Supplementary Material

Learning signalling networks from combinatorial perturbations
by exploiting siRNA off-target effects

Jerzy Tiuryn\textsuperscript{1}, Ewa Szczurek\textsuperscript{1,}\textsuperscript{*}

\textsuperscript{1} Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland,
\textsuperscript{*} E-mail: szczurek@mimuw.edu.pl

Proofs

Proof of Proposition 1

Follows immediately from the inclusion-exclusion principle. The proof is by induction on $k$. Below we show an application of this principle to $k = 2$ and $k = 3$. We omit the details of the proof.

$$
\sum_{x \in V^G_1 \cup V^G_b} v_x = \sum_{x \in V^G_a} v_x + \sum_{x \in V^G_b} v_x - \sum_{x \in V^G_a \cap V^G_b} v_x,
$$

and

$$
\sum_{x \in V^G_1 \cup V^G_b \cup V^G_c} v_x = \sum_{x \in V^G_a} v_x + \sum_{x \in V^G_b} v_x + \sum_{x \in V^G_c} v_x - \sum_{x \in V^G_a \cap V^G_b} v_x - \sum_{x \in V^G_a \cap V^G_c} v_x + \sum_{x \in V^G_b \cap V^G_c} v_x.
$$

Similar equations hold for DAG $G_2$.

Proof of Theorem 1

We prove this theorem by induction on $n \geq 1$. Clearly if $n = 1$, then $G_1 = G_2$ and the statement holds vacuously. Assume now that $n > 1$ and consider several cases.

Case I (Joint leaf)

Assume there is an element $a \in V$ which is a leaf in both $G_1$ and $G_2$. If there is $b \in V$ such that $a \notin V^G_b - V^G_2$, then $V^G_a \cap V^G_2 = \{a\}$, but $V^G_2 \cap V^G_a = \emptyset$. Hence, by Proposition 1, $v_a = 0$ and $v, u$ is not admissible. Therefore, we can assume that for $b \in V$,

$$
a \leftarrow_{G_1} b \iff a \leftarrow_{G_2} b.
$$

It follows that if $G'_1$ is obtained from $G_1$ by removing $a$, and $G'_2$ is obtained from $G_2$ by removing $a$, then $G'_1 \neq G'_2$ and by induction assumption the system $\Sigma^3_{(G'_1, G'_2)}$ has no admissible solution. It also follows that if $v', u'$ is any solution for $\Sigma^3_{(G'_1, G'_2)}$, then the extension $v, u$ of $v', u'$, such that $v_a = u_a = 1$ is a solution of $\Sigma^3_{(G_1, G_2)}$. Therefore, the latter system has no admissible solution.
Case II (Connected pair)
Two nodes $a, b \in V$ are said to form a connected pair in $G_1$ and $G_2$ if either $a, b$ are leaves in $G_1$ and $V^G_a \cap V^G_b \neq \emptyset$, or $a, b$ are leaves in $G_2$ and $V^G_a \cap V^G_b \neq \emptyset$.

Assume that there is a connected pair. Let $a, b$ be leaves in $G_1$ and let $c \in V^G_a \cap V^G_b$. Without loss of generality we may assume that $c$ is a leaf in $G_2$. Then

$$V^G_a \cap V^G_b \cap V^G_c = \emptyset \quad \text{and} \quad V^G_a \cap V^G_b \cap V^G_c = \{c\}.$$ 

Thus, by Proposition 1 $u_c = 0$ and therefore the solution $v, u$ is not admissible.

Case III (Alternating cycles)
We need one more definition for the rest of the proof. An alternating cycle in $(G_1, G_2)$ is a sequence of nodes $a_1, a_2, \ldots, a_{2m}$, for some $m \leq \lfloor n/2 \rfloor$, such that

(i) $a_i \neq a_j$, for $i \neq j$.

(ii) $a_1, a_3, a_5, \ldots$ are leaves in $G_1$.

(iii) $a_2, a_4, a_6, \ldots$ are leaves in $G_2$.

(iv) $a_1 \rightarrow_{G_2} a_2 \rightarrow_{G_1} a_3 \rightarrow_{G_2} a_4 \ldots \rightarrow_{G_2} a_{2m} \rightarrow_{G_1} a_1$.

The length of this alternating sequence is $2m$.

From now on we assume that the DAGs $G_1$ and $G_2$ have no joint leaf and they have no connected pair. Call this property of DAGs ($\alpha$). Before proceeding further we need the following observation

Under the above assumption ($\alpha$) there is an alternating cycle in $(G_1, G_2)$. (1)

For the proof of (1) take any leaf $a_1$ in $G_1$. Since it is not a leaf in $G_2$, there exists a leaf $a_2$ in $G_2$ such that $a_1 \rightarrow_{G_2} a_2$. Since $a_2$ is not a leaf in $G_1$ it follows that there exists a leaf $a_3$ in $G_1$ such that $a_2 \rightarrow_{G_1} a_3$. We can continue and obtain a sequence of elements $a_1 \rightarrow_{G_2} a_2 \rightarrow_{G_1} a_3 \rightarrow_{G_2} a_4 \ldots$ until we get an element $b$ that already occurs earlier in the sequence. Consider two cases:

(A) $b = a_{2i} = a_{2m}$, for some $m > i \geq 1$.

(B) $b = a_{2i+1} = a_{2m+1}$, for some $m > i \geq 0$.

In case (A) we have

$$a_{2i-1} \rightarrow_{G_2} a_{2i} \quad \text{and} \quad a_{2m-1} \rightarrow_{G_2} a_{2m}.$$ 

Hence $a_{2i-1}, a_{2m-1}$ is a connected pair of minimal elements in $G_1$. This proves that (A) is impossible.

In case (B), if $i = 0$ then we have an alternating cycle and we are done. Otherwise $i > 0$ and

$$a_{2i} \rightarrow_{G_1} a_{2i+1} \quad \text{and} \quad a_{2m} \rightarrow_{G_1} a_{2m+1}.$$ 

Thus $a_{2i}, a_{2m}$ forms a connected pair of leaves in $G_2$, which is again impossible. This completes the proof of (1).

Now we return to the main line of the proof and consider the next case.
Case IIIa (Long alternating cycle)
We assume that $G_1, G_2$ satisfy property ($\alpha$) and there exists an alternating cycle of length larger than 2. We have

$$a_1 \rightarrow_{G_2} a_2 \rightarrow_{G_1} a_3 \rightarrow_{G_2} \cdots a_{2m}.$$ 

If $a_1 \leftarrow_{G_1} a_2$, then $a_2, a_{2m}$ is a connected pair, which is impossible. Hence

$$V_{a_1}^{G_1} \cap V_{a_2}^{G_1} = \emptyset.$$ 

On the other hand we have

$$V_{a_1}^{G_2} \cap V_{a_2}^{G_2} = \{a_2\},$$

and by Proposition 1 $u_{a_2} = 0$, thereby proving that $v, u$ is not admissible.

It remains to consider the last case, that is

Case IIIb (Only short alternating cycles) 
We assume that $G_1, G_2$ satisfy property ($\alpha$) and all alternating cycles have length 2. By (1) we know that there must be at least one such cycle. Let $a \rightarrow_{G_1} b \rightarrow_{G_2} a$ be an alternating cycle of length 2, where $b$ is a leaf in $G_1$ and $a$ is a leaf in $G_2$. We claim that

$b$ is the only leaf in $V_{a}^{G_1}$. 

(2)

Suppose $c \in V_{a}^{G_1}$ is another leaf. Since $G_1, G_2$ have no common leaves, it follows that there exists a leaf $d$ in $G_2$ such that $c \rightarrow_{G_2} d$. If $d = a$, then $b, c$ is a connected pair of leaves in $G_1$ (since $a \in V_{b}^{G_2} \cap V_{c}^{G_2}$). The obtained contradiction shows that $a \neq d$ and therefore we have

$$a \rightarrow_{G_1} c \rightarrow_{G_2} d,$$

where $a, c, d$ are all pairwise different. Hence $G_1, G_2$ must contain an alternating cycle of length larger than 2. The obtained contradiction completes the proof of (2).

We also claim that

$$\text{for all } c \in V_{a}^{G_1}, \quad \sum_{x \in V_{c}^{G_1}} v_x = u_a.$$ 

(3)

Indeed, by Proposition 1 we have

$$\sum_{x \in V_{a}^{G_1} \cap V_{c}^{G_1}} v_x = \sum_{x \in V_{a}^{G_2} \cap V_{c}^{G_2}} u_x.$$ 

Since $V_{a}^{G_1} \cap V_{c}^{G_1} = V_{c}^{G_1}$ and $V_{a}^{G_2} \cap V_{c}^{G_2} = \{a\}$, property (3) follows.

Now let $c \in V_{a}^{G_1} - \{b\}$ be a leaf. Then, by (2), we have $V_{c}^{G_1} = \{c, b\}$, and by (3) we obtain

$$v_b = u_a = v_c + v_b.$$ 

Thus $v_c = 0$ and therefore $v, u$ is not admissible. This completes the proof of the Theorem.
Proof of Theorem 2

It follows from the proof of Theorem 1 that the only case when we have to use a perturbation experiment targeting 3 genes is when at least one of the DAGs contains a connected pair. Assume that \( G_1 \) contains a pair of leaves \( a, b \) and \( V_a^{G_2} \cap V_b^{G_2} \neq \emptyset \). We start with the following obvious observation.

\[
\sum_{x \in V_a^{G_2} \cap V_b^{G_2}} u_x = 0. \tag{4}
\]

This follows from Proposition 1 and the equality \( V_a^{G_1} \cap V_b^{G_1} = \emptyset \).

It follows from (4) that if \( a \in V_b^{G_2} \), then \( \sum_{x \in V_a^{G_2}} = 0 \), and thus \( v_a = 0 \). In a similar way we show that if \( b \in V_a^{G_2} \), then \( v_b = 0 \). So from now on we can assume that in \( G_2 \) neither \( a \) is accessible from \( b \), nor \( b \) is accessible from \( a \).

Since the number \( n \) of nodes is less than 6, it follows that \( V_a^{G_2} \cap V_b^{G_2} \) has at most 3 elements. We consider 8 cases of possible locations with respect to each other of the elements of \( V_a^{G_2} \cap V_b^{G_2} \) in \( G_2 \). Those cases are depicted in Figure S1. For example, case (A) represents several possibilities: \( n = 3 \), or \( n = 4 \) with the node that is not present being positioned in \( G_2 \) in any way, except the case when it belongs to \( V_a^{G_2} \cap V_b^{G_2} \) in \( G_2 \), and finally the case \( n = 5 \) with the other two nodes being positioned in any way, except in the above intersection.

Before discussing each case let us make some observations.

If \( y \in V_a^{G_2} \cap V_b^{G_2} \) is a leaf in \( G_2 \), and \( a \not\in V_y^{G_1} \), or \( b \not\in V_y^{G_1} \), then \( u_y = 0 \). \tag{5}

Indeed, if \( a \not\in V_y^{G_1} \), then \( V_a^{G_1} \cap V_y^{G_1} = \emptyset \), and by Proposition 1 \( u_y = 0 \).

If \( y \in V_a^{G_2} \cap V_b^{G_2} \) is a leaf in \( G_2 \), and \( a, b \in V_y^{G_1} \), then \( v_a = v_b = u_y \). \tag{6}

This follows from Proposition 1 and the following equalities: \( V_a^{G_1} \cap V_y^{G_1} = \{a\} \), \( V_a^{G_2} \cap V_y^{G_2} = \{y\} \), and similar equalities for \( b \) and \( y \).

Case (A) follows from (5) and (4). Case (B1) is solved by (5), (6) and (4) (since \( u_c = u_d = v_a = v_b \) and by (4) we have \( u_c + u_d = 0 \), hence \( u_c = u_d = 0 \)). A similar argument for case (C1) shows that \( u_c = u_d = u_e = 0 \).

Case (B2) we handle by observing that by (4) \( u_c + u_d = 0 \). If \( a \in V_c^{G_1} \), then \( V_a^{G_1} \cap V_c^{G_1} = \{a\} \) and therefore \( v_a = u_c + u_d = 0 \). In a similar way we proceed when \( b \in V_c^{G_1} \). If \( a, b \in V_c^{G_1} \), then \( d \not\in V_c^{G_1} \) (since otherwise, by (4), we would obtain \( u_d = 0 \) and therefore \( V_c^{G_1} \cap V_d^{G_1} \) is either \( \emptyset \), or is equal to \( \{c\} \). In the former case we have \( u_d = 0 \), and in the latter case we have \( v_c = u_c + u_d = 0 \).

Now we have to consider the remaining cases (C2)–(C5). By (5) we can assume that for every leaf \( y \) in \( G_2 \) we have \( a, b \in V_y^{G_1} \). Therefore for all these four cases we have the following properties

\[
u_c + u_d + u_e = 0. \tag{7}\]

By applying (6), for cases (C2) and (C5) we have

\[
u_e = v_a = v_b, \tag{8}\]

while for cases (C3) and (C4) we have

\[
u_e = u_d = v_a = v_b. \tag{9}\]
If \( a \in V_{c}^{G_1} \), then by taking the equation that corresponds to intersection \( V_a \) and \( V_c \) we have
\[
v_a = \begin{cases} 
  u_c + u_e, & \text{for cases (C2), (C4), and (C5)} \\
  u_c + u_d + u_e, & \text{for case (C3)}
\end{cases}
\]

Case (C3) is handled by applying (7) to conclude that \( v_a = 0 \), and for cases (C2), (C4) and (C5) we use (8) and (9) to conclude that \( u_c = 0 \).

On the other hand, if \( a \notin V_{c}^{G_1} \), then the equation that corresponds to the above mentioned intersection yields
\[
0 = \begin{cases} 
  u_c + u_e, & \text{for cases (C2), (C4), and (C5)} \\
  u_c + u_d + u_e, & \text{for case (C3)}
\end{cases}
\]

Thus for cases (C2), (C4), and (C5) we obtain from (7) that \( u_d = 0 \).

It remains to consider the case (C3) when \( a \notin V_{c}^{G_1} \). By a similar argument we can assume in that case that \( b \notin V_{c}^{G_1} \). Since \( a, b \in V_{d}^{G_1} \cap V_{e}^{G_1} \), it follows that \( d, e \notin V_{c}^{G_1} \). Hence \( c \) is a leaf in \( G_1 \). Consider the equation for \( c \) and obtain
\[
v_c = u_c + u_d + u_e.
\]

Hence by (7) we obtain \( v_c = 0 \). This completes the proof of the Theorem.

**Proof of Theorem 3**

In order to show that matrix \( S = S(E^{(k,n)}, G) \) has rank \( n \), it suffices to prove that if
\[
Sx = 0,
\]
where \( 0 \in R^m \) is a zero vector, then the only solution of (10) is the trivial solution \( x_1 = \ldots = x_n = 0 \). Let \( \alpha \in R^n \) be any solution of (10). Since \( G \) is a DAG, its vertices, ordered by the accessibility relation, form a finite poset. Let \( V = \{1, \ldots, n\} \) be the vertices of \( G \). We prove by well-founded induction in this poset that for all \( i \in V \), we have
\[
\alpha_i = 0.
\]

In this proof we resort only to the one-element perturbation experiments, and therefore the statement holds for all \( k \geq 1 \). If \( i \) is a leaf, then the equation of the system (10) that corresponds to the perturbation experiment for \( i \) is \( x_i = 0 \). Thus \( \alpha_i = 0 \). Now, take a vertex \( p \) of \( G \) and assume that for all vertices \( i \) in \( V_{p}^{G} - \{p\} \), (11) holds. The equation of the system (10) that corresponds to the perturbation experiment of \( p \) is
\[
\sum_{i \in V_{p}^{G}} x_i = 0.
\]

By induction assumption we obtain \( \alpha_p = 0 \). This completes the proof of (11).
Proof of Theorem 4

Consider the perturbation experiments $E^{(3,n)}$ and take a pair of LEMs $(G_1, u, c_1)$, $(G_2, v, c_2)$ over $n$-element set of vertices. Let $S_1 = S(E^{(3,n)}, G_1)$ and let $S_2 = S(E^{(3,n)}, G_2)$. By Equation (3) from the main text, their likelihood is equal if the measurements expected from the models are equal, i.e., when $S_1u = S_2v$. Indeed, their likelihood depends on this expectation and on the residual variance of the normal noise around it. Note that by definition, 3-distinguishability considered above implies that for any pair of different graphs $G_1$ and $G_2$ with $n$ vertices and any admissible vectors $u$ and $v$ of length $n$, we have $S_1u \neq S_2v$. Moreover, for given data $Y$ and a given graph $G$, Theorem 3 determines that at most one vector $\beta$ can maximize the likelihood. Thus, by Theorems 1, and 3, the LEMs are identifiable from perturbation experiments $E^{(3,n)}$.

Similarly, by Theorems 2, and 3, LEMs are identifiable from experiments $E^{(2,n)}$, if the underlying graphs have less than 6 vertices.

For the proof of the second part of the Theorem we first show that the assumption of transitivity of edges in a LEM is necessary for its identifiability. Assume otherwise, that graphs in LEMs need not be transitive closed. For any pair of different graphs $G_1, G_2$ with the same transitive closure, and for any experiment, the set of nodes perturbed by this experiment via propagation in these two graphs would be the same, i.e., $S_1 = S_2$. Thus, for any $\beta$ and $c$, different models $(G_1, \beta, c)$ and $(G_2, \beta, c)$ would have equal likelihood.

Next we show that the assumption of acyclicity of $G$ is necessary for identifiability of LEMs. Assume otherwise, that graphs in LEMs would be allowed to have cycles. Consider a graph $G$ with a subset of $l$ nodes $L = \{1, \ldots, l\}$ in a cycle. Transitive closure of $G$ results in a clique over the same set of nodes $L$. Take any pair of admissible (real and not zero) vectors $u, v$, such that the sums of the clique entries in these two vectors are the same, i.e., $\sum_{i \in L} u_i = \sum_{i \in L} v_i$, while entries $g$ that correspond to the remaining vertices of $G$ are equal. Note there are infinitely many such pairs of contribution vectors. Moreover, for any such a pair we would have $S(E^{(3,n)}, G)u = S(E^{(3,n)}, G)v$. Thus, for any $c > 0$, pairs of different models $(G, u, c)$, $(G, v, c)$ would have equal likelihood.

Motivation of phenotype correction procedure for inferring LEMs from siRNA data

Consider a given graph $G$ over an $n$-element set $V$ of vertices, where $|V| = n$, with $\rightarrow_G$ the accessibility relation. This graph corresponds to the modeled pathway, consisting of $n$ genes in $V$. Let $N$ be a fixed positive integer, and assume that $V \cap \{1, \ldots, N\} = \emptyset$. The set $\{1, \ldots, N\}$ corresponds to the set of all remaining genes, other than genes in $V$, with $N >> n$. Those remaining genes may be perturbed by siRNA as on- or off-targets and have effect on the measured phenotype. Define a new graph $G^#$ over the set $V \cup \{1, \ldots, N\}$, with the accessibility relation being $\rightarrow_G$, i.e. we do not add any edges that involve the new vertices $\{1, \ldots, N\}$. This corresponds to assuming that all remaining genes other than pathway genes are not connected. This is valid in the case when, as in our application, only the analyzed network was stimulated. In this case, although many remaining genes are involved in some networks, their inter-connections are not active when the experiments are performed.

Let $F$ be a given binary $m \times (n + N)$ perturbation matrix. The entries of the matrix that are equal 1 correspond to the on- and the off-targets of the $m$ siRNAs used in the experiments.
Let $S^# = S(F, G^#)$ and

$$Z = S^# \alpha,$$

(12)

where $Z \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}^{n+N}$. Let us split the matrix $F = [E; E^#]$, where $E$ is $m \times n$ matrix consisting of the first $n$ columns of $F$, and $E^#$ is $m \times N$ matrix of the remaining columns of $F$. Assume that the first $n$ columns of $F$ correspond to the $n$ vertices of the graph $G$. The crucial observation is that for every $G$ we have

$$S^# = [S(E, G); E^#],$$

(13)

i.e. the last $N$ columns of $S^#$ are always independent of $G$.

Now, let $\beta \in \mathbb{R}^n$ be the first $n$ coordinates of $\alpha$ and let $\alpha^# \in \mathbb{R}^N$ be the remaining $N$ coordinates of $\alpha$. It follows from (12) and (13) that

$$Z = S(E, G)\beta + E^# \alpha^#.$$

Therefore

$$S(E, G)\beta = Z - E^# \alpha^#.$$

Hence, if $E$ is a discriminating matrix, then dealing with the corrected phenotype vector $Y = Z - E^# \alpha^#$, we can uniquely solve the LEM problem, inferring graph $G$ and the vector of contributions $\beta$.

We make an additional assumption, that most of the contributions to the phenotype from all genes are zero. This sparsity assumption is justified biologically, as we expect that only a small fraction of all genes may affect the phenotype measured downstream of the modeled pathway. Under this assumption, we apply the Lasso (Tibshirani, 1994) to approximate the contributions $\alpha^#$ of the remaining genes to the measured phenotype. To this end, for a given perturbation matrix $F$, we first solve the following regularized regression problem (using R package glmnet)

$$\hat{\alpha} = \text{argmin}_{\alpha \in \mathbb{R}^{n+N}} \left\{ \frac{1}{m} \|Z - F\alpha\|_2^2 + \lambda \|\alpha\|_1 \right\}.$$

Note that $F = S(F, I)$, where $I$ is a $(n+N) \times (n+N)$ identity matrix representing an empty pathway graph (with nodes in $V \cup \{1, \ldots, N\}$ and no edges). Therefore, the solution to this problem identifies the contributions of all genes to the phenotype $Z$ assuming all genes, also the genes in the pathway, are not connected. Given that $N >> n$, we can consider that $F = S(F, I)$ is a good approximation of $S^# = S(F, G^#)$ and hence that $\hat{\alpha}$ is a good approximation of $\alpha$.

Next, we set $\hat{\alpha}^#$ to the sub-vector of $N$ coordinates of $\hat{\alpha}$ corresponding to the $N$ remaining genes. Finally, as the approximated corrected phenotype we obtain $\hat{Y} = Z - E^# \hat{\alpha}^#$. 
Supplementary Figures

Figure S1: Eight possible DAG configurations in $G_2$ that cover the case when the set $X$ of vertices accessible from both $a$ and $b$ (i.e. $X = V_{a}^{G_2} \cap V_{b}^{G_2}$) is non-empty. Here we assume that in $G_2$ neither $a$ is accessible from $b$, nor $b$ is accessible from $a$. Case (A) is the only possibility with exactly one node in $X$. Cases (B1) and (B2) represent the situations when $X$ has exactly two elements. Cases (C1)-(C5) are for the situation when $X$ has three elements. The DAGs are shown without the transitive edges for clarity.
Figure S2: Accurate parameter estimation and pathway structure learning for enhanced LEMs using the BIC approach. A-C Performance for three-node LEMs and exhaustive search in graph space. A Box plots summarizing distribution (showing 25th, 50th and 75th percentiles -horizontal bars, and 1.5 interquartile ranges -vertical line ends) of the correlation between the true $\beta$ values used to simulate the data and the estimated values (y-axis) for increasing number of experimental repeats (x-axis) and for increasing noise (colors). As it is the case for the Bayesian approach, the estimated $\beta$ values are very close to the true ones for almost all simulations, with only a few outliers. Similarly, both sensitivity (B) and specificity (C) of true edge recovery are close to 1 for almost all simulated graphs, and are lowered only for extreme noise values and for few experimental repeats. D-F The same performance analysis as for three-node LEMs in A-C but for ten-node LEMs using greedy search in graph space. For larger graphs and greedy search the performance of parameter estimation decreases only slightly, with median correlation remaining close to 1. Compared to exhaustive search, the sensitivity and specificity of edge recovery in graphs is also lowered and has more more outliers, with sensitivity affected more than specificity. Still, the median values of both sensitivity and specificity are close to 1. Here, the results are also very similar to the results obtained by the Bayesian approach (Fig. 3 in the main text).
Figure S3: Poor parameter estimation and pathway structure learning using previous LEMs on realistic simulated data, where both positive and negative contributions are allowed. A-C Performance of previous LEMs in the task of parameter estimation and learning three-node graphs using exhaustive search. The performance is much poorer compared to the enhanced LEMs for the same data (Figures 3 in the main text and Figure S2) and compared to their excellent results obtained when the data was simulated using only positive contributions (Szczurek and Beerenwinkel, 2016). A Box plots summarizing distribution (showing 25th, 50th and 75th percentiles -horizontal bars, and 1.5 interquartile ranges -vertical line ends) of the correlation between the true $\beta$ values used to simulate the data and the estimated values ($y$-axis) for increasing number of experimental repeats ($x$-axis) and for increasing noise (colors). With median correlation around 0, parameter estimation using previous LEM is dramatically affected by the fact that the data was simulated allowing negative contributions, which is against its assumptions. B, C The performance of structure learning is also decreased, but still, for low noise levels high sensitivity is obtained. Larger number of experimental repeats increases previous LEM model bias resulting in lower sensitivity (B). Median specificity is on the level of 0.7, regardless of noise and number of repeated experiments (C). D-F The same performance analysis as for three-node previous LEMs, but for ten-node previous LEMs, using greedy search in graph space. For larger graphs and greedy search the performance of parameter estimation is similarly poor, although is more concentrated around 0, indicating that greedy search with previous LEMs rarely succeeds to obtain very high correlation values, but also rarely obtains very low correlation values (D). Compared to exhaustive search, the sensitivity of edge recovery is significantly decreased (E), while specificity increases (F), indicating that greedy search with previous LEMs tends to systematically miss true edges rather than add false ones.