nature methods

Article

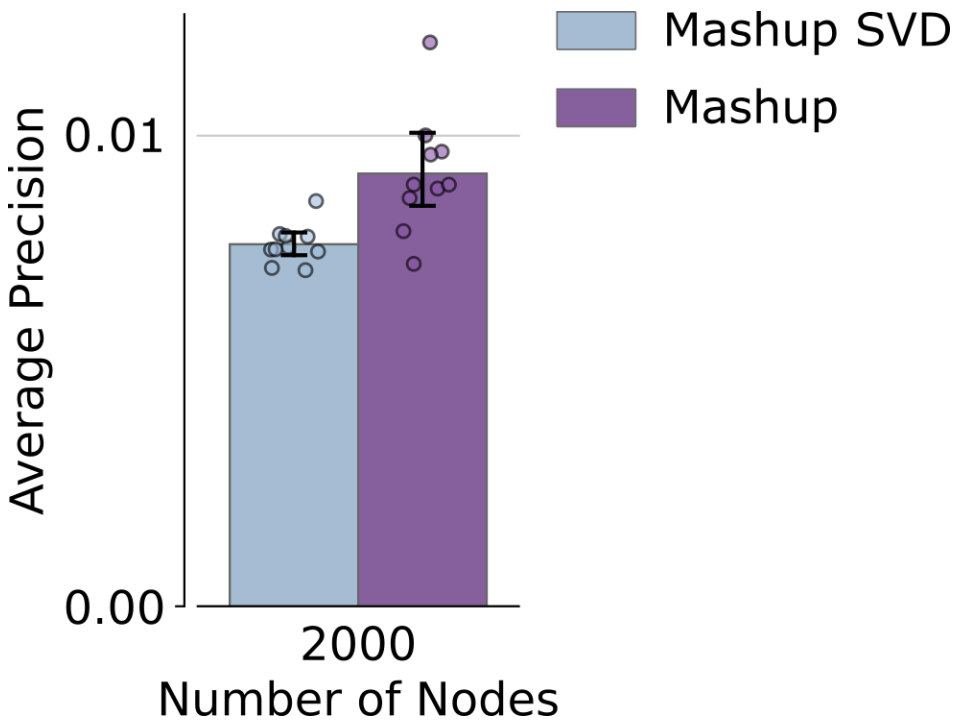
https://doi.org/10.1038/s41592-022-01616-x

BIONIC: biological network integration using convolutions

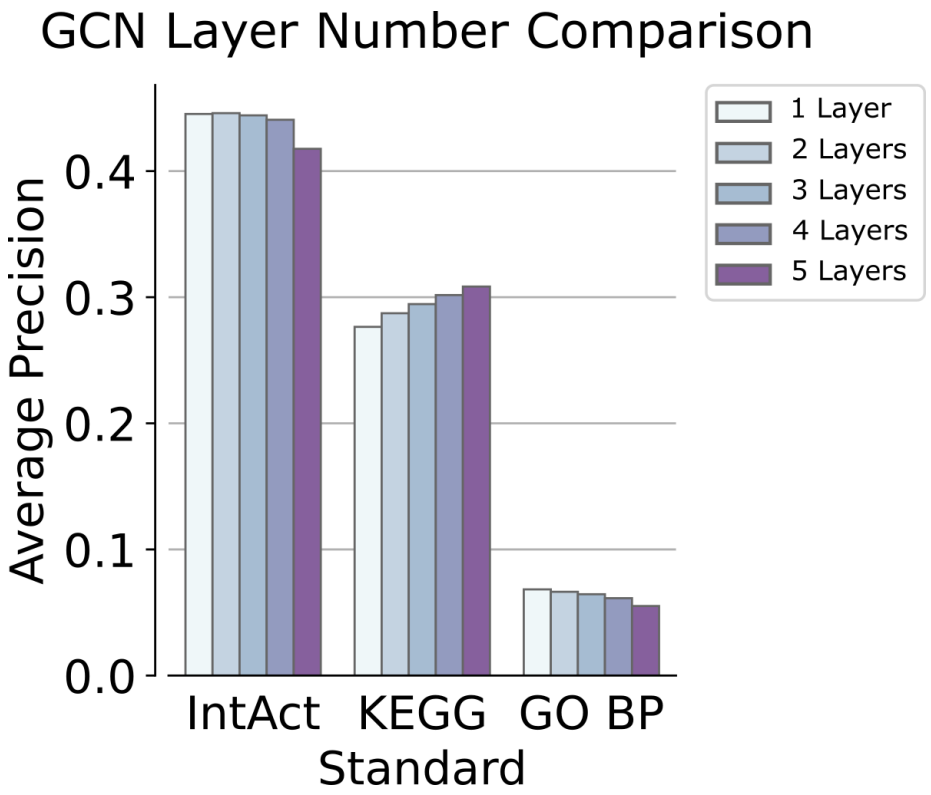
In the format provided by the authors and unedited

Supplementary Information

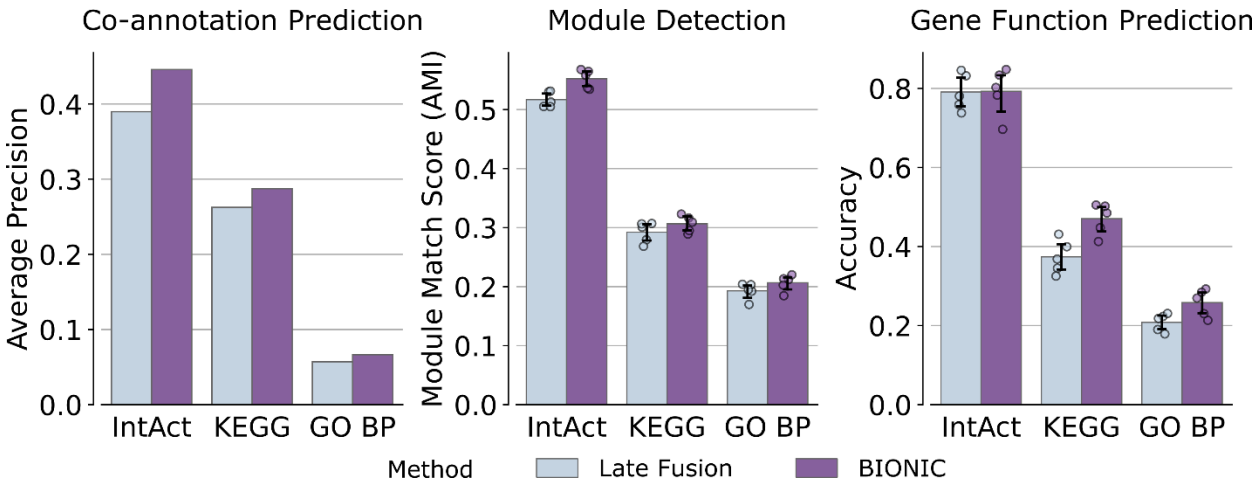
Supplementary Figures



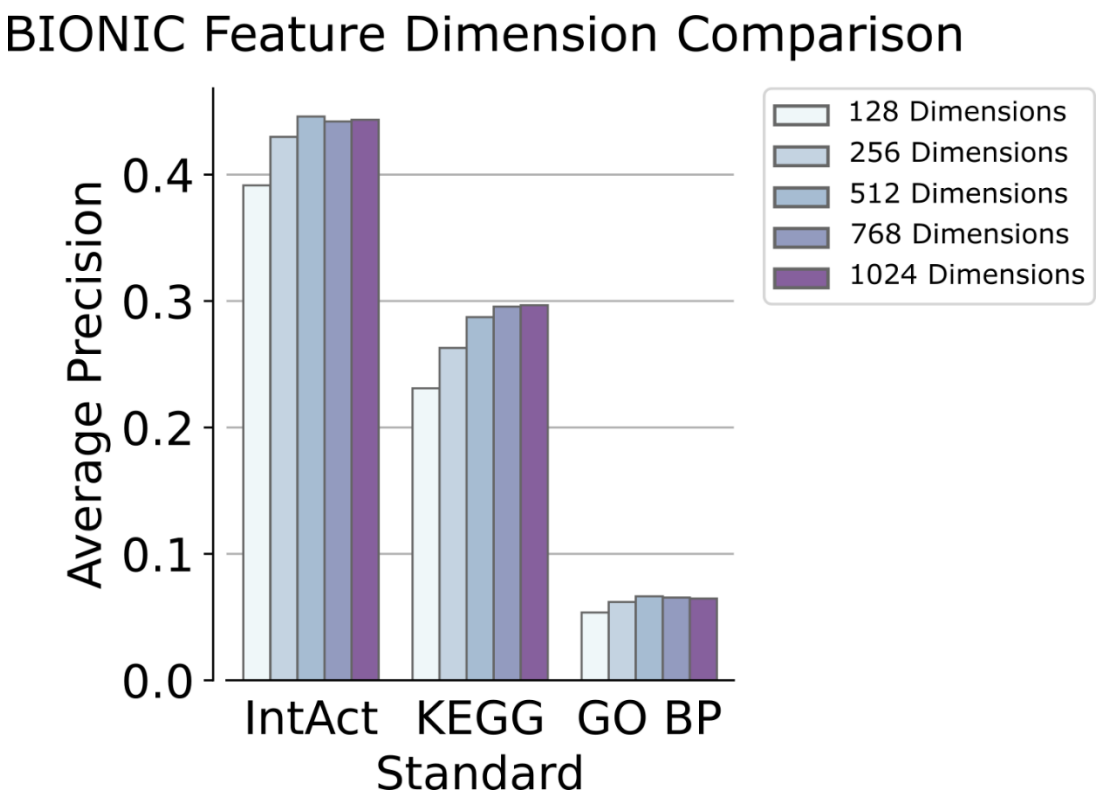
Supplementary Figure 1: Comparison of Mashup and Mashup singular value decomposition approximation. Comparison of Mashup with the author provided singular value decomposition approximation of Mashup (denoted Mashup SVD). This evaluation was performed on the 2000 node scenario from **Fig. 5b**. Data are presented as mean values. Error bars indicate the 95% confidence interval for n=10 independent samples.



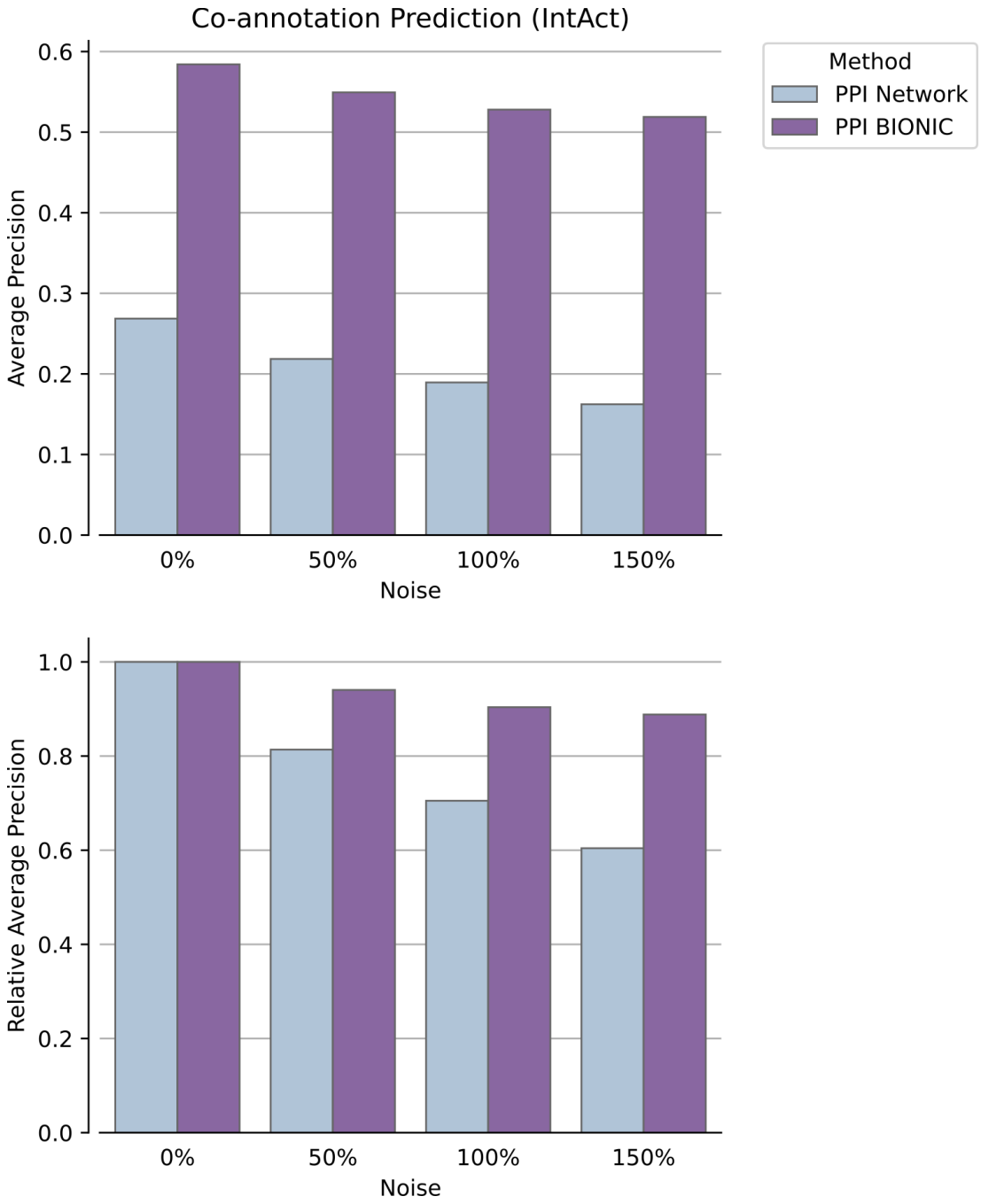
Supplementary Figure 2: BIONIC performance as a function of encoder layers. A co-annotation performance comparison of the unsupervised BIONIC with multiple choices of GCN layers. The number of layers corresponds to the effective neighborhood size (in hops) that is aggregated to update a given node.



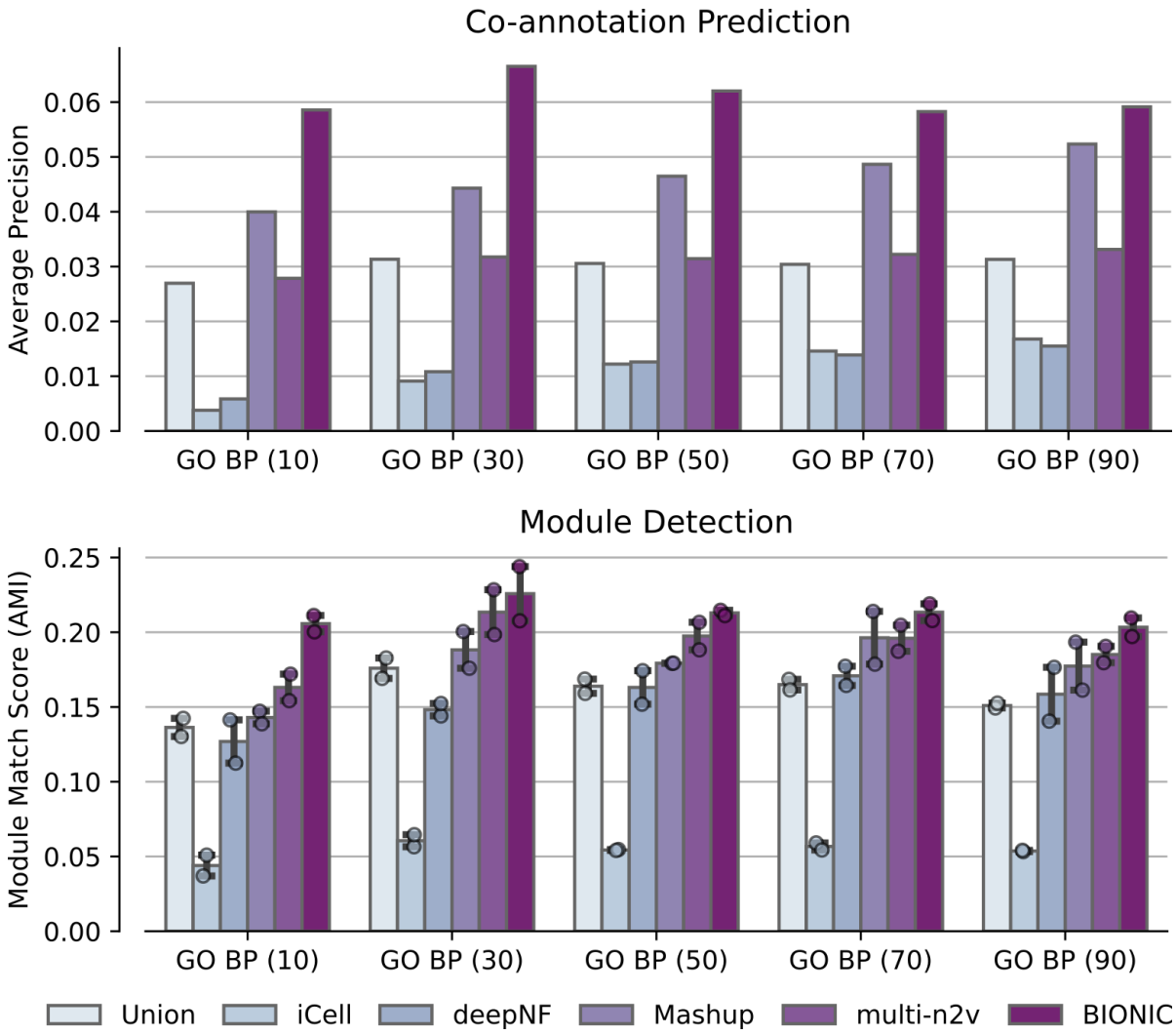
Supplementary Figure 3: BIONIC performance comparison with a late fusion approach. Evaluation of the three yeast network (PPI, COEX, GI) unsupervised BIONIC integration (referred to here as “BIONIC”), and an unsupervised late fusion approach, in which integrated network features are not learned jointly. Data are presented as mean values. Error bars indicate the 95% confidence interval for n=5 independent samples.



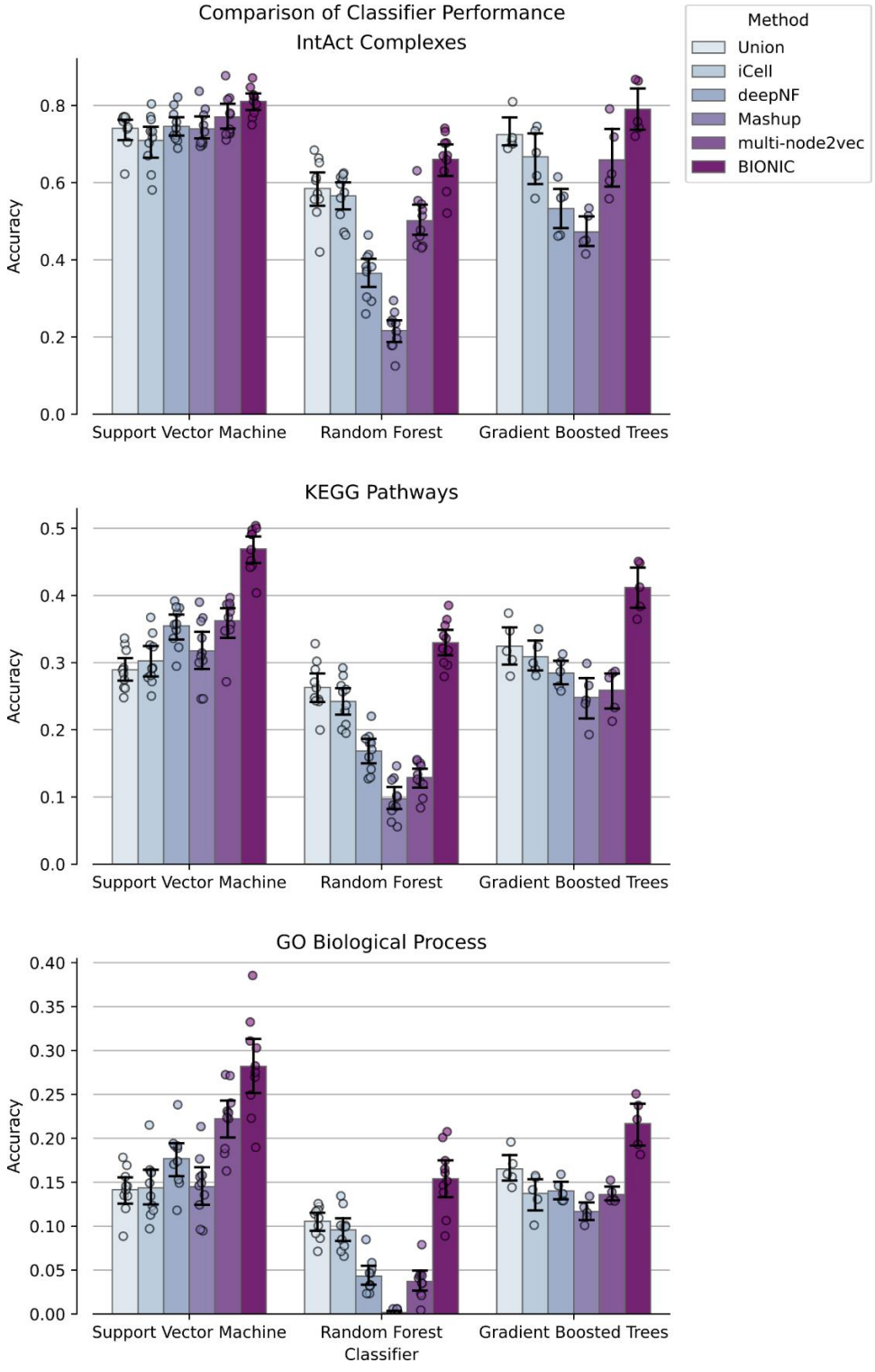
Supplementary Figure 4: BIONIC performance as a function of feature space dimension. Co-annotation prediction performance comparison of different unsupervised BIONIC feature space dimensions. 512 dimensions were used in this work.



Supplementary Figure 5: BIONIC denoising capabilities. Comparison of co-annotation performance (IntAct Complexes) between noisy versions of a yeast PPI network and the unsupervised BIONIC features learned using these networks individually. The top plot shows absolute co-annotation performance and the bottom plot shows the performance relative to a no added noise scenario (i.e. 0% noise). Percentages indicate the amount of added random edges relative to the number of edges in the original PPI network. Here 0% indicates no added random edges, 50% indicates a random edge was added for every two true edges in the original PPI network, and 100% indicates a random edge was added for each true edge.



Supplementary Figure 6: GO term size filtering effect on integration method performance. Comparison of unsupervised integration method performance for various GO Biological Process (BP) term filtering approaches (where numbers in parentheses denote the maximum GO term size). A filtering threshold of 30 was used in this work. Data are presented as mean values. Error bars indicate the 95% confidence interval for n=2 independent samples.



Supplementary Figure 7: Classifier type effect on integration method gene function prediction performance. Comparison of support vector machine, random forest and gradient boosted trees classifiers for unsupervised integration methods and functional standards. Data are presented as mean values. Error bars indicate the 95% confidence interval for n=10 independent samples for the support vector machine and random forest classifiers, and n=5 independent samples for the gradient boosted trees classifier.

Supplementary Notes

Supplementary Note 1

We performed an analysis to examine how the BIONIC features encode information from the input networks. We hierarchically clustered the integrated features over the feature dimensions (rather than over genes, as in **Fig. 2**) and extracted seven clusters of feature dimensions. We then evaluated how accurately these clusters predict edges (gene-gene relationships) in the three input networks (**Extended Data Fig. 6**). Since the number of feature dimensions correlates with performance, we also created a baseline for each cluster by randomly sampling the same number of feature dimensions from the full set of BIONIC features. We hypothesized that large differences in performance between the feature dimension clusters and the corresponding baselines implies BIONIC is using certain groups of feature dimensions to encode certain networks, rather than using all dimensions to encode all networks. In the case of the co-expression network, we see that some clusters show different performance than the baselines, however, this effect is relatively small, and not present for the PPI or genetic interaction network. Additionally, the full set of BIONIC feature dimensions consistently outperforms the clustered feature dimensions on all networks, suggesting that all feature dimensions are used to encode the input network information (albeit at slightly different levels), rather than in a small set of dimensions exclusively.

Supplementary Note 2

To ensure the integration results are consistent under a different set of input networks, and the wealth of yeast-two-hybrid (Y2H) networks available for yeast proteins, we selected the five largest of these networks[2,41–44] to integrate, and then compared the resulting performance of the integration approaches (**Extended Data Fig. 7**). These networks consisted of 453, 1248, 707, 927, and 776 proteins and 3258, 1778, 940, 866, and 784 interactions, respectively. We found that BIONIC substantially outperforms the established integration methods across functional standards and evaluation types.

Supplementary Note 3

We analyzed the utility of labelled data in a scenario where the labels are subjected to random permutation noise (**Extended Data Fig. 8a**). This was done to determine how robust the semi-supervised approach is to noise, compared to the unsupervised approach which uses no labels. We found that, with respect to co-annotation prediction and module detection, the semi-supervised BIONIC outperforms the unsupervised BIONIC for low to moderate amounts of label noise. Interestingly, on the gene function prediction evaluation we find that the unsupervised BIONIC outperforms the semi-supervised BIONIC even for low label noise, despite training on the same set of permuted labels. This is likely because the unsupervised approach does not incorporate incorrect label information into the learned features unlike the semi-supervised approach, so information reflecting true biology is captured more accurately in the unsupervised features leading to better label predictions. We also examined an instance of label noise resulting in the dissolution of a protein complex under the semi-supervised training scenario (**Extended Data Fig. 8b**). Here, the general transcription factor complex (TFIID) is captured more effectively in the semi-supervised case when label noise is low, but it loses members as the label noise increases. For high label noise scenarios, the unsupervised BIONIC is able to more effectively capture the TFIID complex.

Supplementary Note 4

We tested the extent of BIONIC scalability in terms of graphics processing unit (GPU) memory usage and training epoch time (**Extended Data Fig. 9**). This analysis was done with respect to network quantity and network size jointly to determine the relationship between these factors as it pertains to scalability. Here, random networks were generated with varying numbers of nodes such that the average node degree was 30. These were integrated using BIONIC and memory consumption and runtime were recorded. We found that for networks with 8000 nodes or fewer, BIONIC can scale to at least 90 of these networks without exhausting memory or dramatically increasing runtime. For human sized networks (in the worst case consisting of 20,000 nodes) BIONIC can scale to 5-10 networks without considerably increasing runtime, and 20 networks with longer runtimes. Sparser networks than those tested here may lead to further increases in scalability.