# Hypotheses/Objectives:

1. H1 - The nature of hospital diagnoses reporting is different before COVID-19 pandemic, and during pandemic (March - Dec 2020).
   1. The characteristics of the disease co-occurrence network before and during pandemic is different, in terms of density, prominent diagnoses, clusters, etc.
   2. During pandemic, disease co-occurrence increases.
   3. During pandemic, many non-critical diseases are under-reported due to shortage in beds and postponement of treatments.
2. There is a trend/pattern in disease network developed using monthly data suggesting unidirectional changes in hospitalization in US states.

Karthik email to me (5/19/2021):

So let us plan to analyze three networks - for hospital discharge data during 2019-2, 2020-1, and 2020-2 (i.e., pre-covid, peak-covid, during-covid). We can compare the characteristics of the networks and report it. I was also thinking it would be helpful if we could generate some html visualizations of the networks and report the links to the journal paper. It will be helpful for the reviewers to actually try and see the value of comparing disease networks during different time periods.

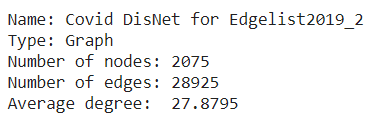
# EDA

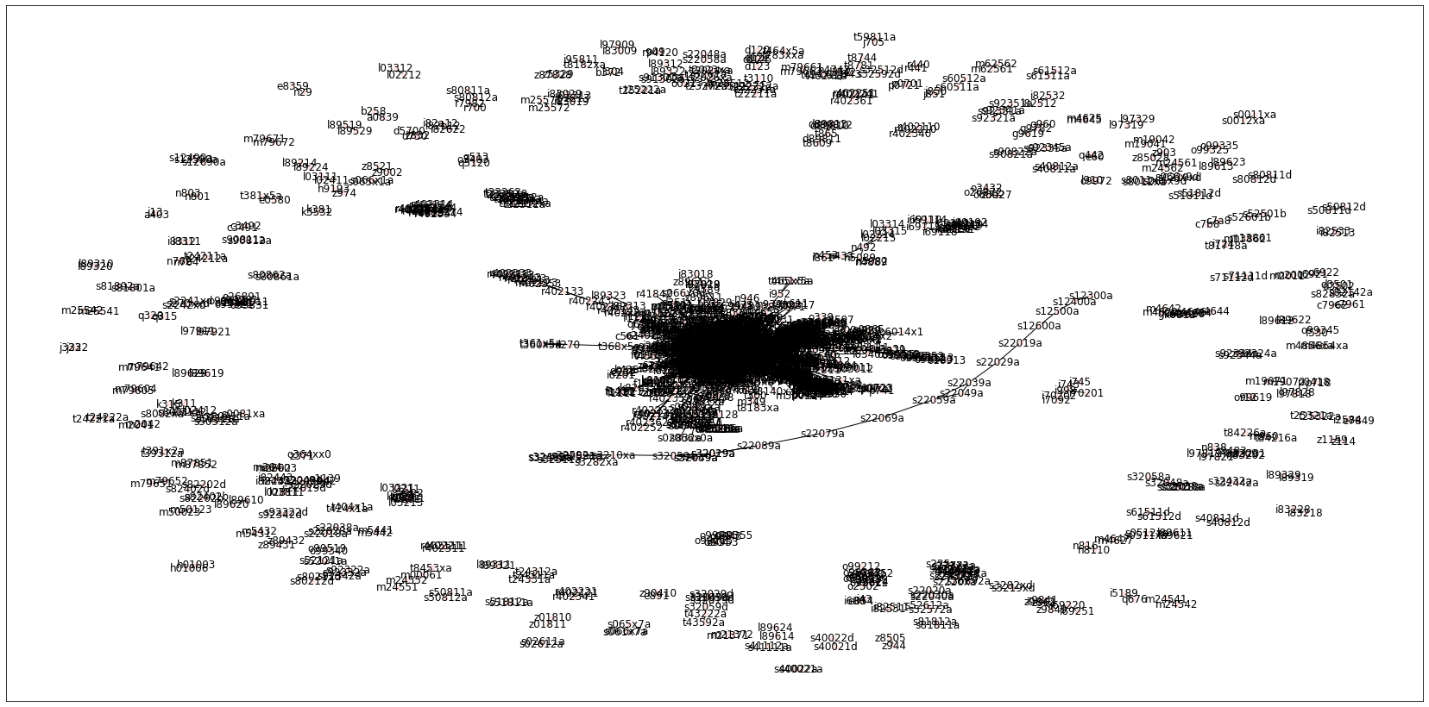
|  |  |  |  |
| --- | --- | --- | --- |
|  | Edgelist2019\_2 | Edgelist2020\_1 | Edgelist2020\_2 |
| **Basic Info** | | | |
| Shape (# of edges) | (28925, 3) | (28980, 3) | (30105, 3) |
| # of Source | 1303 | 1272 | 1330 |
| # of Target | 1737 | 1730 | 1779 |
| # of Nodes | 2075 | 2049 | 2101 |
| Average degree | 27.8795 | 28.2870 | 28.6578 |
| Most similar “z20828” | N/A | (e8342,Hypomagnesemia) | (f419, Anxiety disorder) |
| **Centrality Analyses** | | | |
| degree\_centrality |  |  |  |
| eigenvector\_centrality |  |  |  |
| katz\_centrality |  |  |  |
| **Clustering** | | | |
| # of triangles | 791705.0 | 804995.0 | 842044.0 |
| Transitivity | 0.49345707598767147 | 0.494683603318231 | 0.490647305950 |
| Avg clustering coef | 0.4677968262043043 | 0.468306620960989 | 0.478484940043 |
|  |  |  |  |
|  |  |  |  |
| **Density** | | | |
| Graph Density | 0.013442390584517433 | 0.013811996705710103 | 0.013646562861 |
| Subgraph Density | 0.024968072360538753 | 0.026247212373606877 | 0.025556448085 |
| Z20828 Density | N/A | 0.46855543347459844 | 0.185543067738 |
| **K-means & TSNE** | | | |
| K-means convergence | 20 | 10 | 12 |
| TSNE plot |  |  |  |

Edge1：

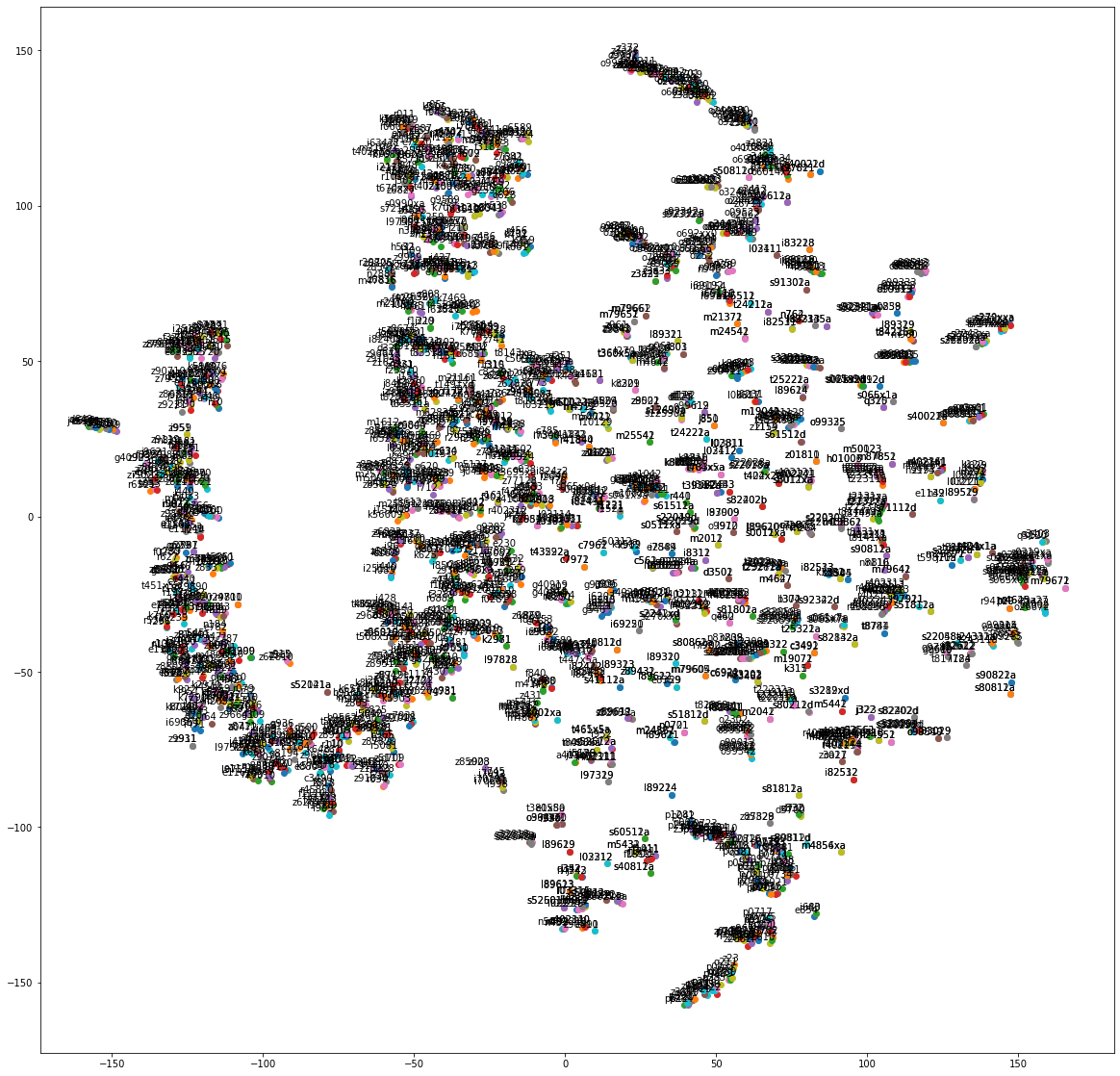
Top 5 most frequent icd10 code：

1. z79899(Other long term (current) drug therapy) (count: 453),
2. z87891(Personal history of nicotine dependence) (count: 420),
3. e785 (Hyperlipidemia, unspecified) (count: 417),
4. i10 (Essential (primary) hypertension) (count: 407),
5. k219(Gastro-esophageal reflux disease without esophagitis) (count: 390)





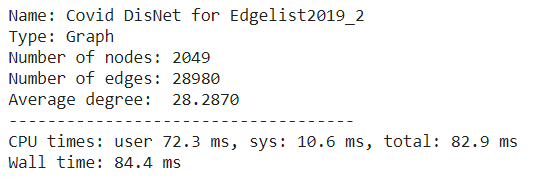
In this list, they do not have z20828 reported.

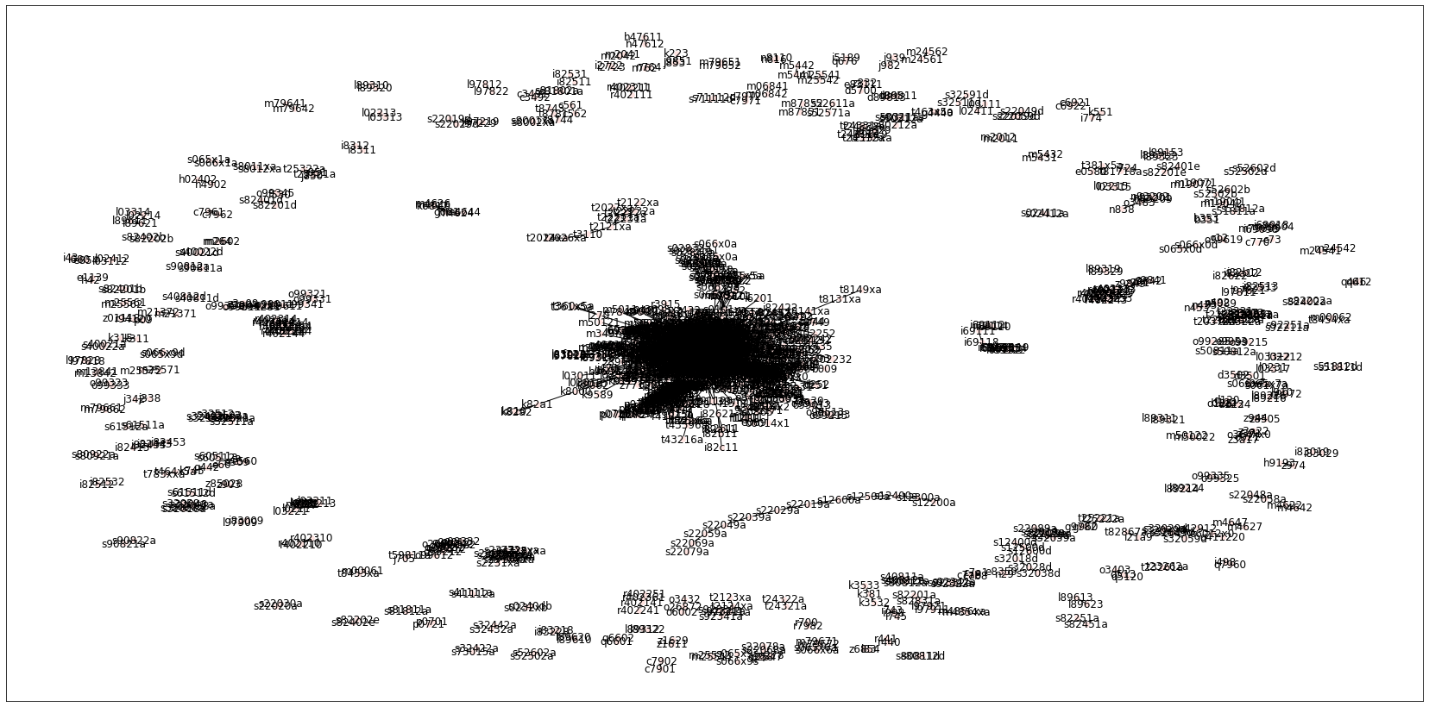


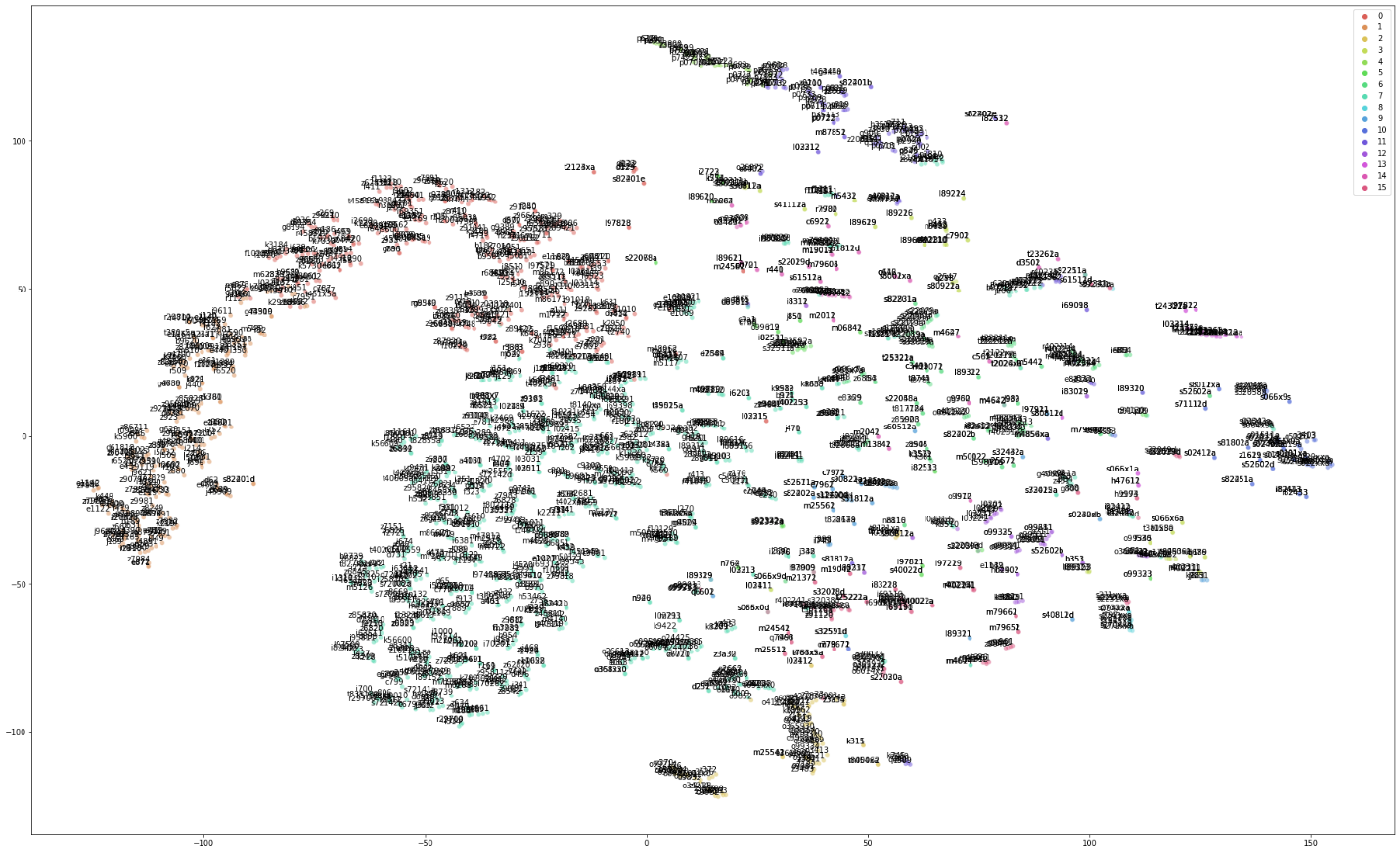
Edge2:

Top 5 most frequent icd10 code：

1. z79899(Other long term (current) drug therapy) (count: 471),
2. i10 (Essential (primary) hypertension) (count: 462),
3. e785 (Hyperlipidemia, unspecified) (count: 428),
4. **i2510 (Atherosclerotic heart disease of native coronary artery without angina pectoris) (count: 379) (prev\_rank: 6),**
5. **n179 (Acute kidney failure, unspecified) (count: 377) (prev\_rank: 14)**







Edge3:

Top 5 most frequent icd10 code:

1. **z20828(Contact with and (suspected) exposure to other viral communicable diseases)(count: 501)(prev\_rank:24),**
2. z79899(Other long term (current) drug therapy)(count: 469)(prev\_rank:1),
3. e785(Hyperlipidemia, unspecified)(count: 465)(prev\_rank:3),
4. i10(Essential (primary) hypertension)(count: 445)(prev\_rank:2),
5. **n179(Acute kidney failure, unspecified)(count: 394)(prev\_rank:5)**

