and are characterized by their demographic data, such as gender, race, and ethnicity, as well as their allergy cohort status. These nodes are created from instances of the Record struct, which includes detailed allergy and demographic information for each patient, capturing attributes like birth year, gender, race, and specific allergy start and end statuses.

Randomized Sample Output of Nodes:

Node: 158045, Edges: ['Peanut Allergy', 'Walnut Allergy', 'Pecan Allergy', 'Pistachio Allergy', 'Almond Allergy', 'Hazelnut Allergy'] Node: 159379, Edges: ['Peanut Allergy', 'Pecan Allergy', 'Almond Allergy'] Node: 159385, Edges: ['Peanut Allergy', 'Pistachio Allergy'] Node: 208921, Edges: ['Peanut Allergy', 'Pecan Allergy', 'Hazelnut Allergy']

On the other hand, NutAllergyStatus nodes represent different types of nut allergies, including Peanut, Tree Nut, and others. Each of these nodes serves as a central point to connect with all individuals who have that particular allergy. The edges in the graph are crucial as they establish the relationship between an Individual node and a NutAllergyStatus node. An edge is created from an Individual node to a NutAllergyStatus node whenever the individual has the corresponding allergy, as determined by the allergy status fields in the Record struct, such as peanut_alg_start. The presence of an edge indicates that the individual has a history of that specific nut allergy.

Additionally, I calculated the degree of centrality to understand the prevalence of nut allergies across various demographic groups, highlighting how certain allergies are more common in specific segments of the population. The outputted centrality scores are shown, which I have organized, below:

Gender-Based Prevalence of Peanut Allergies: **Males (S0)**: Centrality Score for Peanut Allergy - 0.900. **Females (S1)**: Centrality Score for Peanut Allergy - 0.856.

Race-Based Prevalence of Peanut Allergies: **Black (R1)**: Centrality Score for Peanut Allergy - 0.902. **Other (R3)**: Centrality Score for Peanut Allergy - 0.787.

Ethnicity-Based Prevalence of Peanut Allergies: **Non-Hispanic (E0)**: Centrality Score for Peanut Allergy - 0.883. **Hispanic (E1)**: Centrality Score for Peanut Allergy - 0.855.

Payer Factor in Prevalence of Peanut Allergies: **Medicaid (P1)**: Centrality Score for Peanut Allergy - 0.891. **Non-Medicaid (P0)**: Centrality Score for Peanut Allergy - 0.880.

Atopic March Cohort and Peanut Allergy Prevalence: **With Atopic March Cohort**: Centrality Score for Peanut Allergy - 0.879. **Without Atopic March Cohort**: Centrality Score for Peanut Allergy - 0.883.

For gender, both males (S0) and females (S1) showed high prevalence of peanut allergies, with slightly higher scores in males, suggesting a marginally greater prevalence or reporting in this group. Regarding race, the data indicated a significant prevalence of peanut allergy in the Black racial group (R1), with a high centrality score, contrasting with the lower score in the 'Other' group (R3). This variation hints at racial disparities in allergy prevalence or reporting.

Ethnically, both Non-Hispanic (E0) and Hispanic (E1) populations showed high centrality scores for peanut allergy, though it was slightly more prevalent in the Non-Hispanic group. In terms of healthcare coverage, the centrality scores pointed to a higher prevalence or reporting rate of peanut allergies in the Medicaid population (P1) compared to the Non-Medicaid group (P0). Furthermore, the atopic march cohort analysis revealed peanut allergy as prevalent in both groups, but with a slightly higher centrality score in the group without the atopic march cohort.

These results are relevant in guiding healthcare providers and policymakers in developing targeted strategies for managing and preventing nut allergies, particularly focusing on the most affected demographic groups. For future studies, we can utilize decision tree algorithms alongside this dataset, encompassing longitudinal information such as age at onset, types of allergies, and patient demographics, to detect patterns and predict the potential progression of allergies in individuals.