Law of Large Graphs

1 Introduction

We are looking at the scenario where we have M adjacency matrices, A_m, that are sampled from the same edgewise probability matrix, therefore the number of vertices is fixed and there is known vertex correspondence. We are looking at methods that estimate the edge probability matrix P, which include the naive method of averaging over the M adjacency matrices, $\bar{A}$, as well as spectral approaches, where we are taking the spectral decomposition of $\bar{A}$.

2 Model and Notation

1. Stochastic Block Model We are considering a k-dimensional Stochastic Block Model (SBM) for an N vertex undirected hollow graph with block probability matrix $\mathbf{B}$ and a block membership probability vector ρ.

We will often consider a generalized Stochastic Block Model with latent positions distributed according to a mixture of Dirichlet random variables using parameter $r \in [0, \infty)$, where an exact SBM is obtained at $r = \infty$, and $r = 0$ leads to a uniform distribution.

Given the model parameters the edge probability matrix $P \in [0,1]^{N \times N}$ is calculated. M instances of adjacency matrices with known vertex correspondence are sampled with $A_{ij} \sim Bern(P_{ij})$. With these A_m for m in $1, ..., M$, then $\bar{A} = \frac{1}{M} \sum_{m=1}^{M} A_m$.

2. Spectral Embedding and Clustering Since P is positive and semidefinite with rank at most k, P has a spectral decomposition $P = VSV^T$, where $V \in \mathbb{R}^{N \times k}$ has orthonormal columns and S is diagonal with decreasing entries. X is then $VS^{1/2}$, and $P = XX^T$. Let $\bar{A} = U_{\bar{A}} S_{\bar{A}} U_{\bar{A}}^T$ be the full rank decomposition of $\bar{A}$. Then our estimate of X (subject to a rotation) is $\hat{X} = \hat{V}\hat{S}^{1/2}$, where $\hat{S} \in R^{dxd}$ is the diagonal matrix with the d largest magnitude eigenvalues of $\bar{A}$ and $\hat{V} \in \mathbb{R}^{N \times k}$ is the matrix with orthonormal columns of the corresponding eigenvectors.

Further, the rows $\hat{X}$ can be clustered into k clusters by minimizing mean squared error. By taking each row to be the mean of it's respective cluster, we get $\hat{X}_c$.

3. Mean Estimators We are looking at three estimators for the mean:

1. $\bar{A}$, taking the average over M samples.
2. $\hat{P} = \hat{X}\hat{X}^T$
3. $\hat{P}_c = \hat{X}_c\hat{X}_c^{\ T}$

4. Relative Efficiency We will examine the relative efficiency of the ASE vs. Naive approximations of the mean.

$$RE = \frac{Var(\hat{P})}{Var(\bar{A})}$$

3 Theory Questions

Proofs for Relative Efficiency
We would like to have a formal proof for the following formulas that were demonstrated through simulations of an exact SBM:

If k is known and d is chosen to be k, then for high N:

$$RE = \frac{2k}{N}$$

If d is selected to be greater than k

$$RE = \frac{2k + 4(d-k)}{N} = \frac{4d - 2k}{N}$$

4 Simulations

Simulations using 1000 trials for 3 different exact SBM's (with block dimensions k = 2,3,4). The diagonal was perfectly augmented with values from the P matrix
RE was calculated by taking the mean of the component-wise variances.

The following simulations use:

$$B = \begin{bmatrix} .42 & .2 \\ .2 & .7 \end{bmatrix}, \qquad \rho = \begin{bmatrix} .5 & .5 \end{bmatrix} \qquad \text{if k = 2}$$

$$B = \begin{bmatrix} .42 & .2 & .3 \\ .2 & .5 & .15 \\ .3 & .15 & .7 \end{bmatrix}, \qquad \rho = \begin{bmatrix} .33 & .33 & .34 \end{bmatrix} \qquad \text{if k = 3}$$

$$B = \begin{bmatrix} .42 & .2 & .3 & .05 \\ .2 & .49 & .35 & .15 \\ .3 & .35 & .6 & .25 \\ .05 & .15 & .25 & .7 \end{bmatrix}, \qquad \rho = \begin{bmatrix} .25 & .25 & .25 & .25 \end{bmatrix} \qquad \text{if k = 4}$$

Effect of M on RE
We can see RE is constant as a function of M for high M.

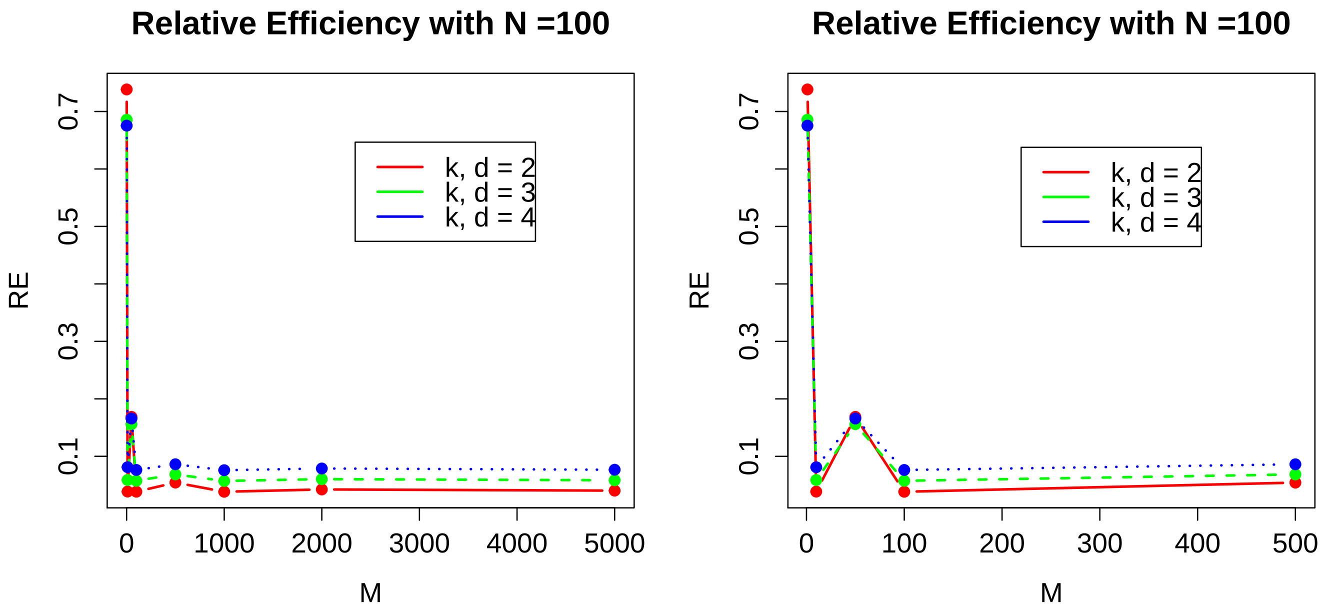

Figure 1: For high M, RE is constant as a function of M. Right image is a zoom in for low M.

As a function of N, when k and d are known, simulations indicate that: $RE = \frac{2k}{N}$

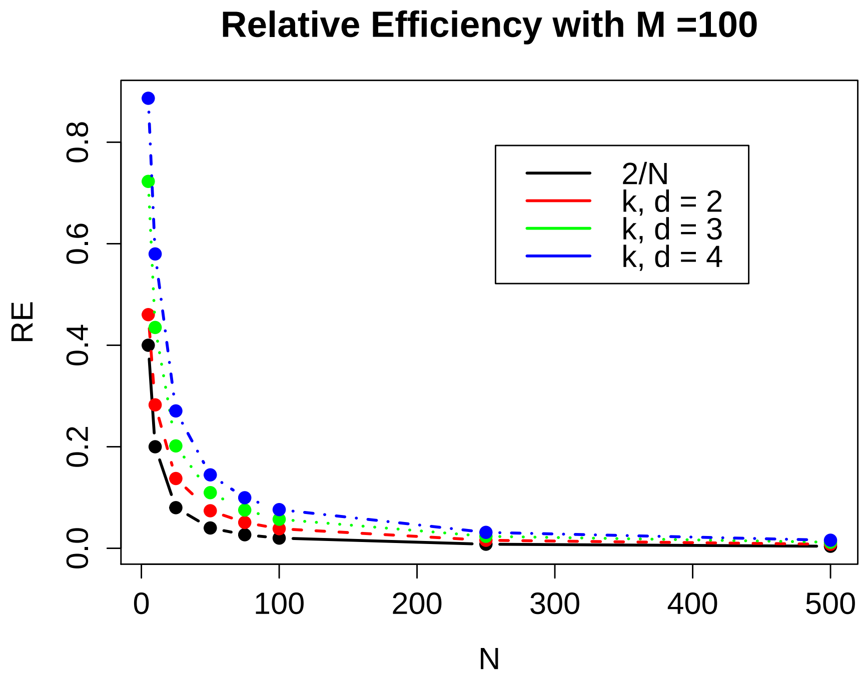

Figure 2: Relative Efficiency as a function of N using SBM's of 3 different dimensions

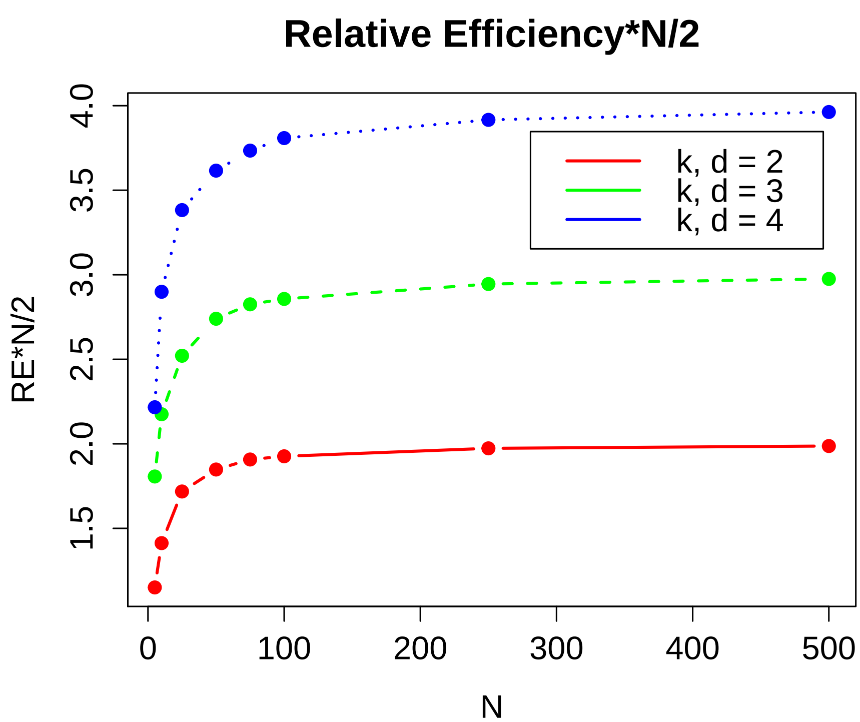

Figure 3: Plot showing that for high N, Relative Efficiency approaches 2k/n

The following simulations observe the effect when d is not known:

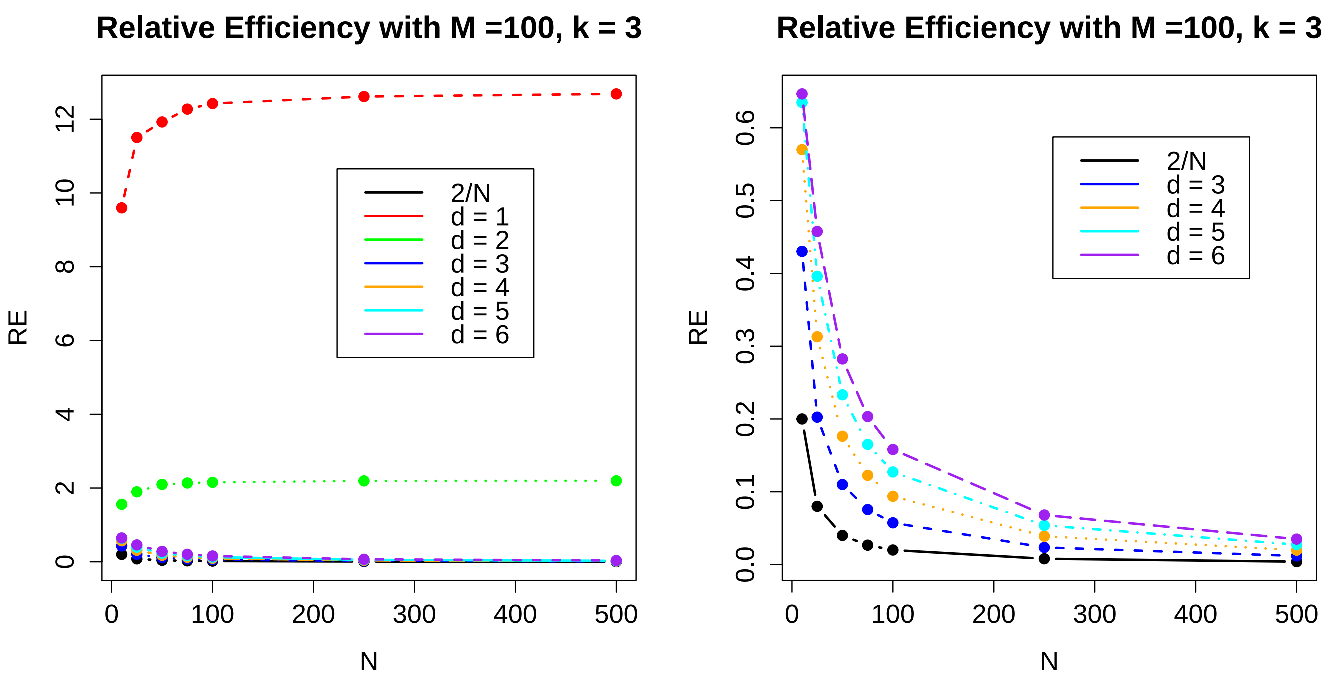

Figure 4: Effect of error in d on RE when k = 3. Right image is a zoom in for $d \geq k$

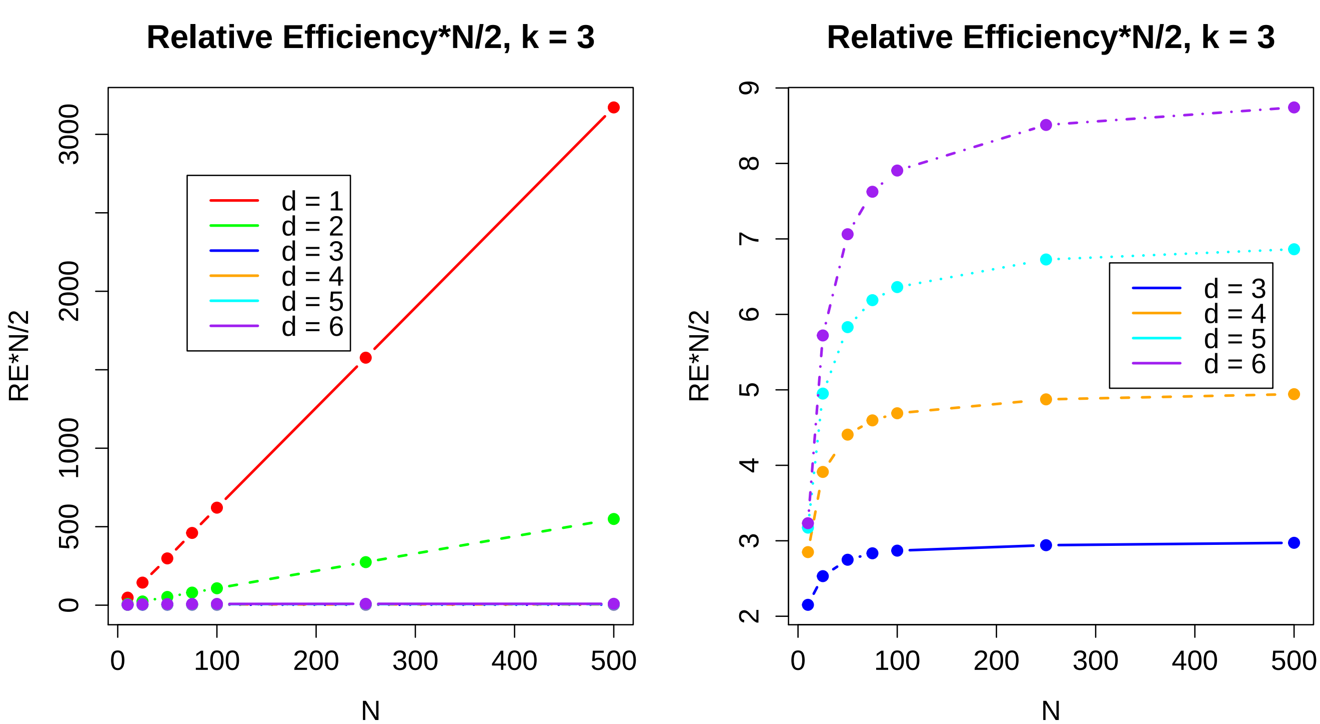

Figure 5: Effect of error in d on constant term when k = 3. Right image is a zoom in for $d \geq k$

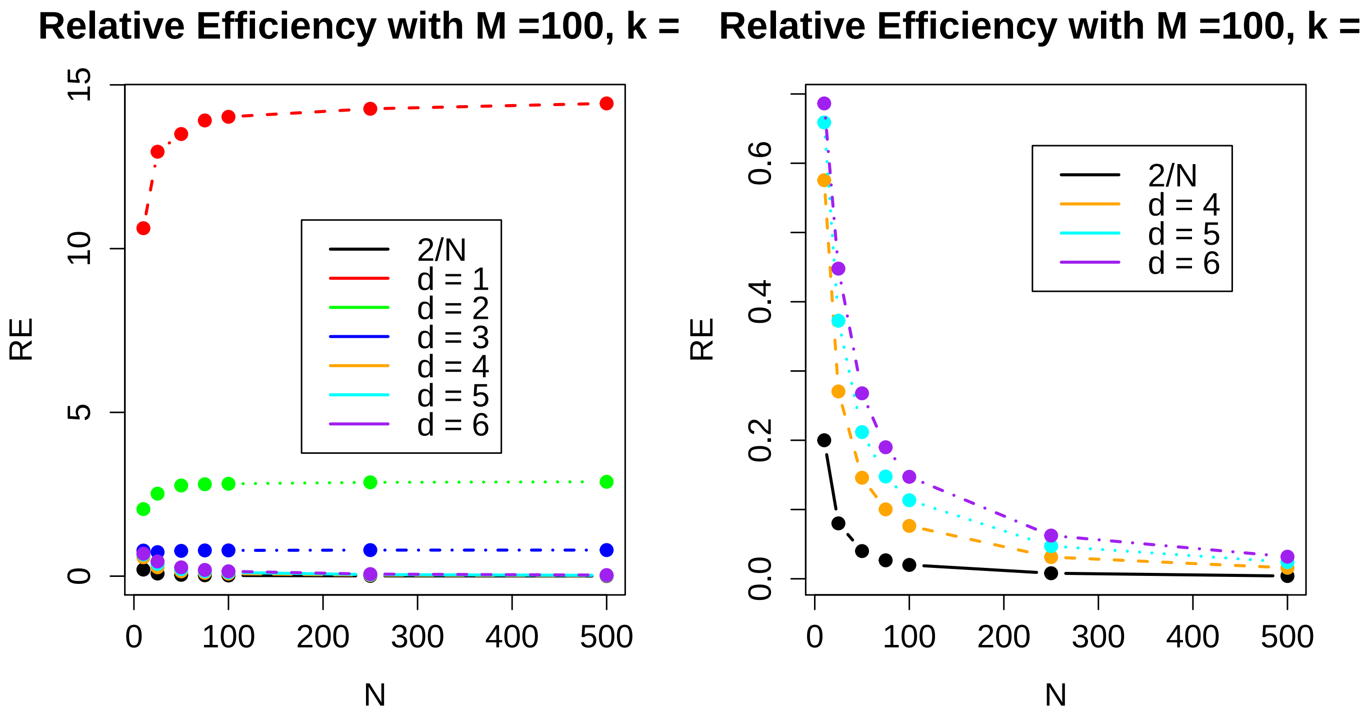

Figure 6: Effect of error in d on RE when k = 4. Right image is a zoom in for $d \geq k$

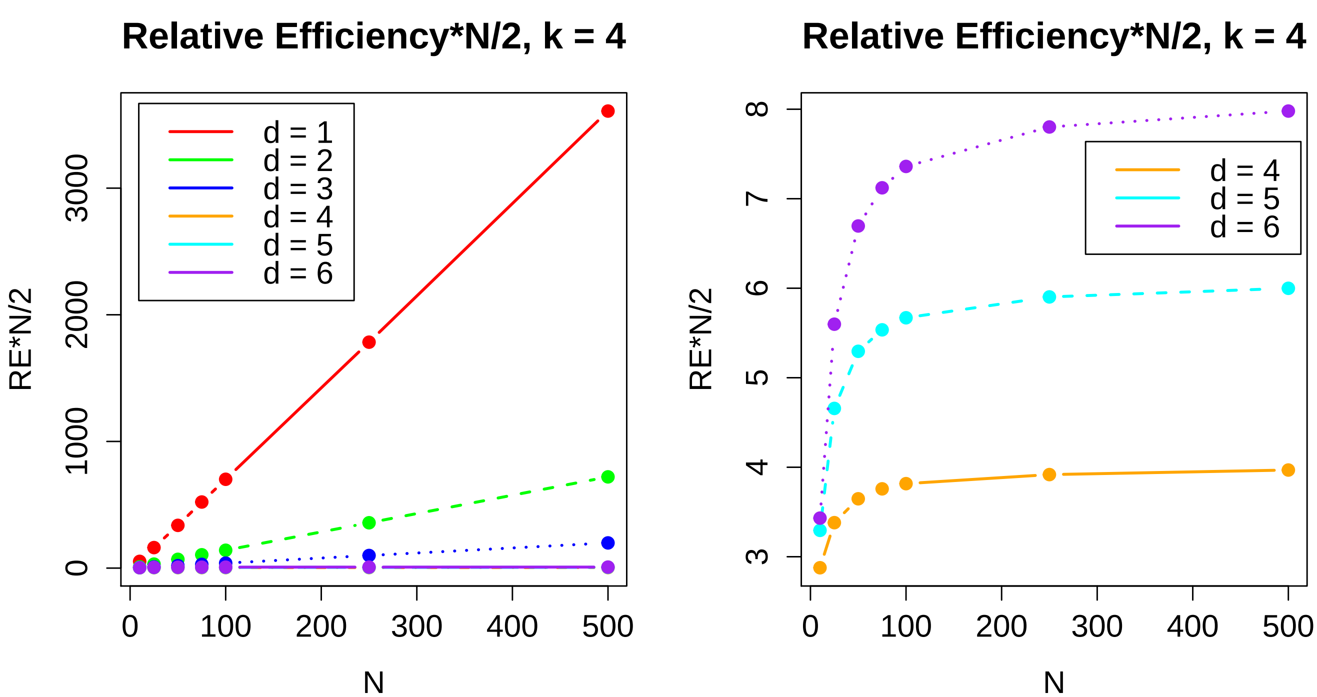

Figure 7: Effect of error in d on constant term when k = 4. Right image is a zoom in for $d \geq k$

Clearly then it is better to overestimate d than underestimate, however, when d is overestimated:

$$RE = \frac{2k + 4(d-k)}{N} = \frac{4d - 2k}{N}$$

Effect of B

The following B matrices were tested:

$$B1 = \begin{bmatrix} .42 & .2 \\ .2 & .7 \end{bmatrix},\ B2 = \begin{bmatrix} .1 & .02 \\ .02 & .07 \end{bmatrix},\ B3 = \begin{bmatrix} .9 & .2 \\ .2 & .8 \end{bmatrix},\ B4 = \begin{bmatrix} .5 & .4 \\ .4 & .6 \end{bmatrix},\ B5 = \begin{bmatrix} .02 & .005 \\ .005 & .03 \end{bmatrix}$$

$$\rho = \begin{bmatrix} .5 & .5 \end{bmatrix}$$

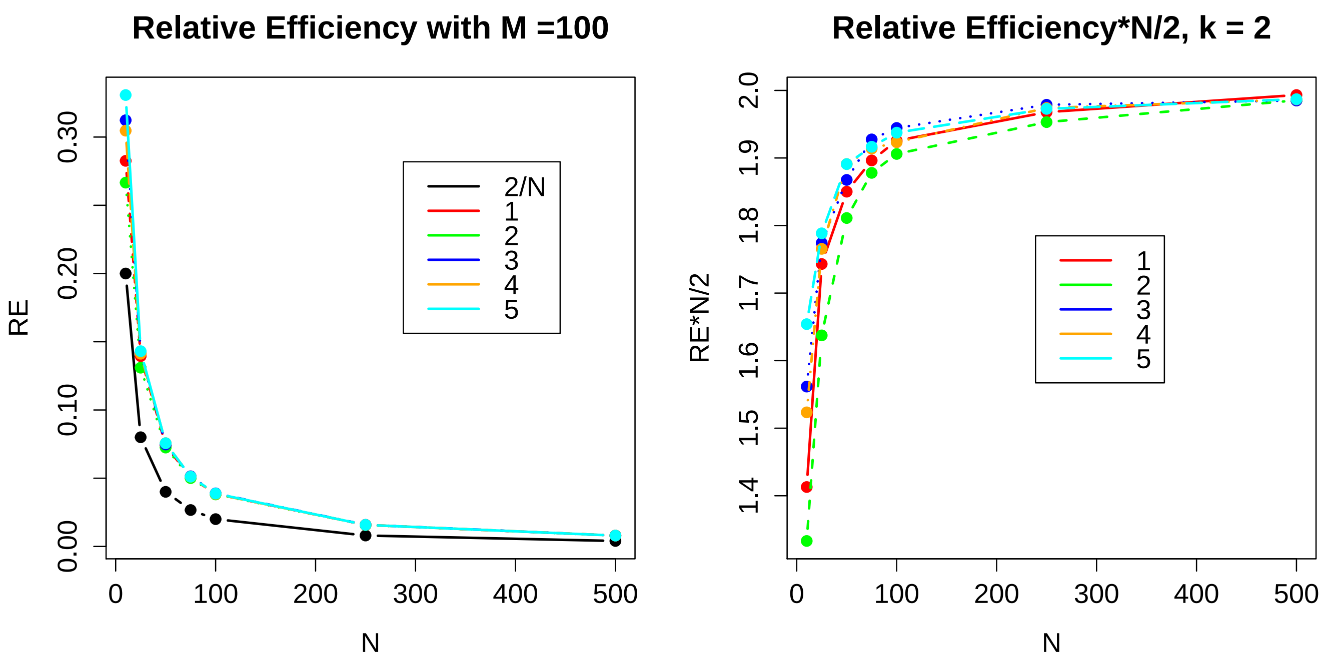

Figure 8: Effects of varying B not seen asymptotically when $\rho = [.5, .5]$.

Effect of ρ

The following ρ vectors were tested:

$$\rho = \begin{bmatrix} .5 & .5 \end{bmatrix}, \ \rho = \begin{bmatrix} .25 & .75 \end{bmatrix}, \ \rho = \begin{bmatrix} .1 & .9 \end{bmatrix}, \ \rho = \begin{bmatrix} .05 & .95 \end{bmatrix}, \ \rho = \begin{bmatrix} .01 & .99 \end{bmatrix}$$

For B:

$$B1 = \begin{bmatrix} .42 & .2 \\ .2 & .7 \end{bmatrix}, \ B2 = \begin{bmatrix} .1 & .02 \\ .02 & .07 \end{bmatrix}$$

Following results show an asymptotic effect of varying B when $\rho \neq [.5, .5]$

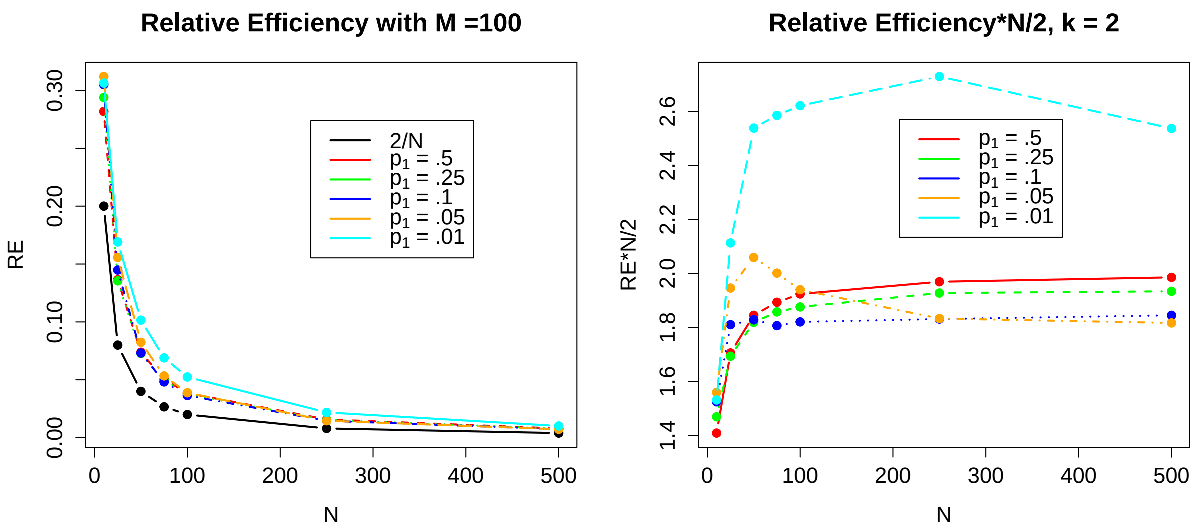

Figure 9: Effect of altering ρ for B1.

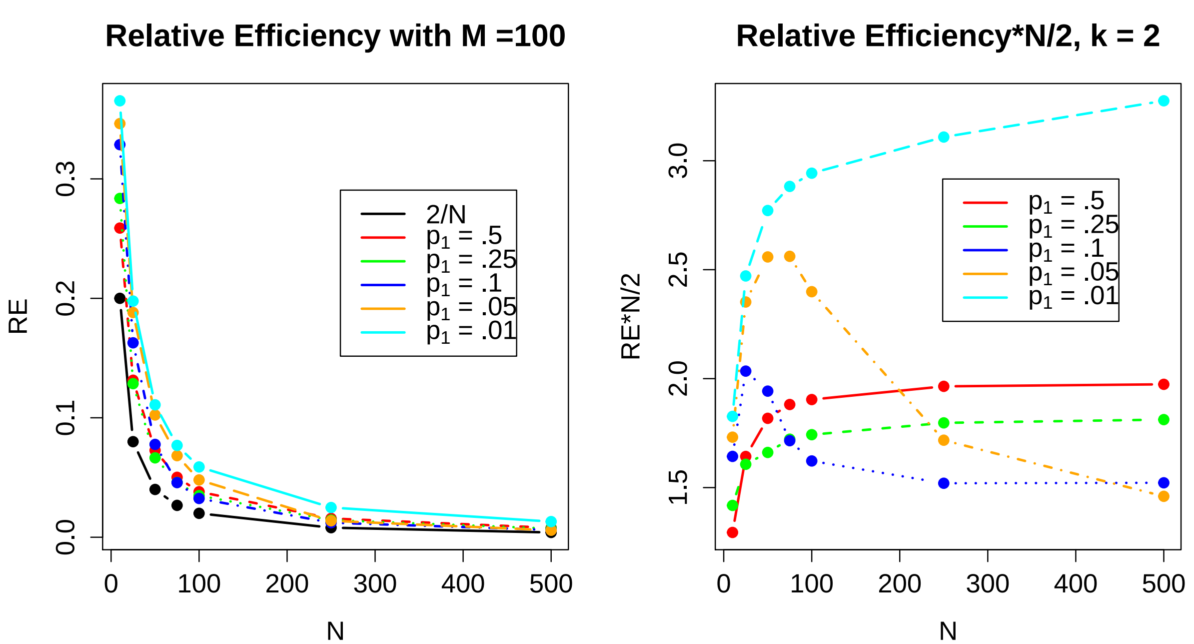

Figure 10: Effect of altering ρ for B2.

Analysis of Per Block Relative Efficiency

$$\rho = \begin{bmatrix} .5 & .5 \end{bmatrix}$$

$$B1 = \begin{bmatrix} .42 & .2 \\ .2 & .7 \end{bmatrix}, \; B2 = \begin{bmatrix} .1 & .02 \\ .02 & .07 \end{bmatrix}$$

Note that efficiency of individual blocks converge to same value when $\rho = [.5, .5]$

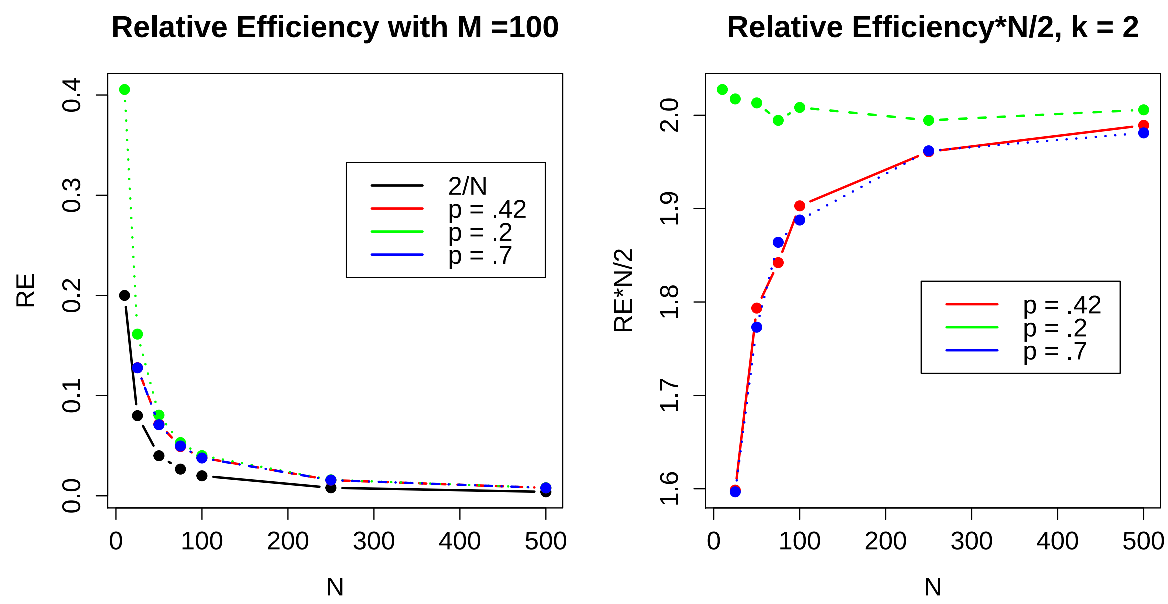

Figure 11: Per block RE for B1.

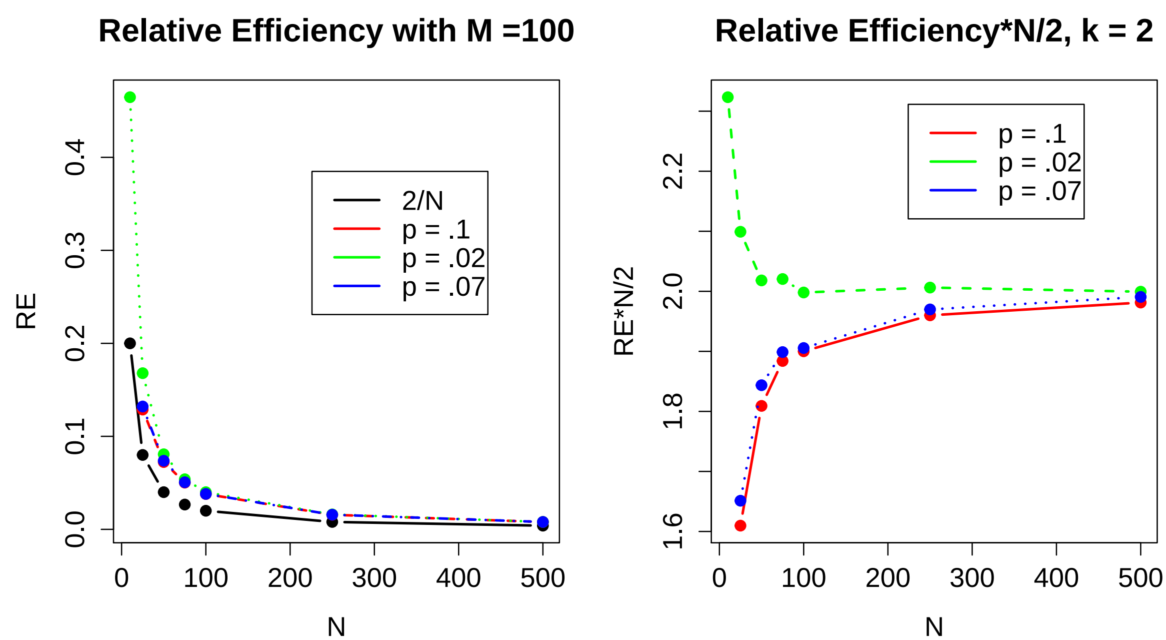

Figure 12: Per block RE for B2.

5 Proving 2k/N result

We would like to show that:

$$RE = \frac{Var(\hat{P}_{ij})}{Var(\bar{A}_{ij})} = \frac{E[(\hat{P}-P)^2_{ij}]}{E[(\bar{A}-P)^2_{ij}]} = \frac{2k}{N} \quad \text{For large N}$$

Since $\bar{A}$ is the mean of M Bernoulli's we know:

$$E[(\bar{A}-P)^2_{ij}] = \frac{P_{ij}(1-P_{ij})}{M} = \frac{X_i^T X_j(1-X_i^T X_j)}{M}$$

Since $\hat{P}$ is generated from a noisy dot product of $\hat{X}_i^T \hat{X}_j$:

$$E[(\hat{P}-P)^2_{ij}] = \frac{\frac{1}{N}(X_i^T \Sigma_j X_i + X_j^T \Sigma_i X_j) + \frac{1}{N^2}(Tr(\Sigma_i \Sigma_j))}{M}$$

Letting the $\frac{1}{N^2 M}$ term quickly approach zero, we have:

$$E[(\hat{P}-P)^2_{ij}] = \frac{X_i^T \Sigma_j X_i + X_j^T \Sigma_i X_j}{NM}$$

From central limit theorem paper:

$$\Sigma_i = \Delta^{-1} E\Big[X_j X_j^T [x_i^T X_j(1-x_i^T X_j)]\Big]\Delta^{-1} \quad \text{where} \quad \Delta^{d\times d} = E[X_j X_j^T]$$

Therefore it comes down to showing that for high N:

$$\frac{X_i^T \Sigma_j X_i + X_j^T \Sigma_i X_j}{N X_i^T X_j(1-X_i^T X_j)} \approx \frac{2k}{N}$$

$$\frac{X_i^T \Sigma_j X_i + X_j^T \Sigma_i X_j}{X_i^T X_j(1-X_i^T X_j)} \approx 2k$$

or

$$\frac{X_i^T \Sigma_j X_i + X_j^T \Sigma_i X_j}{2k} \approx X_i^T X_j(1-X_i^T X_j) = P_{ij}(1-P_{ij})$$