1 Decomposability of NIPD score

Our score for structure learning is based on the pseudo-likelihood of the data given model and requires us to compute the conditional probability distribution of each variable in a condition c. We require that the parameters of this conditional distribution be dependent such that we can pool the data from the different conditions to estimate the parameters. The conditional distribution, $P(X_i|\mathbf{M}_{ci})$ in condition c is defined as a product:

$$P(X_i = x_{id}|\mathbf{M}_{ci} = \mathbf{m_{cid}}) \propto \prod_{\mathbf{E}\in\text{powerset}(\mathcal{C})\ :\ c\in\mathbf{E}} P(X_i = x_{di}|\mathbf{M}^*_{\mathbf{E}i} = \mathbf{m}^*_{\mathbf{E}i}), \tag{1}$$

where d is the data point index and $\mathbf{M}^*_{\mathbf{E}}$ is the Markov blanket (MB) of X_i exclusively in condition set $\mathbf{E}$. The proportionality sign can be eliminated by dividing the product by a normalization constant. Assuming a conditional Gaussian for the form of the conditional probability distribution, this normalization constant can be obtained by writing out the complete RHS of Eq 1, which has the form:

$$\mathcal{N}\left(x_{1id}|\frac{\mu_{1id}\sigma^2_{3i} * \mu_{3id}\sigma^2_{1i}}{\sigma^2_{3i} * \sigma^2_{1i}}; \frac{\sigma_{1i}\sigma_{3i}}{\sqrt{\sigma^2_{1i} * \sigma^2_{3i}}}\right)\mathcal{N}(\mu_{1id}|\mu_{3id}; \sqrt{\sigma^2_{1i} * \sigma^2_{3i}})$$

The first Gaussian gives the normalized conditional probability distribution and the normalization term is $\frac{1}{Z_{1id}} = \mathcal{N}(\mu_{1id}|\mu_{3id}, \sigma^2_{1i} * \sigma^2_{3i}))$, where σ^2_{3i} is the standard deviation from the condition set $\{1, 2\}$, $\mu_{1id} = \mathbf{w}^\mathsf{T}_{1i}\mathbf{m}^*_{1id}$, is the mean of the conditional Gaussian using the d^{th} data point in condition 1.

However, we work directly with the product of conditionals, that is, the un-normalized conditional probability distribution for two reasons: (a) the score improvement can be computed very efficiently, and (b) the second Gaussian acts as a smoothing term over the parameter μ_{1id}. In particular, if we were estimating a new parameter μ_{1id}, the second Gaussian specifies the probability of the new μ_{1id} using a Gaussian centered around the mean computed from the pooled dataset preferring network structures with means μ_{1id} closer to the shared mean μ_{3id}. Our preliminary experiments showed that this score has better performance than if we were to subtract out the normalization term.

2 Per-variable structure comparison

We compared the structure of the networks inferred by a pair of methods (e.g. INDEP vs NIPD) using a per-variable neighborhood comparison. Let $\mathbf{M}_{1i}$ and $\mathbf{M}^{\text{INF}}_{1i}$ denote the neighborhood of X_i in a true and inferred network, respectively. For each variable, X_i, we compare $\mathbf{M}_{1i}$ to $\mathbf{M}^{\text{INF}}_{1i}$ to obtain a *precision*: $P_{1i} = \frac{|\mathbf{M}_{1i}\cap\mathbf{M}^{\text{INF}}_{1i}|}{|\mathbf{M}^{\text{INF}}_{1i}|}$, *recall*: $R_{1i} = \frac{|\mathbf{M}_{1i}\cap\mathbf{M}^{\text{INF}}_{1i}|}{|\mathbf{M}_{1i}|}$, and *F-score* match: $\frac{2P_{1i}R_{1i}}{P_{1i}*R_{1i}}$. We obtain F-scores for q equal-sized partitions of the data, where $q \in \{3, 4, 5, 6, 7, 8, 9, 10\}$, and networks learned for each partition. The training data size decreased with increasing q. We then obtain the number of variables on which one method has a significantly higher F-score than another, as a function of training data size.

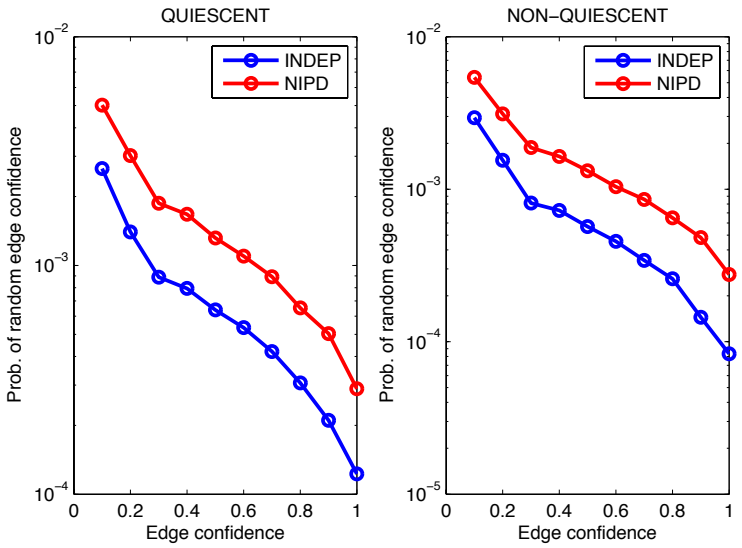

Figure 1: Probability of observing an edge with a particular confidence τ as a function of confidence. The left graph is for quiescent population and the right is for non-quiescent. Probability is estimated by the number of edges that have a confidence of $\geq \tau$ divided by the total number of possible edges subject to the constraint that non-deletion genes can have no more than 8 neighbors.

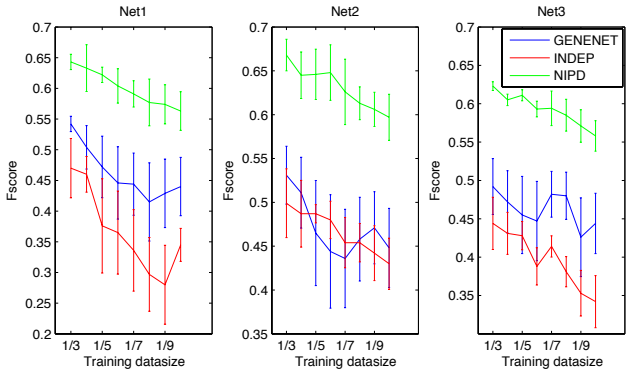

Figure 2: F-score of INDEP, NIPD and GeneNet as a function of decreasing training data for three networks, one per condition. Higher is better.

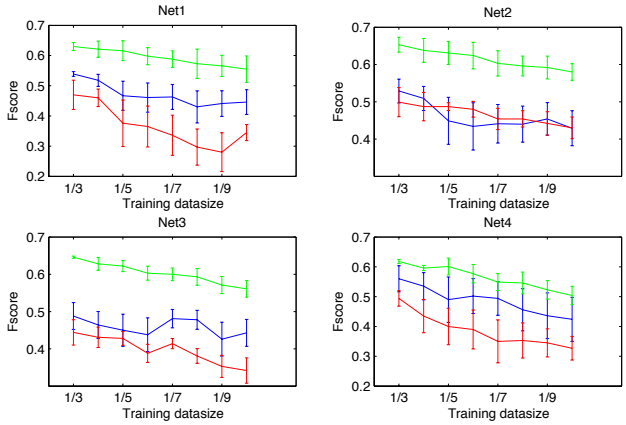

Figure 3: F-score of INDEP, NIPD and GeneNet as a function of decreasing training data for four networks. Higher is better.

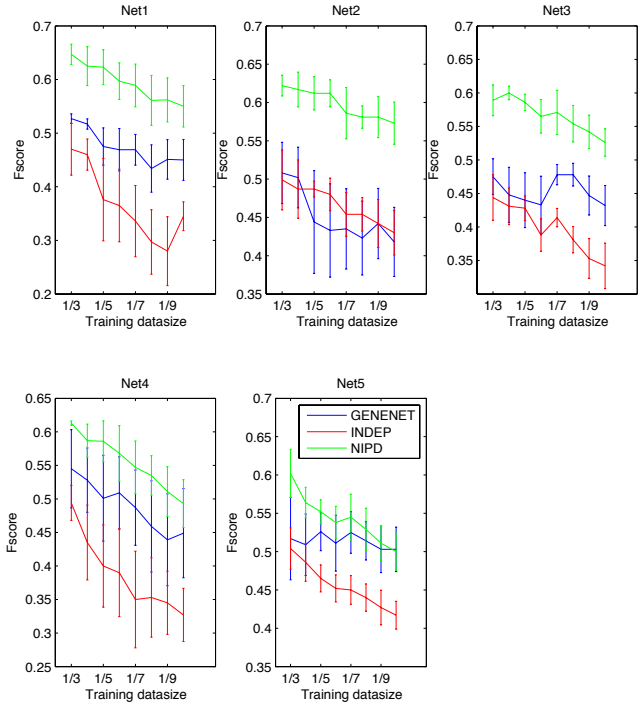

Figure 4: F-score of INDEP, NIPD and GeneNet as a function of decreasing training data for five networks. Higher is better.

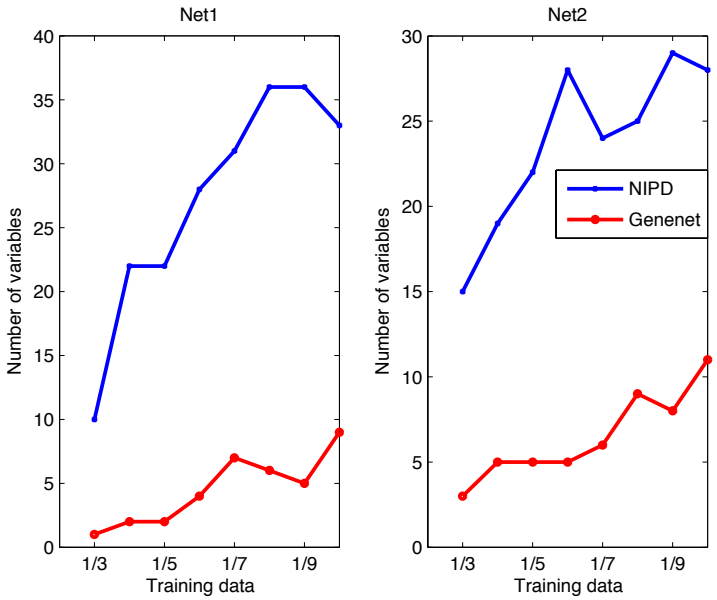

Figure 5: Number of variables on which one algorithm is significantly better than another using datasets for two networks from the HIGHSIM dataset. Comparison for NIPD and GENENET are shown. An algorithm is assessed on the quality of the Markov blanket it identifies for each gene. Specifically, for each fold, we infer a network from which we obtain a collection of Markov blanket match scores, one for each fold and ask if the scores of a variable are significantly better than the scores from another algorithm. The y-axis shows the number of variables with a higher score in one algorithm versus the other.

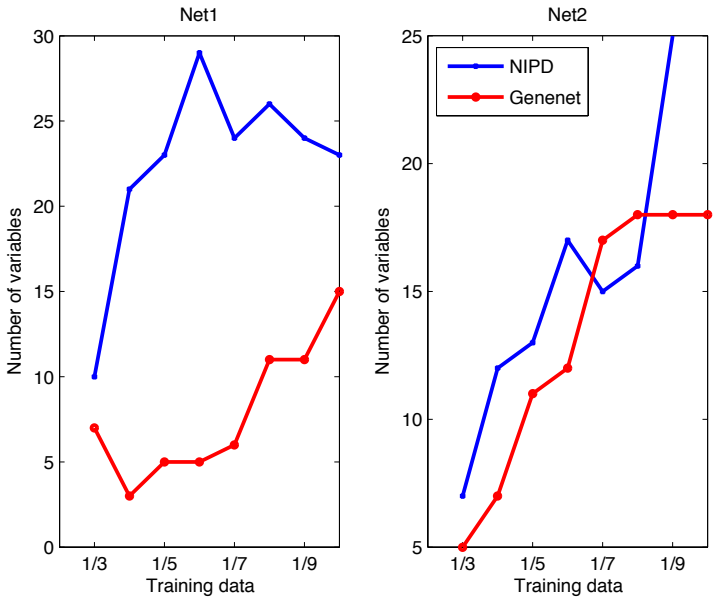

Figure 6: Number of variables on which one algorithm is significantly better than another using datasets for two networks from the LOWSIM dataset. Comparison for NIPD and GENENET are shown. Legend same as Fig 5

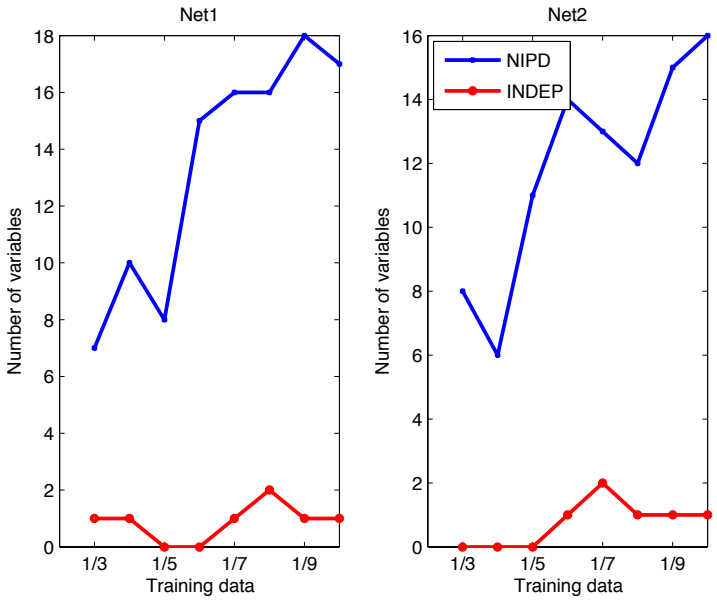

Figure 7: Number of variables on which one algorithm is significantly better than another using datasets for two networks from the HIGHSIM dataset. Comparison for NIPD and INDEP are shown. Legend same as Fig 5

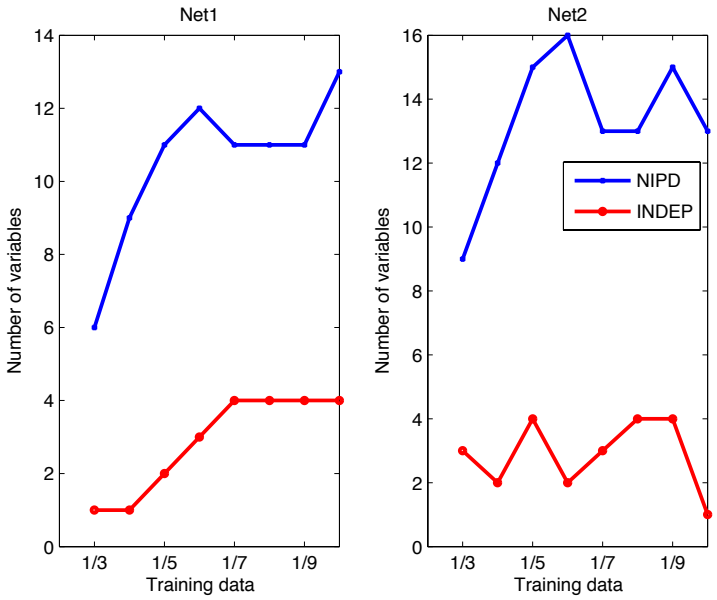

Figure 8: Number of variables on which one algorithm is significantly better than another using datasets for two networks from the LOWSIM dataset. Comparison for NIPD and INDEP are shown. Legend same as Fig 5

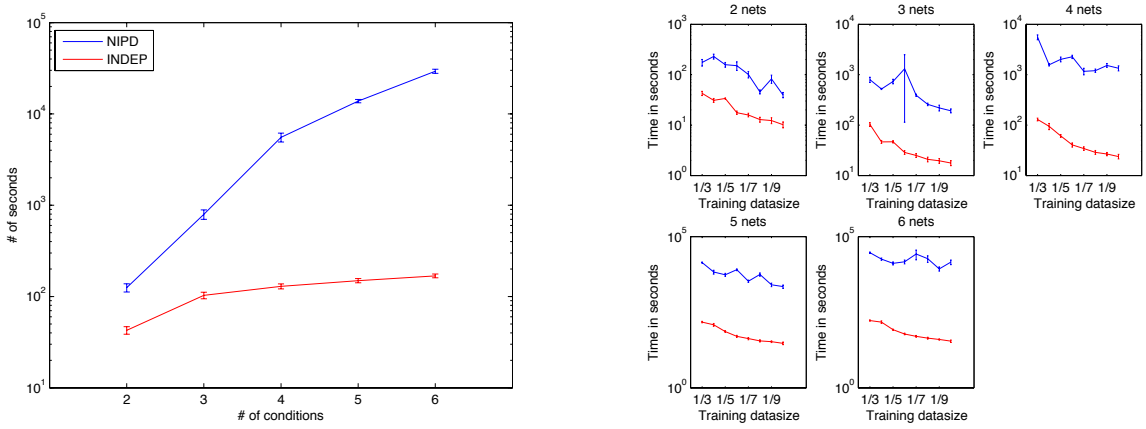

Figure 9: Left: Runtime of NIPD and INDEP as a function of the number of conditions where the dataset per condition was split into three parts and the average runtime per part was measured. Right: Each subplot shows the run time of NIPD and INDEP for different number of conditions (2-6) as a function of the size of training dataset size. The training dataset size decreases as we go to the right.

Figure 10: GO Processes (rows) affected by individual deletion mutants (columns). Colors indicate shared (yellow) affects, quiescent-specific affect (red) and non-quiescent-specific affect (green).

Population	Δ mutant pair	Overlap targets
Q	ATP1*, SDH4*	SDH1,HST2
	ATP3*, COX6*	COX4,COX12
	ETR1*, PST2*	ARA1,PDA1,AYR1
	RIP1*, KGD1*	ATP16,COX5A
	SDH4*, ATP2*	MIR1,NDI1
	AGX1, ATP2*	MIR1,NDI1
	CTA1*, ADH2	ATO3,POX1
	ETR1*, OM14	ARA1,FMP37
	ETR1*, POT1	FAA2,PDA1
	FMP45*, YDR262W	COQ4,ADR1
	QCR7*, ATP18	ILV1,ECM32
	QCR7*, QCR8	YEL076C-A,QCR6
	QCR8, COX7*	QCR6,COX13
	SDH4*, AGX1	MIR1,NDI1
	CTT1, FBP1	DLD1,ACP1
	HXK1, HSP104	HSP42,STI1
	RTN2, ALD3	SOL4,YJL016W
	YDL218W, HSP12	TKL2,GND2
NQ	ATP3*, COX6*	COX4,COX12
	ETR1*, PST2*	ARA1,PDA1,YFR044C,SBP1
	RIP1*, KGD1*	YLL029W,COX5A
	ADY2*, ADH2	ATO3,HMG1
	AGX1, ATP2*	MIR1,NDI1
	COX6*, YLR312C	PSP1,GPX1
	CTA1*, POT1	FOX2,CAT2
	HBT1*, YDR262W	COQ4,SSD1
	HXT5, KGD1*	SYG1,KNS1
	OM45*, ALD4	MDH3,GDH2
	QCR7*, QCR10	YEL076C-A,ILV1
	RTN2, GTT1*	YFL043C,BCY1
	ADH2, ALD4	ATO3,POX1,CAT2,LSC1
	COR1, YDR262W	ADR1,YNL190W
	HSP30, SPI1	SPT2,CPR6
	HXK1, HSP104	HSP42,STI1
	NGR1, SPS100	ALT2,AMS1
	OM14, FBP1	YGR079W,PUT4
	OM14, XBP1	YGR079W,PUT4
	QCR10, QCR8	YEL076C-A,COX13,QCR9
	XBP1, FBP1	YGR079W,PUT4

Table 1: NIPD predictions for double deletions (Δ mutant pair) and their common targets for quiescent and non-quiescent cells *Single gene deletions shown to have a phenotype using viability assays in stationary phase, or in quiescent and non-quiescent cells in Martinez *et al.* 2004, and Aragon *et al.* 2008

.