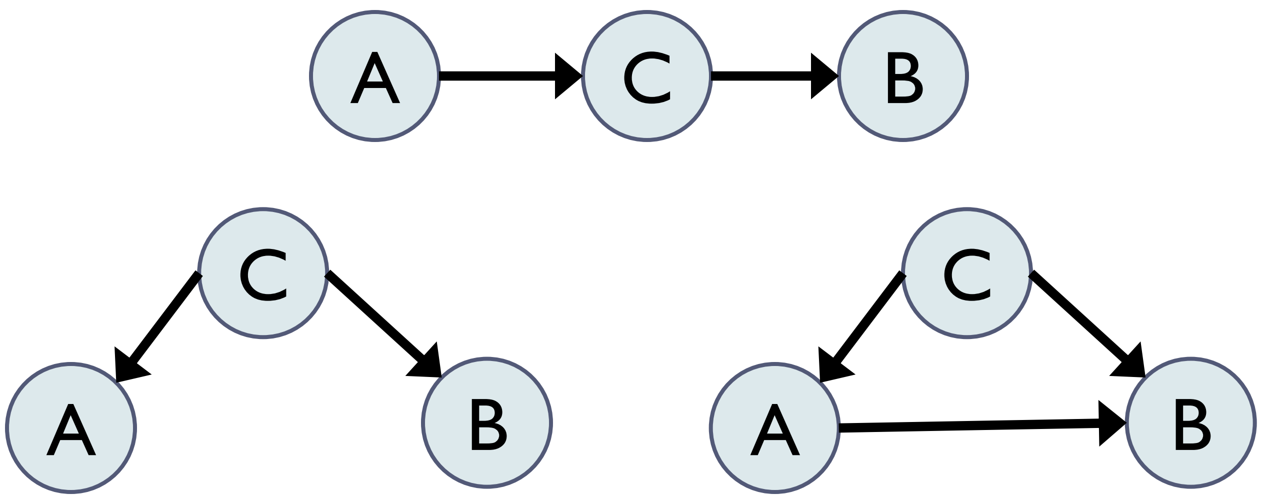
Markov Networks

Doug Downey
Northwestern EECS 395/495 Fall 2014

I-Maps, Perfect Maps, and I-Equivalence

- ***I-Map for S:*** A graph containing at most a set *S* of *independence assertions*, i.e. statements of the form (**X** ⊥ **Y** | **Z**)
- E.g., some I-Maps for *S* = {(A ⊥ B | C)}



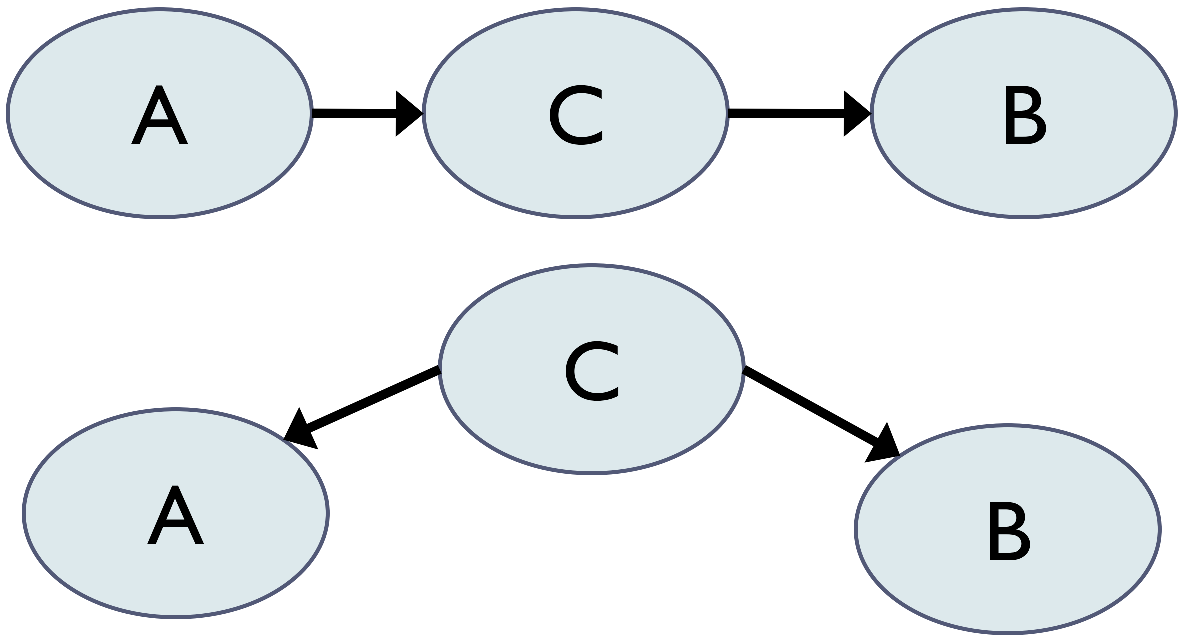
I-Maps: why they matter

- If *G* is an I-Map for the independences in a distribution *P*, then we can represent *P* as a Bayes Net with graph *G*.
  - Whereas we **can't** do so if *G* is not an I-Map for *P*

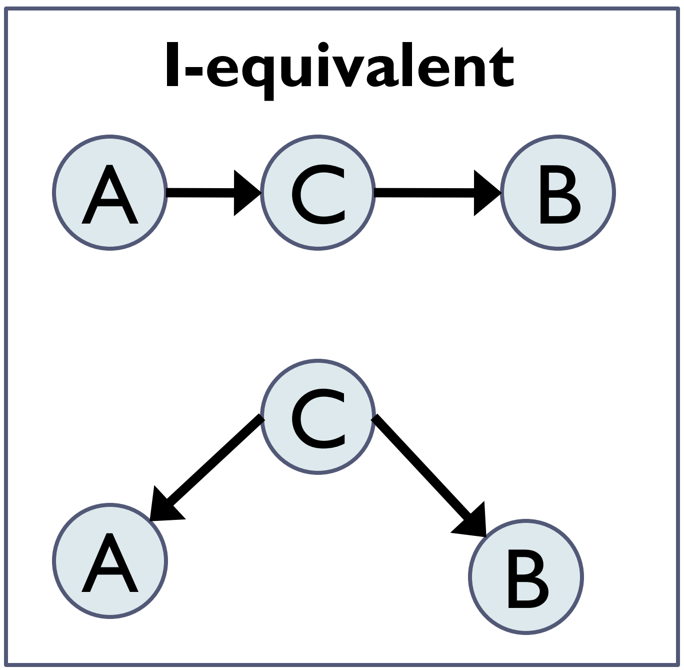
- A given distribution may have many different I-Maps
  - *Minimal I-Map for S*: An I-Map for *S* for which the removal of any edge renders it not an I-Map for *S*
  - *Perfect Map for S:* A graph with *exactly* the set of independencies in *S*

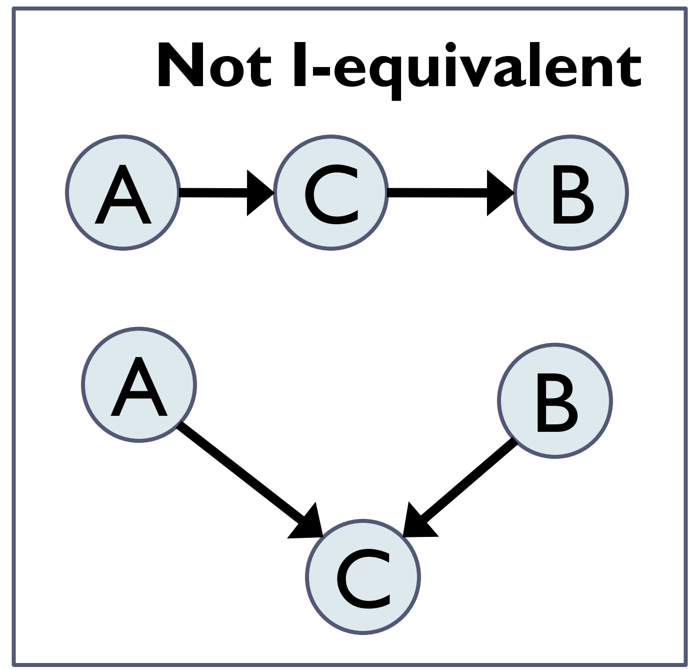
Example

- Two Perfect Maps for $S = \{(A \perp B \mid C)\}$



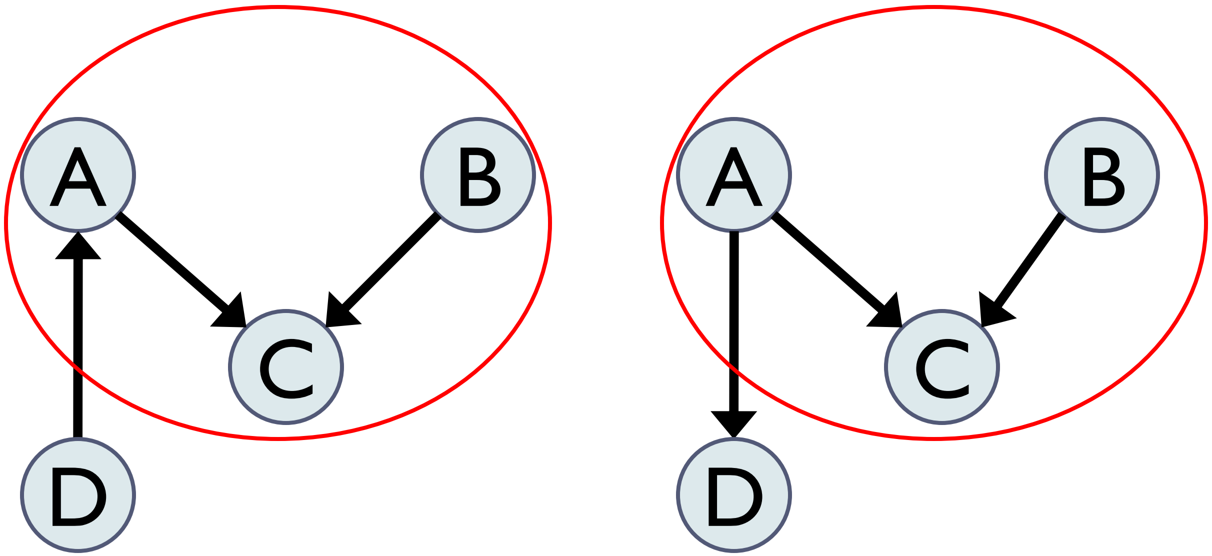
I-Equivalence (1 of 2)

- Two graphs are ***I-Equivalent*** if they imply identical sets of independence assertions
- 



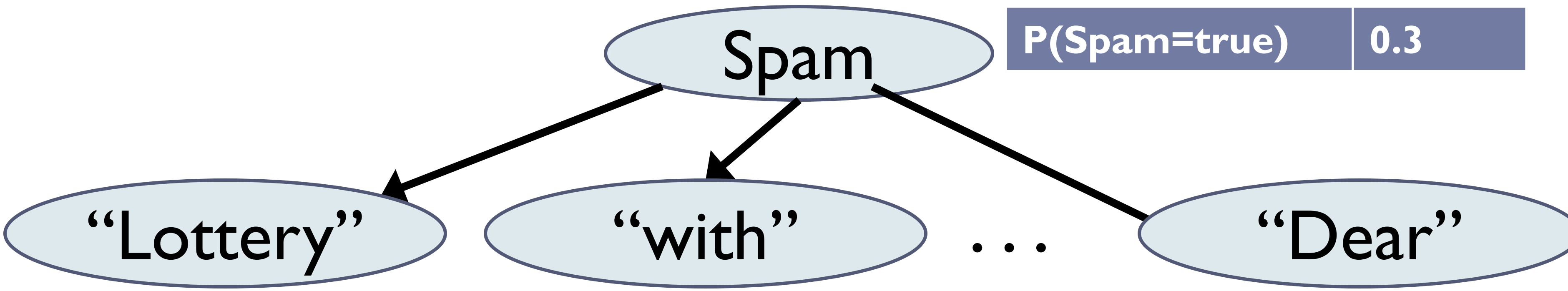
I-Equivalence (2 of 2)

- Two graphs are I-Equivalent *iff* they have the same
  - *Skeleton:* graph ignoring edge direction
  - *Immoralities:* v-structures without direct edge between parents



Sidenote: Naïve Bayes Net

- NB assumes features conditionally indep. given the class:



P(Spam=true)	0.3

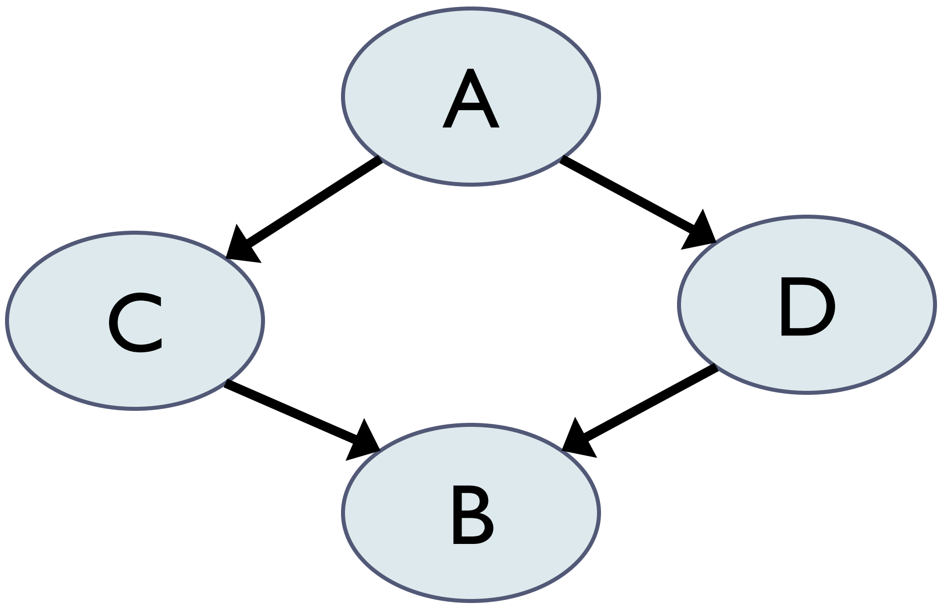
Spam	P(“lottery” | Spam)
true	0.04
false	0.01

Spam	P(“with” | Spam)
true	0.6
false	0.59

Spam	P(“dear” | Spam)
true	0.24
false	0.30

Limitations of Bayesian Networks

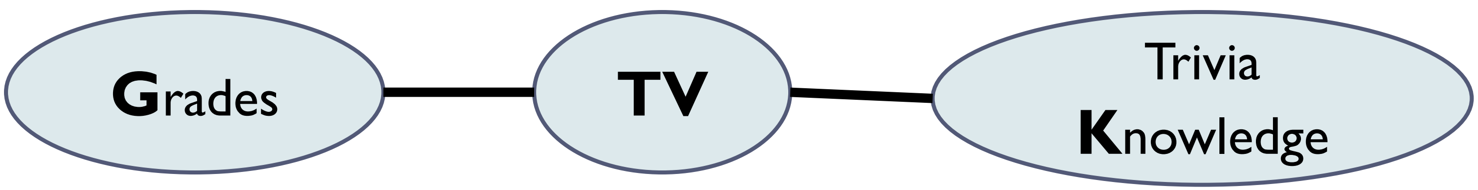
- Perfect Map for $\{(A \perp B \mid C, D), (C \perp D \mid A, B)\}$?



- Not possible! Bayes Nets can't express all possible sets of independence assertions.

Alternative: Markov Networks

- Undirected Graphical Model
  - No CPTs. Uses **potential functions** ϕ_c defined over cliques
- $P(\mathbf{x}) = \frac{\prod_c \phi_c(\mathbf{x}_c)}{Z}$ $\quad Z = \sum_{\mathbf{x}} \prod_c \phi_c(\mathbf{x}_c)$



Grades	TV	ϕ_1(G,TV)
bad	none	2.0
good	none	3.0
bad	lots	3.0
good	lots	1.0

TV	Trivia Knowledge	ϕ_2(TV, K)
none	weak	2.0
lots	weak	1.0
none	strong	1.5
lots	strong	3.0

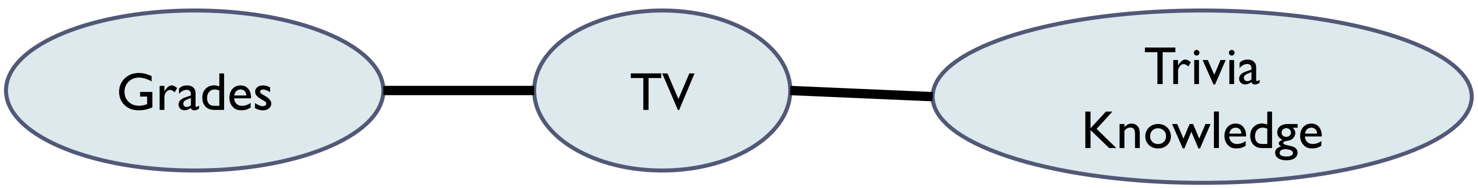
Markov Net Joint Distribution

Grades	TV	Trivia Know.	$\phi_1(G, TV)$	$\phi_2(TV, K)$	$\phi_1(G, TV) * \phi_2(TV, K)$	P(G, TV, K)
bad	none	weak	2.0	2.0	4.0	0.12
good	none	weak	3.0	2.0	6.0	0.18
bad	lots	weak	3.0	1.0	3.0	0.09
good	lots	weak	1.0	1.0	1.0	0.03
bad	none	strong	2.0	1.5	3.0	0.09
good	none	strong	3.0	1.5	4.5	0.13
bad	lots	strong	3.0	3.0	9.0	0.27
good	lots	strong	1.0	3.0	3.0	0.09

$Z = 33.5$

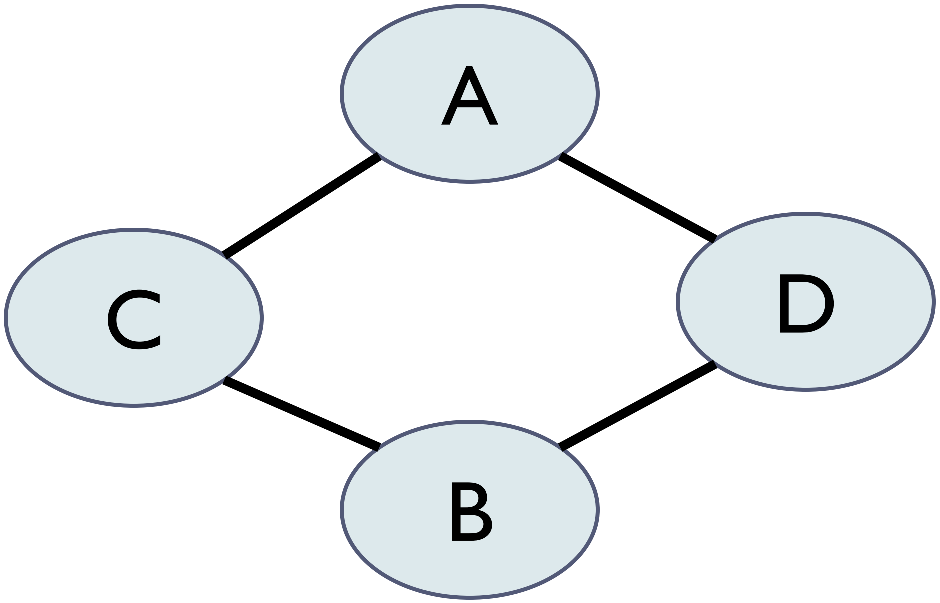
Markov Nets Independence Assertions

- Instead of D-separation, simply graph separation
  - So (Grades $\perp$ Trivia Knowledge | TV)



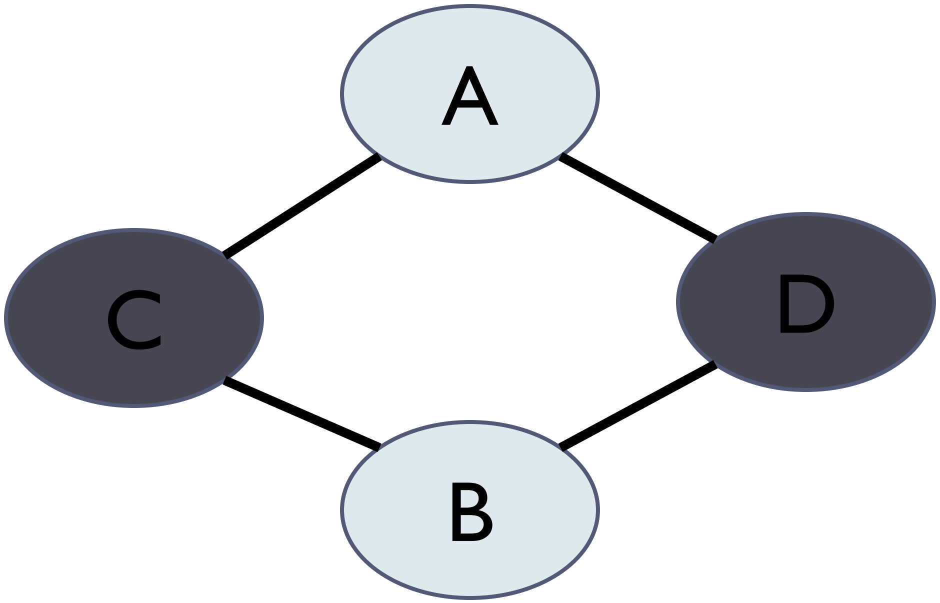
Expressivity of Markov Networks

- Perfect Map for $\{(A \perp B \mid C, D), (C \perp D \mid A, B)\}$?



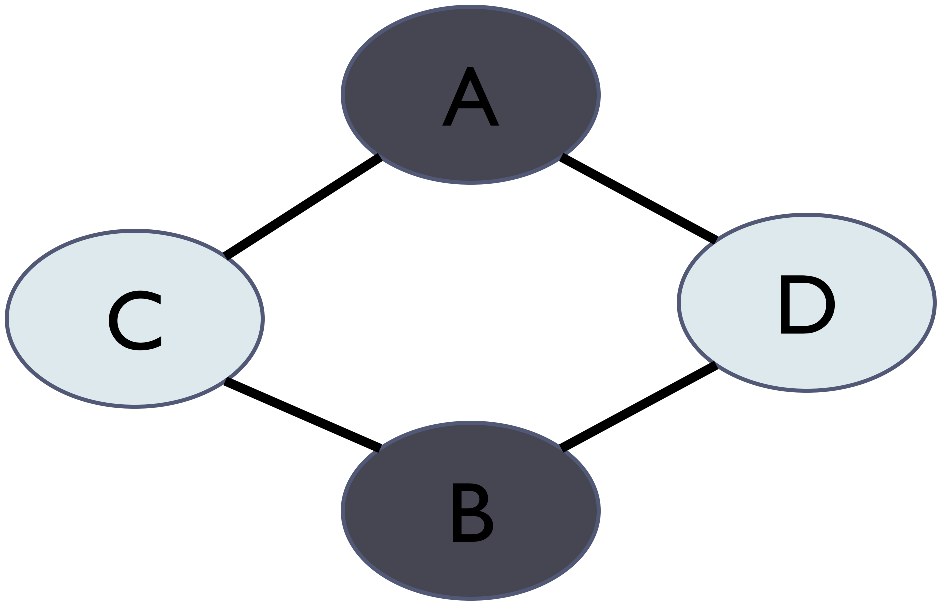
Expressivity of Markov Networks

- Perfect Map for {$(\mathbf{A \perp B \mid C, D})$, $(C \perp D \mid A, B)$}?



Expressivity of Markov Networks

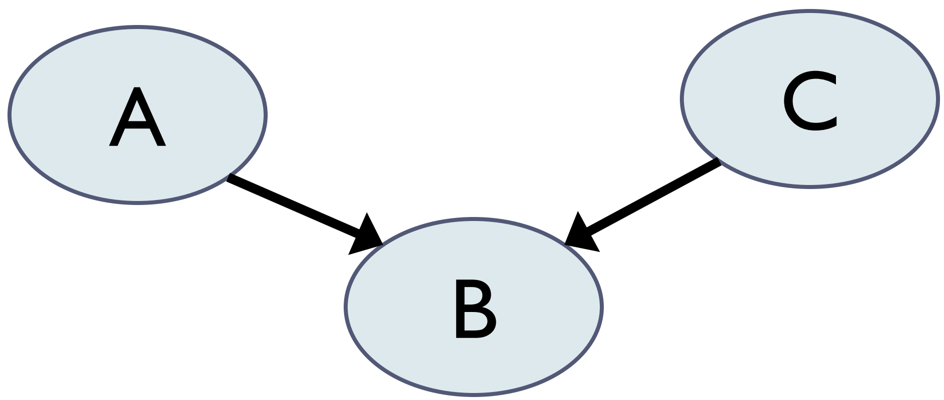
- Perfect Map for {(A ⊥ B | C, D), **(C ⊥ D | A, B)**}?



- Markov Nets *can* capture these independence assertions

But…

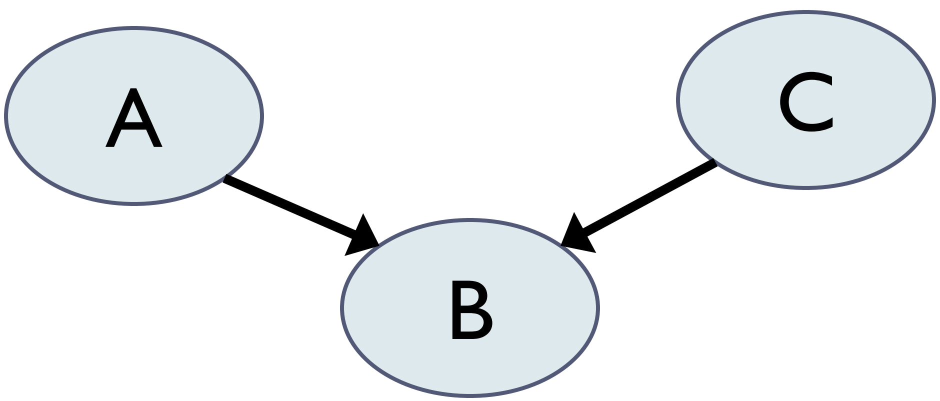
- How about $(A \perp C) \in S$, but $(A \perp C \mid B) \notin S$?



- Can't be captured perfectly in Markov Networks
- If graph separation -> conditional independence, new knowledge can only **remove** dependencies

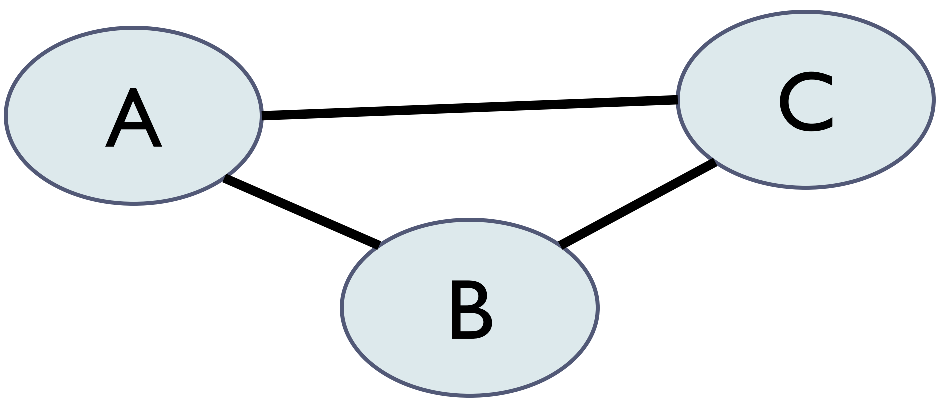
Bayesian Networks => Markov Networks

- Markov Nets can encode independences that Bayes Nets cannot, and vice-versa
- To convert from BN to MN, "moralize":



Bayesian Networks => Markov Networks

- Markov Nets can encode independences that Bayes Nets cannot, and vice-versa
- To convert from BN to MN, "moralize":



Markov Net Applications

- Best when no clear, directed causal structure
  - E.g. statistical physics, text, social networks, image analysis (e.g. segmentation, below)

Zoltan Kato http://www.inf.u-szeged.hu/ipcg/projects/RJMCMC.html