

PROBABILISTIC MODELING OF SYSTEMATIC ERRORS IN TWO-HYBRID EXPERIMENTS

DAVID SONTAG*

ROHIT SINGH*

BONNIE BERGER^{†‡}

*Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge MA 02139*

E-mail: {dsontag, rsingh, bab}@mit.edu

We describe a novel probabilistic approach to estimating errors in two-hybrid (2H) experiments. Such experiments are frequently used to elucidate protein-protein interaction networks in a high-throughput fashion; however, a significant challenge with these is their relatively high error rate, specifically, a high false-positive rate. We describe a comprehensive error model for 2H data, accounting for both random and systematic errors. The latter arise from limitations of the 2H experimental protocol: in theory, the reporting mechanism of a 2H experiment should be activated if and only if the two proteins being tested truly interact; in practice, even in the absence of a true interaction, it may be activated by some proteins – either by themselves or through promiscuous interaction with other proteins. We describe a probabilistic relational model that explicitly models the above phenomenon and use Markov Chain Monte Carlo (MCMC) algorithms to compute both the probability of an observed 2H interaction being true as well as the probability of individual proteins being self-activating/promiscuous. This is the first approach that explicitly models systematic errors in protein-protein interaction data; in contrast, previous work on this topic has modeled errors as being independent and random. By explicitly modeling the sources of noise in 2H systems, we find that we are better able to make use of the available experimental data. In comparison with Bader *et al.*'s method for estimating confidence in 2H predicted interactions, the proposed method performed 5-10% better overall, and in particular regimes improved prediction accuracy by as much as 76%.

Supplementary Information: <http://theory.csail.mit.edu/probmod2H>

1. Introduction

The fundamental goal of systems biology is to understand how the various components of the cellular machinery interact with each other and the environment. In pursuit of this goal, experiments for elucidating protein-protein interactions (PPI) have proven to be one of the most powerful tools available. Genome-wide, high-throughput PPI experiments have started to

*These authors contributed equally to the work

[†]Corresponding author

[‡]Also in the MIT Dept. of Mathematics

provide data that has already been used for a variety of tasks: for predicting the function of uncharacterized proteins; for analyzing the relative importance of proteins in signaling pathways; for new perspectives in comparative genomics, by cross-species comparisons of interaction patterns etc. Unfortunately, the quality of currently available PPI data is unsatisfactory, which limits its usefulness to some degree. Thus, techniques that enhance the availability of high-quality PPI data are of value.

In this paper, we aim to improve the quality of experimentally available PPI data by identifying erroneous datapoints from PPI experiments. We attempt to move beyond current one-size-fits-all error models that ignore the experimental source of a PPI datapoint; instead, we argue that a better error model will also have components tailored to account for the systematic errors of specific experimental protocols. This may help achieve higher sensitivity without sacrificing specificity. This motivated us to design an error model tailored to one of the most commonly-used PPI experimental protocols.

We specifically focus on data from *two hybrid* (2H) experiments^{6,4}, which are one of the most popular high-throughput methods to elucidate protein-protein interaction. Data from 2H experiments forms the majority of the known PPI data for many species: *D. melanogaster*, *C. elegans*, *H. sapiens* etc. However, currently available 2H data also has unacceptably high false-positive rates: von Mering *et al.* estimate that more than 50% of 2H interactions are spurious¹¹. These high rates of error seriously hamper the ability to perform analyses of the PPI data. As such, we believe an error model that performs better than existing models — even if it is tailored to 2H data — is of significant practical value, and may also serve as an example for the development of error models for other biological experiments.

Ideally, the reporting mechanism in a 2H experiment is activated if and only if the pair of proteins being tested truly interact. As in most experimental protocols, there are various sources of random noise. However, there are also systematic, repeatable errors in the data, originating from limitations in the 2H protocol. In particular, there exist proteins that are disproportionately prone to be part of false-positive observations (Fig. 1). It is thought that these proteins either activate the reporting mechanism by themselves or promiscuously bind with many other proteins in the particular setup (promiscuous binding is an experimental artifact— it does not imply a true interaction under plausible biological conditions).

Contributions: The key contribution of this paper is a comprehensive error model for 2H experiments, accounting for both random as well as systematic errors, which is guided by insights into the systematic errors of the 2H experi-

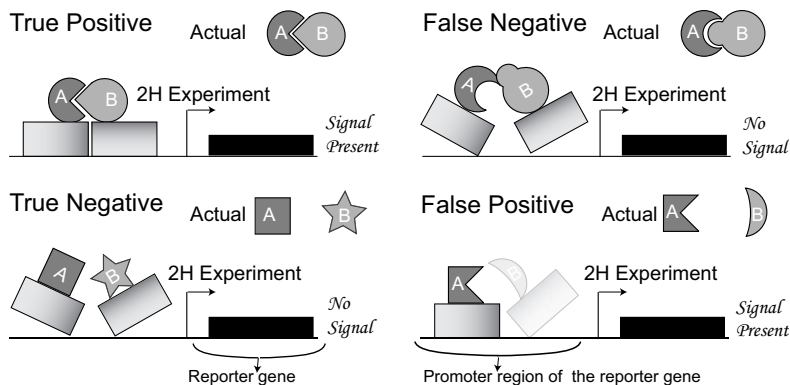


Figure 1: The origin of systematic errors in 2H data. The cartoons shown above demonstrate the mechanism of 2H experiments. Protein A is fused to the DNA binding domain of a particular transcription factor, while protein B is fused to the activation domain of that transcription factor. If A and B physically interact then the combined influence of their respective enhancers results in the activation of the reporter gene. Systematic errors in such experiments may arise: false negatives occur when two proteins which interact *in-vivo* fail to activate the reporter gene under experimental conditions. False positives may occur due to proteins which trigger the reporting mechanism of the system, either by themselves (self-activation) or by spurious interaction with other proteins (promiscuity). Spurious interaction can occur when a protein is grossly over-expressed. In the above figure, protein A in the lower right panel is such a protein: it may either promiscuously bind with B or activate the reporting mechanism even in the absence of B.

mental protocol. We believe this is the first model to account for both sources of error in a principled manner; in contrast, previous work on estimating error in PPI data has assumed that the error in 2H experiments (as in other experiments) is independent and random. Another contribution of the paper are estimates of proteins especially likely to be self-activating/promiscuous (see Supp. Info.). Such estimates of “problem proteins”, may enable the design of 2H experimental protocols which have lower error rates.

We use the framework of Bayesian networks to encode our assumption that a 2H interaction is likely to be observed if the corresponding protein pair truly interacts or if either of the proteins is self-activating/promiscuous. The Bayesian framework allows us to represent the inherent uncertainty and the relationship between promiscuity of proteins, true interactions and observed 2H data, while using all the data available to simultaneously learn the model parameters and predict the interactions. We use a Markov Chain Monte Carlo (MCMC) algorithm to do approximate probabilistic inference in our models, jointly inferring both desired sets of quantities: the probability of interaction, and the propensity of a protein for self-activation/promiscuity.

We show how to integrate our error model into the two most common

probabilistic models used for combining PPI experimental data, and show that our error model can significantly improve the accuracy of PPI prediction. **Related Work:** With data from the first genome-wide 2H experiments (Ito *et al.*⁶, Uetz *et al.*⁴), there came the realization that 2H experiments may have significant systematic errors. Vidalain *et al.* have identified the presence of self-activators as one of the sources of such errors, and described some changes in the experimental setup to reduce the problem¹⁰. Our work aims to provide a parallel, computational model of the problem, allowing post-facto filtering of data, even if the original experiment retained the errors. The usefulness of such an approach was recently demonstrated by Sun *et al.*² (to reconstruct transcriptional regulatory networks).

Previous computational methods of modeling systematic errors in PPI data can be broadly classified into two categories. The first class of methods^{5,11,8} exploits the observation that if two very different experimental setups (e.g. 2H and Co-IP) observe a physical interaction, then the interaction is likely to be true. This is a reasonable assumption to make because the systematic errors of two different experimental setups are likely to be independent. However, this approach requires multiple costly and time consuming genome-wide PPI experiments, and may still result in missed interactions, since the experiments have high false negative rates. Many of these approaches also integrate non-PPI functional genomic information, such as co-expression, co-localization, and Gene Ontology functional annotation.

The second class of methods is based on the topological properties of the PPI networks. Bader *et al.*¹, in their pioneering work, used the number of 2H interactions per protein as a negative predictor of whether two proteins truly interact. Since the prior probability of any interaction is small, disproportionately many 2H interactions involving a particular protein could possibly be explained by it being self-activating or promiscuous. However, such an approach is unable to make fine-grained distinctions: an interaction involving a high-degree protein need not be incorrect, especially if there is support for it from other experiments. Furthermore, the high degree of a promiscuous protein in one experiment (e.g. Ito *et al.*'s) should not penalize interactions involving that protein observed in another experiment (e.g. Uetz *et al.*'s) if the errors are mostly independent (e.g. they use different reporters). Our proposed probabilistic models solve all of these problems.

2. Data Sets

One difficulty with validating any PPI prediction method is that we must have a gold standard from which to say whether two proteins interact or do

not interact. We constructed a gold standard data set of protein-protein interactions in *S. cerevisiae* (yeast) from which we could validate our methods.

Our gold standard test set is an updated version of Bader *et al.*'s data. Bader *et al.*'s data consisted of all published interactions found by 2H experiments; data from experiments by Uetz *et al.*⁴ (the UETZ2H data set) and Ito *et al.*⁶ (the ITO2H data set) comprised the bulk of the data set. They also included as possible protein interactions all protein pairs that were of distance at most two in the 2H network. Bader *et al.* then used published Co-Immunoprecipitation (Co-IP) data to give labels to these purported interactions. When two proteins were found in a bait-hit or hit-hit interaction in Co-IP, they were labeled as having a true interaction. When two proteins were very far apart in the Co-IP network (distance larger than three), they were labeled as not interacting. We updated Bader *et al.*'s data to include all published 2H interactions through February 2006, getting our data from the MIPS⁷ database. We added, for the purposes of evaluation, recently published yeast Co-IP data from Krogan *et al.*³. This allowed us to significantly increase the number of labeled true and false interactions in our data set.

Since the goal of our algorithms is to model the systematic errors in large-scale 2H experiments, we evaluated our models' performance on the test data where at least one of UETZ2H or ITO2H indicated an interaction. We were left with 397 positive examples, 2298 negative examples, and 2366 unlabeled interactions. We randomly chose 397 of the 2298 negative examples to be part of our test set. For all of the experiments we performed 4-fold cross validation on the test set, hiding one fourth of the labels while using the remaining labeled data during inference.

3. Probabilistic Models

We show how to integrate our model of systematic errors into the two most common probabilistic models used for PPI prediction. Our first model is complementary to the relational probabilistic model proposed by Jaimovich *et al.*⁸, and can be easily integrated into their approach. Our second model is an extension of Bader *et al.*'s, and will form the basis of our comparison.

Our models also adjust to varying error rates in different experiments. For instance, while we account for random noise and false negatives in our error model for both UETZ2H and ITO2H, we only model self-activation/promiscuity for ITO2H observations. The UETZ2H data set was smaller and included only one protein with degree larger than 20; ITO2H had 36 proteins with degree larger than 30, including one with degree as high as 285. Thus, while modeling promiscuity made a big difference for the ITO2H

data, it did not significantly affect our results on the UETZ2H data.

3.1. Generative model

We begin with a simplified model of PPI interaction (Fig. 2). We represent the uncertainty about a protein interaction as an indicator random variable X_{ij} , which is 1 if proteins i and j truly interact, and 0 otherwise. For each experiment, we construct corresponding random variables (RVs) indicating if i and j have been observed to interact under that experiment. Thus, U_{ij} is the observed^a random variable (RV) representing the observation from UETZ2H, and I_{ij} is the observed RV representing the observation from ITO2H. The arrow from X_{ij} to I_{ij} indicates the dependency of I_{ij} on X_{ij} . The box surrounding the three RVs indicates that this template of three RVs is repeated for all $i, j = 1, \dots, N$ (i.e. all pairs of proteins), where N is the number of proteins. In all models of this type, the I_{ij} RVs are assumed to be independent of one another.

If an experiment provides extra information about each observation, the model can be correspondingly enriched. For instance, for each of their observed interactions Ito *et al.* provide the number of times the interaction was discovered (called the number of IST hits). Rather than making I_{ij} binary, we have it equal the number of IST hits, or 3 if $\text{IST} > 3$. We will refer to the portion of ITO2H observations with $\text{IST} \geq 3$ as ITOCORE.

The model is called “generative” because the ground truth about the interaction, X_{ij} , generates the observations in the 2H experiments, I_{ij} and U_{ij} . To our knowledge, all previous generative models of experimental interactions made the assumption that I_{ij} depended only on X_{ij} . They allowed for false positives by saying that $\text{Pr}(I_{ij} > 0 | X_{ij} = 0) = \delta_{fp}$, where δ_{fp} is a parameter of their model. Similarly, they allowed for false negatives by saying that $\text{Pr}(I_{ij} = 0 | X_{ij} = 1) = \delta_{fn}$, for another parameter δ_{fn} . However, these models are missing much of the picture. For example, many experiments have particular difficulty testing the interactions of proteins along the membrane. For these proteins, δ_{fn} should be significantly higher. In the 2H experiment, for interactions that involve self-activating/promiscuous proteins, δ_{fp} will be significantly higher.

In Fig. 3, we propose a novel probabilistic model in which the self-activating/promiscuous tendencies of particular proteins are explicitly modeled. The latent Bernoulli RV F_k is 1 if protein k is believed to be promiscuous or self-activating. In the context of our data set, this RV applies specifically to the ITO2H data; if self-activation/promiscuity in multiple ex-

^aClear nodes are unobserved (latent) RVs, and shaded nodes are observed RVs.

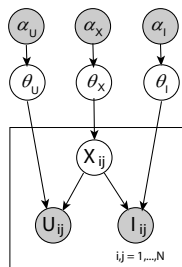


Figure 2: **Generative model.**

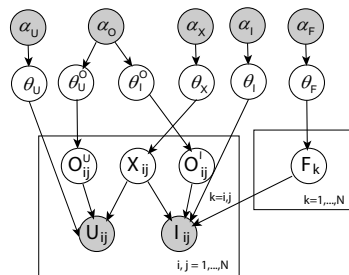


Figure 3: **Generative model, with noise variables.**

periments is to be modeled, we may introduce multiple such variables F_k^H (for protein k and experiment H). The I_{ij} RV thus depends on F_i and F_j . Intuitively, I_{ij} will be > 0 if either $X_{ij} = 1$ or $F_k = 1$. As we show later in the Results section, this model of noise is significantly more powerful than the earlier model, because it allows for the “explaining away” of false positives in ITO2H. Furthermore, it allows evidence from data sets other than ITO2H to influence (through the X_{ij} RVs) the determination of the F_k RVs. We also added the latent variables O_{ij}^U and O_{ij}^I , which will be 1 if the Uetz *et al.* and Ito *et al.* experiments, respectively, have the capacity to observe a possible interaction between proteins i and j . These RVs act to explain away the false negatives in UETZ2H and ITO2H. We believe that these RVs will be particularly useful for species where we have relatively little PPI data. The distributions in these models all have Dirichlet priors (θ) with associated hyperparameters α (see Supp. Info. for more details).

There are many advantages to using the generative model described in this section. First, it can easily handle missing data without adding complexity to the inference procedure. This is important for when integrating additional experimental data into the model. Suppose, for example, that we use gene expression correlation as an additional signal of protein interaction, by introducing new RVs E_{ij} (indicating coexpression of genes i and j) and corresponding edges $X_{ij} \rightarrow E_{ij}$. If, for a pair of proteins, the coexpression data is unavailable, we simply omit the corresponding E_{ij} from this model. In Bader *et al.*’s model, and the second model that we propose below, we would need to integrate over possible values of the missing datapoint, a potentially complicated task. Second, this generative model can be easily extended: e.g., we could easily combine this model with Jaimovich *et al.*’s in order to model the common occurrence of transitive closure in PPIs.

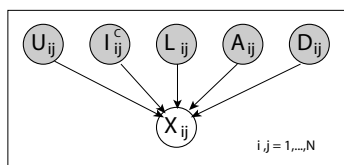


Figure 4: Bader *et al.*'s logistic regression model (BaderLR).

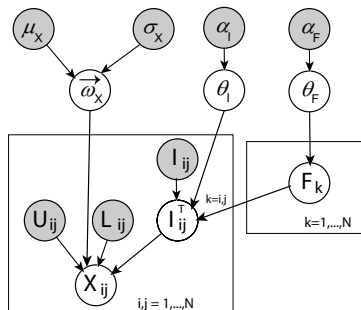


Figure 5: Our Bayesian logistic model, with noise variables (BayesLR).

3.2. Bayesian logistic model

In Fig. 4 we show Bader *et al.*'s model (BADERLR); it includes three new variables in addition to the RVs already mentioned, whose values are pre-calculated using the 2H network. Two of these encode topological information: variable A_{ij} is the number of adjacent proteins in common between i and j , and variable D_{ij} is $\ln(d_i + 1) + \ln(d_j + 1)$, where d_i is the degree of protein i . Variable L_{ij} is an indicator variable for whether this protein interaction has been observed in any low-throughput experiments. In Bader *et al.*'s model, I_{ij}^C is an indicator variable representing whether the interaction between proteins i and j was in the ITOCORE data set ($IST \geq 3$). X_{ij} 's conditional distribution is given by the logistic function:

$$p(X_{ij} = 1) = \frac{1}{1 + \exp(w_{offset} + U_{ij}w_U + I_{ij}^C w_I + L_{ij}w_L + A_{ij}w_A + D_{ij}w_D)}.$$

The weights w are discriminatively learned using the Iterative Re-weighted Least Squares (IRLS) algorithm, which requires that all of the above quantities are observed in the training data.

In Fig. 5 we propose a new model (BAYESLR), with two significant differences. First, we no longer use the two proteins' degree, D_{ij} , and instead integrate our noise model in the form of the F_k random variables. Second, instead of learning the model using IRLS, we assign the weights uninformative priors and do inference via Markov Chain Monte Carlo (MCMC). This will be necessary because X_{ij} will have an unobserved parent, I_{ij}^T . The new RV I_{ij}^T will be 1 when the Ito *et al.* experiment should be considered for predicting X_{ij} . Intuitively, its value should be $(I_{ij} > 0) \wedge \neg(F_i \vee F_j)$. However, to allow greater flexibility, we give the conditional distribution for I_{ij}^T a Dirich-

let prior, resulting in a noisy version of the above logical expression. The RVs O_{ij} are not needed in this logistic model because the parameterization of the X_{ij} conditional distribution induces a type of noisy OR distribution in the posterior. Thus, logistic models can easily handle false negatives.

Because we wanted to highlight the advantages of modeling the experimental noise, we omitted A_{ij} (one-hop) from both the models, BAYESLR and BADERLR. The one-hop signal, gene expression, co-localization, *etc.* can be easily added to any of the models to improve their prediction ability.

3.3. Inference

As is common in probabilistic relational models, the parameters for the conditional distributions of each RV are shared across all of their instances. For example, in the generative model, the prior probability $Pr(X_{ij} = 1)$ is the same for all i and j . With the exception of X_{ij} in BAYESLR, we gave all the distributions a Dirichlet prior. In BAYESLR, the conditional distribution of X_{ij} is the logistic function, and its weights are given Gaussian priors with mean $\mu_X = 0$ and variance $\sigma_X^2 = .01$. Note that by specifying these hyperparameters (e.g. μ_X, σ_X^2), we never need to do learning of the parameters (i.e., weights). Given the relational nature of our data, and the relatively small amount of it, we think that this Bayesian approach is well-suited. We prevent the models from growing too large by only including protein pairs where at least one experiment hinted at an interaction.

We used BUGS⁹ to do inference via Gibbs sampling. We ran 12 MCMC chains for 6000 samples each, from which we computed the desired marginal posterior probabilities. The process is simple enough that someone without much knowledge of machine learning could take our probabilistic models (which we provide in the Supplementary Information) and use them to interpret the results of their 2H experiments. We also tried using loopy belief propagation instead of MCMC to do approximate inference in the generative model of Fig. 3. These results (see Supp. Info.) were very similar, showing that we are likely not being hurt by our choice of approximate inference method. Furthermore, our implementation of the inference algorithm (in Java) takes only seconds to run, and would easily scale to larger problems.

4. Results

We compared the proposed Bayesian logistic model (BAYESLR) with the model based on Bader *et al.*'s work (BADERLR). Both models were trained and tested on the new, updated version of Bader *et al.*'s gold standard data set. We show in Fig. 6 that BAYESLR achieves 5-10% higher accuracy

at most points along the ROC curve. We then checked to see that the improvement was really coming from the noise model, and not just from our use of unlabeled data and MCMC. We tried using a modified BAYESLR model (called Bayesian Bader) which has D_{ij} RVs instead of the noise model, and which uses ITOCORE instead of ITO2H. As expected, it performed the same as BADERLR. We also tried modifying this model to use ITO2H, and found that the resulting performance was much worse.

Investigating this further, we found that the average maximum a posteriori (MAP) weights for BAYESLR were $\{w_U = -2.32, w_L = -10.85, w_I = -4.26, \text{ and } w_{offset} = 7.34\}$. The weight corresponding to ITO2H is almost double the weight for UETZ2H. Interestingly, this is a similar ratio of weights as would be learned had we only used the ITOCORE data set, as in BADERLR. In the last of the above-mentioned experiments, the MAP weight for ITO2H was far smaller than the weight for UETZ2H, which indicates that UETZ2H was a stronger signal than ITO2H. Overall, these experiments demonstrate that we can get significantly better performance using data with many false positives (ITO2H) and a statistical model of the noise than by using pre-filtered data (ITOCORE) and no noise model.

In all regimes of the ROC curve, BAYESLR performs at least as well as BADERLR; in some, it performs significantly better (Fig. 8). The examples that follow demonstrate the weaknesses inherent in BADERLR and show how the proposed model BAYESLR solves these problems. When IRLS learns the weight for the degree variable (in BADERLR), it must trade off having too high a weight, which would cause other features to be ignored, and having too low a weight, which would insufficiently penalize the false positives caused by self-activation/promiscuity. In BADERLR, a high degree D_{ij} penalizes positive predictors from all the experiments (U_{ij}, I_{ij}, L_{ij}) . However, the degree of a protein in a particular experiment (say, Ito *et al.*'s) only gives information about self-activation/promiscuity of the protein in that experiment. Thus, if a protein has a high degree in one experiment, even if that experiment did not predict an interaction (involving some other protein), the degree will negatively affect any predictions made by other experiments on that protein. Our proposed models solve this problem by giving every experiment a different noise model, and by having each noise model be conditionally independent given the X_{ij} variables. Thus, we get the desired property that noise in one experiment should not affect the influence of other experiments on the X_{ij} variables.

Fig. 8(a) illustrates this by showing the prediction accuracy for the test points where $D_{ij} > 4$ and $U_{ij} = 1$ or $L_{ij} = 1$ (called the ‘medium’ degree

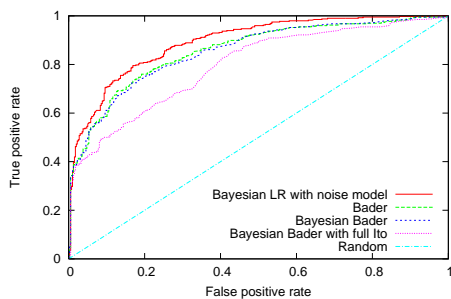


Figure 6: Comparison of logistic models.

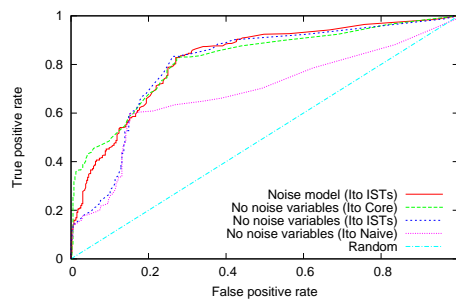
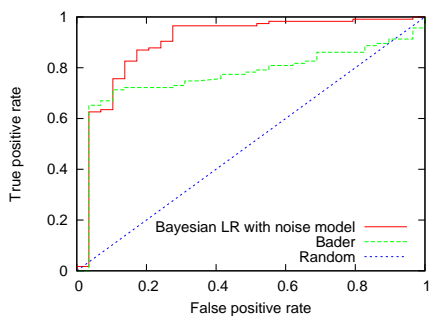
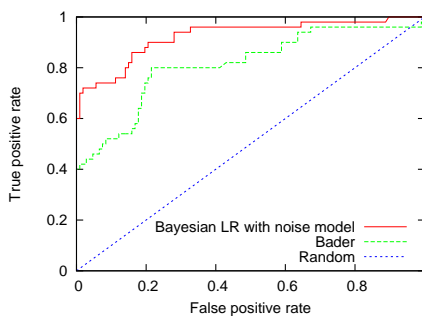


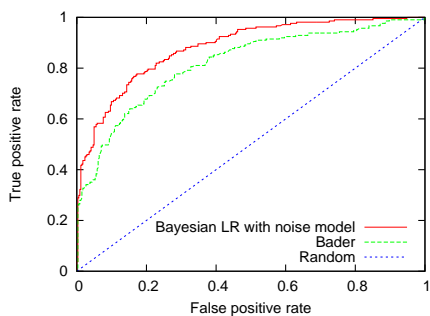
Figure 7: Comparison of generative models.



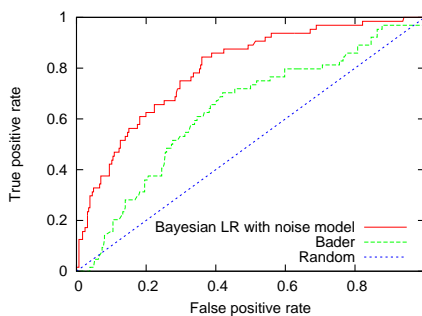
(a) Medium degree and UETZ2H = 1 or Lit=1 (29 neg and 115 pos)



(b) High degree (107 neg and 50 pos)



(c) No ItoCORE (342 neg and 211 pos)



(d) No signal (286 neg and 64 pos)

Figure 8: Examples of regimes where the noise model is particularly helpful. In parentheses we give the number of test cases that fall into each category.

range). When the degree of a protein is very high, BADERLR will always classify interactions involving it as false positives. Fig. 8(b) shows the setting of $D_{ij} > 6^b$. With a false positive rate of less than 1%, BADERLR detects ^bRecall that D_{ij} is on a log-scale, and is the sum for both proteins.

42% of the true interactions, while BAYESLR detects 74% of the true interactions, a 76% improvement. Bader *et al.* found that they got better performance by using only a subset (where $IST \geq 3$) of the interactions in ITO2H. Our noise model allows us to make use of all of the predicted interactions, without hurting our overall results. As a result, our predictions for the proteins pairs where Bader *et al.*'s model ignored ITO2H's interactions (i.e. $IST < 3$) are highly more accurate. This is illustrated in Fig. 8(c). Finally, we show in Fig. 8(d) that at the very extreme when neither ITOCORE, nor the low-throughput 2H experiments (Lit), nor UETZ2H showed an interaction, we can still make meaningful predictions, using a combination of the noise model and the observed interactions in Ito *et al.* where $IST < 3$.

We next compared the various generative models, with the results shown in Fig. 7. Naively implementing the generative model of Fig. 2, using an indicator variable for whether the interaction was observed in ITO2H, results in the worst performance. Changing the indicator variable to a discretized IST count significantly improves performance. Using our noise model (i.e. the model from Fig. 3) provides further improvements, especially in the lower left corner, where the previous two had performed poorly. However, if we remove the noise model and instead pre-filter the data as Bader *et al.* did, using an indicator variable for whether $IST \geq 3$ in ITO2H, we can get almost as good performance using the simple generative model of Fig. 2. The noise model still does better in the upper half of the ROC curve, which is arguably where it matters the most. It is also interesting that our noise model is able to recover the accuracy of the hand-filtered $IST \geq 3$ criterion.

We then applied the BAYESLR model to the full data set to identify proteins in the ITO2H data which are likely to be self-activating/promiscuous (see Supplementary Information). As expected, most proteins with high degree in ITO2H (e.g. YPR086W, degree 99) had a high probability of being self-activating/promiscuous. However, three of the proteins with high degree (YER022W, degree 98; YGL127C, degree 68; and YGR218W, degree 34) had very low probabilities. These differences in promiscuity estimates make sense: for example, there were no positive labeled 2H examples involving YPR086W, while there were five involving YER022W. This propagation of information is precisely what we hoped to capture by using our Bayesian framework. When applying this model to new species where no labeled data is available, the inclusion of additional signals (e.g. co-expression) should result in the same effect. (Note that when no labeled data is available, it might be helpful to fix the model parameters to their MAP values from experiments on related species.)

5. Conclusion

In this paper, we have presented a principled approach to modeling the random and systematic sources of error in two-hybrid experiments, and showed how to integrate our noise models into the two most common probabilistic models for integrating PPI data. Comparisons with previous work demonstrate that explicit modeling of the sources of error can improve protein-protein interaction prediction, making better use of experimental data.

Future work could involve discriminative training of the generative models, investigation of systematic sources of noise in other biological experiments such as Co-IP, and applying noise models to the Markov networks of Jaimovich *et al.* and possibly even in a first-order probabilistic model, where more intricate properties of proteins can be described and jointly predicted.

Acknowledgments: The authors thank Chris Bakal, Leslie Kaelbling, Luke Zettlemoyer, Dan Roy, and Tommi Jaakkola for useful comments. D.S. and R.S. were partially supported by a NSF Graduate Research Fellowship and NSF grant ITR (ASE + NIH)-(dms)-0428715, respectively.

References

1. J.S. Bader *et al.* Gaining confidence in high-throughput protein interaction networks. *Nat Biotech*, 22(1):78–85, January 2004.
2. N. Sun *et al.* Bayesian error analysis model for reconstructing transcriptional regulatory networks. *PNAS*, 103(21):7988–7993, 2006.
3. Nevan J. Krogan *et al.* Global landscape of protein complexes in the yeast *saccharomyces cerevisiae*. *Nature*, 2006.
4. Peter Uetz *et al.* A comprehensive analysis of protein-protein interactions in *saccharomyces cerevisiae*. *Nature*, 403(6770):623–627, February 2000.
5. R. Jansen *et al.* A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, 302(5644):449–453, October 2003.
6. Takashi Ito *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *PNAS*, 98(8):4569–4574, 2001.
7. U. Guldener *et al.* Cygd: the comprehensive yeast genome database. *Nucleic Acids Research*, 33:Database issue:D364–8, 2005.
8. Ariel Jaimovich, Gal Elidan, Hanah Margalit, and Nir Friedman. Towards an integrated protein-protein interaction network: A relational Markov network approach. *Journal of Computational Biology*, 13(2):145–164, 2006.
9. D. J. Lunn, A. Thomas, N. G. Best, and D. J. Spiegelhalter. Winbugs – a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing*, 10:321–333, 2000.
10. Pierre-Olivier Vidalain, Mike Boxem, Hui Ge, Siming Li, and Marc Vidal. Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods*, 32:363–370, 2004.
11. Christian von Mering *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:399–403, 2002.