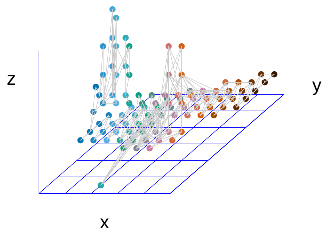
# Outline of
# "Testing independence between networks and nodal attributes via multiscale metrics"

Youjin Lee

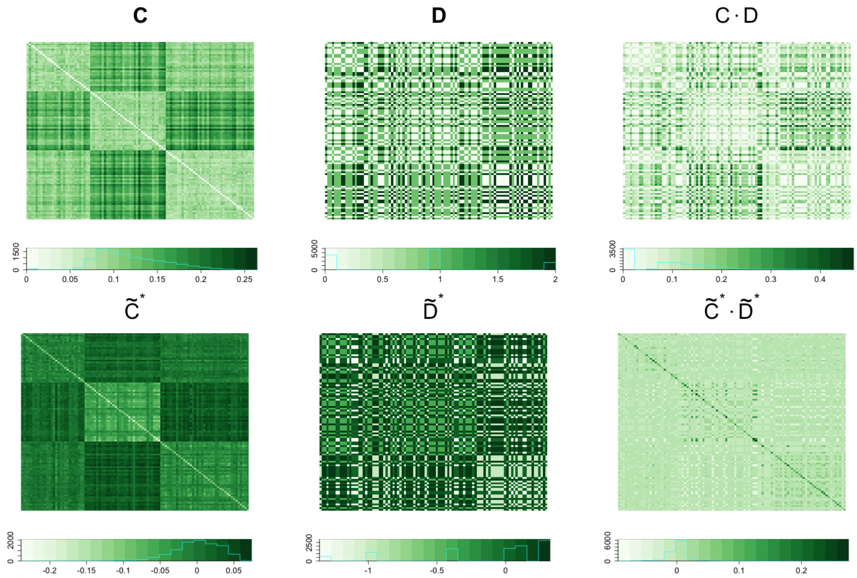
November 4, 2016

## 1. Introducing network topology



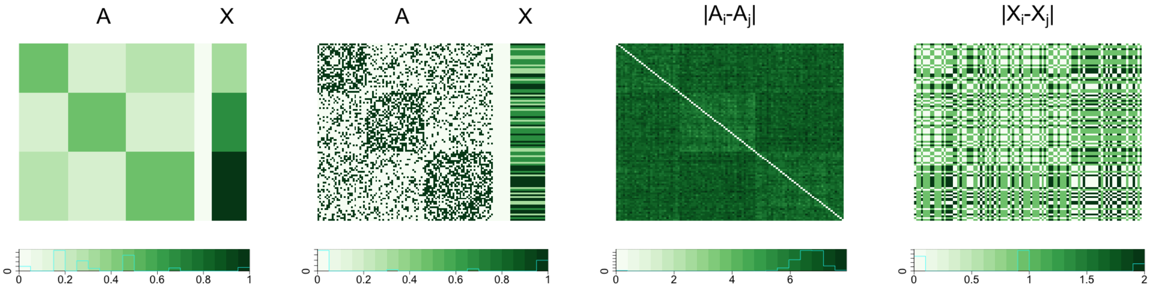

We introduce a concept of each node's location over their underlying network to bring up the problem of testing independence between distance in terms of network in distance-based test.

## 2. Multiscale Generalized Correlation



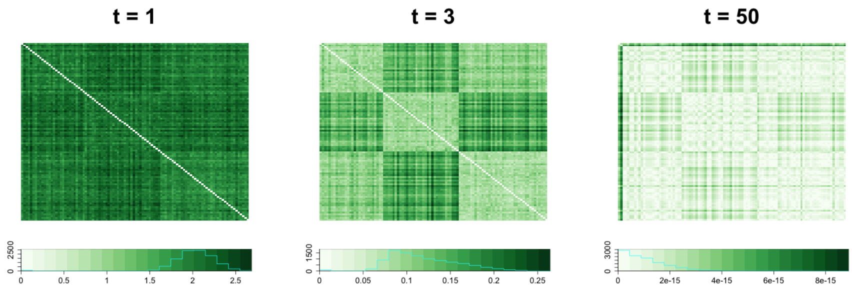

As an efficient distance-based test in existence of nonlinear dependency and high-dimensionality, we adopt using local scale distance correlation which truncates each component of distance matrices up to a certain rank.

## 3. Problems in adopting valid distance metric defined over network



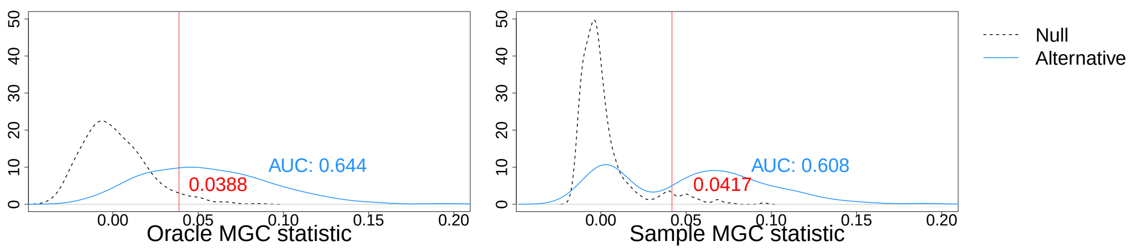

Every information on edge distribution is denoted in an adjacency matrix so we can consider its Euclidean distance matrix as an ingredient of the test statistics as well as that of nodal attributes.

## 4. demonstrate the validity of diffusion matrix



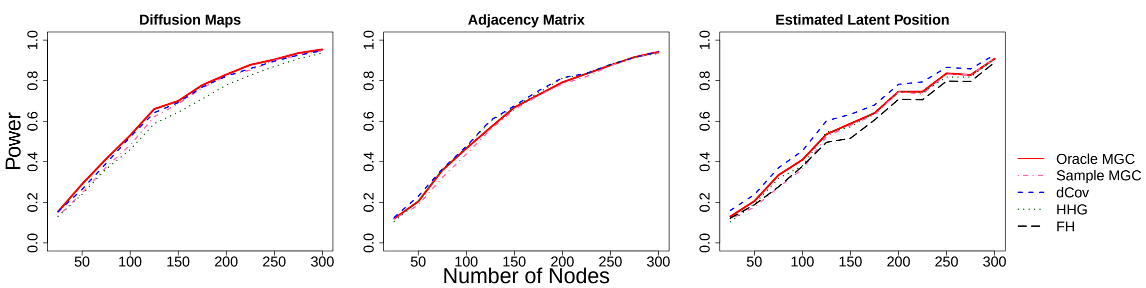

Out of concerns on theoretical and also practical shortcomings of using an adjacency matrix, we introduce a one-parameter family of network metrics called *diffusion matrix* which keeps every information of adjacent relation and also effectively captures the clustering of networks.

## 5. Introducing the simulation study



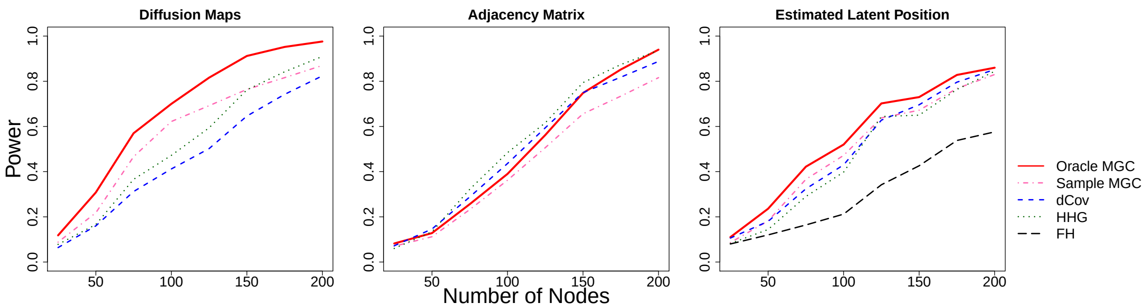

Throughout the simulation study, we are going to make a comparison between the proposed statistics from null simulated networks and also attribute-dependent simulated networks to calculate the empirical power, based on `Oracle MGC` and `Sample MGC`.

## 6. Simplest Stochastic Block Model



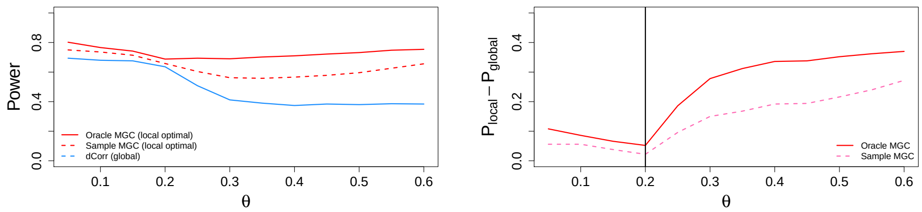

First simplest SBM with two blocks illustrates the typical case of linear-dependence so that we have similar results for each distance-based tests as well as `FH`-test.

## 7. Stochastic Block Model with non-linearly Dependent Attributes



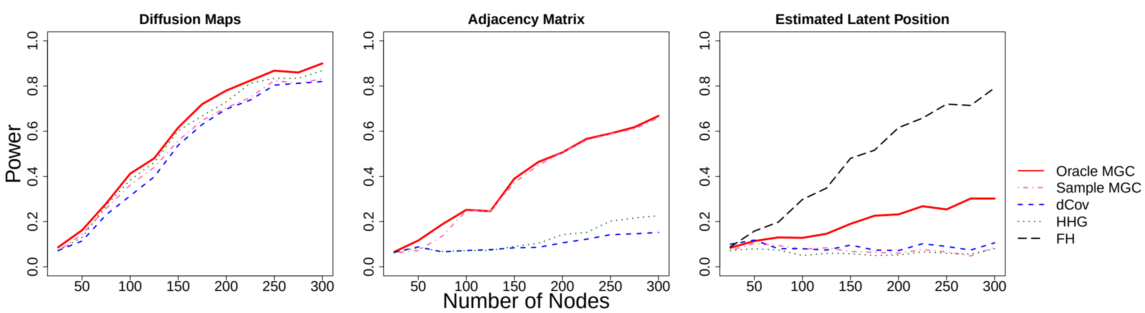

Next SBM is our punch line that MGC shows its superiority over other statistics especially in diffusion maps metrics and also results higher power in estimated latent position metric than FH.

## 8. Superiority of the proposed method under non-linear dependency



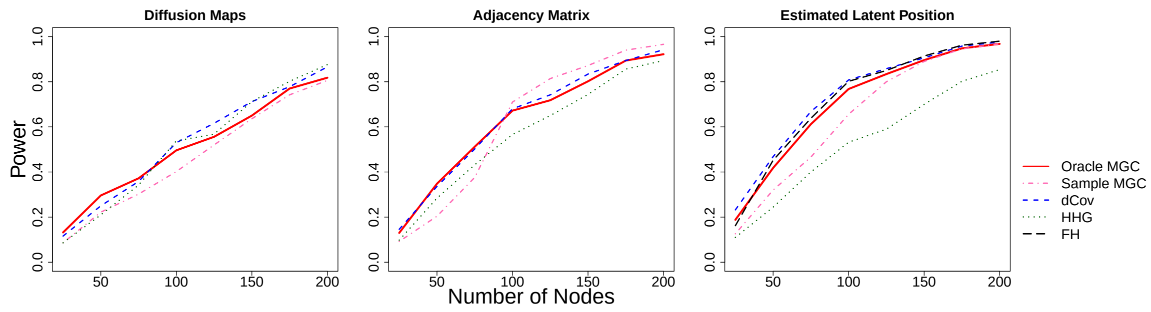

This plot deeps into when exactly our proposed tests exerts better performance; in the existence of non-linearity which can be formalized into conditional distribution of $A_{ij}$ given Euclidean distance between $\mathbf{X}_i$ and $\mathbf{X}_j$.

## 9. Degree-corrected SBM with increased variability in node distribution



Since previous two SBMs might not demonstrate the real example, we introduce a SBM but with increase variability in edge distribution and especially claim the improved power of diffusion maps when variance in an adjacency matrix is relatively higher.

## 10. Validity of the method even under competitor's model




However it would be fair to include others' mode-based tests and we can still suggest using estimated latent factors when the model is correct; but within that metrics our method does as good as their method.

## 11. Node Contribution

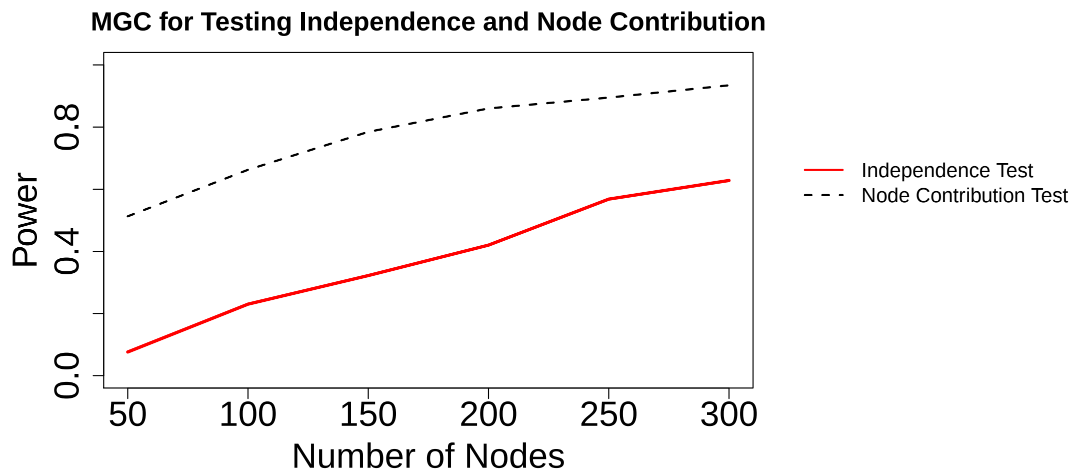



**Figure 1:** Change of empirical power and inclusion rate at each total sample size $n$. You can see that inclusion rate of $c(v)$ increases as empirical power of `MGC` increases.