Statistical Inferences from Noisy and Incomplete Protein Interaction Network Data

Michael P.H. Stumpf

Centre for Bioinformatics & Institute for Mathematical Sciences,
Imperial College London. London SW7 2AZ, UK.

Present protein interaction datasets are plagued by large false-positive and false-negative rates. In their present state they moreover represent only a fraction of the true interaction network. Here we show how an error model for protein interaction data can be used to (i) assess and compare different datasets, (ii) assign statistical weights to reported interactions, and (iii) construct confidence intervals for network analyses. We furthermore show how this error model combined with multi-model inference (MMI) and Bayesian model averaging can be used to infer properties of a true interaction network from present noisy and incomplete network datasets. A detailed mathematical analysis of the properties of noisy and incomplete networks shows that incompleteness rather than reported levels of noise affect our inferences more profoundly. The power of the MMI approach for inferring properties of networks from noisy and incomplete subnets is illustrated by applications to *S.cerevisiae* and human protein interaction data.

References

Stumpf, Wiuf & May, PNAS **102** (2005), 4221-4224.
Wiuf, Brameier, Hagberg & Stumpf, PNAS **103** (2005), 7566-7570.
de Silva et al. BMC Biology **4** (2006), 39.