

## A Network-fused Dirichlet-multinomial Model for Clustering Multivariate Count Data

The code is developed for three methods mentioned in our paper, i.e., DMNet, DMNet+ and DMNet++.

### Functions Explanation

- **func.R**  
This package contains some of the functions used.
- **simul.R**  
An example for running programs.
- **gen.R**  
This package is used to generate data.
- **plot.R**  
This package plots figures in our paper.
- **cora.R**  
A package for running real dataset Cora.
- **citeseer.R**  
A package for running real dataset CiteSeer.
- **simul**  
This function generates samples from the Dirichlet-multinomial density under different network settings.  
*Usage:* simul(true.G, Mmin=80, Mmax=120, alpha1, alpha2, graph.type = "edge", nei = NULL, prob = NULL)
- **admm.norm**  
This function computes the predictive Dirichlet-multinomial parameter and the clustering results in the DM, DM+ and DM++ model.  
*Usage:* admm.norm(w, E, Ebynode, xi, lambda, rho, method = 'DM', epsilon1 = 1.e-8, epsilon2 = 1.e-14, s = 1.e-5)
- **measures**  
This function computes the evaluation index of clustering.  
*Usage:* measure(labels)
- **compare.plot**  
Given the true parameters, this function plots the log probabilities assigned to  $D$  texts by the likelihood of DM distribution compared to that of DM+ (red) and DM++ (blue) respectively.  
*Usage:* compare.plot(w, gamma1, gamma2, gamma3)

### Publicly Available Dataset

- Citeseer dataset and Cora dataset can be downloaded from the website <https://linqs.soe.ucsc.edu/data>.