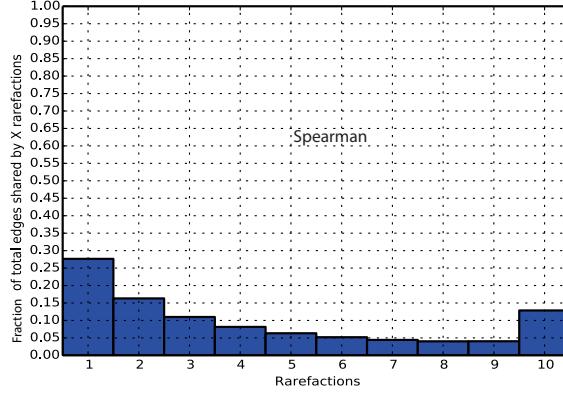
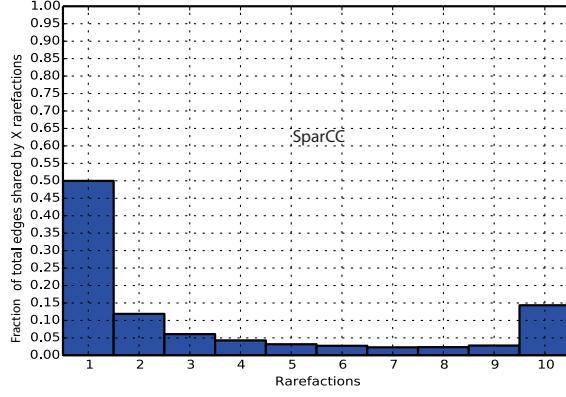
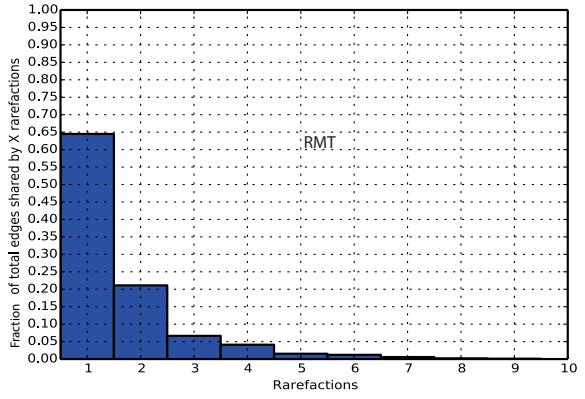
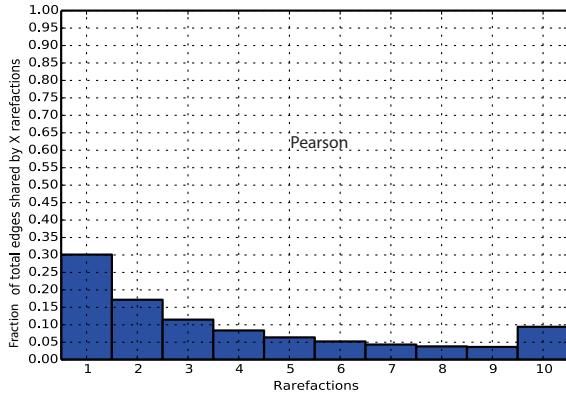
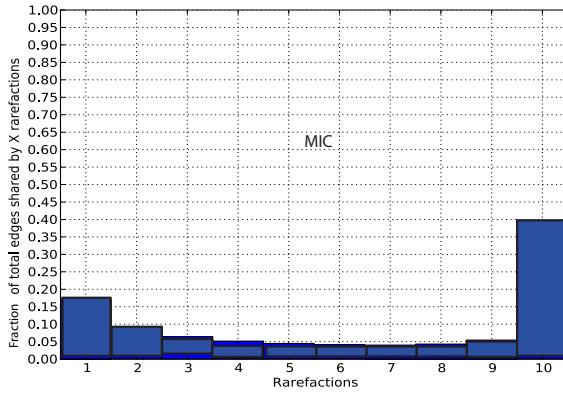
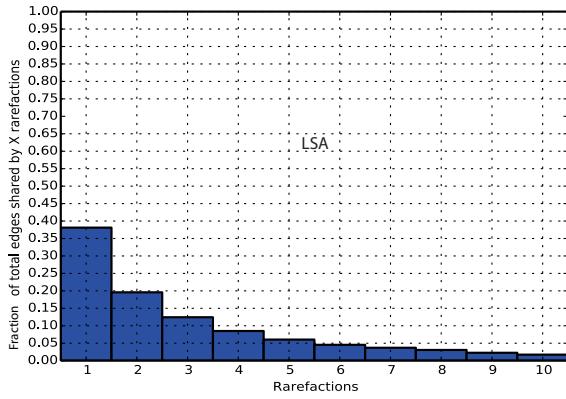
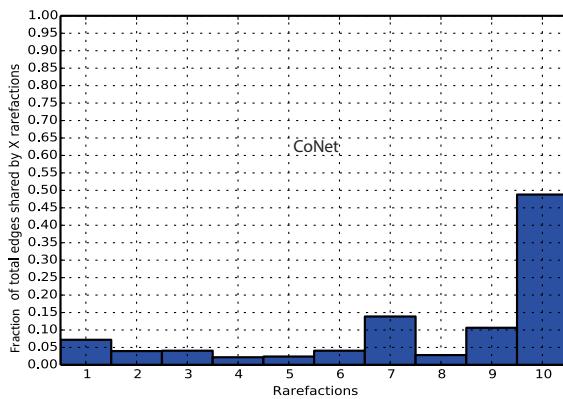
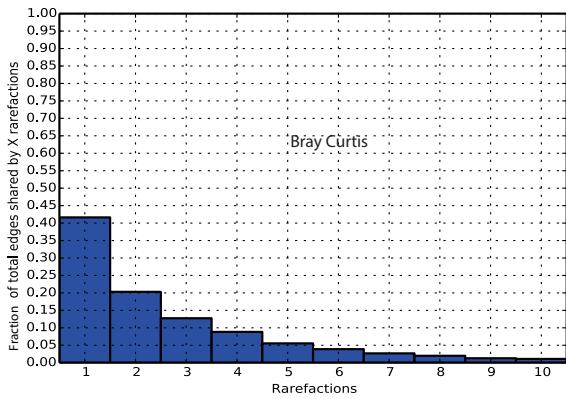


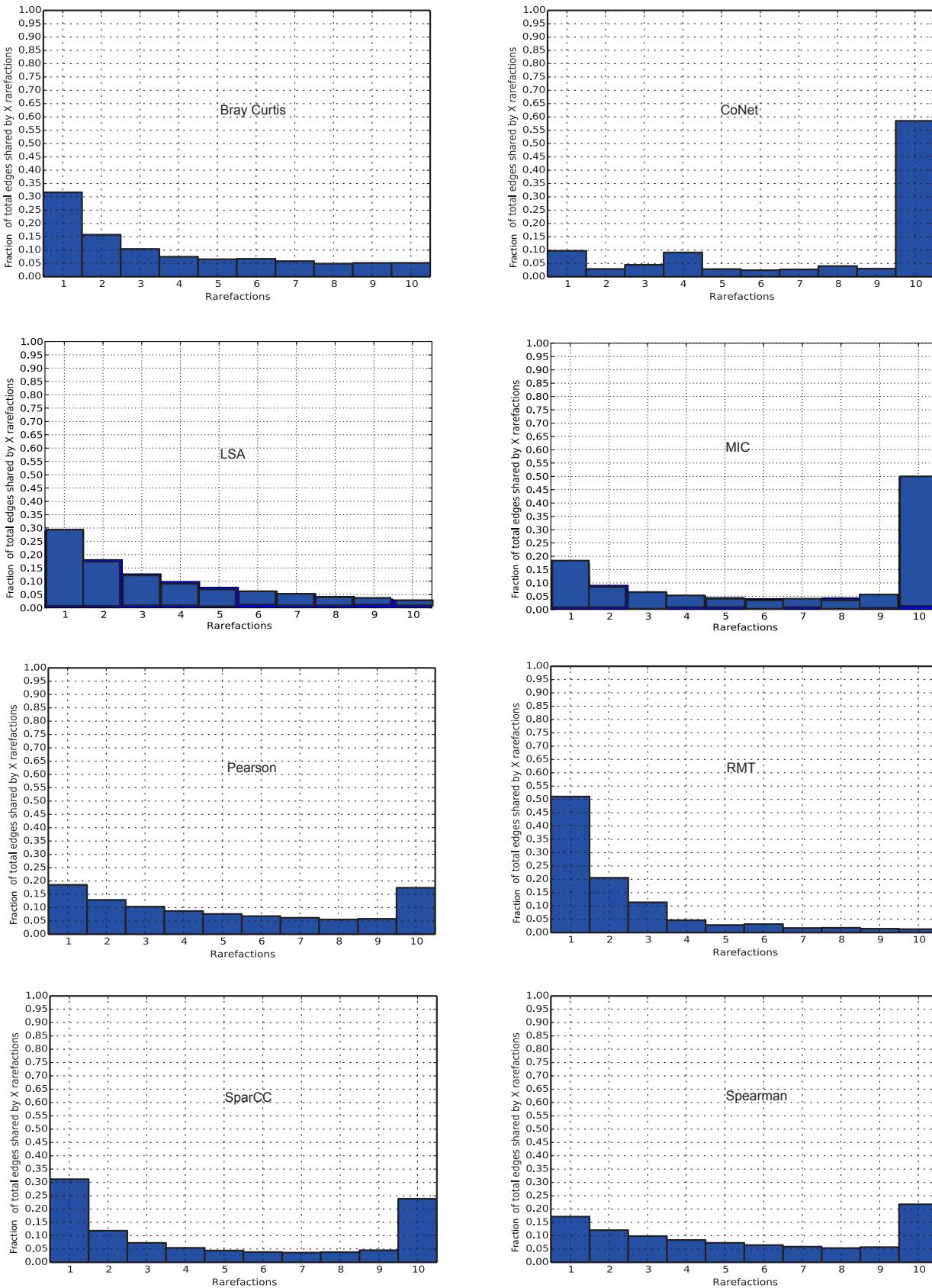
Supplementary Figure 1:

The fraction shared edges between the tools on all evaluated tables. Each cell i,j represents the shared edges on all tables (excluding the time series tables 3.34-3.43). The x-axis of each subpanel represents the percentage of tool i 's edges that are shared by both tools, and the y-axis represents the percentage of tool j 's edges that are shared by both.



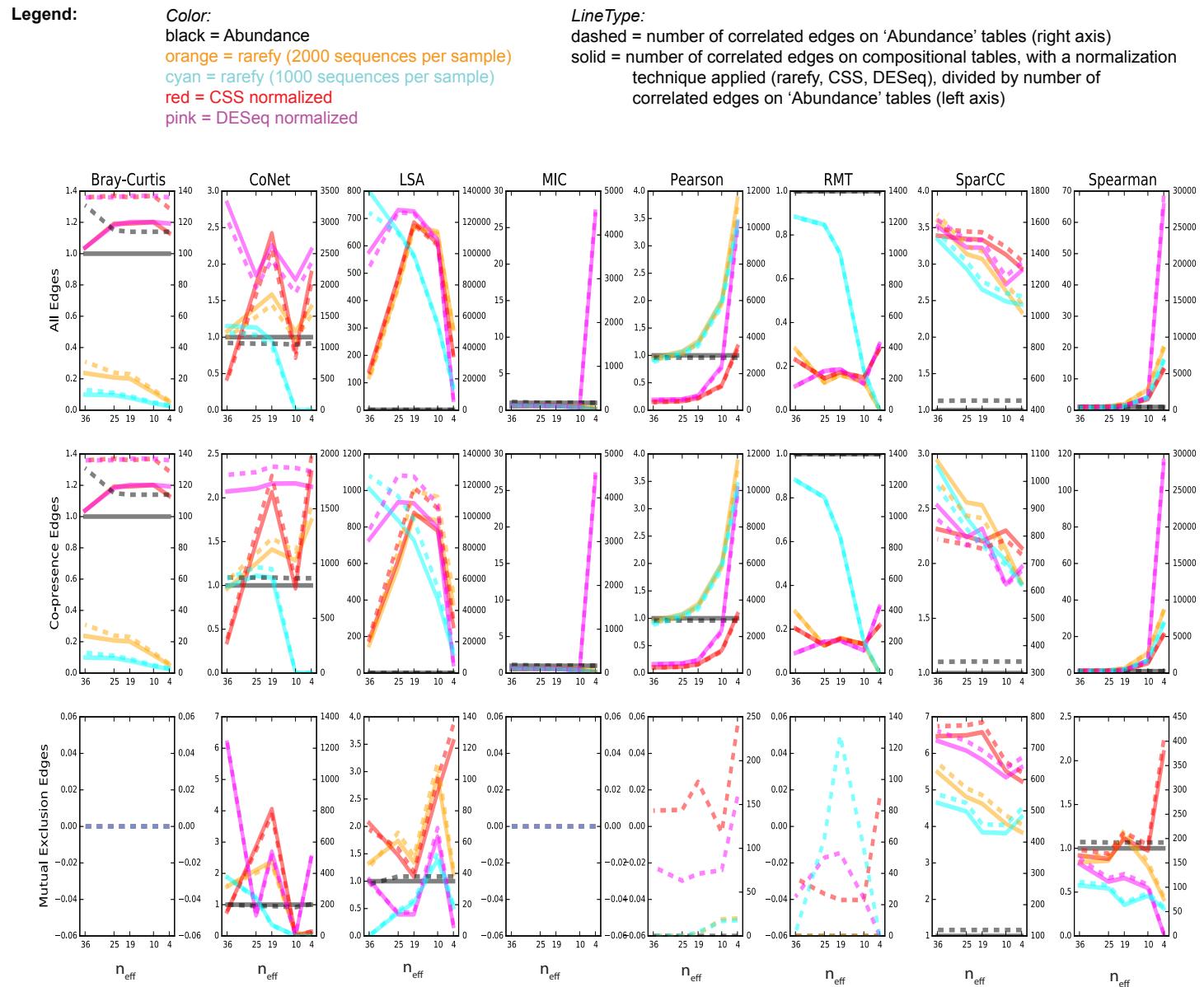
Supplementary Figure 2:

The fraction shared feature pairs in X/10 rarefactions of 1000 sample count depth for a given technique.



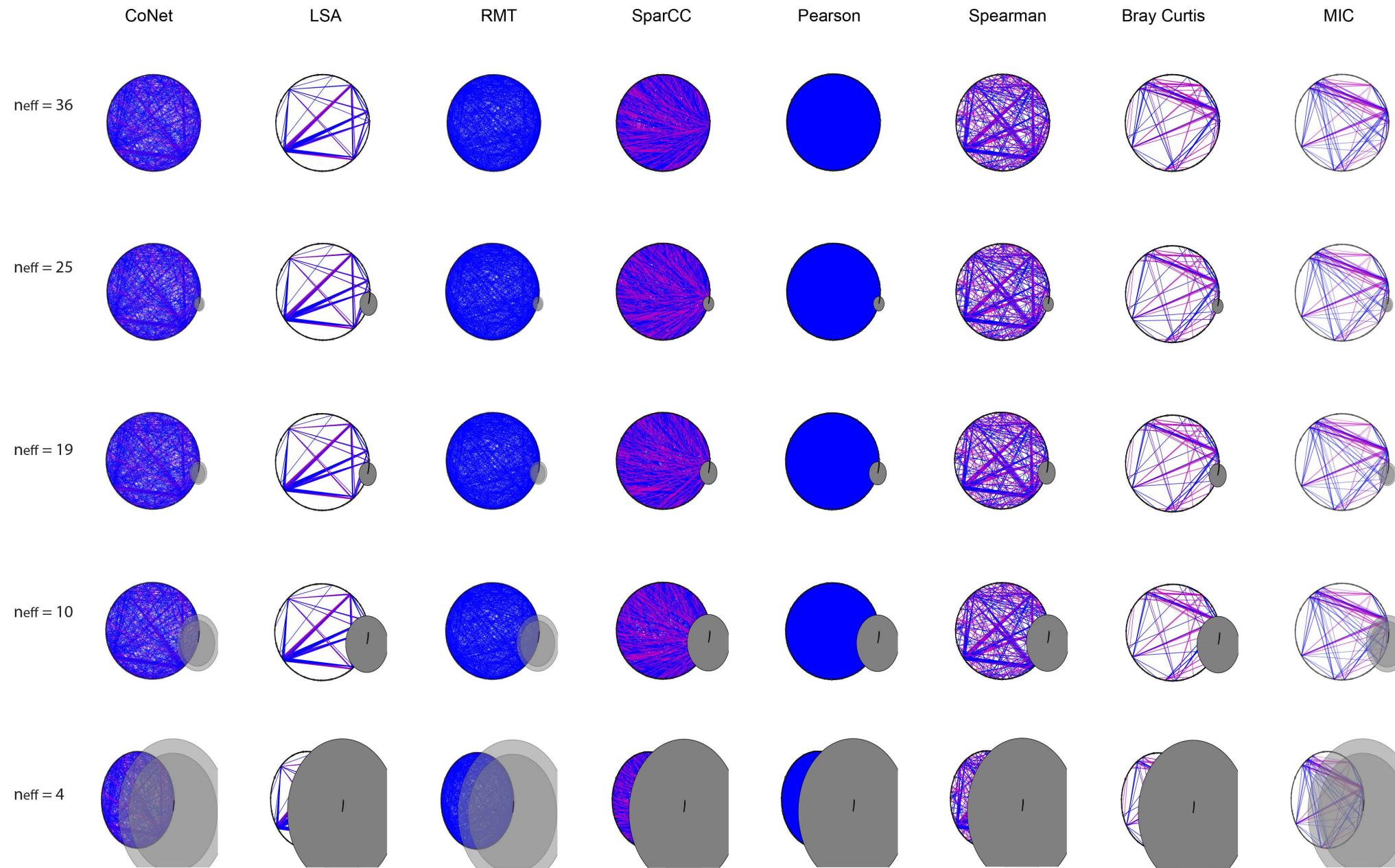
Supplementary Figure 3:

The fraction shared feature pairs in $X/10$ rarefactions of 2000 sample count depth for a given technique.



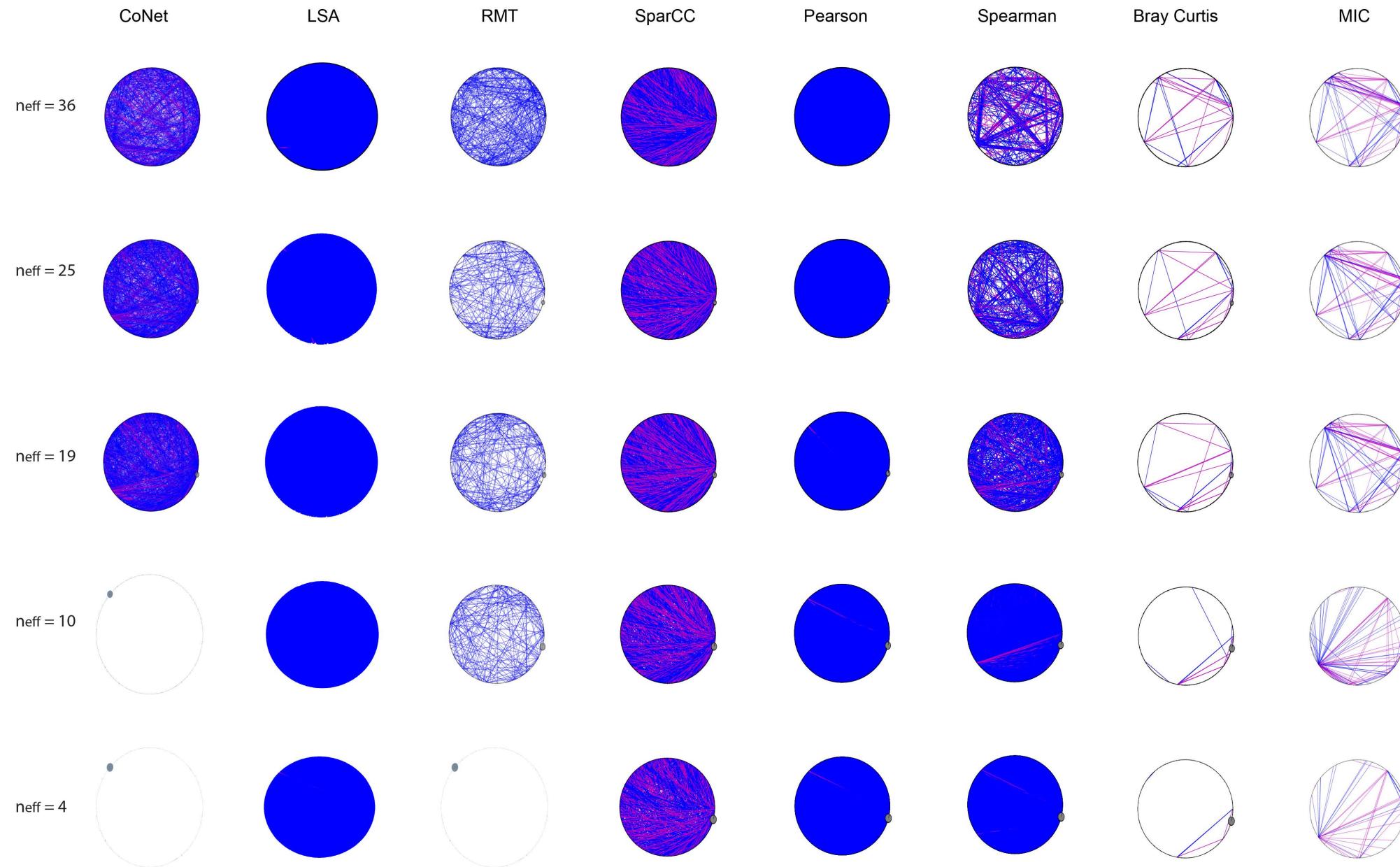
Supplementary Figure 4:

The impact of varying n_{eff} and library-size adjustment strategies on resulting significant edges. We created five copula tables whose ranked correlation structure and marginal distributions were the same, but where one pair of species was multiplied by an increasing factor to decrease the n_{eff} – these are the same tables as Fig. 2. The left axis in each plot is fractional and corresponds to the solid lines, thus, the black solid lines are always at 1.



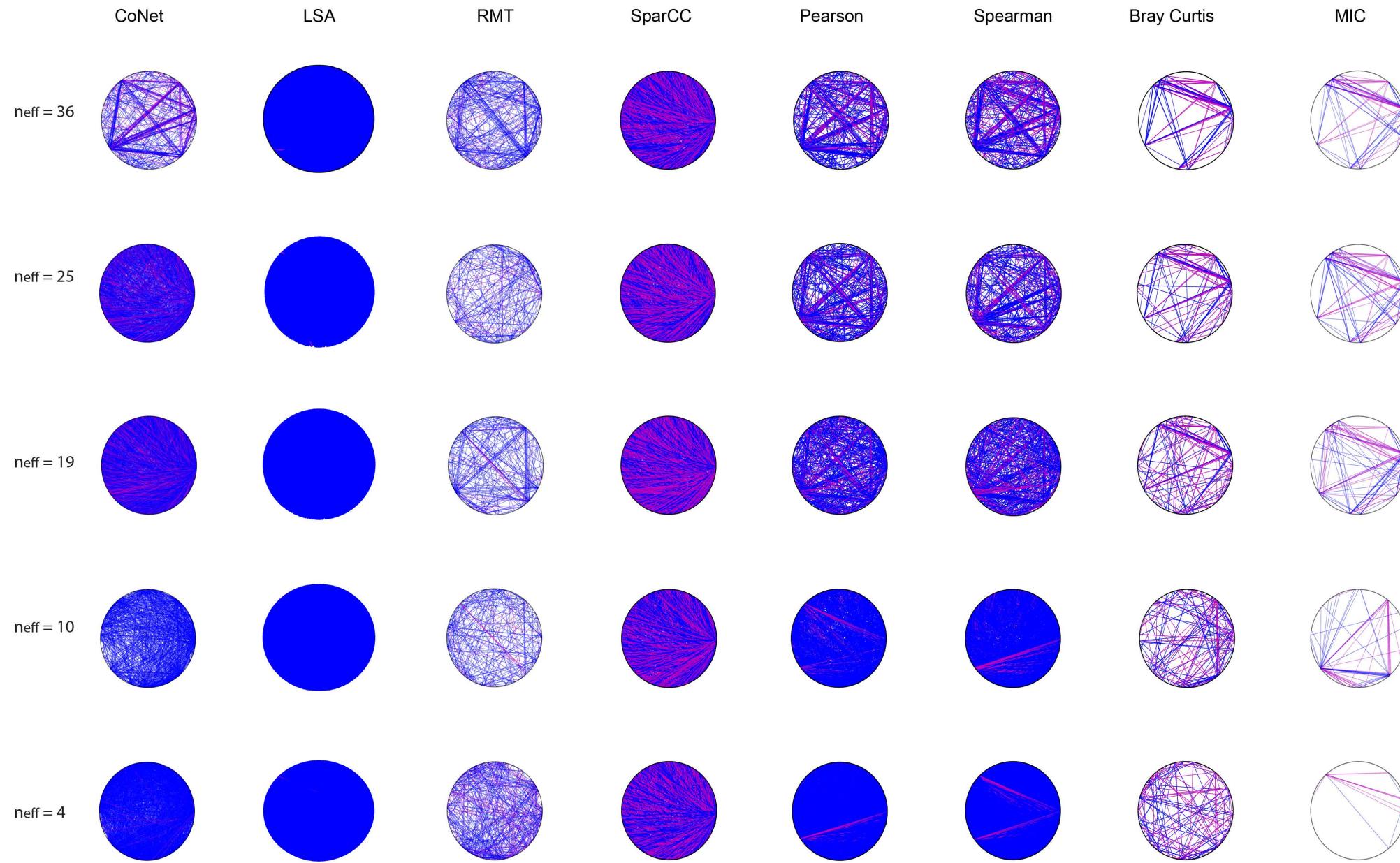
Supplementary Figure 5:

Visual depiction of significant ($p < 0.001$) edges found by each technique on raw abundance tables with decreasing n_{eff} : series 1 of 4. This corresponds to the 'Abundance' data of Fig. 2 and to the black lines in Supplementary Fig. 4. The significant edges from each tool graphed (correlated in blue, anti-correlated in pink). p -value thresholds determining a significant edge were set at .001 for all but RMT and CoNet. Nodes are displayed in gray and size is proportional to mean abundance. N_{eff} was calculated using inverse Simpson.



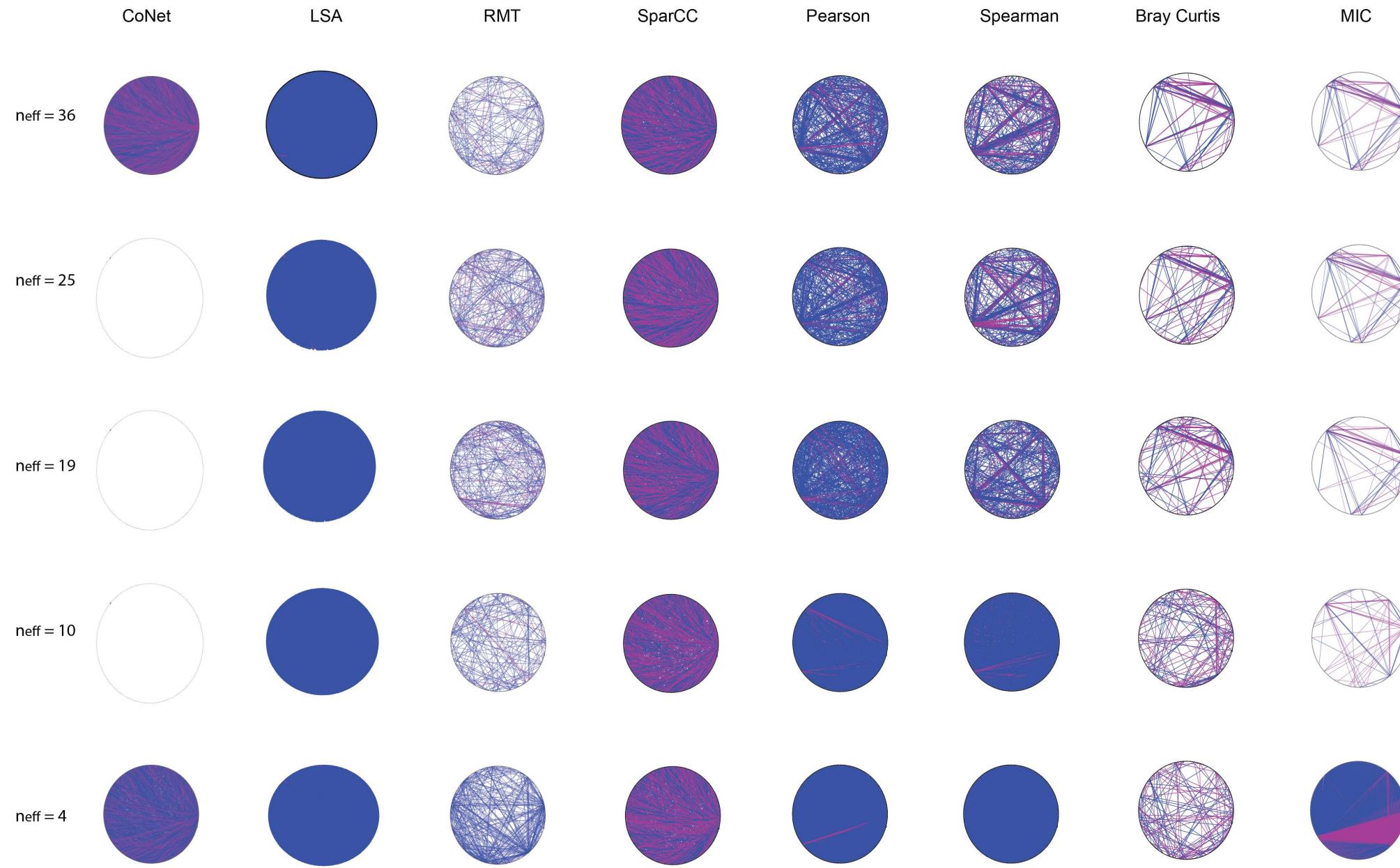
Supplementary Figure 6:

Visual depiction of significant ($p < 0.001$) edges found by each technique on compositional tables with decreasing n_{eff} : series 2 of 4. This corresponds to the 'rarefy_2000' data of Fig. 2 and to the orange lines in Supplementary Fig. 4. Label corresponds to Supplementary Fig. 5.



Supplementary Figure 7:

Visual depiction of significant ($p < 0.001$) edges found by each technique on compositional tables with decreasing n_{eff} : series 3 of 4. This corresponds to the 'CSS' data of Fig. 2 and to the red lines in Supplementary Fig. 4. Label corresponds to Supplementary Fig. 5.



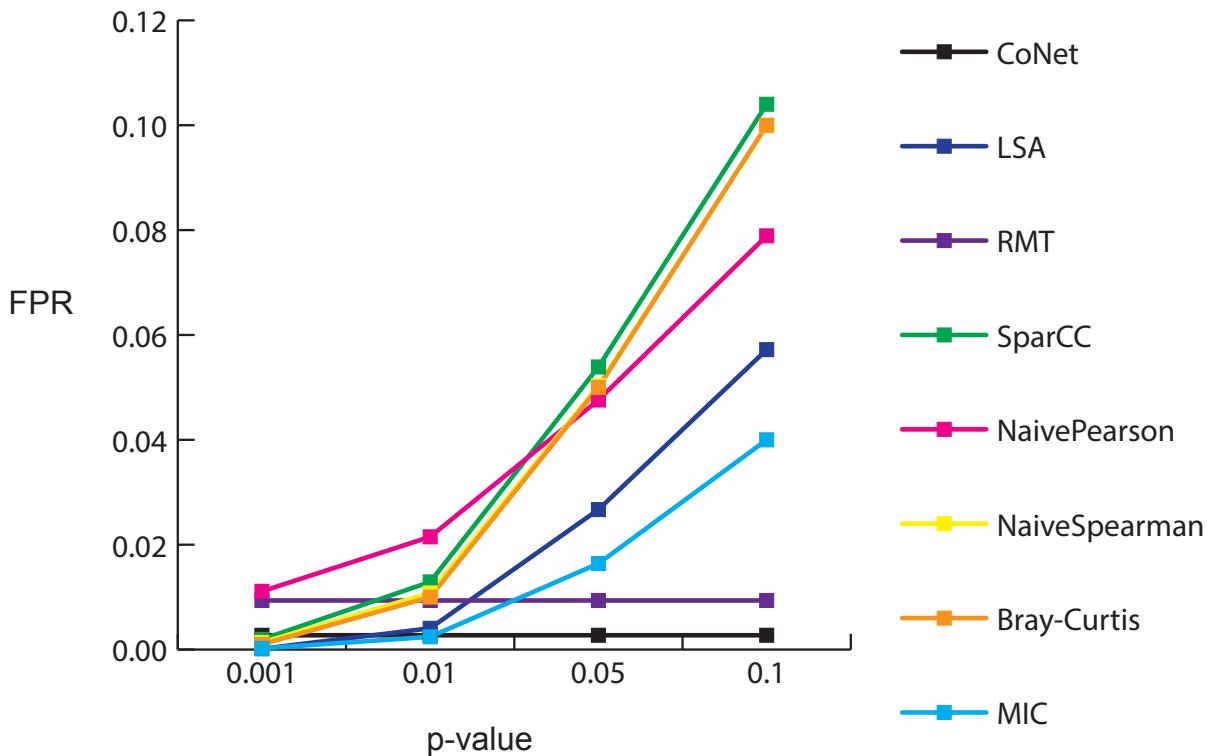
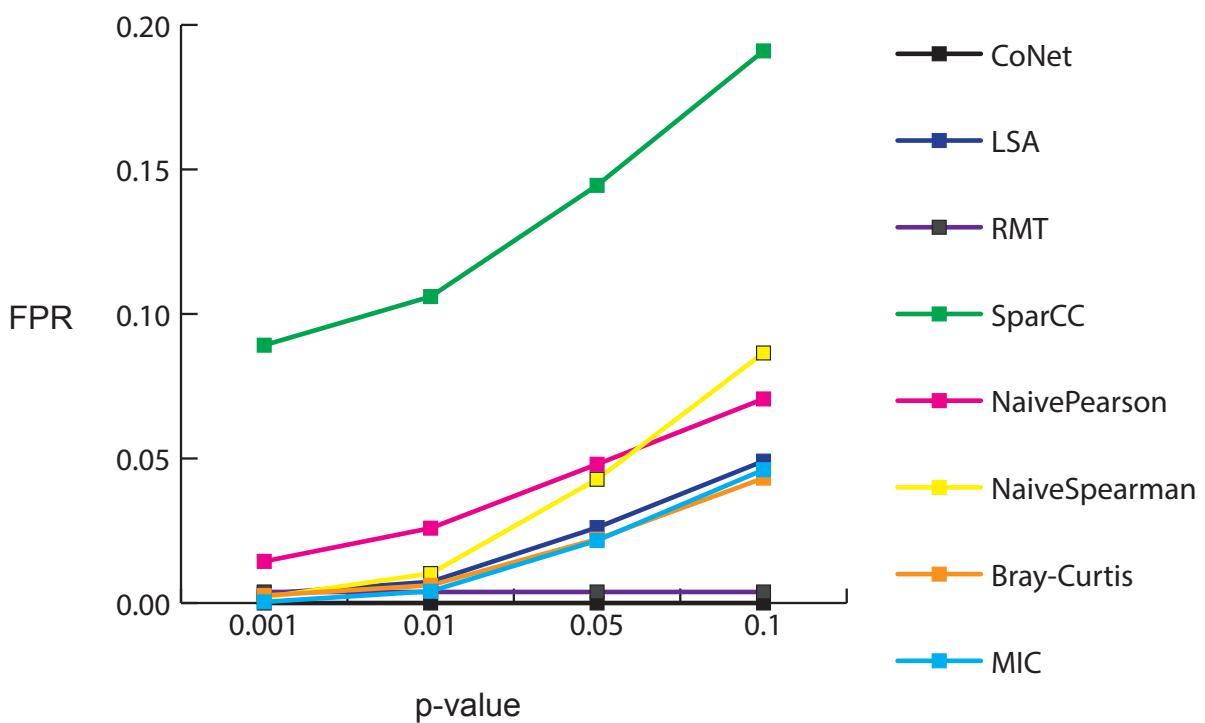
Supplementary Figure 8:

Visual depiction of significant ($p < 0.001$) edges found by each technique on compositional tables with decreasing n_{eff} : series 4 of 4. This corresponds to the 'DESeq' data of Fig. 2 and to the pink lines in Supplementary Fig. 4. Label corresponds to Supplementary Fig. 5.



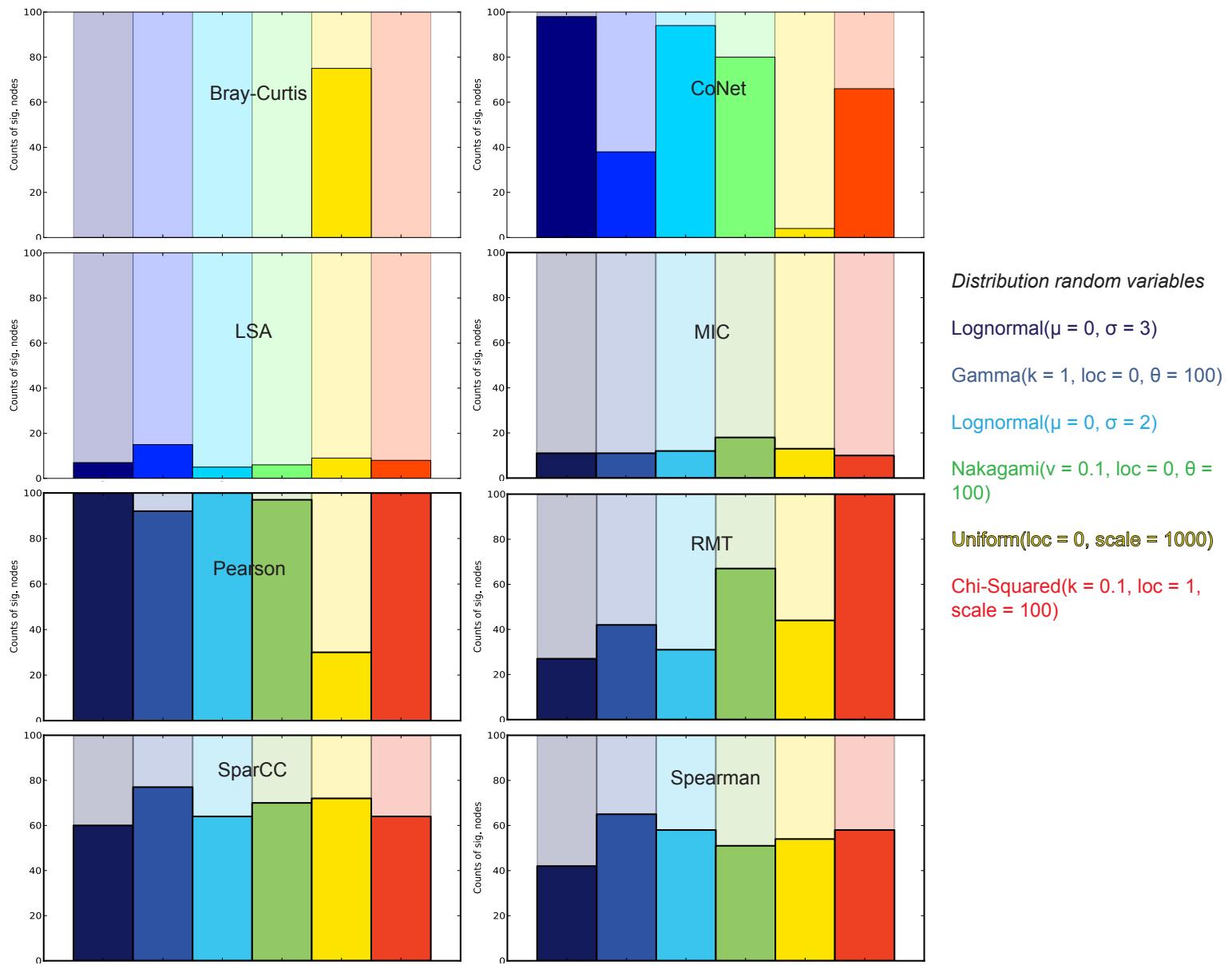
Supplementary Figure 9:

Network overlap in response to compositionality. The same data as Fig. 2b, except using Jaccard index as a measure of network overlap. Tool predictions were benchmarked against the edges found by the same tool on absolute 'Abundance' data. SparCC was benchmarked against the log-transformed Pearson correlations on 'Abundance' data, since that is what it seeks to estimate.

a**b**

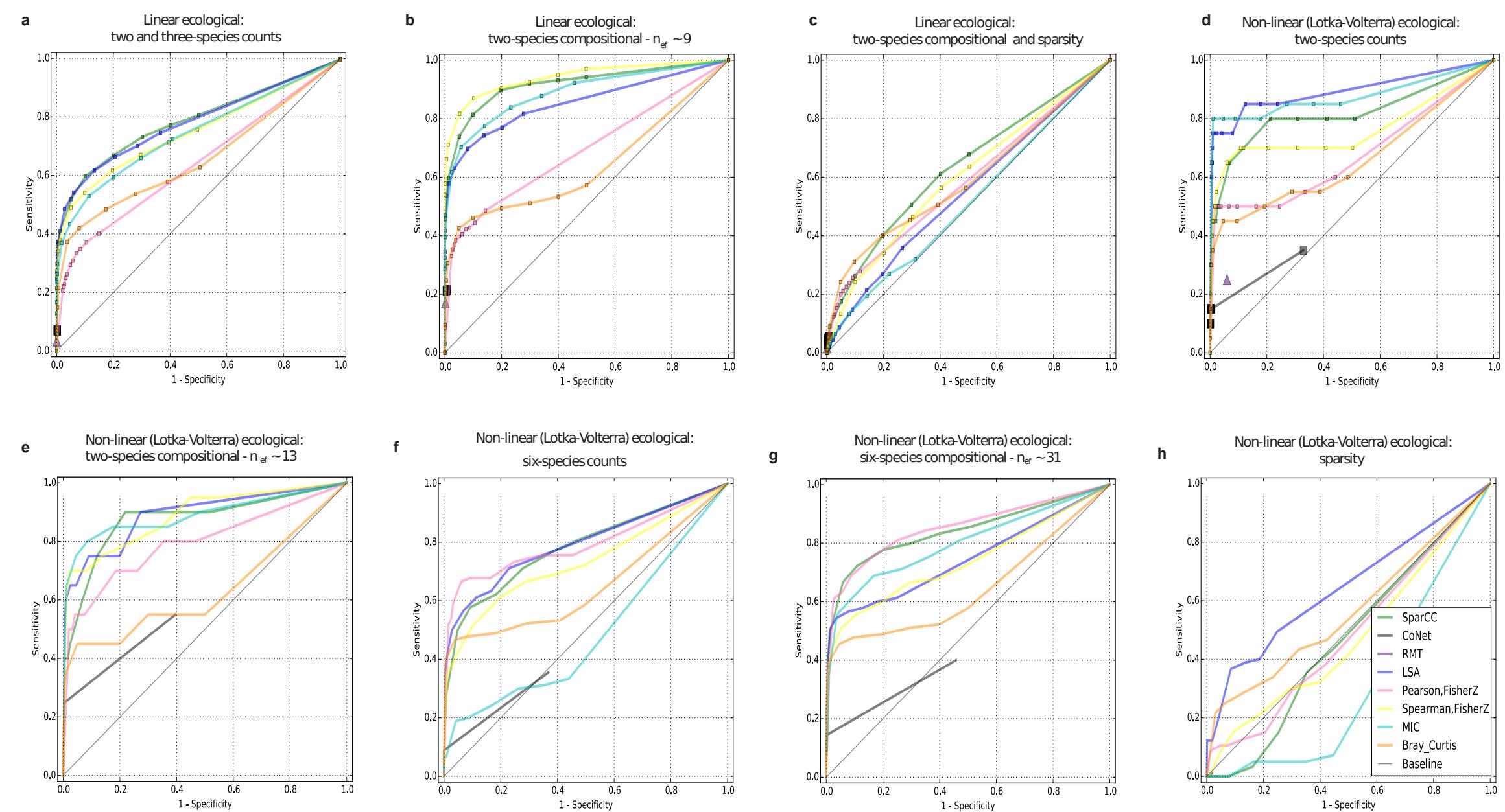
Supplementary Figure 10:

The false positive rate (FPR) plotted at different p-values for the different metrics and toolkits. (a) The FPR for the null table with features drawn for different distributions. (b) The FPR for the null table created from random samplings of a Dirichlet distribution modeled on real data.

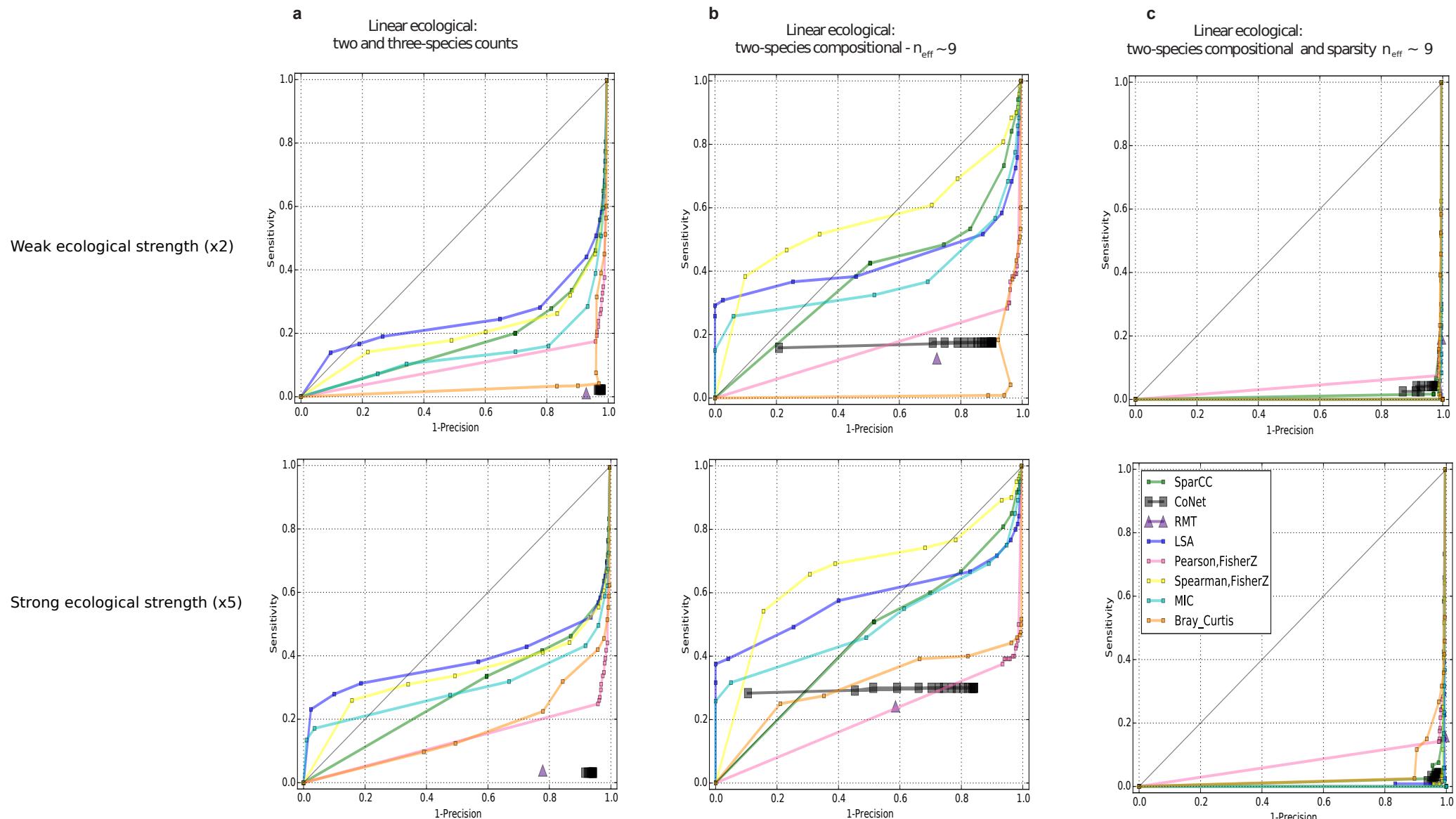


Supplementary Figure 11:

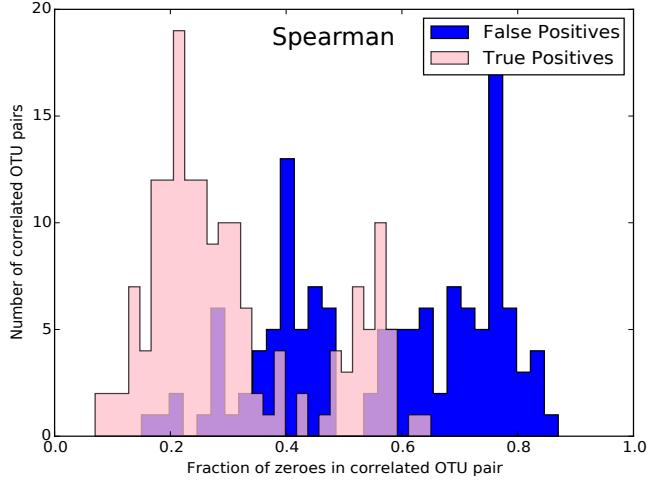
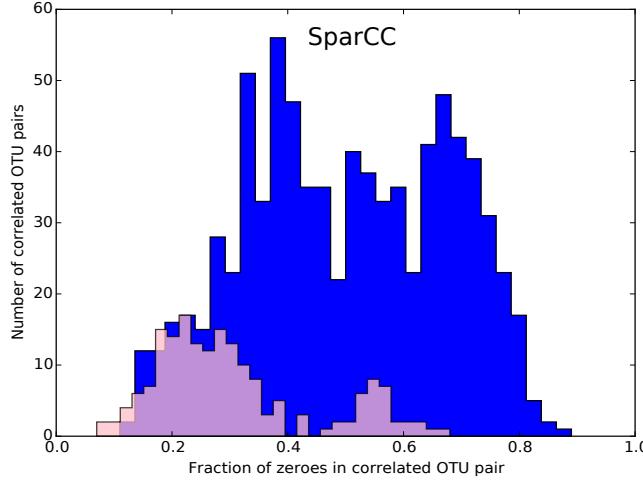
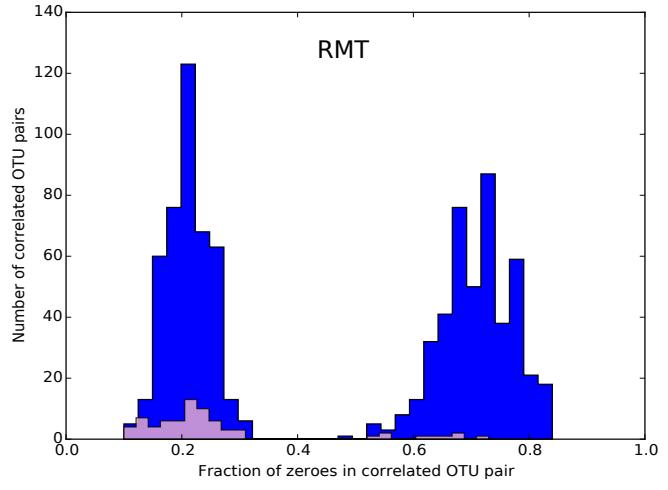
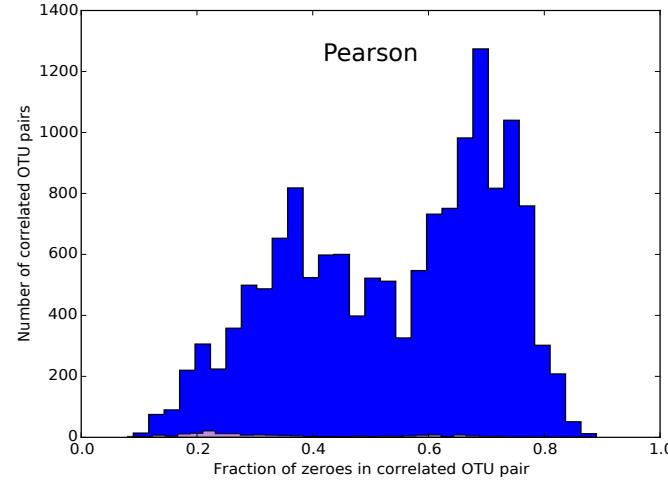
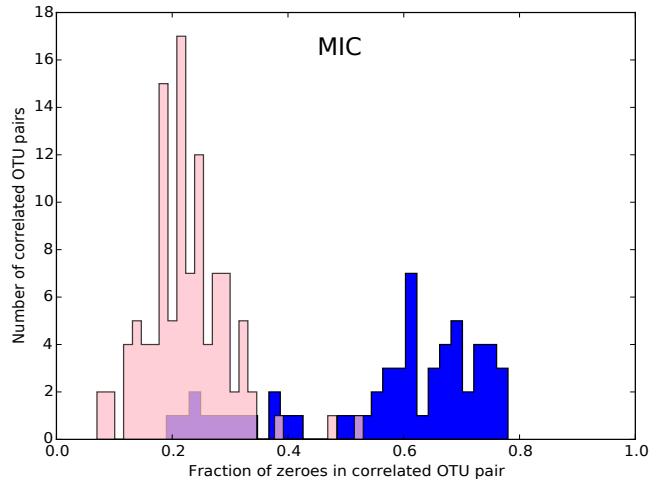
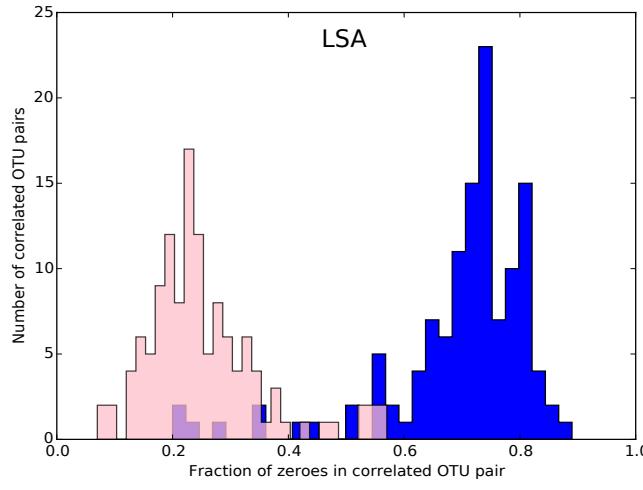
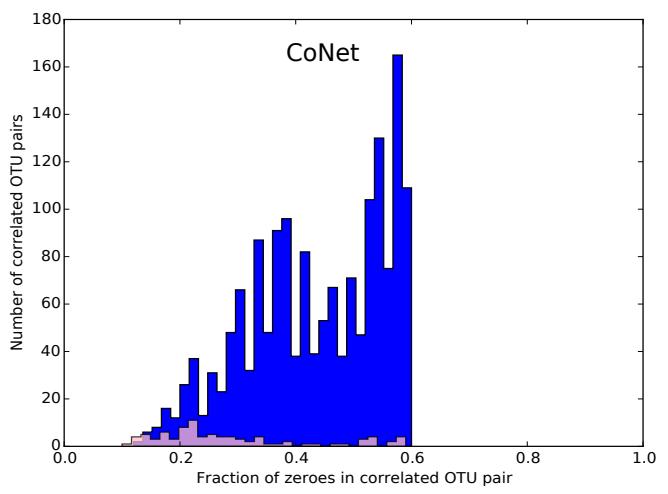
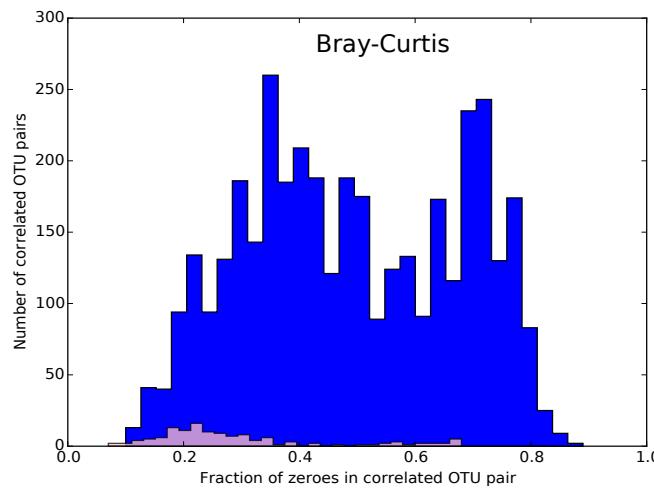
Number of features from each null distribution type deemed significant by the given metric or toolkit. One hundred features were drawn from each distribution.



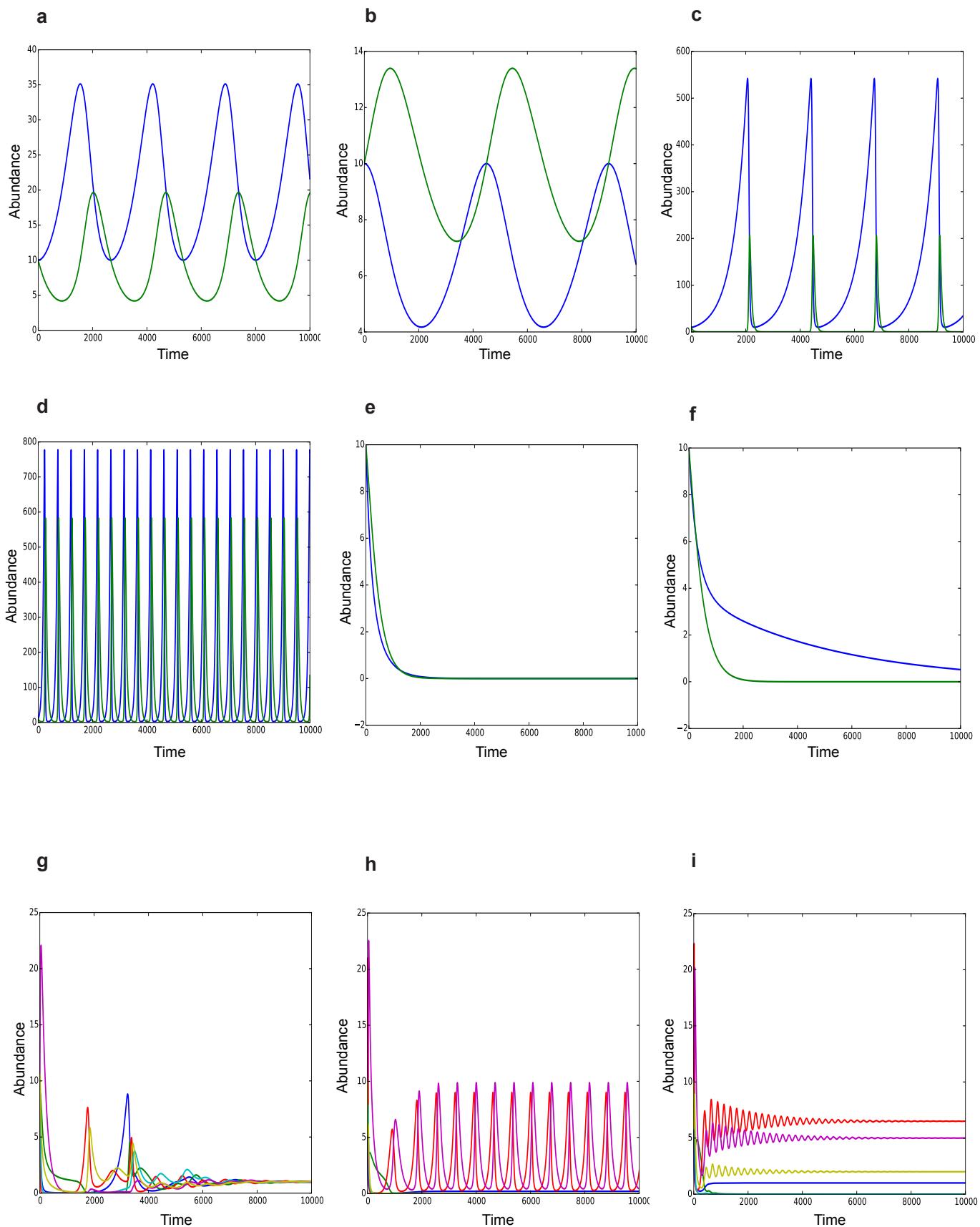
Supplementary Figure 12: Receiver Operating Characteristic (ROC) curves for linear ecological relationships (a-c) and non-linear/Lotka-Volterra ecological relationships (d-h). All tables were approximately 40% sparse, except (c) and (h), which were 70% sparse. Labels as in Fig. 4.



Supplementary Figure 13: Tool precision is extremely low under realistic microbiome dataset conditions. Precision vs. recall (sensitivity) curves for linear (a-c) ecological relationships. Analogous to Fig. 4 and Supplementary Fig 12, except constructed for strong and weak engineered correlations separately.

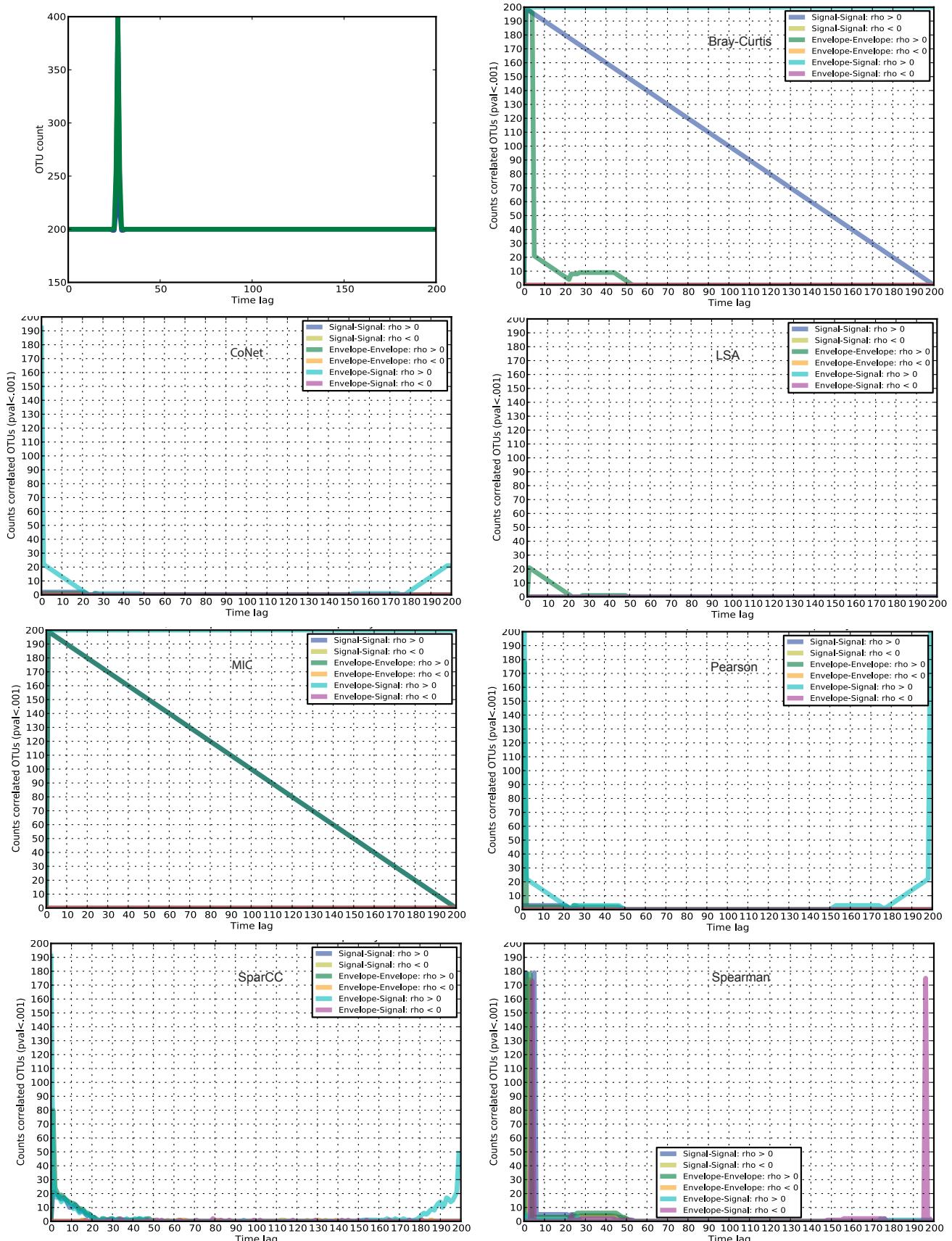


Supplementary Figure 14: OTU sparsity is a significant factor in true/false positive determination. Linear ecological data of Fig. 4b-c, Supplementary Figures 12b-c, and 13b-c. A p-value threshold of 0.001 was used, except for RMT. A threshold of 0.0001 was used for MIC and Spearman, since it increased resolution, otherwise the plots looked similar to SparCC's. Changing the p-value of any other technique did not result in a difference.



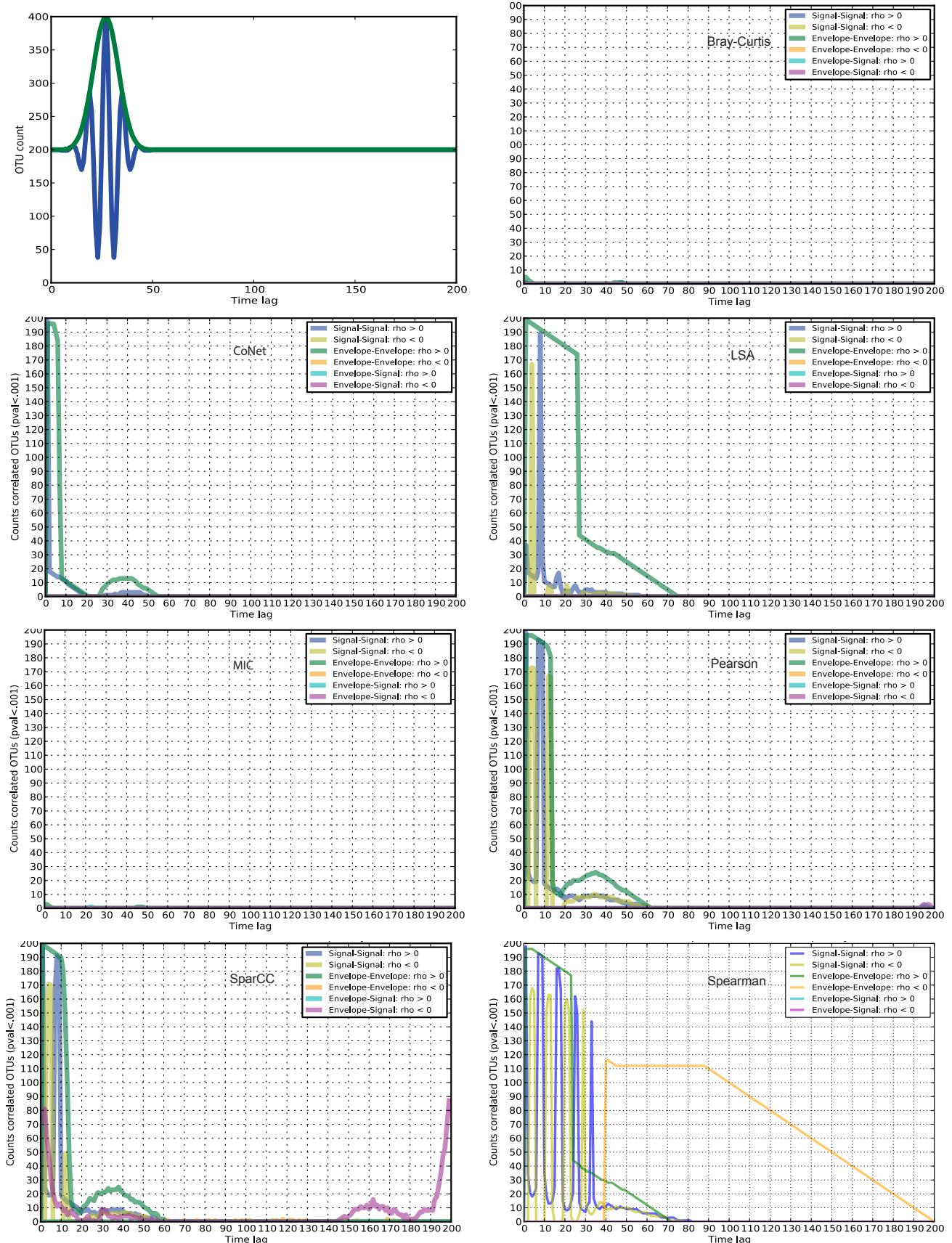
Supplementary Figure 15:

Visual depictions of feature count dynamics over time of the ecological Lotka-Volterra equation systems tested. (a-f) Two-species Lotka-Volterra relationships. The green line is the values of feature 1, and the blue line is feature 2 generated from the coupled differential equations. (g-i) Six-species Lotka-Volterra relationships. The abundance values of the six species over time are shown in different colors.



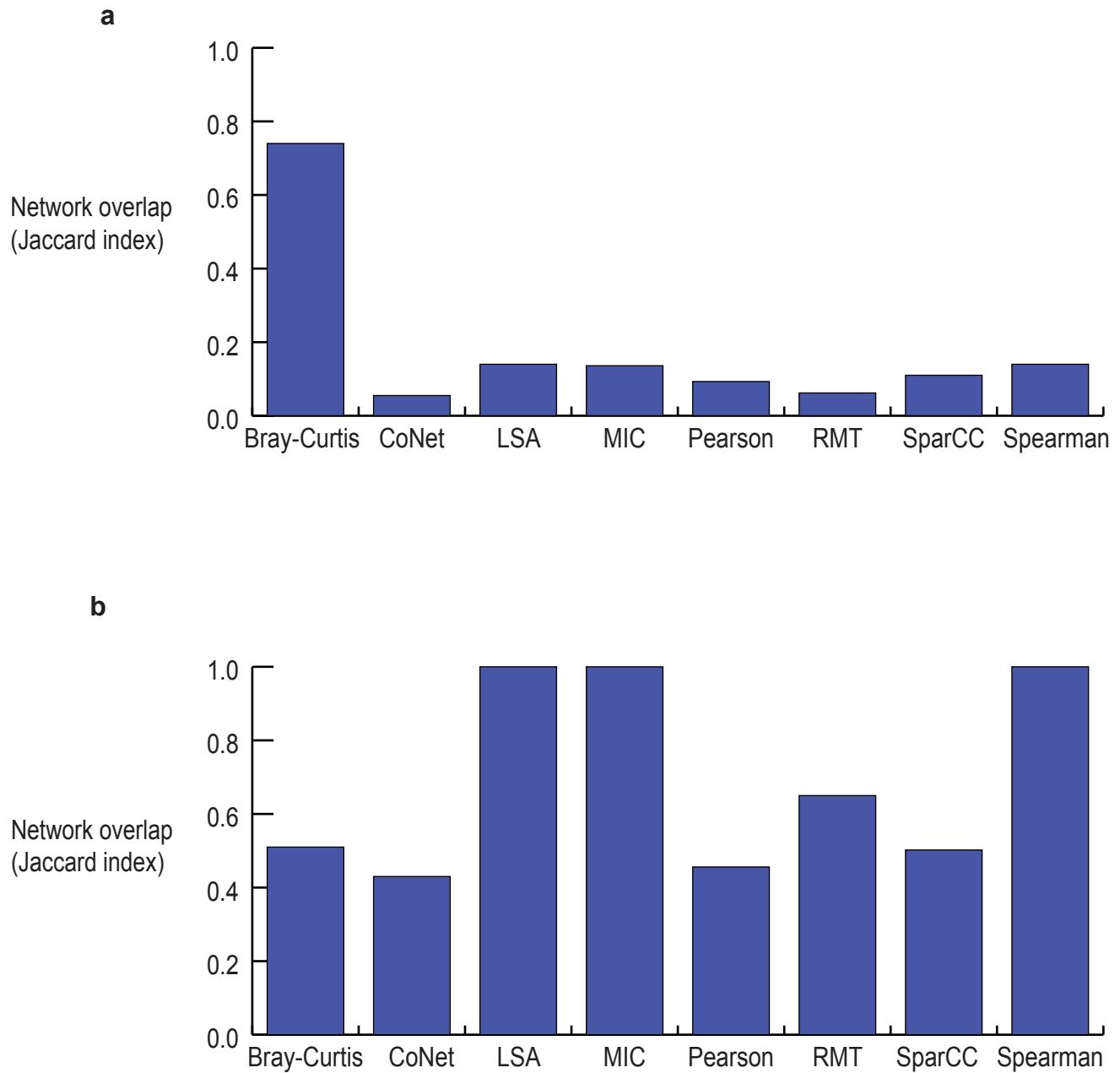
Supplementary Figure 16:

Time series correlations: data containing features with 10Hz abundance pulse at some point in time. Top left panel: Example pulse (blue) and envelope (green) features exhibiting a 10Hz spike in abundance. There were 200 pulse features with the abundance peak at times 1-200 of 200 time point samples. There were also 200 envelope features with the same placement throughout time as the pulses. Other panels show the count of feature pairs (e.g. signal-signal) with a given time lag and correlation sign for each tool. RMT is not included as these signals were too noisy for the technique.



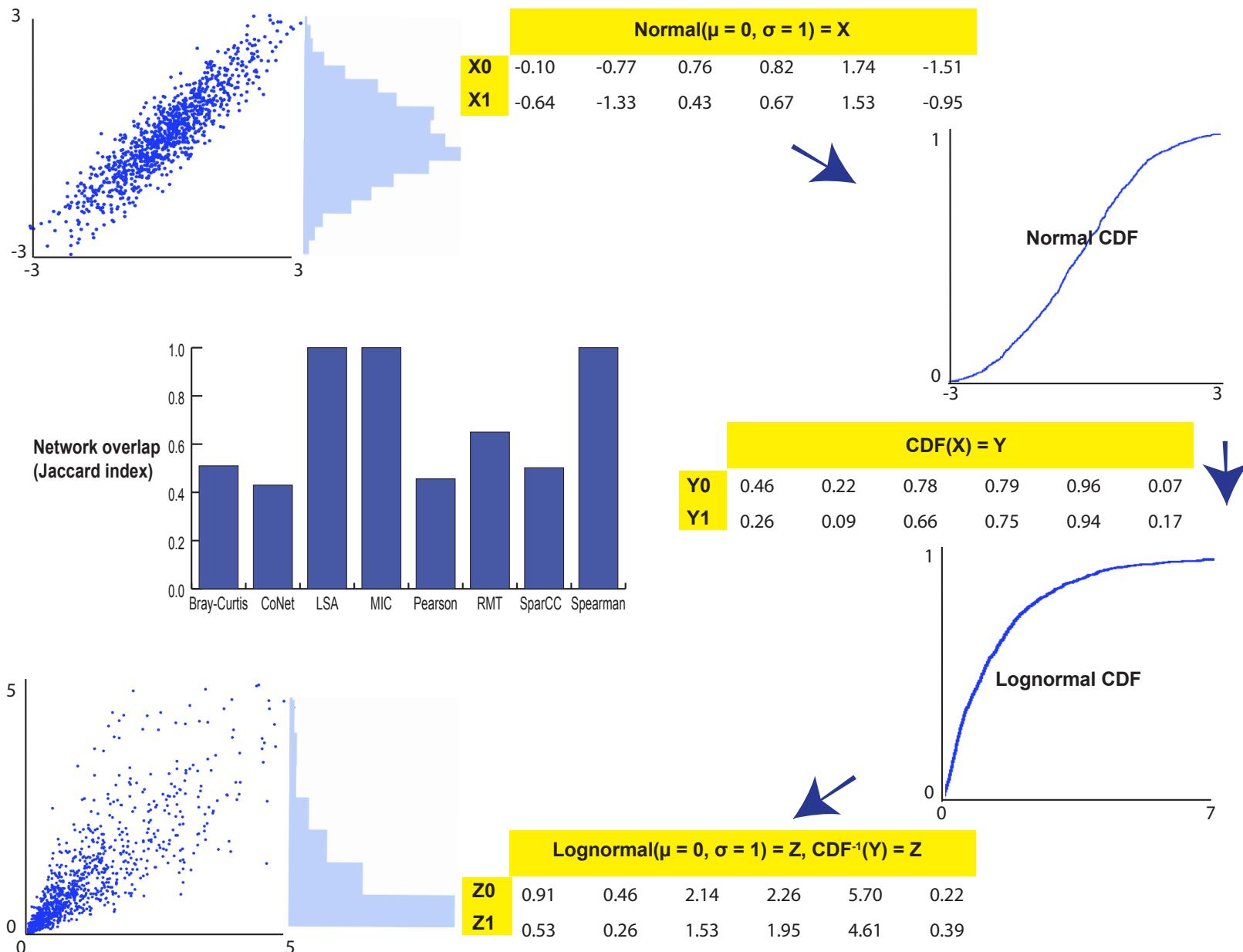
Supplementary Figure 17:

Time series correlations: data containing features with 1Hz abundance pulse at some point in time. Top left panel: Example pulse (blue) and envelope (green) features exhibiting a 1Hz pulse in abundance. There were 200 pulse features with the abundance peak at times 1-200 of 200 time point samples. There were also 200 envelope features with the same placement throughout time as the pulses. Other panels show the count of feature pairs (e.g. signal-signal) with a given time lag and correlation sign for each technique. RMT is not included as these signals were too noisy for the technique.



Supplementary Figure 18:

Sequencing technology, and therefore distribution significantly affects inferred correlation networks. (a) Jaccard index (edge intersection/edge union) showing network overlap by the same technique on 454 sequencing and Illumina datasets. The only difference between the datasets was sequencing technology; they were normalized in the same manner and then filtered to contain only the same OTUs (b) Correlation network overlap on datasets with the same rank correlation matrix but different distributions; generated by the copula methodology. Bray-Curtis only detected one or zero edges, causing more variability in the Jaccard index.



Supplementary Figure 19:

Distribution affects some correlation network tools, even if the underlying rank correlation matrix is the same. In the copula methodology, two features with normally distributed scores are converted to their cumulative distribution function (CDF) value for the normal distribution (top). The corresponding CDF score on the lognormal CDF is converted to its lognormal distribution value (bottom). The average fraction of shared feature pairs between copula tables with the same correlation matrix but different distributions is determined (center left) for each tool. This is the same plot as Supplementary Fig. 18b.