Node-and-Edge Graphs

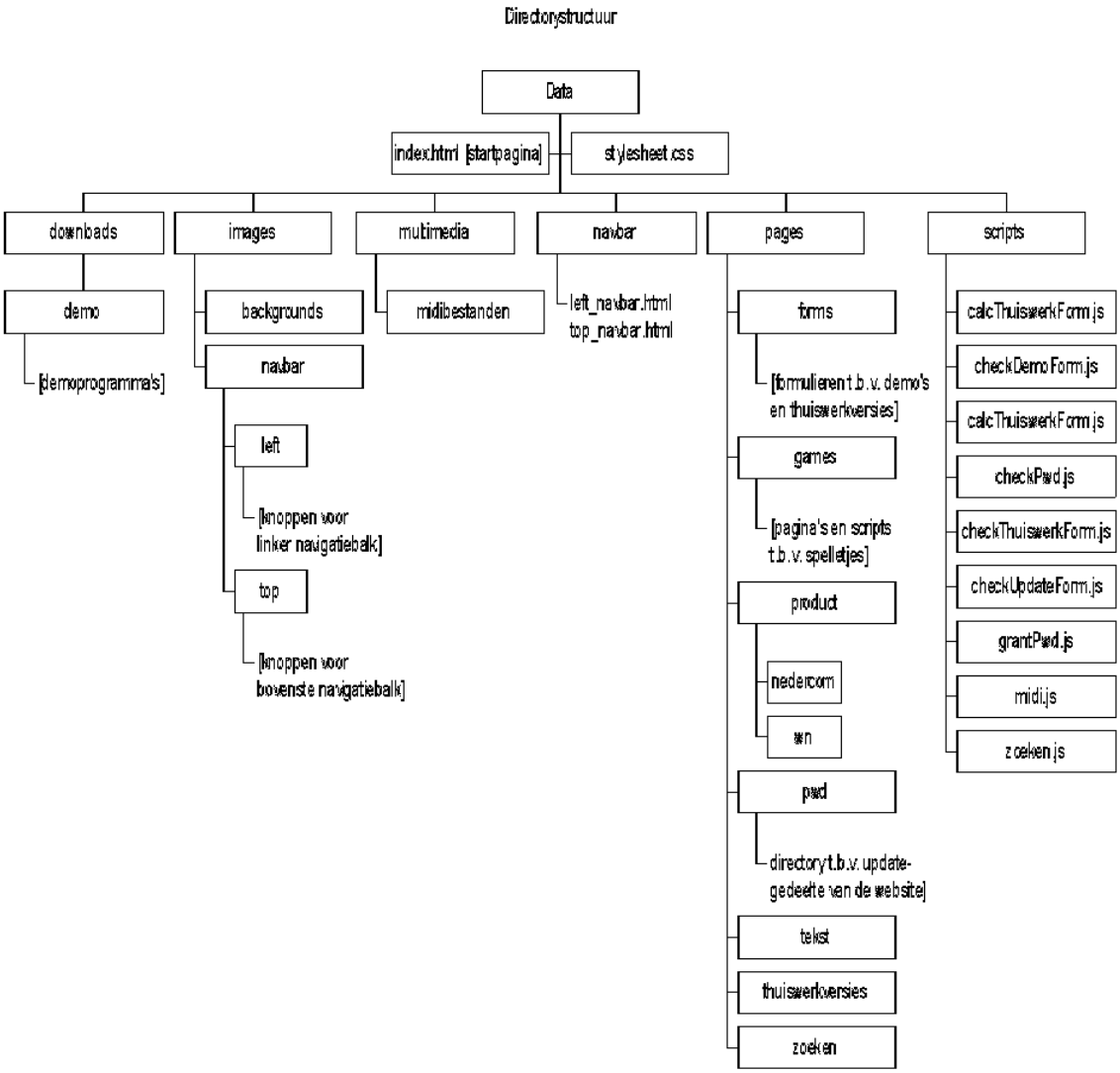
GC

Motivation: effective information transmission

- family trees
  - terminology
  - individuals
  - species
  - languages, etc.
- hierarchical structures
  - syntactic structure
  - directories in file systems
  - organizational charts
- classification
  - dualism: properties/individuals
  - decision trees
- state diagrams
- schematic maps
- internet communication
- chemical structure — beyond trees!

File Systems

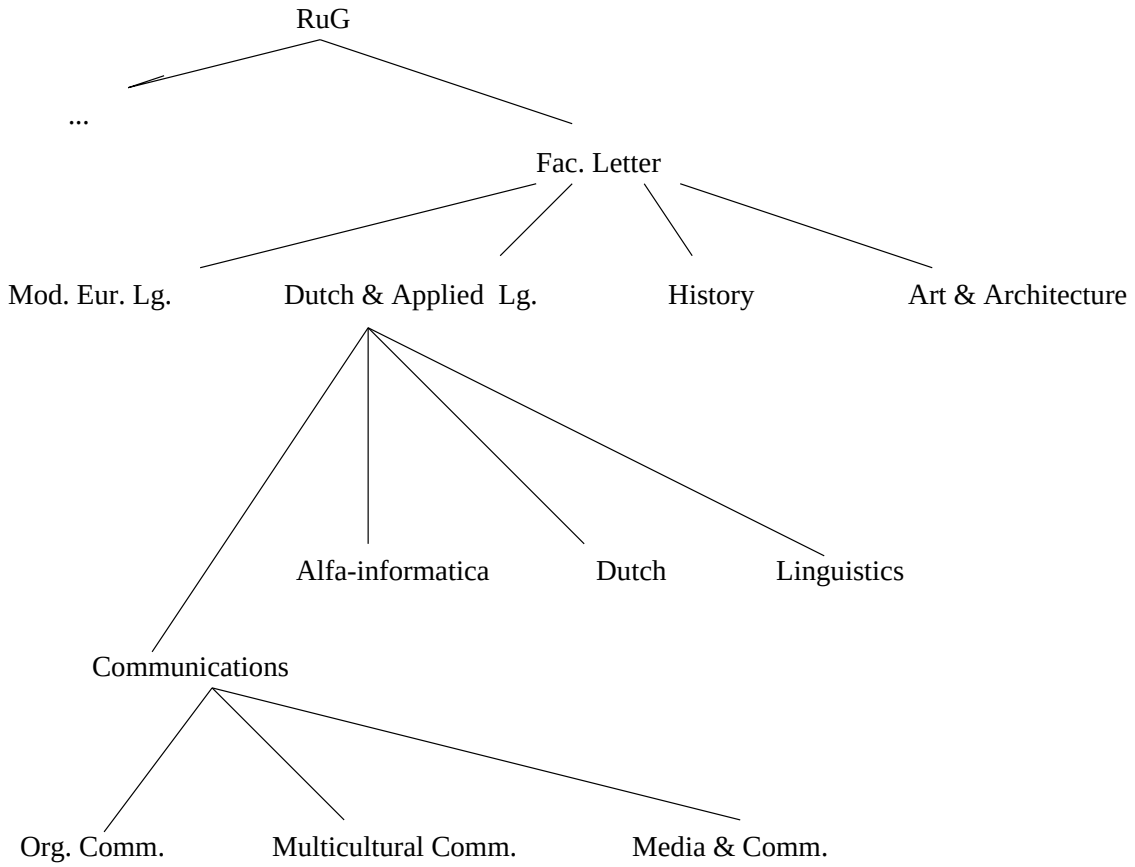
GC



RuG

Organizational Charts

GC



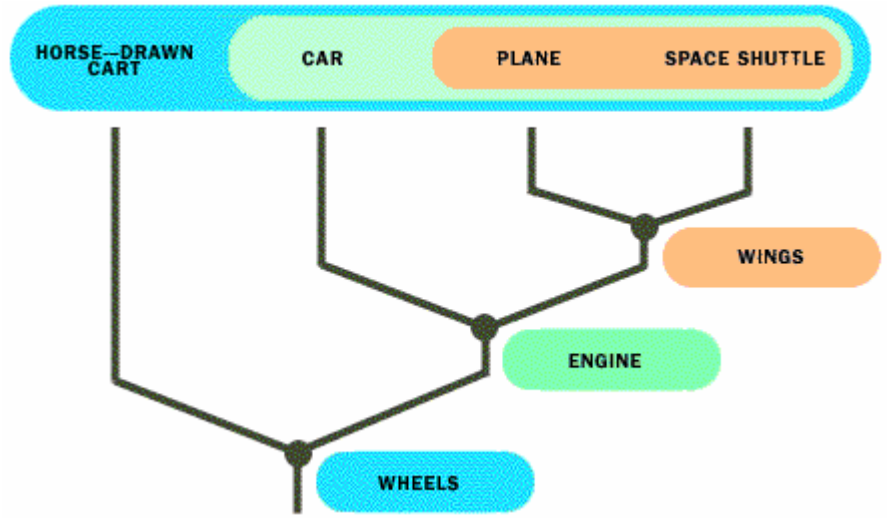
see course web site for extended example analyzing RuG "Org Chart"

- difference between 'is a part of' vs. 'is subordinate to'
- 'is subordinate to' = 'reports to' or 'is supervised by'

often an excellent way to understand an organization is to draw an org chart (based on other information — i.e., not simply copying the org chart one finds)

Classification

usually, the parts of hierarchies are distinguished by different properties (maybe not mentioned)



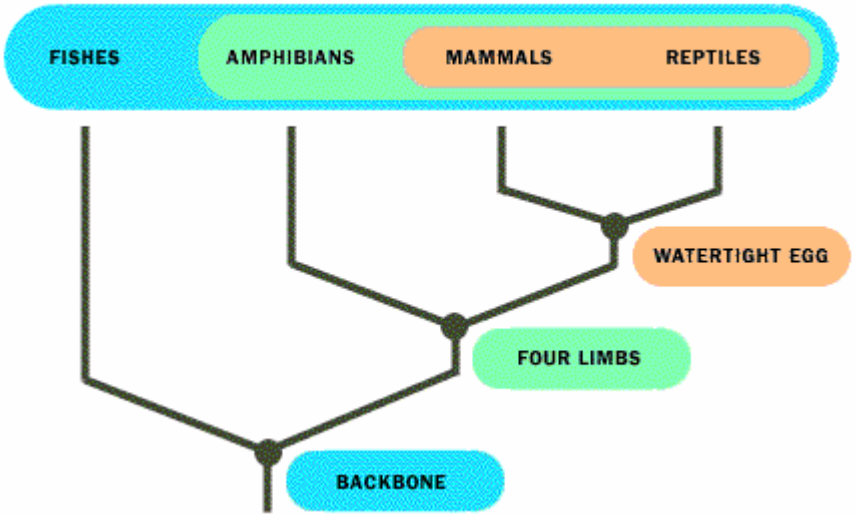
the property 'has wings' distinguishes the planes and space shuttles from the other vehicles

moving from mother to daughter corresponds to

- smaller groups of objects referred to
- larger groups of properties

Classification

similarly, the biological examples seen earlier



notice that as one moves from parent to child, properties accumulate

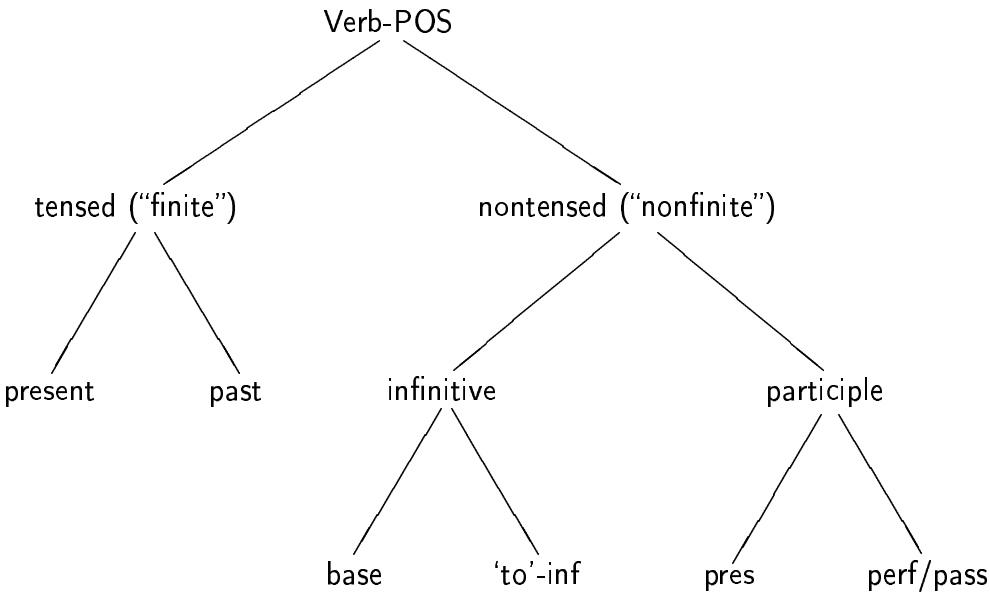
- reptiles have all the properties:
  backbones, four limbs, watertight eggs
- fish have only the one: backbones

sometimes this kind of hierarchy is referred to as an *inheritance hierarchy*

Classification of Words

We distinguished verb forms as those which showed tense (vs. those unmarked for tense), and among the latter, between the infinitive forms and the participles.

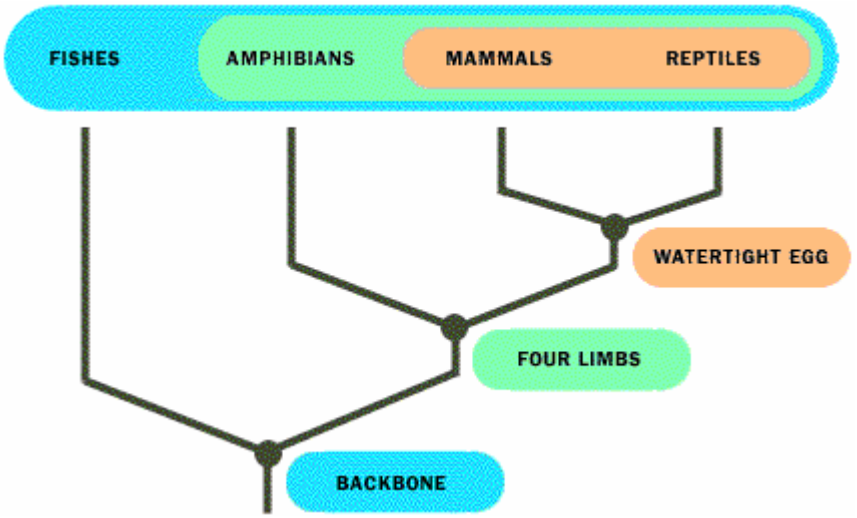
These can be organized as an inheritance hierarchy as well.



"Lexical Hierarchies"

Decision Trees

if you imagine that you are trying to classify a novel object, then the inheritance hierarchy functions as a DECISION TREE



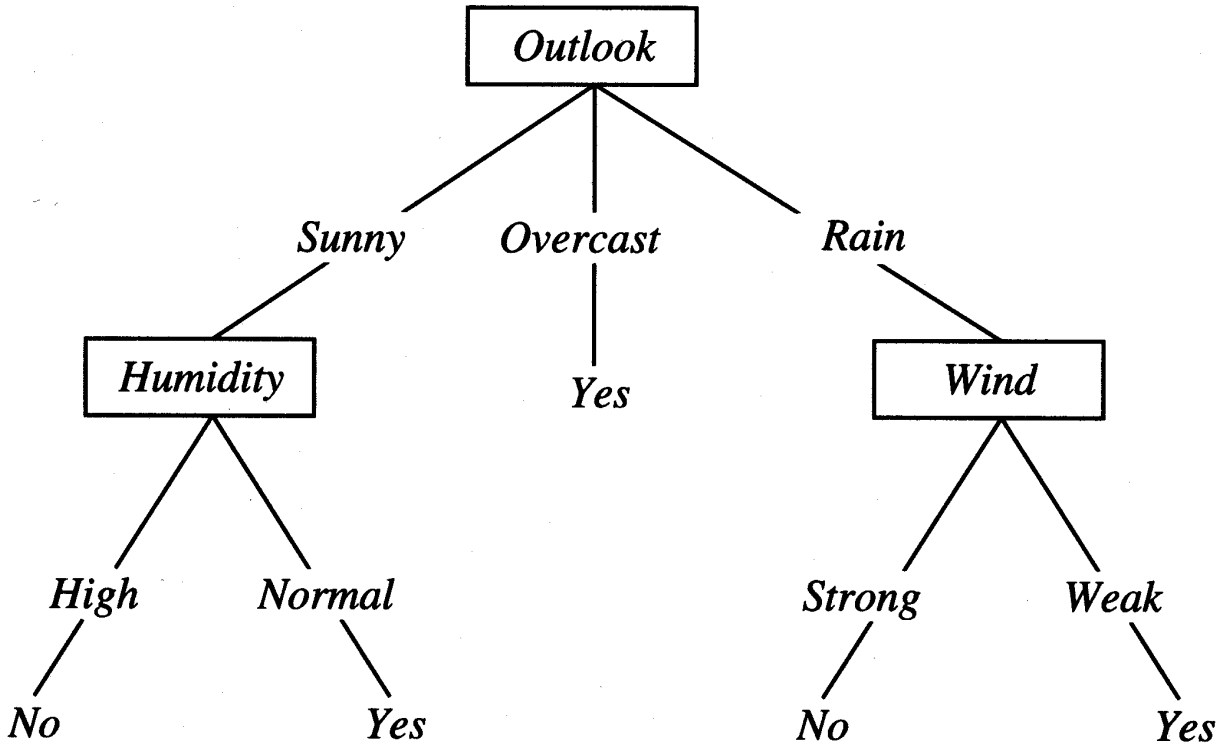
e.g., if you're trying to identify a fossil, and it has a backbone, but not four limbs, then you should decide that it corresponds to a *fish*

suppose it *does* have four limbs, what is the further information you need to decide on its classification?

Decision Trees

GC

If we have only the discriminating properties, we have the classical decision tree. —Shall we play tennis?



The information can also be expressed symbolically.

If outlook is sunny, and humidity is high, then *no* tennis.
If outlook is sunny, and humidity normal, then tennis is on.
etc.

pro's and cons?

Graph Theory

Euler developed Graph Theory from a puzzle in Königsberg.

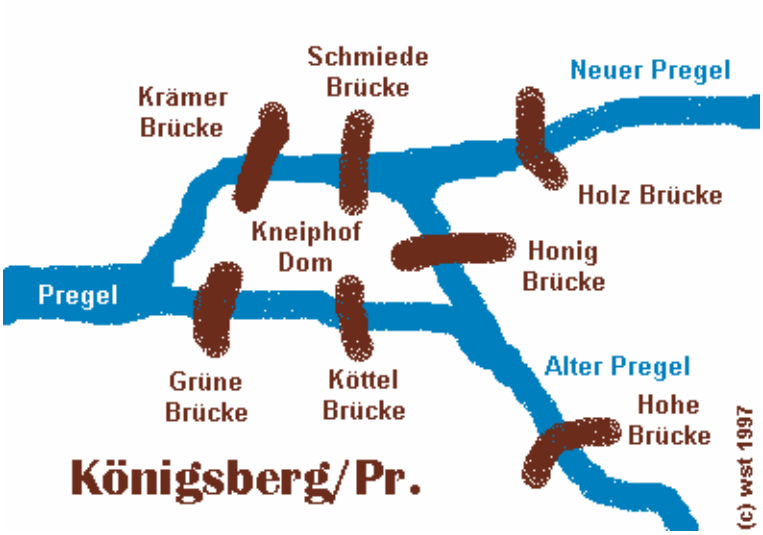
There are seven bridges over the different parts of the Pregel.

The puzzle: Kann man über alle Brücken spazieren, ohne eine zweimal zu betreten?

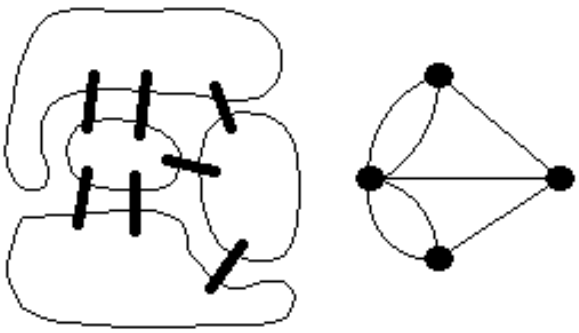
Graph Theory

Kann man über alle Brücken spazieren, ohne eine zweimal zu betreten?

The bridges are sketched SCHEMATICALLY below.



The size of the different land masses is irrelevant, since they're only connnected by the bridges.
So the puzzle can be expressed by nodes and edges.



Graph Theory

Lots of problems can be formulated in graph theory.

- problem-solving

  nodes are states, edges are steps

- speech understanding

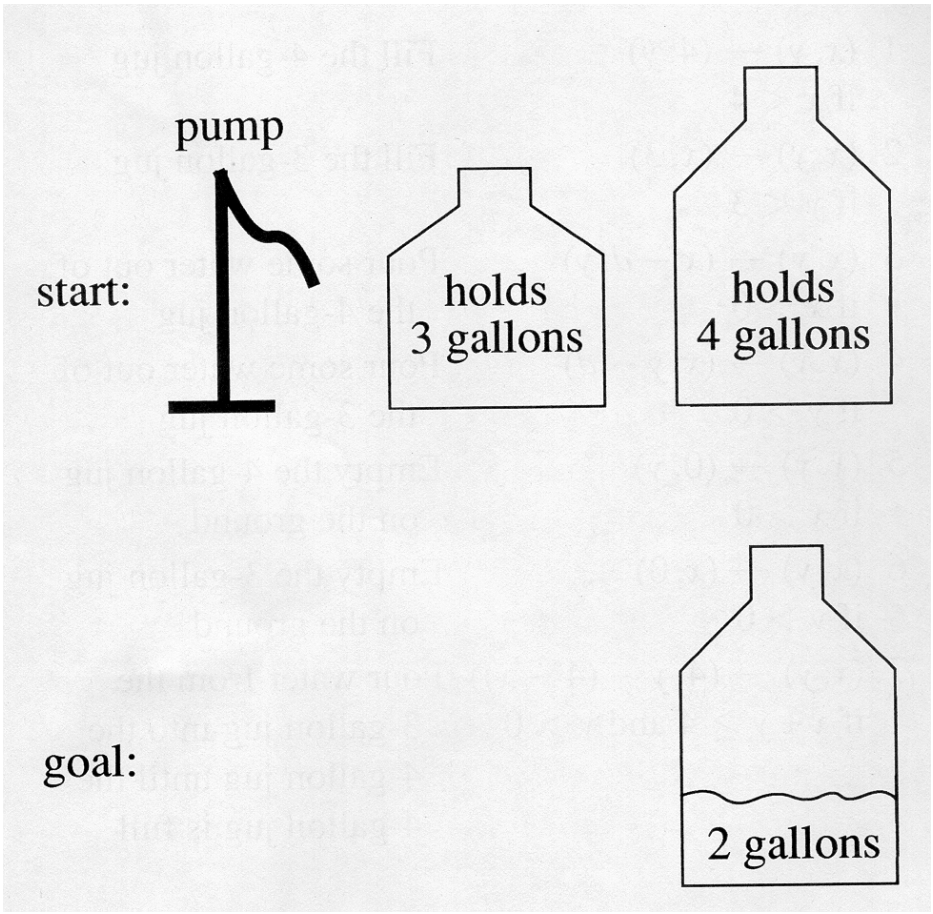
  nodes are sound-hypotheses, edges form sequences
  find mostly likely sequence (conforming to grammar)

- internet information

  nodes are sites, edges are hyperlinks
  find authoritative sites, find "communities"

Problem Solving

You have two jugs, one holding 3 gal and one holding 4. Fill one of them with exactly two gallons.



You can fill or empty jugs or pour one into another (until the latter is full).

Problem Solving

You have two jugs, one holding 3 gal and one holding 4. Fill one of them with exactly two gallons.

nodes are states, edges are steps

node is amount of water in each jug, e.g., (1/3, 4/4) indicates that there's 1 gal in the 3-gal. jug, and 4 in the 4-gal. jug.

you begin at (0/3,0/4)

edges show what's allowed in one step

- fill either container from the pump
  (a/3,b/4) $\rightarrow$ (3/3,b/4) or (a/3,4/4)
- fill either container from the other
  we have to look at two cases

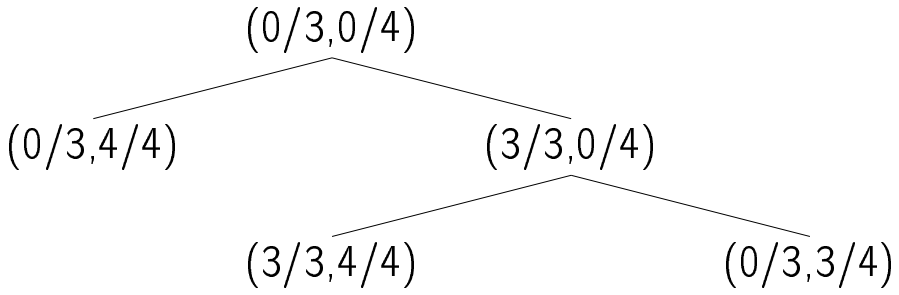
$a + b \geq$ **capacity-target** then

into 3-gal	$(a/3, b/4)$	$\rightarrow$	$(3/3, (b - (3 - a))/4)$
into 4-gal			$((a - (4 - b))/3, 4/4)$

$a + b <$ **capacity-target** then

into 3-gal	$(a/3, b/4)$	$\rightarrow$	$((a + b)/3, (b - (3 - a))/4)$
into 4-gal			$((a - (4 - b))/3, (a + b)/4)$

Jug Problem

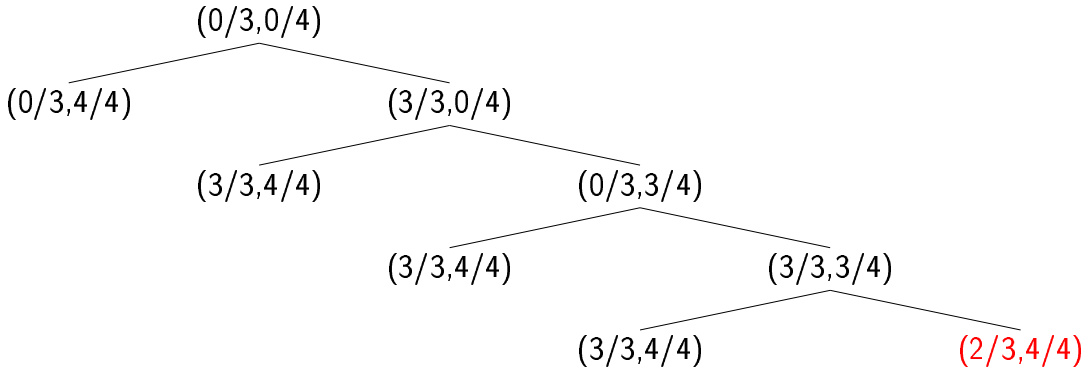


Each edge represents a manipulation corresponding to a legitimate acction in the puzzle — fill/empty one of the jugs or pour one into another (until latter is full).

The puzzle is solve when either jug contains 2 gal., i.e., at (2/3,x/4) or (x/3,2/4).

The graph (tree) now represents *process*.

Jug Problem



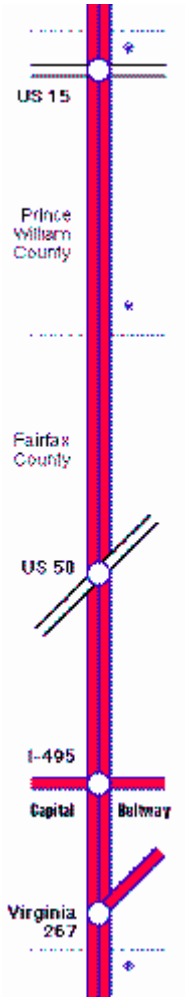
the lowest right state contains a jug with 2 gal. — a solution!

the path from the top to the solution state shows the *steps* need to reach the solution

- many optimizations, refinements possible
  - fuse states where jugs same $\longrightarrow$ dag
  - choose which edge to explore next
    depth-first, breadth-first
  - introduce "evaluation" of states
    help in choice of edge
- "mechanization" of intelligent behavior (problem-solving)

Schematic Maps

SCHEMATIC MAPS are related to state diagrams. They don't reflect geography, but rather steps in travel.



map shows travel "steps"

traveling from US 15 to the Capital Beltway involves crossing US 50

but no information on distance!

also *process*

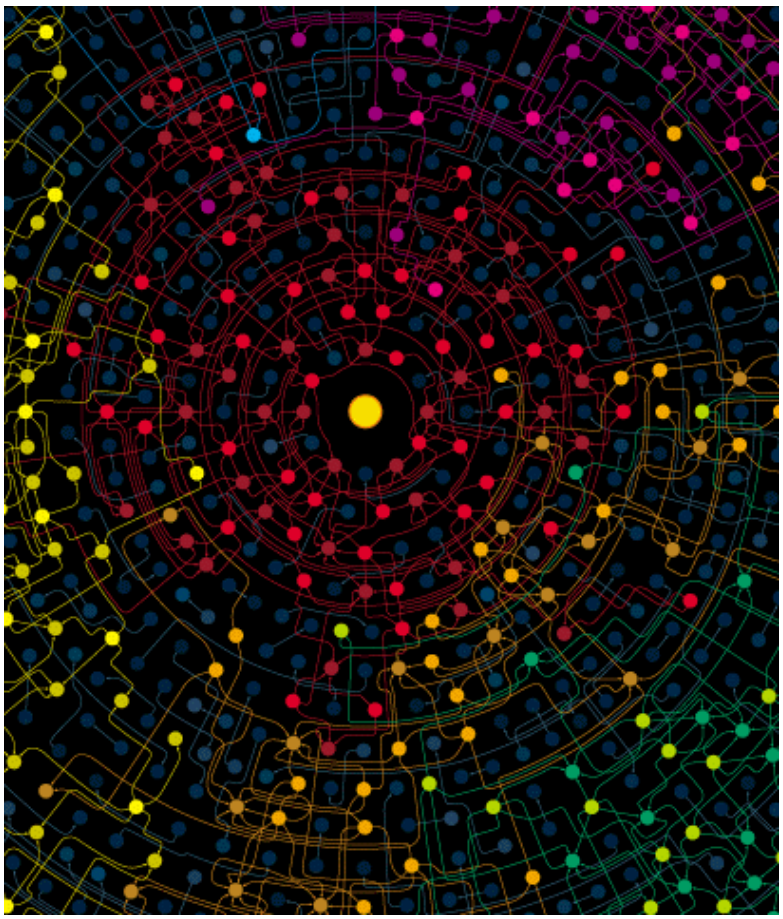
Google (Stanford)

Citation Analysis

- nodes $\sim$ web sites, edges $\sim$ hyperlinks to other sites
- sites often pointed to are RED
- $\Rightarrow$ Web pages: good sites are linked to by many others
- Google randomly traverses the web building a list of frequently encountered sites
- Finds universally popular sites, e.g. New York Times
- Favors pages on these sites in ranking search results
- Garfield: importance of journal article proportional to the number of citations it receives (arrowheads)

Dynamic Analysis

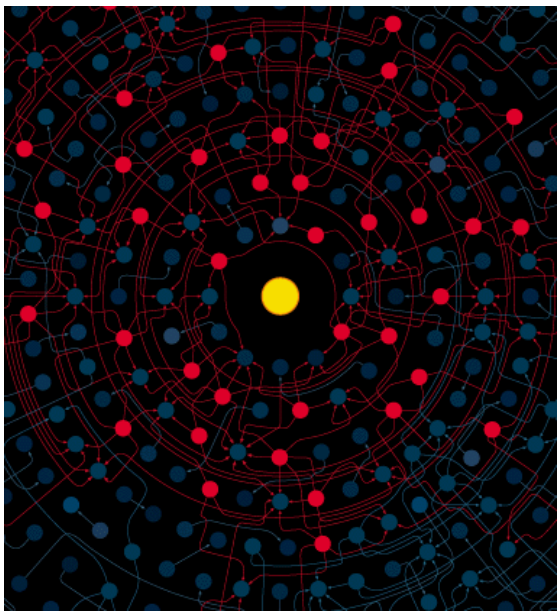
Clustered Links indicate "Web Communities"



- sites which (mostly) point to each other
  - oil spills in Japan
  - resources for Turks living in US
  - fire fighting in Australia

Dynamic Analysis

IBM Clever System



- Like Google but distinguishes between:
  - Hub pages: lists of links (red)
  - Authority pages: sites with content (blue)
    —worth pointing to
- A good hub points to many good authorities, and vice versa
- "Circular definition" utilised by an iterative algorithm to rank results of standard search
- Good hubs and authorities are near the top