Dynamic Network Analysis with Missing Data: Theory and Methods

Zack W. Almquist Carter T. Butts

University of Minnesota *University of California, Irvine*

Supplementary Material

Extended Analysis and Simulation Results

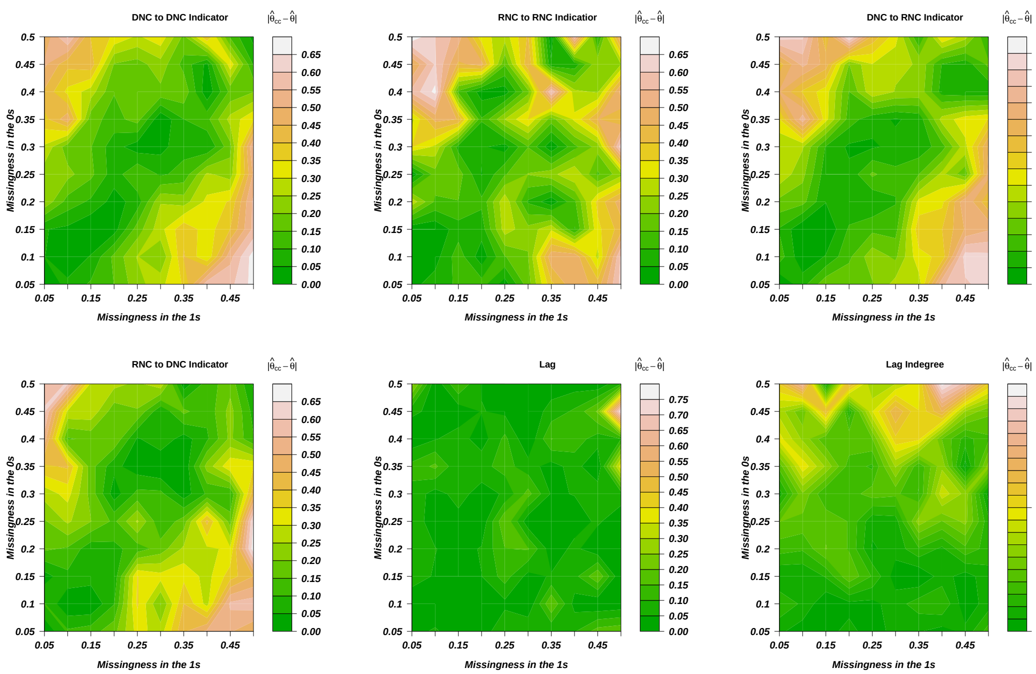

Figure 1: Missingness was induced at increments of 0.05 from 0.05 to 0.50 for each time point within the dynamic network (in this case we used 50 time points from the blog data). All combinations are plotted with the absolute difference between the "true" value and the complete-case estimate. Parameter estimates were generated for DNR both for the case of no missing data and from missing data cases using the complete case MAP with 0-imputation.

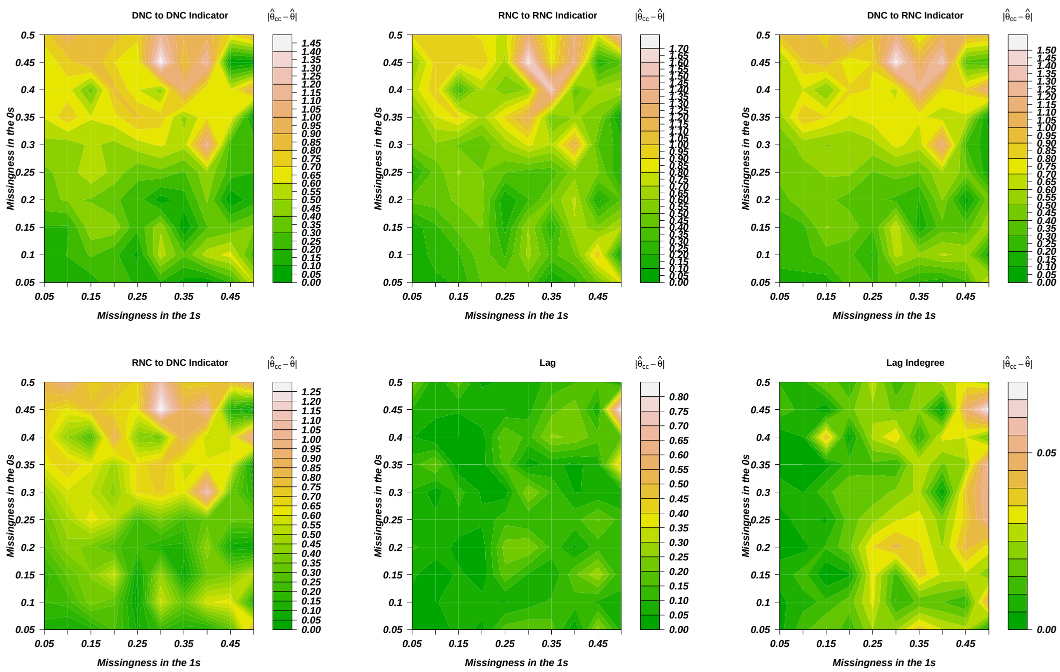

Figure 2: Missingness was induced at increments of 0.05 from 0.05 to 0.50 for each time point within the dynamic network (in this case we used 50 time points from the blog data). Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. Note that the MAP estimates and PIs under the approximate complete-case method quickly converge to the parameter estimates for the full data as the level of missingness becomes a smaller fraction of the total data observed. Parameter estimates were generated for DNR both for the case of no missing data and from missing data cases using the complete case with 1-imputation.

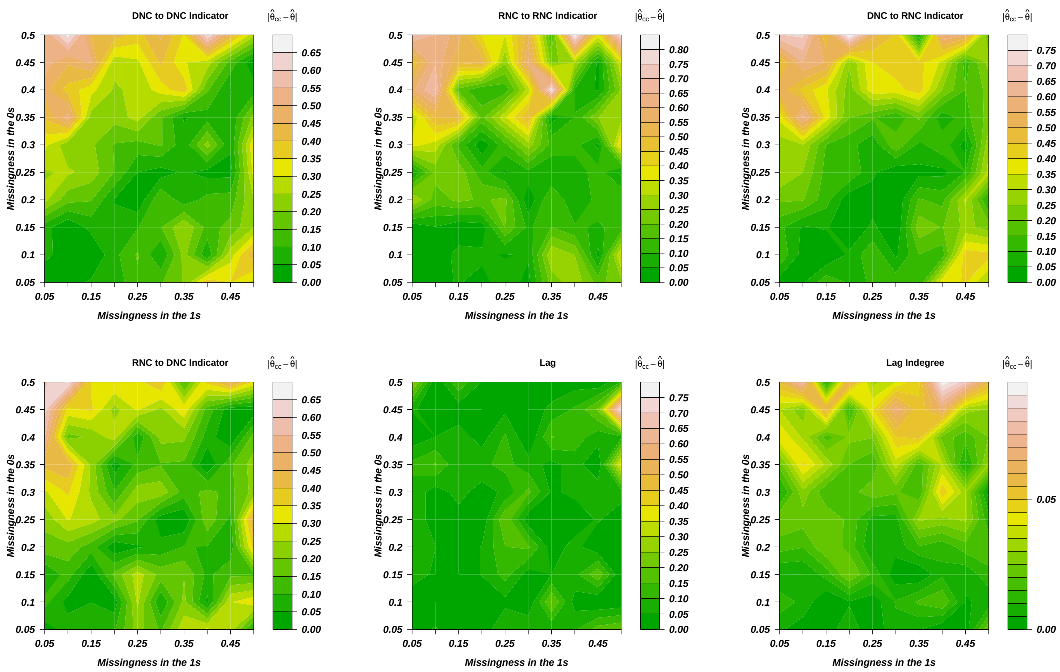

Figure 3: Missingness was induced at increments of 0.05 from 0.05 to 0.50 for each time point within the dynamic network (in this case we used 50 time points from the blog data). Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. Note that the MAP estimates and PIs under the approximate complete-case method quickly converge to the parameter estimates for the full data as the level of missingness becomes a smaller fraction of the total data observed. Parameter estimates were generated for DNR both for the case of no missing data and from missing data cases using the complete case with δ-imputation.

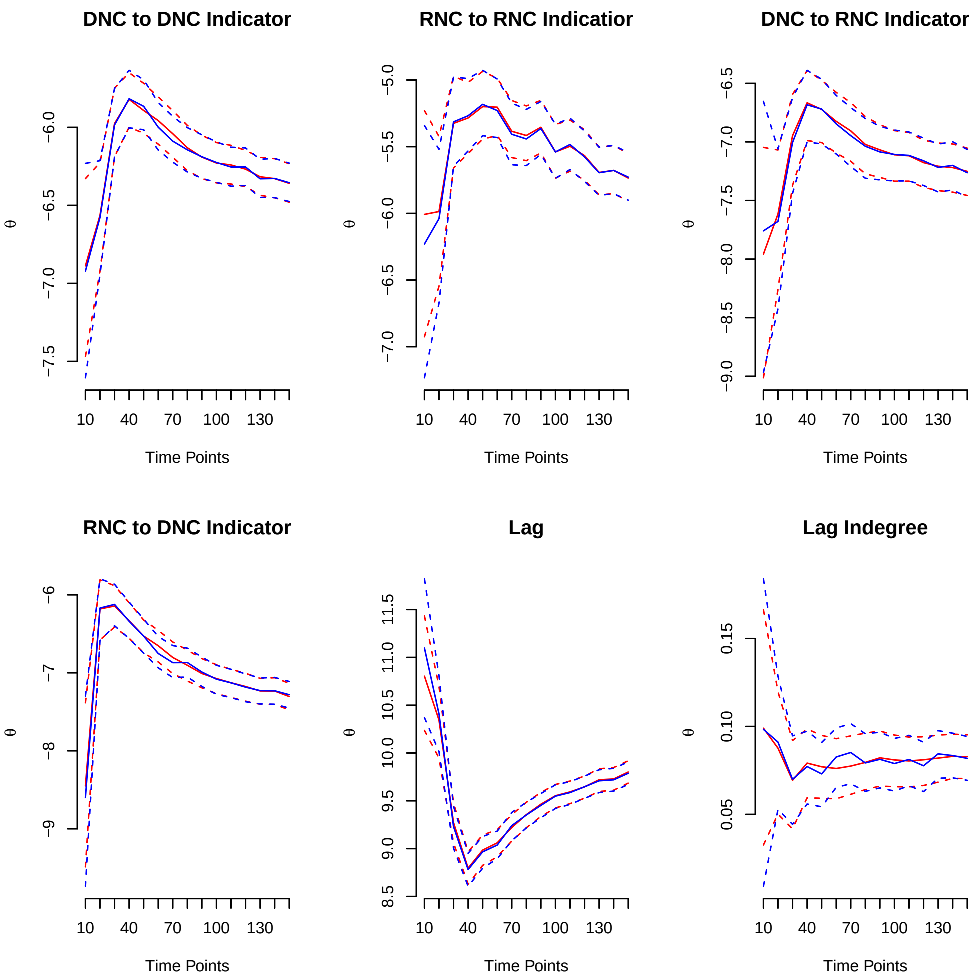

Figure 4: Missingness maintained at 5 random time-points with 10 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly. Parameter estimates were then generated for DNR under complete case with 0-imputation. Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. Note that the MAP estimates and PIs under the approximate complete-case method quickly converge to the parameter estimates for the full data as the level of missingness becomes a smaller fraction of the total data observed.

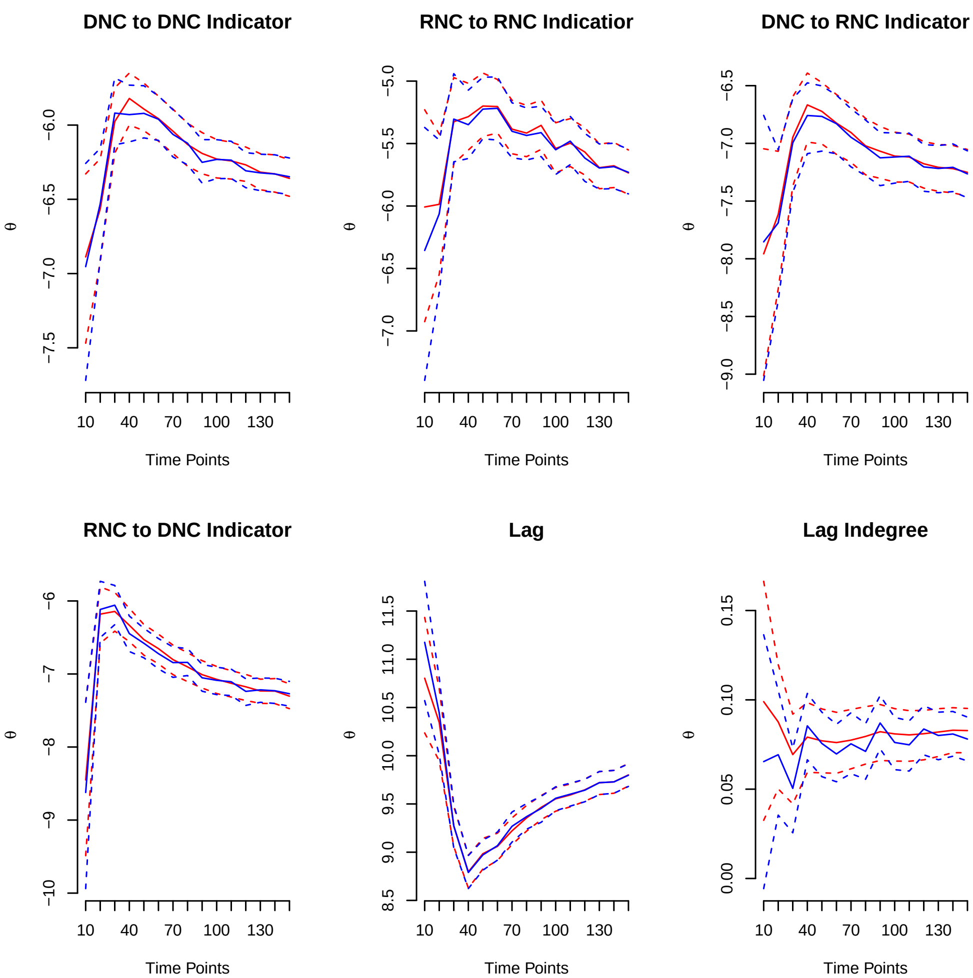

Figure 5: Missingness maintained at 5 random time-points with 10 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly. Parameter estimates were then generated for DNR under complete case with 1-imputation. Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. It appears that in this case the 1-imputation is much less efficient that 0-imputation, thought it is worth noting that the network under observation relatively sparse.

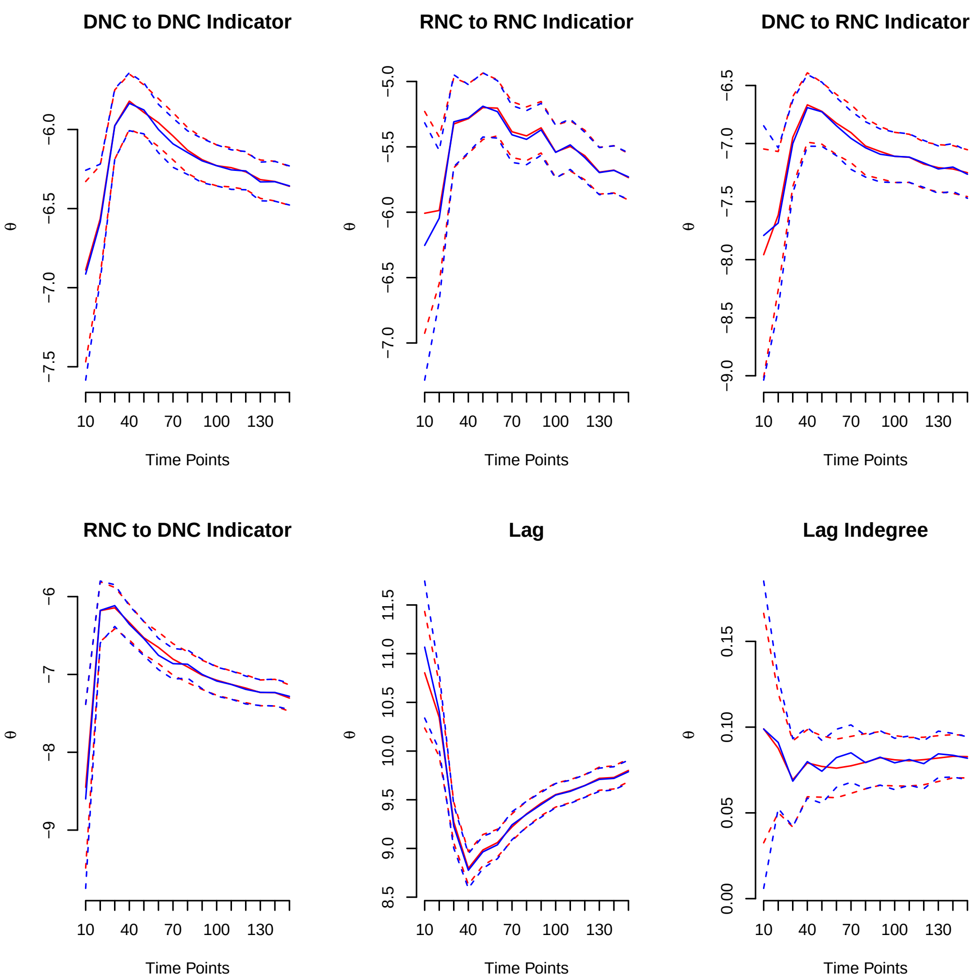

Figure 6: Missingness maintained at 5 random time-points with 10 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly. Parameter estimates were generated for DNR under complete case with d-imputation. Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. Note that the MAP estimates and PIs under the approximate complete-case method quickly converge to the parameter estimates for the full data as the level of missingness becomes a smaller fraction of the total data observed. We see similar results for the 0-imputation case, likely due to the relatively sparse nature of the dynamic blog network.

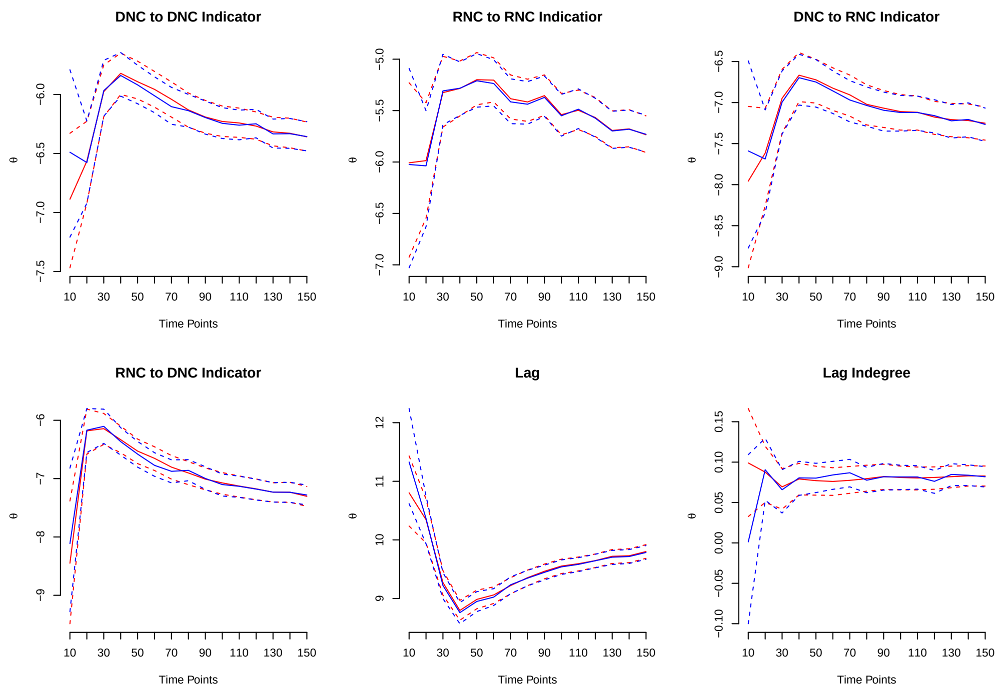

Figure 7: Missingness maintained at 5 random time-points with 10 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly. Parameter estimates were then generated for DNR under complete case with 0-imputation. Red is the “true” parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. Note that the MAP estimates and PIs under the approximate complete-case method quickly converge to the parameter estimates for the full data as the level of missingness becomes a smaller fraction of the total data observed. Notice that a number of the parameters are relatively biased when the fraction of missingness is high, but appears to be more efficient than the other three imputation schemes.

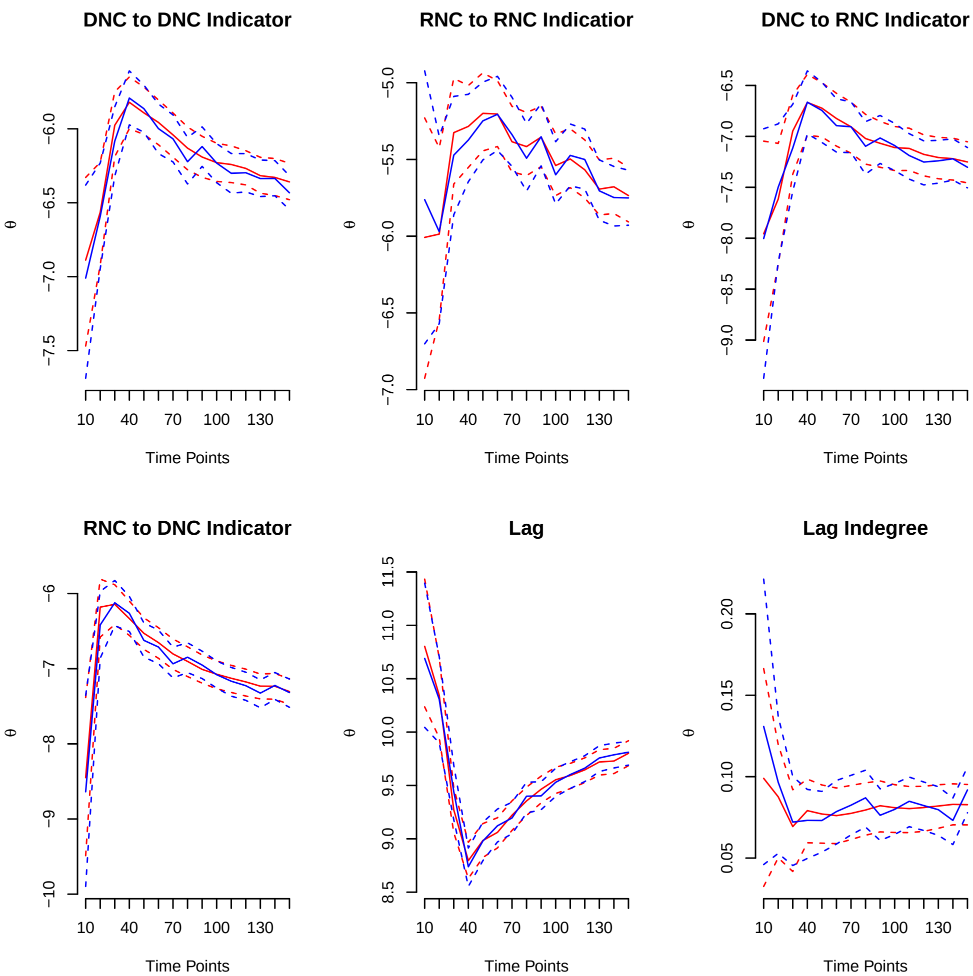

Figure 8: Missingness starting at 5 random time-points with 15 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly and the number of time points with missing was increased by 2 linearly. Parameter estimates were then generated for DNR under complete case with 0-imputation. Red is the "true" parameter and PI (i.e., MAP estimate from the complete data) and the blue line is the estimated parameter and PI from the data with simulated missingness. The key thing to observe here is the MAP stabilize to the "true" parameter values of the non-missing MAP as the level of missingness becomes a smaller fraction of the total data when the amount of missingness is increasing at a slower rate than the amount of observations.

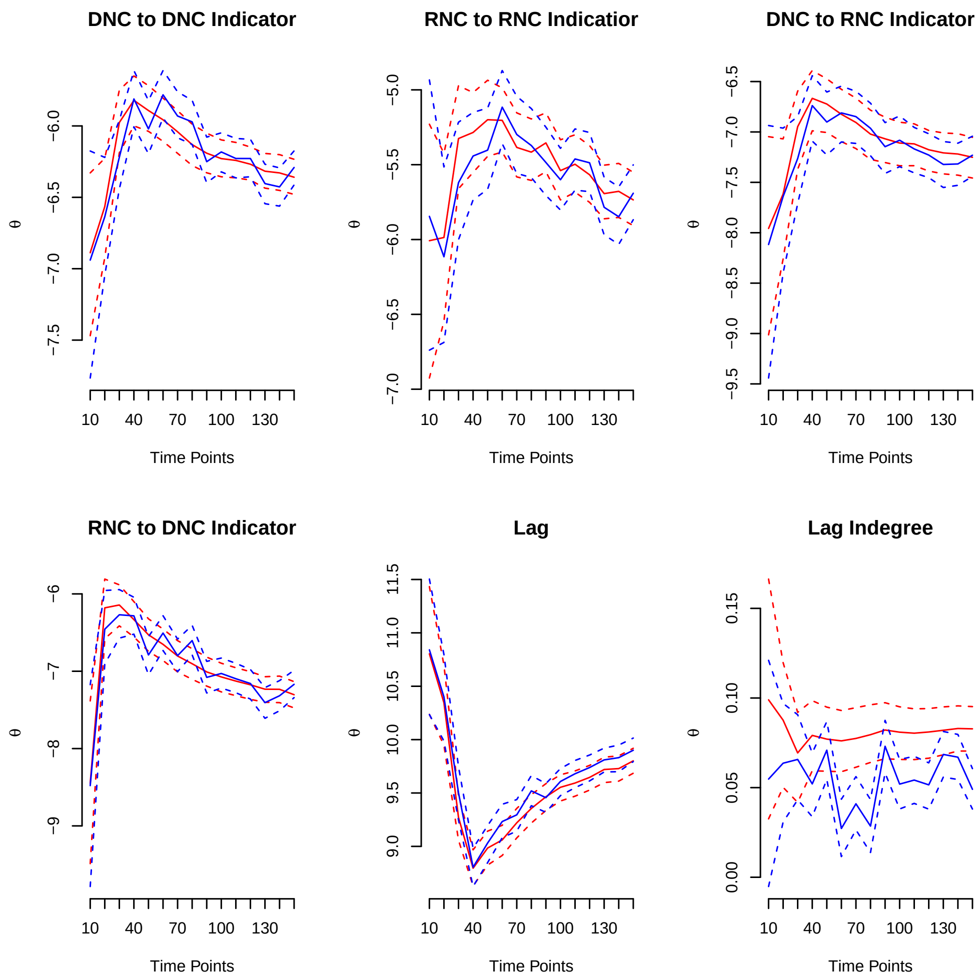

Figure 9: Missingness starting at 5 random time-points with 15 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly and the number of time points with missing was increased by 2 linearly. Parameter estimates were then generated for DNR under complete case with 1-imputation. Red is the "true" parameter and PI (i.e., MAP from the complete data) and the blue line is the estimated parameter and Pl from the data with simulated missingness. The key thing to observe here is the MAP and PI stabilize to the "true" parameter values of the non-missing MAP and PI as the level of missingness becomes a smaller fraction of the total data when the amount of missingness is increasing at a slower rate than the amount of observations except for the 1-imputed data where MAP behaves very poorly.

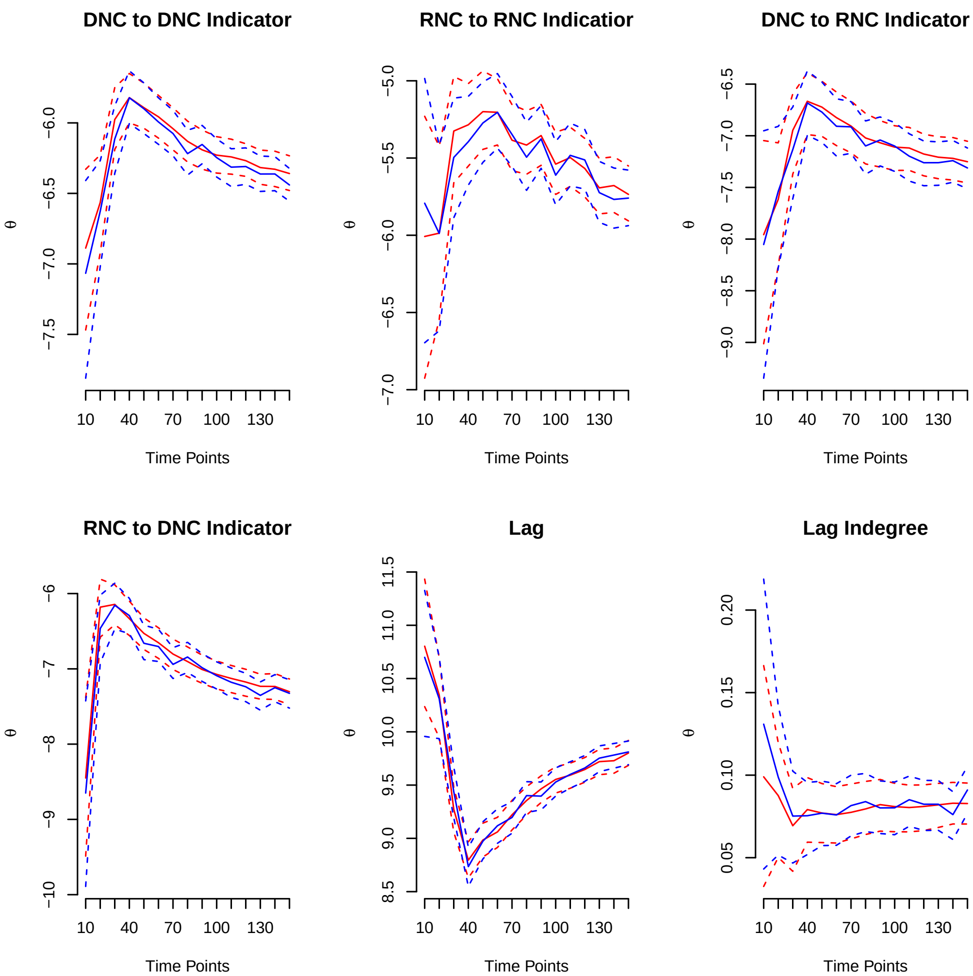

Figure 10: Missingness starting at 5 random time-points with 15 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly and the number of time points with missing was increased by 2 linearly. Parameter estimates were then generated for DNR under complete case with d-imputation. Red is the "true" parameter and PI (i.e., MAP from the complete data) and the blue line is the estimated parameter and Pl from the data with simulated missingness. The key thing to observe here is the MAP and PI stabilize to the "true" parameter values of the non-missing MAP and PI as the level of missingness becomes a smaller fraction of the total data when the amount of missingness is increasing at a slower rate than the amount of observations.

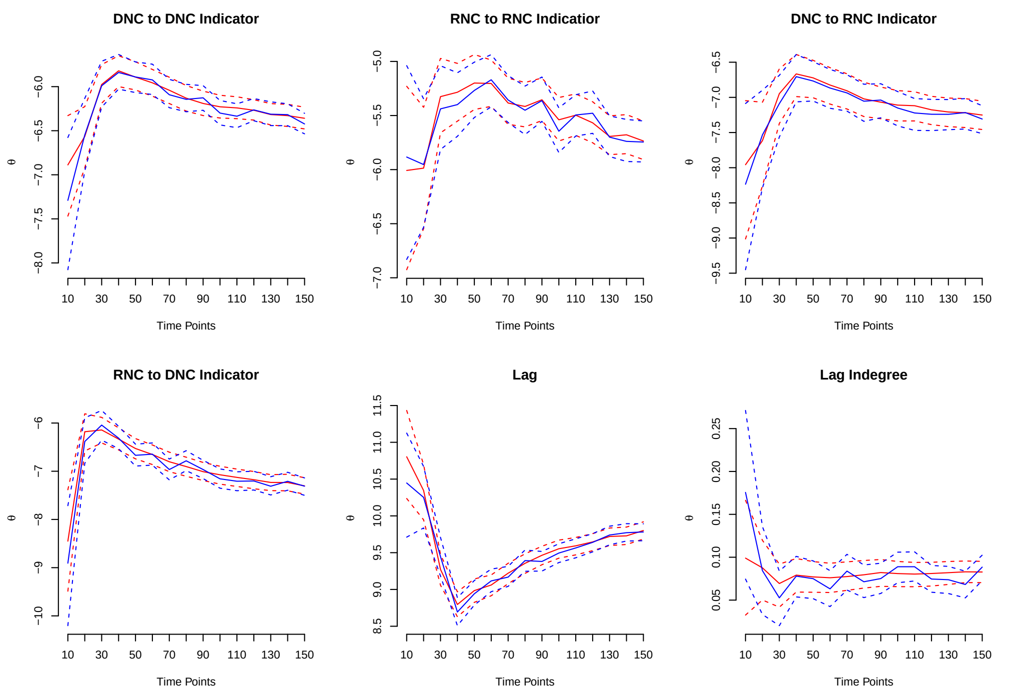

Figure 11: Missingness starting at 5 random time-points with 15 percent missing in the zeros and 30 percent in the 1s. The number of observed time points was then increased by 10 time points linearly and the number of time points with missing was increased by 2 linearly. Parameter estimates were then generated for DNR under complete case with R-imputation. Red is the "true" parameter and PI (i.e., MAP from the complete data) and the blue line is the estimated parameter and Pl from the data with simulated missingness. The key thing to observe here is the MAP and PI stabilize to the "true" parameter values of the non-missing MAP and PI as the level of missingness becomes a smaller fraction of the total data when the amount of missingness is increasing at a slower rate than the amount of observations, though not as well as the 0 and d-imputation methods – this may be due to the sparsity in these graphs.

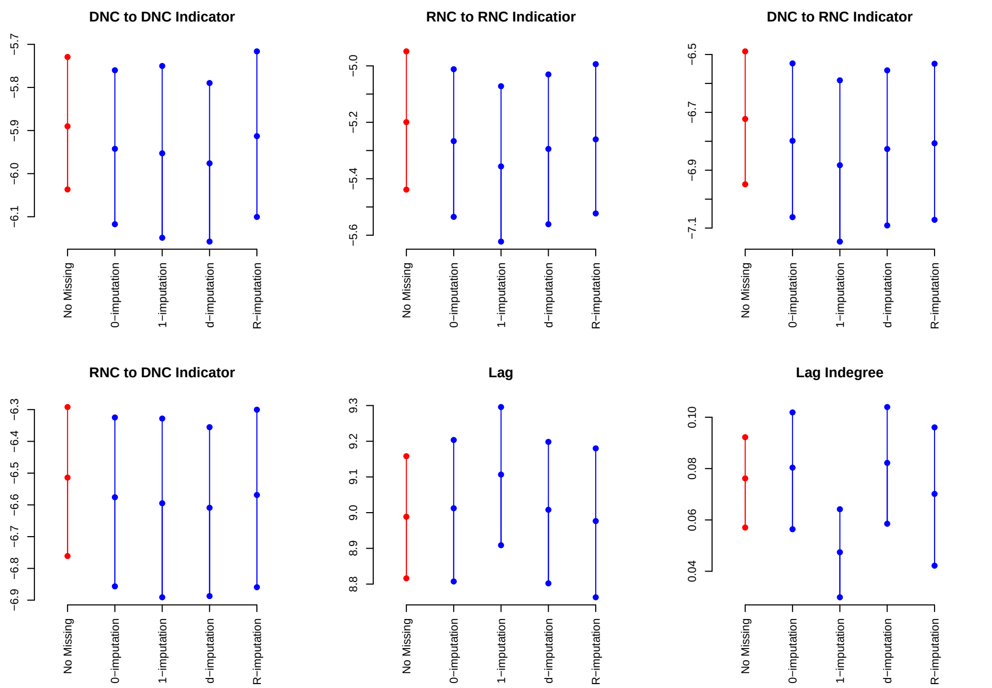

Figure 12: Missingness in 25 out of 50 observations. Each missing observation contained 15 percent missing in 0s and 30 percent in the 1s. Simulated 100 times, parameters and Bayesian PI averaged. Red is the “true” parameter and PI estimate and blue is the model with missingness in the data.

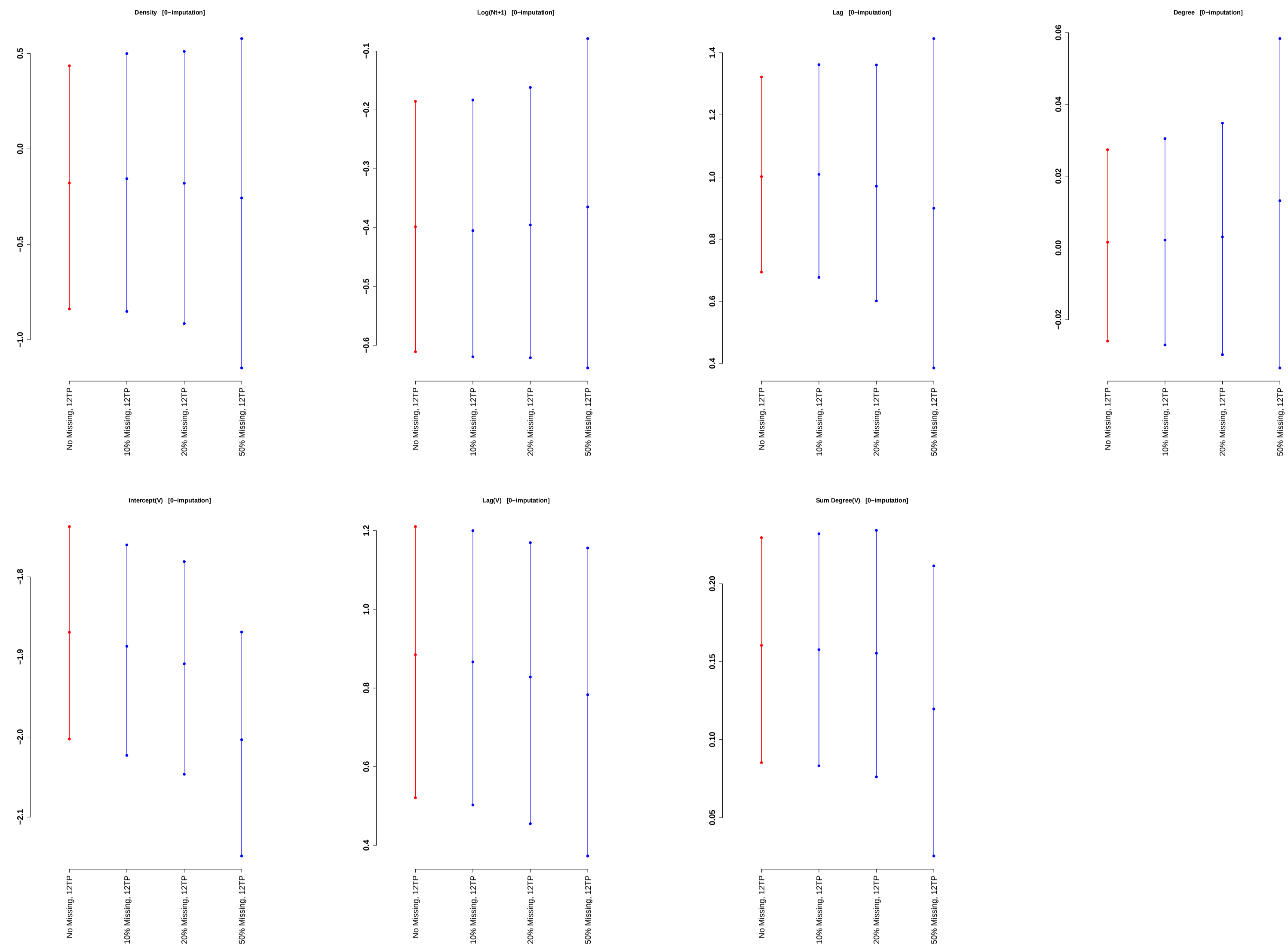

Figure 13: Missingness at varying rates in the vertex set over 12 time points. Simulated 100 times, parameters and Bayesian PI averaged. Red is the "true" parameter and PI estimate and blue is the model with missingness in the data.

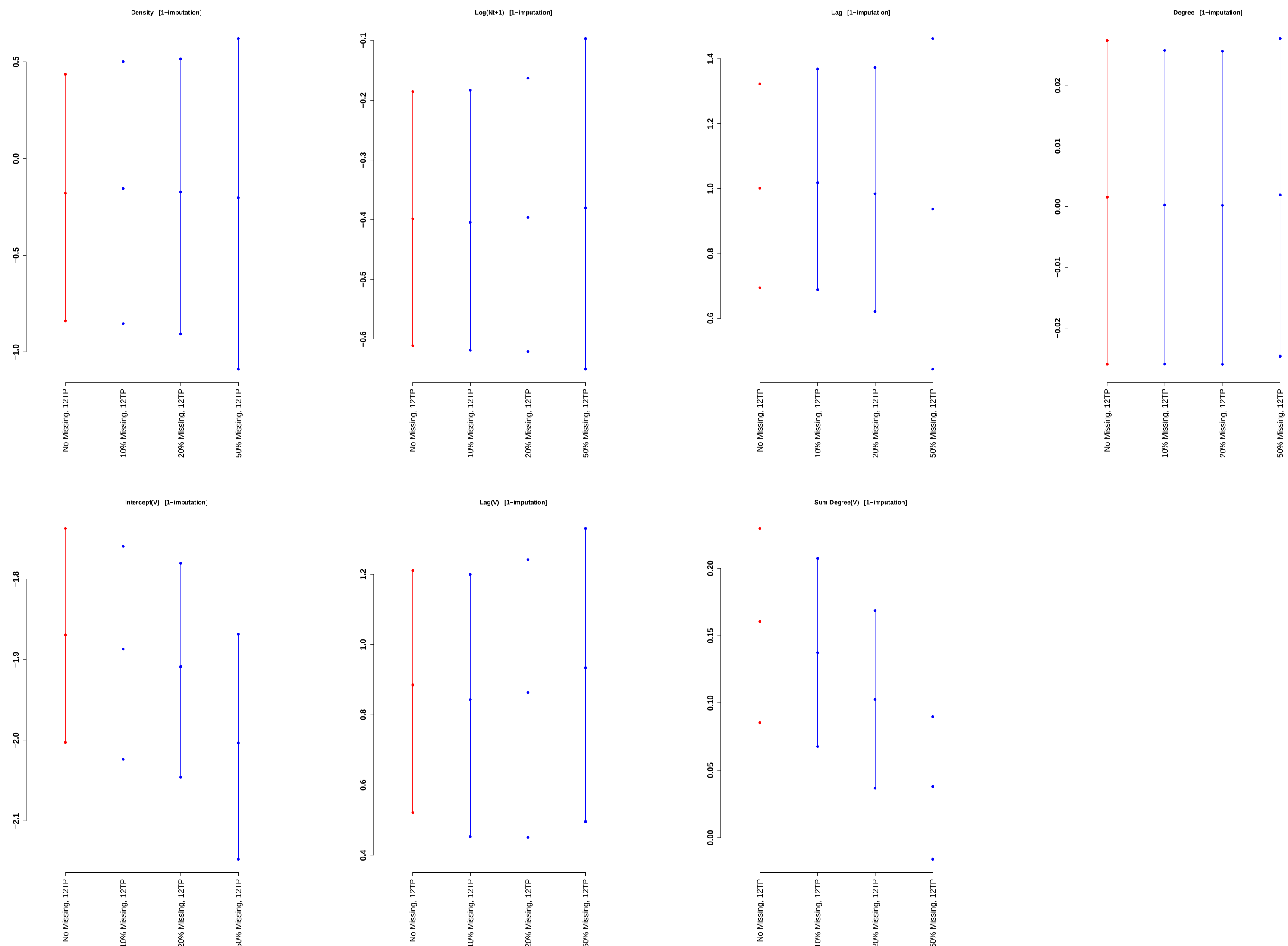

Figure 14: Missingness at varying rates in the vertex set over 12 time points. Simulated 100 times, parameters and Bayesian PI averaged. Red is the “true” parameter and PI estimate and blue is the model with missingness in the data.

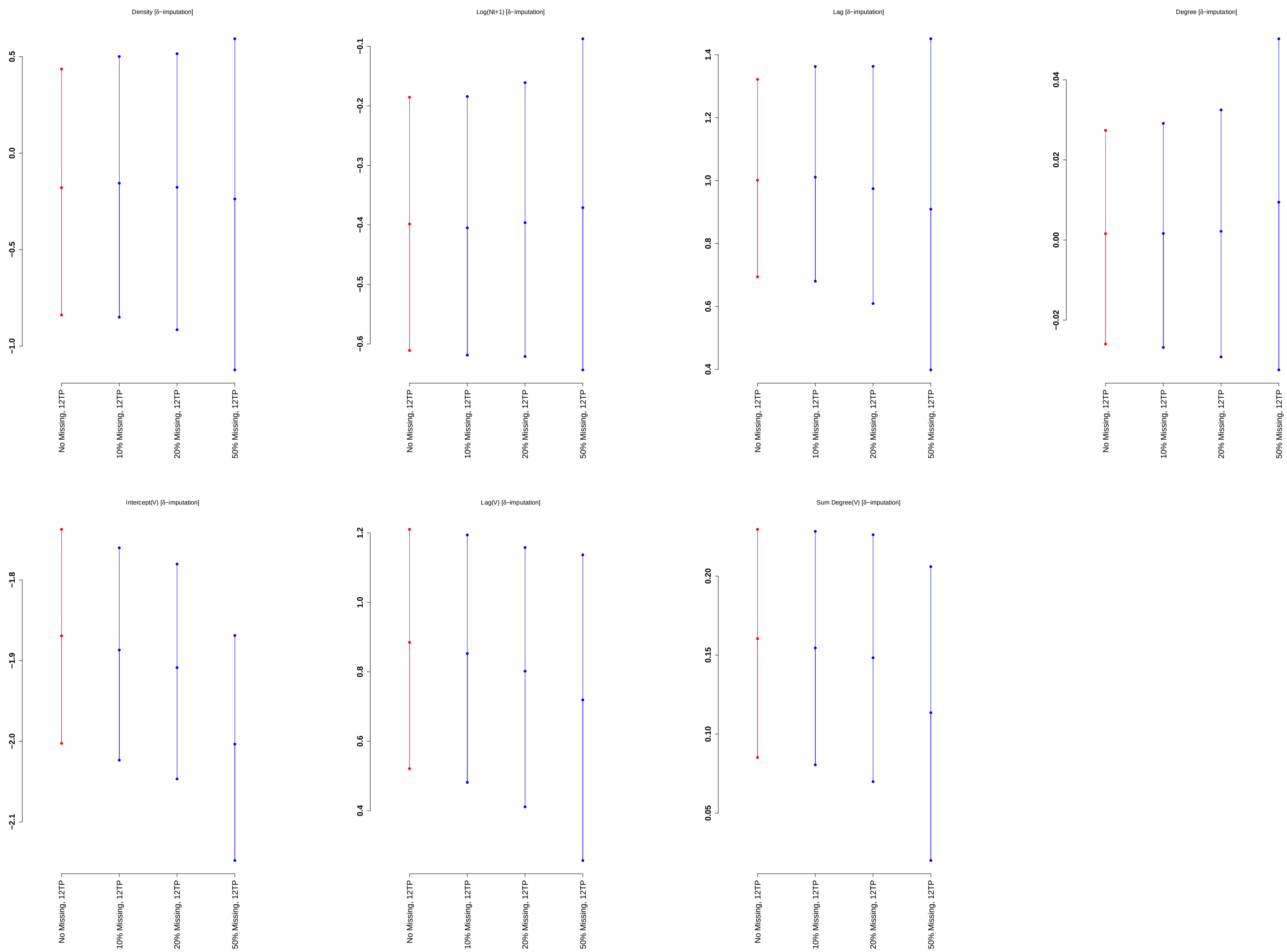

Figure 15: Missingness at varying rates in the vertex set over 12 time points. Simulated 100 times, parameters and Bayesian PI averaged. Red is the “true” parameter and PI estimate and blue is the model with missingness in the data.