# Abstract

This manuscript deals with the study of epidemic inference in the framework of Statistical Physics. Epidemics are treated as stochastic processes on graphs. The inference task consists in a probabilistic reconstruction of a specific epidemic cascade (so called *planted*) using partial and noisy knowledge of the contact network and of individuals' infection state. The reconstruction process is reframed in this thesis as the computation of observables over a high-dimensional probability distribution, known as the *posterior*. In the Introduction, connections are drawn between posterior computation and Statistical Physics. Specifically, a parallel is portrayed between inference and spin glass theory. Special attention is given to the Nishimori conditions, which play a central role in both Spin Glass theory and (epidemic) inference. The computation of the epidemic posterior marginals is shown to be an NP-hard problem (as shown in the manuscript). Thus, some approximate methods are required. The Causal Variational Approach is introduced for this purpose. It allows sampling without rejection from a distribution which approximates the posterior. This method surpasses previously existing machine learning-based techniques, as well as some Mean-Field approximations, in terms of accuracy. An attempt to characterize the difficulty of inference tasks involves computing theoretical bounds on algorithmic performance as functions of epidemic parameters. This is the objective of *Epidemble*, introduced in Chapter 3 of this manuscript. Epidemble (Epidemic Ensemble) is a semi-analytical tool based on the Replica Symmetric Cavity Method. This technique allows to compute, in the limit of large-sized graphs, what a perfect (exact) algorithm would find. In particular, Epidemble finds the values of statistical estimators (e.g., Area Under the ROC, Minimum Mean Squared Error, Maximum Mean Overlap) as functions of epidemic parameters such as infection rate, patient zero density, and the quantity and quality of clinical tests. These results are provided in the form of phase diagrams which can be interpreted as upper bounds to real inference alogrithms' performances.