iPMR

Recap

- What did we learn from the last lecture?

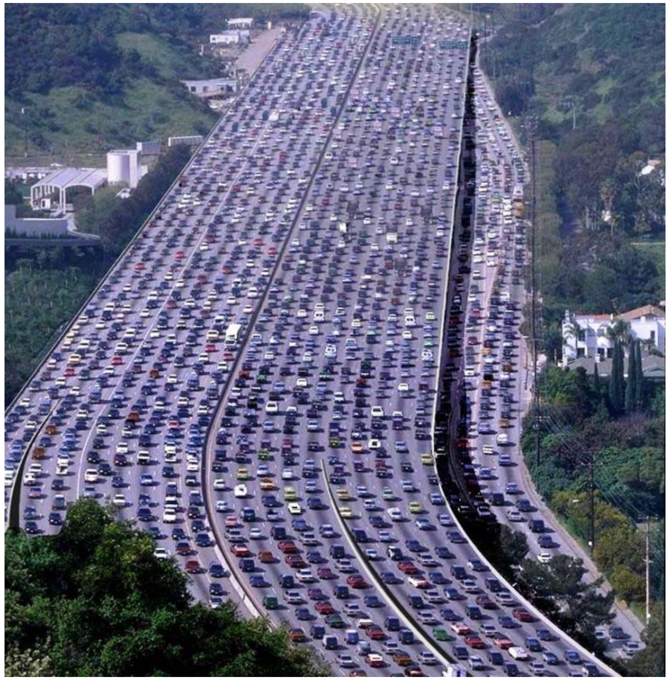
Reminder

- Website: Work to go through after each lecture.

Markov Blanket

- Two Questions:

Copyright Francisco Osorio, Creative Commons

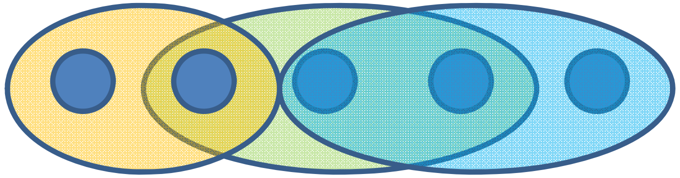


Copyright Samuel Lio, Creative Commons

Factor Graphs

- Factor Graphs

  We can write any probability distribution as



$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(-E(\mathbf{x})\right)$$

where Z is a normalisation constant. Convince yourself that is true.

The question is, "What form does *E* take?"

E can be additive in a number of different components. (Trivial statement – components could be identical).

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{N_f} \phi_i(\mathbf{x})\right)$$

But often each component can be dependent only on subset of variables

$$\phi_i(\mathbf{x}) = \phi_i(\mathbf{x}_{C_i}) \qquad \text{e.g. } \phi_i(x_1, x_2, x_3, x_4, x_5) = \phi_i(x_1, x_2)$$

C_i denotes the set of indices of the variables that occur in the *i*th factor.

Building a Factor Graph

- We could also write

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N_f}\psi_i(\mathbf{x}_{C_i}) \quad \text{(equivalent to)} \quad P(\mathbf{x}) = \frac{1}{Z}\exp\left(\sum_{i=1}^{N_f}\phi_i(\mathbf{x}_{C_i})\right)$$

- Let χ_i collect the variables indexed by C_i. E.g. $\chi_1 = (x_1, x_3, x_7, x_8)$. Then we can write

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N_f}\psi_i(\chi_i)$$

Why?

Why?

- Lots of different motivations.
- E.g. 1

Consider independent variables x_1, x_2, x_3, x_4, x_5.

$$P(\mathbf{x}) = P(x_1)P(x_2)P(x_3)P(x_4)P(x_5)$$

Now consider random binary features f_1, f_2, f_3 of the form $P(f_1|x_1, x_2)$, $P(f_2|x_2, x_3, x_4)$, $P(f_3|x_3, x_4, x_5)$.

Suppose we now use those features as constraints: $f_1 = 1, f_2 = 1, f_3 = 1$.

Then conditioning on those constraints we have the distribution $P(\mathbf{x}|\mathbf{f})$. We can write

$$P(\mathbf{x}|\mathbf{f}) \propto P(x_1)P(x_2)P(x_3)P(x_4)P(x_5)P(f_1|x_1, x_2)P(f_2|x_2, x_3, x_4)P(f_3|x_3, x_4, x_5)$$

Looking carefully we notice

$$P(\mathbf{x}|\mathbf{f}) = \frac{1}{Z}\Psi(x_1, x_2)\Psi(x_2, x_3, x_4)\Psi(x_3, x_4, x_5)$$

Why?

- Lots of different motivations.
- E.g. 2. Exponential family distributions
- Many distributions we use fall in the class of exponential family distributions. They can be written as

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{N_f} w_i \phi_i(\mathbf{x}) \right)$$

Why?

- More examples later.

Building a Factor Graph

- Reminder:

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N_f}\psi_i(\mathbf{x}_{C_i})$$

- Let χ_i collect the variables indexed by C_i. E.g. $\chi_1 = (x_1, x_3, x_7, x_8)$. Then we can write

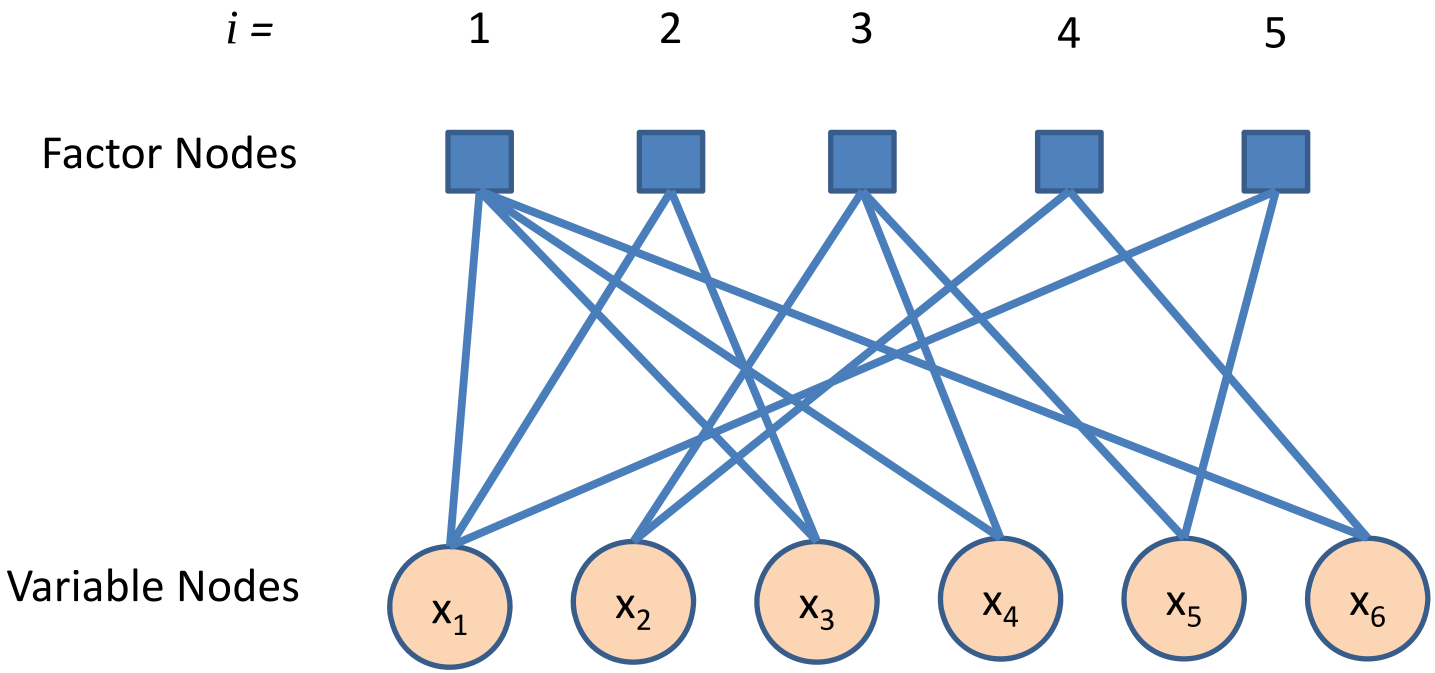
$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N_f}\psi_i(\chi_i)$$

- We can build a graphical representation to capture this. It is called a factor graph. ψ are factors.
- *Bipartite* graph with variable and factor nodes.

Building a Factor Graph

$$P(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N_f} \psi_i(\chi_i)$$

$$\chi_1 = (x_1, x_3, x_4, x_6)$$
$$\chi_2 = (x_1, x_3)$$
$$\chi_3 = (x_2, x_4, x_5)$$
$$\chi_4 = (x_2, x_6)$$
$$\chi_5 = (x_5, x_1)$$



Examples

- Draw the minimal factor graphs for the general form of the following distributions

$$P(\mathbf{x}) \propto (x_1 - x_2)^2 x_3^2 x_4^2 \text{ for } 0 \leq x_i \leq 1$$

$$P(\mathbf{x}) = \frac{1}{Z} \exp(-\mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x}) \text{ for } \mathbf{x} \text{ length } 4$$

$$P(\mathbf{x}, \mathbf{v}) = P(x_0) \prod_{i=1}^{5} P(x_i | x_{i-1}) P(v_i | x_i)$$

$$P(\mathbf{x}) = \frac{1}{Z} \exp(-[\mathbf{x}^T \mathbf{W} \mathbf{x}]^2) \text{ for } \mathbf{x} \text{ length } 4$$

A special case

- Recall the chain rule of probability

e.g. for $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$

$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1)P(x_4|x_3, x_2, x_1)$$

- It may be that some conditional independence relationships hold:

$$P(x_3|x_2, x_1) = P(x_3|x_2)$$
$$P(x_4|x_3, x_2, x_1) = P(x_4|x_1)$$

- We can also see that each bit of the chain rule is a factor.
- The conditional dependencies cause removed edges in the factor graph. Draw the factor graph.

$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2)P(x_4|x_1)$$

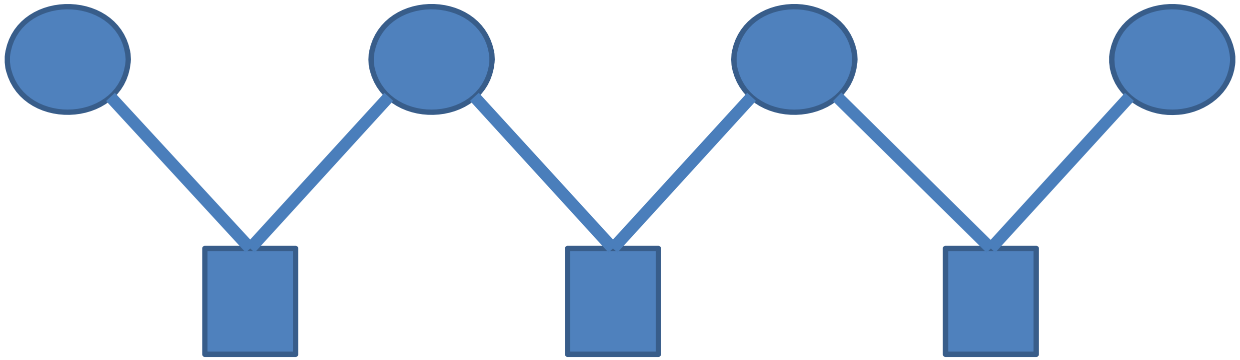
Summary

Factor Graph – structure of distribution. Explain why.

BREAK

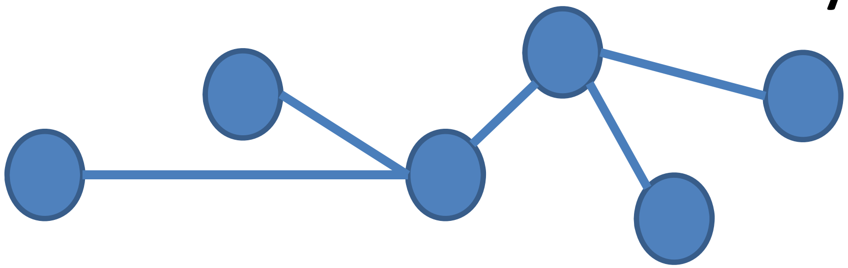
Special Factor Graphs

- Chain factor graphs



Tree Structured Networks

- An (undirected) network is a *tree* if the network has no undirected cycles. (*)



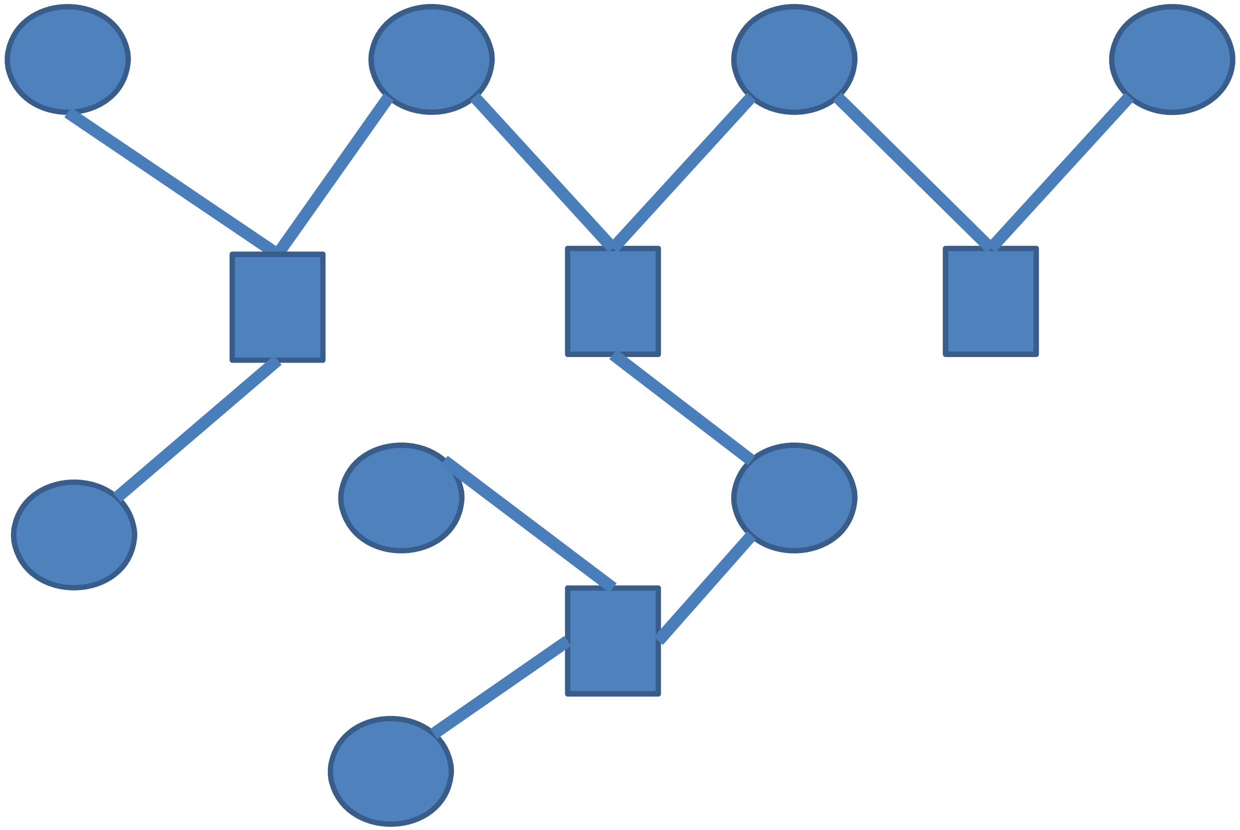
- The simplest type of tree is a chain

- Trees are special. We will see trees are easy to do inference with…

* Strictly it is a polytree – there could be many isolated trees. WLOG any method that applies to trees also applies to polytrees by applying it to each tree in the polytree. We simply refer to trees henceforth without further worry about this matter.

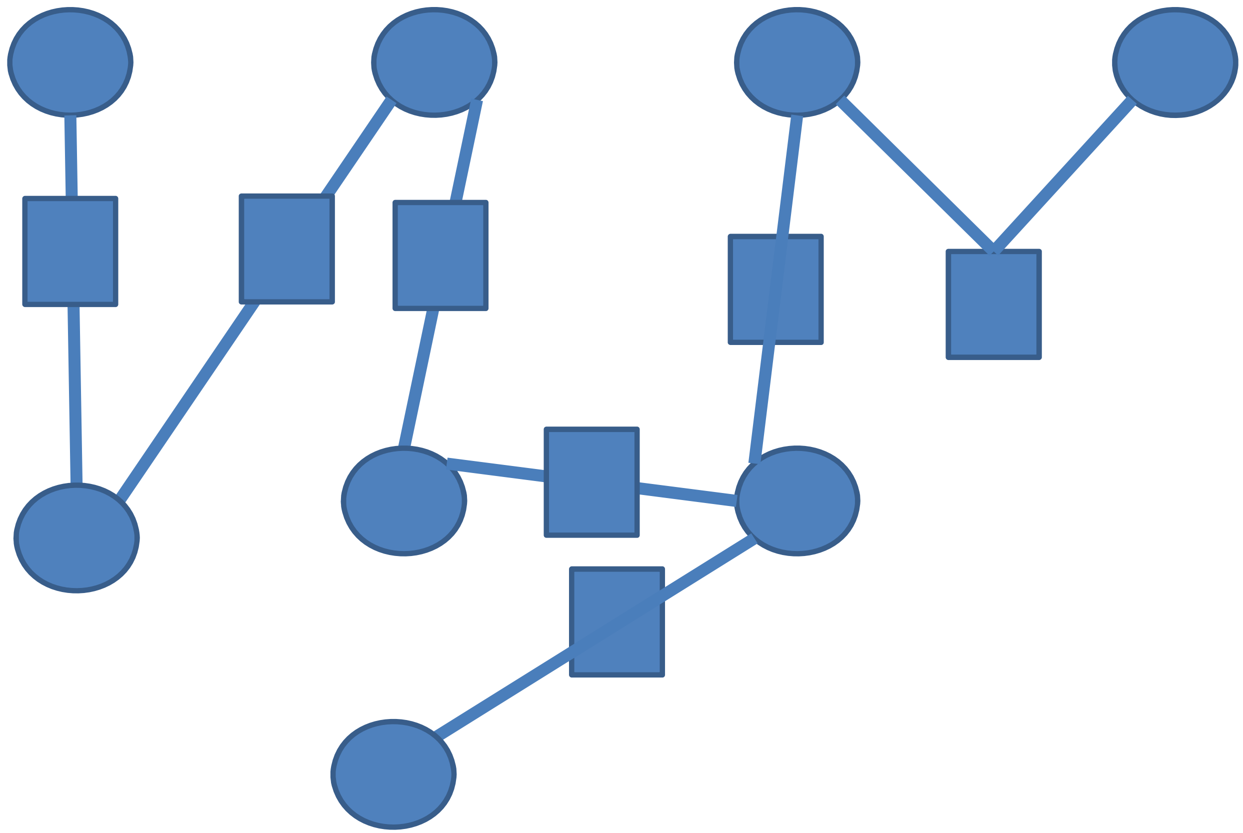
Special Factor Graphs

- Tree structured factor graphs



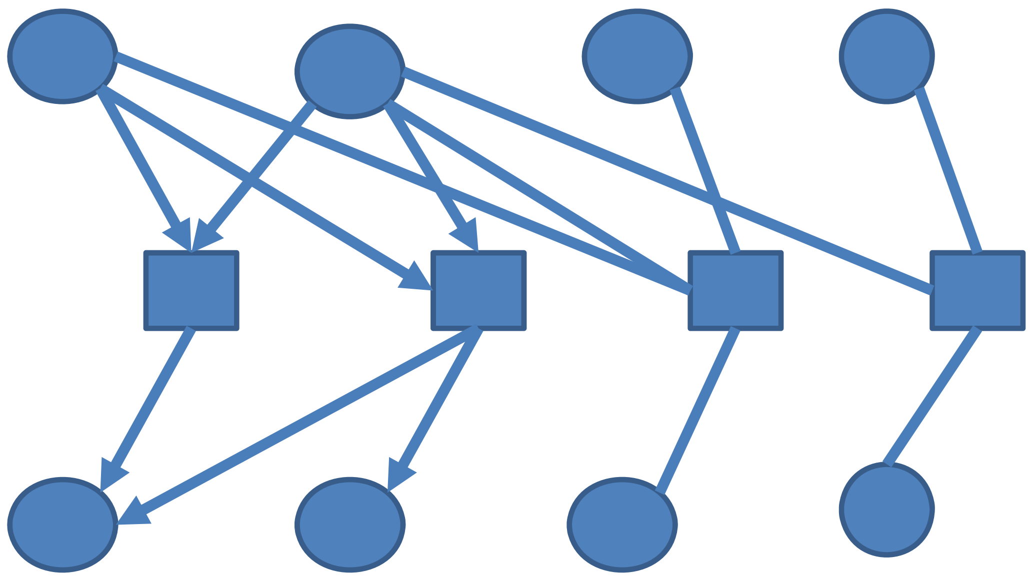
Special Factor Graphs

- Tree structured factor graphs



Directed Factor Graph

- If a factor takes the form P(x,y|r,s), then
  - use directed edges from r,s to the factor,
  - use directed edges from the factor to x,y.
- Can have mixed graphs.



Independence...

- Two random variables are independent if we can write

$$P(\mathrm{x}, \mathrm{y}) = P(\mathrm{x})P(\mathrm{y})$$

- We can extend this to sets of random variables

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$$

- We can use the rule of conditioning to get

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x})P(\mathbf{y})}{P(\mathbf{y})} = P(\mathbf{x})$$

…Independence

- Is Toothache **independent of** Cavity **below?**

$P(\text{Toothache}, \text{Cavity})$ is

	Toothache = true	Toothache = false
Cavity = true	0.04	0.06
Cavity = false	0.01	0.89

$P(\text{Cavity}|\text{Toothache})$ is

	Toothache = true	Toothache = false
Cavity = true	0.8	0.063
Cavity = false	0.2	0.937

Conditional Independence

- Rule of independence extends to conditional probabilities,

$$P(\mathbf{x}|\mathbf{y},\mathbf{z}) = \frac{P(\mathbf{x}|\mathbf{z})P(\mathbf{y}|\mathbf{z})}{P(\mathbf{y}|\mathbf{z})} = P(\mathbf{x}|\mathbf{z})$$

- This is conditional independence and is notated by $I(\mathbf{x},\mathbf{y}|\mathbf{z})$
- Some comments on handling conditional probabilities…
  - Each variable must appear either on the right hand side or left hand side of | not both.
  - Conditional independence means you can drop a variable from the right side.

Taster: In Factor Graphs

- Factor graph ➔ conditional independence relationships.
  - Directed ✓
  - Undirected ✓
  - Partially directed ✓

Two variables are guaranteed conditionally independent **given** a set of conditioned variables **if** all paths connecting the two variables are blocked.

A path is blocked if one or more of the following conditions is satisfied:

- One of the variables in the path is in the conditioning set.
- One of the variables or factors in the path has two incoming edges that are part of the path, and neither the variable/factor nor any of its descendants are in the conditioning set.

- Notes: if but not only if. Incoming: must be directed. Descendants: must all be directed.
- Frey 2003 Extending Factor Graphs so as to Unify Directed and Undirected Graphical Models

Our Journey

Graphical Models

- Lecture 1: Introduce PMR
  - Focus: Distributions, Joint Models, Unsupervised.
- Lecture 2: Introduce Factor Graphs
  - Distributions → Factor Graphs
  - Content v Form
    - Structure of distributions
    - Intro: Read off conditional independence
  - Next lecture: Conditional Independence in Factor Graphs.