We propose to use the natural gradient descent to optimize
graph neural networks. The GNN is a generalization of the
NN to the graph domain. The GCN is a GNN
with a linear approximation followed by a nonlinear activation function.
The Hessian is the matrix of second-order partial derivatives of
the loss function. The Fisher information matrix is the inverse
of the empirical Hessian. The results are very promising. The
Adam algorithm is the best of the four, and the
SGD-KFAC algorithm is the worst. The Adam-KFAC algorithm is the
best of the two-layer algorithms.