# Supplementary material
# HypBench: Hyperbolic Benchmark for Graph Neural Network Performance

Roya Aliakbarisani, Robert Jankowski, M. Ángeles Serrano, Marián Boguñá

## S.1 Degree distribution and clustering control in Geometric Network Model

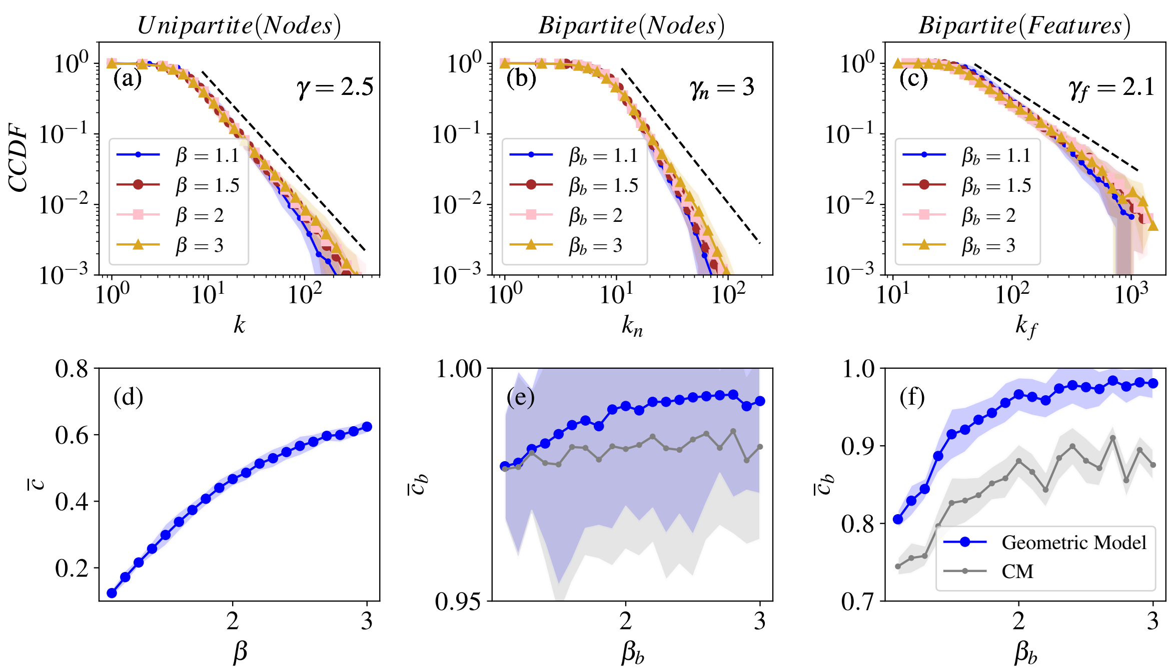


Fig. S1: Independent control of degree distribution and clustering coefficient in synthetic networks generated by geometric network model. (a)-(c) Complementary cumulative degree distribution of nodes in $\mathcal{G}_n$ ($N = 2000$, $\gamma = 2.5$, and $\langle k \rangle = 10$) and those of nodes and features in $\mathcal{G}_{n,f}$ ($N_n = 2000$, $N_f = 200$, $\langle k_n \rangle = 10$, $\gamma_n = 3$, and $\gamma_f = 2.1$) for different values of $\beta$ and $\beta_b$. The dashed black lines are guides for the eyes corresponding to the power-law exponents. (d)-(e) Clustering and bipartite clustering coefficients as a function of $\beta$ and $\beta_b$. The gray curves in (e) and (f) illustrate the bipartite clustering in the randomized versions of the bipartite synthetic networks, generated according to the configuration model (CM).

## S.2 Hyperparameters of the machine learning models

TABLE 1: Hyperparameters of the machine learning models.

| Model | Dropout | Weight Decay | Optimizer | Activation | Layers | Dimensions |
|---|---|---|---|---|---|---|
| MLP | 0.2 | 0.001 | Adam | - | 2 | 16 |
| HNN | 0.2 | 0.001 | Adam | - | 2 | 16 |
| GCN | 0.2 | 0.0005 | Adam | ReLU | 2 | 16 |
| GAT | 0.2 | 0.0005 | Adam | ReLU | 2 | 16 |
| HGCN | 0.2 | 0.0005 | Adam | ReLU | 2 | 16 |

## S.3 Empirical validation of Geometric Network Model

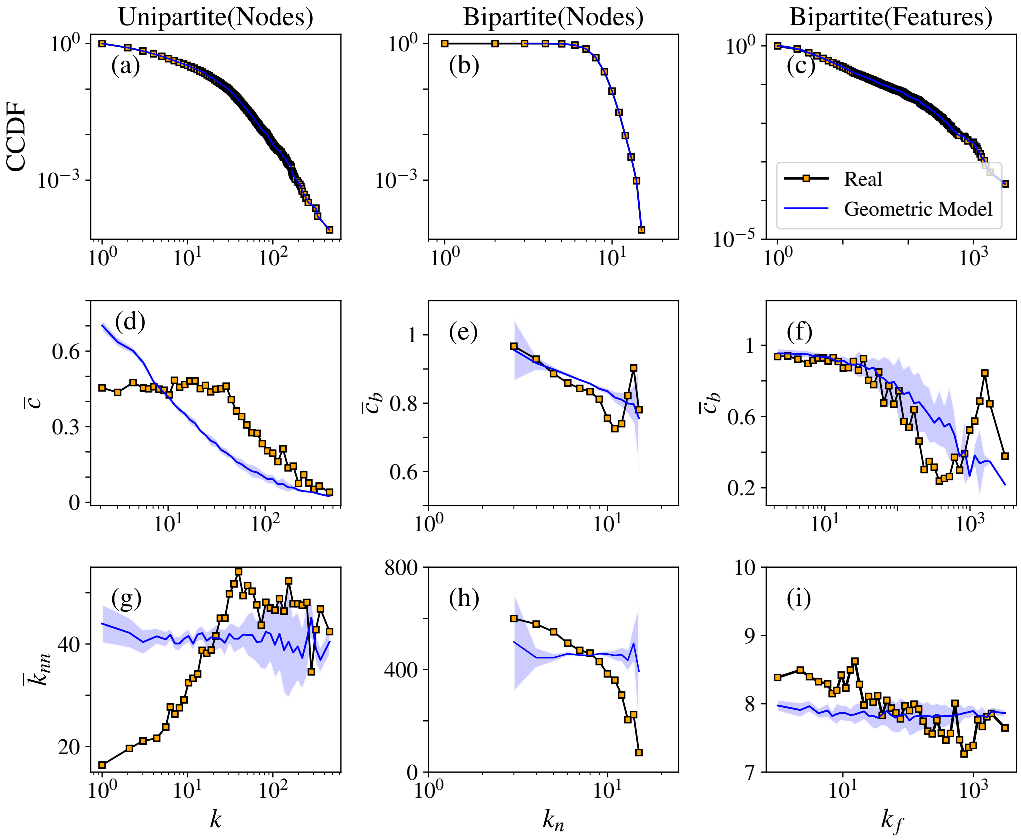


Fig. S2: Topological properties of Facebook dataset and its synthetic counterpart generated by geometric network model. (a)-(c) Complementary cumulative degree distribution of nodes in $\mathcal{G}_n$ and those of nodes and features in $\mathcal{G}_{n,f}$. (d)-(f) clustering spectrum of nodes as a function of node degrees in $\mathcal{G}_n$ and bipartite clustering sepctrum of nodes and features as a function of nodes and features degrees in $\mathcal{G}_{n,f}$. (g)-(i) average nearest neighbors degree function in $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$. The clustering coefficient, $c$, for each node in $\mathcal{G}_n$ is calculated as the fraction of connected pairs among its neighboring nodes relative to the total possible pairs of neighbors. To compute the bipartite clustering, $c_b$, for a given node, each pair of its neighboring features is considered connected if they share at least one common node, apart from the node in question. The clustering coefficient is then computed similarly to the clustering coefficient in a unipartite network. This definition also applies to the bipartite clustering coefficient of features. Finally, $k_{nn}$ in both $\mathcal{G}_n$ and $\mathcal{G}_{n,f}$ quantifies the average degree of the neighbors for a given node or feature. Exponential binning is used in the computation $\overline{c}$, $\overline{c}_b$ and $\overline{k}_{nn}$. The blue shaded region indicates the two-$\sigma$ intervals around the mean, derived from 100 realizations of geometric network model.

## S.4 Exploring the parameters' space

Here, we explore the impact of parameters within the HypBench framework including the parameters in $\mathbb{S}^1/\mathbb{H}^2$ and bipartite-$\mathbb{S}^1/\mathbb{H}^2$ models, as well as the combinations of both on the performance of graph machine learning techniques.

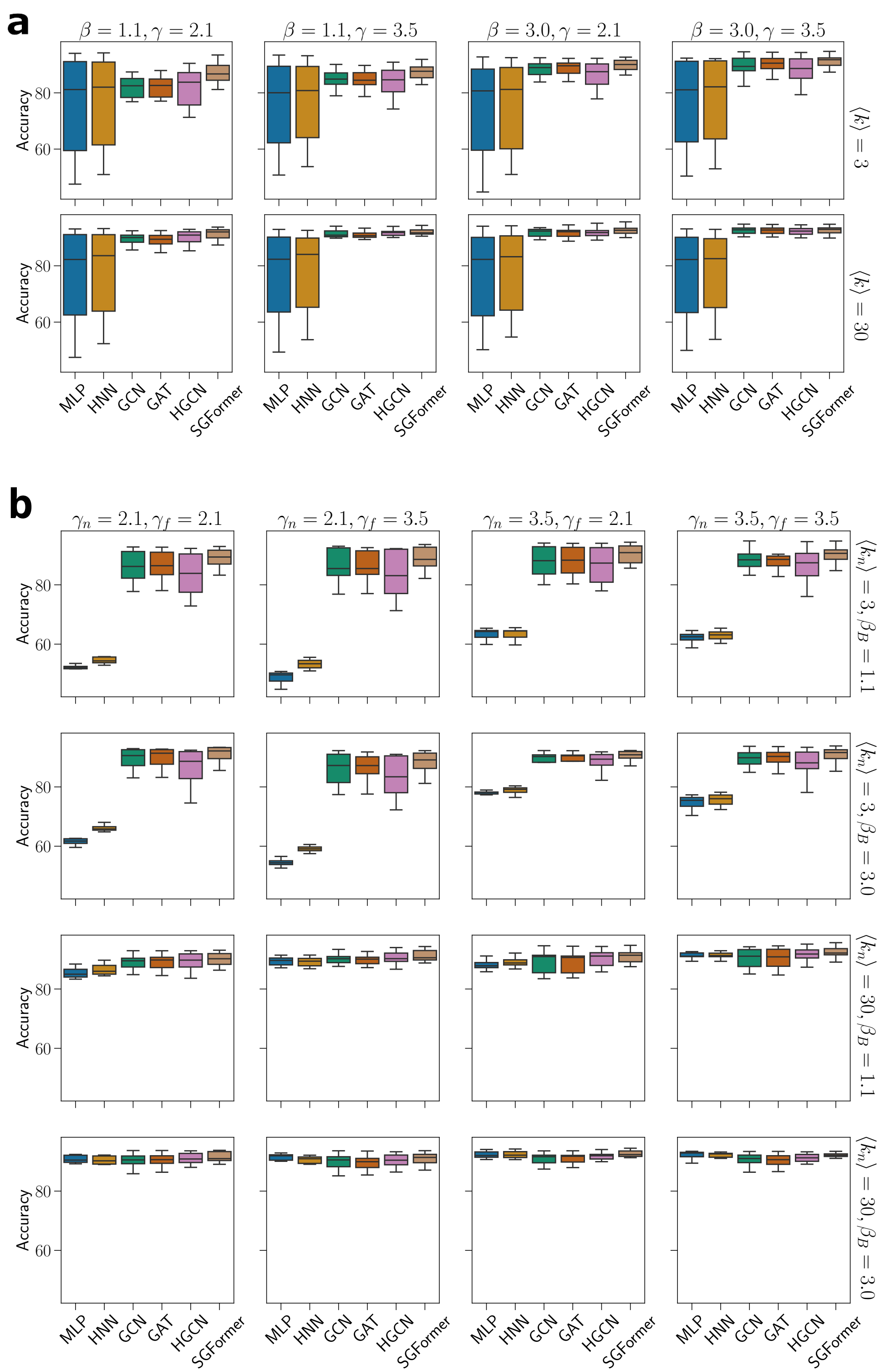


Fig. S3: Performance comparison of machine learning models on NC task with respect to (a) $\mathbb{S}^1/\mathbb{H}^2$ model parameters and (b) bipartite-$\mathbb{S}^1/\mathbb{H}^2$ model parameters. We set $\mathcal{N}_L = 6$ and $\alpha = 10$.

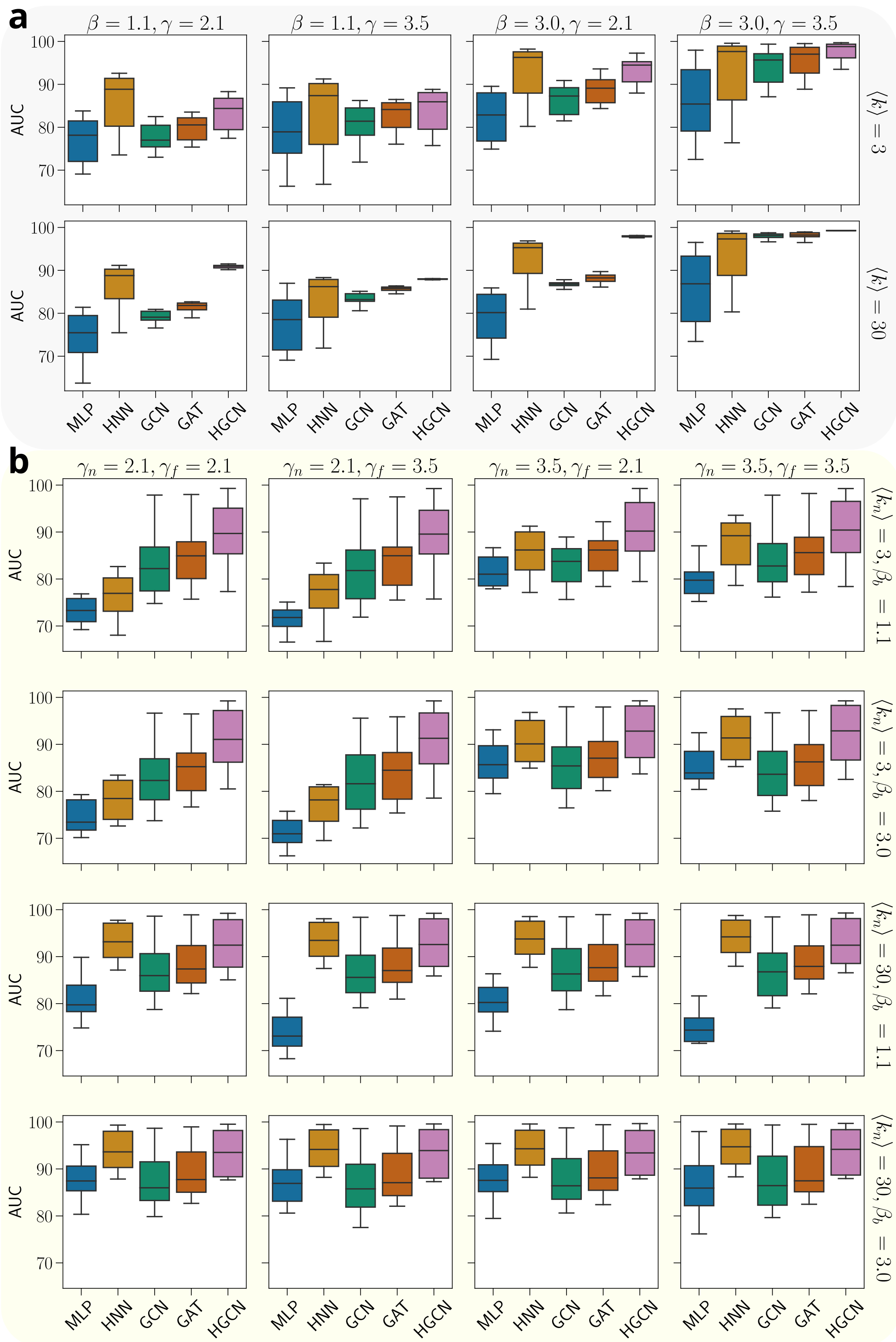


Fig. S4: Performance comparison of machine learning models on LP task with respect to (a) $\mathbb{S}^1/\mathbb{H}^2$ model parameters and (b) bipartite-$\mathbb{S}^1/\mathbb{H}^2$ model parameters.

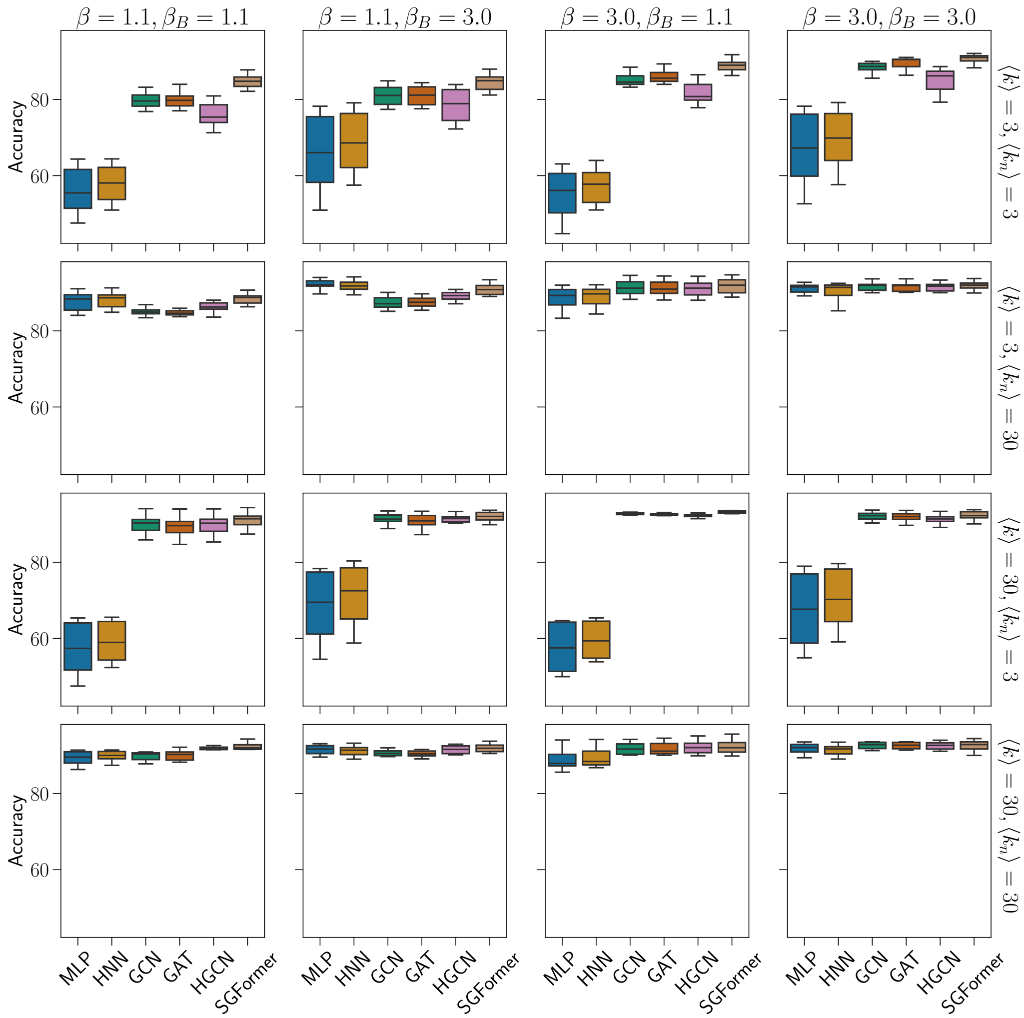


Fig. S5: Performance comparison of machine learning models on NC task with respect to a combination of parameters in the $\mathbb{S}^1/\mathbb{H}^2$ and bipartite-$\mathbb{S}^1/\mathbb{H}^2$ models. We set $\mathcal{N}_L = 6$ and $\alpha = 10$.

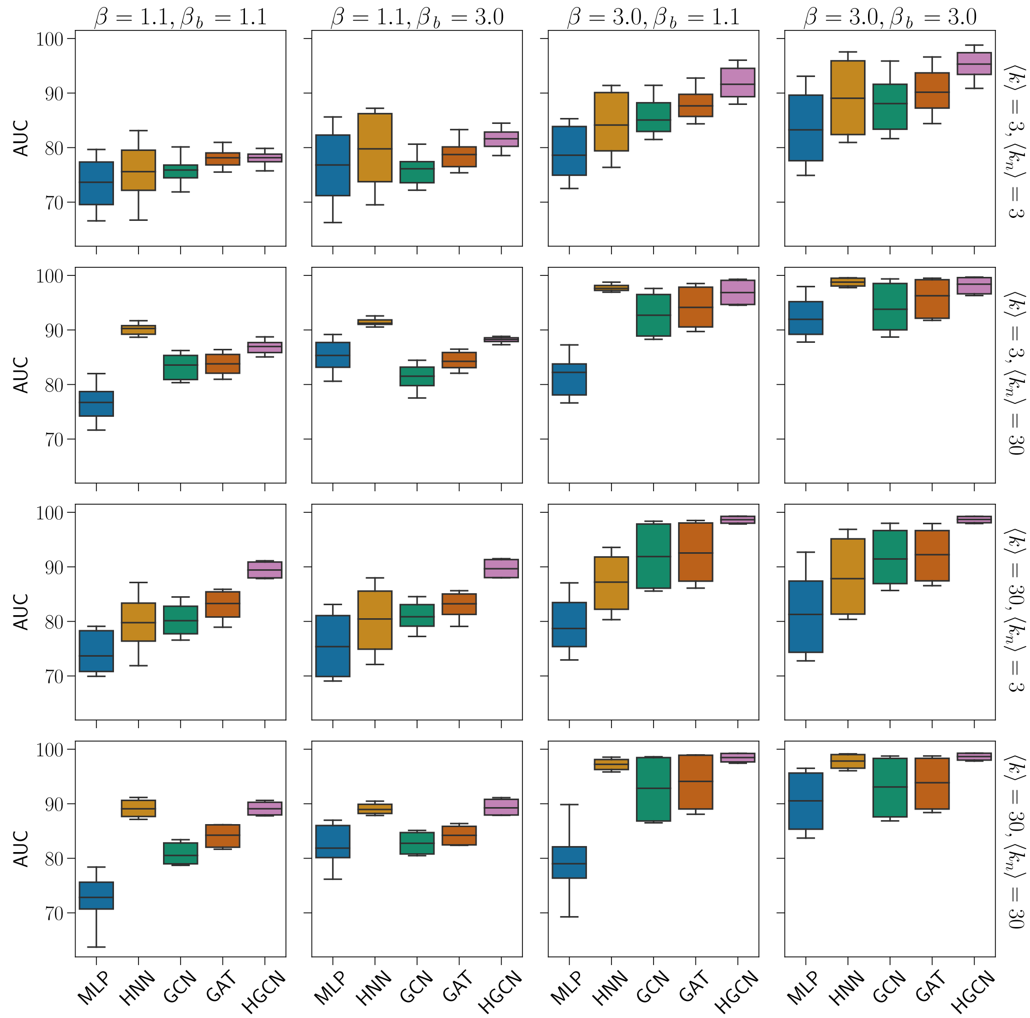


Fig. S6: Performance comparison of machine learning models on LP task with respect to a combination of parameters in the $\mathbb{S}^1/\mathbb{H}^2$ and bipartite-$\mathbb{S}^1/\mathbb{H}^2$ models.

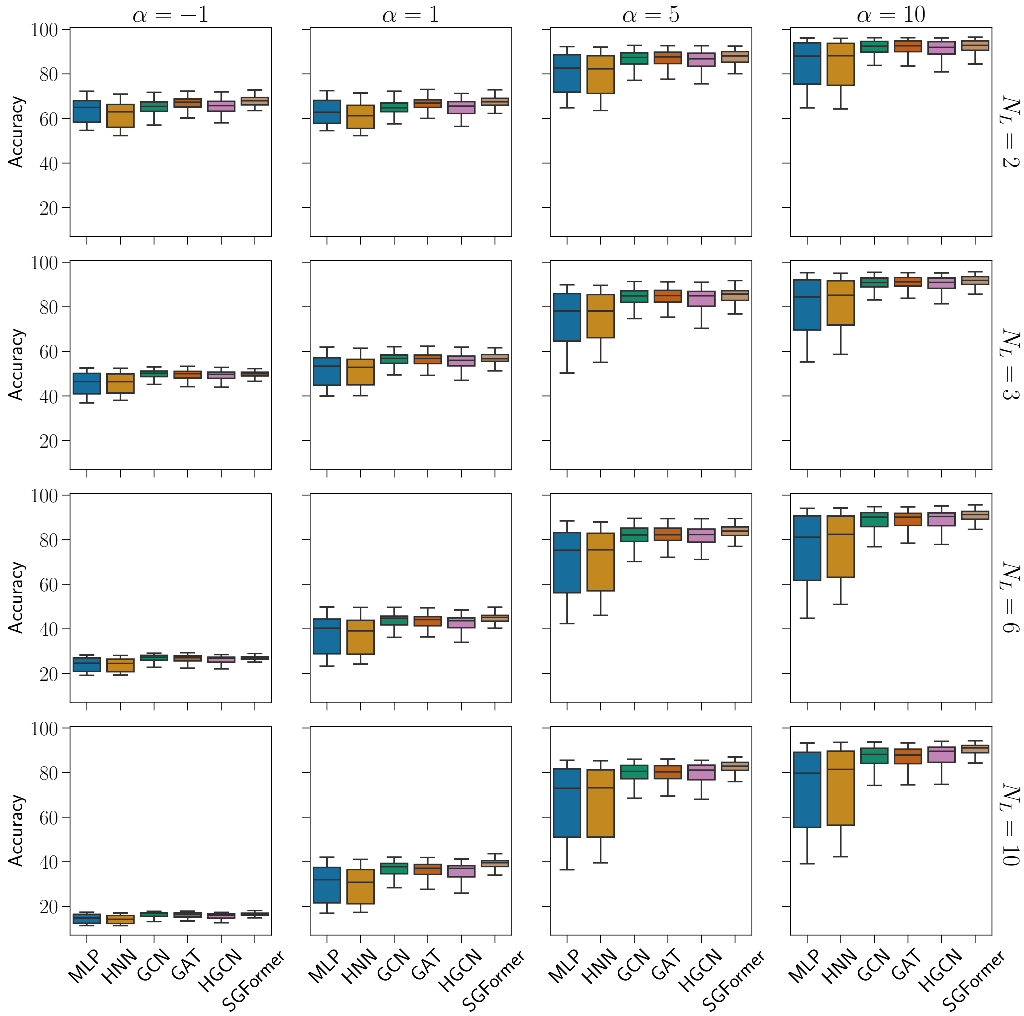


Fig. S7: Performance comparison of machine learning models on the NC task with respect to a combination of $\mathcal{N}_l$ and $\alpha$ parameters averaged over all other parameters.

## S.5 Time complexity analysis

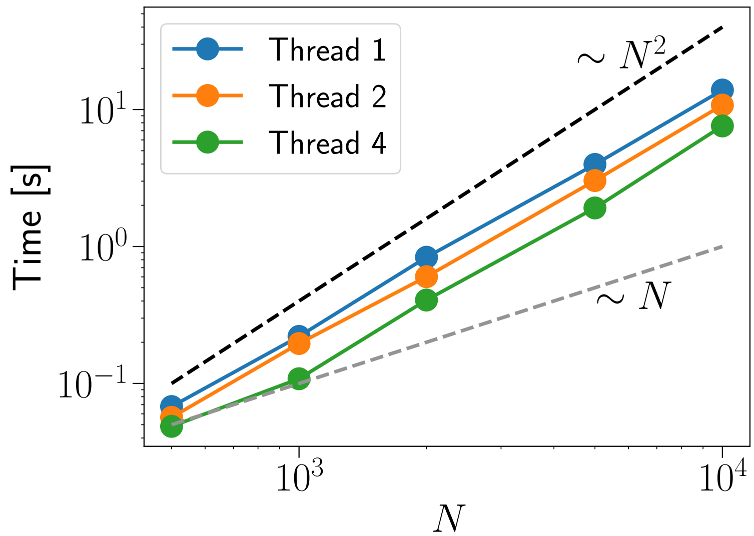


Fig. S8: Time complexity analysis. We generated synthetic networks from the geometric soft configuration model. We simultaneously increased the size of the network and the feature vector. The remaining parameters are: $(\beta, \gamma, \langle k \rangle, \beta_b, \gamma_n, \gamma_f, \langle k_n \rangle)$ = $(3, 2.7, 10, 3, 2.7, 2.7, 10)$. The results are averaged over 50 realizations. Simulations were conducted on an Intel i7-7700K (8 cores, 4.5 GHz) with 16GB of RAM.