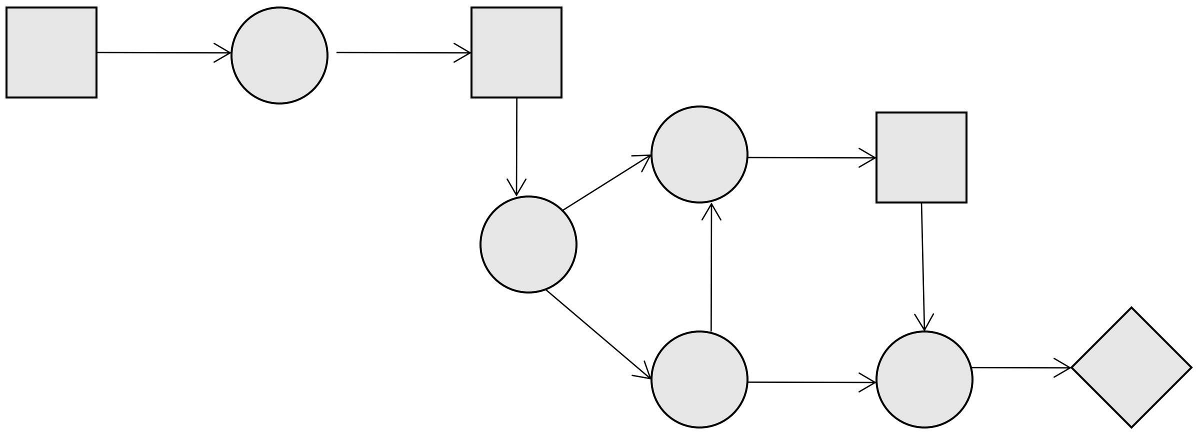
Modelli Probabilistici per le Decisioni

Enza Messina

Reti Bayesiane

DISCo
Università degli Studi di Milano-Bicocca
Viale Sarca, 336
20126 Milano
messina@disco.unimib.it



Reti Bayesiane

I concetti di *indipendenza* e *indipendenza condizionata* tra variabili offrono la possibilità di memorizzare e trattare in modo efficiente distribuzioni congiunte di probabilità con dimensione elevata.

Ora presenteremo il paradigma computazionale ideato da *Judea Pearl* che va sotto il nome di *Rete Bayesiana* e che traduce in termini concreti i diversi concetti che abbiamo presentato fino ad ora

Una *Rete Bayesiana* è spesso riferita anche con i seguenti termini

- *Bayesian Belief Network* (BBN)
- *Probabilistic Network* (PN)
- *Casual Network* (CN)

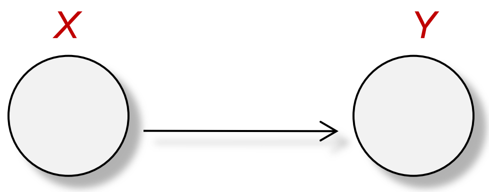
Le *Reti Bayesiane* appartengono ad una classe più ampia di modelli che viene riferita nella letteratura specializzata con il termine di classe dei *Modelli Grafici* (*Graphical Models*)

Reti Bayesiane

Una *Rete Bayesiana* è *un* grafo in cui i nodi sono annotati con una informazione *quantitativa* (tabelle di probabilità condizionata) e i cui archi definiscono dipendenza e indipendenza condizionale tra le variabili rappresentate dai nodi.

Il grafo orientato è costituito da:

- *Nodo:* associato ad una variabile, relazione 1 a 1 tra nodo e variabile
- *Arco orientato:* collega due nodi e traduce di norma una relazione di causalità diretta



Di norma (ma non sempre) diremo che *X* è causa diretta di *Y*.

Diremo inoltre che *X* è *genitore* di *Y* e che *Y* è *figlio* di *X* .

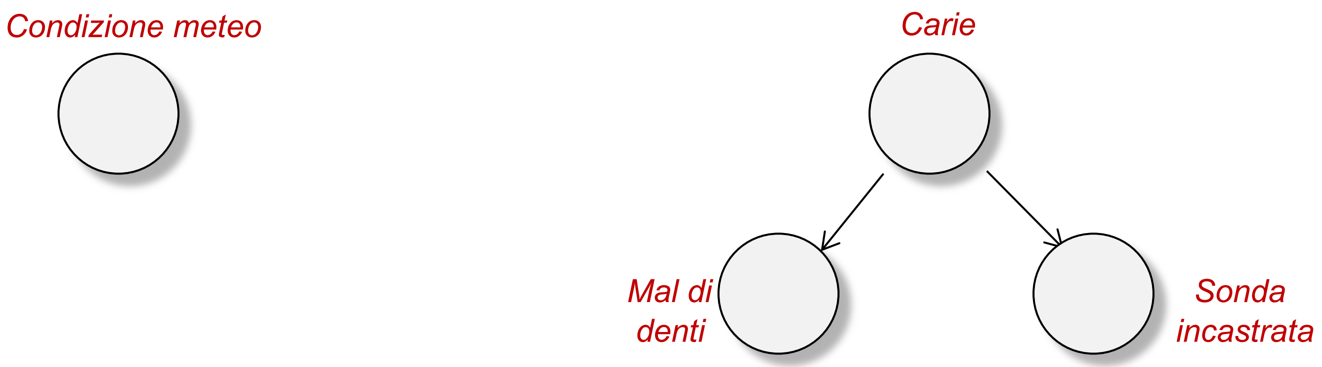
Le *variabili* possono essere *continue* o *discrete*.

La topologia della rete e le probabilità condizionate dei nodi dati i loro genitori, sono sufficienti a specificare (implicitamente) la distribuzione congiunta di tutte le variabili.

Reti Bayesiane

Tornando all'esempio del dentista abbiamo concluso che

- *Condizione meteo* è indipendente da *Carie*, *Mal di denti* e *Sonda incastrata*,
- *Mal di denti* e *Sonda incastrata* sono *condizionalmente indipendenti* data la conoscenza di *Carie*.

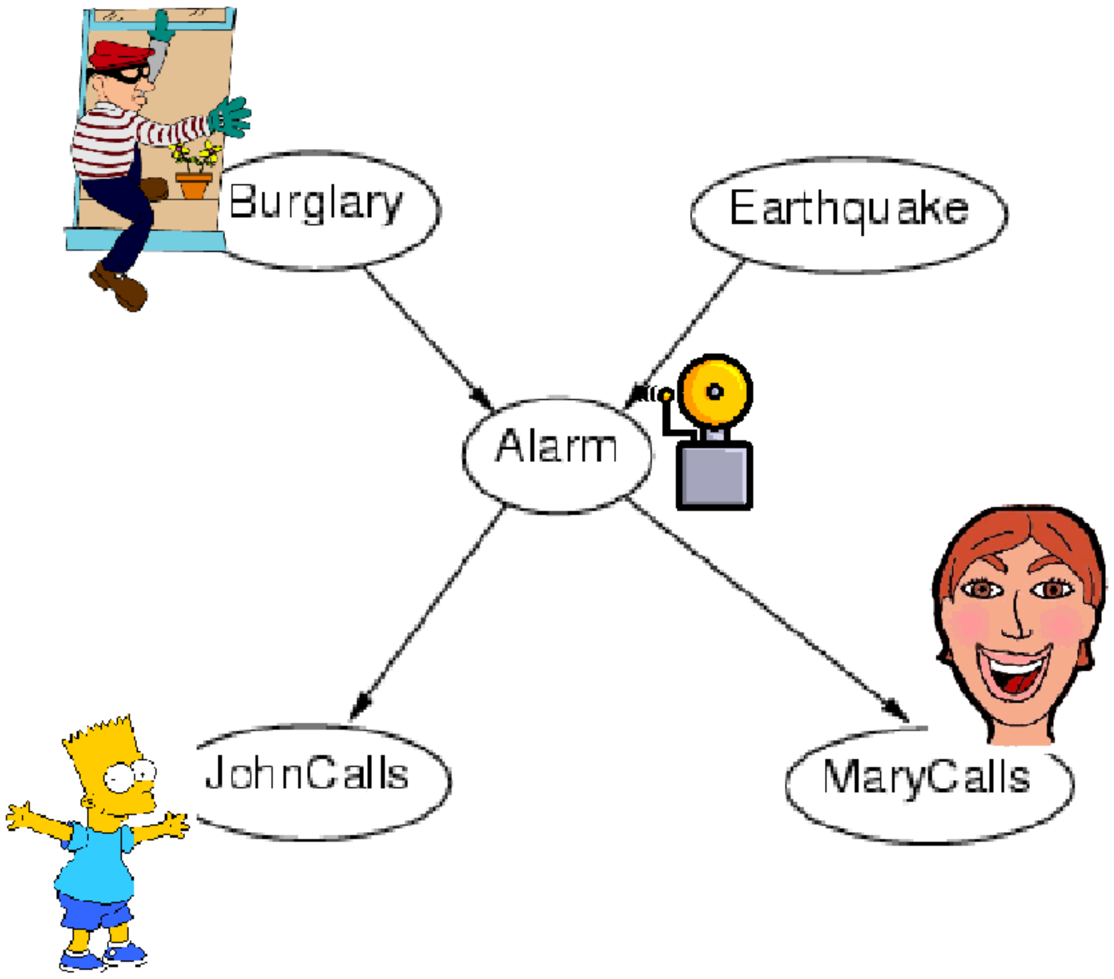


La componente topologica della *Rete Bayesiana* traduce queste due condizioni come segue:

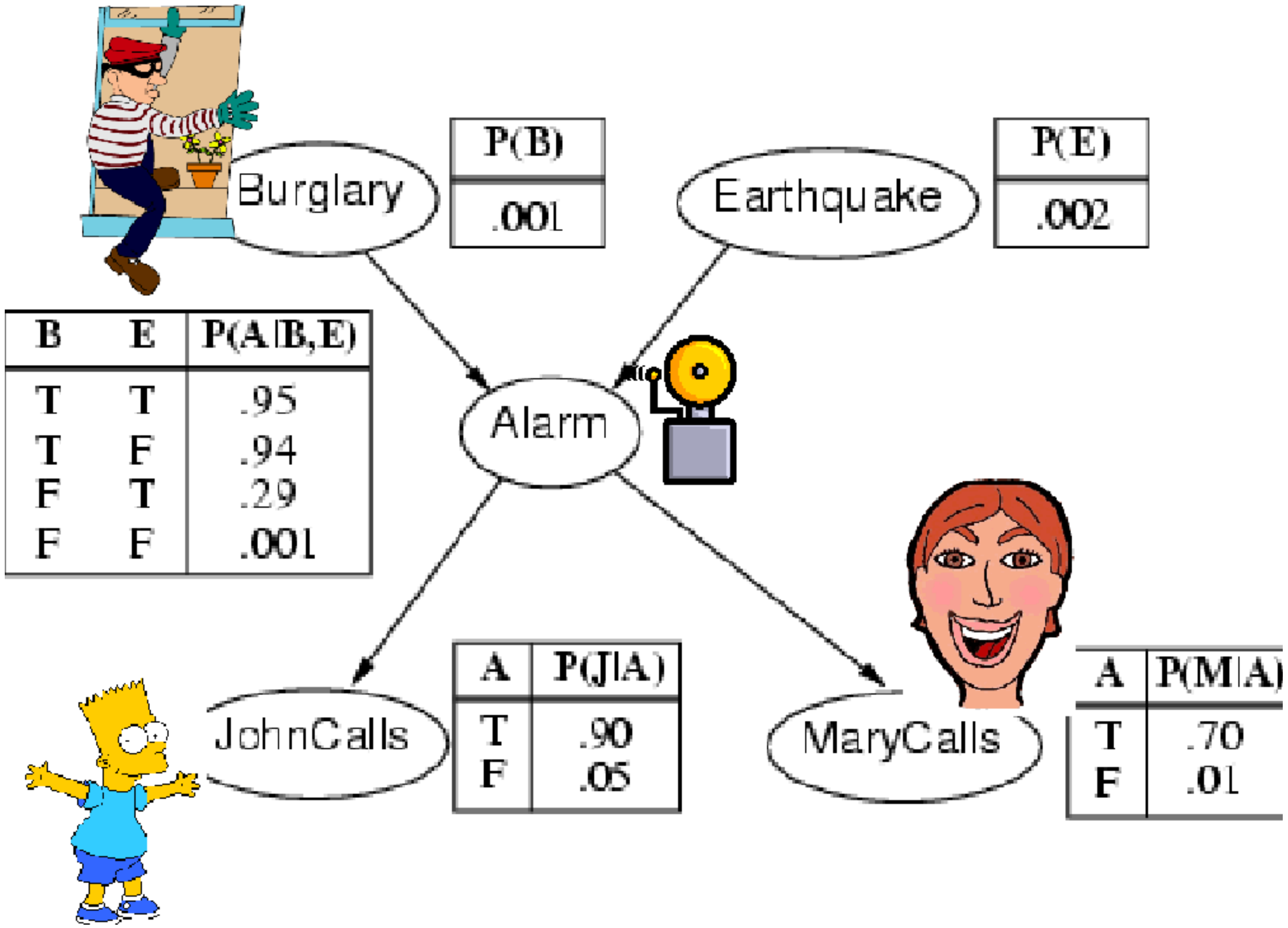
- assenza di collegamenti tra *Condizione meteo* e le altre varibili,
- assenza di collegamenti tra *Mal di denti* e *Sonda incastrata*.

Inoltre la *Rete Bayesiana* stabilisce che *Carie* è causa diretta per *Mal di denti* e *Sonda incastrata*, e tra queste due variabili *non esiste una relazione diretta di causalità*.

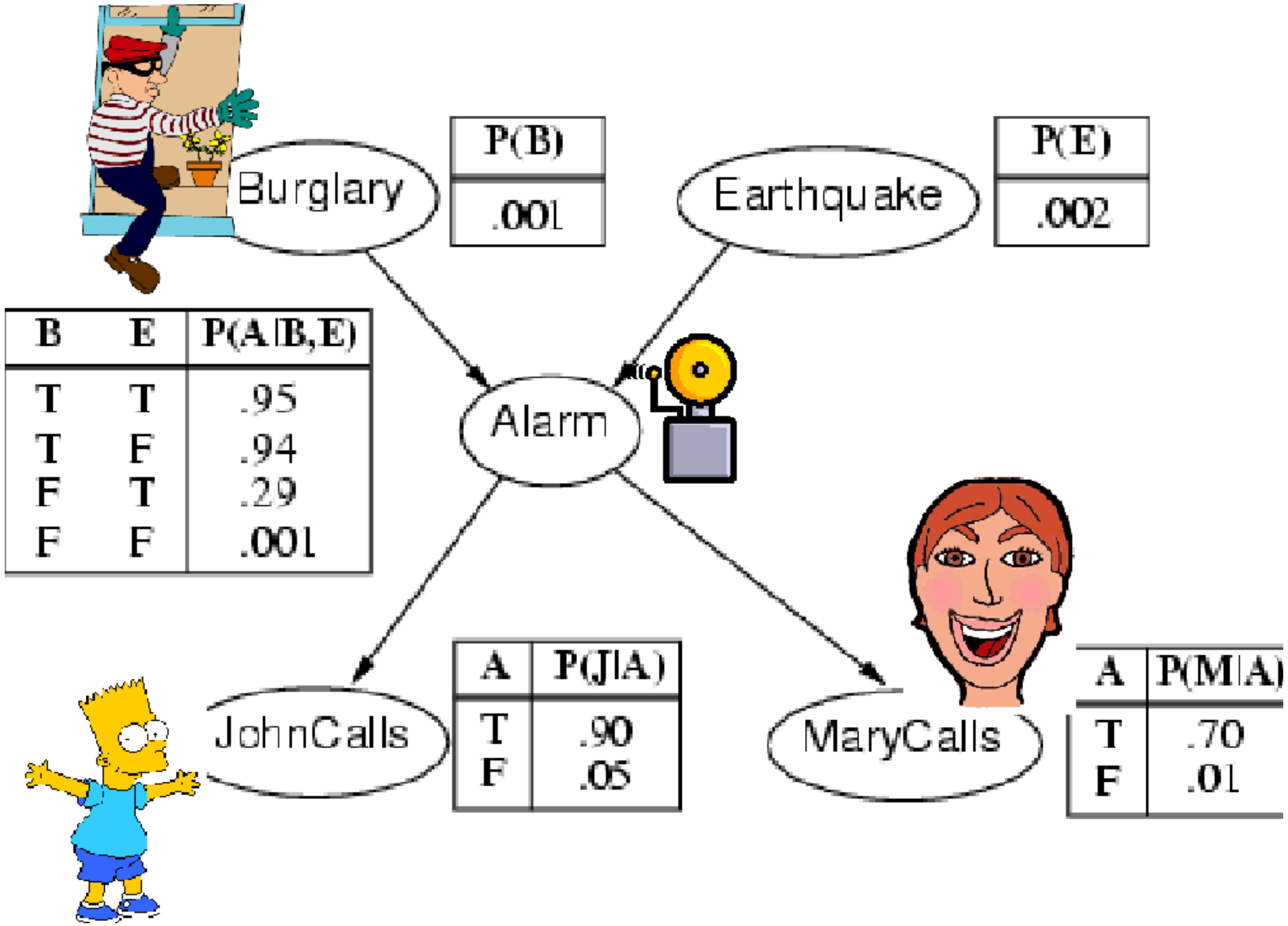
Reti Bayesiane



Reti Bayesiane



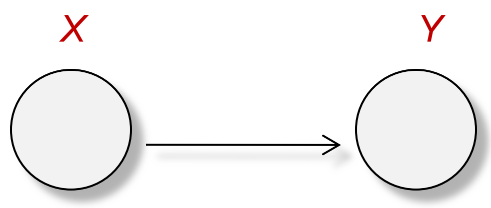
Reti Bayesiane



Reti Bayesiane

La *componente quantitativa* è costituita da un insieme di *tabelle di probabilità condizionale*:

- Ogni nodo (variabile) ha associata una *Conditioned Probability Table* (CPT),
- La CPT di ogni variabile traduce l'impatto dei genitori sulla variabile stessa.



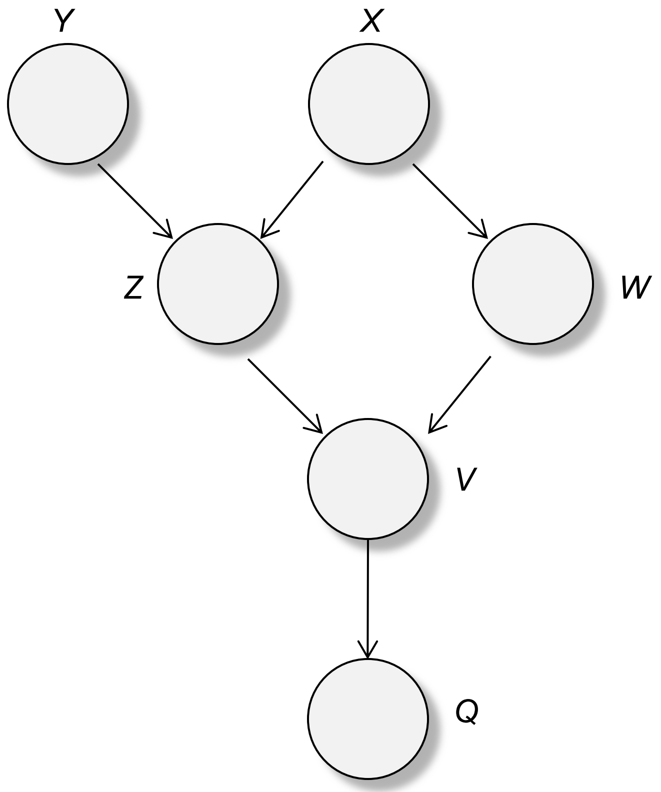
La variabile *Y* ha associata una CPT che traduce l'influenza del suo genitore *X* su di essa.

$$Parents(Y) = X$$

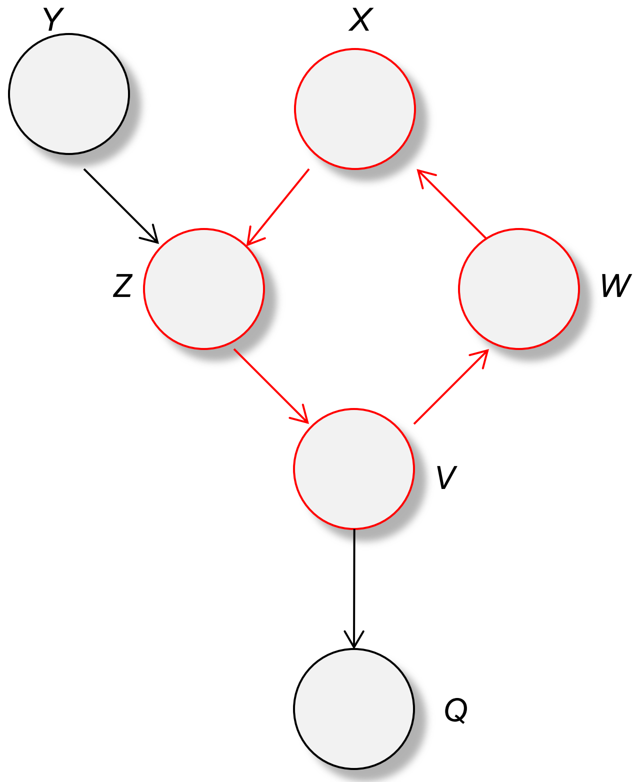
$$Parents(X) = \emptyset.$$

Reti Bayesiane

È mandatorio che il *grafo orientato* non contenga cicli ovvero che la *Rete Bayesiana* sia un *Directed Acyclic Graph* (DAG); non è possibile che una variabile influenzi (causi) se stessa.



Directed Acyclic Graph → Rete Bayesiana



Directed Graph → NO Rete Bayesiana

Reti Bayesiane

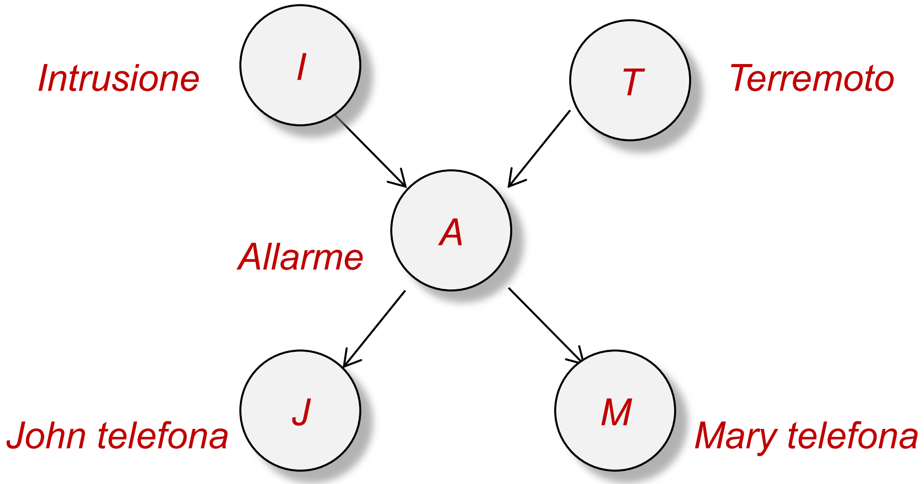
Di seguito riprendiamo il modello di Rete Bayesiana dell'esempio dell'allarme.

Concentriamo l'attenzione sulla componente qualitativa, osservando quanto segue:

i. *Intrusione e Terremoto influenzano direttamente la probabilità che l'Allarme suoni,*

ii. *Il fatto che John o Mary telefonino dipende solo dall'Allarme,*

iii. *Il modello grafico rappresenta l'assunzione implicita che John e Mary non percepiscono il tentativo di Intrusione così come non percepiscono lievi scosse di Terremoto ed inoltre John e Mary non si consultano per decidere se telefonarci al lavoro oppure no.*



Reti Bayesiane

Consideriamo la *componente quantitativa* della *Rete Bayesiana*, ovvero l'insieme di *Conditional Probability Tables* (CPTs) associate ad ogni nodo.

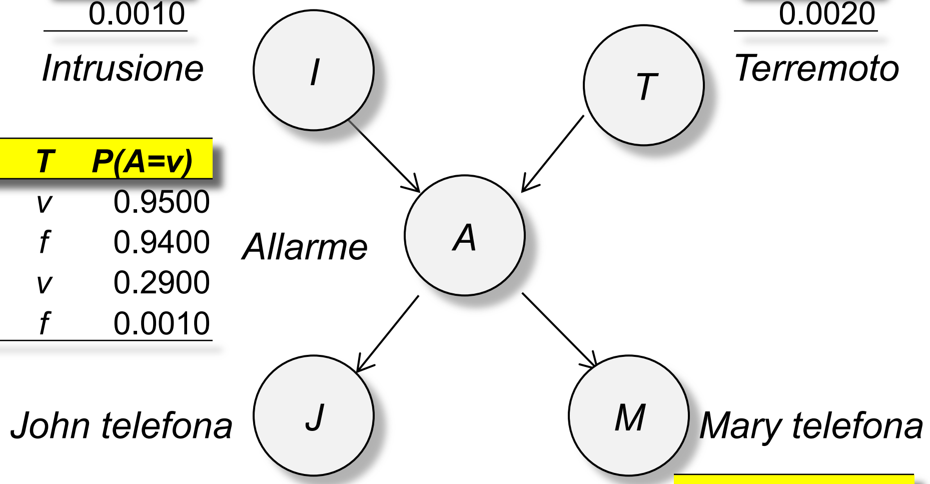
i. *ogni nodo (variabile) ha associata una CPT,*

ii. *la CPT descrive la probabilità condizionata della variabile dato lo stato delle sue variabili genitore,*

iii. *ogni riga della CPT somma ad uno,*

iv. *la CPT di una variabile Booleana con k variabili genitore Booleane contiene* 2^k *valori di probabilità che possono essere specificati indipendentemente,*

v. *una variabile senza genitori (nodo radice) ha una CPT costituita da una sola riga che contiene i valori di probabilità a priori per ogni possibile valore che la variabile può assumere.*



P(I=v)
0.0010

P(T=v)
0.0020

I	T	P(A=v)
v	v	0.9500
v	f	0.9400
f	v	0.2900
f	f	0.0010

A	P(J=v)
v	0.9000
f	0.0500

A	P(M=v)
v	0.7000
f	0.0100

Semantica delle reti Bayesiane

La *semantica* delle *Reti Bayesiane* può essere presentata e compresa in base alle seguenti chiavi di lettura:

- la *rete rappresenta* una *distribuzione congiunta* di probabilità
- la *rete codifica* un insieme di *relazioni di indipendenza condizionale*.

Le due chiavi di lettura sono semanticamente equivalenti anche se :

i. la prima risulta particolarmente importante per comprendere come sia possibile *costruire un Modello di Rete Bayesiana*,

ii. la seconda ricopre importanza centrale per comprendere come *progettare ed implementare procedure di inferenza*.

Semantica delle reti Bayesiane

Semantica delle Reti Bayesiane: *rappresentazione della distribuzione congiunta*

Ogni *Rete Bayesiana* costituisce una descrizione completa del dominio che rappresenta e pertanto ogni elemento della distribuzione di probabilità congiunta può essere calcolato a partire dall'informazione contenuta nella rete.

Un generico elemento della distribuzione di probabilità congiunta è associato ad una realizzazione congiunta delle variabili (nodi) presenti nella rete:

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$

che rappresenteremo in forma abbreviata come segue

$$P(x_1,\ldots,x_n)$$

Ogni elemento della distribuzione congiunta di probabilità può essere calcolato sfruttando la seguente *formula di fattorizzazione della distribuzione congiunta di probabilità*.

.

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

Semantica delle reti Bayesiane

Semantica delle Reti Bayesiane: *rappresentazione della distribuzione congiunta*

Ogni *Rete Bayesiana* costituisce una descrizione completa del dominio che rappresenta e pertanto ogni elemento della distribuzione di probabilità congiunta può essere calcolato a partire dall'informazione contenuta nella rete.

Un generico elemento della distribuzione di probabilità congiunta è associato ad una realizzazione congiunta delle variabili (nodi) presenti nella rete:

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$

che rappresenteremo in forma abbreviata come segue

$$P(x_1,\ldots,x_n)$$

Ogni elemento della distribuzione congiunta di probabilità può essere calcolato sfruttando la seguente *formula di fattorizzazione della distribuzione congiunta di probabilità*.

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

.

Semantica delle reti Bayesiane

Si faccia attenzione alla notazione

$$parents(X_i)$$

che sta a rappresentare la realizzazione (congiunta) delle variabili genitore di Xi

$$Parents(X_i)$$

Ogni elemento della distribuzione congiunta di probabilità è rappresentato tramite il prodotto (fattorizzazione) delle opportune componenti delle CPTs.

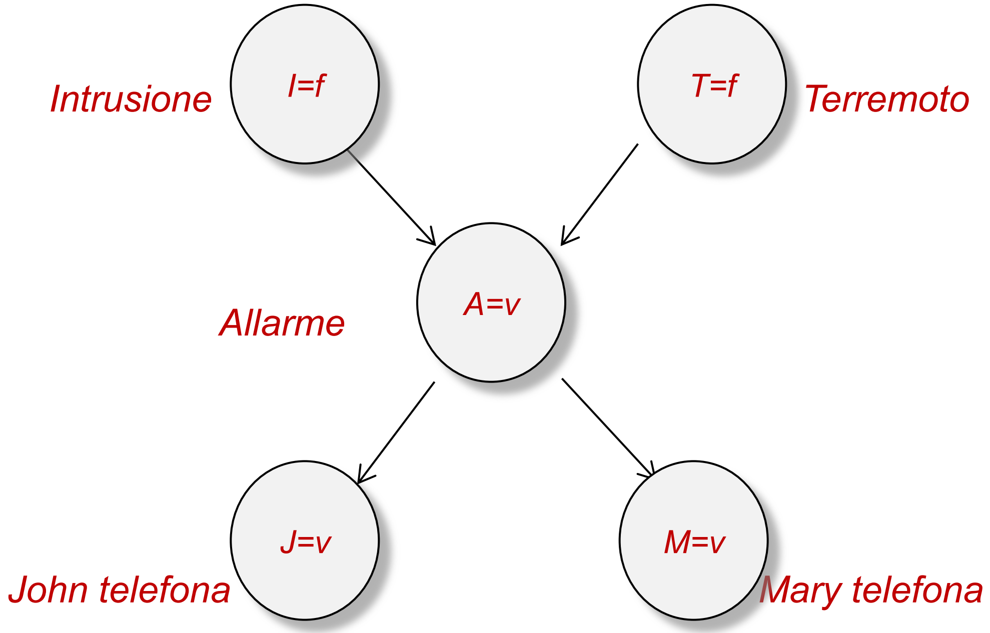
Le CPTs costituiscono quindi una rappresentazione decomposta della distribuzione di probabilità congiunta.

Semantica delle reti Bayesiane

Esempio: computiamo la probabilità che l'*Allarme* stia suonando, sapendo che non si tratta di un tentativo di Intrusione, non si è verificata nessuna scossa di *Terremoto*, e che *John telefona* e *Mary telefona*.

In particolare, abbiamo il seguente evento congiunto:

- *Intrusione* = *falso* (*I=f*)
- *Terremoto* = *falso* (*T=f*)
- *Allarme* = *vero* (*A=v*)
- *John telefona* = *vero* (*J=v*)
- *Mary telefona* = *vero* (*M=v*)



e desideriamo conoscerne la probabilità, ovvero desideriamo computare il valore della seguente probabilità congiunta

$$P(I = f, T = f, A = v, J = v, M = v)$$

Semantica delle reti Bayesiane

Grazie alla specifica istanza di fattorizzazione applicata al modello in esame è possibile scrivere la seguente uguaglianza:

$$P(I = f, T = f, A = v, J = v, M = v) = P(I = f)P(T = f)P(A = v \mid I = f, T = f)P(J = v \mid A = v)P(M = v \mid A = v)$$

È importante rimarcare ancora una volta che per ogni nodo, in base alla fattorizzazione, è rilevante solo la conoscenza dello stato dei suoi genitori, infatti abbiamo:

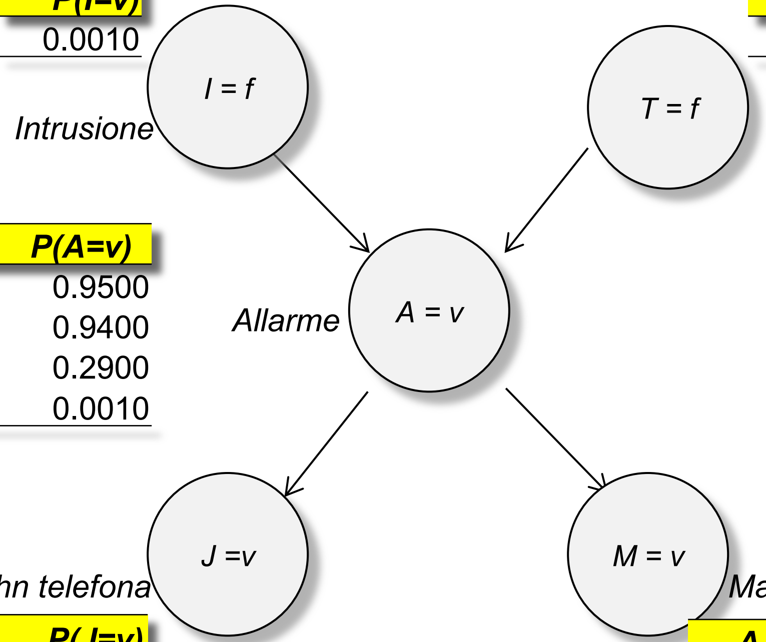
$Parents(I) = \varnothing \Rightarrow P(I = f)$

$Parents(T) = \varnothing \Rightarrow P(T = f)$

$Parents(A) = \{I, T\} \Rightarrow P(A = v \mid I = f, T = f)$

$Parents(J) = \{A\} \Rightarrow P(J = v \mid A = v)$

$Parents(M) = \{A\} \Rightarrow P(M = v \mid A = v)$



P(I=v)
0.0010

P(T=v)
0.0020

I	T	P(A=v)
v	v	0.9500
v	f	0.9400
f	v	0.2900
f	f	0.0010

A	P(J=v)
v	0.9000
f	0.0500

A	P(M=v)
v	[illegible]
f	[illegible]

Semantica delle reti Bayesiane

Sostituendo i valori di probabilità ricavati dalle CPTs possiamo scrivere:

$$P(I = f, T = f, A = v, J = v, M = v) = P(I = f)P(T = f)P(A = v \mid I = f, T = f)P(J = v \mid A = v)P(M = v \mid A = v)$$

$$= 0.9990 \cdot 0.9980 \cdot 0.0010 \cdot 0.9000 \cdot 0.7000$$

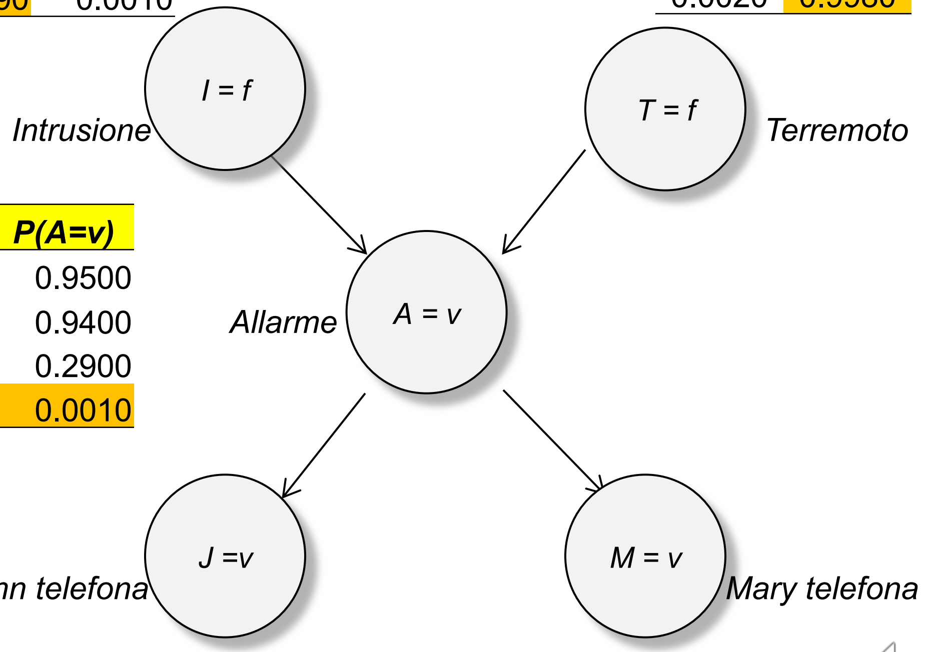
$$= 0.00062$$

Nelle lezioni precedenti abbiamo visto come la disponibilità dell'intera distribuzione di probabilità congiunta consenta di fornire risposte quantitative a tutte le possibili query che intendiamo porre sul dominio sotto studio ed analisi.

Pertanto, se una *Rete Bayesiana* rappresenta una distribuzione congiunta di probabilità allora essa può essere utilizzata per rispondere a qualsiasi query relativa al dominio che descrive, tramite il meccanismo di marginalizzazione

	P(I=v)
0.9990	0.0010

P(T=v)	
0.0020	0.9980



I	***T***	***P(A=v)***
v	*v*	0.9500
v	*f*	0.9400
f	*v*	0.2900
f	*f*	0.0010

A	***P(J=v)***
v	0.9000
f	0.0500

A	***P(M=***
v	0.700
f	0.0100