Exploring Network Fusion methods for Multi-omic data integration

Ramón Reszat[1], Rachel Cavill[1]
[1]Dept. Data Science and Knowledge Engineering, Maastricht University, Maastricht, The Netherlands
E-mail: r.reszat@student.maastrichtuniversity.nl, rachel.cavill@maastrichtuniverisity.nl

1. Introduction

With studies increasingly being designed using multiple experimental techniques, a large amount of heterogenous data including transcriptomics, metabolomics and other characteristics of the samples are being generated. This presents a challenge for existing data analysis methods, as different feature scales and distributions make it difficult to discover relations between the samples. One approach for dealing with multi-omics data is through graph-based data integration. However, there are only a limited number of tools available for this purpose and they haven't been compared extensively so far. Therefore, studying this approach can lead to improvements in these tools for multi-omics data fusion.

2. Approach

We compared two tools Similarity Network Fusion (SNF) and Affinity Network Fusion (ANF) on a large-scale public dataset collected from the Cancer Cell Line Encyclopedia (CCLE). First, the influence of the scaling parameters, that measure the similarity within the two methods is measured. Thereafter, we analyse the affinity graphs for cluster separation and cohesion using the spectrum of the graph Laplacian and labelled data.

Subsequently, clustering algorithms are applied on the network representation to discover subtypes of cancer cell lines in a semi-supervised manner and the clusters are validated using the labels provided by the CCLE. Additionally, the clusters are screened for interesting clinical features using the Cancer Therapeutic Response Portal (CTRP) drug response dataset corresponding to the cell lines from the CCLE.

3. Results

Both methods, SNF and ANF construct networks with an algebraic connectivity in a comparable range. However, the representations produced by ANF tend to have sparser edge weights within the clusters.

We compared the fused networks generated from multiple omics sources with networks generated from a single omics dataset and were able to show increased clustering in the former.

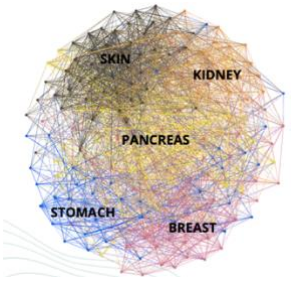

Figure 1. Fused affinity graph of carcinoma cell lines.

We analysed the networks using techniques from spectral graph theory and cohesion metrics, which showed that the hyperparameters of SNF and ANF have a range, in which they can drastically influence network topology. Using a k-means algorithm adopted to work on the graph structure, we were able to separate subtypes of lung carcinoma with a purity of 87% and to distinguish clear cell renal carcinoma from regular renal cell carcinoma with a purity of 73%. Some of these clusters were shown to have significant associations with drug resistance to selected anticancer compounds.

4. Discussion

Network-based data fusion methods have been used successfully by clinical researchers to analyse cell and tissue samples. SNF and ANF proof to be sensitive to the choice of their hyperparameters. On one hand increasing algebraic connectivity in these methods can decrease clustering quality. On the other hand, decreasing connectivity through the scale parameter reduces cohesion of the clusters. This represents a new approach of estimating the scale parameter in SNF and ANF.

Overall, our fused networks were shown to have biological relevance both in terms of relations to cancer subtypes and with respect to drug resistance.

References

1. B. Wang, A.M. Mezlini, F.Demir, M.Fiume, Z. Tu, M. Brudno, B. Haibe-Kains and A. Goldenberg "Similarity network fusion for aggregating data types on a genomic scale". *Nature methods* 11(3):333-337, 2014.
2. T. Ma and A. Zhang "Affinity Network Fusion and semi-supervised learning for cancer patient clustering". *Methods* 145:16-24, 2018.