1 PARAMETER SETTING

The definitions of the parameters can be found in their original papers.

- MTransE: $\lambda = 0.01$, $\alpha = 5$, $k = 75$, $epoch = 1000$.
- KECG: $K_1 = 25$, $K_2 = 2$, $\lambda = 0.005$, $\gamma_1 = 3.0$, $\gamma_2 = 3.0$, dimension set to 128, epoch set to 1000.
- MuGNN: $\gamma_1 = 1.0$, $\gamma_2 = 1.0$, $\gamma_r = 0.12$, embedding size set to 128, learning rate set to 0.001, L2 set to 0.01, dropout set to 0.2, epoch set to 500.
- RSNs: $\alpha = 0.9$, $\beta = 0.9$, learning rate set to 0.003, embedding size set to 256, batch size set to 512, length set to 15.
- AliNet: $\lambda = 1.5$, $\alpha_1 = 0.1$, $\alpha_2 = 0.01$, $L = 2$, $k = 2$, learning rate set to 0.001.
- MRAEA: $d = 100$, $\lambda = 3$, $K = 2$, $l = 2$, drop out rate set to 0.3, learning rate set to 0.005.
- TransEdge: $\gamma_1 = 0.2$, $\gamma_2 = 2.0$, $\alpha = 0.8$, $s = 0.7$, $d = 75$, learning rate set to 0.01, early stopping.
- SSP: $\beta_1 = 0.2$, $\beta_2 = 2.0$, $\alpha = 0.8$, embedding size set to 300.
- RREA: $d = 100$, $\lambda = 3$, $l = 2$, $\mu = 0.3$, $\gamma = 0.005$.
- JAPE: $d = 75$, $\alpha = 0.1$, $\beta = 0.05$, $\delta = 0.05$. For SE, learning rate is set to 0.01, early stopping; For AE, learning rate is set to 0.1, epochs set to 100.
- GCN-Align: $d_s = 300$, $d_a = 100$, $\gamma_s = 3$, $\gamma_a = 3$, $\beta = 0.9$.
- RDGCN: $\beta_1 = 0.1$, $\beta_2 = 0.3$, $\gamma = 1.0$, $d = 300$, $d' = 600$, $\tilde{d} = 300$, $\kappa = 125$, learning rate set to 0.001.
- HGCN: $\gamma = 1$, $\beta = 20$, $\kappa = 125$, learning rate set to 0.001.
- NMN: $\beta = 0.1$, $\gamma = 1.0$, embedding size set to 300.
- DAT: $d_s = 300$, $d_g = 300$, $\gamma = 0.8$, $\theta = 0.05$, $\zeta = 0.05$, learning rate set to 0.1.
- RNM: $\gamma = 1$, $\delta_e = 5$, $\delta_r = 3$, $\lambda = 0.001$, $\lambda_e = 10$, $\lambda_r = 200$, embedding size set to 300, learning rate set to 0.001.
- CEAFF: $\epsilon = 3$, $\theta_1 = 0.99$, $\theta_2 = 0.48$, $\gamma = 0.9$, $\alpha = 0.001$, $\beta = 0.01$, $\tau = 10$, embedding size set to 300.

For the rest of the methods, since we do not have access to their source codes and our implemented results are much worse than those reported in the original papers, we directly adopt the results from the papers (which might be missing on some datasets) and omit the parameter settings.