Probabilistic Inference and Monadic Second Order Logic

Hans L. Bodlaender

Universiteit Utrecht

This presentation

- Main result: combination of two famous and classic results from different fields:
  - Courcelle’s theorem (algorithmic graph theory)
  - Lauritzen-Spiegelhalter algorithm (decision support systems, probabilistic networks)
  - Both use *treewidth* and *tree decompositions*
- (Small print: Notational imprecisions...)

Contents

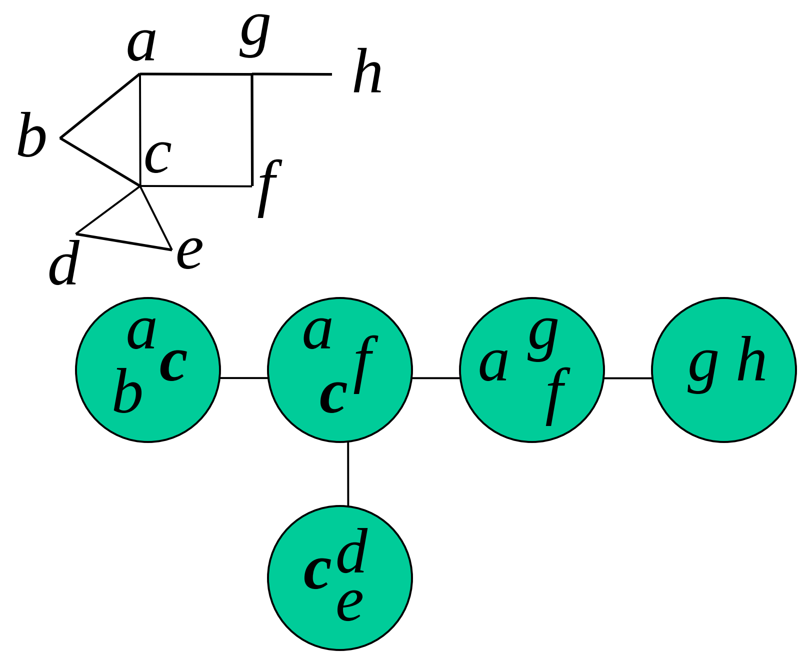
- Treewidth
- Courcelle's theorem and *finite state properties*
- Probabilistic networks
- Statement of main result
- Sketch of proof
- Conclusions and remarks

Treewidth

- In 198*, several authors discovers that many problems are solvable efficiently (linear time) when restricted to graphs with a certain *tree structure* (Bern, Lawler, Wong; Arnborg, Proskurowski; Wimer; Borie; Scheffler, Seese; Courcelle; Lautemann; …)
  - Robertson, Seymour: treewidth and tree decompositions (work on graph minors)
- Many (algorithmic and nonalgorithmic) applications

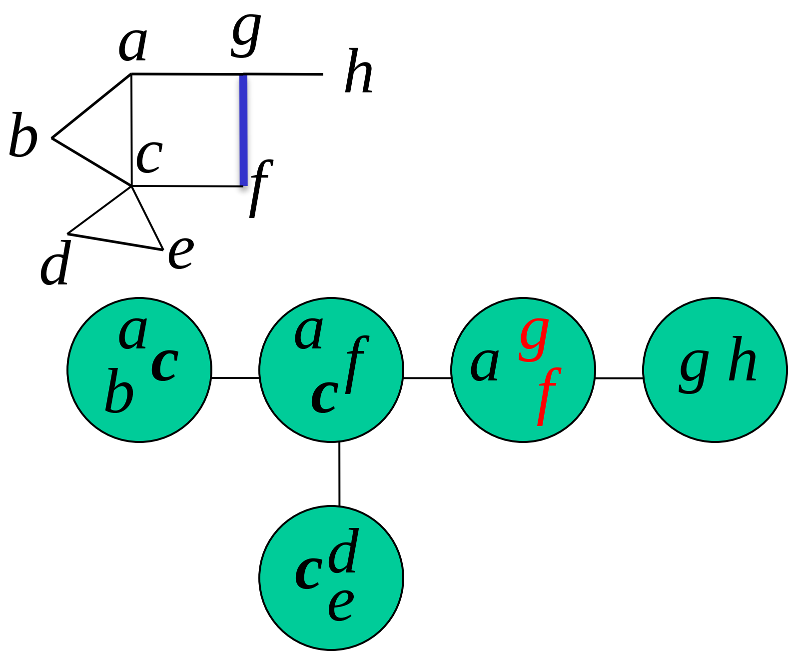
Tree decomposition

- A tree decomposition:
  - Tree with a vertex set called bag associated to every node.
  - For all edges {*v*,*w*}: there is a set containing both *v* and *w*.
  - For every vertex *v*: the nodes that contain *v* form a connected subtree.



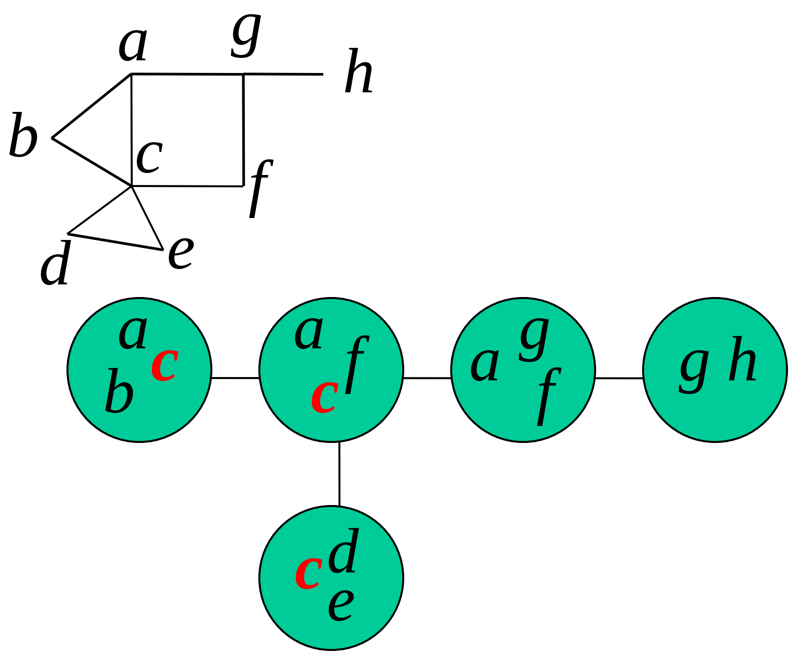
Tree decomposition

- A tree decomposition:
  - Tree with a vertex set called bag associated to every node.
  - For all edges {*v*,*w*}: there is a set containing both *v* and *w*.
  - For every vertex *v*: the nodes that contain *v* form a connected subtree.



Tree decomposition

- A tree decomposition:
  - Tree with a vertex set called bag associated to every node.
  - For all edges {*v*,*w*}: there is a set containing both *v* and *w*.
  - For every vertex *v*: the nodes that contain *v* form a connected subtree.

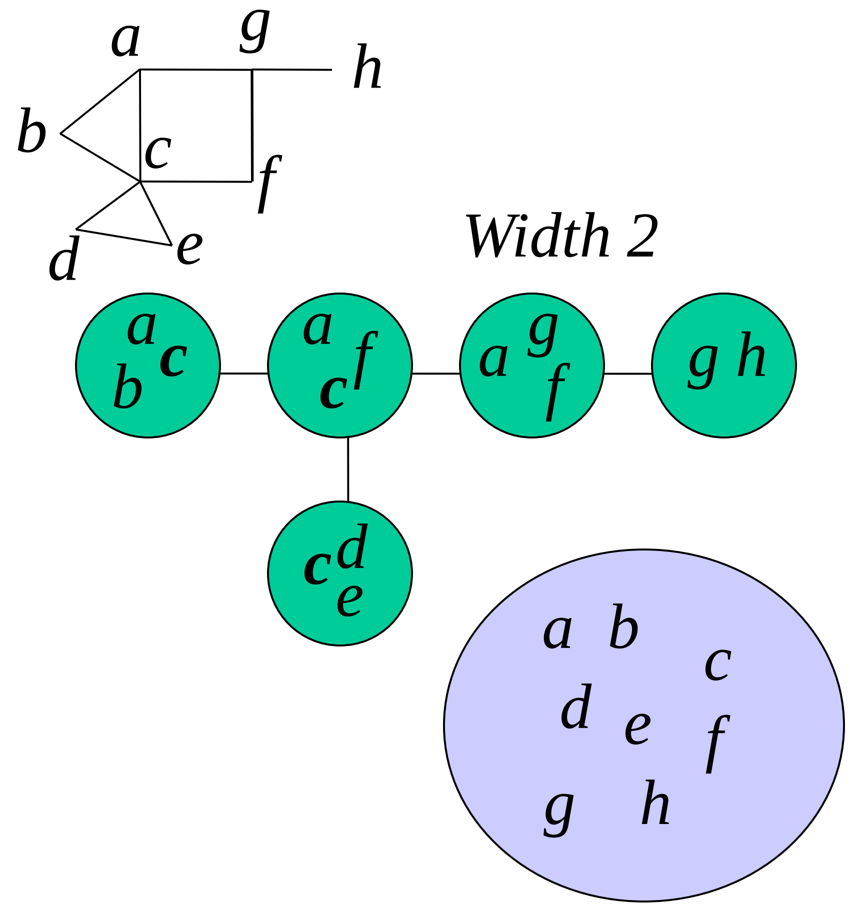


Treewidth (definition)

- Width of tree decomposition:

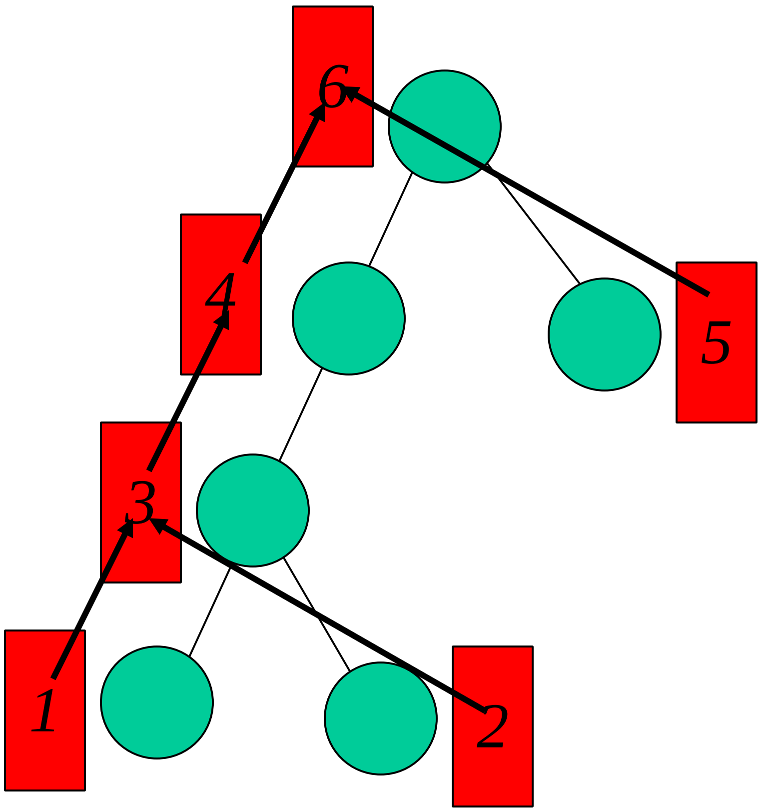
$$\max_{i \in I} |X_i| - 1$$

- Treewidth of graph G: tw(G)= minimum width over all tree decompositions of G.



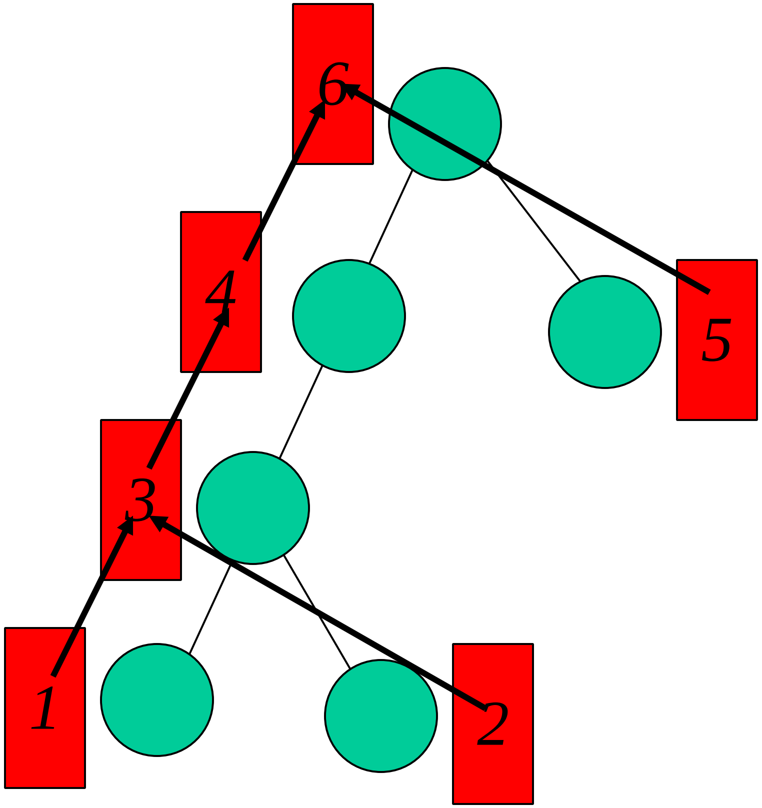
Many algorithmic results

- Many problems (including many NP-hard ones) become linear time solvable on graphs of bounded treewidth
  - Find tree decomposition of bounded width (O(*n*) time, B, 1993)
  - Run a dynamic programming algorithm bottom-up in the tree



Form of many DP algorithms on tree decompositions – Finite state

- For each bag, a table is computed:
  - Contains information on subgraph formed by vertices in bag and bags *below* it
  - Often limited information needed
- Computing a table needs only tables of children and local information
- If each table needs O(1) bits, we say the problem is ***finite state***



Courcelle’s theorem

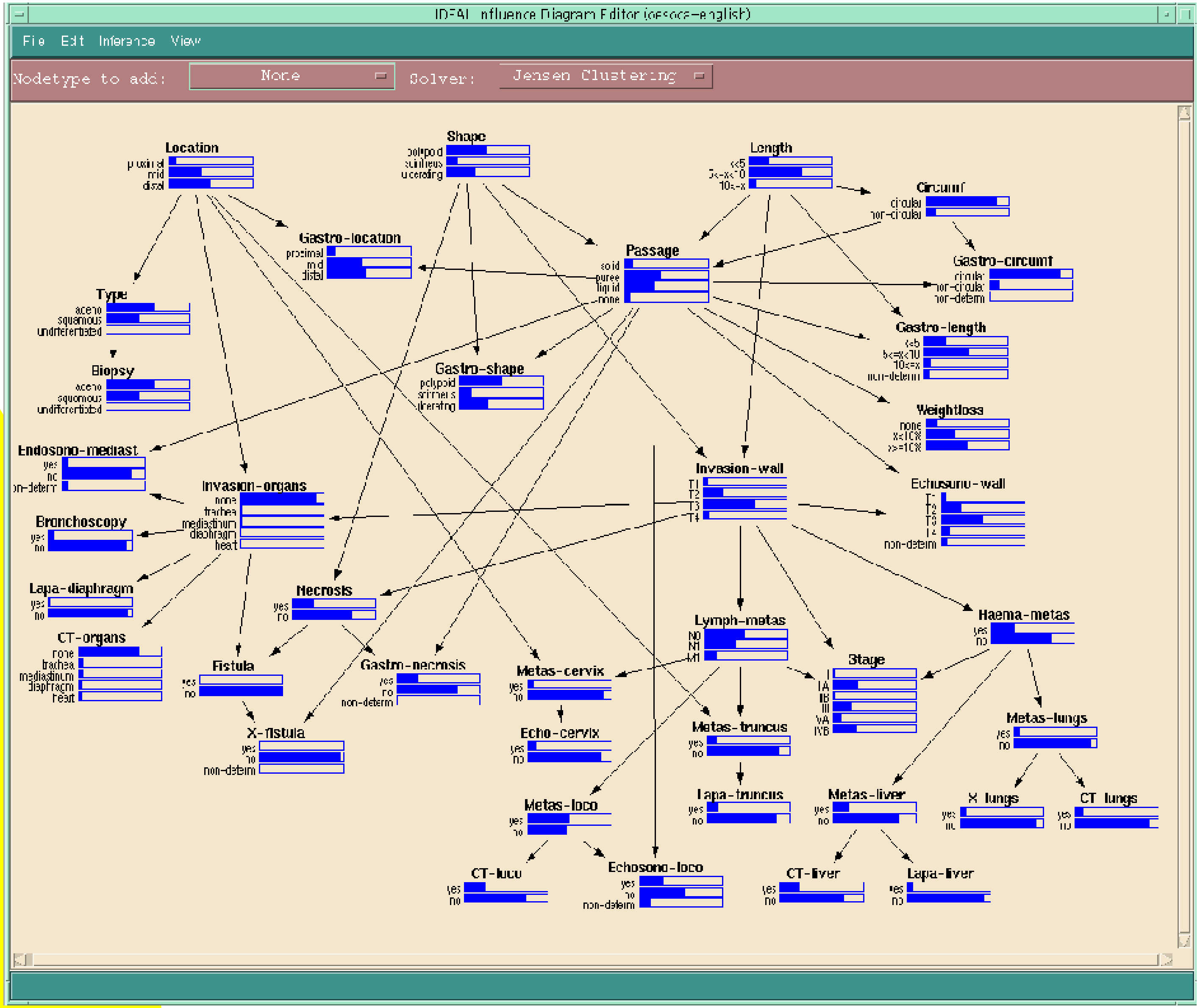
- Courcelle, 1990: Every graph problem that can be formulated in *Counting Monadic Second Order Logic* is finite state, and thus can be solved in linear time on graphs with bounded treewidth
  - MSOL: Language with constructions:
    - Quantification over vertices, edges, sets of vertices, sets of edges (for all vertex sets W there exists an edge e, such that ...)
    - Membership tests, adjacency tests ($\{v,w\}$ in E),
    - Logical operations (and, or, not, ...)
  - CMSOL: variant with extra operations |W| mod $a = b$, e.g., expresses that a set has odd size

Form of Coucelle's algorithm

- First find a tree decomposition of bounded width
- Dynamic programming algorithm on the tree decomposition; resembles *finite state tree automaton*:
  - Set of states S (actually tables with O(1) bits)
  - For each node in the tree decomposition, we compute a state
    - Bottom-up in the tree
    - State depends on states of children and local info from the node
  - State of root gives answer

Probabilistic networks

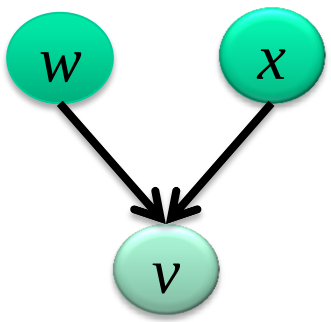
- Technology underlying many decision support systems
  - E.g., medical or vetinary diagnosis
- Representation of statistical variables and (in)dependencies by a graph



File Edit Inference View
Nodetype to add: None
Solver: Jensen Clustering
Location
Shape
Length
Circumf
Gastro-location
Passage
Gastro-circumf
Type
Biopsy
Gastro-shape
Gastro-length
Weightloss
Endosono-mediast
Invasion-organs
Invasion-wall
Echosono-wall
Bronchoscopy
Lapa-diaphragm
Necrosis
CT-organs
Fistula
Gastro-necrosis
Metas-cervix
Lymph-metas
Stage
Haema-metas
X-fistula
Echo-cervix
Metas-truncus
Metas-lungs
Lapa-truncus
Metas-liver
X-lungs
CT-lungs
Metas-loco
CT-loco
Echosono-loco
CT-liver
Lapa-liver

Probabilistic network

- Directed acyclic graph G=(V,E)
- For each vertex *v*, a set of possible values (here: True or False)
- For each vertex *v*, a conditional probability distribution of its values, given the values of its parents in G
  - Pr(*v* = true | w = true and x is true)
  - Pr(*v* = true | w = true and x is false)
  - Pr(*v* = true | w = false and x is true)
  - Pr(*v* = true | w = false and x is false)

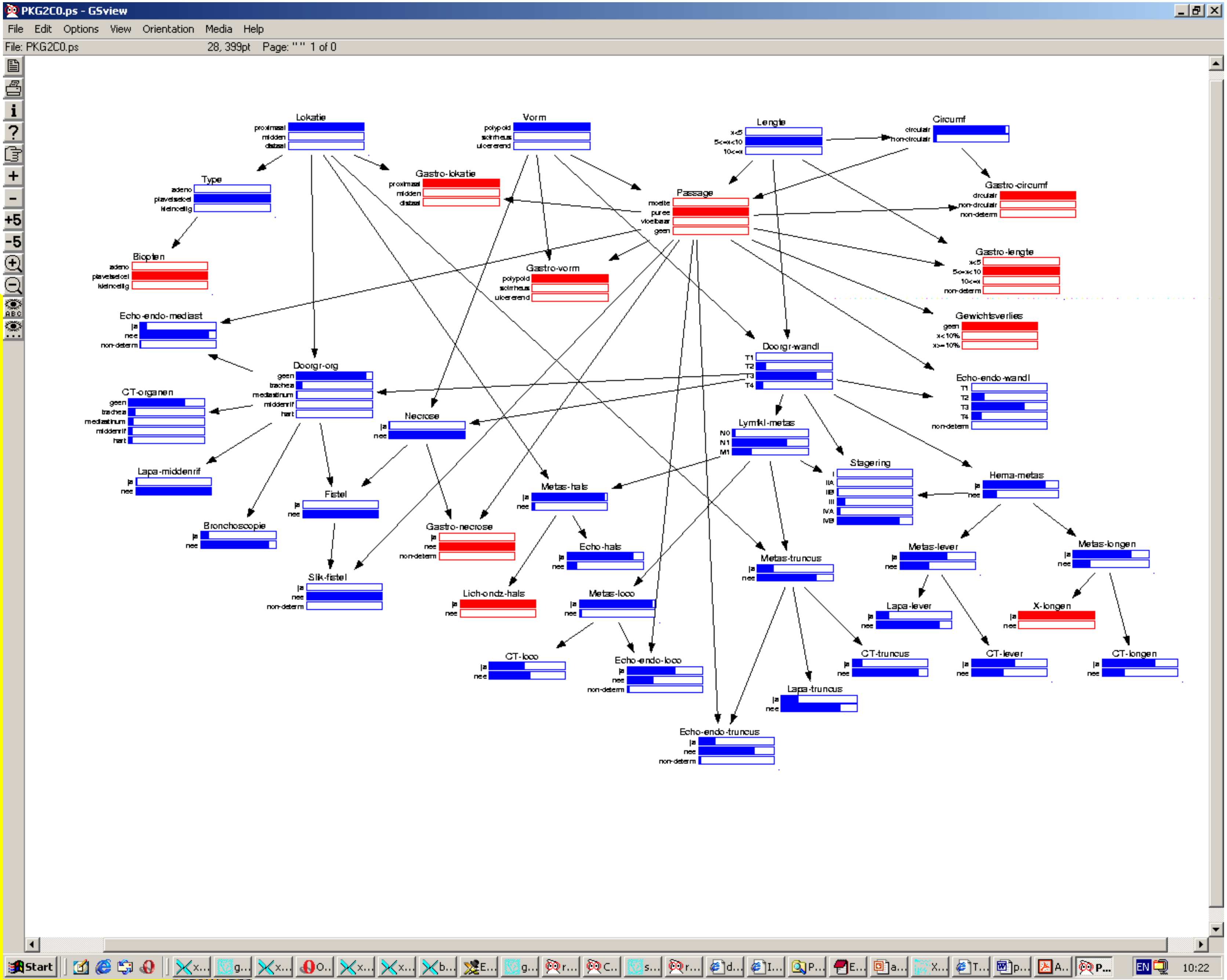


Inference

- The conditional probability distribution tables define a probability distribution over the variables
- Inference:
  - Given are values for some variables
  - Asked is probability distribution for another variable
  - E.g.: given information on a patient, what is the probability of some diagnosis?
- Probabilistic inference is #P-complete (Cooper, ...)

Some terminology

- Configuration of vertex/variable set W: c_W -> {true, false}
- The network defines a probability for each configuration

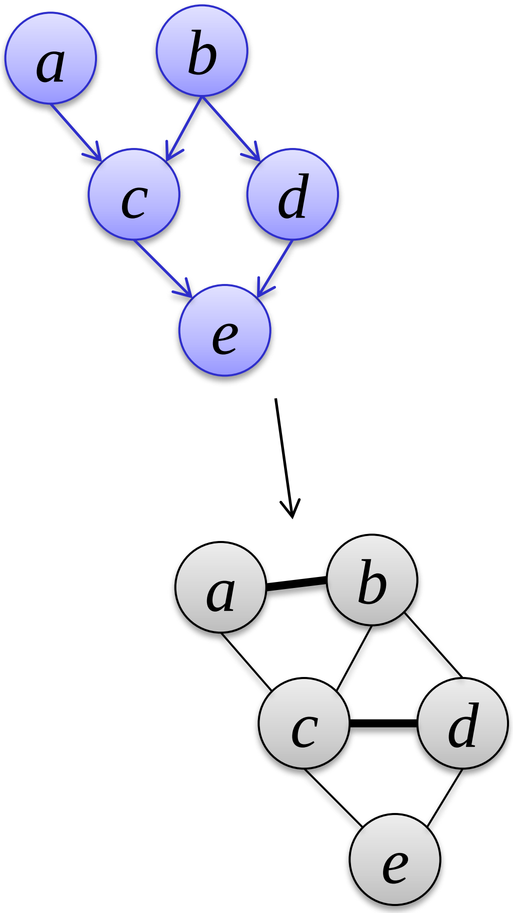

PKG2C0.ps - GSview
File Edit Options View Orientation Media Help
File: PKG2C0.ps 28, 399pt Page: "" 1 of 0
Lokatie
proximaal
midden
distaal
Vorm
polypoid
scirrheus
ulcererend
Lengte
x<5
5<=x<10
10<=x
Circumf
circulair
non-circulair
Type
adeno
plaveiselcel
kleincellig
Gastro-lokatie
proximaal
midden
distaal
Passage
moeite
puree
vloeibaar
geen
Gastro-circumf
circulair
non-circulair
non-determ
Biopten
adeno
plaveiselcel
kleincellig
Gastro-lengte
x<5
5<=x<10
10<=x
non-determ
Gastro-vorm
polypoid
scirrheus
ulcererend
Echo-endo-mediast
ja
nee
non-determ
Gewichtsverlies
geen
x<10%
x>=10%
Doorgr-wandl
T1
T2
T3
T4
Doorgr-org
geen
trachea
mediastinum
middenrif
hart
Echo-endo-wandl
T1
T2
T3
T4
non-determ
CT-organen
geen
trachea
mediastinum
middenrif
hart
Necrose
ja
nee
Lymfkl-metas
N0
N1
M1
Stagering
I
IIA
IIB
III
IVA
IVB
Lapa-middenrif
ja
nee
Fistel
ja
nee
Metas-hals
ja
nee
Hema-metas
ja
nee
Bronchoscopie
ja
nee
Gastro-necrose
ja
nee
non-determ
Echo-hals
ja
nee
Metas-truncus
ja
nee
Metas-lever
ja
nee
Metas-longen
ja
nee
Slik-fistel
ja
nee
non-determ
Lich-ondz-hals
ja
nee
Metas-loco
ja
nee
Lapa-lever
ja
nee
X-longen
ja
nee
CT-loco
ja
nee
Echo-endo-loco
ja
nee
non-determ
CT-truncus
ja
nee
CT-lever
ja
nee
CT-longen
ja
nee
Lapa-truncus
ja
nee
Echo-endo-truncus
ja
nee
non-determ
Start
10:22

Algorithms for Inference

- Several, including:
- Pearl’s algorithm: uses *loopcut set* (variant of *feedback vertex set*)
- Lauritzen-Spiegelhalter algorithm (1988 “*junction tree algorithm*”, “*treewidth of moralized graph*”)

Moralized graph

- The moralized graph of a dag G is the graph obtained by:
  - Adding an edge between each pair of vertices that share a child
  - Dropping directions



Lauritzen-Spiegelhalter result

- Let (G,Pr) be a probabilistic network such that the moralized graph of G has bounded treewidth. Then Inference can be solved in linear time.
  - Linear: exponential in treewidth, linear in number of vertices/variables

Result of this paper

- Given a probabilistic network
  - whose moralization has bounded treewidth and
- a property P of the network and the values of the variables that can be formulated in (counting) monadic second order logic,
  - CMSOL extended with operations:
    Value(*v*) = true ; Value(*v*) = false
- one can determine in linear time the probability that P holds.

Examples

- *Suggested causality*: for two variables *x* and *y*, what is the probability that *x* and *y* hold *and* there is a path from *x* to *y* with all variables on the path true?
- *Independence*: what is the probability that no true variables are adjacent?
- But also questions like: what is the probability that the subgraph formed by false variables is 3-colorable?

Sketch of proof

1. Make (using Courcelle's theorem) FSTA with states S, for the property of G and the value assignment cV: V -> {true, false}.
2. For each node *i* in the tree decomposition, we compute a value in [0,1] (*probability*) for each pair ($c_{X(i)}$,*s*) : $c_{X(i)}$ a configuration of the bag X(i) of *i*; and *s* a state
   - Vaguely, the probability that Courcelle's algorithm would be in state *s* for this node, but ...
   - We only take into account values of the conditional probability distributions for some vertices, namely
     - Those *v* with *v* and all parents in a bag that is a descendant of *i* or in X(*i*)

More precisely

- For a bag *i*
  - Let V(*i*) be the vertices in bag X(*i*) and all bags below *i*
  - Let Y(*i*) be all vertices *v* in V(*i*) such that all parents of *v* also belong to V(*i*)
- Compute for a state *s* and configuration $c_{x(i)}$ of X(*i*): sum over all configurations $c_{V(i)}$ of V_i that extend $c_{x(i)}$ of the value

$$\prod_{v \in Y(i)} \Pr(v \text{ as in } c_{Y(i)} \mid parents(v) \text{ as in } c_{Y(i)})$$

Algorithm

- Compute these tables bottom up
- Details are tedious, technical, but mainly just combine the ideas from the two algorithms of Lauritzen/Spiegelhalter and of Courcelle

Extensions

- Non-binary variables
- Edges and arcs with different roles
- Conditional probabilities (giving observations, etc.)

Discussion

- Running time: often not yet useful in this form ... Optimizations by recent work by Courcelle and Durand may help?
- Real applications?
- Open problems:
  - Other problems on probabilistic networks, e.g., finding configurations of maximum probability (MAP problem)
  - Optimization problems (finite integer index)
  - Is Inference on probabilistic networks with bounded treewidth of moralized graph in LOGSPACE?