Using Networks to find patterns in Healthcare Payer Data

F Thompson
BCBSM
This work comes out of working with the Underwriting and Actuarial Pricing teams to understand what is driving our costs. I’m using Network Analysis to visualize and isolate interesting features in the payer claims data. Today I’ll walk through a couple basic examples.

1. Members who travel out of state for services.

2. Condition Comorbidity for high cost members.
First Example: Traveling

Cost for services rendered out of state are a large driver in year over year trend. Understanding the traveling patterns can help contain this trend.
Networks in Healthcare Payer Data

Travel Costs

Must look at the data first!

Claims by where they came from and where they went!

For all claims rendered in Ohio, this plots the total claims by rendering zip, and also the same claims by subscriber zip.

Size and color are proportional to total claim dollars for ZIP. Buckets were based on straight deciles. The top two deciles were emphasized, the other diminished.

Cleveland
Toledo
Columbus
Cincinnati
I have connections in the data, connections between zips where people live and zips where people get services.

I have a graph. Like most real world graphs there are lots of vertices and edges, and most don’t mean a thing.

To distill meaning out of the apparent chaos, we do three things:

1. Subset the data. Here, we consider travel to particular states individually. Someone will go to Wisconsin for different reasons than going to Ohio.
2. We prune. Erase all edges less than $8,000 in this case.
3. Run algorithms to find communities and visualize graph neatly (few crossing edges, plot strongly connected vertices next to each other).

Subscribers from Royal Oak had $137,150 worth of claims in Chicago.
People traveling to Ohio generally go to four distinct locations: Cleveland, Cincinnati, Toledo and Columbus. These cities are the focus of the four communities found by the algorithm.

Are Cincinnati and Toledo connected?

Very few connections between Cleveland and Cincinnati. Does not seem random.

Is the Columbus grouping stable over time. Trend is the issue, not raw cost per se.
Cleveland and Cincinnati are disconnected.

Are Toledo and Cincinnati connected in some sort of referral network? Why are they in the same community with consistency?
Second Example: Condition Comorbidity

High cost members often have very clinically complex cases. A large fraction of the cost is driven by easily identifiable single conditions, i.e. metastatic cancer. Others are driven by multiple conditions. Understanding which baskets of conditions drive the cost helps inform where and when to have care management programs.

This example looks at the conditions in complex cases for high cost members.
Networks in Healthcare Payer Data

Condition Comorbidity

Each condition is visualized as a node, vertex, or bubble on the graph.

Here is a rough example:

- Diabetes
- COPD
- Cancer
Each node is assigned a color and size

The color relates to the average cost of all high cost members that have that condition

The size is proportional to probability of the members with that condition repeating as high cost

If a member has three conditions, her total cost would be used in the average of all three conditions

In general, we have double, triple, counted everyone's cost

These charts are not to divide the whole pool of cost into buckets, but to discover relationships between conditions
Now, to account for co-morbidity, we add edges or lines.

Each line represents the cost for members with both the connected conditions.

The thickness of the line is proportional to that the cost.
Subset members with high cost without top tier conditions.

192 edges

24 nodes
After pruning we can look at high degree nodes.

Chronic Liver Disease is a major driver.

Major Amputation is also, but is perhaps a lagging indicator.
Like to thank my team.

Tia
Paul
Felice
Jane

And my boss.

Raghu