

Some questions about GNN:

1. **For building a network, instead of using KNN, we search for all the neighbors of a node to form a network. Is this network method applicable?**

This method is a bit like the process of expanding the search in circles. If possible, I plan to expand it in two circles.

e. g. For the three neighbors of node i, J, K, L.

And the neighbors of J other than i are J1, J2...

The same is true for K and L (there may be repeated neighbors, only take once)

So for node i, the neighbors used to form a network should be:

First-level neighbors: J, K, L (high weight for attribute contribution)

Second-level neighbors: J1, J2..., K1, K2..., L1, L2... (low contribution)

2. **Split the model test set and training set**

According to Stephen's paper, I plan to try the following two methods

- A. Random Split
- B. Select several administrative divisions as test sets only, and do not participate in training.
-- Test generalization ability

3. **Now only anxiety disorder is tested as the outcome variable.**

For the others, should we use existing models to predict the results one by one?

Or is there any way to aggregate the results?

4. **How to select independent variables?**

According to Sanja's guidance, I will try first extract the listed control variables as a reference for expert experience. Then I will select more variables based on correlation to prescription and check VIF before model testing.

Is this more reasonable than extracting based on correlation now?

About Border Vacuum:

1. Draw the gradient distribution of the following variables to facilitate understanding and further discussion

A: Socioeconomic variables: such as income

B: Prescription variables

C: Prescription prediction value

2. The boundary finding idea provided by Daniele.

My current understanding:

1. For i and i 's neighbour $J1$, add the attribute $\text{DIFF}_{i,J1}$ (maybe use absolute value to eliminate the directionality?)
2. Quartile grouping, extract the edges of Quartile 1 and Quartile 4 for high and low difference groups.
3. For example, for the region pair (l,m) , we hope to use the edge B_{lm} to predict $\text{DIFF}_{l,m}$.

Questions

1. Are GNN part and finding the vacuum boundary an integral task or two separate tasks?
It seems that using this idea to explore the vacuum boundary, we can directly use the true value from the dataset, regardless of the prediction of GNN?
2. What should we use to construct the independent variables of edge B_{lm} to predict DIFF ?
For example, if we want to predict $\text{DIFF}_{l,m}$, we should give some environmental features of nodes l and m on the edge B_{lm} ? (For example, large rivers, vegetation, impervious areas, because these variables are related to vacuum boundaries?)