Introducing network topology
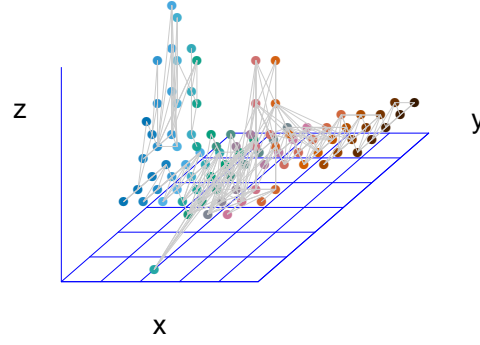

Figure 1: You can embed each subject (node) connected via edges upon Euclidean space, e.g. xyz 3D-space as above, according to their possessing attributes, e.g. physical location; while how to embed subjects from network into Euclidean space is not intuitive, which should consider their distance with respect to network relationship.

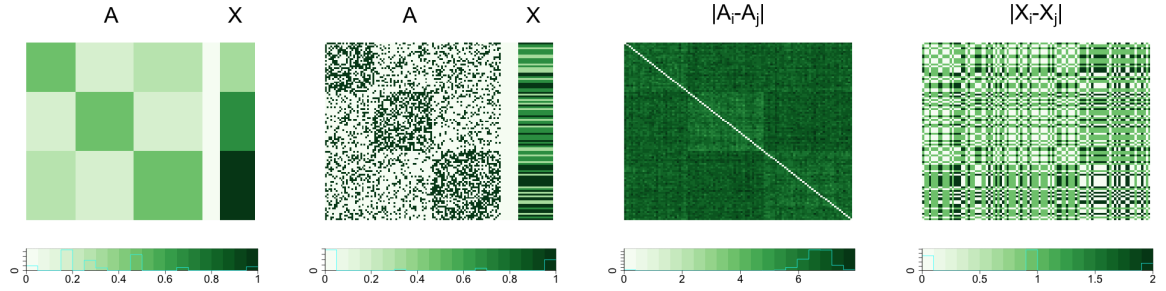# Problems in adopting valid distance metric defined over network

Figure 2: Assume that a set of edges $\{A\}$ follow certain stochastic block model, also depending on the distribution function of nodal attributes X (a), then with some amount of noise we have a realized adjacency matrix and a set of attribute outcomes (b) of which Euclidean distances (c & d) are constructed to be possibly used in standard distance-based independence test.
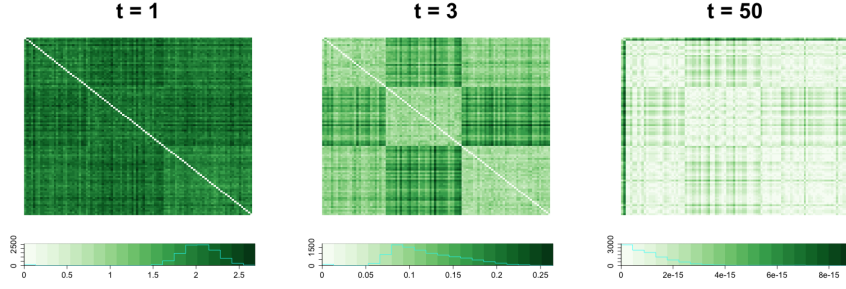
introduce a family of network distance matrices

Figure 3: *Diffusion matrix*, as a proposed alternative for Euclidean distance of A, provides one-parameter family of network-based distance where at early stage, e.g. at $t = 1$, distance matrix is very similar to Euclidean distance of A but as time goes by the pattern shown in the distance matrix changes, and our optimal time point t^* choose the diffusion time when diffusion distance matrix exhibits most dependence to distance matrix of $\mathbf{X}$.

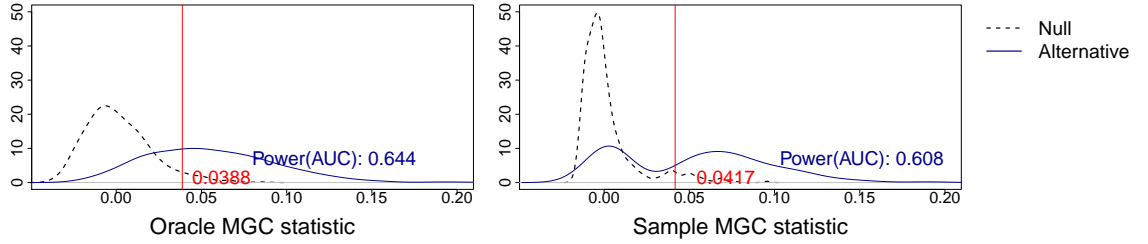# Empirical power of Oracle MGC/Sample MGC

Figure 4: In the left panel we have empirical Null distribution of `Oracle MGC` illustrated by dotted line of which 95% sample quantile determines testing power of `Oracle MGC` by calculating area under the curve (AUC) of the empirical distribution under alternative beyond that quantile, and AUC of `Oracle MGC` (0.644) looks similar to that of `Sample MGC` (0.608), as presented in the right panel, even though the shape of its distributions under null and alternative look different, which supports the use of `Sample MGC` as a substitute for `Oracle MGC` in real data.
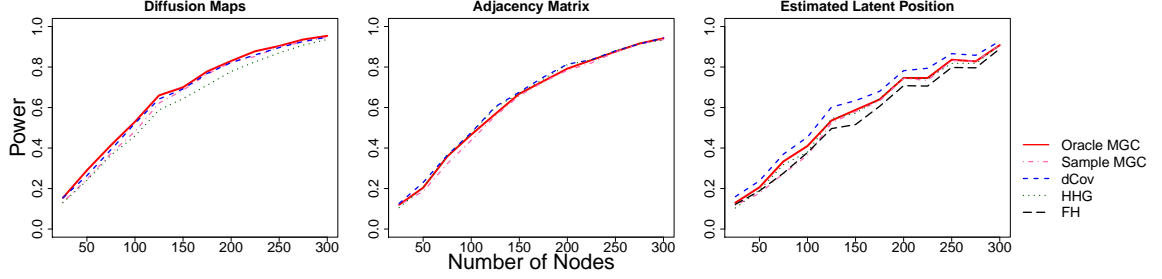
2

Simplest Stochastic Block Model

Figure 5: Above three figures are showing power of network independence test under two block SBM using diffusion maps, adjacency matrix, and estimated latent position as a network distance measure where power of `MGC` and global distance-based tests (`dCov, HHG`) are similar in each case and they are doing as good as model-based `FH` tests under latent factor metrics but still overall power under diffusion maps is the highest.
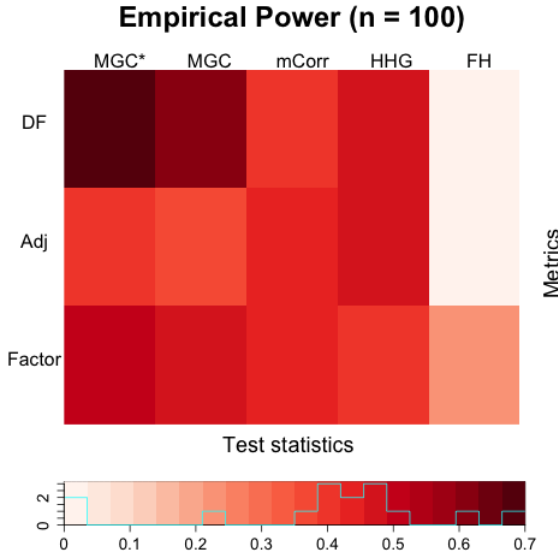
Highest power of MGC under diffusion maps

Figure 6: This power heatmap demonstrate superior power of `Oracle MGC`(MGC*) or `Sample MGC` (MGC) under diffusion distance matrix under three SBM, compared to adjacency matrix distance or latent factor distance where four distance-based tests show similar power whereas `FH` test results much less power than the others.

3

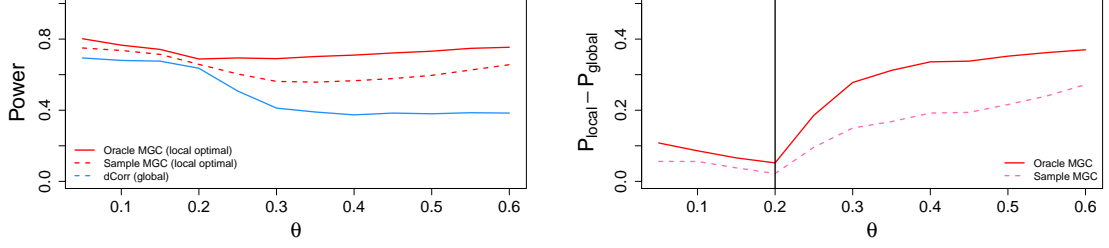# Superiority of the proposed method under non-linear dependency

Figure 7: X-axis of θ controls the existence/amount of non-linear dependency and in this particular case non-linearity exists when $\theta > 0.2$ and gets larger as it increases, and you can see the discrepancy in power between global and local scale tests also gets larger accordingly, mostly due to decreasing power of global test but relatively stable power of MGC under non-linear dependency as presented in the left panel.

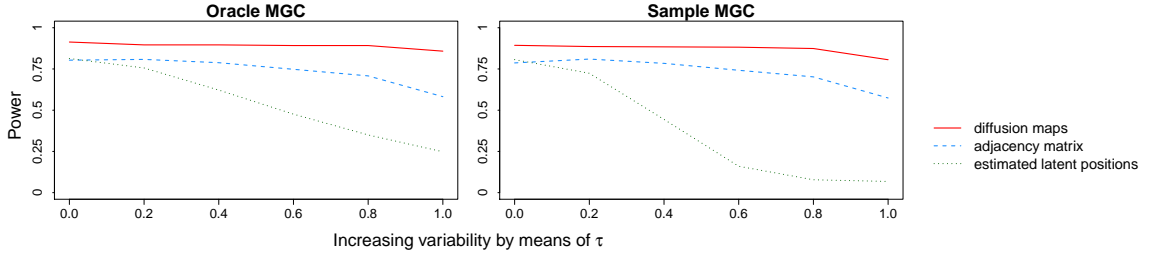# Degree-corrected SBM with increased variability in node distribution

Figure 8: In degree-corrected SBM where the variability in degree distribution increases as τ increases, power of diffusion maps are more likely to be robust against increasing variability compared to adjacency matrix and latent positions.

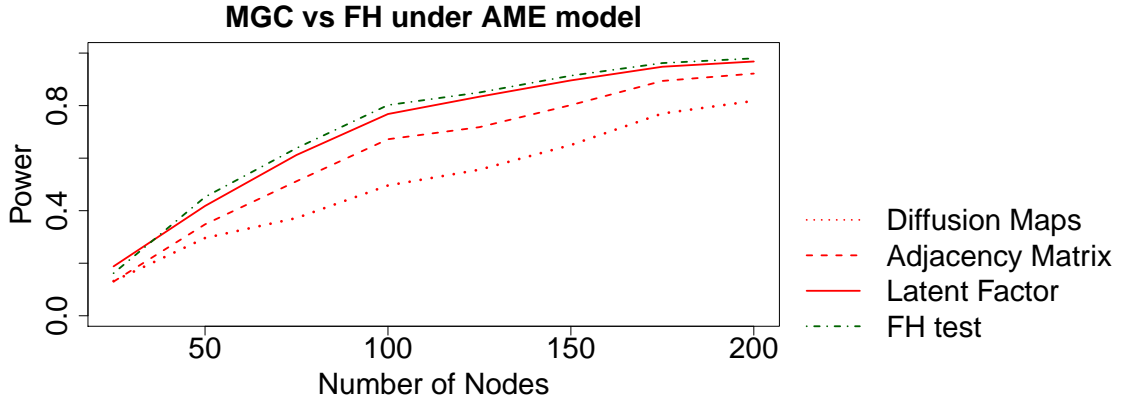# Validity of the method even under competitor's model

MGC vs FH under AME model

Figure 9: Even under additive and multiplicative model which favors estimated latent position metrics, `MGC` lost some power when using diffusion maps and adjacency matrix but `MGC` does as good as `FH` tests under latent factor metrics, which supports excellent ability of `MGC` in diverse, nearly true network metrics.

Node Contribution

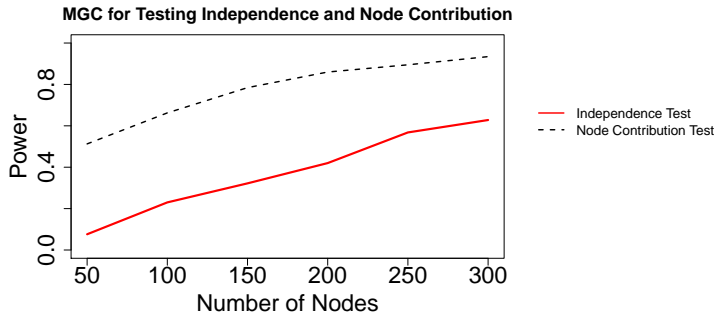### MGC for Testing Independence and Node Contribution

Figure 10: This plot describes both power of `MGC` and the rate of correctly-ranked node contribution increase as the number of nodes increases when only half of the nodes for each simulation actually are set to contribute to the independence test, which validates the use of node contribution measure in independence test.

Political Network

Collaborative networks and organization types
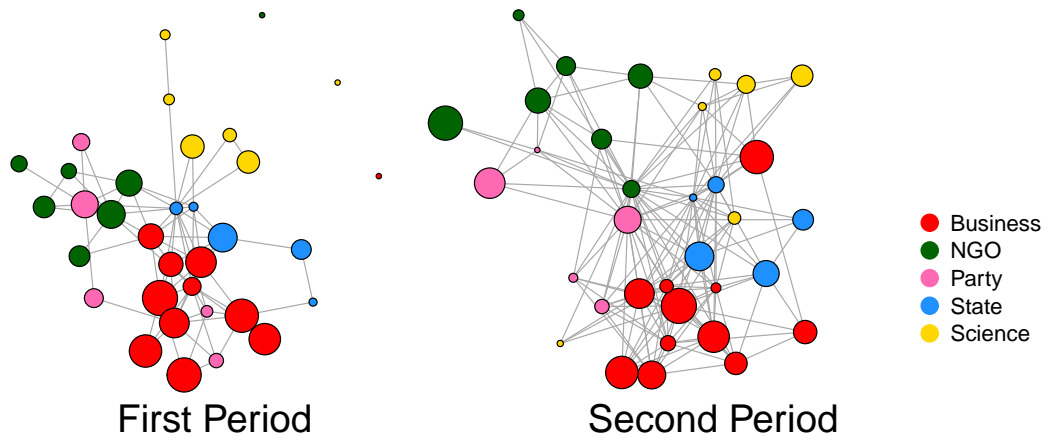

First Period   Second Period

Figure 11: Both networks depict the collaboration network during the two time periods comprised of five types of political organizations of which node size is proportional to the rank in terms of contribution to the network independence test.