

## APPENDIX D: RESULTS AND DISCUSSION

This appendix provides the results and discussion of the five datasets IoS edges, Israel nodes, QLD nodes, QLD temporal edges, and QLD temporal and spatial edges.

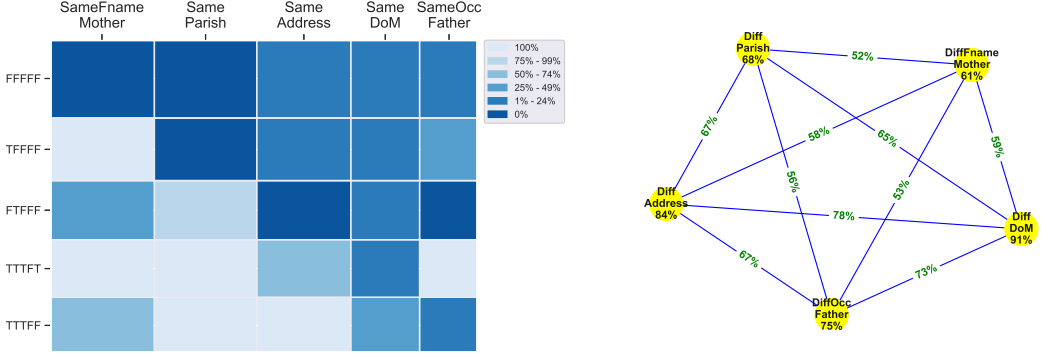


Fig. 1. Tile and association plots of the IoS edges dataset.

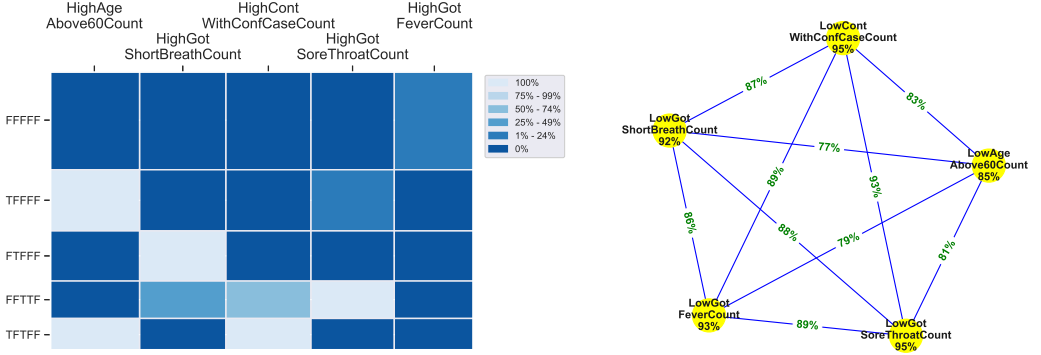


Fig. 2. Tile and association plots of the Israel nodes dataset.

Fig. 1 shows identified abnormal patterns and associations for the IoS edges dataset. As can be seen from the tile plot, the largest pattern represents birth records identified as siblings with mismatching mother's first name, parish, address, parent's date of marriage, and father's occupation. It is interesting to note here that most of the abnormal edges do not have a matching parent's date of marriage. While this can possibly represent the mother getting remarried or unmarried parents, domain expertise would be required to determine whether there are any other reasons behind such a mismatch. Furthermore, the first pattern represents siblings with a different mother's first name. This can be due to a remarried father or an error in the record linkage process, where our approach can be used to obtain insights about the underlying record linkage process. As per the association plot, 56% of the edges are abnormal due to *DiffParish* and *DiffOccFather*. It can be assumed that individuals have migrated between parishes, and therefore their occupations have changed. If assessed, these scenarios will relate useful information about the lifestyle of people who lived on the Isle of Skye in the mid to late 1800s.

Fig. 2 shows identified abnormal patterns and associations for the Israel nodes dataset. As can be seen from the tile plot, the majority of weeks have had no considerable increase for the features under consideration. Shifting the focus to the second and last patterns, there is an increase in the number of people above the age of 60 years who got tested, which can possibly be due to an outbreak in an elderly home. Furthermore, the third pattern reflects weeks with patients reported due to shortness of breath and the fourth cluster with a 100% increase in patients with sore throat. All the above cases require domain expertise to answer questions such as, if an increase in patients with shortness of breath reflects those weeks when people suffering from respiratory issues were getting tested? If yes, does it mean that people with respiratory issues are more vulnerable to this pandemic? As per the association plot, the anomalous behavior of the identified binary patterns are due to low values occurring for the five features under consideration.

Fig. 3 shows identified abnormal patterns and associations for the QLD nodes dataset. As can be seen from the tile plot, only a few features have a high color intensity, clearly resembling the incidents that require domain expertise. The second and last patterns show an increase above the average in the count of old aged people contracting the disease, while the third to the fifth patterns show an increase in the count of people who have acquired the disease overseas. This information can aid in identifying weeks in which repatriation flights have arrived, postcodes that have elderly homes, parents with children living abroad, and so on. As per the association plot, the abnormal binary patterns have a decrease in all features when compared to their mean value, thus all features are named *Low*. While domain expertise is required to assess incidents that are higher than average, as discussed above, insights are required to assess these low incidents in detail.

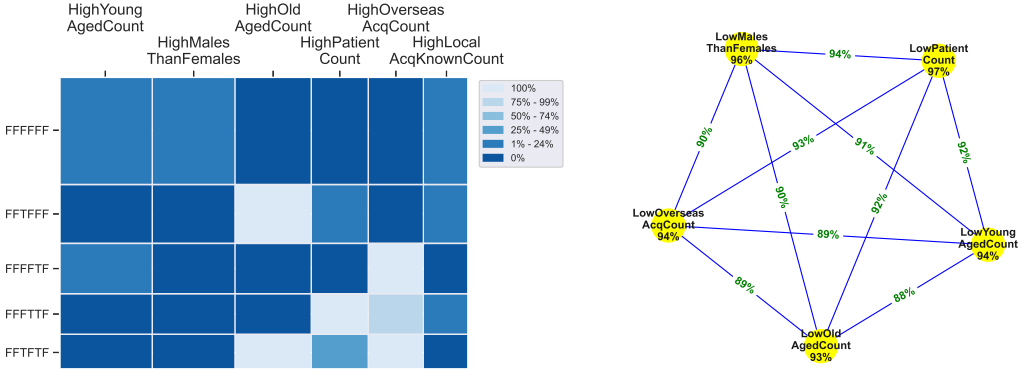


Fig. 3. Tile and association plots of the QLD nodes dataset.

Fig. 4 shows identified abnormal patterns and associations for the QLD temporal edges dataset. As can be seen from the tile plot, the second, third and fourth patterns show an increase in the count of indigenous people contracting the disease. While this aids in identifying postcodes with indigenous people, it also raises the question whether indigenous people have less immunity compared to non-indigenous individuals. The fourth pattern shows an increase in *HighOverseas-AcqCount* possibly reflecting a period when repatriation flights have arrived. Furthermore, the last pattern indicates a significant increase in the death count. Analyzing these binary patterns will allow local health authorities to better plan the utilization of their medical resources.

As per the association plot, having a node support of 78%, the feature *HighLocalUnidenInter-TravCount* is the most influential, stating that there is an increase in locally acquired unidentified

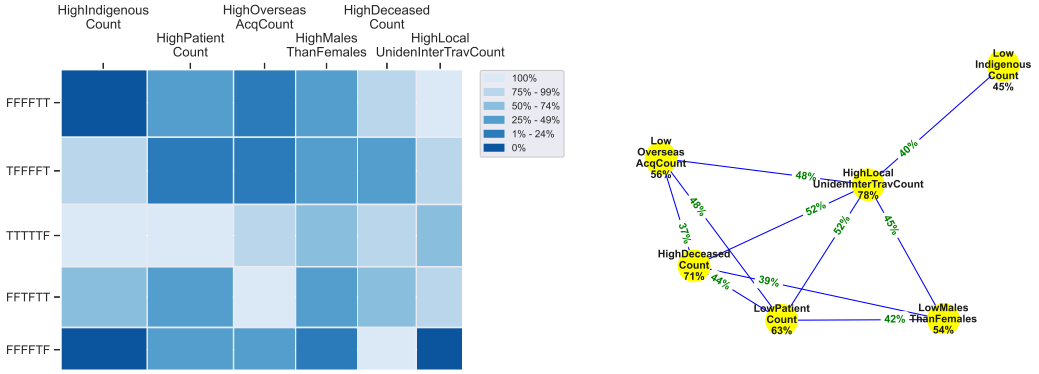


Fig. 4. Tile and association plots of the QLD temporal edges dataset.

interstate travel count. Furthermore, 52% of the edges are abnormal due to both *HighDeceased-Count* and *HighLocalUnidenInterTravCount*. Analyzing these would be beneficial to determine the existence of community transmission or a new variant of the virus.

Fig. 5 shows the identified abnormal patterns and associations for the QLD temporal and spatial edges dataset. As can be seen from the tile plot, considering all the patterns in general, domain expertise is required to determine the reasons behind the considerable increase in ventilated count, ICU admitted count, locally acquired unidentified interstate travel count and under investigation count, as they may resemble an outbreak in a particular postcode or a long weekend over which people have traveled interstate. As per the association plot, all features being low reflects that a particular postcode in week 1 has a low count in comparison to the neighboring postcode in week 2. When assessing these patterns, it is important to understand the nature of the postcodes such as residential or business as that affects the people count.

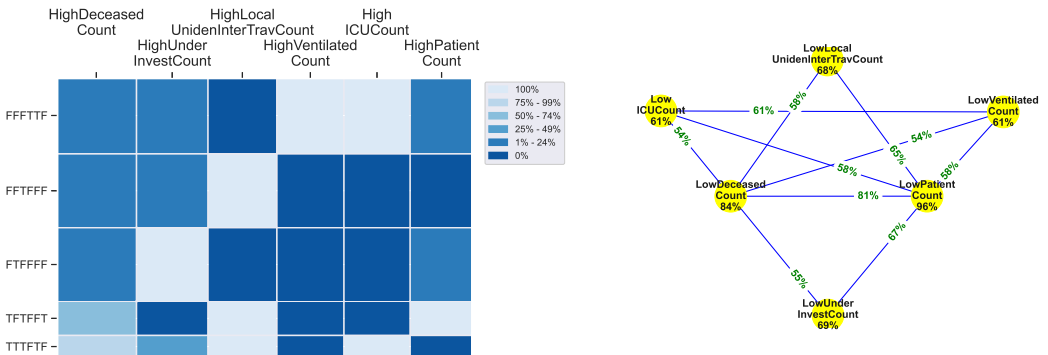


Fig. 5. Tile and association plots of the QLD temporal and spatial edges dataset.