Using imsets for learning Bayesian networks

Jiří Vomlel and Milan Studený

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic (AV ČR)

Liblice, September 15–18, 2007

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and
- person's gender, denoted by G.

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and
- person's gender, denoted by G.

We can describe relations between these three variables as follows:

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and
- person's gender, denoted by G.

We can describe relations between these three variables as follows:

- Seeing the length of hair of a person will tell us more about his/her gender and conversely. It means, the value of G is dependent on the value of H.

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and
- person's gender, denoted by G.

We can describe relations between these three variables as follows:

- Seeing the length of hair of a person will tell us more about his/her gender and conversely. It means, the value of G is dependent on the value of H.
- Knowing more about the gender will focus our belief on his/her stature - S is dependent on G and (through G) also on H.

Conditional Independence

Example (CI model)

Assume three variables:

- person's length of hair, denoted by H,
- person's stature, denoted by S, and
- person's gender, denoted by G.

We can describe relations between these three variables as follows:

- Seeing the length of hair of a person will tell us more about his/her gender and conversely. It means, the value of G is dependent on the value of H.
- Knowing more about the gender will focus our belief on his/her stature - S is dependent on G and (through G) also on H.
- Nevertheless, if we know the gender of a person then length of hair of that person gives us no extra clue on his/her stature - H is independent of S given G.

Conditional Independence Statements

Definition (CI statement)

Let A, B, C be pairwise disjoint subsets of a set of variables N. Then the statement "A is conditionally independent of B given C" is a CI statement (over N), written as $I(A, B, C)$.

Conditional Independence Statements

Definition (CI statement)

Let A, B, C be pairwise disjoint subsets of a set of variables N. Then the statement "A is conditionally independent of B given C" is a CI statement (over N), written as $I(A, B, C)$.

Example (CI statement)

In Example 1 we have indicated only one CI statement, $I(H, S, G)$. On the other hand, we have indicated two dependence statements, namely $\neg I(G, H) = \neg I(G, H, \emptyset)$ and $\neg I(S, G)$.

Conditional Independence (CI) model

Definition (CI in PDs)

Let P be a discrete probability distribution over N. Given any $A \subseteq N$, let

- $\mathbf{x}_A$ denote a configuration of values of variables $\mathbf{X}_A = \{X_i\}_{i \in A}$ and
- for $B \subseteq N \setminus A$ let $P(\mathbf{x}_A \mid \mathbf{x}_B)$ denote the conditional probability of $\mathbf{X}_A = \mathbf{x}_A$ given $\mathbf{X}_B = \mathbf{x}_B$.

The CI statement $I(A, B, C)$ is induced by probability distribution P over N if for all $\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C$ such that $P(\mathbf{x}_C) > 0$

$$P(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x}_C) \;=\; P(\mathbf{x}_A \mid \mathbf{x}_C) \cdot P(\mathbf{x}_B \mid \mathbf{x}_C) \;.$$

Conditional Independence (CI) model

Definition (CI in PDs)

Let P be a discrete probability distribution over N. Given any $A \subseteq N$, let

- $\mathbf{x}_A$ denote a configuration of values of variables $\mathbf{X}_A = \{X_i\}_{i \in A}$ and
- for $B \subseteq N \setminus A$ let $P(\mathbf{x}_A \mid \mathbf{x}_B)$ denote the conditional probability of $\mathbf{X}_A = \mathbf{x}_A$ given $\mathbf{X}_B = \mathbf{x}_B$.

The CI statement $I(A, B, C)$ is induced by probability distribution P over N if for all $\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C$ such that $P(\mathbf{x}_C) > 0$

$$P(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x}_C) \;=\; P(\mathbf{x}_A \mid \mathbf{x}_C) \cdot P(\mathbf{x}_B \mid \mathbf{x}_C) \;.$$

Example

In Example 1 we have indicated one CI statement - $I(H, S, G)$. For all values h, s, g of variables H, S, G it holds that

$P(h, s \mid g) \;=\; P(h \mid g) \cdot P(s \mid g)$ or, equivalently

Conditional Independence (CI) model

Definition (CI in PDs)

Let P be a discrete probability distribution over N. Given any $A \subseteq N$, let

- $\mathbf{x}_A$ denote a configuration of values of variables $\mathbf{X}_A = \{X_i\}_{i \in A}$ and
- for $B \subseteq N \setminus A$ let $P(\mathbf{x}_A \mid \mathbf{x}_B)$ denote the conditional probability of $\mathbf{X}_A = \mathbf{x}_A$ given $\mathbf{X}_B = \mathbf{x}_B$.

The CI statement $I(A, B, C)$ is induced by probability distribution P over N if for all $\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C$ such that $P(\mathbf{x}_C) > 0$

$$P(\mathbf{x}_A, \mathbf{x}_B \mid \mathbf{x}_C) \;=\; P(\mathbf{x}_A \mid \mathbf{x}_C) \cdot P(\mathbf{x}_B \mid \mathbf{x}_C) \;.$$

Example

In Example 1 we have indicated one CI statement - $I(H, S, G)$. For all values h, s, g of variables H, S, G it holds that

$P(h, s \mid g) \;=\; P(h \mid g) \cdot P(s \mid g)$ or, equivalently $P(h \mid g, s) \;=\; P(h \mid g)$

What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

- the path contains a node $c \in C$, in which edges **do not meet "head-to-head"** or

What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

- the path contains a node $c \in C$, in which edges **do not meet "head-to-head"** or
- the path contains a node c in which edges **meet "head-to-head"** and neither c nor any of its descendants belong to C.

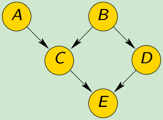
What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

- the path contains a node $c \in C$, in which edges **do not meet "head-to-head"** or
- the path contains a node c in which edges **meet "head-to-head"** and neither c nor any of its descendants belong to C.

Example



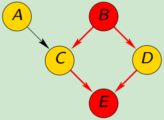
What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

- the path contains a node $c \in C$, in which edges **do not meet "head-to-head"** or
- the path contains a node c in which edges **meet "head-to-head"** and neither c nor any of its descendants belong to C.

Example



$I(A, D, B)$

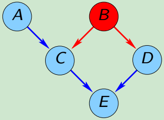
What CI-statements are represented by a DAG?

Definition (d-separation criteria)

Two nodes a and b in a DAG G are d-separated by a set C if for all paths between a and b there is a node c ($c \neq a$ and $c \neq b$) such that either:

- the path contains a node $c \in C$, in which edges **do not meet "head-to-head"** or
- the path contains a node c in which edges **meet "head-to-head"** and neither c nor any of its descendants belong to C.

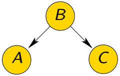
Example

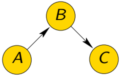


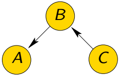
$$\neg I(A, D, \{B, E\})$$

Equivalence classes of Bayesian networks

What independence statements are represented by these three models?

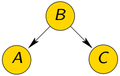




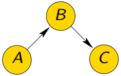


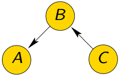
Equivalence classes of Bayesian networks

What independence statements are represented by these three models?



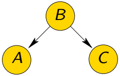
$I(A, C, B)$



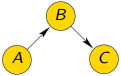


Equivalence classes of Bayesian networks

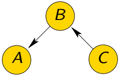
What independence statements are represented by these three models?



$I(A, C, B)$

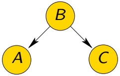


$I(A, C, B)$

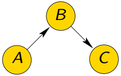


Equivalence classes of Bayesian networks

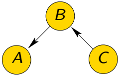
What independence statements are represented by these three models?



$I(A, C, B)$



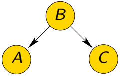
$I(A, C, B)$



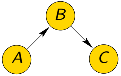
$I(A, C, B)$

Equivalence classes of Bayesian networks

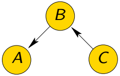
What independence statements are represented by these three models?



$I(A, C, B)$



$I(A, C, B)$

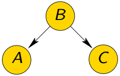


$I(A, C, B)$

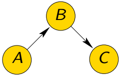
Different graphs may represent the same set of CI-statements!

Equivalence classes of Bayesian networks

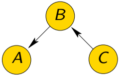
What independence statements are represented by these three models?



$I(A, C, B)$



$I(A, C, B)$



$I(A, C, B)$

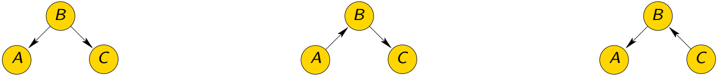
Different graphs may represent the same set of CI-statements!

Definition (Equivalence class)

We say that Bayesian networks with DAGs representing the same set of CI-statements belong to an equivalence class.

Equivalence classes of Bayesian networks

What independence statements are represented by these three models?



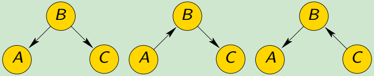
$I(A, C, B)$ $\quad$ $I(A, C, B)$ $\quad$ $I(A, C, B)$

Different graphs may represent the same set of CI-statements!

Definition (Equivalence class)

We say that Bayesian networks with DAGs representing the same set of CI-statements belong to an equivalence class.

Example



belong to the same equivalence class.

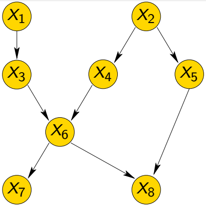
Equivalence classes of Bayesian networks

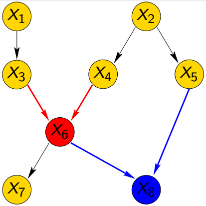
Definition (Immorality)

An immorality in a DAG G is an induced subgraph of G for a set $\{A, B, C\}$, where A, B, C are distinct nodes of G such that there are edges $A \rightarrow C$ and $B \rightarrow C$ and there is no edge between A and B in G.

Definition (Immorality)

An immorality in a DAG G is an induced subgraph of G for a set $\{A, B, C\}$, where A, B, C are distinct nodes of G such that there are edges $A \rightarrow C$ and $B \rightarrow C$ and there is no edge between A and B in G.





Equivalence classes of Bayesian networks

Definition (Immorality)

An immorality in a DAG G is an induced subgraph of G for a set $\{A, B, C\}$, where A, B, C are distinct nodes of G such that there are edges $A \rightarrow C$ and $B \rightarrow C$ and there is no edge between A and B in G.

Definition (Underlaying graph)

An underlaying graph of a DAG is the undirected graph that has the same set of nodes and all directed edges $A \rightarrow B$ are replaced by undirected edges $A - B$.

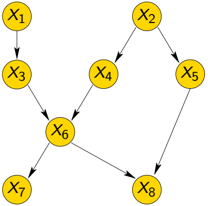
Equivalence classes of Bayesian networks

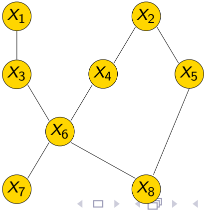
Definition (Immorality)

An immorality in a DAG G is an induced subgraph of G for a set $\{A, B, C\}$, where A, B, C are distinct nodes of G such that there are edges $A \rightarrow C$ and $B \rightarrow C$ and there is no edge between A and B in G.

Definition (Underlaying graph)

An underlaying graph of a DAG is the undirected graph that has the same set of nodes and all directed edges $A \rightarrow B$ are replaced by undirected edges $A - B$.





Equivalence classes of Bayesian networks

Definition (Immorality)

An immorality in a DAG G is an induced subgraph of G for a set $\{A, B, C\}$, where A, B, C are distinct nodes of G such that there are edges $A \to C$ and $B \to C$ and there is no edge between A and B in G.

Definition (Underlaying graph)

An underlaying graph of a DAG is the undirected graph that has the same set of nodes and all directed edges $A \to B$ are replaced by undirected edges $A - B$.

Theorem

Bayesian networks belong to the same equivalence class iff they have the same underlaying graph and the same set of immoralities.

Essential graphs

Definition (Essential graph)

The essential graph G^* of an equivalence class $\mathcal{G}$ of DAGs over N is a hybrid graph over N defined as follows:

- $a \to b$ in G^* if $a \to b$ in G for every $G \in \mathcal{G}$,

Essential graphs

Definition (Essential graph)

The essential graph G^* of an equivalence class $\mathcal{G}$ of DAGs over N is a hybrid graph over N defined as follows:

- $a \rightarrow b$ in G^* if $a \rightarrow b$ in G for every $G \in \mathcal{G}$,
- $a - b$ in G^* if $\exists\, G_1, G_2 \in \mathcal{G}$ such that $a \rightarrow b$ in G_1 and $a \leftarrow b$ in G_2.

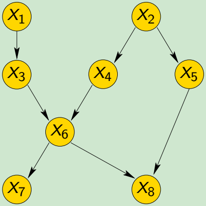
Essential graphs

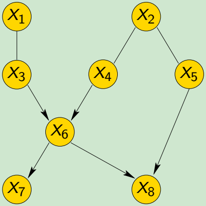
Definition (Essential graph)

The essential graph G^* of an equivalence class $\mathcal{G}$ of DAGs over N is a hybrid graph over N defined as follows:

- $a \longrightarrow b$ in G^* if $a \longrightarrow b$ in G for every $G \in \mathcal{G}$,
- $a - b$ in G^* if $\exists\, G_1, G_2 \in \mathcal{G}$ such that $a \longrightarrow b$ in G_1 and $a \longleftarrow b$ in G_2.

Example





What is an imset? (formal definition)

- N ... a finite set
- $\mathcal{P}(N)$ ... power set of N
- $\mathbb{Z}$ ... set of all integers

What is an imset? (formal definition)

- N ... a finite set
- $\mathcal{P}(N)$ ... power set of N
- $\mathbb{Z}$ ... set of all integers

Definition

Imset u is a function $u : \mathcal{P}(N) \mapsto \mathbb{Z}$.

What is an imset? (formal definition)

- N ... a finite set
- $\mathcal{P}(N)$ ... power set of N
- $\mathbb{Z}$ ... set of all integers

Definition

Imset u is a function $u : \mathcal{P}(N) \mapsto \mathbb{Z}$.

Function $m : \mathcal{P}(N) \mapsto \mathbb{N}$ is sometimes called multiset. Thus, imset is an abbreviation from **I**nteger valued **M**ulti**SET**. Studený (2001)

What is an imset? (an example)

Let $N = \{a, b, c\}$. An imset u over N is

$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a, b\}$	$\{b, c\}$	$\{a, c\}$	$\{a, b, c\}$
0	0	+1	0	-1	-1	0	+1

What is an imset? (an example)

Let $N = \{a, b, c\}$. An imset u over N is

$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{b,c\}$	$\{a,c\}$	$\{a,b,c\}$
0	0	+1	0	-1	-1	0	+1

A convention:

$$\delta_A(B) = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{otherwise} \end{cases}$$

What is an imset? (an example)

Let $N = \{a, b, c\}$. An imset u over N is

$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a, b\}$	$\{b, c\}$	$\{a, c\}$	$\{a, b, c\}$
0	0	+1	0	-1	-1	0	+1

A convention:

$$\delta_A(B) = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{otherwise} \end{cases}$$

$$\forall B \subseteq N: \ u(B) = \sum_{A \subseteq N} u(A) \cdot \delta_A(B)$$

What is an imset? (an example)

Let $N = \{a, b, c\}$. An imset u over N is

$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{b,c\}$	$\{a,c\}$	$\{a,b,c\}$
0	0	+1	0	-1	-1	0	+1

A convention:

$$\delta_A(B) = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{otherwise} \end{cases}$$

$$\forall B \subseteq N: \; u(B) = \sum_{A \subseteq N} u(A) \cdot \delta_A(B)$$

$$u = \sum_{A \subseteq N} c_A \cdot \delta_A$$

What is an imset? (an example)

Let $N = \{a, b, c\}$. An imset u over N is

$\emptyset$	$\{a\}$	$\{b\}$	$\{c\}$	$\{a,b\}$	$\{b,c\}$	$\{a,c\}$	$\{a,b,c\}$
0	0	+1	0	-1	-1	0	+1

A convention:

$$\delta_A(B) = \begin{cases} 1 & \text{if } A = B \\ 0 & \text{otherwise} \end{cases}$$

$$\forall B \subseteq N: \ u(B) = \sum_{A \subseteq N} u(A) \cdot \delta_A(B)$$

$$u = \sum_{A \subseteq N} c_A \cdot \delta_A$$

Using the convention we will write

$$u = \delta_{\{b\}} - \delta_{\{a,b\}} - \delta_{\{b,c\}} + \delta_{\{a,b,c\}}$$

Elementary imset

Definition (Elementary imset)

Let $K \subseteq N$, $a, b \in N \setminus K$, and $a \neq b$.
Elementary imset is defined by the formula

$$u_{\langle a,b|K\rangle} \quad = \quad \delta_{\{a,b\}\cup K} + \delta_K - \delta_{\{a\}\cup K} - \delta_{\{b\}\cup K}$$

Elementary imset

Definition (Elementary imset)

Let $K \subseteq N$, $a, b \in N \setminus K$, and $a \neq b$.
Elementary imset is defined by the formula

$$u_{\langle a,b|K\rangle} \quad = \quad \delta_{\{a,b\}\cup K} + \delta_K - \delta_{\{a\}\cup K} - \delta_{\{b\}\cup K}$$

Definition

The standard imset for a DAG G is given by the formula

$$u_G \quad = \quad \delta_N - \delta_\emptyset + \sum_{a \in N} \left\{ \delta_{pa_G(a)} - \delta_{\{a\}\cup pa_G(a)} \right\} \quad .$$

Elementary imset

Definition (Elementary imset)

Let $K \subseteq N$, $a, b \in N \setminus K$, and $a \neq b$.
Elementary imset is defined by the formula

$$u_{\langle a,b|K\rangle} \quad = \quad \delta_{\{a,b\}\cup K} + \delta_K - \delta_{\{a\}\cup K} - \delta_{\{b\}\cup K}$$

Definition

The standard imset for a DAG G is given by the formula

$$u_G \quad = \quad \delta_N - \delta_\emptyset + \sum_{a\in N} \left\{ \delta_{pa_G(a)} - \delta_{\{a\}\cup pa_G(a)} \right\} \quad .$$

Standard imset is another uniquely determined representative of an equivalence class of Bayesian networks.

Inclusion neigbourhood

$\mathcal{M}_G$ will denote the set of CI-statements generated by a DAG G.

Inclusion neigbourhood

$\mathcal{M}_G$ will denote the set of CI-statements generated by a DAG G.

Definition

Given two DAGs K, L over N, we say that they are inclusion neighbors and write $\mathcal{M}_K \sqsubset \mathcal{M}_L$ if

- $\mathcal{M}_K \subset \mathcal{M}_L$ and

Inclusion neigbourhood

$\mathcal{M}_G$ will denote the set of CI-statements generated by a DAG G.

Definition

Given two DAGs K, L over N, we say that they are inclusion neighbors and write $\mathcal{M}_K \sqsubset \mathcal{M}_L$ if

- $\mathcal{M}_K \subset \mathcal{M}_L$ and
- there is no DAG G such that $\mathcal{M}_K \subset \mathcal{M}_G \subset \mathcal{M}_L$.

Inclusion neigbourhood

$\mathcal{M}_G$ will denote the set of CI-statements generated by a DAG G.

Definition

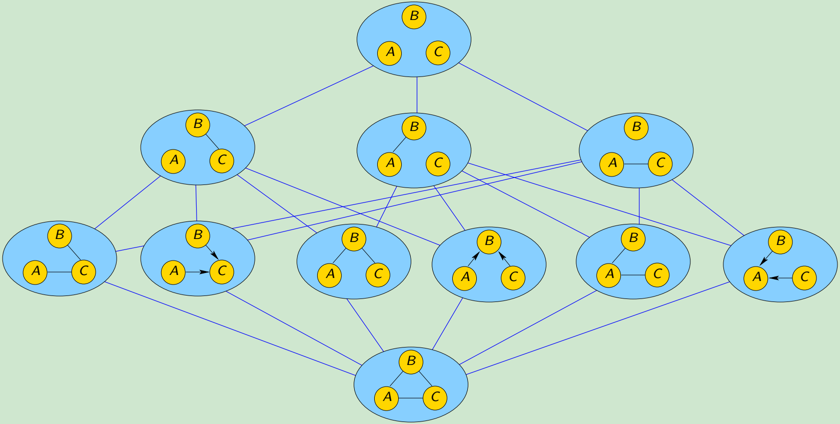
Given two DAGs K, L over N, we say that they are inclusion neighbors and write $\mathcal{M}_K \sqsubset \mathcal{M}_L$ if

- $\mathcal{M}_K \subset \mathcal{M}_L$ and
- there is no DAG G such that $\mathcal{M}_K \subset \mathcal{M}_G \subset \mathcal{M}_L$.

We say that $\mathcal{M}_L$ is an upper neighbor of $\mathcal{M}_K$ or, dually, that $\mathcal{M}_K$ is a lower neighbor of $\mathcal{M}_L$.

Search space for models of three variables

Example



Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,

Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,
- $X_i, i = 1, \ldots, N$ be the variables and nodes of the graph G of Bayesian network,

Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,
- $X_i, i = 1, \ldots, N$ be the variables and nodes of the graph G of Bayesian network,
- $r(i)$ denote number of states of variable X_i,

Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,
- $X_i, i = 1, \ldots, N$ be the variables and nodes of the graph G of Bayesian network,
- $r(i)$ denote number of states of variable X_i,
- $q(i, G)$ denote number of parent configurations for parents $\mathbf{X}_{pa(i)}$ of variable X_i, and

Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,
- $X_i, i = 1, \ldots, N$ be the variables and nodes of the graph G of Bayesian network,
- $r(i)$ denote number of states of variable X_i,
- $q(i, G)$ denote number of parent configurations for parents $\mathbf{X}_{pa(i)}$ of variable X_i, and
- $N(i, j, k)$ denote ocurrance of the corresponding configuration in the learning dataset D.

Likelihood of data

- Let $D = \{\mathbf{x}^m, m = 1, \ldots, M\}$ be the learning dataset, where $\mathbf{x}$ is the vector of values of variable $\mathbf{X} = \{X_i\}_{i=1}^N$,
- $X_i, i = 1, \ldots, N$ be the variables and nodes of the graph G of Bayesian network,
- $r(i)$ denote number of states of variable X_i,
- $q(i, G)$ denote number of parent configurations for parents $\mathbf{X}_{pa(i)}$ of variable X_i, and
- $N(i, j, k)$ denote ocurrance of the corresponding configuration in the learning dataset D.

The likelihood of D given G is the probability of data D being generated from the Bayesian network model with the structure given by directed acyclic graph G and representing joint probability distribution P is

$$P(D|G) = \prod_{m=1}^{M} P(\mathbf{X} = \mathbf{x}^m)$$

Scores

Lemma (Maximum loglikelihood)

The maximum log-likelihood for a given Bayesian network with graph G is

$$MLL(G|D) = \sum_{i=1}^{N} \sum_{k=1}^{r(i)} \sum_{j=1}^{q(i,G)} N(i,j,k) \log \frac{N(i,j,k)}{N(i,j)}$$

Lemma (Maximum loglikelihood)

The maximum log-likelihood for a given Bayesian network with graph G is

$$MLL(G|D) \;=\; \sum_{i=1}^{N}\sum_{k=1}^{r(i)}\sum_{j=1}^{q(i,G)} N(i,j,k)\log\frac{N(i,j,k)}{N(i,j)}$$

Let $d(G)$ be the number of free parameters in the Bayesian network model with graph G. It is given by $d(G) = \sum_{i=1}^{N}(r(i)-1)q(i,G)$.

Scores

Lemma (Maximum loglikelihood)

The maximum log-likelihood for a given Bayesian network with graph G is

$$MLL(G|D) = \sum_{i=1}^{N} \sum_{k=1}^{r(i)} \sum_{j=1}^{q(i,G)} N(i,j,k) \log \frac{N(i,j,k)}{N(i,j)}$$

Let $d(G)$ be the number of free parameters in the Bayesian network model with graph G. It is given by $d(G) = \sum_{i=1}^{N}(r(i) - 1)q(i,G)$.

Definition (Bayesian Information Criterion)

$$BIC(G|D) = MLL(G|D) - \frac{\log M}{2} d(G)$$

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

In each step of the GES algorithm:

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

In each step of the GES algorithm:

- we search only in the inclusion neigborhood,

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

In each step of the GES algorithm:

- we search only in the inclusion neigborhood,
- select the model that maximizes the criteria, and

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

In each step of the GES algorithm:

- we search only in the inclusion neigborhood,
- select the model that maximizes the criteria, and
- if there is no better model than the current one we start the second stage or terminate if we are in the second stage,

Greedy Equivalence Search (GES) algorithm

The GES algorithm starts with an empty graph and has two stages:

1. deleting CI-statements (which corresponds to edge additions and directing some edges)
2. adding CI-statements (which corresponds to edge removal and undirecting some edges).

In each step of the GES algorithm:

- we search only in the inclusion neigborhood,
- select the model that maximizes the criteria, and
- if there is no better model than the current one we start the second stage or terminate if we are in the second stage,

Theorem

In the limit of large datasets, if the CI-statements that hold in the dataset are exactly those of a Bayesian network then the algorithm terminates in the essential graph of this Bayesian network.

Comparison of Chickering's GES with our imset version

Chickering's version of GES:

Our imset version of GES:

Comparison of Chickering's GES with our imset version

Chickering's version of GES:

- Two stages:
  (1) edge additions (+ directing some edges) and
  (2) edge removal (+ undirecting some edges).

Our imset version of GES:

- Two stages:
  (1) elementary CI statements (their corresponding imsets) are subtracted and
  (2) elementary CI statements (their corresponding imsets) are added.

Comparison of Chickering's GES with our imset version

Chickering's version of GES:

- Two stages:
  (1) edge additions (+ directing some edges) and
  (2) edge removal (+ undirecting some edges).
- It requires a test of validity of resulting structures.

Our imset version of GES:

- Two stages:
  (1) elementary CI statements (their corresponding imsets) are subtracted and
  (2) elementary CI statements (their corresponding imsets) are added.
- This allows a very easy recomputation of the criteria (two values of a data imset are added and two subtracted).

Comparison of Chickering's GES with our imset version

Chickering's version of GES:

- Two stages:
  (1) edge additions (+ directing some edges) and
  (2) edge removal (+ undirecting some edges).
- It requires a test of validity of resulting structures.
- Some models from the same equivalence class are generated twice.

Our imset version of GES:

- Two stages:
  (1) elementary CI statements (their corresponding imsets) are subtracted and
  (2) elementary CI statements (their corresponding imsets) are added.
- This allows a very easy recomputation of the criteria (two values of a data imset are added and two subtracted).
- From each equivalence class no more than one model is generated.

Experiments with the imset version of GES

We implemented the imset version of GES in R.

The code is freely available from: http://www.utia.cz/vomlel/imset

Experiments with the imset version of GES

We implemented the imset version of GES in R.

The code is freely available from: `http://www.utia.cz/vomlel/imset`

Model	# variables	# selected nb.	# evaluated nb.
abcde - lower nb.	5	4	60
- upper nb.		0	4
asia - lower nb.	8	8	389
- upper nb.		0	8
abcedefghi - lower nb.	9	10	621
- upper nb.		0	10
boerlage92 - lower nb.	23	35	19607
- upper nb.		0	47
alarm - lower nb.	37	52	65787
- upper nb.		1	128

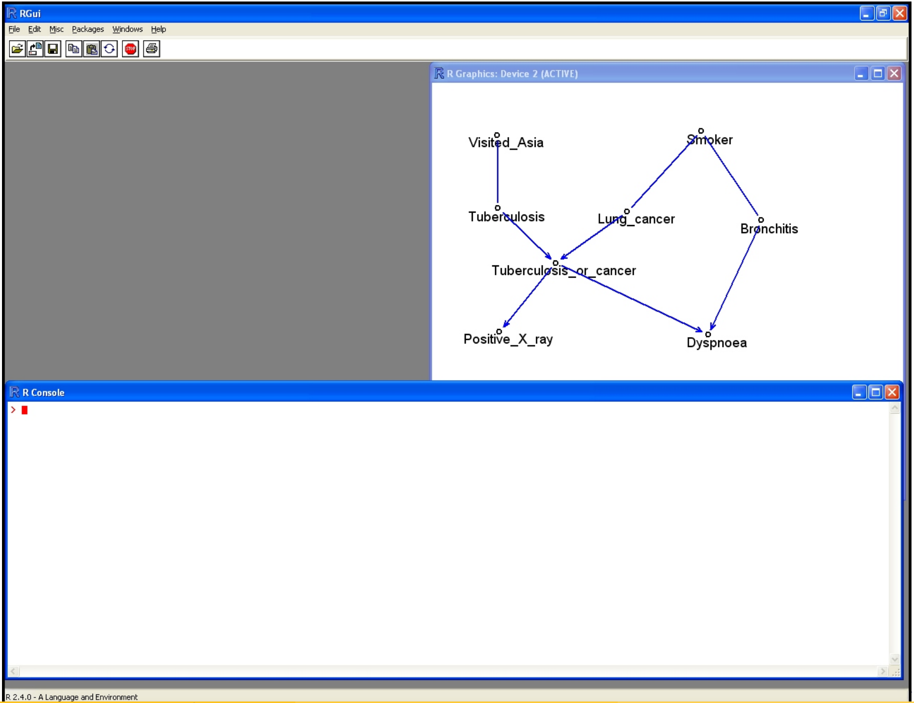
Experiments with the imset version of GES

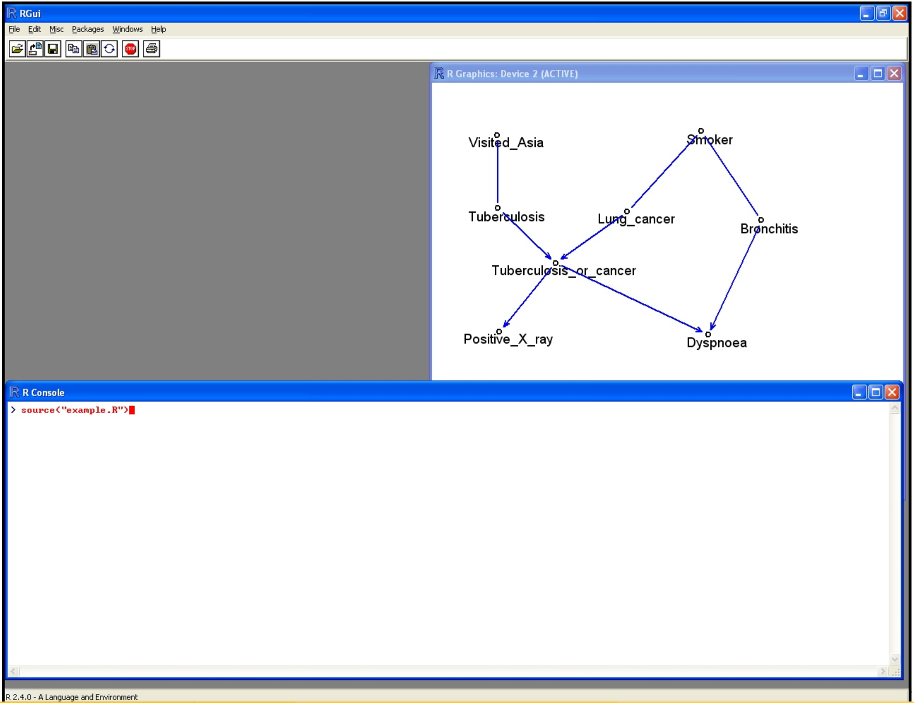
We implemented the imset version of GES in R.

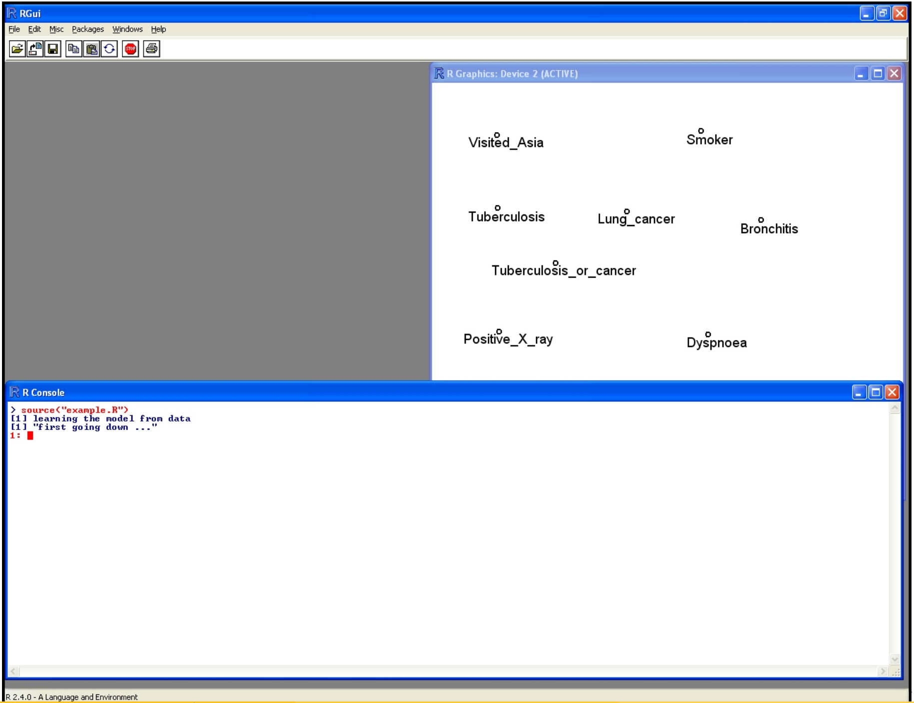
The code is freely available from: `http://www.utia.cz/vomlel/imset`

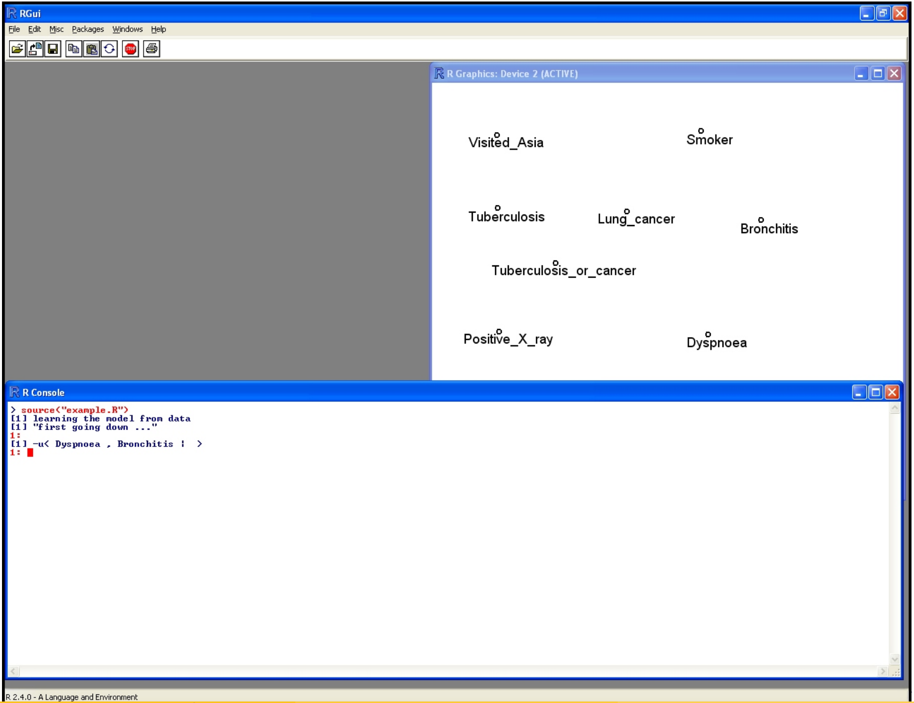
Model	# variables	# selected nb.	# evaluated nb.
abcde - lower nb.	5	4	60
- upper nb.		0	4
asia - lower nb.	8	8	389
- upper nb.		0	8
abcedefghi - lower nb.	9	10	621
- upper nb.		0	10
boerlage92 - lower nb.	23	35	19607
- upper nb.		0	47
alarm - lower nb.	37	52	65787
- upper nb.		1	128

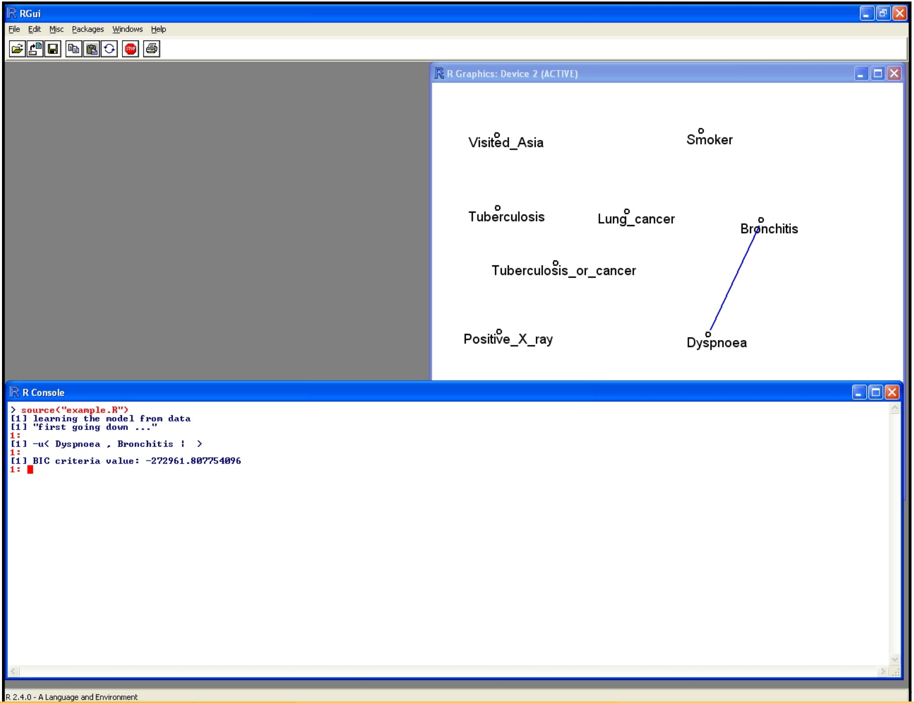
For the alarm model there are more than $2^{\frac{37 \cdot 36}{2}} = 2^{666}$ essential graphs.

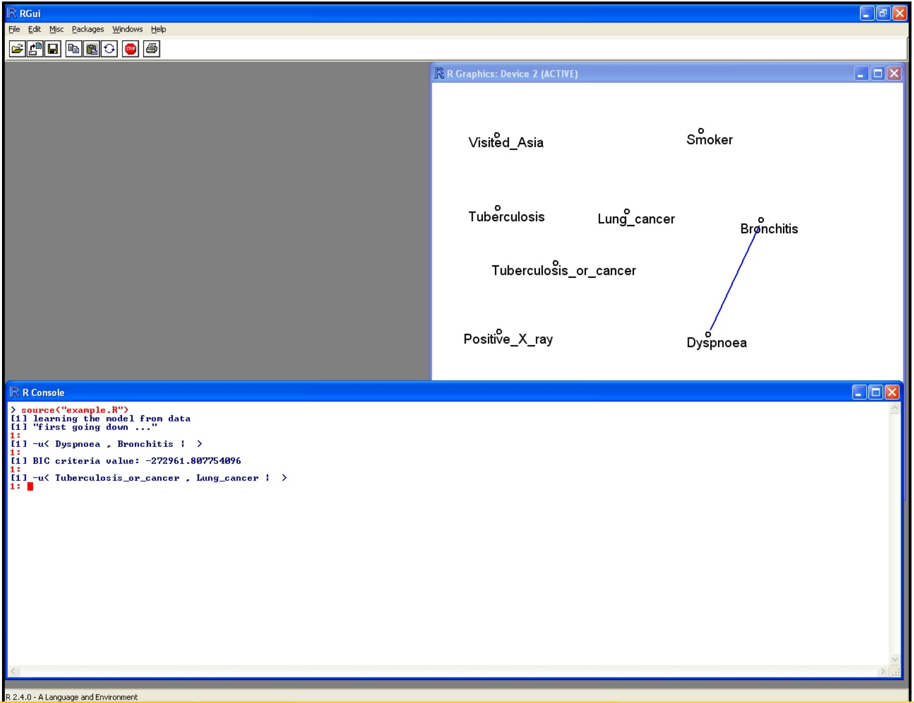

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
R 2.4.0 - A Language and Environment

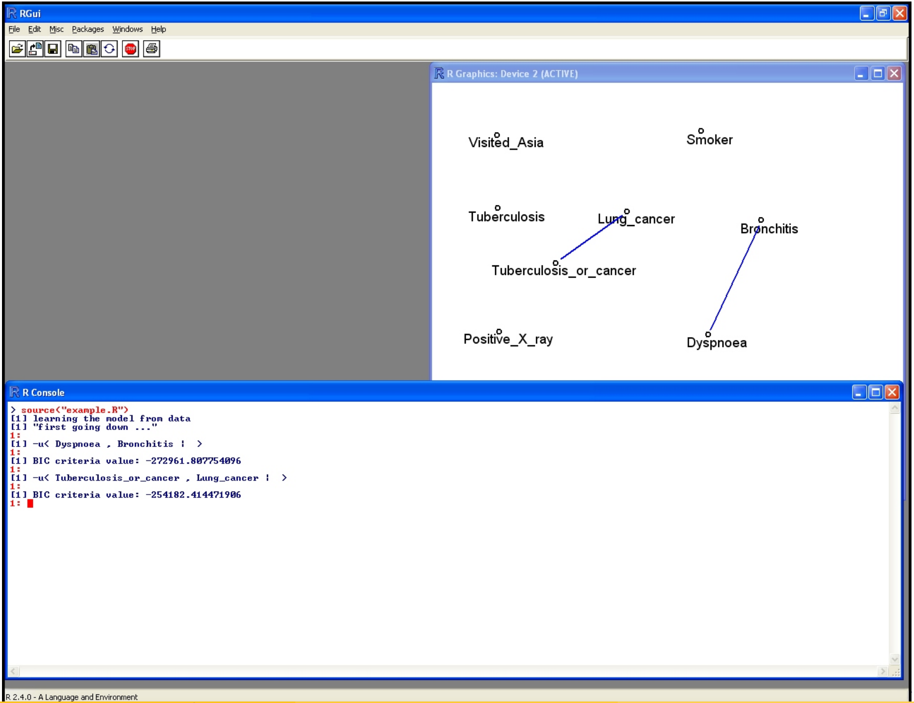

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
R 2.4.0 - A Language and Environment

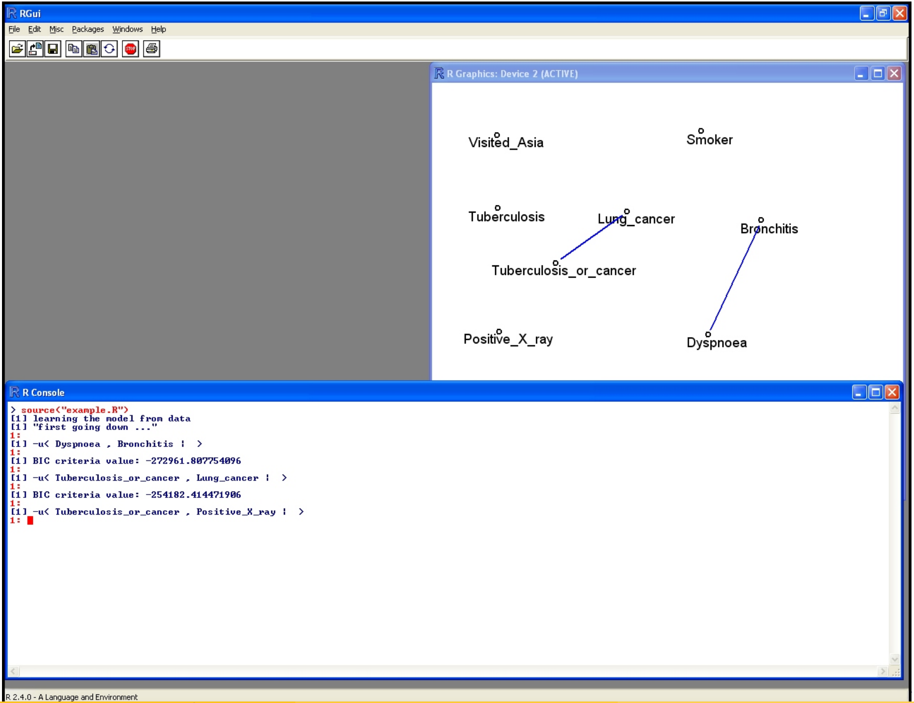

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
R 2.4.0 - A Language and Environment

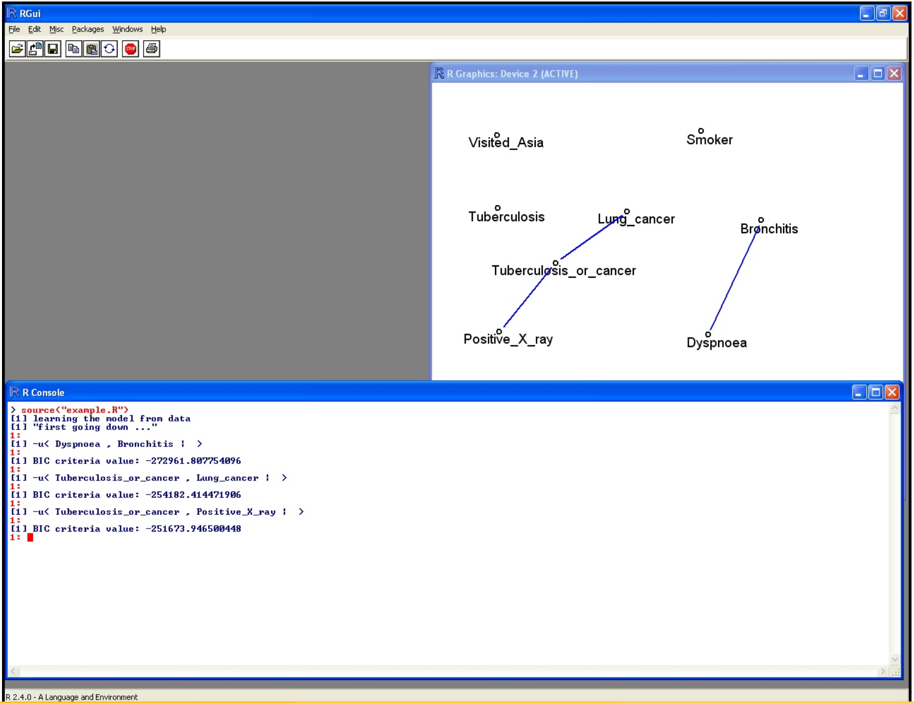

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
R 2.4.0 - A Language and Environment

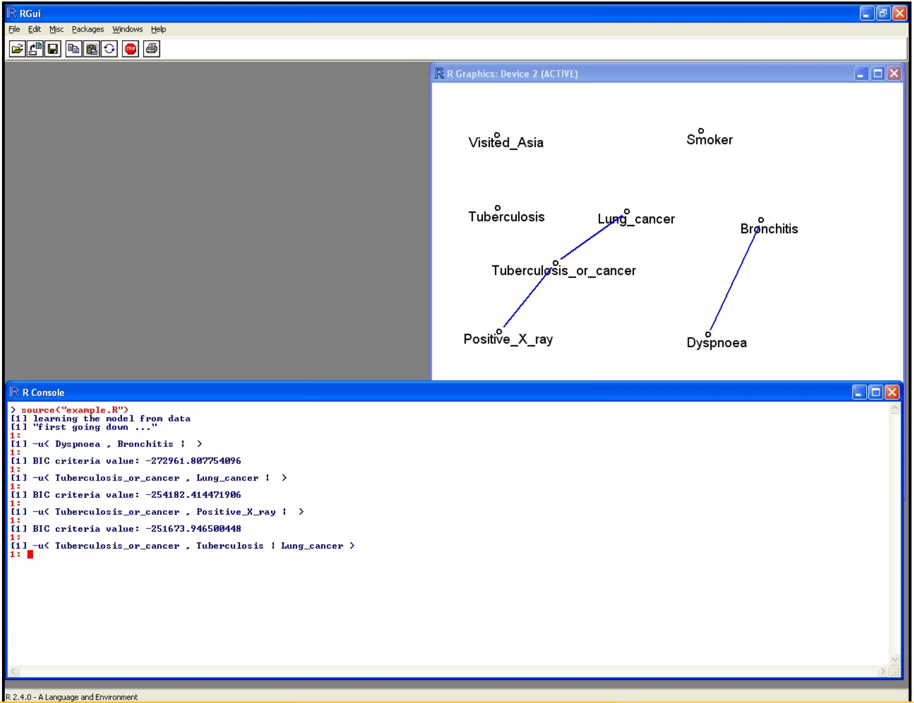

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
R 2.4.0 - A Language and Environment

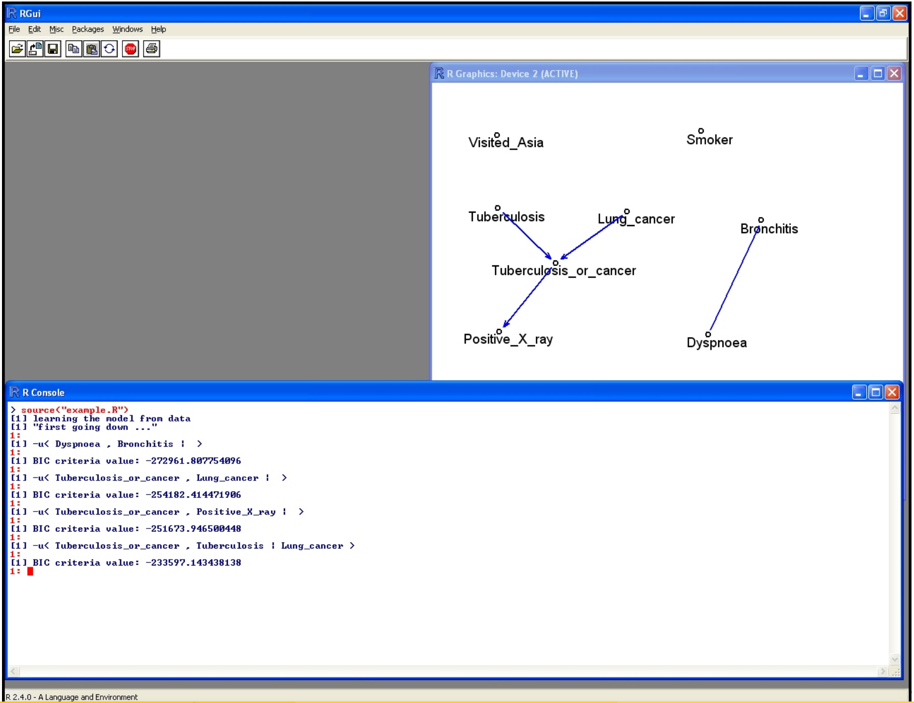

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
R 2.4.0 - A Language and Environment

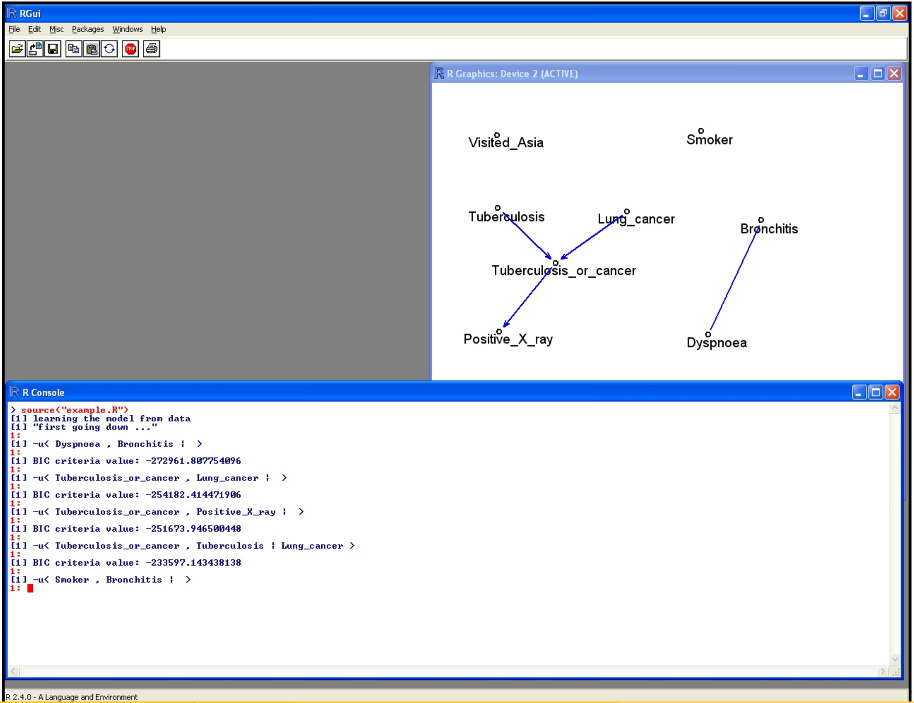

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
R 2.4.0 - A Language and Environment

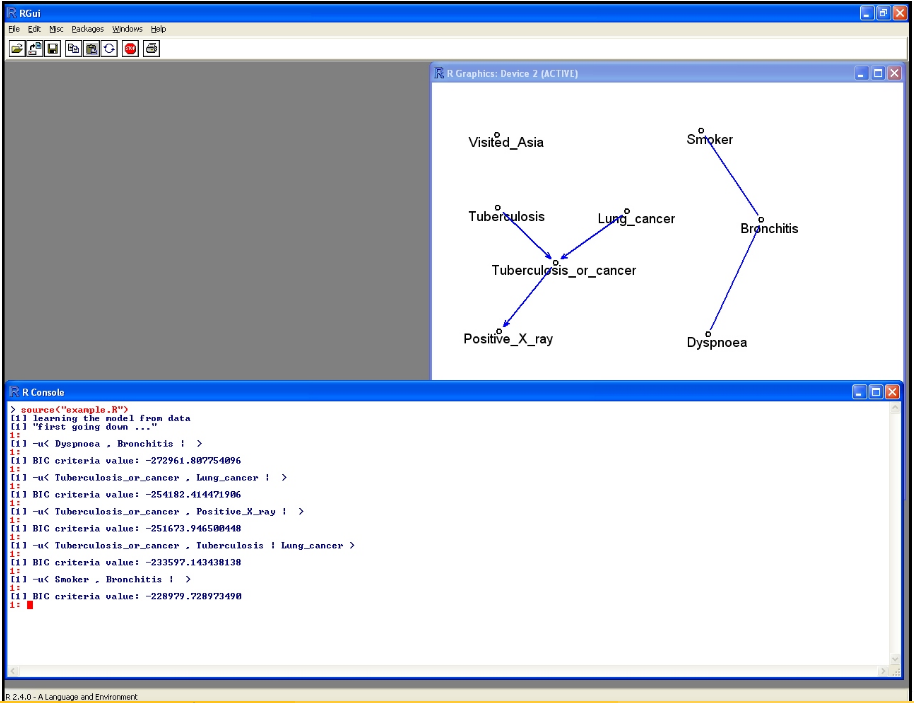

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
R 2.4.0 - A Language and Environment

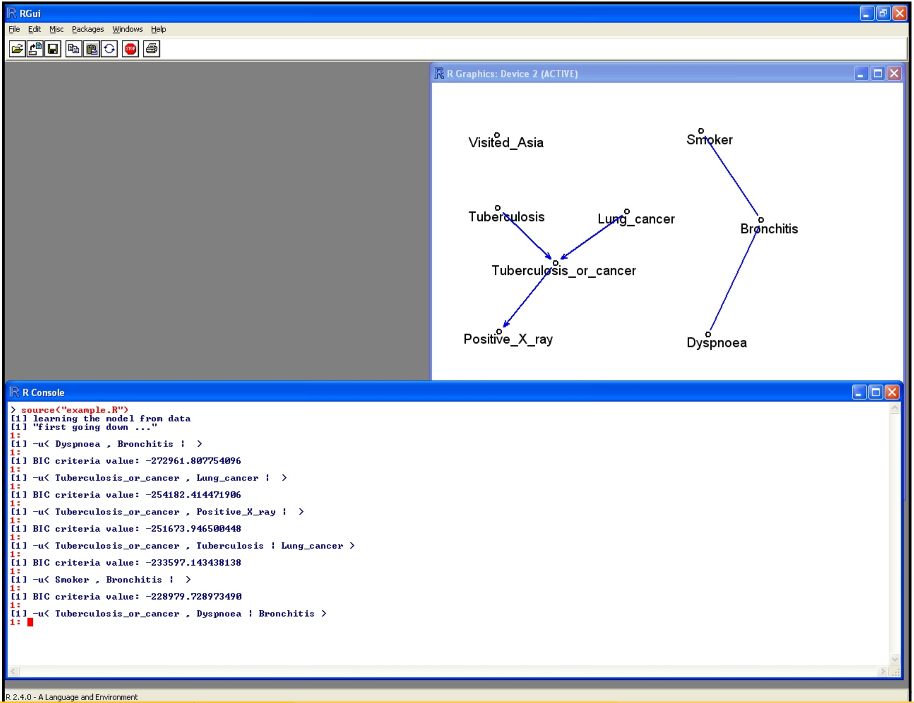

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
R 2.4.0 - A Language and Environment

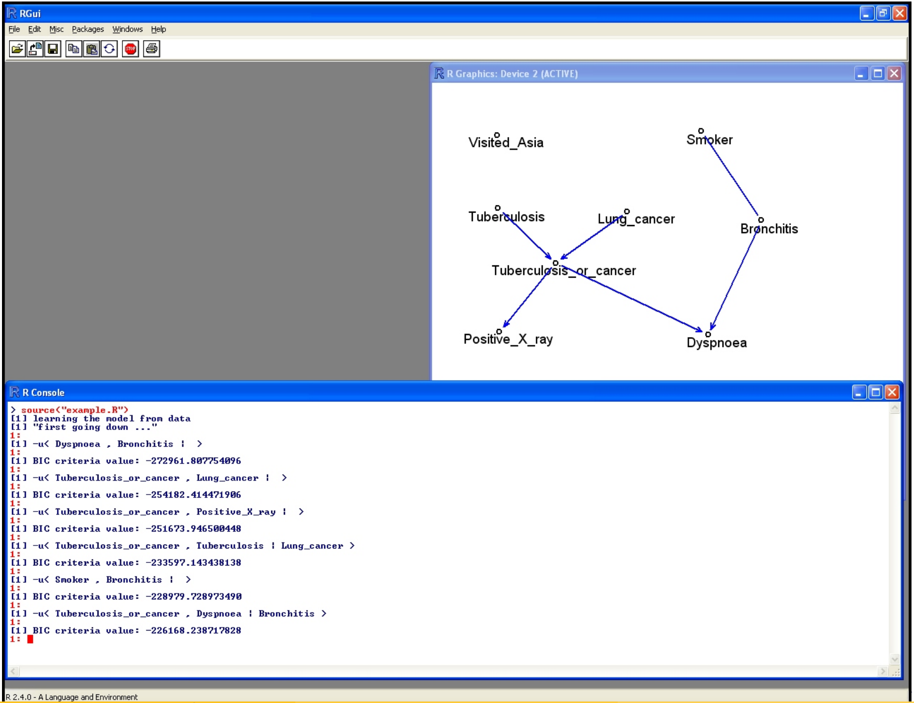

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
R 2.4.0 - A Language and Environment

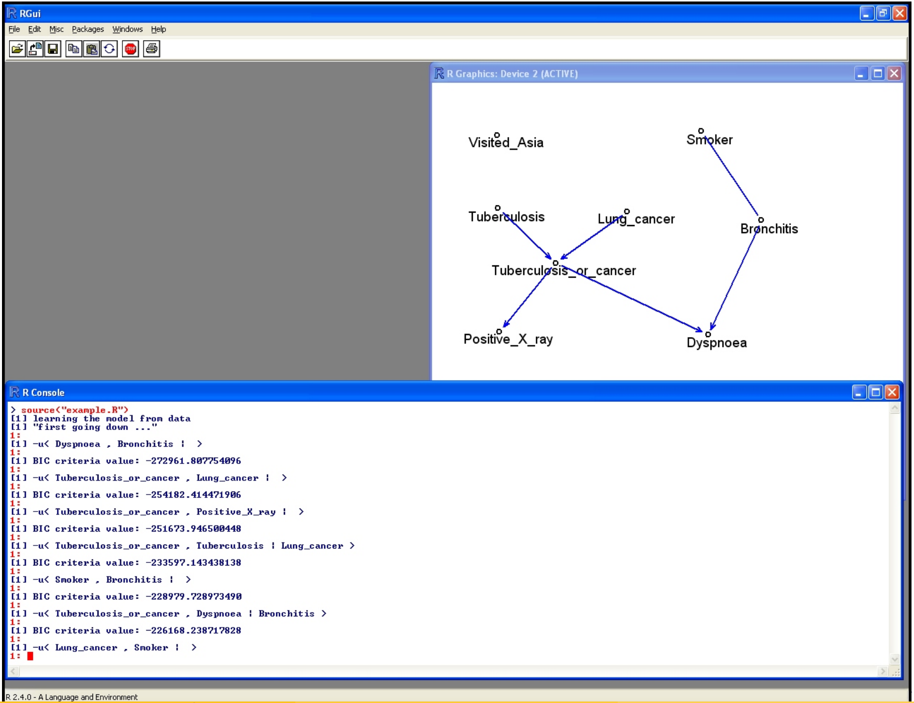

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
R 2.4.0 - A Language and Environment

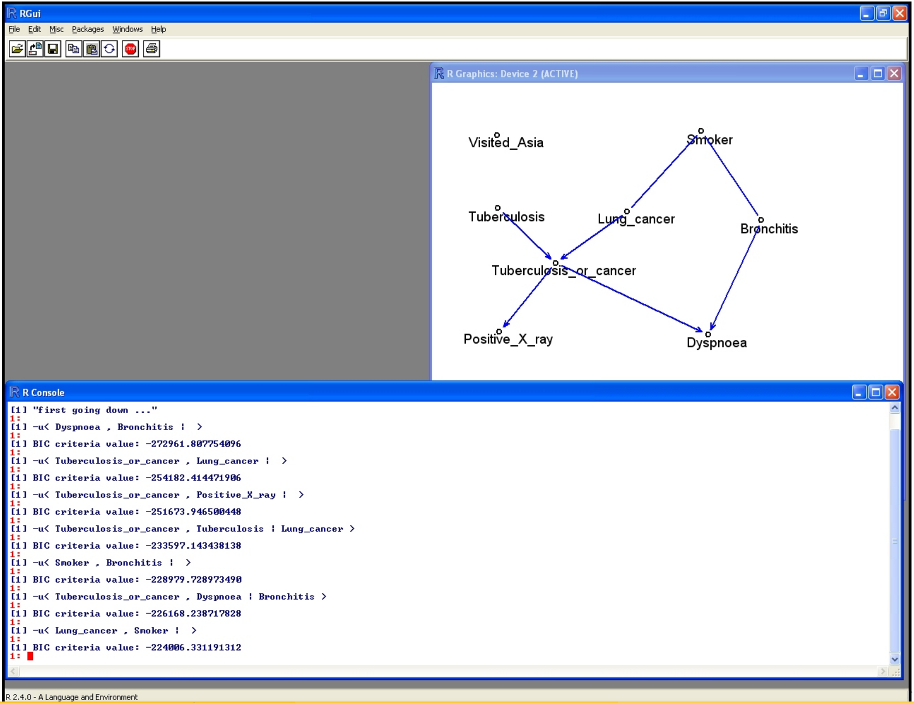

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
R 2.4.0 - A Language and Environment

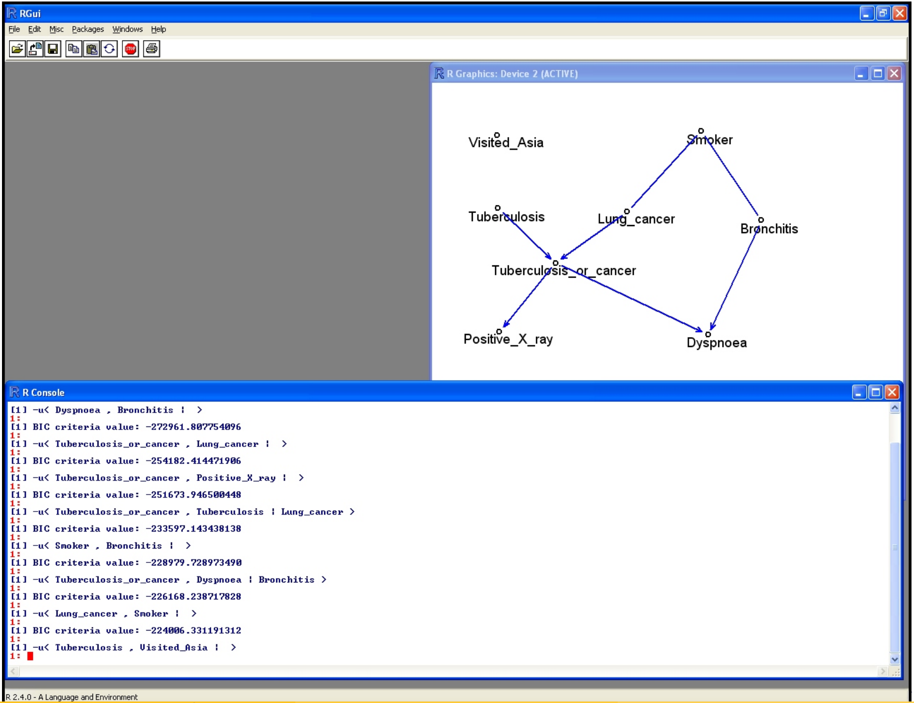

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
R 2.4.0 - A Language and Environment

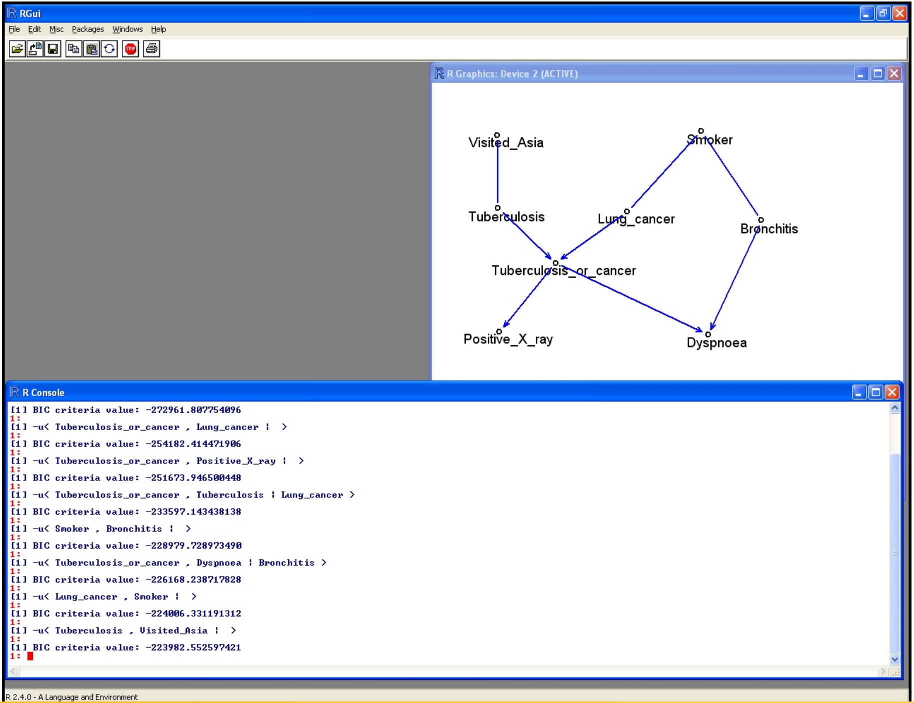

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
R 2.4.0 - A Language and Environment

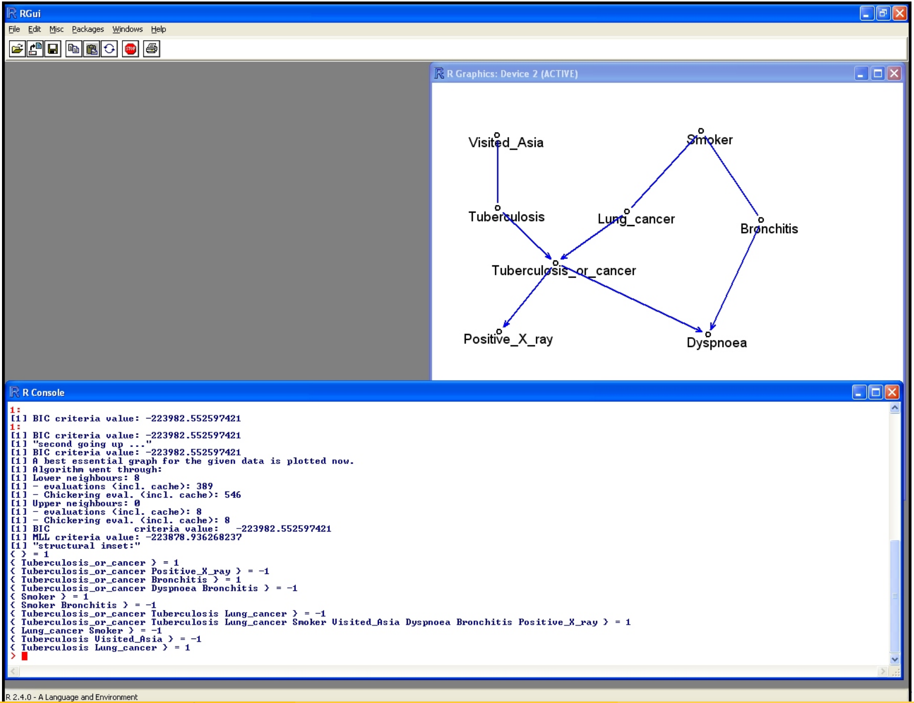

RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
[1] BIC criteria value: -226168.238717828
1:
R 2.4.0 - A Language and Environment


RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
> source("example.R")
[1] learning the model from data
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
[1] BIC criteria value: -226168.238717828
1:
[1] -u< Lung_cancer , Smoker | >
1:
R 2.4.0 - A Language and Environment


RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
[1] "first going down ..."
1:
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
[1] BIC criteria value: -226168.238717828
1:
[1] -u< Lung_cancer , Smoker | >
1:
[1] BIC criteria value: -224006.331191312
1:
R 2.4.0 - A Language and Environment


RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
[1] -u< Dyspnoea , Bronchitis | >
1:
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
[1] BIC criteria value: -226168.238717828
1:
[1] -u< Lung_cancer , Smoker | >
1:
[1] BIC criteria value: -224006.331191312
1:
[1] -u< Tuberculosis , Visited_Asia | >
1:
R 2.4.0 - A Language and Environment


RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
[1] BIC criteria value: -272961.807754096
1:
[1] -u< Tuberculosis_or_cancer , Lung_cancer | >
1:
[1] BIC criteria value: -254182.414471906
1:
[1] -u< Tuberculosis_or_cancer , Positive_X_ray | >
1:
[1] BIC criteria value: -251673.946500448
1:
[1] -u< Tuberculosis_or_cancer , Tuberculosis | Lung_cancer >
1:
[1] BIC criteria value: -233597.143438138
1:
[1] -u< Smoker , Bronchitis | >
1:
[1] BIC criteria value: -228979.728973490
1:
[1] -u< Tuberculosis_or_cancer , Dyspnoea | Bronchitis >
1:
[1] BIC criteria value: -226168.238717828
1:
[1] -u< Lung_cancer , Smoker | >
1:
[1] BIC criteria value: -224006.331191312
1:
[1] -u< Tuberculosis , Visited_Asia | >
1:
[1] BIC criteria value: -223982.552597421
1:
R 2.4.0 - A Language and Environment


RGui
File Edit Misc Packages Windows Help
R Graphics: Device 2 (ACTIVE)
Visited_Asia
Smoker
Tuberculosis
Lung_cancer
Bronchitis
Tuberculosis_or_cancer
Positive_X_ray
Dyspnoea
R Console
1:
[1] BIC criteria value: -223982.552597421
1:
[1] BIC criteria value: -223982.552597421
[1] "second going up ..."
[1] BIC criteria value: -223982.552597421
[1] A best essential graph for the given data is plotted now.
[1] Algorithm went through:
[1] Lower neighbours: 8
[1] - evaluations (incl. cache): 389
[1] - Chickering eval. (incl. cache): 546
[1] Upper neighbours: 0
[1] - evaluations (incl. cache): 8
[1] - Chickering eval. (incl. cache): 8
[1] BIC criteria value: -223982.552597421
[1] MLL criteria value: -223878.936268237
[1] "structural imset:"
{ } = 1
{ Tuberculosis_or_cancer } = 1
{ Tuberculosis_or_cancer Positive_X_ray } = -1
{ Tuberculosis_or_cancer Bronchitis } = 1
{ Tuberculosis_or_cancer Dyspnoea Bronchitis } = -1
{ Smoker } = 1
{ Smoker Bronchitis } = -1
{ Tuberculosis_or_cancer Tuberculosis Lung_cancer } = -1
{ Tuberculosis_or_cancer Tuberculosis Lung_cancer Smoker Visited_Asia Dyspnoea Bronchitis Positive_X_ray } = 1
{ Lung_cancer Smoker } = -1
{ Tuberculosis Visited_Asia } = -1
{ Tuberculosis Lung_cancer } = 1
>
R 2.4.0 - A Language and Environment