Knowledge and reason

Lecture 2 of “Mathematics and AI”

Data structures & inferences

Lecture 2 of “Mathematics and AI”

Outline

1. Why care about knowledge representation?
2. Tabular data structures
3. Graphical data structures
   a. Deterministic graphical models
   (semantic nets and knowledge graphs)
   b. Probabilistic graphical models
   (Bayesian networks and Markov random fields)

Why care about knowledge representation?

Knowledge is a key aspect of intelligence

intelligence *noun*

in·tel·li·gence in-ˈte-lə-jən(t)s

Synonyms of *intelligence* >

1 a (1) : the ability to learn or understand or to deal with new or trying situations : **REASON**

also : the skilled use of reason

(2) : the ability to apply knowledge to manipulate one's environment or to think abstractly as measured by objective criteria (such as tests)

b : mental acuteness : **SHREWDNESS**

c Christian Science : the basic eternal quality of divine Mind

2 a : **INFORMATION**, **NEWS**

b : information concerning an enemy or possible enemy or an area

also : an agency engaged in obtaining such information

3 : the act of understanding : **COMPREHENSION**

Source: www.merriam-webster.com

Knowledge representations in the era of ML

- Representation of prior knowledge for learning
- Efficient AI
  (reduce training)
- Reliable artificial intelligence
  (reduce hallucinations)
- Interpretable machine learning

Image sources: https://www.weka.io/learn/ai-ml/what-is-ai/
https://dgraph.io/blog/post/why-google-needed-graph-serving-system/

Challenges in knowledge representation

- Translating a complex world into a formal language
- Handling conflicts in knowledge
- Handling uncertainties in knowledge

Tabular data structures

Tabular knowledge representation

- Examples:
  - Tables / Data Frames
    (in python: pandas package)
  - Arrays
    (in python: numpy package)
  - Matrices

Name	Position	Office	Age	Start date
Airi Satou	Accountant	Tokyo	33	2008/11/28
Angelica Ramos	Chief Executive Officer (CEO)	London	47	2009/10/09
Ashton Cox	Junior Technical Author	San Francisco	66	2009/01/12
Bradley Greer	Software Engineer	London	41	2012/10/13
Brenden Wagner	Software Engineer	San Francisco	28	2011/06/07
Brielle Williamson	Integration Specialist	New York	61	2012/12/02
Bruno Nash	Software Engineer	London	38	2011/05/03
Caesar Vance	Pre-Sales Support	New York	21	2011/12/12
Cara Stevens	Sales Assistant	New York	46	2011/12/06
Cedric Kelly	Senior Javascript Developer	Edinburgh	22	2012/03/29

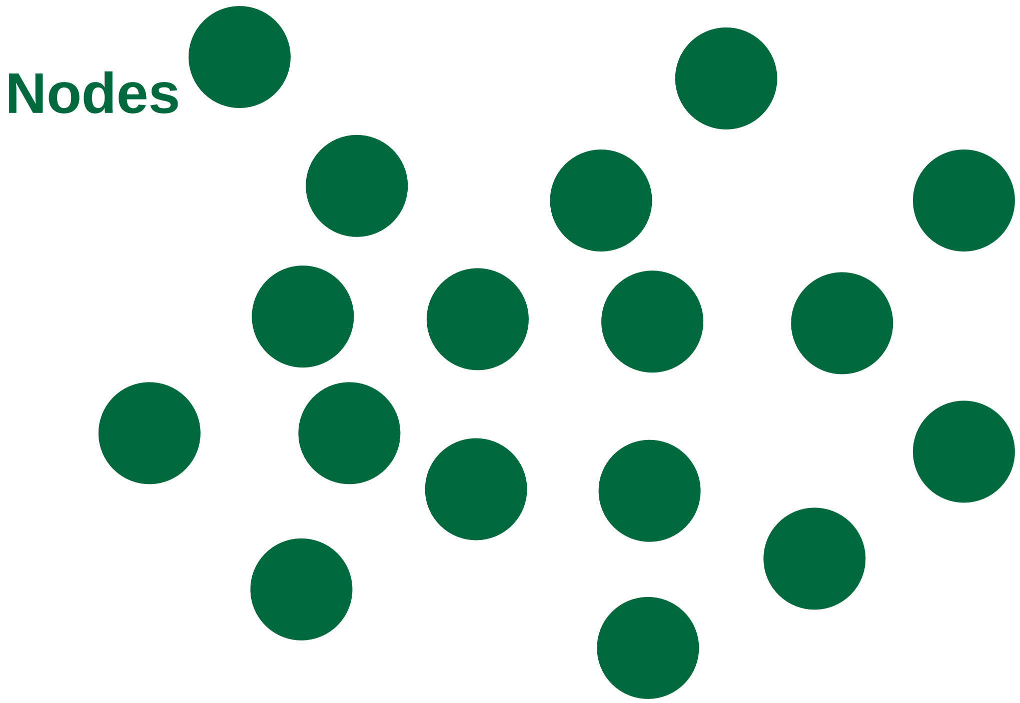
Image source: https://poovarasu.medium.com/pros-cons-in-datatables-ff9a4fa83f17

Tabular knowledge representation

- Advantages:
  - Standardized formats for use in programming
  - Tables with numerical values facilitate use of standard data science methods (e.g., PCA)
- Limitations:
  - All objects have a value for all attributes
  - Attributes cannot have attributes
  - Cannot capture complex situations and/or processes

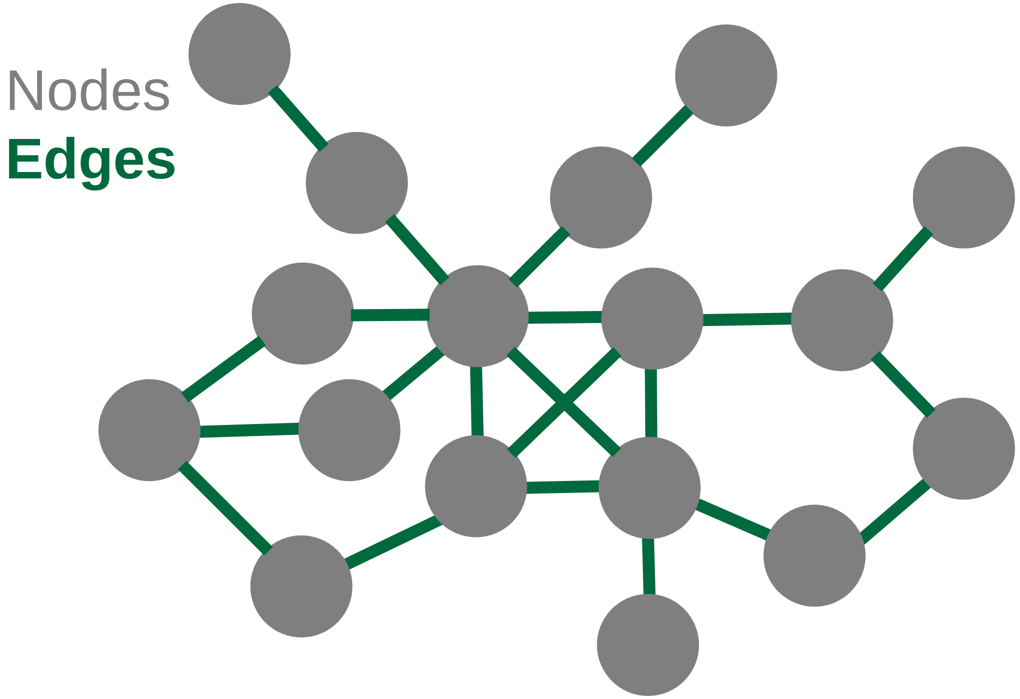
Graphical data structures

Graphs



- Node set V

Graphs



- Node set *V*
- Edge set *E*

$\Rightarrow$ Graph (V,E)

Graphs

- Node set *V*
- Edge set *E*

$\Rightarrow$ Graph *(V,E)*

- Some important substructures:
  - Cliques

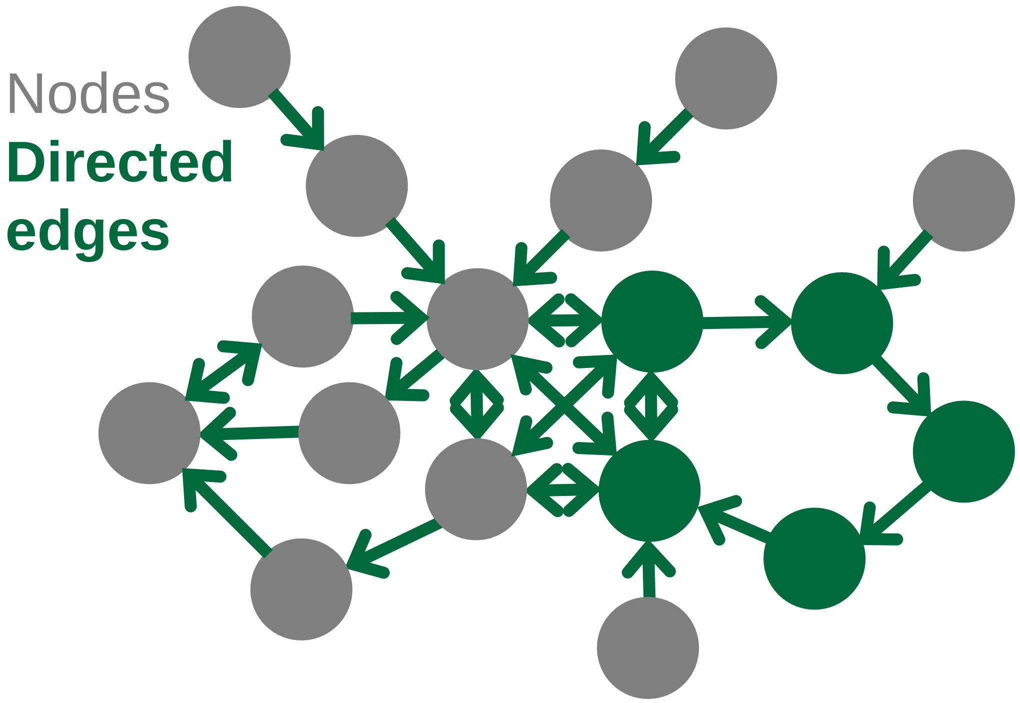
Graphs

- Node set V
- Edge set E

$\Rightarrow$ Graph (V,E)

- Some important substructures:
  - Cliques
  - Cycles

Graphs

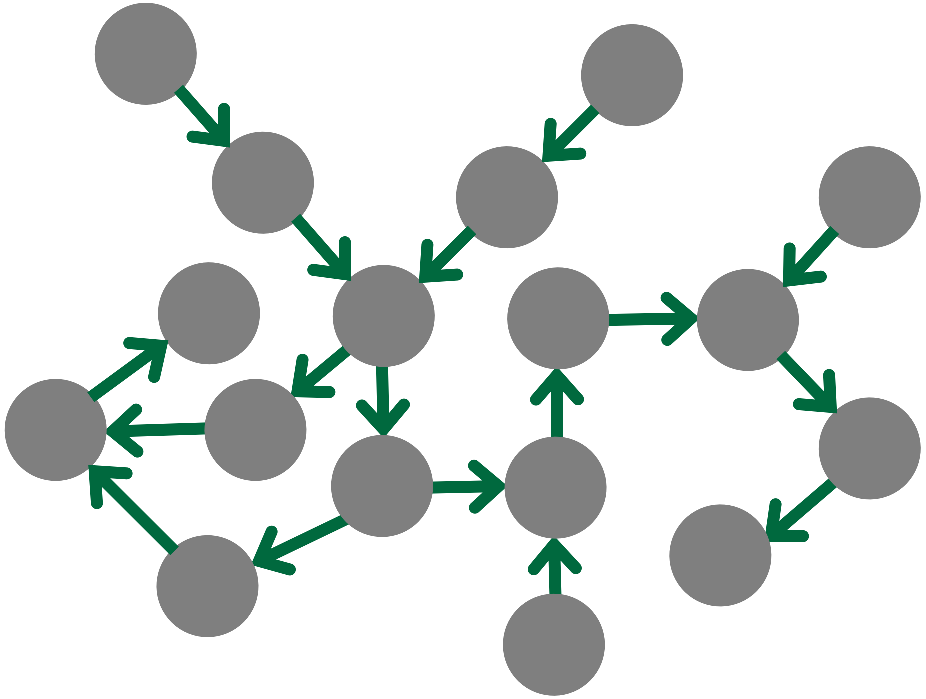


- Node set *V*
- Edge set *E*

=> Directed Graph *(V,E)*

- Some important substructures:
  - Cliques
  - Directed cycles

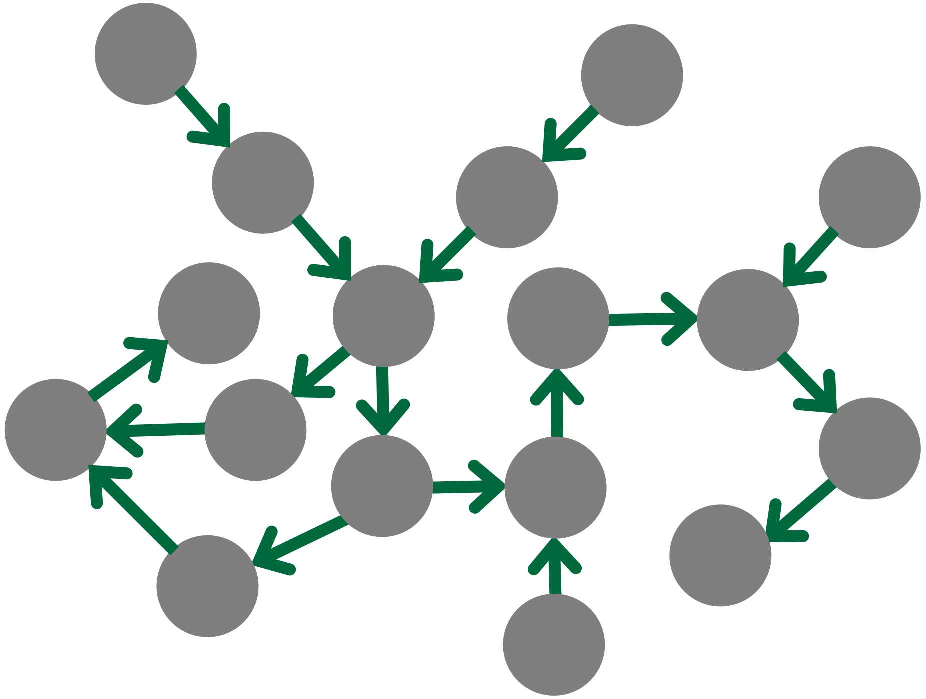
Directed Acyclic Graphs (DAG)



- No directed cycles
- Example uses:
  - hierarchical ontologies
  - causal event structure without feedback loops

Graphical data structures for certain knowledge

Semantic networks



- Nodes represent entities or attributes
- Edges represent relationships (e.g. “is member of”)
- Large semantic networks are sometimes called knowledge graphs

Exercise 1

“All of probability” in 4 slides

Probability

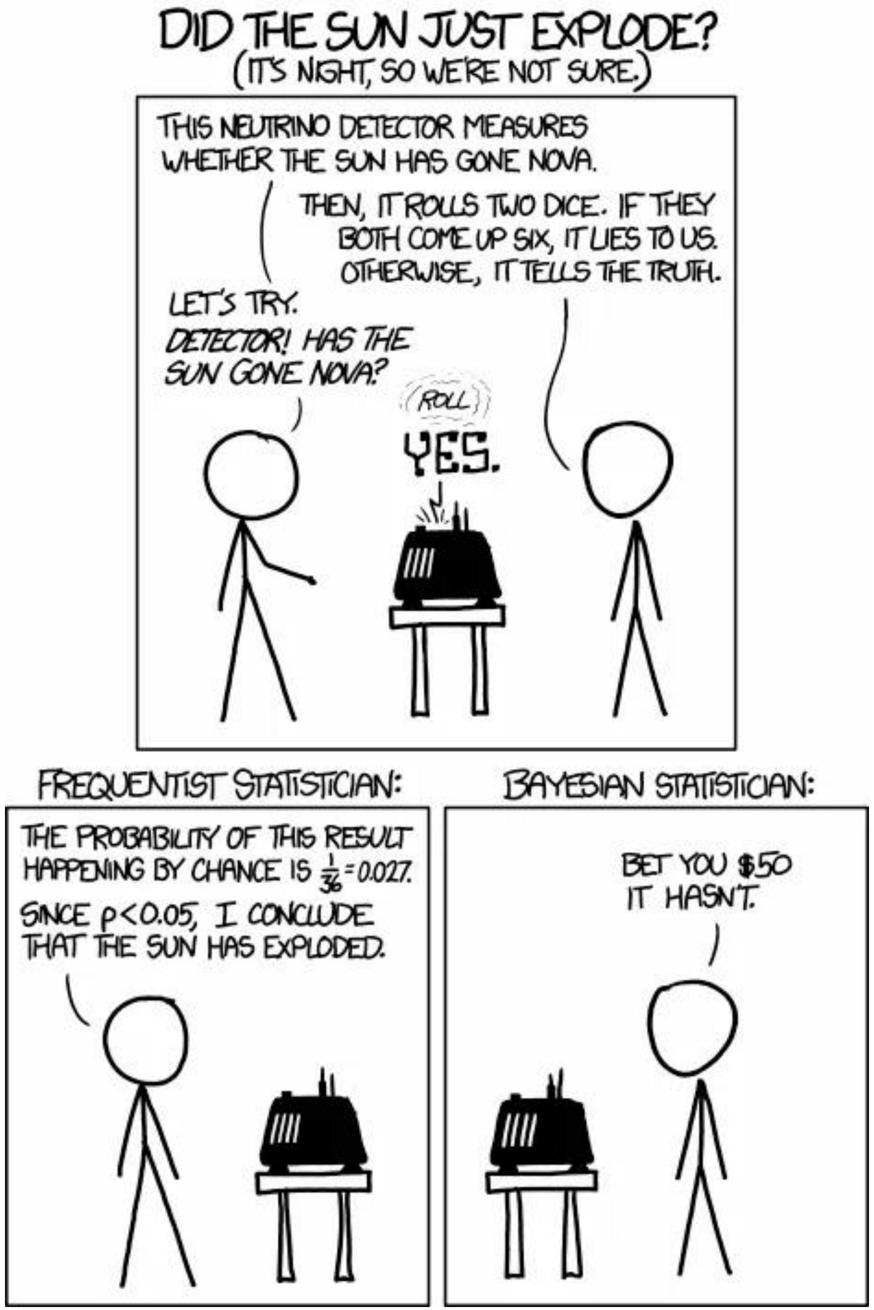
- Probability $p(x) \in [0,1]$
- Probability distribution $p(x)$, $x \in D$ with $\sum_x p(x) = 1$
- Independent probabilities $p(x,y) = p(x)p(y)$ (otherwise dependent)
- Conditional probabilities $p(x|y) = p(x,y)/p(y)$
- Conditional probability distribution $p(x|y) = p(x,y)/p(y)$, $x \in D$
- Marginal probability distribution $p(x) = \sum_y p(x|y)\,p(y)$, $x \in D$

DARTMOUTH

Bayes rule

- How should our estimates of probabilities change when we receive new data?
- Bayes rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



Random variables

- The **random variable** X has a probability distribution $p(x)$, $x \in D(X)$
- **Samples** of the random variable X are elements of D (e.g., numbers)
- **Expectation** of X: $\mathrm{E}[\mathrm{X}] = \sum_x xp(x)$
- **Variance** of X: $\mathrm{Var}[\mathrm{X}] = \mathrm{E}[(X - E[X])^2] = \mathrm{E}[X^2] - \mathrm{E}[\mathrm{X}]^2$

The Markov property

- A Markov process in (a model of) a memory-less process
- Markov property

$$p(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, \dots) = p(X_{t+1} = x_{t+1} | X_t = x_t)$$