

NETWORK ANALYSIS CASE STUDY

Warning: Please remove data from your laptop when the exercise is completed. There are serious consequences for your instructors and Galvanize for violation. Thank you.

Problem: You are provided two datasets, one has Provider Ids and Provider specialties, the other has Provider Ids and Procedure Codes for procedures each doctor performed. Not all Providers have their specialties identified. Your task is to label the unknown specialties using community detection in an appropriate network.

Suggested steps:

1. Access data -- read in the CSV files
2. Merge to create a combined file
3. Perform EDA. You will see that some common procedure codes do not help in the analysis. You may have to do some NLP to discard some procedure codes.
4. Create a network with appropriate Nodes and Edges
5. Apply a community detection algorithm (that you created yesterday)
6. Check to see if communities correspond to specialties
7. Identify unknown specialties
8. Create a CSV of nodes and edges that can be input to Gephi
9. Visualize the network and communities
10. Check to see if unknown specialties can be identified