This is a note for the paper: Bayesian Semi-supervised Learning with Graph Gaussian Process (2018 NIPS)

Define a Gaussian Process on graph: any subset of nodes is jointly Gaussian distributed. The covariance kernel function is hand-crafted. A label of a node is decided by a hidden variable which is the mean of values of f(pdf of nodes) over 1-hop neighbourhood. The probability of p(y|h) is given by robust-max likelihood(???). We use a set of inducing random variables to refresh the posterior kernel function. The inducing random varibles are decided by labeled nodes. Scalable variational inference for GP are leveraged here which I do not understand.