Modelli Probabilistici per le Decisioni

Enza Messina

Reti Bayesiane

DISCo

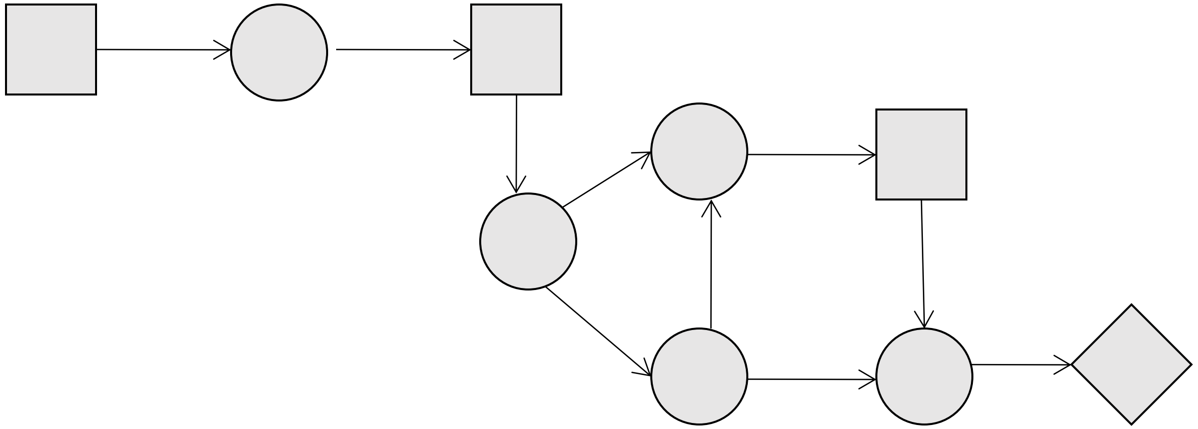
Università degli Studi di Milano-

Bicocca

Viale Sarca, 336

20126 Milano

messina@disco.unimib.it



Semantica delle reti Bayesiane

Semantica delle Reti Bayesiane: *un metodo per la costruzione di una Rete Bayesiana*

La *formula di fattorizzazione della distribuzione della probabilità congiunta*

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid parents\ (X_i))$$

Come costruire una *Rete Bayesiana* ?

- la distribuzione congiunta di probabilità che essa induce deve fornire una buona rappresentazione per un dato dominio applicativo.

La *formula di fattorizzazione* implica relazioni di *indipendenza condizionale* che possono essere sfruttate per determinare la *componente topologica* della *Rete Bayesiana*.

Semantica delle reti Bayesiane

La regola della probabilità congiunta la possiamo scrivere:

$$P(x_1,\ldots,x_n) = P(x_n \mid x_{n-1},\ldots,x_1) \cdot P(x_{n-1} \mid x_{n-2},\ldots,x_1) \cdot \ldots \cdot P(x_2 \mid x_1) \cdot P(x_1) = \prod_{i=1}^{n} P(x_i \mid x_{n-1},\ldots,x_1)$$

Questa uguaglianza è vera per ogni insieme di variabili aleatorie ed è nota con il termine di *chain rule*.

Confrontando la *chain rule* con la *formula di fattorizzazione* è possibile verificare che la specificazione della distribuzione di probabilità congiunta è equivalente all'asserzione generale che, per ogni variabile X_i,

$$\mathbf{P}(x_i \mid x_{i-1},\ldots,x_1) = \mathbf{P}(x_i \mid \mathit{parents}\ (X_i))$$

a patto che:

$$\mathit{Parents}\ (x_i) \subseteq \{x_{i-1},\ldots,x_1\}$$

Semantica delle reti Bayesiane

$$\mathbf{P}(x_i \mid x_{n-1},\ldots,x_1) = \mathbf{P}(x_i \mid \textit{parents}\ (X_i))$$

Una *Rete Bayesiana* rappresenta correttamente un dominio solo a condizione che *ogni nodo* risulti *condizionalmente indipendente dai suoi predecessori*, per un dato ordinamento, *dati i suoi genitori*.

Pertanto, per costruire una *Rete Bayesiana* che abbia la corretta struttura del dominio da modellare è necessario scegliere, per ogni nodo, i nodi genitore in modo che tale proprietà risulti verificata.

Intuitivamente, l'insieme dei genitori per ogni nodo X_i, ovvero tutti i nodi che influenzeranno direttamente il nodo X_i, devono poter essere scelti tra $X_1,\ldots,X_{i-1}$

Semantica delle reti Bayesiane

Una possibile *procedura* per la *costruzione incrementale* della *componente topologica* di una *Rete Bayesiana* è sintetizzabile nei seguenti passi:

1. Selezionare un insieme di variabili $\{X_1, \ldots, X_n\}$ da utilizzare per descrivere il dominio da modellare,
2. Scegliere un ordinamento delle variabili $\{X_{(1)}, \ldots, X_{(n)}\}$, (*molto importante*)
3. Inizializzare il numero di nodi aggiunti alla rete ad uno; $i = 1$,
4. Selezionare la variabile $X_{(i)}$ e aggiungere il nodo corrispondente alla rete,

   a) porre $Parents(X_{(i)})$ uguale all'insieme minimale di nodi, attualmente appartenenti alla rete $\{X_{(1)}, \ldots, X_{(i-1)}\}$, che soddisfa la proprietà di indipendenza condizionale:

   $$\mathbf{P}(X_{(i)} \mid X_{(i-1)},\ldots,X) = \mathbf{P}(X_{(i)} \mid Parents\ (X_{(i)}))$$

   b) computare la CPT per la variabile $X_{(i)}$,

5. Incrementare il numero di nodi aggiunti alla rete; $i = i + 1.$ Se si sono aggiunte tutte le variabili alla rete ($i > n$), allora la procedura termina, in caso contrario tornare al passo 4.

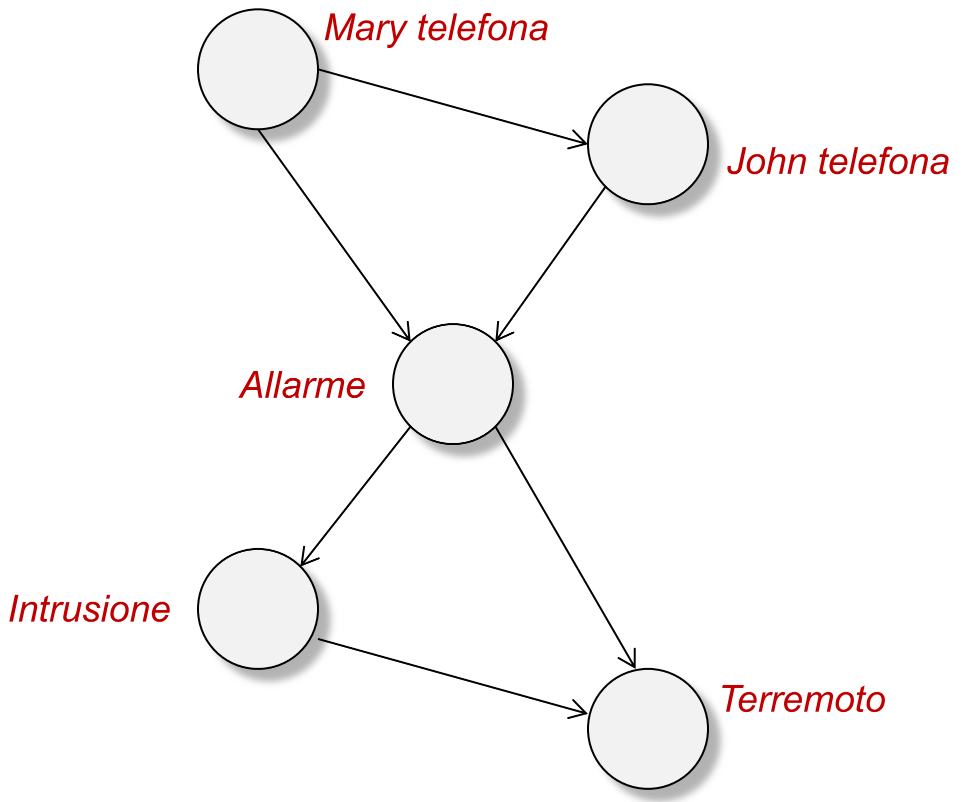
Semantica delle reti Bayesiane

Cosa accade se al passo 2 della procedura, di costruzione incrementale, viene scelto un ordinamento delle varabili $\{X_{(1)}, \ldots, X_{(n)}\}$ che risulta essere errato ?

Supponiamo che nell'esempio venga scelto il seguente ordinamento delle variabili:

- *Mary telefona*
- *John telefona*
- *Allarme*
- *Intrusione*
- *Terremoto*

Applicando la procedura di costruzione incrementale della componente qualitativa otterremo un modello più complesso di *Rete Bayesiana*, che viene illustrato a destra.



Semantica delle reti Bayesiane

La *procedura di costruzione incrementale* si svolgerebbe come segue:

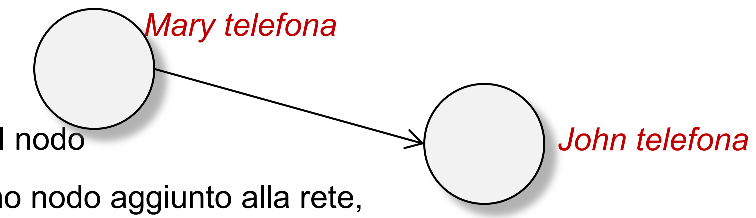
1. Selezioniamo $n=5$ variabili per descrivere il dominio

   $\{X_1, \ldots, X_5\}$ = {*Allarme, Intrusione, Terremoto, Mary telefona, John telefona*}

1. Scegliamo il seguente ordinamento delle variabili

   $\{X_{(1)}, \ldots, X_{(5)}\}$ = {*Mary telefona, John telefona, Allarme, Intrusione, Terremoto*}

1. Poniamo il numero di nodi aggiunti alla rete pari a uno; $i = 1$,



2. Selezioniamo la variabile $X_{(1)}$ = *Mary telefona e* aggiungiamo il nodo corrispondente alla rete; non ha alcun genitore in quanto primo nodo aggiunto alla rete,

1. Si incrementa il numero di nodi aggiunti alla rete; $i=i+1$, $i=1+1=2$. Si controlla la condizione di terminazione; $i=2 \leq 5=n$ per cui si va al passo 4,

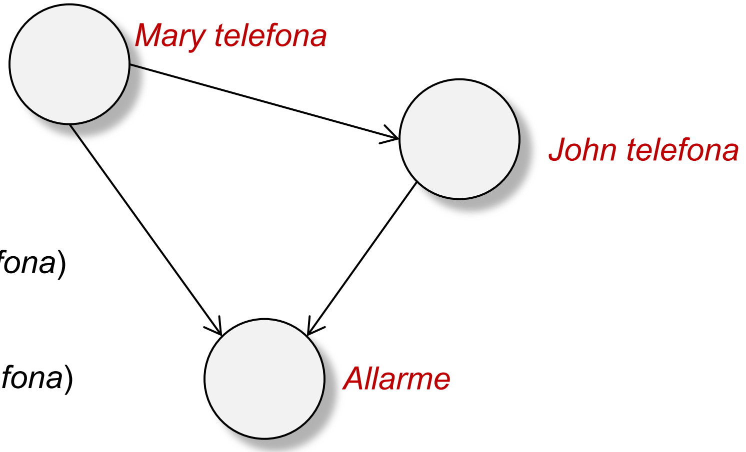
2. Selezioniamo la variabile $X_{(2)}$ = *John telefona* e aggiungiamo il nodo corrispondente alla rete; se Mary ci telefona (*Mary telefona* = *v*) probabilmente è perché l'allarme si è attivato (*Allarme* = *v*) *il* che rende più verosimile che anche John ci telefoni (*John telefona* = *v*). Pertanto, concludiamo che la variabile *John telefona* ha come genitore la variabile *Mary telefona*.

$$\mathbf{P}(\textit{John telefona} \mid \textit{Mary telefona}) \neq \mathbf{P}(\textit{John telefona})$$

Semantica delle reti Bayesiane

7. Si incrementa il numero di nodi aggiunti alla rete; *i=i+1, i=2+1=3*. Si controlla la condizione di terminazione; *i=3 ≤ 5=n* per cui si va al passo 4,

8. Selezioniamo la variabile $X_{(3)}$ *= Allarme;* se Mary e John ci telefonano (*Mary telefona = v, John telefona = v*) è più verosimile che l'allarme stia suonando (*Allarme = v*) che non se uno solo dei due ci telefona oppure se entrambe non ci telefonano. Pertanto, concludiamo che la variabile *Allarme* ha come genitori le variabili *Mary telefona* e *John telefona.*



P(*Allarme* | *Mary telefona,John telefona*) ≠**P**(*Allarme*)

P(*Allarme* | *Mary telefona,John telefona*) ≠ **P**(*Allarme* | *Mary telefona*)

P(*Allarme* | *Mary telefona,John telefona*) ≠ P(*Allarme* | *John telefona*)

9. Si incrementa il numero di nodi aggiunti alla rete; *i=i+1, i=3+1=4*. Si controlla la condizione di terminazione; *i=4≤5=n* per cui si va al passo 4,

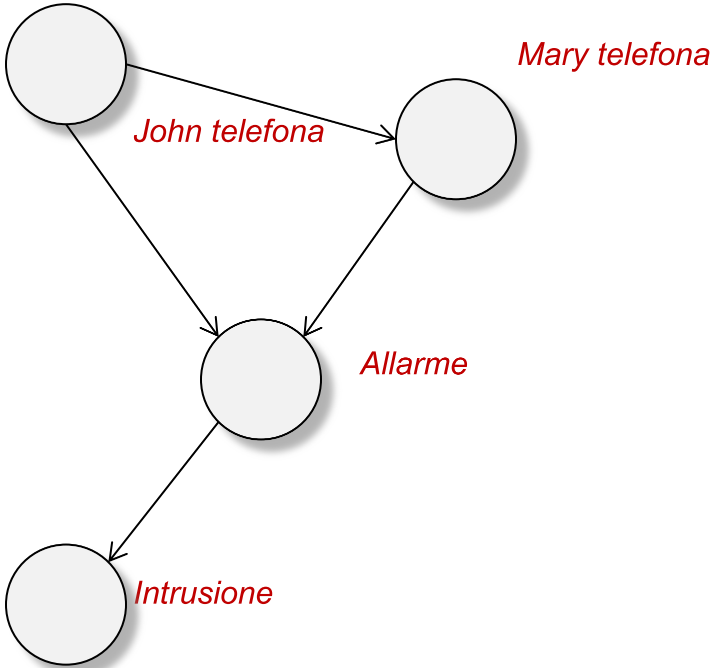
Semantica delle reti Bayesiane

10. Selezioniamo la variabile $X_{(4)}$ = *Intrusione*; se siamo a conoscenza dello stato della variabile *Allarme* allora il fatto che Mary o John ci telefonino (*Mary telefona* = *v* o *John telefona* = *v*) fornisce informazione riguardo al fatto che Mary stia ascoltando musica oppure che il nostro telefono di casa stia suonando. Comunque in tale situazione le variabili *Mary telefona* e *John telefona* non forniscono nessuna informazione circa l'intrusione (*Intrusione*). Possiamo allora scrivere:

$$\mathbf{P}(\textit{Intrusione} \mid \textit{Allarme}, \textit{John telefona}, \textit{Mary telefona}) = \mathbf{P}(\textit{Intrusione} \mid \textit{Allarme})$$

    Concludiamo allora che *Intrusione* ha solo *Allarme* come nodo genitore.

11. Si incrementa il numero di nodi aggiunti alla rete; *i=i+1, i=4+1=5*. Si controlla la condizione di terminazione; *i=5≤5=n* e si va al passo 4,
12. Selezioniamo la variabile $X_{(5)}$ = *Terremoto*; se l'allarme sta suonando (*Allarme* = *v*), è più verosimile che si sia verificata una scossa di terremoto (*Terremoto* = *v*). Se però siamo a conoscenza del fatto che vi è stata un'intrusione (*Intrusione=v*) allora questo evento spiega l'attivazione dell'allarme, mentre la probabilità di una scossa di terremoto risulterebbe solo marginalmente incrementata.



Semantica delle reti Bayesiane

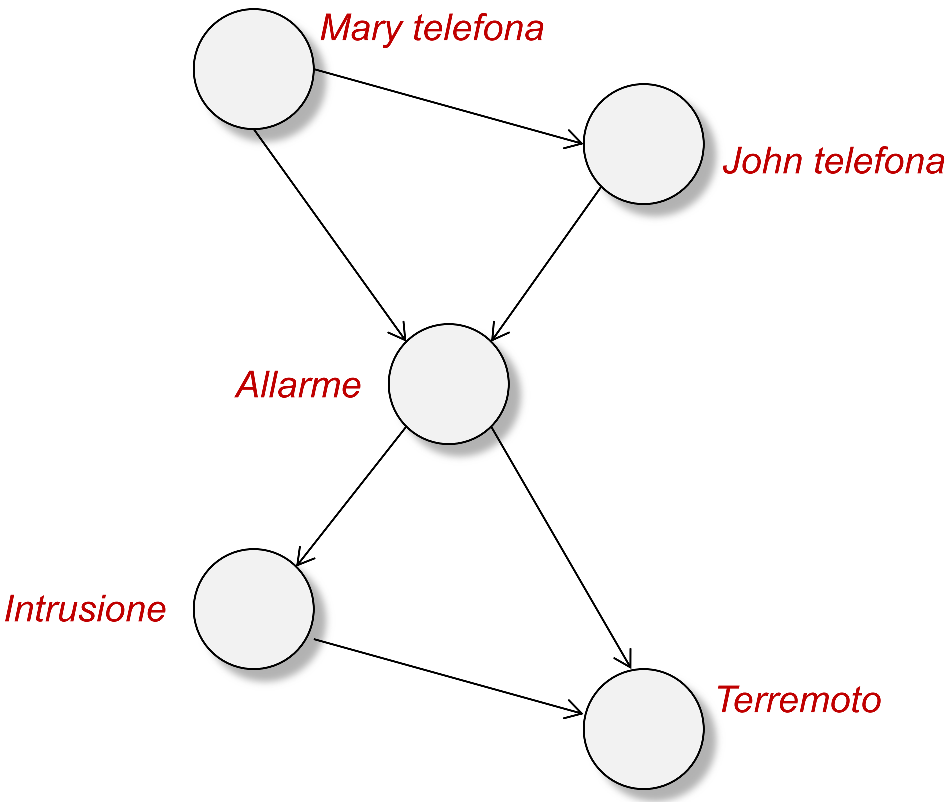
In definitiva la variabile *Terremoto* ha bisogno avere come nodi genitore sia la variabile *Allarme* che la variabile *Intrusione.*

P(*Terremoto* | *Mary telefona,John telefona,Allarme,Intrusione*) = **P**(*Terremoto* | *Allarme,Intrusione*)

11. Si incrementa il numero di nodi aggiunti alla rete; *i=i+1, i=5+1=6*. Si controlla la condizione di terminazione;

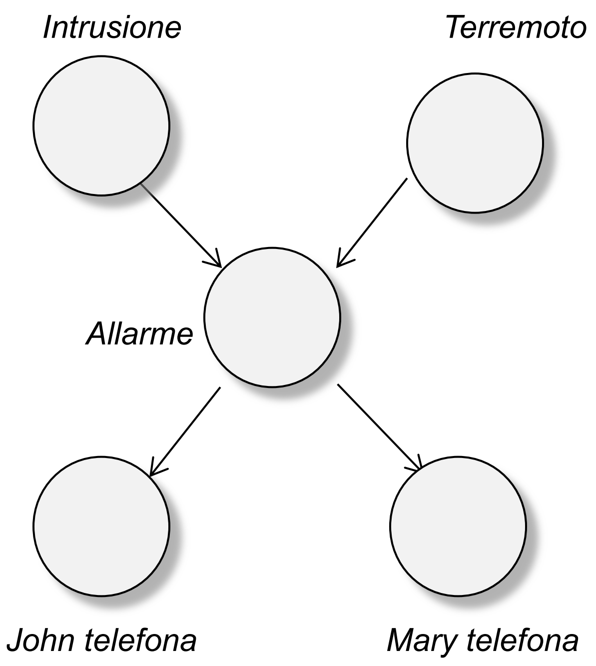
    i=6>5=n

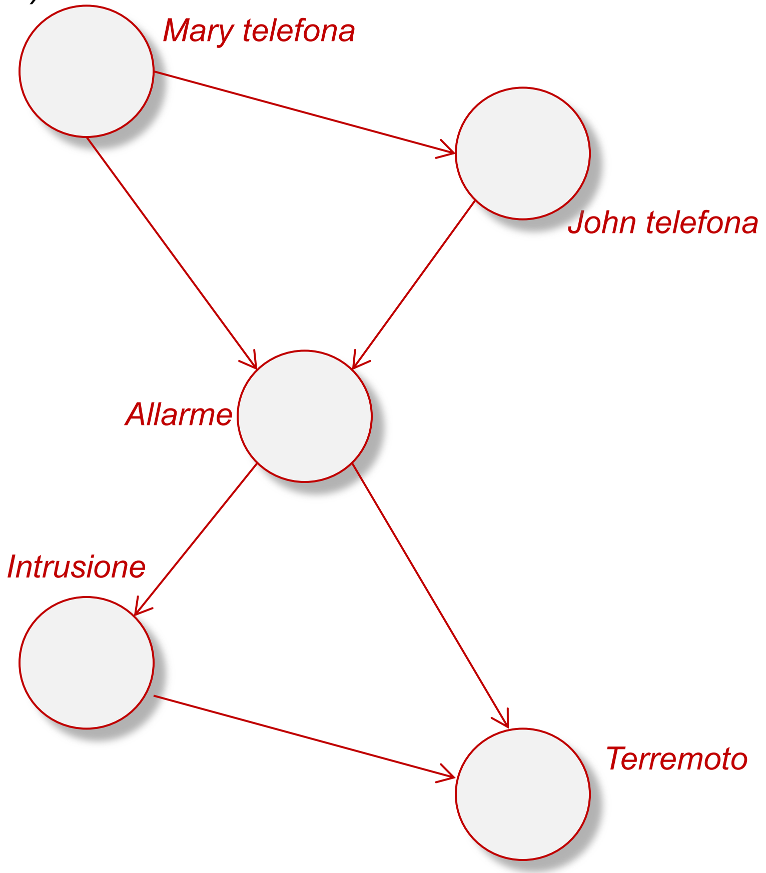
    per cui la procedura di costruzione incrementale della componente qualitativa della Rete Bayesiana termina.



Semantica delle reti Bayesiane

Rispetto a quello originale, il modello ottenuto ha un maggior numero di collegamenti causali, richiede di specificare un numero maggiore di valori di probabilità e, cosa peggiore, è caratterizzato dal fatto che diverse relazioni causali sono molto tenui e richiedono un giudizio complesso ed innaturale dei corrispettivi valori di probabilità, per esempio **P**(*Terremoto* | *Allarme,Intrusione).*

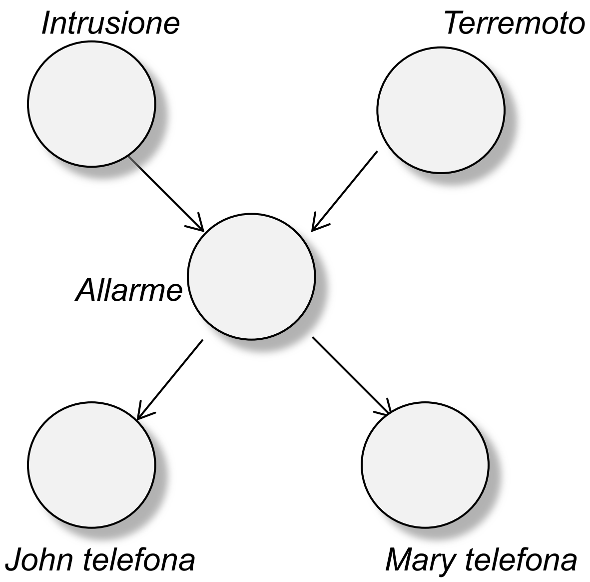


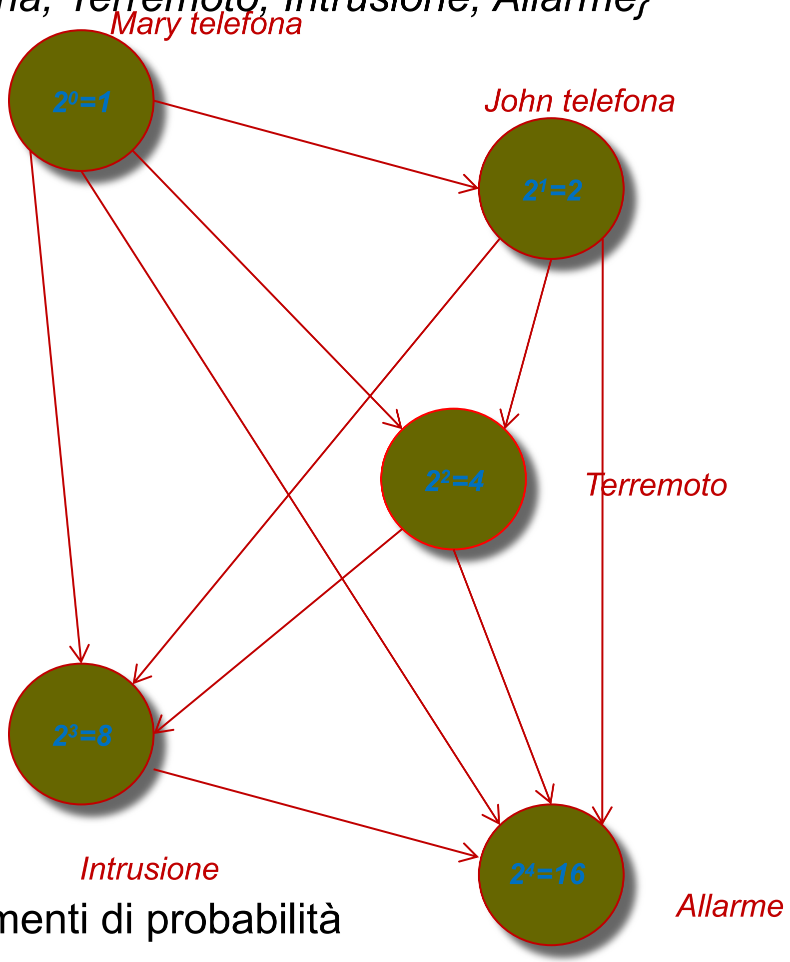


Semantica delle reti Bayesiane

Il modello di *Rete Bayesiana* riportato a destra è ottenuto a partire dal seguente ordinamento:

$\{X_{(1)}, \ldots, X_{(5)}\}$ = {*Mary telefona, John telefona, Terremoto, Intrusione, Allarme*}



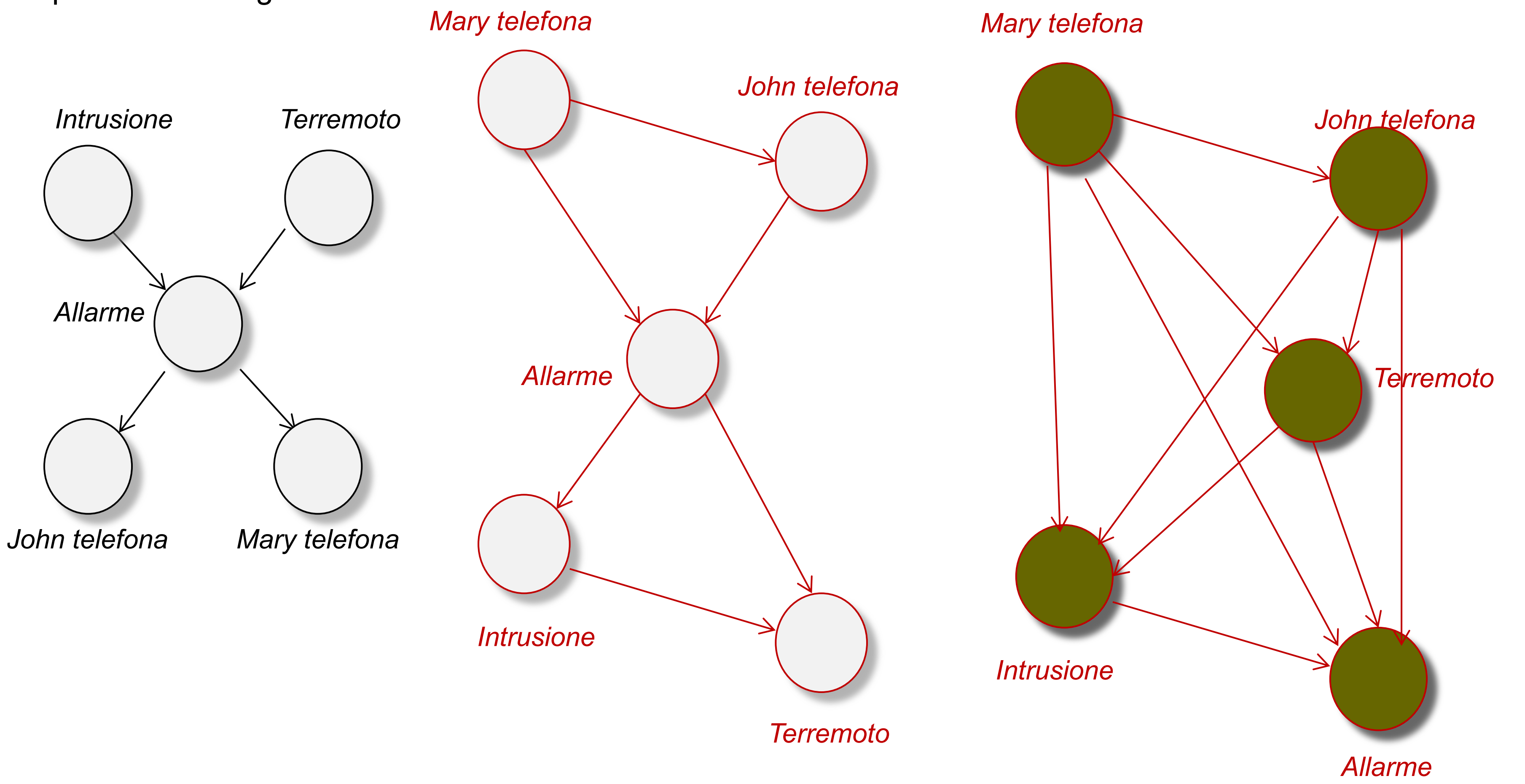


Richiede di specificare 31 (1+2+4+8+16) distinti elementi di probabilità

(stesso numero richiesto per specificare l'intera distribuzione di probabilità congiunta).

Semantica delle reti Bayesiane

È importante sottolineare che i tre modelli sottostanti possono rappresentare la medesima distribuzione di probabilità congiunta.



I due modelli a destra non rappresentano tutte le relazioni di indipendenza condizionale e pertanto richiedono di specificare informazione ridondante per descrivere la medesima distribuzione di probabilità congiunta descritta dal modello corretto (a sinistra).

Semantica delle reti Bayesiane

Semantica delle Reti Bayesiane: *compattezza e ordinamento dei nodi*

Oltre a costituire una rappresentazione completa e non ridondante di un dominio una *Rete Bayesiana* è spesso molto più compatta dell'intera distribuzione di probabilità congiunta.
Questo rende possibile il trattamento di domini caratterizzati da un numero molto elevato di variabili.

La *compattezza delle Reti Bayesiane* è un esempio della proprietà dei sistemi *strutturati localmente* o *sparsi*.

In ogni sistema *strutturato localmente* ogni sottocomponente interagisce solo con un numero limitato di altre componenti, indipendentemente dal numero totale di componenti del sistema.
La strutturazione locale è di norma associata ad un fattore di crescita della complessità lineare e non esponenziale.

Nel caso di una *Rete Bayesiana* è ragionevole ipotizzare che *ogni variabile* sia *direttamente influenzata* da *al massimo k variabili* (*k* costante).

Semantica delle reti Bayesiane

Nel caso in cui si consideri una *Rete Bayesiana* costituita da *n variabili* (nodi) *Booleane* avremo che la quantità di *informazione necessaria* per *specificare* una qualsiasi CPT è limitata superiormente da:

$$2^k \text{ numeri}$$

per cui la rete completa richiede di specificare al più

$$n \cdot 2^k \text{ numeri}$$

Mentre specificare l'intera distribuzione congiunta di probabilità richiede

$$2^n \text{ numeri}$$

Consideriamo il caso in cui abbiamo n=30 nodi e ogni nodo può avere *al massimo k=5 genitori*

In questo caso la Rete Bayesiana richiede di specificare

$$n \cdot 2^k = 30 \cdot 2^5 = 960 \text{ parametri (numeri)}$$

mentre specificare l'intera distribuzione di probabilità congiunta richiederebbe

$$2^n = 2^{30} = 1{,}073{,}741{,}824 \text{ parametri.}$$