# Discrete Graphical Models with One Hidden Variable

Marco Valtorta

Department of Computer Science and Engineering

University of South Carolina

October 7, 2011

*Joint work with:*

*Elizabeth Allman, John Rhodes (Mathematics, U. Fairbanks)*

*Elena Stanghellini (Statistics, U. Perugia)*

*Support from AIM is gratefully acknowledged.*

# Contents

- Identifiability problems
- Kruskal's Theorem and Its Application
- Examples
- Summary and conclusion

# What is Identifiability?

- The sufficient parameters for discrete Bayesian network with hidden and observable nodes are the conditional probability tables (CPTs) for each family of nodes

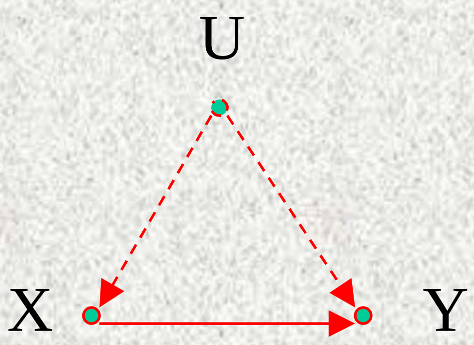
1. **Unidentifiability_1: The ability to determine whether the CPTs can be computed from observable data alone and, if so, to compute them**
2. **Unidentifiability_2: The ability to determine whether the causal effect of a set of observable variables on another observable variable in a causal Bayesian network with hidden nodes can be computed from observable data alone, and, if so, to compute it**

- An Example of case 2 follows

# Unidentifiability_2 Example(1)

- All the variables are binary.
- P(U=0) = 0.5,
- P(X=0|U) = (0.6,0.4),
- P(Y=0|X,U) =

| Y=0 | X =0 | X= 1 |
|---|---|---|
| U =0 | 0.7 | 0.2 |
| U=1 | 0.2 | 0.7 |



# Unidentifiability_2 Example(2)

- Note that

$$P(X,Y)=\sum_U P(Y\mid X,U)P(X\mid U)P(U)$$

- We get:

| | X =0 | X= 1 |
|---|---|---|
| Y =0 | 0.25 (=0.7x0.6x0.5+ 0.2x0.4x0.5) | 0.25 |
| Y=1 | 0.25 | 0.25 |

- Because of the excision semantics, the link from U to X is removed, and we have:

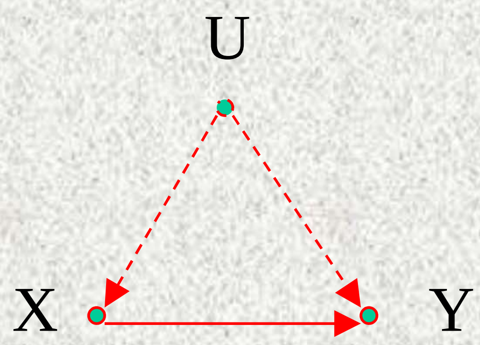
$$P_X(Y)=\sum_U P(Y\mid X,U)P(U)$$

- So, $P_{X=0}$ (Y=0) = (0.7x0.5) + (0.2x0.5) = 0.45

# Unidentifiability_2 Example(3)

- All the variables are still binary.
- P(U=0) = 0.5
- P(X=0|U) = (0.7,0.3)
- P(Y=0|X,U) =

| Y=0 | X =0 | X= 1 |
|---|---|---|
| U =0 | 0.65 | 0.15 |
| U=1 | 0.15 | 0.65 |



# Unidentifiability_2 Example(4)

- Using

$$P(X,Y)=\sum_{U}P(Y\mid X,U)P(X\mid U)P(U)$$

- We still get:

| | X =0 | X= 1 |
|---|---|---|
| Y =0 | 0.25 | 0.25 |
| Y=1 | 0.25 | 0.25 |

- From

$$P_X(Y)=\sum_{U}P(Y\mid X,U)P(U)$$

- We have $P_{X=0}$ (Y=0) = (0.65x0.5) + (0.35x 0.5) = 0.4 <> 0.45
- So, $P_X$(Y) is unidentifiable in this model

# The Identifiability_2 Problem

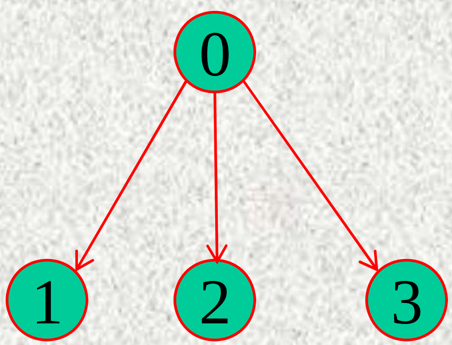
- For a given causal Bayesian network, decide whether $P_t(s)$ (i.e., $P(S \mid do(T))$ is identifiable or not
- If $P_t(s)$ is identifiable, give a closed-form expression for the value of $P_t(s)$ in term of distributions derived from the joint distribution of all observed quantities, $P(n)$

# Contents

- Identifiability problems
- Kruskal's Theorem and Its Application
- Examples
- Summary and conclusion

# Kruskal's Theorem

- Model with one hidden variable (r states) and three observable variables (s1, s2, s3 states)
- Provided that s1, s2, s3 are "large enough" relative to r, the parameters are generically identifiable_1
- In this presentation, we assume that all variables are binary



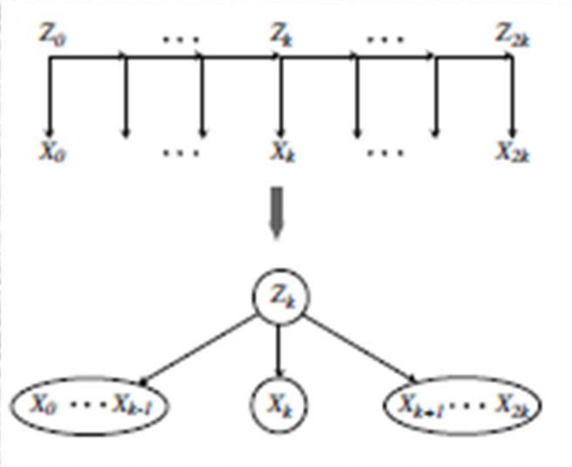

Kruskal Graph

# Application of Kruskal Theorem

Kruskal theorem can be applied to more complicated graphs:

1. Clumping several variables (all hidden or all observed) into a single one, with larger state space
2. **Conditioning on the state of an observed variable**
3. **Marginalizing over an observed variable (making it hidden)**

**Operations 2 and 3 are novel in this context**

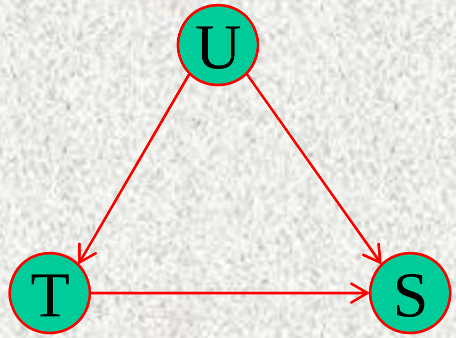


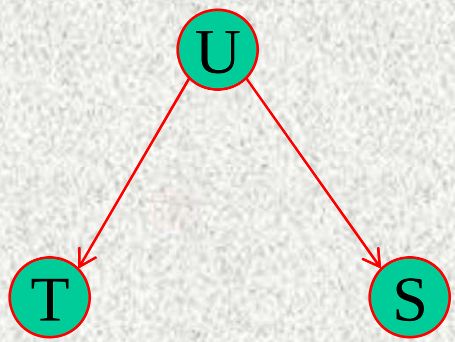
# Contents

- Identifiability problems
- Kruskal's Theorem and Its Application
- Examples
- Summary and conclusion

# Two Observable Variables, One Hidden

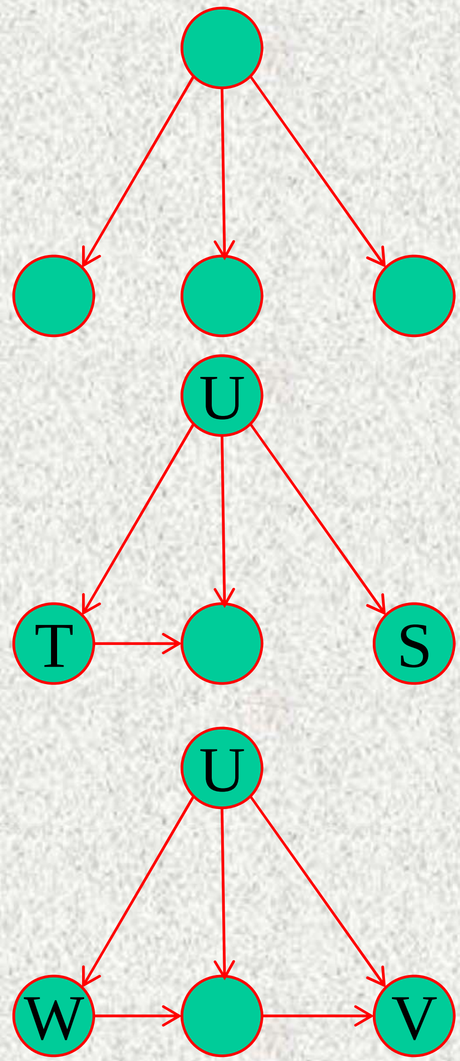
- Neither of the two possible models is identifiable_1
- P(S|do(T)) is unidentifiable_2 in the top model
- P(S|do(T)) is identifiable_2 in the bottom model
  - Effects are independent given their common cause, so when we marginalize out U, the effect of T is eliminated



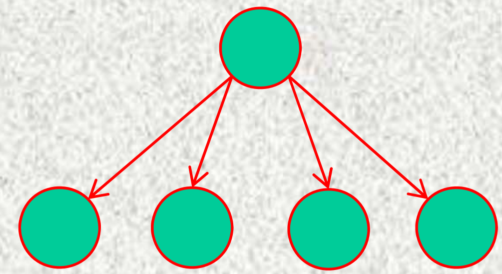
# Three Observed Variables

- The model of the original Kruskal Theorem (top) is (obviously) identifiable_1
- The causal effect of any leaf on any other leaf is identifiable_2
- If any edges are added, the model is unidentifiable_1
- P(S | do(T)) is identifiable_2
- P(V | do(W)) is unidentifiable_2

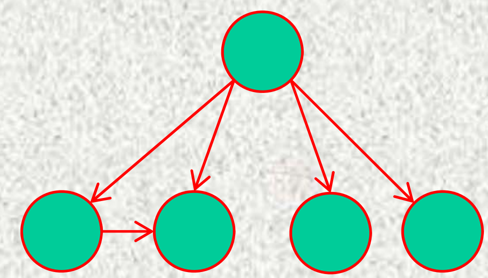


# Four Observed Variables

- Identifiable_1
  - By clumping two observable variables together

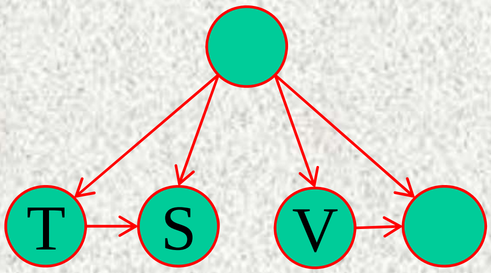


- Identifiable_1
  - By clumping the two observable variables that are connected by an arc



# Four Observed Variables (ctd.)

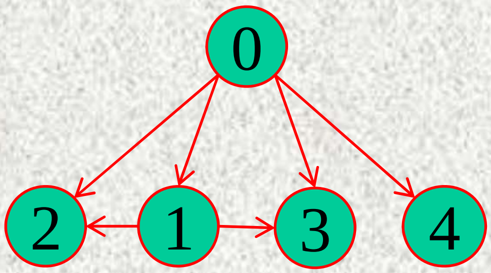
- We conjecture that this is unidentifiable_1, and so are variants where the horizontal arcs are oriented in different ways
- P(S | do(T)) is unidentifiable_2, but
- P(V | do(T)) is identifiable_2



# Four Observed Variables

1. Condition on the states of 1
2. The resulting distributions arise from the Kruskal graph with 0 as the central node
3. Obtain the CPT 4|0 using Kruskal's theorem
4. Obtain 1,2,3,4|0 by inverting 4|0

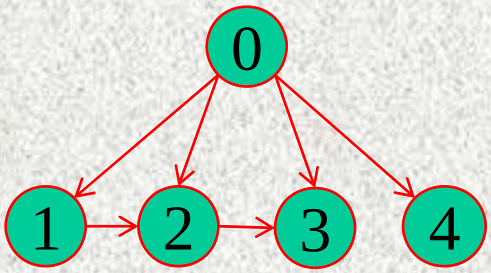
There are a few other ways of obtaining the parameters; one starts by marginalizing out 1




Two edges with a common source: Identifiable_1

# Four Observed Variables

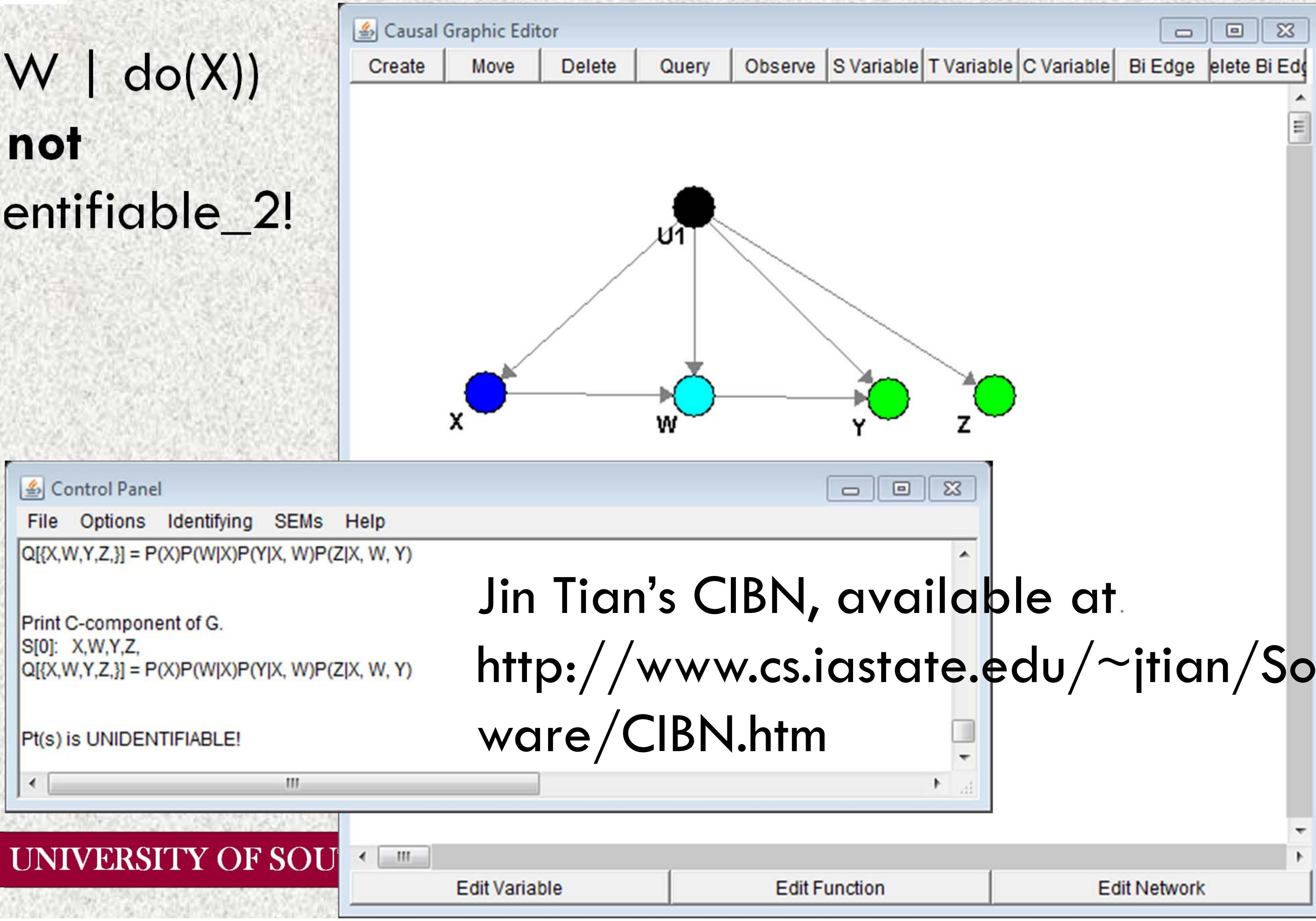
- Condition on 2
- The resulting distribution arise from a Kruskal BN with 0 as the central node
- Apply Kruskal, obtaining the CPTs of 0 and 4|0
- Continue as in the previous case
- Marginalizing over 2 does not seem to work




Two edges forming a directed path: Identifiable_1

# A Surprise

P(W | do(X)) is **not** identifiable_2!




Jin Tian's CIBN, available at http://www.cs.iastate.edu/~jtian/Software/CIBN.htm

# Contents

- Identifiability problems
- Kruskal's Theorem and Its Application
- Examples
- Summary and conclusion

# Comments

- We obtained additional results on graphs with five observables
- I omitted the important issue of generic vs. absolute identifiability.  Our results for identifiability_1 are generic.  The results for identifiability_2 are absolute.
- Some heuristics have emerged, e.g., when both conditioning and marginalization lead to a result, marginalization is more efficient

# Comments (ctd.)

- In some cases, by assuming a hidden variable is binary, a model may go from unidentifiable to identifiable for generic parameter values
- In these cases, it appears that the one needs not rational formulas, but algebraic ones, in order to solve for parameter values
- It appears that for identifiability_2, one always can obtain rational formulas for parameter values, when they are identifiable

# References


- Elizabeth S. Allman, Catherine Matias, and John A. Rhodes "Identifiability of parameters in latent structure models with many observed variables." *Annals of Statistics*, 37 (2009), 3099-3132.
- Yimin Huang and Marco Valtorta. "On the completeness of an identifiability algorithm for semi-Markovian models." *Annals of Mathematics and Artificial Intelligence*, 54 (2008), 363-408.
- Jin Tian and Ilya Shpitser. "On Identifying Causal Effects." In:Rina Dechter, Hector Geffner, and Joseph Halpern (eds.) *Heuristics, Probability, and Causality: A Tribute to Judea Pearl.* College Press (2010), 523-543.
- Elena Stanghellini, Barbara Vantaggi "On the identification of discrete graphical models with hidden nodes." arXiv:1009.4279v1 (2008).

- These slides are available through http://www.cse.sc.edu/~mgv/talks/index.html

# Questions?