Probabilistic Machine Learning:
6. Belief Networks: basic notation

Tomasz Kajdanowicz, Przemysław Kazienko (substitution)

Department of Computational Intelligence
Wroclaw University of Technology

HR EXCELLENCE IN RESEARCH

Wrocław University
of Science and Technology

The presentation was inspired by Chapter 2 and 3 of D. Barber book "Bayesian Reasoning and Machine Learning", 2012.

Already covered

We have covered:

- inferring a distribution over a discrete variable drawn from a finite hypothesis space
- inferring the probability that a coin shows up heads and dice has given value
- given a series of discrete observations

Let’s focus now on much more complex probabilities!

Graphs

Graph

A graph G consists of nodes (also called vertices) and edges (also called links) between the nodes. Edges may be directed (they have an arrow in a single direction) or undirected. Edges can also have associated weights. A graph with all edges directed is called a directed graph, and one with all edges undirected is called an undirected graph

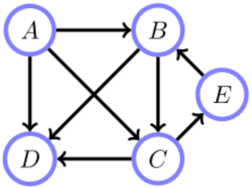


Figure: An directed graph G consists of directed edges between nodes

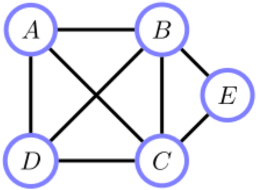


Figure: An undirected graph G consists of undirected edges between nodes

What do we need graphs for?

- to form connection between directed graphs and probability
- two variables will be independent if they are not linked by a path on the graph

Path, ancestors, descendants

Path

A *path* $A \rightarrow B$ from node A to node B is a sequence of nodes that connects A to B. That is, a path is of the form $A_0, A_1, \ldots, A_{n-1}, A_n$, with $A_0 = A$ and $A_n = B$ and each edge (A_{k-1}, A_k), $k = 1, \ldots, n$ being in the graph. A directed path is a sequence of nodes which when we follow the direction of the arrows leads us from A to B.

Ancestor

In directed graphs, the nodes A such that $A \rightarrow B$ and $B \not\rightarrow A$ are the *ancestors* of B.

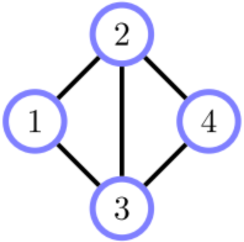
Descendant

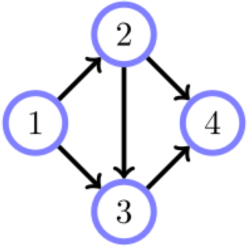
The nodes B such that $A \rightarrow B$ and $B \not\rightarrow A$ are the *descendants* of A.

Cycle

Cycle

A cycle is a directed path that starts and returns to the same node $a \to b \to \ldots \to z \to a$

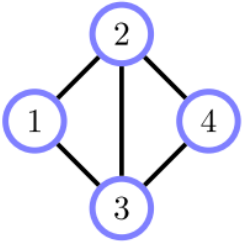


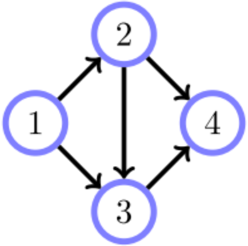


Loop

Loop

A loop is a path containing more than two nodes, irrespective of edge direction, that starts and returns to the same node.

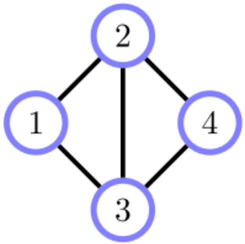


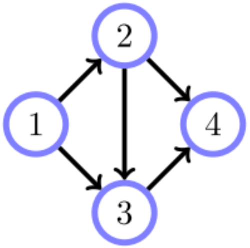


Chord

Chord

A chord is an edge that connects two non-adjacent nodes in a loop.





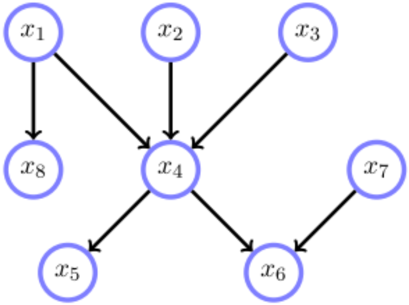
Directed Acyclic Graph (DAG)

DAG

A DAG is a graph G with directed edges (arrows on each link) between the nodes such that by following a path of nodes from one node to another along the direction of each edge no path will revisit a node.

In a DAG the ancestors of B are those nodes who have a directed path ending at B. Conversely, the descendants of A are those nodes who have a directed path starting at A.

Relationships in a DAG



parents, children, family

- the parents of x_4 are $pa(x_4) = \{x_1, x_2, x_3\}$
- the children of x_4 are $ch(x_4) = \{x_5, x_6\}$
- the family of a node is itself and its parents

Markov blanket

The Markov blanket of a node is its parents, children and the parents of its children, excluding itself. (Markov blanket of x_4 is $x_1, x_2, x_3, x_5, x_6, x_7$

Neighbour, Clique

Neighbour

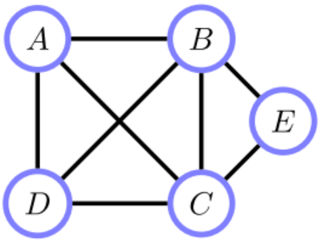
For an undirected graph G the neighbours of x, $ne(x)$ are those nodes directly connected to x.

Clique

Given an undirected graph, a clique is a fully connected subset of nodes.

- all the members of the clique are neighbours
- for a maximal clique there is no larger clique that contains the clique

Clique example



- the graph has two maximal cliques, $C1 = \{A, B, C, D\}$ and $C2 = \{B, C, E\}$
- A, B, C are fully connected, but this is a non-maximal clique
- there is a larger fully connected set: A, B, C, D

Cliques play a role in:

- modelling - describe variables that are all dependent on each other
- inference - describe sets of variables with no simpler structure that makes the relationship between them and for which no simpler efficient inference procedure is likely to exist

Connected graph

Connected graph

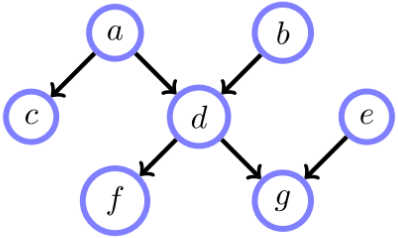
An undirected graph is connected if there is a path between every pair of nodes (i.e. there are no isolated islands). For a graph which is not connected, the connected components are those subgraphs which are connected.

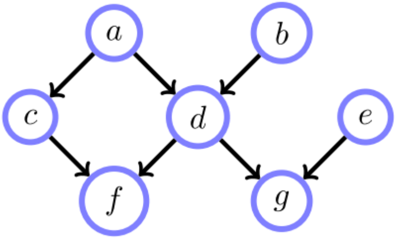
Singly Connected Graph

A graph is *singly connected* if there is only one path from any node A to any other node B. Otherwise the graph is multiply connected (for directed and undirected graphs)

An alternative name for a singly connected graph is a tree. A multiply-connected graph is also called loopy.

Singly, multiply connected

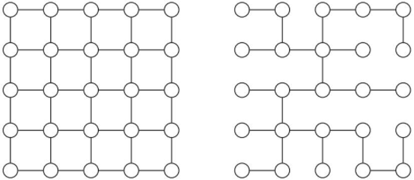




Spanning Tree

Spanning Tree

A spanning tree of an undirected graph G is a singly-connected subset of the existing edges such that the resulting singly-connected graph covers all nodes of G



A maximum weight spanning tree is a spanning tree such that the sum of all weights on the edges of the tree is at least as large as any other spanning tree of G.

Spanning Tree

Finding a maximal weight spanning tree

find a spanning tree with maximal weight()

- pick the next candidate edge which has the largest weight and add this to the edge set
- if this results in an edge set with cycles, then reject the candidate edge and propose the next largest edge weight
- pick the edge with the largest weight and add this to the edge set

There may be more than one maximal weight spanning tree.

Numerically Encoding Graphs

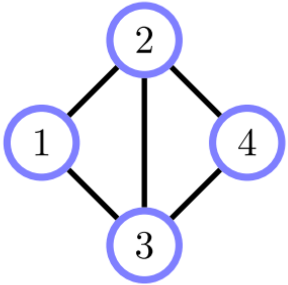
- edge list
- adjacency matrix
- clique matrix
- graph2vec etc.

Edge list

Edge list

Edge list simply lists which node-node pairs are in the graph.

$L = \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3)\}$



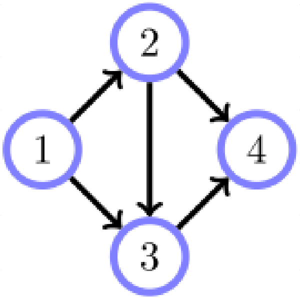
Undirected edges are listed twice, once for each direction

Adjacency matrix

Adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

where $A_{ij} = 1$ if there is an edge from node i to node j in the graph, and 0 otherwise. An undirected graph has a symmetric adjacency matrix.



Adjacency matrix

Adjacency matrices may seem wasteful since many of the entries are zero, but...

Adjacency matrix powers

For an $N \times N$ adjacency matrix A, powers of the adjacency matrix $[A^k]_{ij}$ specify how many paths there are from node i to node j in k edge hops.

Clique matrix

For an undirected graph with N nodes and maximal cliques $C_1, \ldots, C_K$ a clique matrix is an $N \times K$ matrix in which each column c_k has zeros except for ones on entries describing the clique.

$$C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$