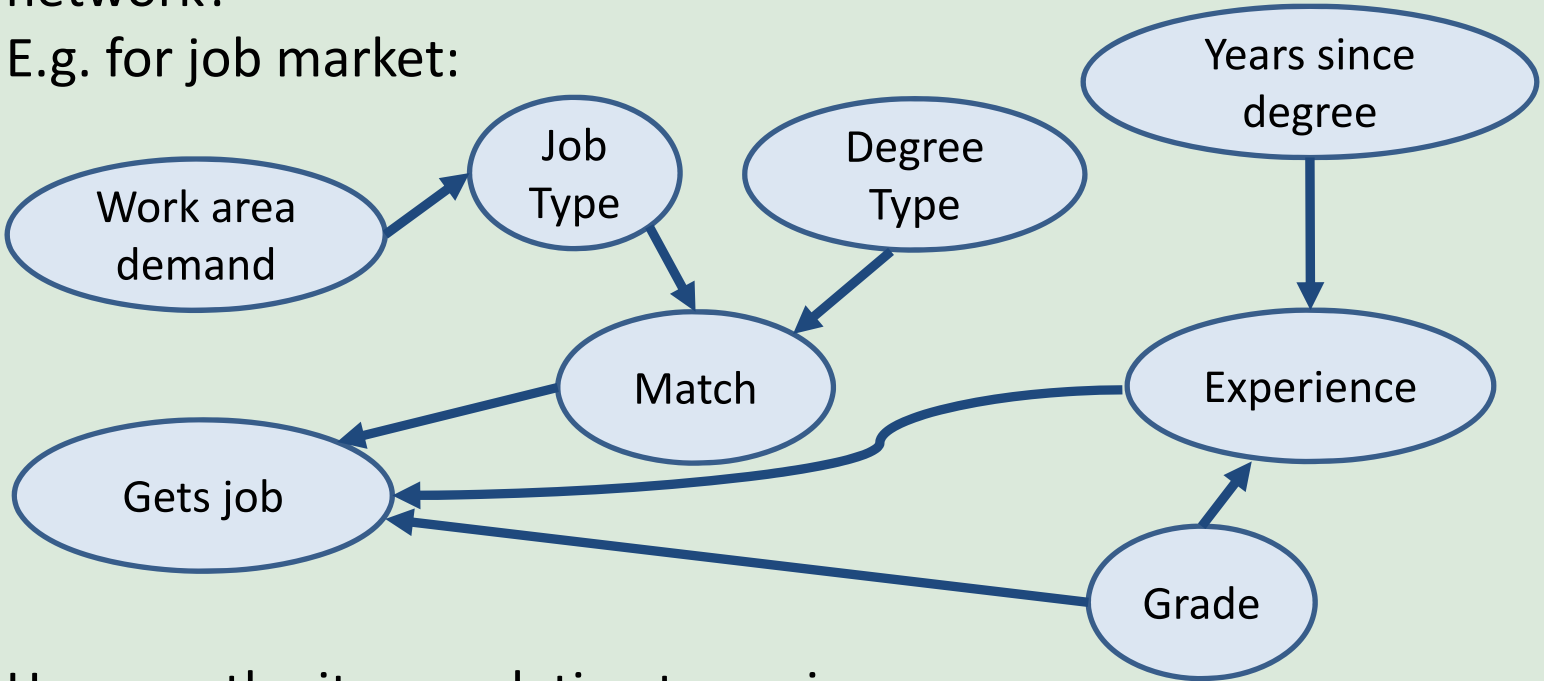
iPMR

Exercise

- Consider the variables below. By considering what items are directly dependent on other items can you build a dependency network?
- E.g. for job market:



- Here are the items relating to car insurance:
- Car colour, Sex of driver, Age, Driver Safety, Accident within previous three years, place of work, place of residence, average annual mileage, car make, car has modifications, accident within a year.

Go for it…

- Car colour (C), Sex of driver (S), Age (A), Driver Safety (DS), Accident within previous three years (3), place of work (W), place of residence (H), average annual mileage (D), car make (M), car has modifications (P), accident within a year (£)

Belief Networks

- **Belief Networks** (aka Bayesian Networks, Bayes Nets) represent the structure of probability distributions in ways that relate to the idea of a dependency network.
- **Starting Point:** The joint probability distribution.

$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2,x_1)P(x_4|x_3,x_2,x_1)\ldots$$

or more formally

$$P(\mathbf{x}) = P(x_1)\prod_{i=2}^{M} P(x_i|x_{<i})$$

Notation:

- I use *M* rather than *D* (which Barber uses) for the dimensionality of a variable. (D means dataset).
- $<i$ means the set of all the indices that are less than *i*.
- *x* with a set or vector subscript means the collection of *x* values with subscripts in that set/vector.

Conditional Probability

- Consider

$$P(x_6 | x_5, x_4, x_3, x_2, x_1)$$

- Suppose all x_i were binary.
- How would you encode this probability?

Conditional Probability Tables

- Let us look at the simple case

For $P(\text{Toothache}, \text{Cavity})$ we can write

	Toothache = true	Toothache = false
Cavity = true	0.04	0.06
Cavity = false	0.01	0.89

For $P(\text{Cavity}|\text{Toothache})$ we can write

	Toothache = true	Toothache = false
Cavity = true	0.8	0.063
Cavity = false	0.2	0.937

- But what if conditioning on many items?
- Multidimensional table. Very costly.

Conditional Independence

- Conditional dependence in belief networks

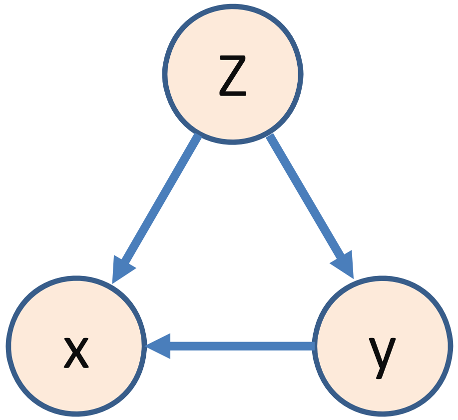
$$P(\mathbf{x}) = P(\mathrm{x}_1)P(\mathrm{x}_2|\mathrm{x}_1)P(\mathrm{x}_3|\mathrm{x}_2,\cancel{\mathrm{x}_1})P(\mathrm{x}_4|\mathrm{x}_3,\cancel{\mathrm{x}_2},\mathrm{x}_1)\ldots$$

- Utilising conditional independence in the chain rule gives us a more compact representation.
- Think back to the dependency network you built earlier. What were you constructing?

Graphically

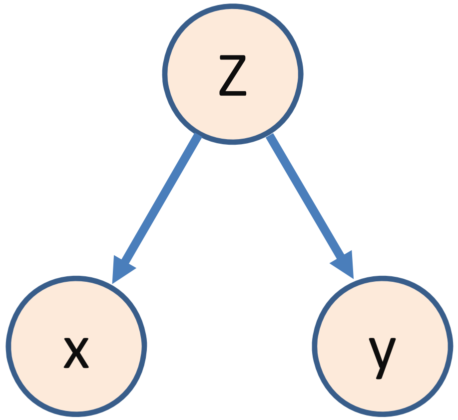
- No independence:

$$P(x, y, z) = P(z)P(y|z)P(x|y, z)$$



- I(x,y|z):

$$P(x, y, z) = P(z)P(y|z)P(x|z)$$



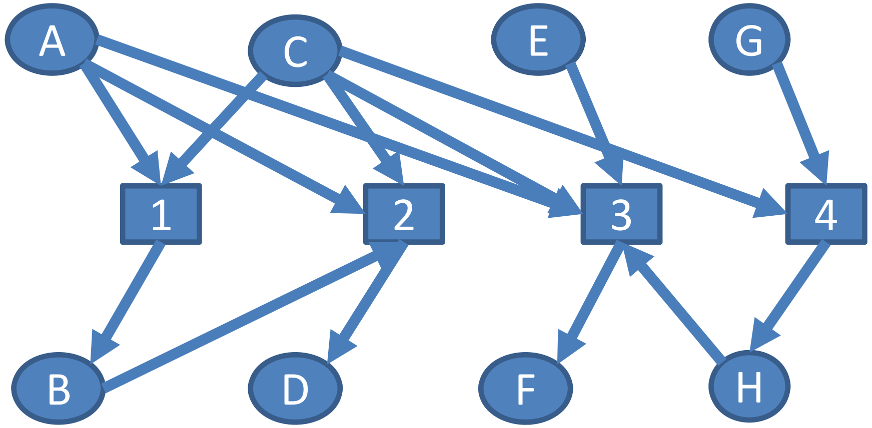
Belief Networks

- Graphical notation that represents various conditional independence assertions for a joint probability distribution.
- How?
  - A Directed Acyclic Graph (DAG) with one node per variable.
  - Look at chain rule expansion.
  - Include all edges except where a variable is dropped from the conditional probability.
  - If $P(\mathrm{r}|\mathrm{s,t,u,v})$ appears in chain rule, but

$$P(\mathsf{r}|\mathsf{s},\mathsf{t},\mathsf{u},\mathsf{v}) = P(\mathsf{r}|\mathsf{s},\mathsf{u})$$

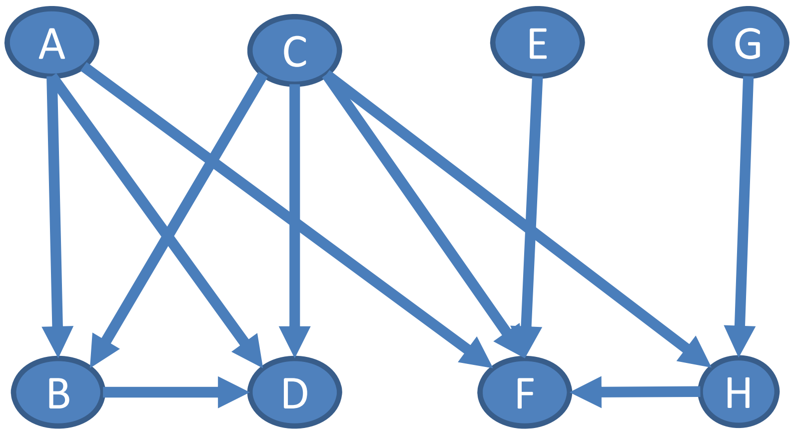
  - then drop directed edge (arrow) from t to r and from v to r.
- Reminder – you should have looked through the Barber notes on Graph Theory.

Belief Networks from Factor Graphs



If:
All edges are directed
All variables have one input*
No **directed** cycles

$$P(\mathbf{x}) = P(\mathrm{x}_1)P(\mathrm{x}_2|\mathrm{x}_1)P(\mathrm{x}_3|\mathrm{x}_2,\mathrm{x}_1)P(\mathrm{x}_4|\mathrm{x}_3,\mathrm{x}_2,\mathrm{x}_1)\ldots$$



Collapse factors onto nodes that are pointed at to get a **directed acyclic graph**

= Belief network

Special case of a general factor graph.
No loss of representational power*
Reversible: just uncollapse.

* Multi-input variables can be handled but complicated and map to BN involves loss of representational power

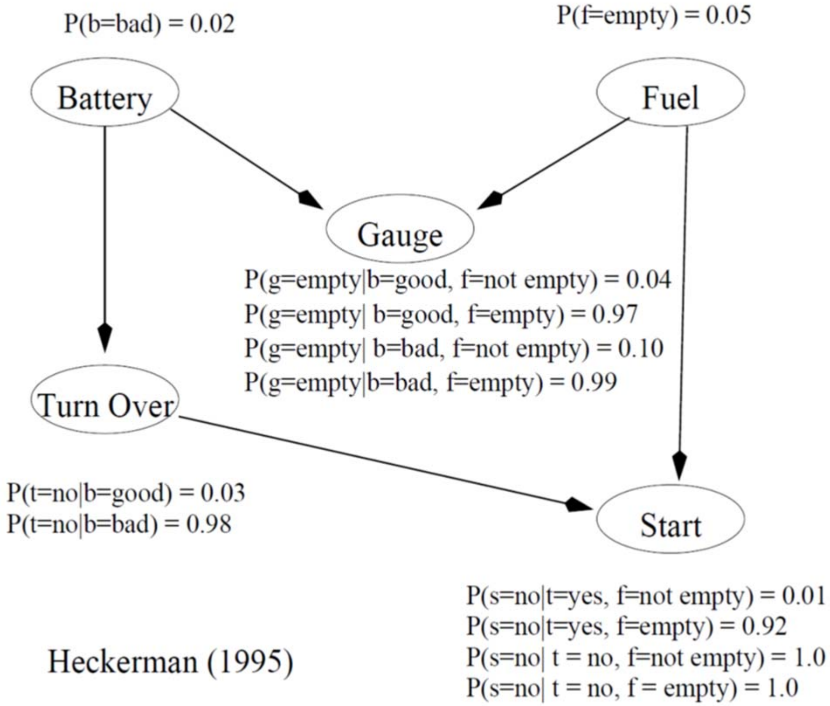
Graphs

- Directed Acyclic Graphs
  - A Directed Acyclic Graph (DAG) is a graph with only directed edges between nodes, and where there are no directed cycles.
  - (Directed) Acyclicity: Can number the nodes so no edge can go from a node to a node with a lower number.

  - Relate this to the chain rule.

Working with Joint Probabilities

- Consider



- Assume for now we are given the probabilities.
- Then we can build the full probability distribution.
- Again we can ask inferential questions
  - E.g. What is $P(\mathsf{f}|g=\mathsf{empty}, s=\mathsf{no})$ ← Inference

Break

Belief Network

Definition 1 *Belief Network. A Belief Network is a distribution of the form*

$$P(\mathbf{x}) = \prod_i P(\mathbf{x}_i | \mathbf{x}_{Pa(i)})$$

along with a corresponding directed graph with arrows pointing to i from $Pa(i)$. Here, $Pa(i)$ denotes the parents of the node corresponding to x_i in the graph.

- Some questions.
  - Does a particular distribution correspond to one belief network?
  - Given a set of independence relationships encoded in a network, is that network representation unique?
  - Can we always encode all conditional independencies using a belief network?
  - Is it right to interpret a belief network causally?

No, No, No, No (not in general), but useful to construct belief nets causally…

Constructing Belief Networks

- Choose a set of variables (those relevant to the domain).
- Choose an order to those variables.
- For each variable in turn
  - Add a node to the graph for that variable.
  - Add directed edges *from* all the existing nodes the variable *directly* depends on *to* the new node.
  - Add the corresponding conditional probability to the chain rule.
- Note the sensitivity to the order.
- If we choose a less good order, we may end up encoding fewer conditional independencies (and so have costly encoding).
- Hint: Choose order causally, from cause to effect, to naturally capture the most independence relationships in the graph.
- But always remember that a belief network does not necessarily encode a causal order.
  - A belief network that does encode causal order is called a *causal graph.*
- Some examples.

Independence Relationships

So a Belief Network encodes conditional independence relationships, right?

Yes…

So it must be trivial to read off the conditional independence relationships from a belief net, right?

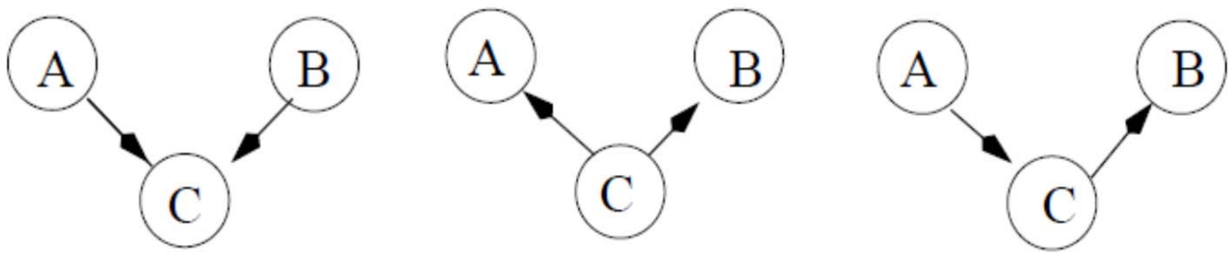
Um, Er, Um, Er, well kind of…

- No brainer strategy: convert it to a factor graph.
- Use the factor graph separation rule.

- Alternatively: use D-Separation rule.

Rules of D-Separation

- To find out if things are independent we use D-Separation.
- If *every* path from a set of nodes *X* to a set of nodes *Y* is blocked by set *Z*, then we have $I(\mathbf{x}_X, \mathbf{x}_Y | \mathbf{x}_Z)$
- A path is blocked iff
  - A node in *Z* is on the path and is head to tail wrt the path.
  - A node in Z is on the path and is tail to tail wrt the path.
  - There is a node on the path that is head to head, and neither that node nor any of its descendants is in *Z.*



C is head-to-head   C is tail-to-tail   C is head-to-tail

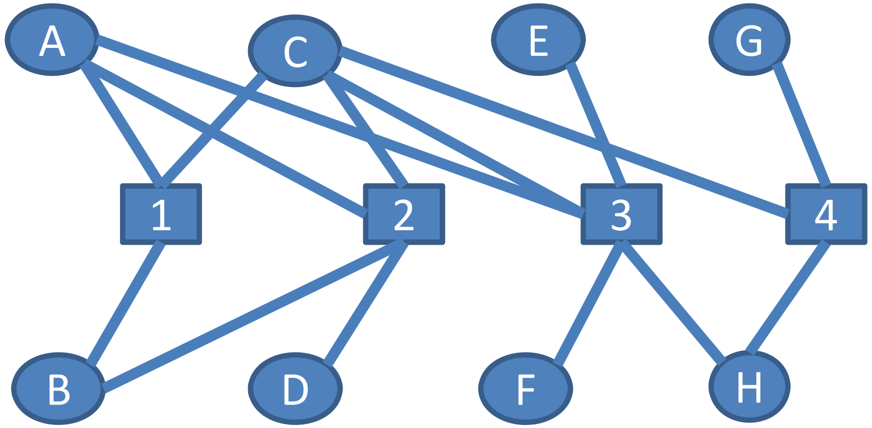
Advice

- Build your belief network
- Use the equivalent directed factor graph for computing conditional independence.
- Use the equivalent undirected factor graph for solving inferential problems.

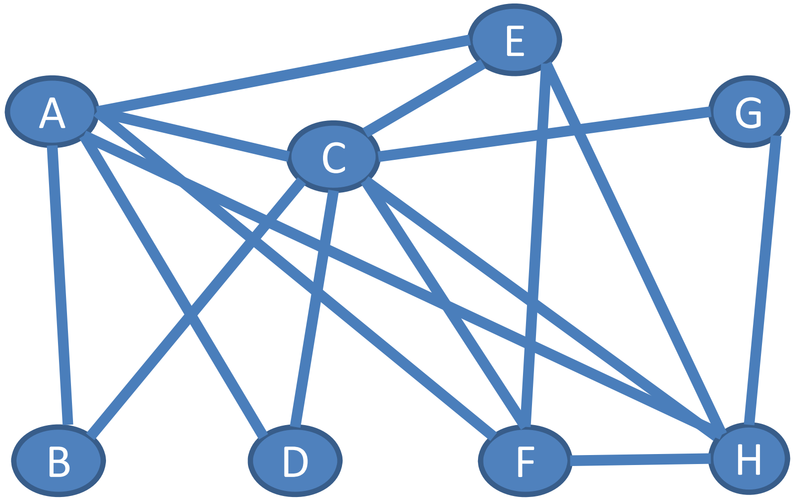
Markov Network

- A Markov network is an undirected graphical model.
- It is built from a factor graph by linking together all the nodes in each factor.
- A link in a Markov network encodes a direct dependence:
  - An edge i -- j is missing implies $I(x_i, x_j | \mathbf{x}_{\setminus\{i,j\}})$
  - The second term denotes all the *x* variables apart from the *i*th and *j*th term.
- Markov networks are a special case of factor graphs.

Terminology: a Markov Random Field is a Markov Network.



If:
All edges are undirected



Connect all nodes adjacent to each factor
Potential loss of representational power.

Reverse:

Find all maximal cliques (biggest fully connected subgraphs)
Make factor for each clique.

Independence Relationships

So a Markov Network encodes conditional independence relationships, right?

Yes…

So it must be trivial to read off the conditional independence relationships from a belief net, right?

Well kind of…

- No brainer strategy: convert it to an undirected factor graph.
- Use the factor graph separation rule.

- Alternatively: use U-Separation rule.

U-Separation

- Conditional independence in Markov networks is much simpler.
- Separation Method:
  - Consider I(A,B|C).
  - Remove all the nodes C from the graph, and all links connecting node C.
  - If there is now no path from A to B, the independence relationship holds.

Conversion

- Converting between network types.
- Directed/Mixed Factor Graph → Undirected Factor Graph
- Directed Factor Graph →Belief Network
- Belief Network → Directed Factor Graph
- Undirected Factor Graph →Markov Network
- Markov Network →Undirected Factor Graph (Lossy)
  - Can you think of a Factor Graph s.t.

    FG → MN → FG

    returns something different from what you started with?
- Undirected Factor Graph → Directed Factor Graph/Belief Network = **Inference (Last Lecture)**
- Markov Network → Directed Factor Graph/Belief Network = **Inference**

Advice

- Work with factor graphs where possible.
- They are more general than both belief networks and Markov Networks.
- Use the undirected factor graph for solving inferential problems.

Our Journey

Graphical Models

- Lecture 2-4: Introduce Factor Graphs
  - Distributions → Factor Graphs
  - Conditional Independence in Factor Graphs
  - Inference in Factor Graphs
- Lecture 5: Belief Networks and Markov Networks
- Next Lecture: Learning as Inference