

## CHAPTER 1

# Conclusion and further perspective

---

In the chapter, we shall summarize our findings regarding network deconvolution and estimation of mutual information. In particular, we shall conclude on the observed properties of the framework including shortcomings, wherefrom these originate and potential fixes. Moreover, we shall state further perspectives and possible future studies. This includes ways to correct for some of these shortcomings, other methods for estimating mutual information and applications of the framework

### 1.1 Conclusion

Using network deconvolution, we observed that for linear chains and networks that can be represented by a directed acyclic graph, the causal structure could be recovered perfectly when the true correlation between pairs of random variables were known as well as the topological structure. Namely, for any network where a node  $X_i$  is given by  $X_i = \epsilon_i + \sum_{j \in N_i^-} \tilde{\rho}_{j,i} X_j$  where  $N_i^-$  denotes the parents or in-neighbors of  $X_i$  and  $\epsilon_i$  is some independent noise, we observe that we can recover the coefficients  $\tilde{\rho}_{j,i}$ . However, removing the assumption of a topological order, we observed a bleeding effect, where especially for chains, this resulted

in an inability to perfectly recover the causal structure. In particular, the weak links in the chain with small  $\tilde{\rho}_{j,i}$  broke in the deconvolved network, whereas the more strongly connected subchains were observed to generate new edges, connecting random variables that should otherwise not be directly connected. This issue was not observed to the same extent when a more complex causal structure was used to generate the samples. Thus, when using correlations as the measure of similarity, we conclude that an assumption of the topological order of the random variables is important for reliable results, especially in the presence of chain-like substructures.

Furthermore, using only 400 samples from a network with 10 nodes, we show that the noise in the estimates of the correlations is so small that after a threshold is applied to the resulting deconvolved similarity matrix  $G_{dir}$ , we can in practice perfectly regain the network structure. From this, we show that when using mutual information instead of correlation, we introduce errors such that the underlying assumption does not hold true for how information is convolved. Namely, for mutual information, it no longer holds that  $G_{obs} = \sum_{k=1}^{\infty} G_{dir}^k$ . The error from this assumption is exemplified in the case of Gaussian chains which are observed to result in the largest errors. We find that as long as the pairwise correlations in the chain is at most 0.9 in terms of absolute value, the errors introduced by using mutual information are so small that the inferred causal structures are near identical to those of using correlation. In particular, we conclude that if the underlying causal structure is a linear DAG with Gaussian noise, and the observed random variables are homeomorphic to the underlying random variables such that the mutual information between the observed random variables is the same as for the underlying causal structure, we can in most cases perfectly recover the causal structure.

- Long chains can be a problem when the links have varying strength. We observe that the chain might break at the weak points and more strongly connected parts bleed into neighboring nodes resulting in *dense* subgraphs.
- Long chains using MI with high correlation if symmetric seem to be connected  $i - i + 1$  and  $i - i + 2$ . If very large MI might want to treat differently.
- Works well, through experimentation, on linear networks. If  $X_j = f(\sum X_i)$  for in-neighbors it is not necessarily well, however as MI is independent of marginals, we expect better performance on these systems than if one had used correlation. In particular, if it is a chain, transforming each variable gives exactly the same result.
- Although [?] use normalized mutual information, we have not done this. However, it will be easy to extend the code we have produced to use this

---

instead as we already compute the entropies. However, it is no longer invariant to marginal transformations and ideally, the entropy is 0 for marginals as are transformed to be uniform. However if instead we compute the entropy for the marginals without transforming, we can use the constructed framework in extension to calculate the mutual information and from that the normalized mutual information

## 1.2 Further perspectives