Modeling Discrete Interventional Data using Directed Cyclic Graphical Models

Mark Schmidt and Kevin Murphy

Department of Computer Science
University of British Columbia

June 21, 2009

Outline

Motivating Problem: Modeling Biological Networks

Recently, Sachs et al. [2005] analyzed an intracellular multivariate flow cytometry data set that:

- simultaneously measures multiple molecules
- collects a large number of samples
- collects both observational and interventional data.

The difference between conditioning by observation and conditioning by intervention in the 'hungry at work' problem:

- If I see that my watch says 11:55, then it's almost lunch time
- If I set my watch so it says 11:55, it doesn't help

Motivating Problem: Modeling Biological Networks

Recently, Sachs et al. [2005] analyzed an intracellular multivariate flow cytometry data set that:

- simultaneously measures multiple molecules
- collects a large number of samples
- collects both observational and interventional data.

The difference between conditioning by observation and conditioning by intervention in the 'hungry at work' problem:

- If I see that my watch says 11:55, then it's almost lunch time
- If I set my watch so it says 11:55, it doesn't help

Motivating Problem: Modeling Biological Networks

Recently, Sachs et al. [2005] analyzed an intracellular multivariate flow cytometry data set that:

- simultaneously measures multiple molecules
- collects a large number of samples
- collects both observational and interventional data.

The difference between conditioning by observation and conditioning by intervention in the 'hungry at work' problem:

- If I see that my watch says 11:55, then it's almost lunch time
- If I set my watch so it says 11:55, it doesn't help

Motivating Problem: Modeling Biological Networks

Recently, Sachs et al. [2005] analyzed an intracellular multivariate flow cytometry data set that:

- simultaneously measures multiple molecules
- collects a large number of samples
- collects both observational and interventional data.

The difference between conditioning by observation and conditioning by intervention in the 'hungry at work' problem:

- If I see that my watch says 11:55, then it's almost lunch time
- If I set my watch so it says 11:55, it doesn't help

Motivating Problem: Modeling Biological Networks

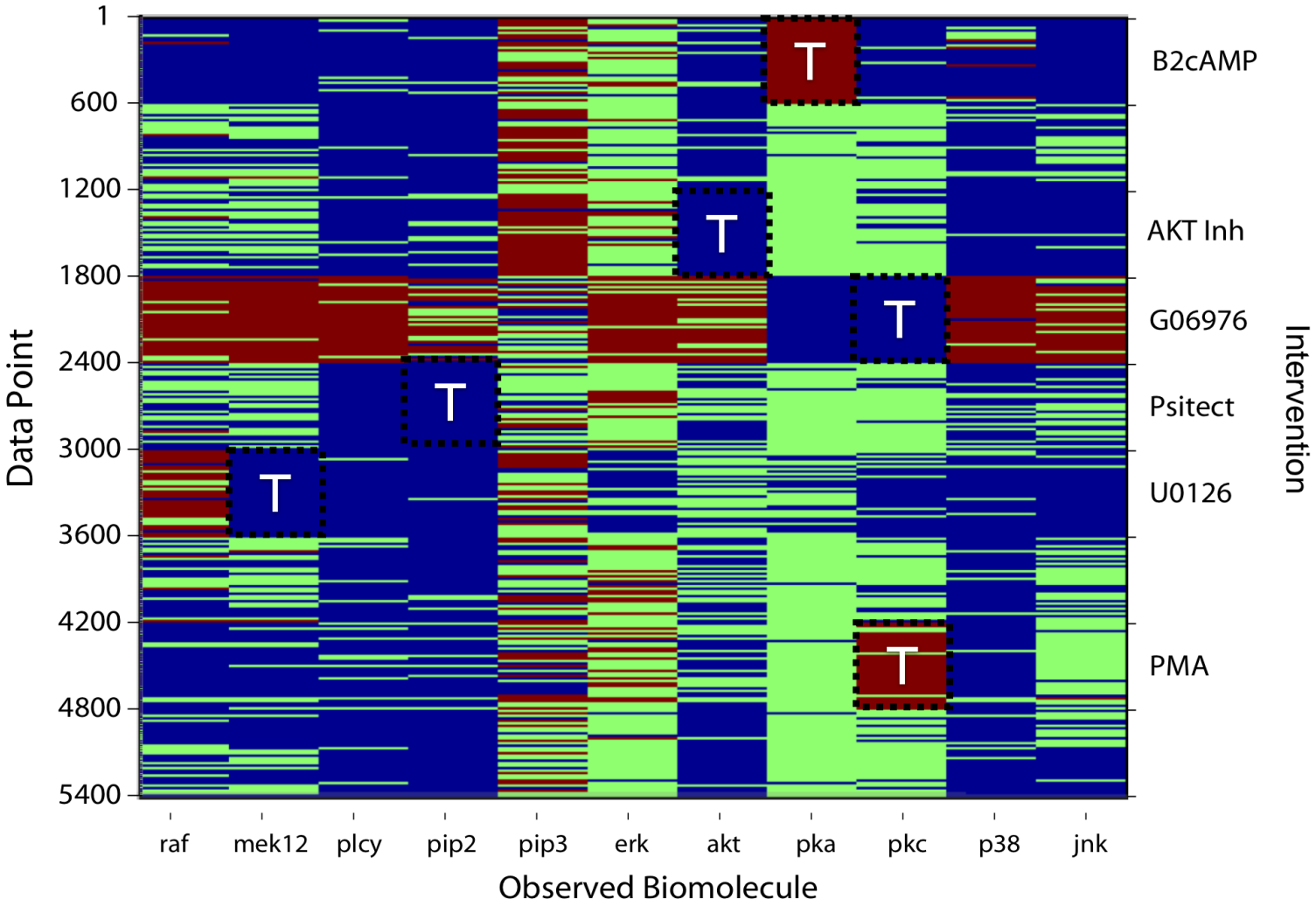
Recently, Sachs et al. [2005] analyzed an intracellular multivariate flow cytometry data set that:

- simultaneously measures multiple molecules
- collects a large number of samples
- collects both observational and interventional data.

The difference between conditioning by observation and conditioning by intervention in the 'hungry at work' problem:

- If I see that my watch says 11:55, then it's almost lunch time
- If I set my watch so it says 11:55, it doesn't help

Motivating Problem: Modeling Biological Networks



Drawbacks of Directed Acyclic and Undirected Models

So what kind of graphical model should we use for this data?

We could use directed acyclic graphical (DAG) models:

- DAGs can model effects of interventions
- But DAGs do not allow the model to have cycles
  (most biological networks contain feedback cycles)

We could use undirected graphical (UG) models:

- UGs allow the model to have cycles
- But UGs do not model effects of interventions
  (there is no difference between 'seeing' and 'doing')

Drawbacks of Directed Acyclic and Undirected Models

So what kind of graphical model should we use for this data?

We could use directed acyclic graphical (DAG) models:

- DAGs can model effects of interventions
- But DAGs do not allow the model to have cycles
  (most biological networks contain feedback cycles)

We could use undirected graphical (UG) models:

- UGs allow the model to have cycles
- But UGs do not model effects of interventions
  (there is no difference between 'seeing' and 'doing')

Drawbacks of Directed Acyclic and Undirected Models

So what kind of graphical model should we use for this data?

We could use directed acyclic graphical (DAG) models:

- DAGs can model effects of interventions
- But DAGs do not allow the model to have cycles
  (most biological networks contain feedback cycles)

We could use undirected graphical (UG) models:

- UGs allow the model to have cycles
- But UGs do not model effects of interventions
  (there is no difference between 'seeing' and 'doing')

Our Contribution: A Cyclic Model for Interventional Data

This talk presents a representation for discrete distributions that:

- can model the effects of interventions
- allows the model to have cycles

We do this by factorizing the distribution in terms of globally normalized interventional potential functions.

The work is closely related to several other branches of research:

- Conditionally-specified distributions
- Structural equations models
- Chain graphs and directed factor graphs

Our Contribution: A Cyclic Model for Interventional Data

This talk presents a representation for discrete distributions that:

- can model the effects of interventions
- allows the model to have cycles

We do this by factorizing the distribution in terms of globally normalized interventional potential functions.

The work is closely related to several other branches of research:

- Conditionally-specified distributions
- Structural equations models
- Chain graphs and directed factor graphs

Our Contribution: A Cyclic Model for Interventional Data

This talk presents a representation for discrete distributions that:

- can model the effects of interventions
- allows the model to have cycles

We do this by factorizing the distribution in terms of globally normalized interventional potential functions.

The work is closely related to several other branches of research:

- Conditionally-specified distributions
- Structural equations models
- Chain graphs and directed factor graphs

Outline

Interventional Potential Representation

We represent the distribution of discrete x_i as a globally normalized product of interventional potential functions

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi(x_i | x_{\pi(i)}),$$

where each $\phi(x_i | x_{\pi(i)})$ assigns a non-negative potential to each joint configuration of x_i and its parents $x_{\pi(i)}$.

Relation to Undirected Graphical Models

Interventional potential representation:

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi(x_i | x_{\pi(i)}),$$

In contrast, undirected graphical models represent the distribution with potential functions defined on cliques,

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi(x_c).$$

With clique potentials, we visualize the structure in the model as an undirected graph.

With interventional potentials, we visualize the structure in the model as a directed graph, with edges from parents to children.

Relation to Undirected Graphical Models

Interventional potential representation:

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi(x_i | x_{\pi(i)}),$$

In contrast, undirected graphical models represent the distribution with potential functions defined on cliques,

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi(x_c).$$

With clique potentials, we visualize the structure in the model as an undirected graph.

With interventional potentials, we visualize the structure in the model as a directed graph, with edges from parents to children.

Relation to Undirected Graphical Models

Interventional potential representation:

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi(x_i | x_{\pi(i)}),$$

In contrast, undirected graphical models represent the distribution with potential functions defined on cliques,

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi(x_c).$$

With clique potentials, we visualize the structure in the model as an undirected graph.

With interventional potentials, we visualize the structure in the model as a directed graph, with edges from parents to children.

Relation to Undirected Graphical Models

Interventional potential representation:

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi(x_i | x_{\pi(i)}),$$

In contrast, undirected graphical models represent the distribution with potential functions defined on cliques,

$$p(x_1, \ldots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi(x_c).$$

With clique potentials, we visualize the structure in the model as an undirected graph.

With interventional potentials, we visualize the structure in the model as a directed graph, with edges from parents to children.

Relation to Directed Acyclic Models

We would obtain DAG models if the graph is acyclic and each node is locally 'normalizable':

$$\exists c_i, \forall_{x_{\pi(i)}} \sum_{x_i} \phi(x_i|x_{\pi(i)}) = c_i$$

In this case, $p(x_i|x_{\pi(i)}) \propto \phi(x_i|x_{\pi(i)})$.

Unlike DAG models:

- We don't enforce local normalization
- $p(x_i|x_{\pi(i)})$ will not generally be proportional to $\phi(x_i|x_{\pi(i)})$
- The graph is allowed to have directed cycles

Relation to Directed Acyclic Models

We would obtain DAG models if the graph is acyclic and each node is locally 'normalizable':

$$\exists c_i, \forall_{x_{\pi(i)}} \sum_{x_i} \phi(x_i|x_{\pi(i)}) = c_i$$

In this case, $p(x_i|x_{\pi(i)}) \propto \phi(x_i|x_{\pi(i)})$.

Unlike DAG models:

- We don't enforce local normalization
- $p(x_i|x_{\pi(i)})$ will not generally be proportional to $\phi(x_i|x_{\pi(i)})$
- The graph is allowed to have directed cycles

Relation to Directed Acyclic Models

We would obtain DAG models if the graph is acyclic and each node is locally 'normalizable':

$$\exists c_i, \forall_{x_{\pi(i)}} \sum_{x_i} \phi(x_i|x_{\pi(i)}) = c_i$$

In this case, $p(x_i|x_{\pi(i)}) \propto \phi(x_i|x_{\pi(i)})$.

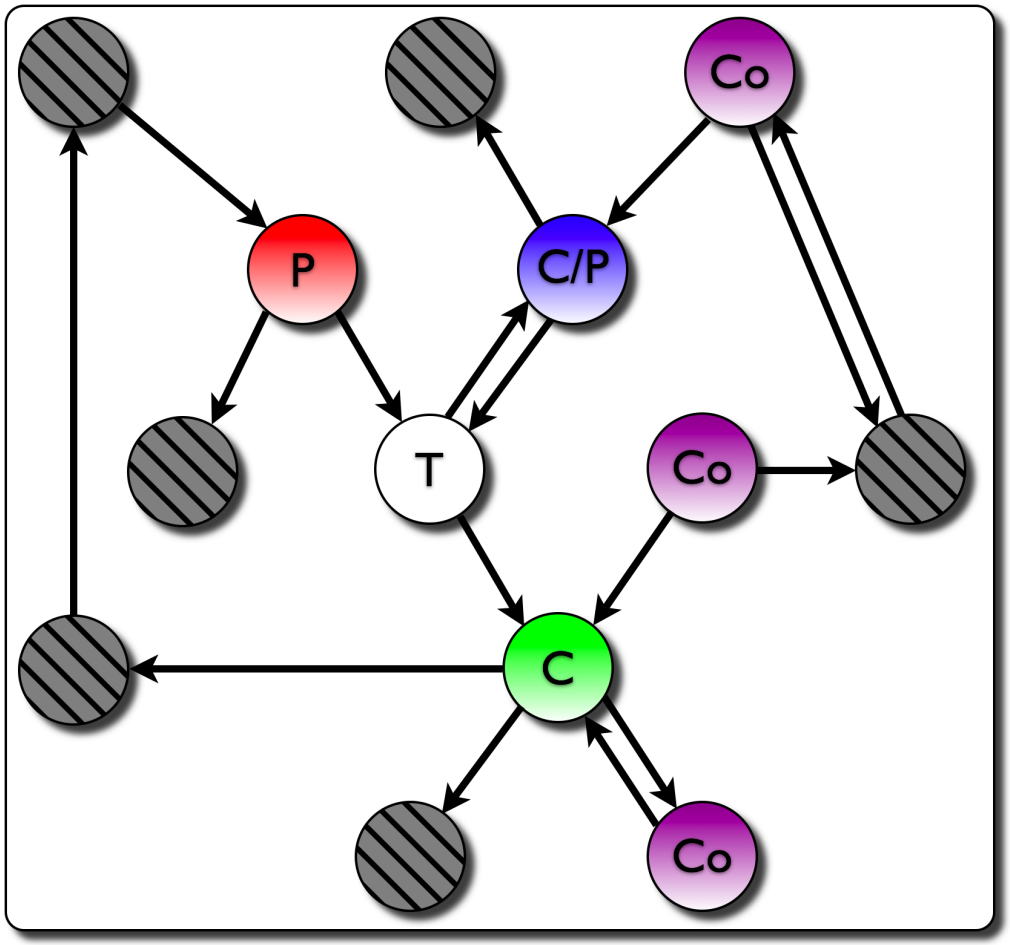
Unlike DAG models:

- We don't enforce local normalization
- $p(x_i|x_{\pi(i)})$ will not generally be proportional to $\phi(x_i|x_{\pi(i)})$
- The graph is allowed to have directed cycles

Outline

Local Markov Properties

The factorization implies that each node is independent of all other nodes given its Markov blanket (parents, children, and co-parents)



Global Markov Properties

We can use graphical operations to answer more general independence queries using moralization and graph separation:

Given a graph structure and the query $P \perp Q|R$:

1. Place an undirected edge between unconnected co-parents
2. Replace all 2-cycles with an undirected edge
3. Remove directions on all edges
4. Test whether R blocks all paths between P and Q

If no path exists, the factorization implies $P \perp Q|R$.

Global Markov Properties

We can use graphical operations to answer more general independence queries using moralization and graph separation:

Given a graph structure and the query $P \perp Q|R$:

1. Place an undirected edge between unconnected co-parents
2. Replace all 2-cycles with an undirected edge
3. Remove directions on all edges
4. Test whether R blocks all paths between P and Q

If no path exists, the factorization implies $P \perp Q|R$.

Global Markov Properties

We can use graphical operations to answer more general independence queries using moralization and graph separation:

Given a graph structure and the query $P \perp Q|R$:

1. Place an undirected edge between unconnected co-parents
2. Replace all 2-cycles with an undirected edge
3. Remove directions on all edges
4. Test whether R blocks all paths between P and Q

If no path exists, the factorization implies $P \perp Q|R$.

Outline

Conditioning by Intervention

Are interventional potentials equivalent to using clique potentials?

Under most data generating processes for undirected graphs there is no difference between conditioning by observation and conditioning by intervention [Lauritzen and Richardson, 2002]

In our representation, we define the effects of interventions by analogy with interventions in DAGs.

Conditioning by Intervention

Are interventional potentials equivalent to using clique potentials?

Under most data generating processes for undirected graphs there is no difference between conditioning by observation and conditioning by intervention [Lauritzen and Richardson, 2002]

In our representation, we define the effects of interventions by analogy with interventions in DAGs.

Conditioning by Intervention

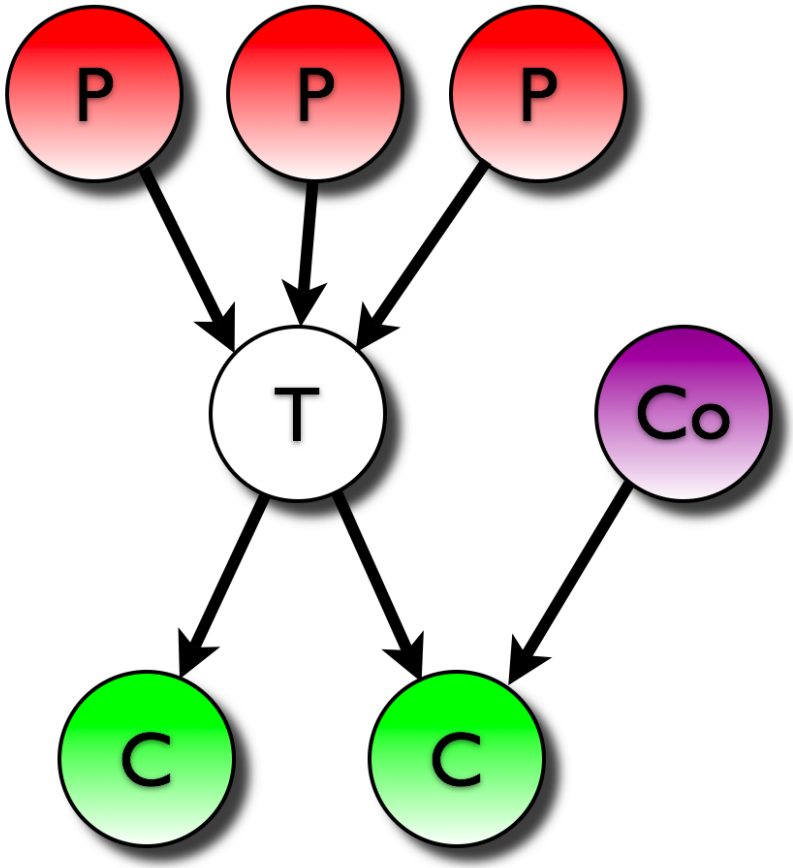
Are interventional potentials equivalent to using clique potentials?

Under most data generating processes for undirected graphs there is no difference between conditioning by observation and conditioning by intervention [Lauritzen and Richardson, 2002]

In our representation, we define the effects of interventions by analogy with interventions in DAGs.

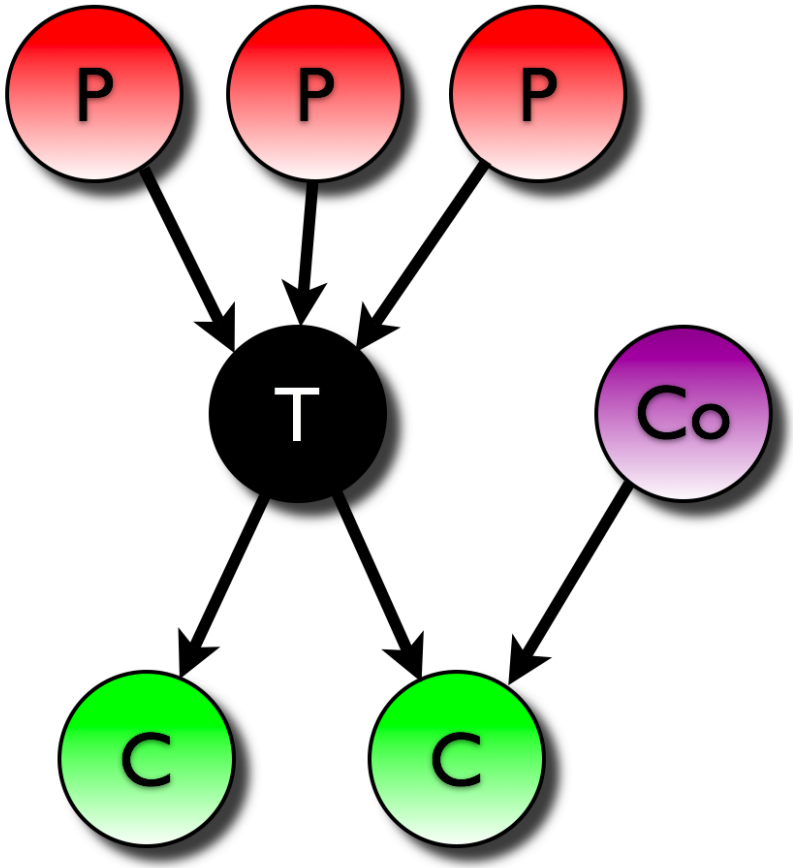
Interventions in Directed Acyclic Graphical Models

Consider the following DAG:



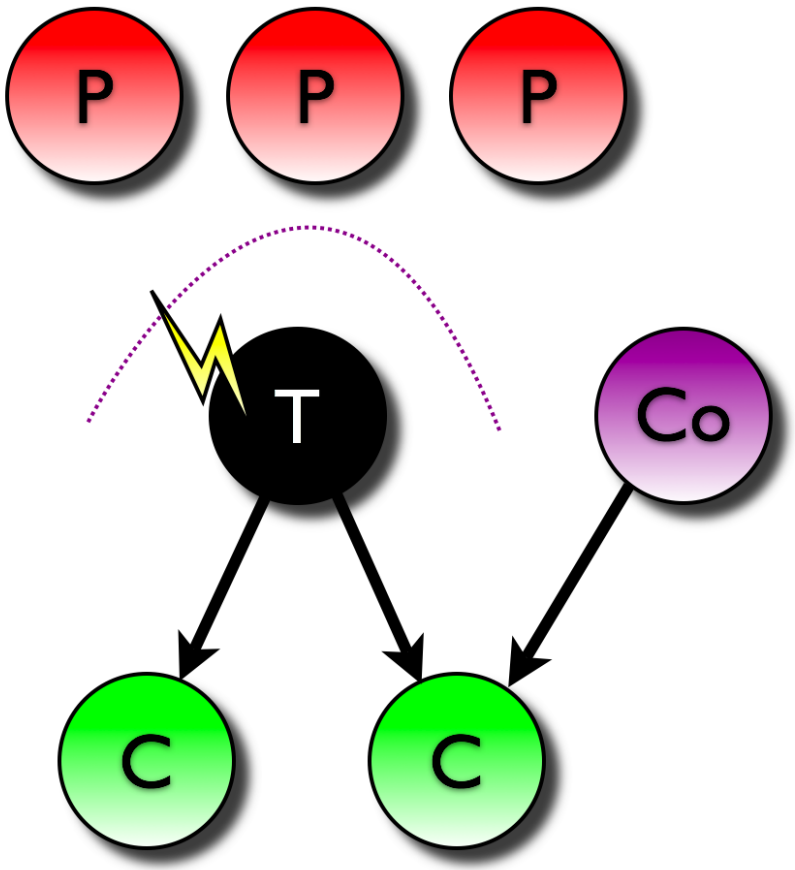
Interventions in Directed Acyclic Graphical Models

Conditioning on T by observation:



Interventions in Directed Acyclic Graphical Models

Conditioning on T by intervention:



Interventions in Directed Cyclic Graphical Models

- In DAG models, interventions remove $p(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.
- In DCG models, interventions remove $\phi(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.

It is possible to interpret the model and these interventions in terms of the equilibrium distribution of a stochastic process.

Interventions in Directed Cyclic Graphical Models

- In DAG models, interventions remove $p(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.
- In DCG models, interventions remove $\phi(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.

It is possible to interpret the model and these interventions in terms of the equilibrium distribution of a stochastic process.

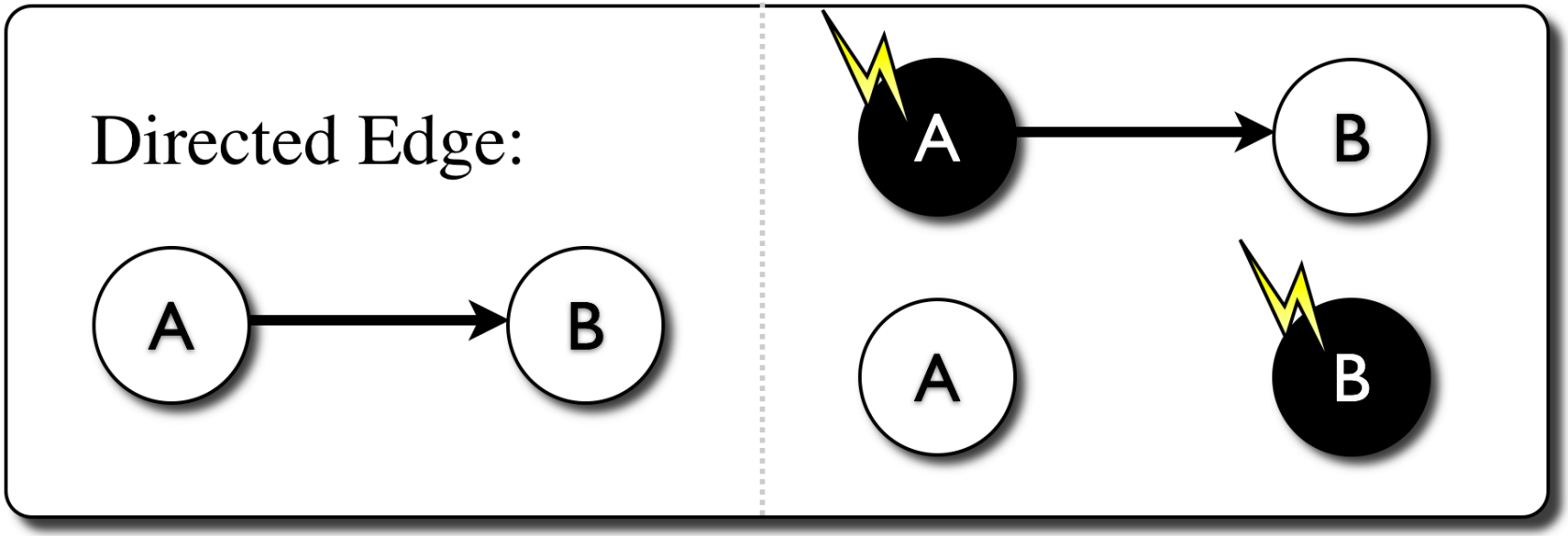
Interventions in Directed Cyclic Graphical Models

- In DAG models, interventions remove $p(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.
- In DCG models, interventions remove $\phi(x_i|x_{\pi(i)})$ from the factorization when we intervene on x_i.

It is possible to interpret the model and these interventions in terms of the equilibrium distribution of a stochastic process.

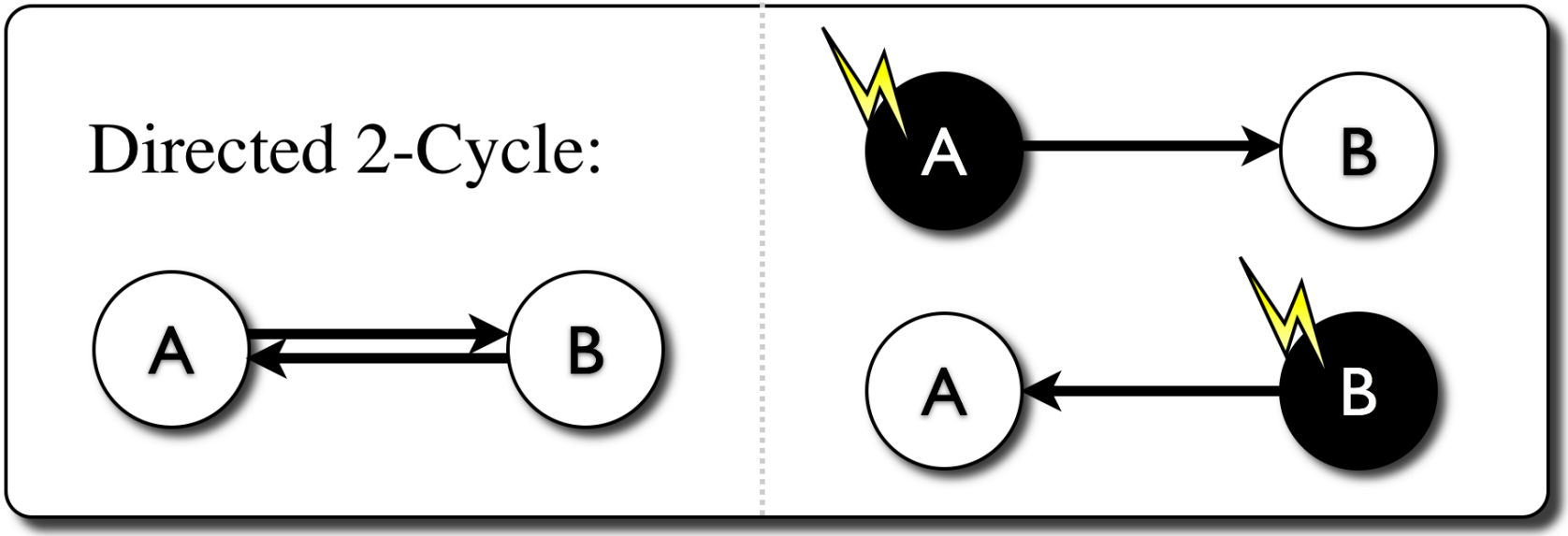
Interventions in Directed Cyclic Graphical Models

Graphical effect of interventions on a single directed edge:



Interventions in Directed Cyclic Graphical Models

Graphical effect of interventions on a directed 2-cycle:



Outline

Exponential Family Parameterization

An appealing parameterization of the graphical model is with interventional potential functions of the form

$$\phi(x_i|x_{\pi(i)},\theta) = \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

Under this parameterization, parameter estimation can be formulated as a convex optimization problem.

Unfortunately, computing the objective/gradient requires inference.

When inference is intractable we could implement a pseudo-likelihood or Gibbs sampler, or consider fancier methods.

Exponential Family Parameterization

An appealing parameterization of the graphical model is with interventional potential functions of the form

$$\phi(x_i|x_{\pi(i)},\theta) = \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

Under this parameterization, parameter estimation can be formulated as a convex optimization problem.

Unfortunately, computing the objective/gradient requires inference.

When inference is intractable we could implement a pseudo-likelihood or Gibbs sampler, or consider fancier methods.

Exponential Family Parameterization

An appealing parameterization of the graphical model is with interventional potential functions of the form

$$\phi(x_i|x_{\pi(i)},\theta) = \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

Under this parameterization, parameter estimation can be formulated as a convex optimization problem.

Unfortunately, computing the objective/gradient requires inference.

When inference is intractable we could implement a pseudo-likelihood or Gibbs sampler, or consider fancier methods.

Exponential Family Parameterization

An appealing parameterization of the graphical model is with interventional potential functions of the form

$$\phi(x_i|x_{\pi(i)},\theta) = \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

Under this parameterization, parameter estimation can be formulated as a convex optimization problem.

Unfortunately, computing the objective/gradient requires inference.

When inference is intractable we could implement a pseudo-likelihood or Gibbs sampler, or consider fancier methods.

Addressing Non-Identifiability with ℓ_2-Regularization

Unfortunately, the model has too many parameters to be uniquely identified:

- We can re-scale potentials without changing the likelihood.
- We can move weight between the bias and edge potentials.
- Identifying the parameters of a 2-cycle requires interventions.

To make the parameters identifiable, we do MAP parameter estimation with a small ℓ_2-regularizer:

$$\min_{\theta} -\log p(X|\theta) + \lambda_2 ||\theta||_2^2,$$

This is strictly convex, leading to a unique parameterization.

Addressing Non-Identifiability with ℓ_2-Regularization

Unfortunately, the model has too many parameters to be uniquely identified:

- We can re-scale potentials without changing the likelihood.
- We can move weight between the bias and edge potentials.
- Identifying the parameters of a 2-cycle requires interventions.

To make the parameters identifiable, we do MAP parameter estimation with a small ℓ_2-regularizer:

$$\min_{\theta} -\log p(X|\theta) + \lambda_2 ||\theta||_2^2,$$

This is strictly convex, leading to a unique parameterization.

Outline

Convex Relaxation of Structure Learning

In many applications we do not know the graph. We could consider putting a penalty on the number of edges

$$\min_{\theta,G} -\log p(X|\theta) + \lambda \mathcal{E}(G),$$

We can relax the cardinality penalty (and avoid searching over graphs) using a group ℓ_1-regularizer on the edge weights

$$\min_{\theta} -\log p(X|\theta) + \lambda \sum_{e} ||w_{.,.,e}||_2,$$

We re-write this with a linear penalty and second-order cone constraints, and solve it with a projected quasi-Newton method.

Convex Relaxation of Structure Learning

In many applications we do not know the graph. We could consider putting a penalty on the number of edges

$$\min_{\theta, G} -\log p(X|\theta) + \lambda \mathcal{E}(G),$$

We can relax the cardinality penalty (and avoid searching over graphs) using a group ℓ_1-regularizer on the edge weights

$$\min_{\theta} -\log p(X|\theta) + \lambda \sum_{e} ||w_{.,.,e}||_2,$$

We re-write this with a linear penalty and second-order cone constraints, and solve it with a projected quasi-Newton method.

Convex Relaxation of Structure Learning

In many applications we do not know the graph. We could consider putting a penalty on the number of edges

$$\min_{\theta,G} -\log p(X|\theta) + \lambda \mathcal{E}(G),$$

We can relax the cardinality penalty (and avoid searching over graphs) using a group ℓ_1-regularizer on the edge weights

$$\min_{\theta} -\log p(X|\theta) + \lambda \sum_{e} ||w_{.,.,e}||_2,$$

We re-write this with a linear penalty and second-order cone constraints, and solve it with a projected quasi-Newton method.

Outline

Summary of Experiments

We performed two sets of experiments:

- Synthetic data: we compared models on data generated from a DCG model (sanity check)
- Real data: we compared models on the intracellular multivariate flow cytometry data [Sachs et al., 2005]

We measure performance using test set negative log-likelihood, to measure generalization ability in an objective way.

Summary of Experiments

We performed two sets of experiments:

- Synthetic data: we compared models on data generated from a DCG model (sanity check)
- Real data: we compared models on the intracellular multivariate flow cytometry data [Sachs et al., 2005]

We measure performance using test set negative log-likelihood, to measure generalization ability in an objective way.

Summary of Experiments

We performed two sets of experiments:

- Synthetic data: we compared models on data generated from a DCG model (sanity check)
- Real data: we compared models on the intracellular multivariate flow cytometry data [Sachs et al., 2005]

We measure performance using test set negative log-likelihood, to measure generalization ability in an objective way.

Summary of Experiments

We performed two sets of experiments:

- Synthetic data: we compared models on data generated from a DCG model (sanity check)
- Real data: we compared models on the intracellular multivariate flow cytometry data [Sachs et al., 2005]

We measure performance using test set negative log-likelihood, to measure generalization ability in an objective way.

Experimental Set-Up

We compared the performance of several graphical models, each using a linear exponential family representation:

- DAG:

$$p(x_i|x_{\pi(i)},\theta) \propto \exp(b_{i,x_i} + \sum_{j\in\pi(i)} w_{x_i,x_j,e}),$$

- UG:

$$p(x_1,\ldots,x_n|\theta) \propto \exp(\sum_{i=1}^{n} b_{i,x_i} + \sum_{e:\{<i,j>\in\mathcal{E}\}} w_{x_i,x_j,e}),$$

- DCG:

$$\phi(x_i|x_{\pi(i)},\theta) \propto \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

In all cases, we use exact inference and group ℓ_1-regularization of the edge parameters to learn a sparse structure.

Experimental Set-Up

We compared the performance of several graphical models, each using a linear exponential family representation:

- DAG:

$$p(x_i|x_{\pi(i)},\theta) \propto \exp(b_{i,x_i} + \sum_{j\in\pi(i)} w_{x_i,x_j,e}),$$

- UG:

$$p(x_1,\ldots,x_n|\theta) \propto \exp(\sum_{i=1}^{n} b_{i,x_i} + \sum_{e:\{<i,j>\in\mathcal{E}\}} w_{x_i,x_j,e}),$$

- DCG:

$$\phi(x_i|x_{\pi(i)},\theta) \propto \exp(b_{i,x_i} + \sum_{e\in\{<i,j>:j\in\pi(i)\}} w_{x_i,x_j,e}),$$

In all cases, we use exact inference and group ℓ_1-regularization of the edge parameters to learn a sparse structure.

Experimental Set-Up

We compared the performance of several different representations:

- DAG: An interventional DAG model, searching over all node orderings to select parents.
- UG-observe: A UG model that ignores interventions. (it maximizes $p(x_1, \ldots, x_n)$ over the training samples)
- UG-condition: A UG model that treats interventions as fixed observations during training. (it maximizes $p(x_1, \ldots, x_n)$ on observational samples, and maximize $p(x_{-k}|x_k)$ when we intervene on node k)
- DCG: the proposed model, modeling interventions by removing the appropriate interventional potential.

Experimental Set-Up

We compared the performance of several different representations:

- DAG: An interventional DAG model, searching over all node orderings to select parents.
- UG-observe: A UG model that ignores interventions. (it maximizes $p(x_1, \ldots, x_n)$ over the training samples)
- UG-condition: A UG model that treats interventions as fixed observations during training. (it maximizes $p(x_1, \ldots, x_n)$ on observational samples, and maximize $p(x_{-k}|x_k)$ when we intervene on node k)
- DCG: the proposed model, modeling interventions by removing the appropriate interventional potential.

Experimental Set-Up

We compared the performance of several different representations:

- DAG: An interventional DAG model, searching over all node orderings to select parents.
- UG-observe: A UG model that ignores interventions. (it maximizes $p(x_1, \ldots, x_n)$ over the training samples)
- UG-condition: A UG model that treats interventions as fixed observations during training. (it maximizes $p(x_1, \ldots, x_n)$ on observational samples, and maximize $p(x_{-k}|x_k)$ when we intervene on node k)
- DCG: the proposed model, modeling interventions by removing the appropriate interventional potential.

Experimental Set-Up

We compared the performance of several different representations:

- DAG: An interventional DAG model, searching over all node orderings to select parents.
- UG-observe: A UG model that ignores interventions. (it maximizes $p(x_1, \ldots, x_n)$ over the training samples)
- UG-condition: A UG model that treats interventions as fixed observations during training. (it maximizes $p(x_1, \ldots, x_n)$ on observational samples, and maximize $p(x_{-k}|x_k)$ when we intervene on node k)
- DCG: the proposed model, modeling interventions by removing the appropriate interventional potential.

Experimental Set-Up

We compared the performance of several different representations:

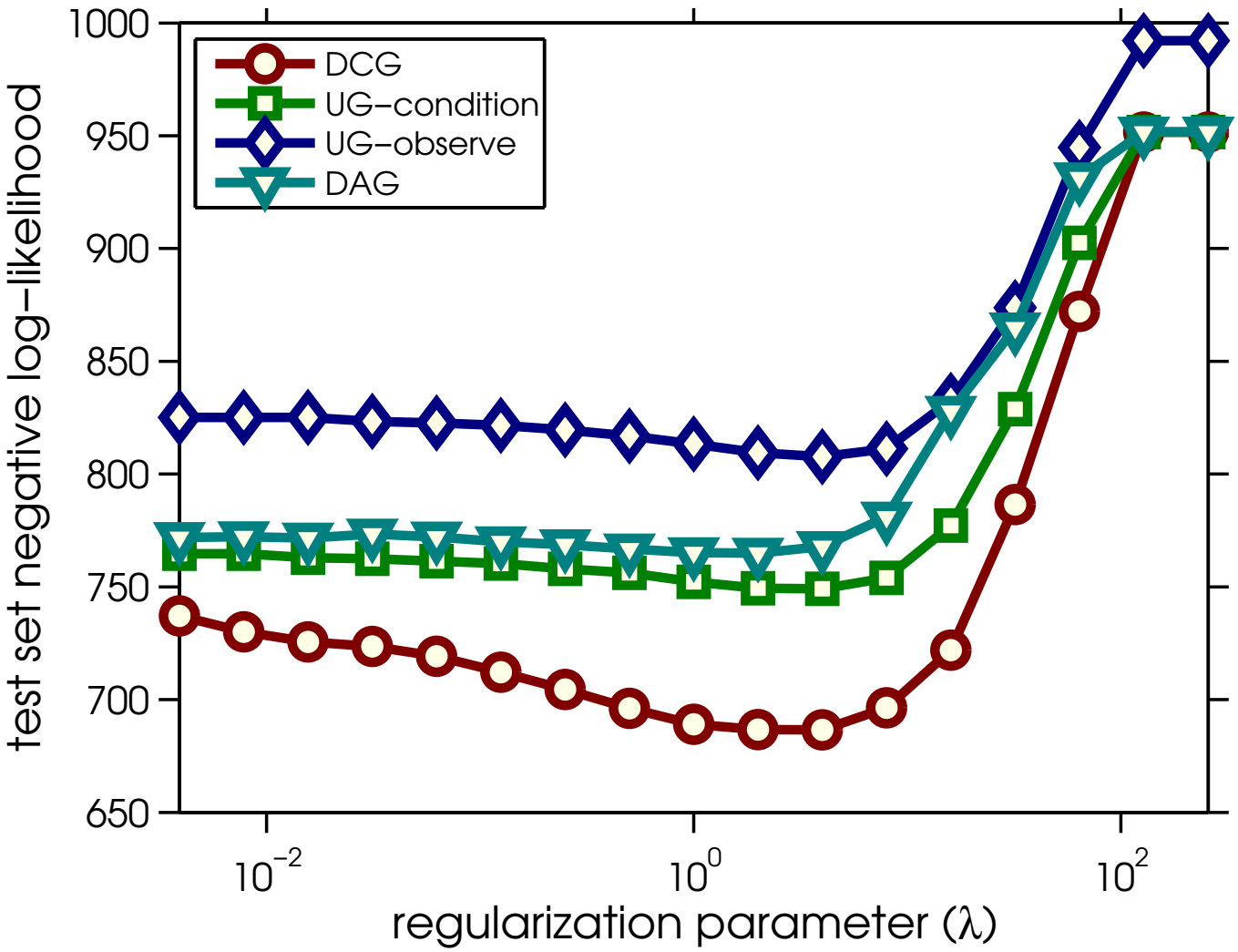
- DAG: An interventional DAG model, searching over all node orderings to select parents.
- UG-observe: A UG model that ignores interventions. (it maximizes $p(x_1, \ldots, x_n)$ over the training samples)
- UG-condition: A UG model that treats interventions as fixed observations during training. (it maximizes $p(x_1, \ldots, x_n)$ on observational samples, and maximize $p(x_{-k}|x_k)$ when we intervene on node k)
- DCG: the proposed model, modeling interventions by removing the appropriate interventional potential.

Experiments on Synthetic Data

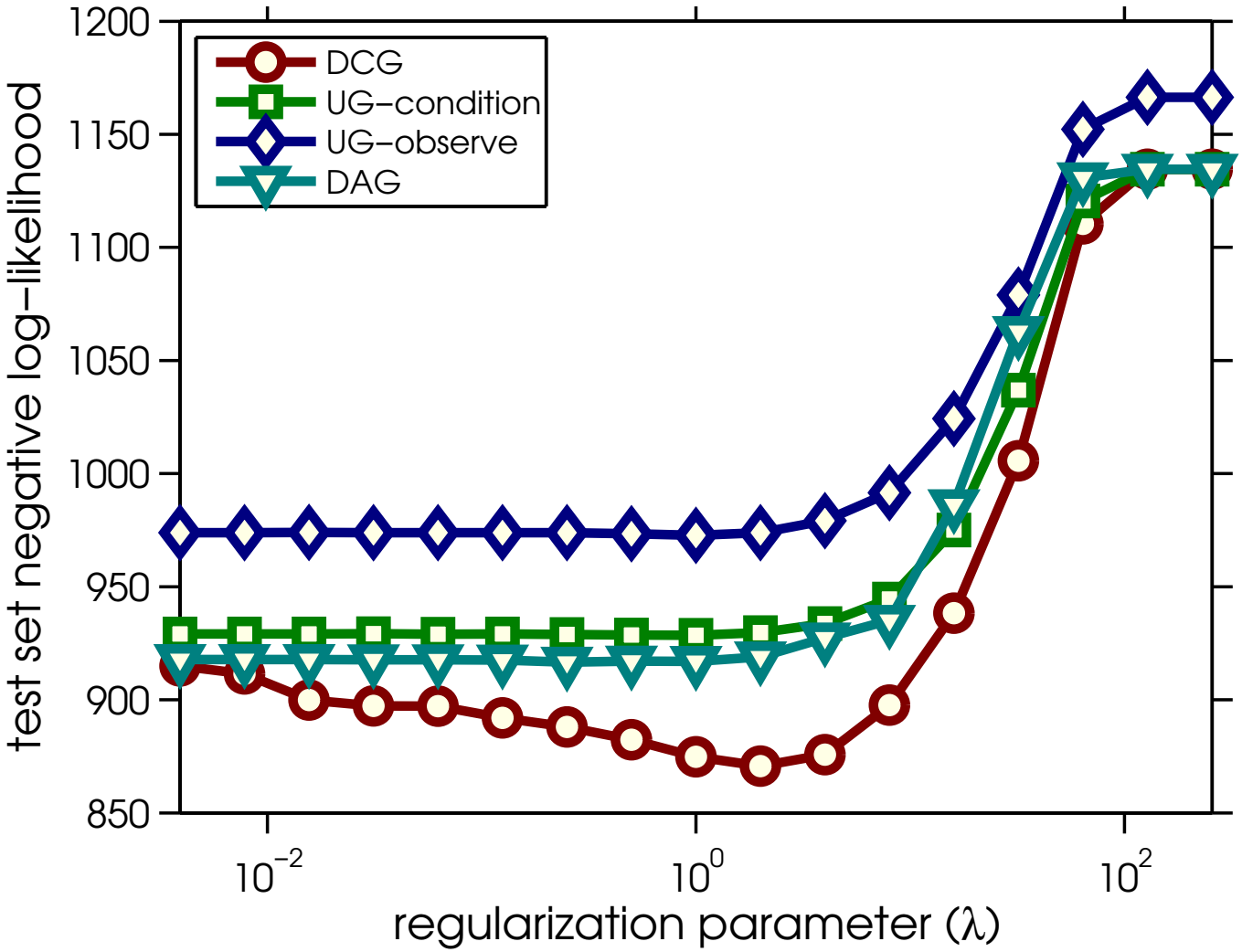
We generated 10 synthetic binary 10 node data sets:

- We included each possible directed edge with probability 0.5
- We generated node and edge parameters from $\mathcal{N}(0, 1)$
- We used an inverse-CDF method to generate 1000 samples
- With probability 1/11 we generate an observational sample
- With probability 10/11 we set a random node by intervention
- We trained on the first 500 samples and evaluated on the remaining 500

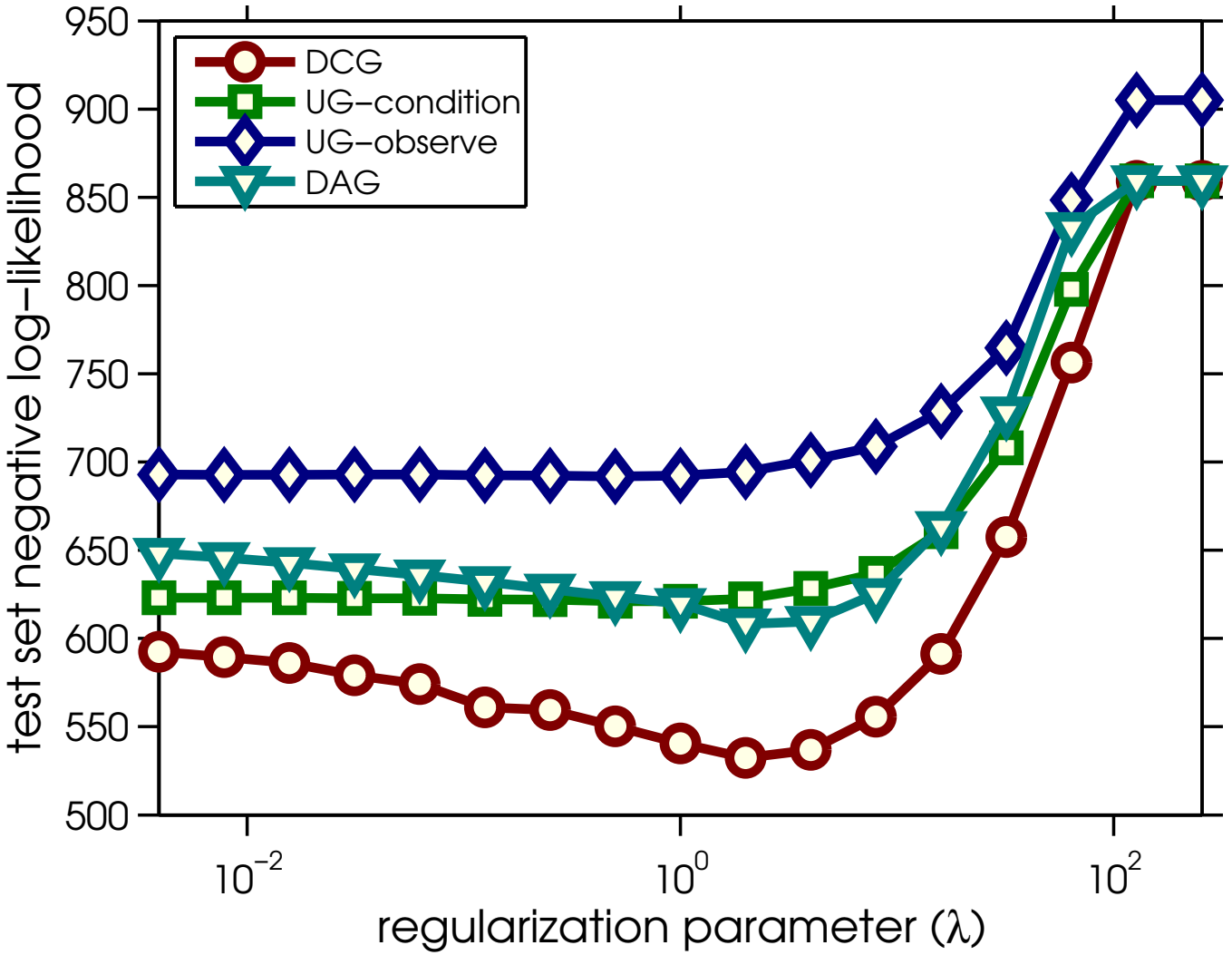
Results on Synthetic Data: Data Set 1



Results on Synthetic Data: Data Set 2

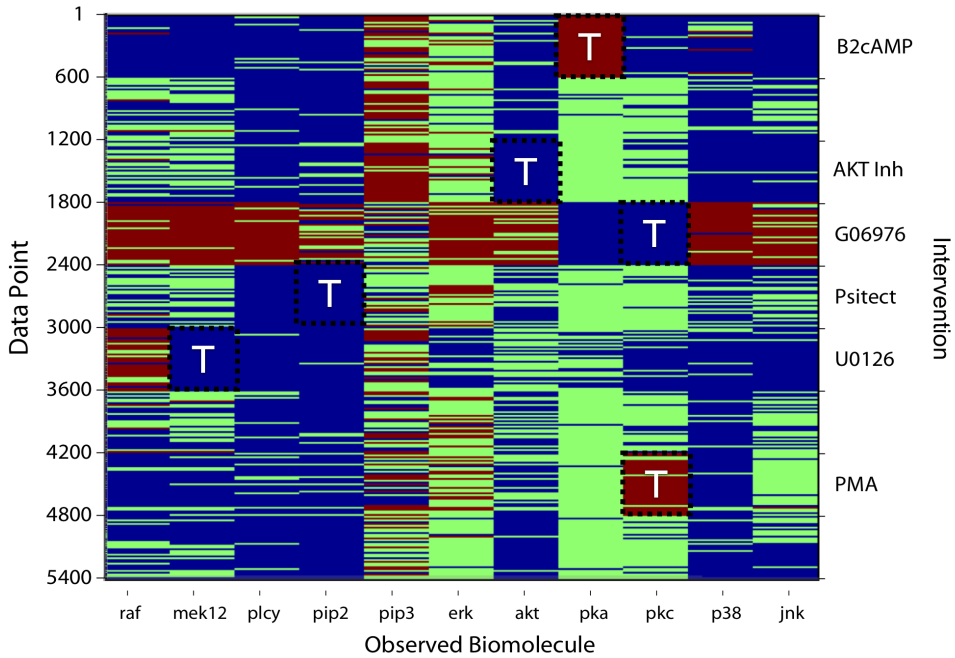


Results on Synthetic Data: Data Set 3



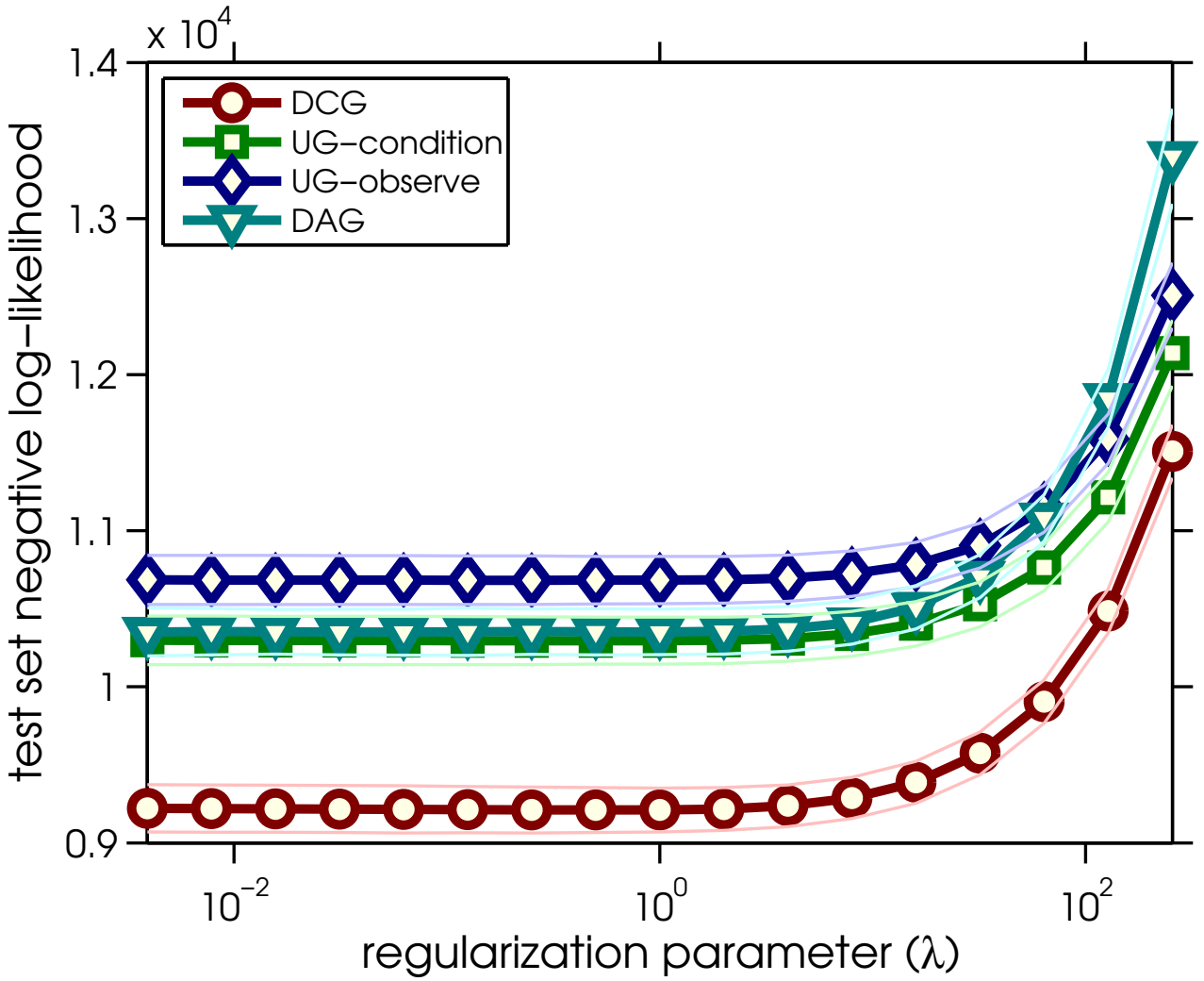
Experiments on Cell Signaling Network Data

We next compared the 4 models on the interventional multivariate flow cytometry data [Sachs et al., 2005].



We trained on half the samples, and tested on the other half. We repeated this with ten random splits to assess variability.

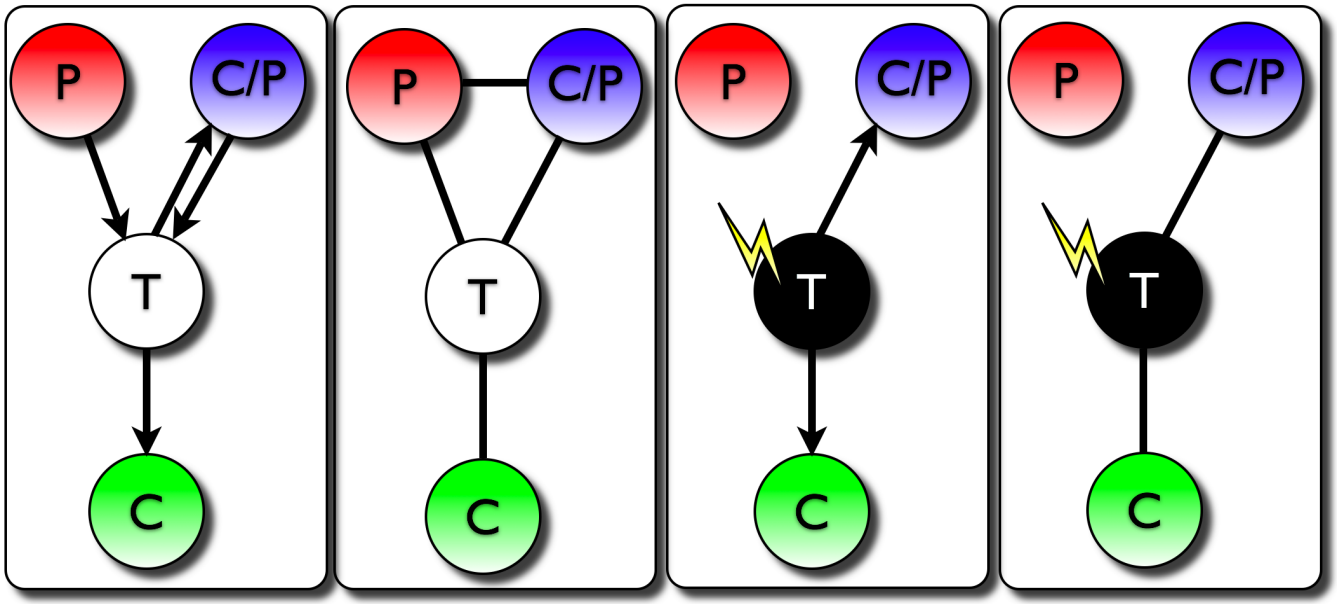
Results on Cell Signaling Network Data



Outline

Discussion: Do we have the right idea?

Weird effect: intervention on a child can remove co-parent edge.

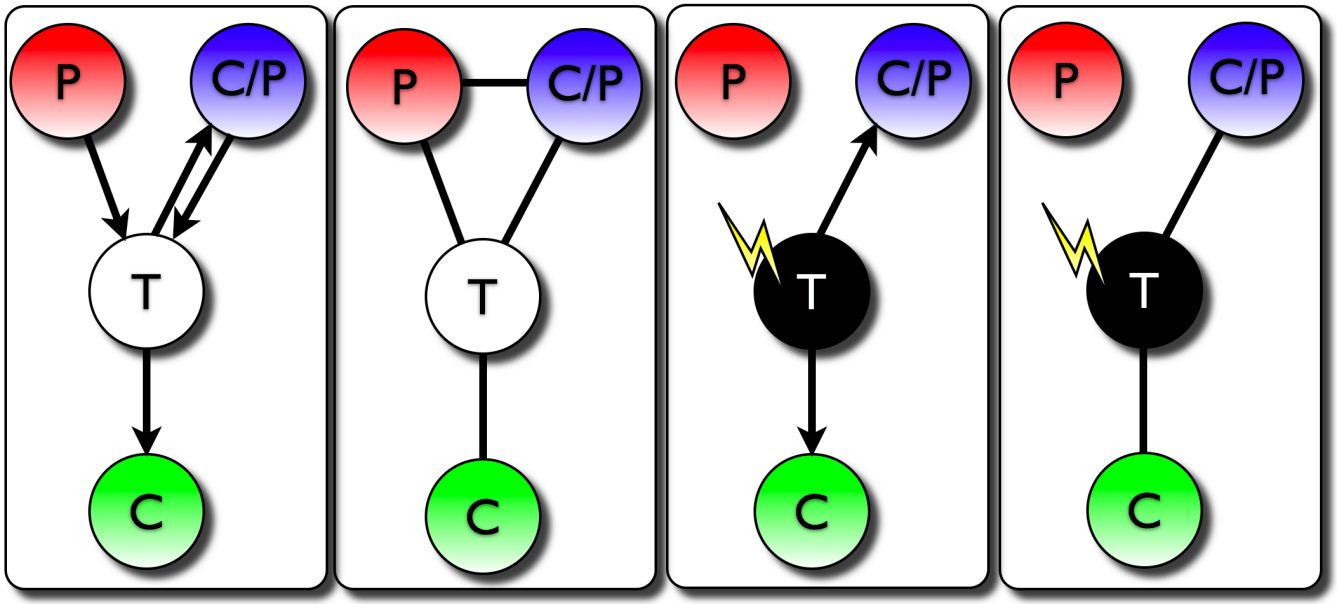


Not relevant to our experiments since edges are not induced between co-parents

But, are interventional potentials the 'right' way to model interventions in undirected models?

Discussion: Do we have the right idea?

Weird effect: intervention on a child can remove co-parent edge.

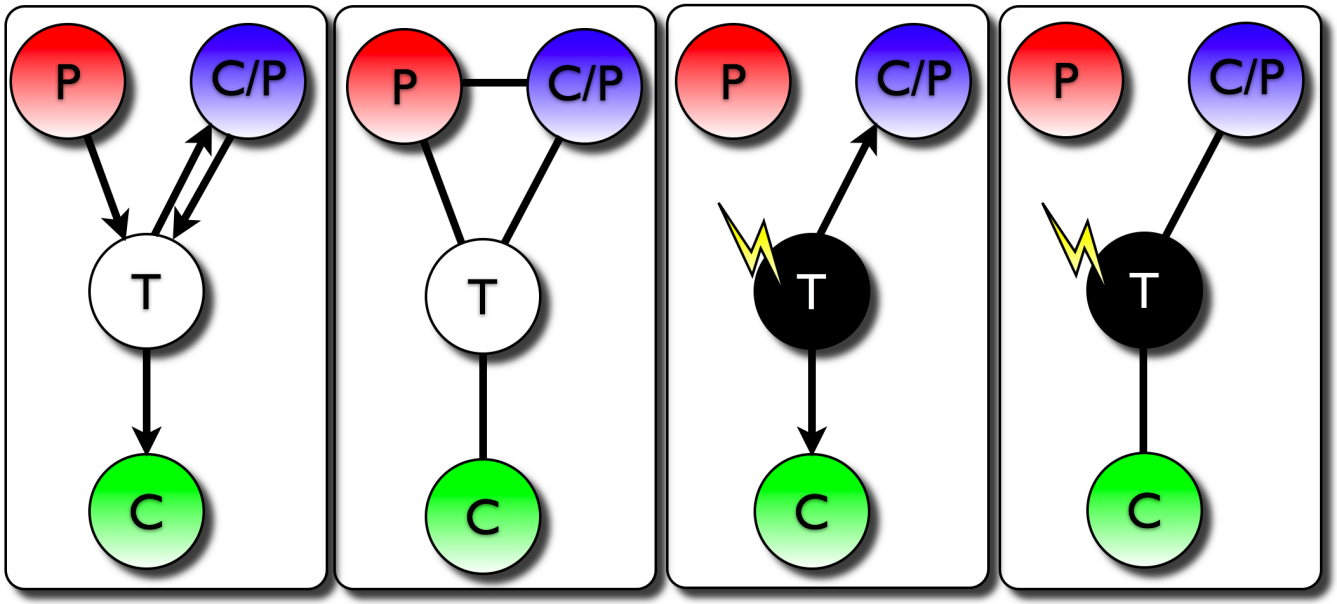


Not relevant to our experiments since edges are not induced between co-parents

But, are interventional potentials the 'right' way to model interventions in undirected models?

Discussion: Do we have the right idea?

Weird effect: intervention on a child can remove co-parent edge.



Not relevant to our experiments since edges are not induced between co-parents

But, are interventional potentials the 'right' way to model interventions in undirected models?

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Discussion: Do we have the right idea?

An equivalent model (for our parameterization) would be to use clique potentials, where we define a 'target' for each potential.

For pairwise interactions, we could consider 4 types of cliques:

- Directed edges: the edge is affected by intervention on the child but not the parent
- Directed 2-cycles: the dependency remains after intervention, but works differently
- Undirected edges: the dependency is not affected by intervention (like regular UGs)
- Unstable edges: the effect is removed after intervention on either node (hidden common cause)

Summary

- We presented a model for interventional data that allows cycles
- We outlined the Markov properties and effects of interventions in the model
- We looked at an exponential family parameterization and a convex approach to structure learning
- We showed some promising results on a real data set
- There are still a lot of issues to explore
  (code online soon)

Thank you to the anonymous reviwers

Summary

- We presented a model for interventional data that allows cycles
- We outlined the Markov properties and effects of interventions in the model
- We looked at an exponential family parameterization and a convex approach to structure learning
- We showed some promising results on a real data set
- There are still a lot of issues to explore
  (code online soon)

Thank you to the anonymous reviwers

Summary

- We presented a model for interventional data that allows cycles
- We outlined the Markov properties and effects of interventions in the model
- We looked at an exponential family parameterization and a convex approach to structure learning
- We showed some promising results on a real data set
- There are still a lot of issues to explore (code online soon)

Thank you to the anonymous reviwers

Summary

- We presented a model for interventional data that allows cycles
- We outlined the Markov properties and effects of interventions in the model
- We looked at an exponential family parameterization and a convex approach to structure learning
- We showed some promising results on a real data set
- There are still a lot of issues to explore
  (code online soon)

Thank you to the anonymous reviwers

Summary

- We presented a model for interventional data that allows cycles
- We outlined the Markov properties and effects of interventions in the model
- We looked at an exponential family parameterization and a convex approach to structure learning
- We showed some promising results on a real data set
- There are still a lot of issues to explore
  (code online soon)

Thank you to the anonymous reviwers