# Bayesian Reasoning:

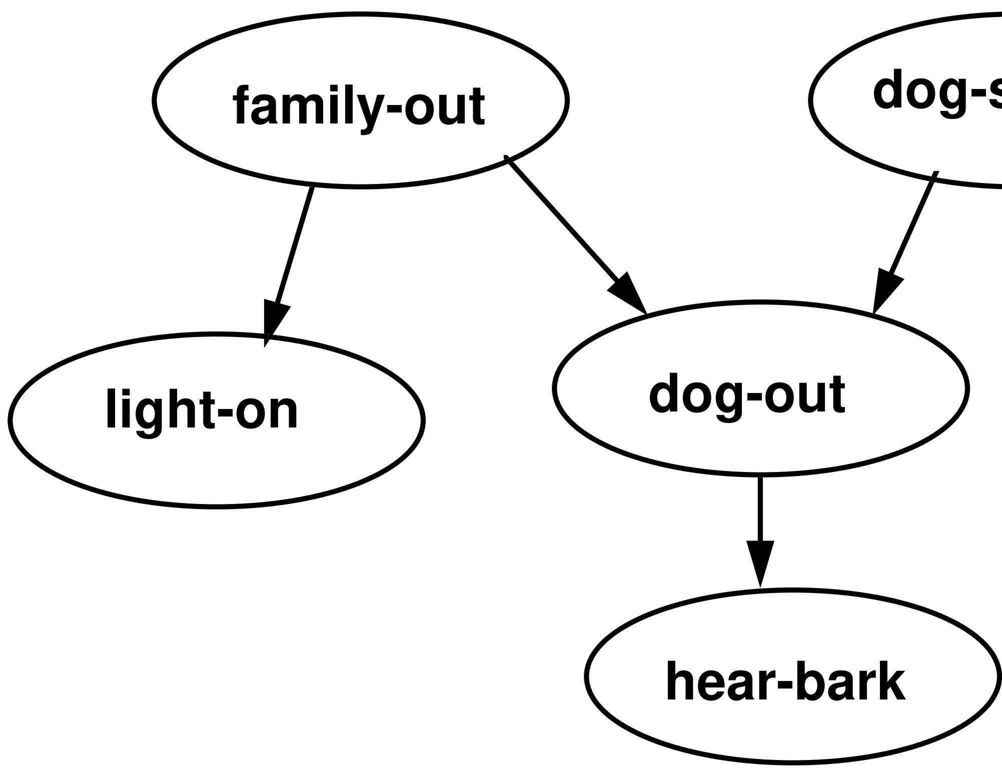
foils on Eugene Charniak's
"Bayesian Networks without Tears"
*AI Magazine, Winter 1991*

L. Morgenstern
CS 4725
April 21 & 28, 1999

Bayesian networks are used in situations where:

- we are modelling some sort of causal reasoning
- our understanding of what is going on is incomplete

Example: (all examples are from Charniak)




We can use this graph for:

- <u>prediction</u>: reasoning about what will happen (if family leaves, they'll put the dog out)
- <u>infer causes; explain</u>: if the light is on and the dog is out, the family is probably out

**In a Bayesian network,**
**the nodes represent states of affairs (random variables);**

**In this case, the random variables are Boolean,**
**but often they can have a range of values**

**the links represent causal connections**

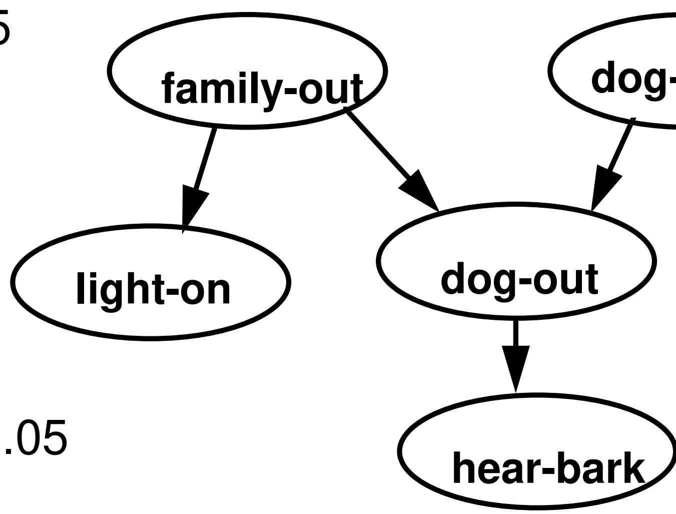
**In a Bayesian network, one must give**

- **the prior probabilities of all root nodes**
- **the conditional probabilities of all non-root nodes, given all possible combinations of their direct predecessors**

The following is a fully specified Bayesian network:

P(fo) = .15

P(ds) = .01




P(lo| fo) = .6
P(lo | ~fo) = .05

P(do | fo ds) = .99
P(do | fo ~ds) = .90
P(do | ~fo ds) = .97
P(do | ~fo ~ds) = .3

This says, e.g., that the probability of the dog being out if the family is not out and the dog is sick is .97

P(hb | do) = .7
P(hb | ~do) = ..01

What are the benefits of using a Bayesian network?

1. Need to specify fewer conditional probabilities.
Without Bayesian network, must specify full joint probability table, 2**n (-1) values. For this network, we need 10 as opposed to 31 values; for larger networks of similar structure, savings can be and usually are much greater.
Exception: worst-case Bayesian network, a fully connected graph.

2. Easier to calculate.
Calculating probabilities in general is painful.
Bayesian networks generally reduce the problem.

What makes this saving possible?
Independence assumptions

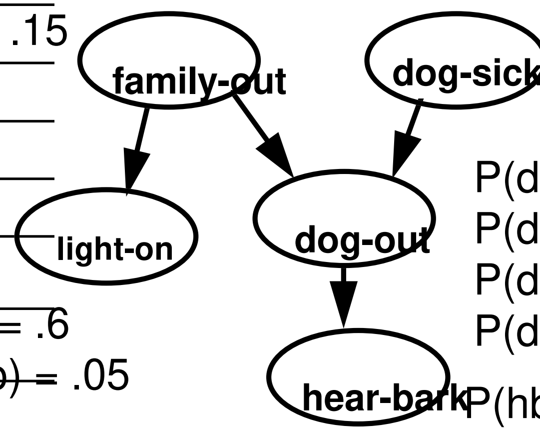
## 1. Need to specify fewer conditional probabilities

**Compare joint probability distribution table to Bayesian network .....**

| | |
|---|---|
| P(fo, lo, do, ds, hb) | |
| P(fo, lo, do, ds, ~hb) | |
| P(fo, lo, do, ~ds, hb) | |
| P(fo, lo, do, ~ds, ~hb) | |
| P(fo, lo, ~do, ds, hb) | |
| P(fo, lo, ~do, ds, ~hb) | |
| P(fo, lo, ~do, ~ds, hb) | |
| P(fo, lo, ~do, ~ds, ~hb) | |
| ...................... | |
| P(~fo, ~lo, do, ds, hb) | |
| P(~fo, ~lo, do, ds, ~hb) | |
| P(~fo, ~lo, do, ~ds, hb) | |
| P(~fo, ~lo, do, ~ds, ~hb) | |
| P(~fo, ~lo, ~do, ds, hb) | |
| P(~fo, ~lo, ~do, ds, ~hb) | |
| P(~fo, ~lo, ~do, ~ds, hb) | |
| P(~fo, ~lo, ~do, ~ds, ~hb) | |

P(fo) = .15

P(ds) = .01




P(lo| fo) = .6
P(lo | ~fo) = .05

P(do | fo ds) = .99
P(do | fo ~ds) = .90
P(do | fo ~ds) = .97
P(do | ~fo ~ds) = .3

P(hb | do) = .7
P(hb | ~do) = ..01

## 2. Easier to calculate probabilities in Bayesian networks

To see this, let's go through some examples ....

First, the basics:

**The Chain Rule:**

P(a1 .... an) = P(a1| a2 .... an) * P(a2|a3 .... an) * P(an-1|an) * P(an)

**Bayes's Rule:**

$$P(h \mid e) = \frac{P(e \mid h) * P(h)}{P(e)}$$

Note: we are usually given the evidence,
and are interested in finding the probability
of the hypothesis given the evidence.
So we'd like to figure out P(h | e)

**More basics:**

We have the following variation of the chain rule

**The Chain Rule:**
P(a1 .... an) = P(a1| a2 .... an) * P(a2|a3 .... an) * P(an-1|an) * P(an)

**The variation:**
P(a1 ... an) = P(a1 | dp(a1) * P(a2 | dp(a2) ............ * P(an | dp(an))

dp(ai) are the direct predecessors of random variable ai.
In a Bayesian network, the direct predecessor of a node are the parents of the node.
In a standard joint probability distribution, we can define
the direct predecessors of a node as follows:
Assume n random variables. Let R = the set of random variables - ai.
Let DP be the set such that P(ai | DP) is the same no matter what values
we have for the variables in R - DP.
Then DP is the set of direct predecessors of ai.

In other words, given DP, ai is independent of all variables in R - DP.

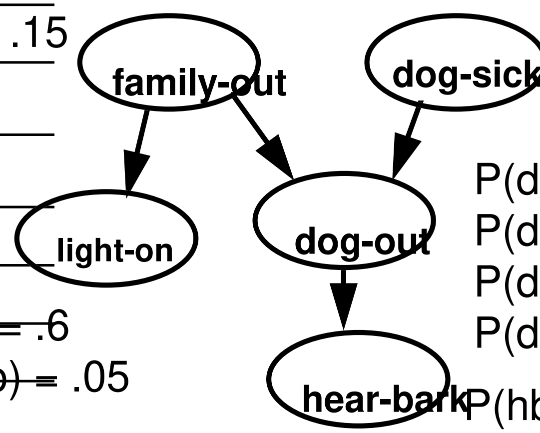
| | |
|---|---|
| P(fo, lo, do, ds, hb) | v |
| P(fo, lo, do, ds, ~hb) | v |
| P(fo, lo, do, ~ds, hb) | v |
| P(fo, lo, do, ~ds, ~hb) | v |
| P(fo, lo, ~do, ds, hb) | |
| P(fo, lo, ~do, ds, ~hb) | |
| P(fo, lo, ~do, ~ds, hb) | |
| P(fo, lo, ~do, ~ds, ~hb) | |
| ...................... | |
| P(~fo, ~lo, do, ds, hb) | |
| P(~fo, ~lo, do, ds, ~hb) | |
| P(~fo, ~lo, do, ~ds, hb) | |
| P(~fo, ~lo, do, ~ds, ~hb) | |
| P(~fo, ~lo, ~do, ds, hb) | |
| P(~fo, ~lo, ~do, ds, ~hb) | |
| P(~fo, ~lo, ~do, ~ds, hb) | |
| P(~fo, ~lo, ~do, ~ds, ~hb) | |

Here, we can see from the table that fo and lo are the direct predecessors of do, because P(do | fo lo) is the same no matter what the values of ds and hb

Here, we can see from
the table that fo and lo are
the direct predecessors of do,
because P(do | fo lo) is the
same no matter what the values
of ds and hb

| | |
|---|---|
| P(fo, lo, do, ds, hb) | v |
| P(fo, lo, do, ds, ~hb) | v |
| P(fo, lo, do, ~ds, hb) | v |
| P(fo, lo, do, ~ds, ~hb) | v |
| P(fo, lo, ~do, ds, hb) | |
| P(fo, lo, ~do, ds, ~hb) | |
| P(fo, lo, ~do, ~ds, hb) | |
| P(fo, lo, ~do, ~ds, ~hb) | |
| ..................... | |
| P(~fo, ~lo, do, ds, hb) | |
| P(~fo, ~lo, do, ds, ~hb) | |
| P(~fo, ~lo, do, ~ds, hb) | |
| P(~fo, ~lo, do, ~ds, ~hb) | |
| P(~fo, ~lo, ~do, ds, hb) | |
| P(~fo, ~lo, ~do, ds, ~hb) | |
| P(~fo, ~lo, ~do, ~ds, hb) | |
| P(~fo, ~lo, ~do, ~ds, ~hb) | |

The same information is
easily visible in the Bayesian net




P(fo) = .15

P(ds) = .01

P(lo| fo) = .6
P(lo | ~fo) = .05

P(do | fo ds) = .99
P(do | fo ~ds) = .90
P(do | fo ~ds) = .97
P(do | ~fo ~ds) = .3

P(hb | do) = .7
P(hb | ~do) = ..01

## We use these basic rules to calculate probablilities

**For example,**

P(fo, lo, ds, do, hb) =

P(fo) * P(lo | fo) * P(ds) * P(do | fo ds) * P(hb | do)

**What about when we are only interested in a subset of the variables?**

We sum over the possible values of the other variables

**For example,**

P(lo do) =

$\Sigma_{hb} \Sigma_{fo} \Sigma_{ds}$ P(lo,do)

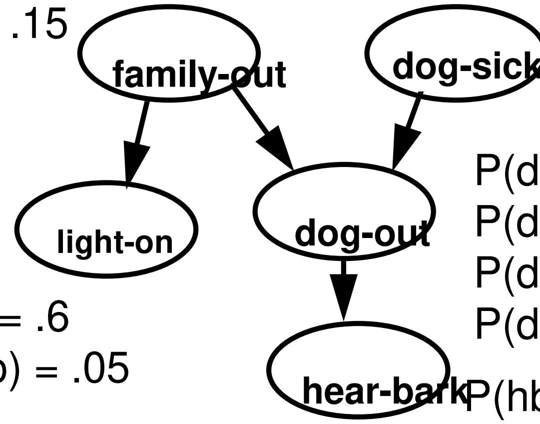
=

P(lo | fo) * P(do | fo ds) * P(hb | do) * P(fo) * P(ds)
P(lo | fo) * P(do | fo ds) * P(hb | do) * P(fo) * P(~ds)
P(lo | fo) * P(do | fo ds) * P(hb | do) * P(~fo) * P(ds)
P(lo | fo) * P(do | fo ds) * P(hb | do) * P(~fo) * P(~ds)
P(lo | fo) * P(do | fo ds) * P(~hb | do) * P(fo) * P(ds)
P(lo | fo) * P(do | fo ds) * P(~hb | do) * P(fo) * P(~ds)
P(lo | fo) * P(do | fo ds) * P(~hb | do) * P(~fo) * P(ds)
P(lo | fo) * P(do | fo ds) * P(~hb | do) * P(~fo) * P(~ds)

**Example of a calculation using the network:**

P(fo) = .15

P(ds) = .01




P(lo| fo) = .6
P(lo | ~fo) = .05

P(do | fo ds) = .99
P(do | fo ~ds) = .90
P(do | fo ~ds) = .97
P(do | ~fo ~ds) = .3

P(hb | do) = .7
P(hb | ~do) = ..01

**What is P(fo | lo) ?**

By Bayes Rule:

$$P(fo \mid lo) = \frac{P(lo \mid fo) * P(fo)}{P(lo)}$$

We can read off P(lo | fo) and P(fo) straight from the net.
But we still need to calculate P(lo).
To do this, we note that
P(fo | lo) + P(~fo | lo) = 1

Thus, by Bayes Rule:

$$\frac{P(lo \mid fo) * P(fo)}{P(lo)} + \frac{P(lo \mid {\sim}fo) * P({\sim}fo)}{P(lo)} = 1$$

So (multiplying both sides by P(lo)), we get:
P(lo | fo) * P(fo) + P(lo | ~fo)*P(~fo) = P(lo)
We now can just read the values off the net and calculate.

$$\frac{(.6)(.15)}{(.6)(.15) + (.05)(.85)} = .09/.1325 = .57$$

**Example of a more complicated calculation:**

**Calculate P(fo | lo do)**

**1. without wisely using the independence assumptions:**

First, note that by the chain rule,

P(fo lo do) = P(fo | lo do) * P(lo do)

*The chain rule states that*

*P(fo lo do) = P(fo | lo do) * P(lo| do) * P(do)*

*but note that also by the chain rule,*

*P(lo do)= P(lo | do) * P(do)*

So, P(fo | lo do) = P(fo lo do) / P(lo do)

Now, P(fo lo do) =

P(fo) * P(lo | fo) * P(do | fo ds) * P(hb | do) * P(ds) +

P(fo) * P(lo | fo) * P(do | fo ~ds) * P(hb | do) * P(~ds) +

P(fo) * P(lo | fo) * P(do | fo ds) * P(~hb | do) * P(ds) +

P(fo) * P(lo | fo) * P(do | fo ~ds) * P(~hb | do) * P(~ds)

We can factor this to get:

P(fo)* P(lo | fo)

[P(do|fo ds) * P(hb | do) * P(ds)

P(do|fo ~ds) * P(hb | do) * P(~ds)

P(do|fo ds) * P(~hb | do) * P(ds)

P(do|fo ~ds) * P(~hb | do) * P(~ds)

## Calculation 1, continued:

We have:

P(fo)* P(lo | fo)

[P(do|fo ds) * P(hb | do) * P(ds) 1

P(do|fo ~ds) * P(hb | do) * P(~ds) 2

P(do|fo ds) * P(~hb | do) * P(ds) 3

P(do|fo ~ds) * P(~hb | do) * P(~ds) 4

term 1 + 3 can be factored:

= P(do | fo ds) * P(ds) (P(hb | do) + P(~hb | do))

= P(do| fo ds) * P(ds) * 1

term 2 +4 can be factored:

= P(do | fo ~ds) * P(~ds) (P(hb | do) + P(~hb | do)

= P(do | fo ~ds) * P(~ds)

So we have

P(fo)* P(lo | fo) *( P(do | fo ds) * P(ds) + P(do | fo ~ds) * P(~ds))

**Calculation 1, continued:**

So far, we have
$P(fo) * P(lo \mid fo) *( P(do \mid fo\ ds) * P(ds) + P(do \mid fo\ \sim ds) * P(\sim ds))$

We can simplify this further:

$P(ds) = P(ds \mid fo)$ because ds and fo are independent
$P(do \mid fo\ ds) = P(do\ ds \mid fo)$

So we get
$P(fo) * P(lo \mid fo) *( P(do\ ds \mid fo) * P(ds \mid fo) + P(do\ \sim ds \mid fo) * P(\sim ds \mid fo))$
$= P(fo) * P(lo \mid fo) * P(do \mid fo)$ (chain rule, plus identity)

**Now, wasn't this complicated?**

## Calculation 2: Using independence assumptions (from the beginning)

$$P(fo \mid lo\ do) = \frac{P(lo\ do \mid fo) * P(fo)}{P(lo\ do)}$$

Now, lo and do are independent given fo,
so P(lo do | fo) = P(lo | fo) * P*(do | fo)

So we get

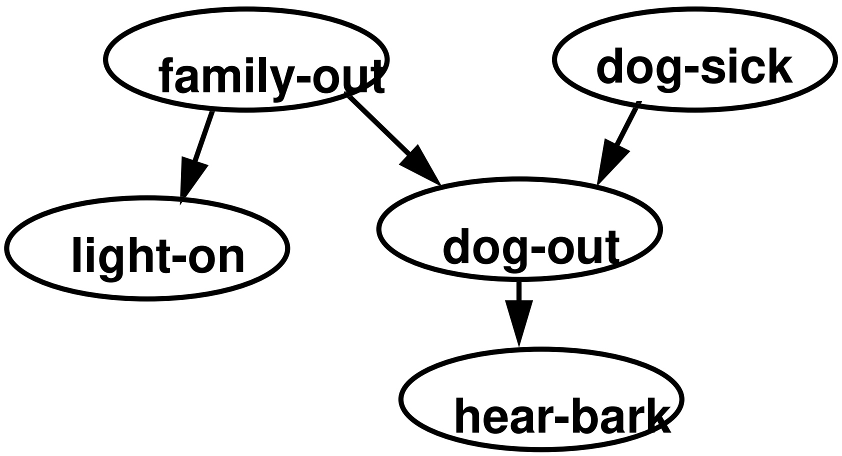
$$\frac{P(lo \mid fo) * P(do \mid fo) * P(fo)}{P(lo\ do)}$$

**much shorter than the last calculation!**

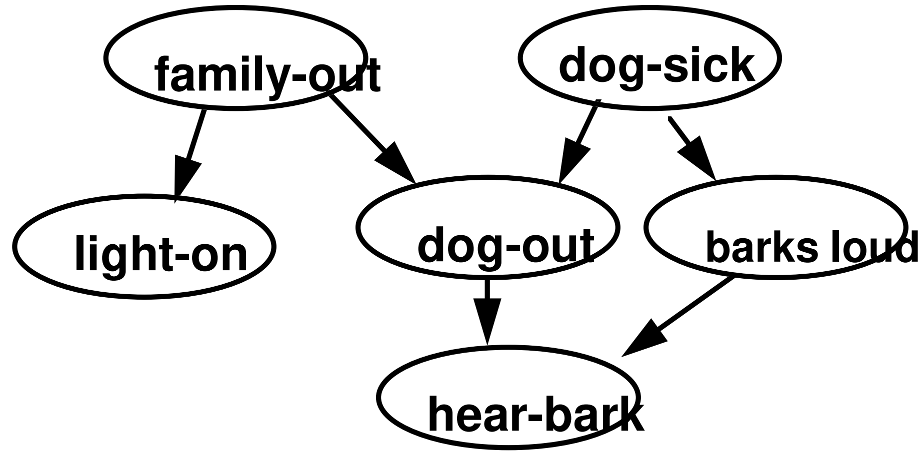
**This works because the graph is singly connected; there is only one path between any two nodes; otherwise it would be multiply connected.**

## Multiply vs. singly connected graphs

**Singly connected:**




**Multiply connected:**



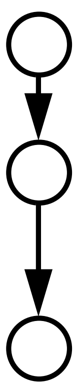

**Singly connected: between any two nodes, only 1 path**
**Multiply connected: may be more than 1 path between 2 nodes**
**ruins independence assumptions; now hear bark is <u>not</u> independent of dog sick given dog out.**

**Independence assumptions:**

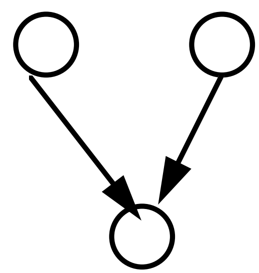
- **In general, we talk of random variable a being independent of or dependent on random variable b relative to some body of evidence.**

- **a is dependent on b given evidence E = [e1, ...., en] if there is a d-connecting path from a to be given E.**
  **If a is not dependent on b given E, a is independent of b given E.**

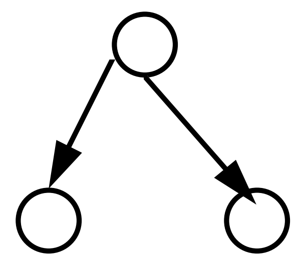
- **A path from q to r is d-connecting with respect to the evidence nodes E if every interior node n is the path has the property that either**
  1. **it is linear or diverging and not a member of E**
     **or**
  2. **it is converging, and either n or one of its descendents is in E**



**Linear**

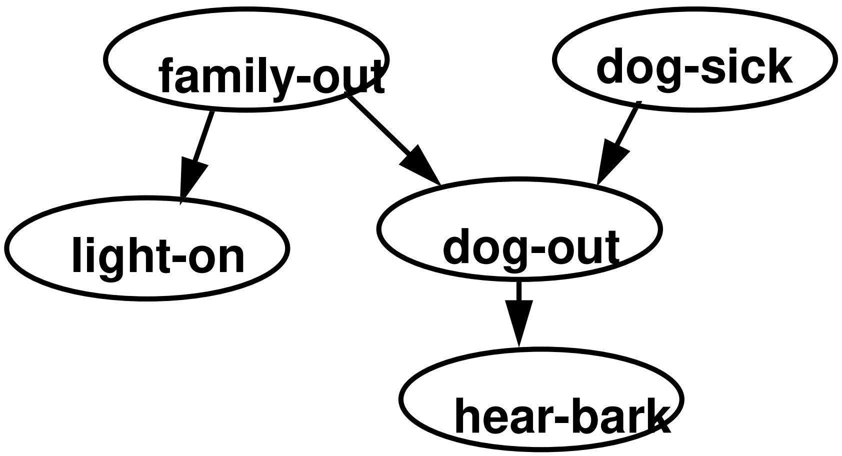


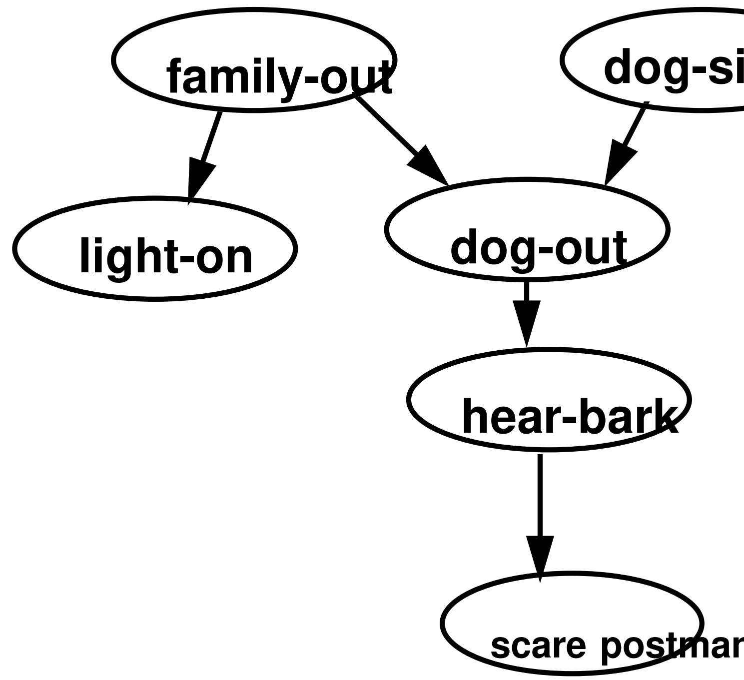
**Converging**



**Diverging**

## Examples:




**d-connecting path from**
**dog-sick to hear-bark wrt to dog-out**
**since every interior node (dog-out)**
**is converging and a member of E (dog-out)**

**no d-connecting path from dog-out to scare-postman wrt hear-bark;**
**every interior node (hear-bark) is linear but is a member of E.**