Computational Cognitive Science

Structure
"form"

F

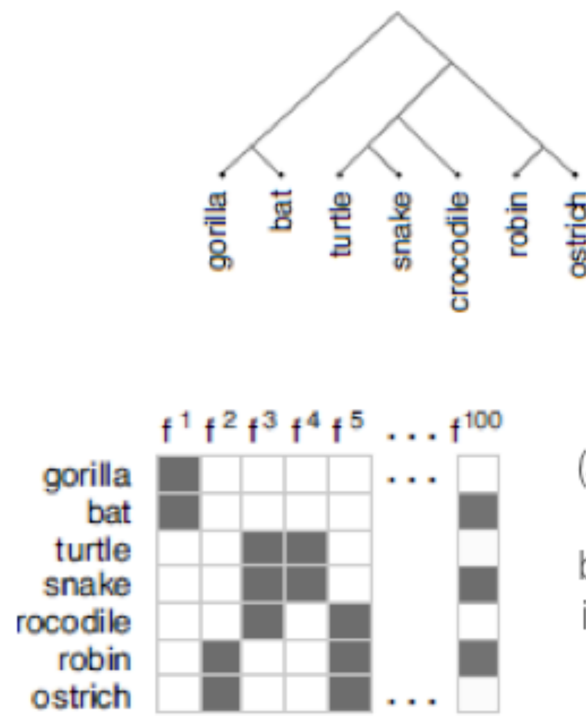
Hierarchy

Structure

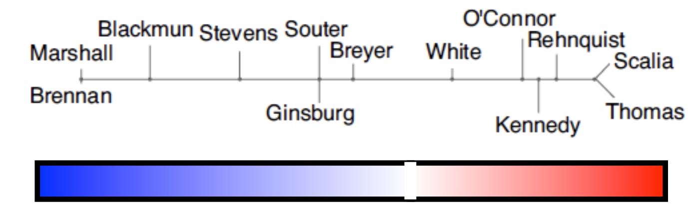
S

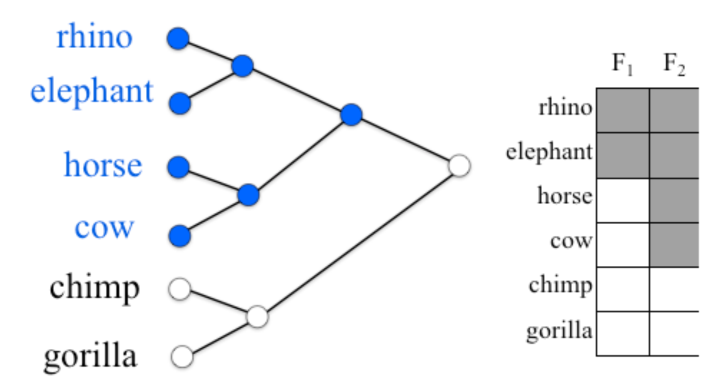
Object-feature
matrix

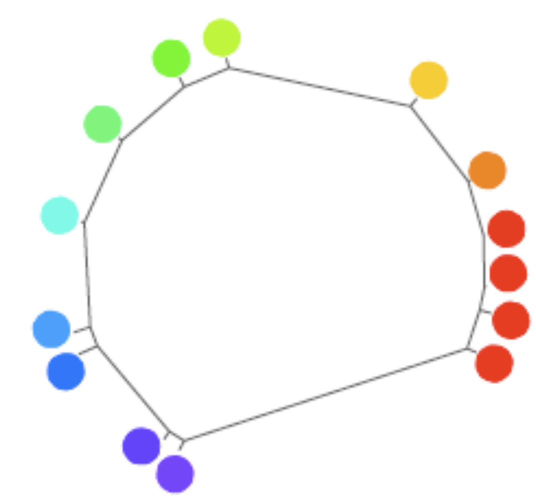
D



(also works with
similarity data,
but feature data
is more intuitive
to explain)



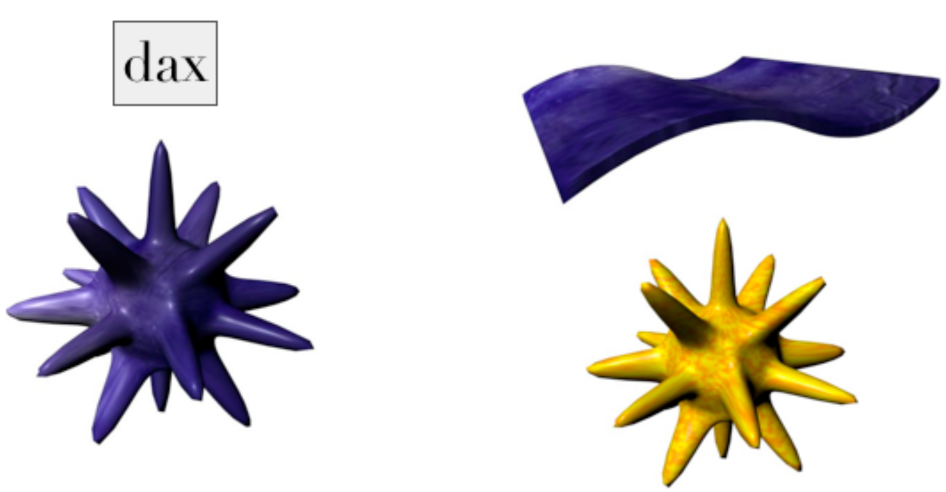


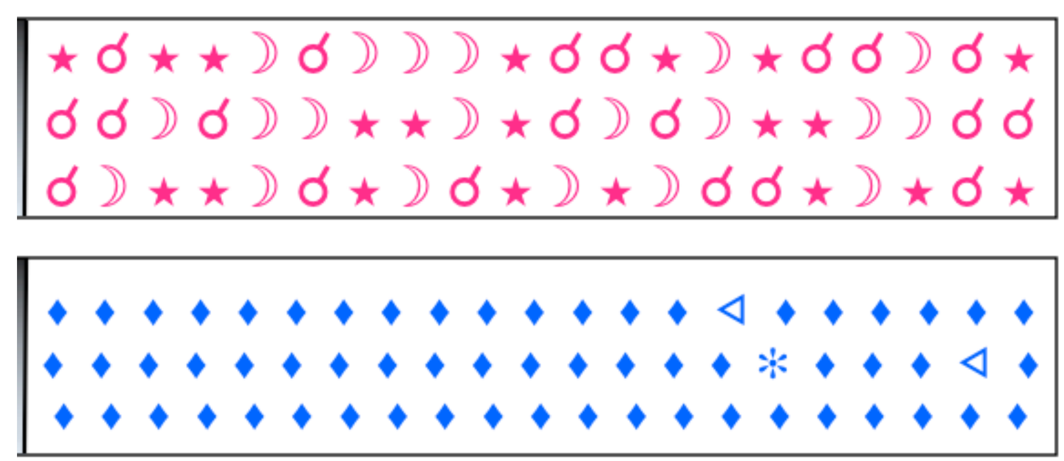


Lecture 13: Higher order
knowledge 3

Last few lectures

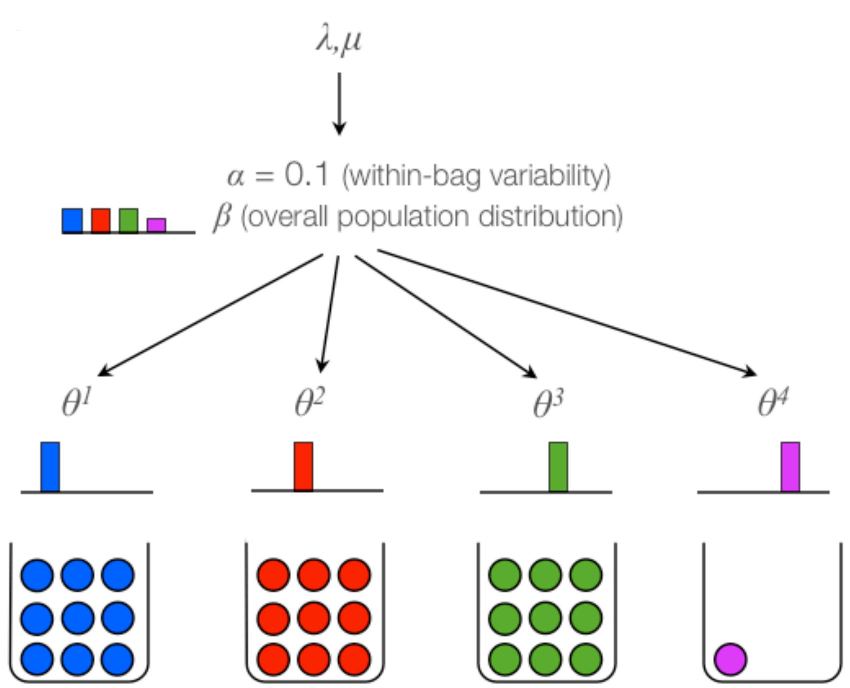
- We’ve seen several examples of instances where people can learn *overhypotheses* -- making higher-order inferences about the variability or distribution of items within categories

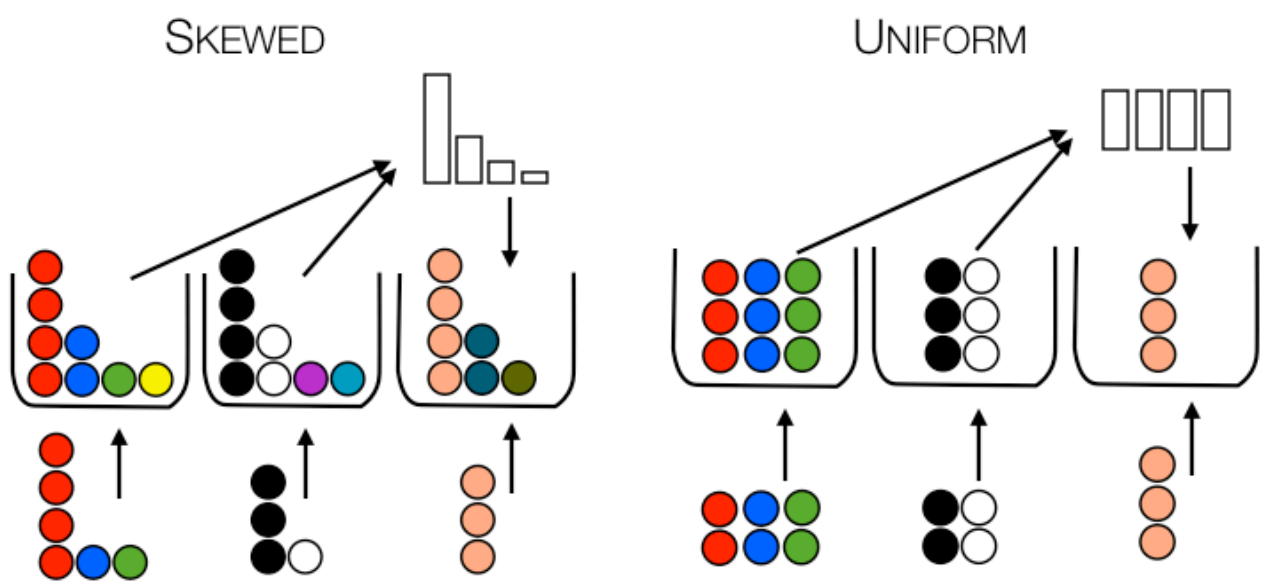




Last few lectures

- We've seen several examples of instances where people can learn *overhypotheses* -- making higher-order inferences about the variability or distribution of items within categories
- We also saw models that can capture this learning





Last few lectures

- We’ve seen several examples of instances where people can learn *overhypotheses* -- making higher-order inferences about the variability or distribution of items within categories
- We also saw models that can capture this learning
- Today: one additional kind of learning: structure learning

Lecture outline (next three lectures)

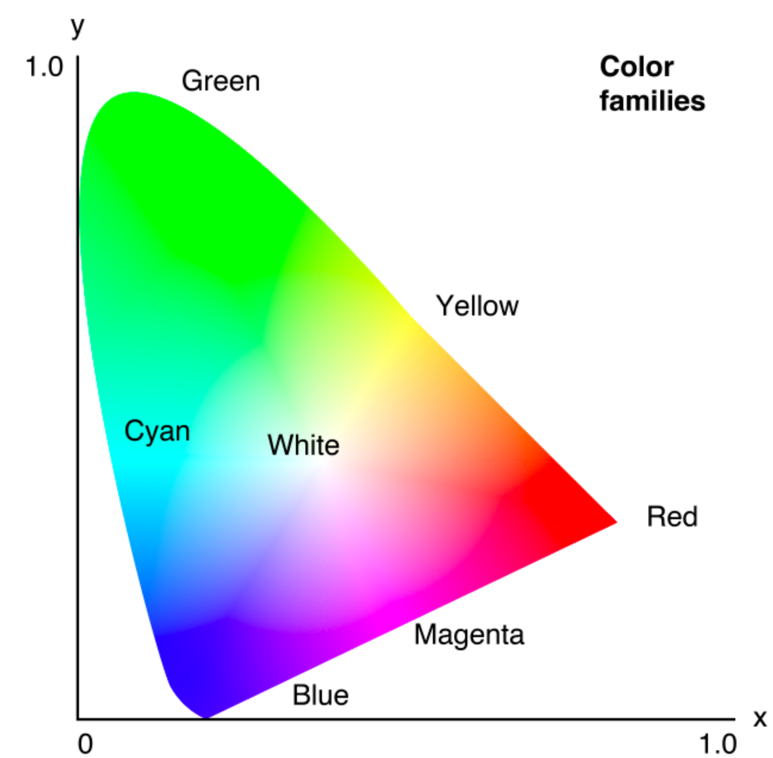
- Lecture 11: Learning about category variability
  - This kind of learning in children and adults
  - A model for this kind of learning
  - Limitations of this model
- Last time: Learning about distributions of categories
  - This kind of learning in adults
  - Failure of current models
  - A model for this kind of learning
- Today: Learning about category structure
  - This kind of learning in people
  - A model for this kind of learning

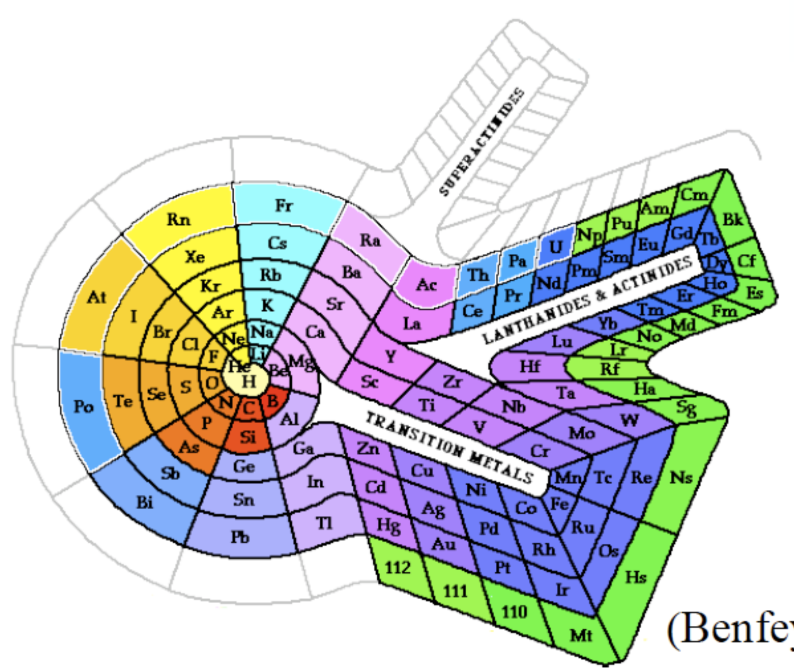
Lecture outline (next three lectures)

- Lecture 11: Learning about category variability
  - This kind of learning in children and adults
  - A model for this kind of learning
  - Limitations of this model
- Last time: Learning about distributions of categories
  - This kind of learning in adults
  - Failure of current models
  - A model for this kind of learning
- ➡ Today: Learning about category structure
  - ➡ This kind of learning in people
  - A model for this kind of learning

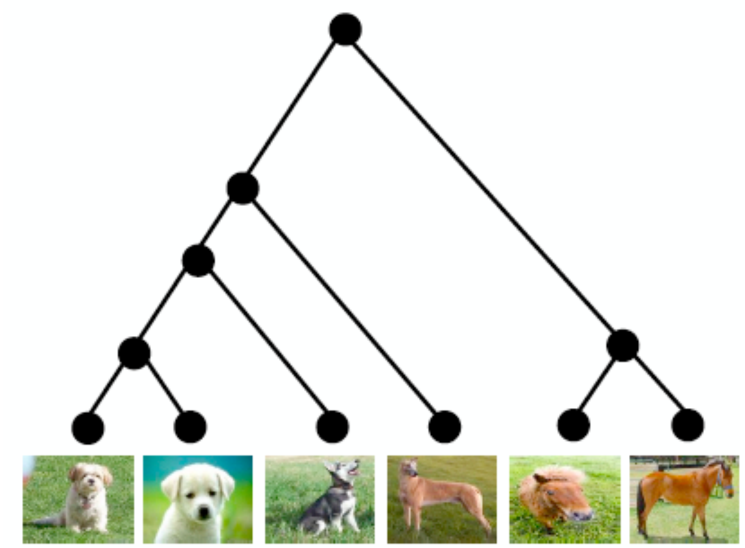
What is the problem of structure learning?

We've seen already that different domains appear to have different structures



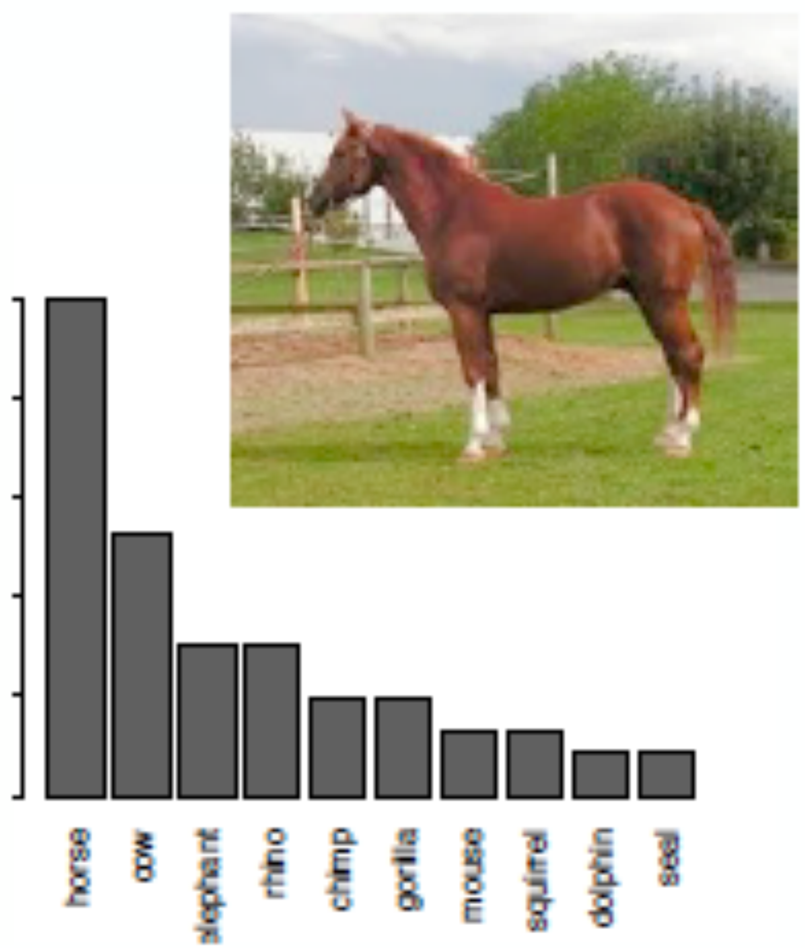


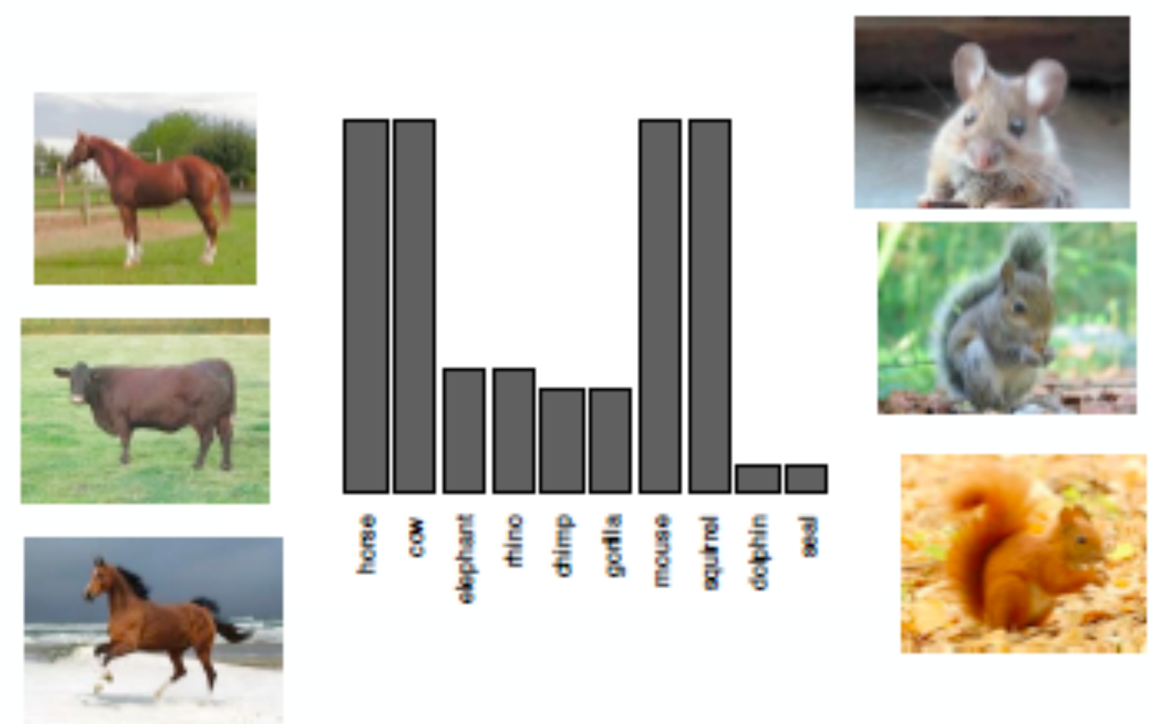
(Benfey, 1960)



What is the problem of structure learning?

... and that structure matters for the inferences one makes





What is the problem of structure learning?

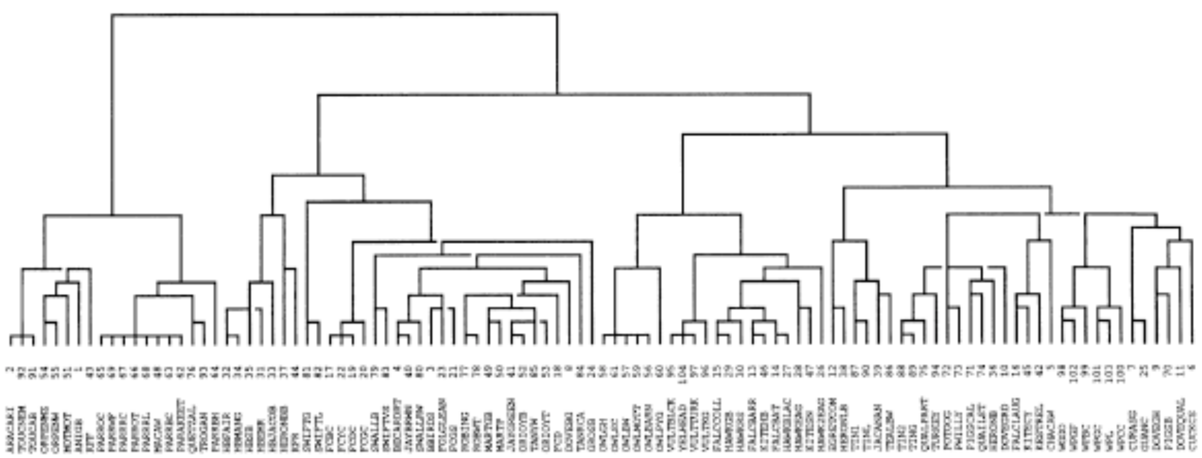
... and that structure matters for the inferences one makes

> "One can predict the discovery of many new elements, for example, analogues of Si and Al with atomic weights of 65-75."
>
> "A few atomic weights will probably require correction; for example Te cannot have the atomic weight 128, but rather 123-126."
>
> \- Mendeleev

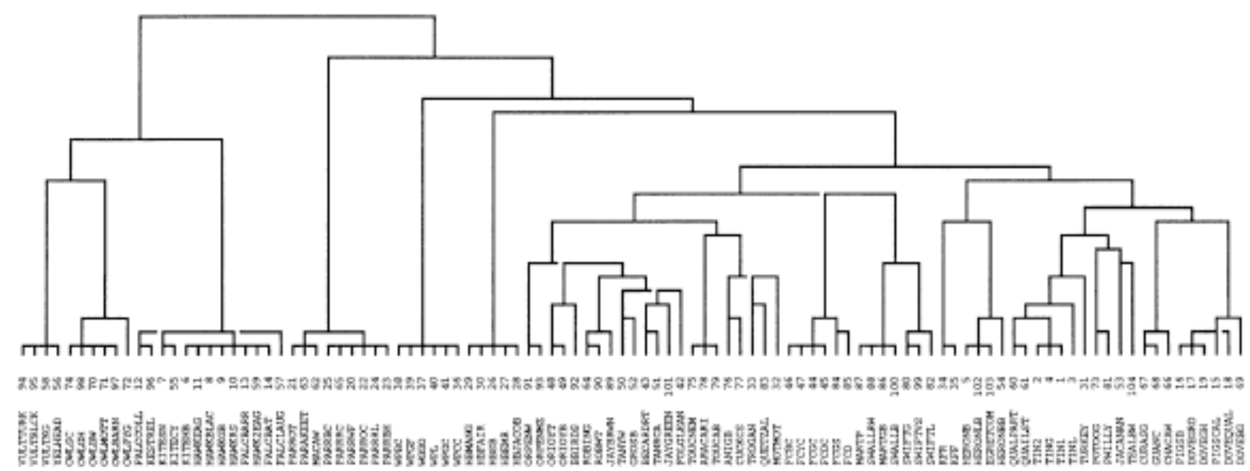
Structure in different domains: biology

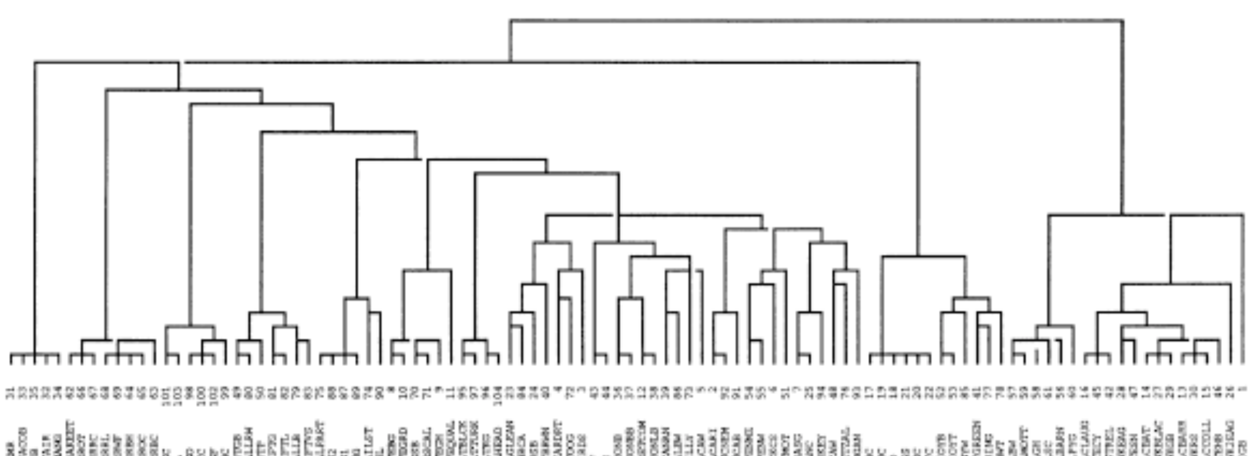
Cultures all over the world group animals into taxonomic trees

US non-experts - Tikal birds



US experts - Tikal birds



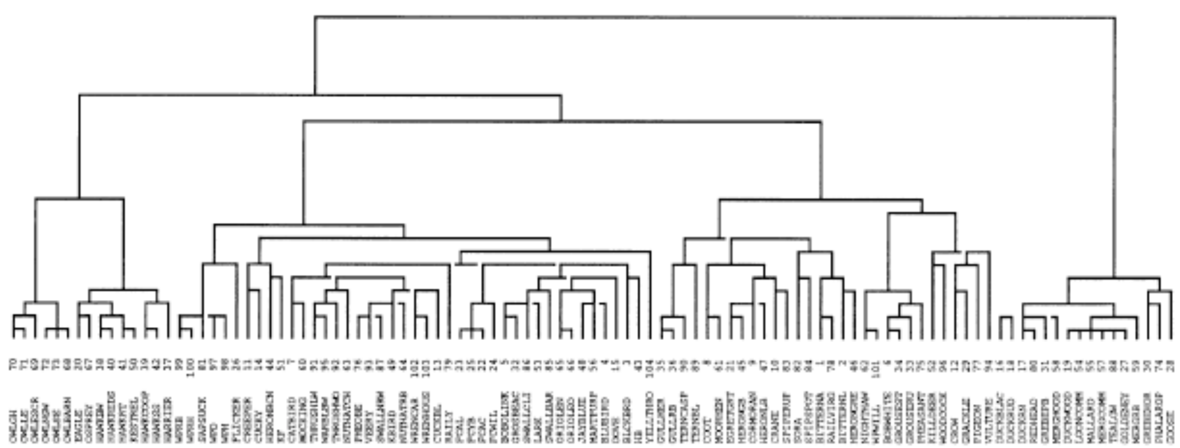


Itza' Maya - Tikal birds

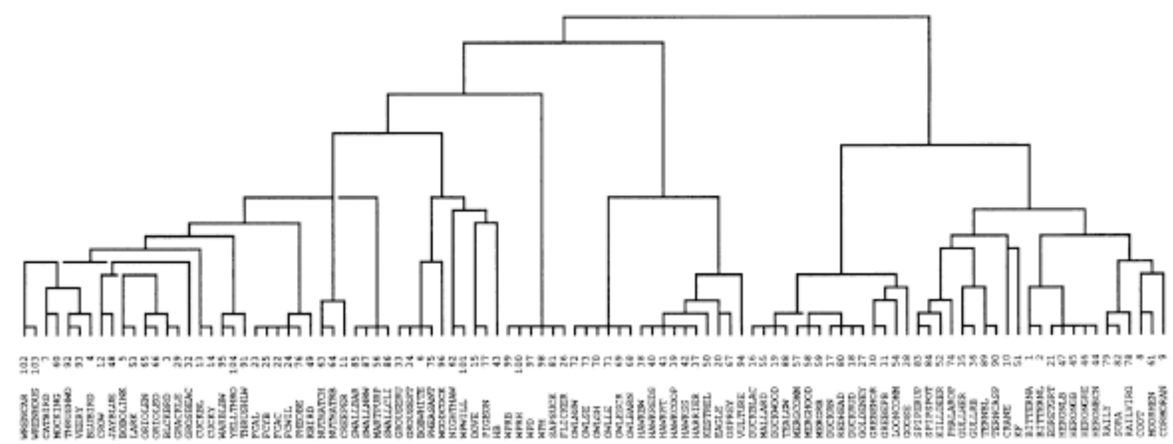
Structure in different domains: biology

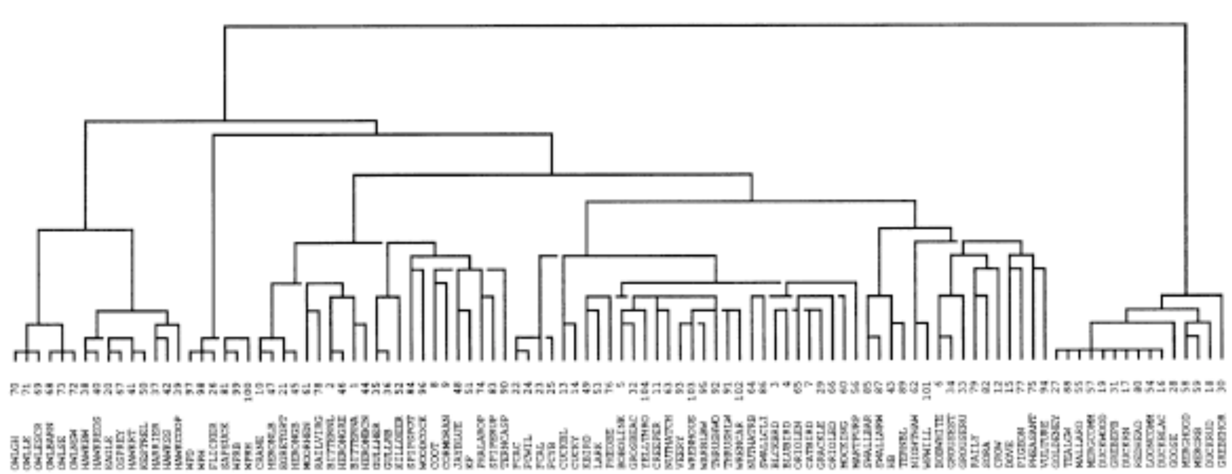
Cultures all over the world group animals into taxonomic trees

US non-experts - US birds



US experts - US birds



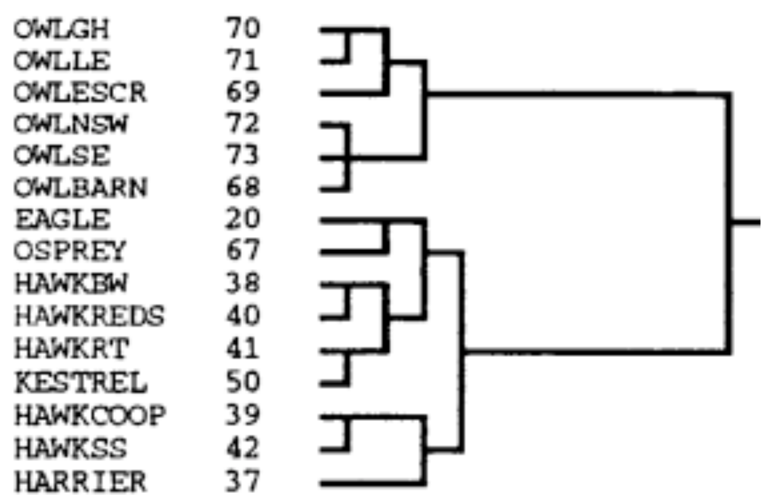


Itza’ Maya - US birds

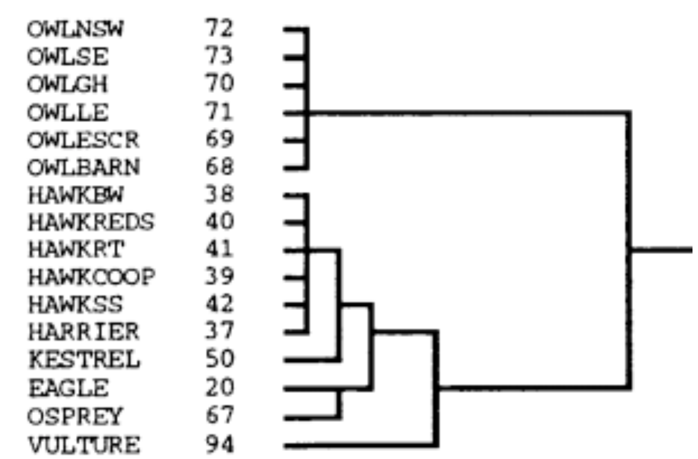
Structure in different domains: biology

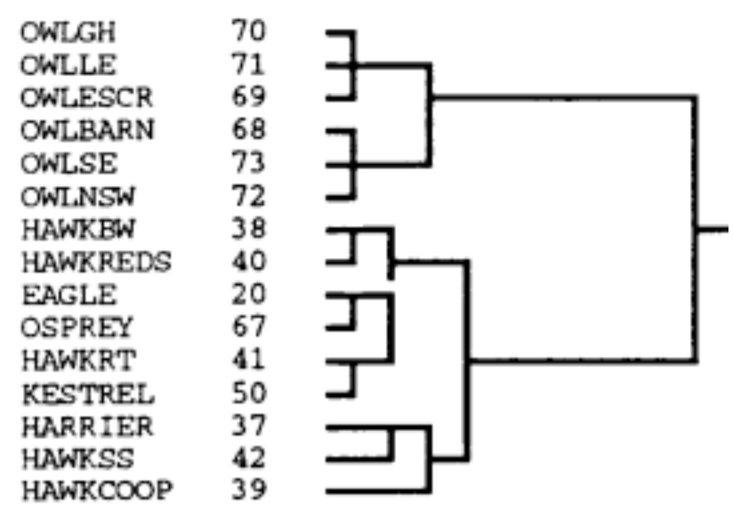
Cultures all over the world group animals into taxonomic trees... although details may differ

US non-experts - US birds



US experts - US birds



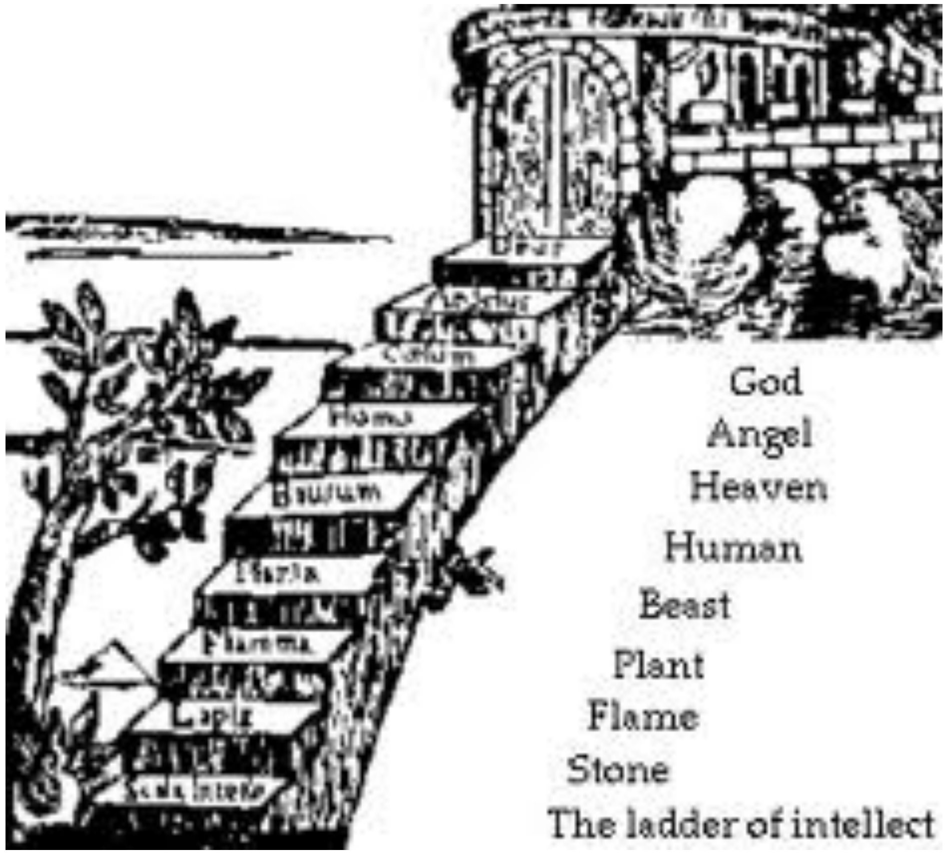


Itza' Maya - US birds

Structure in different domains: biology

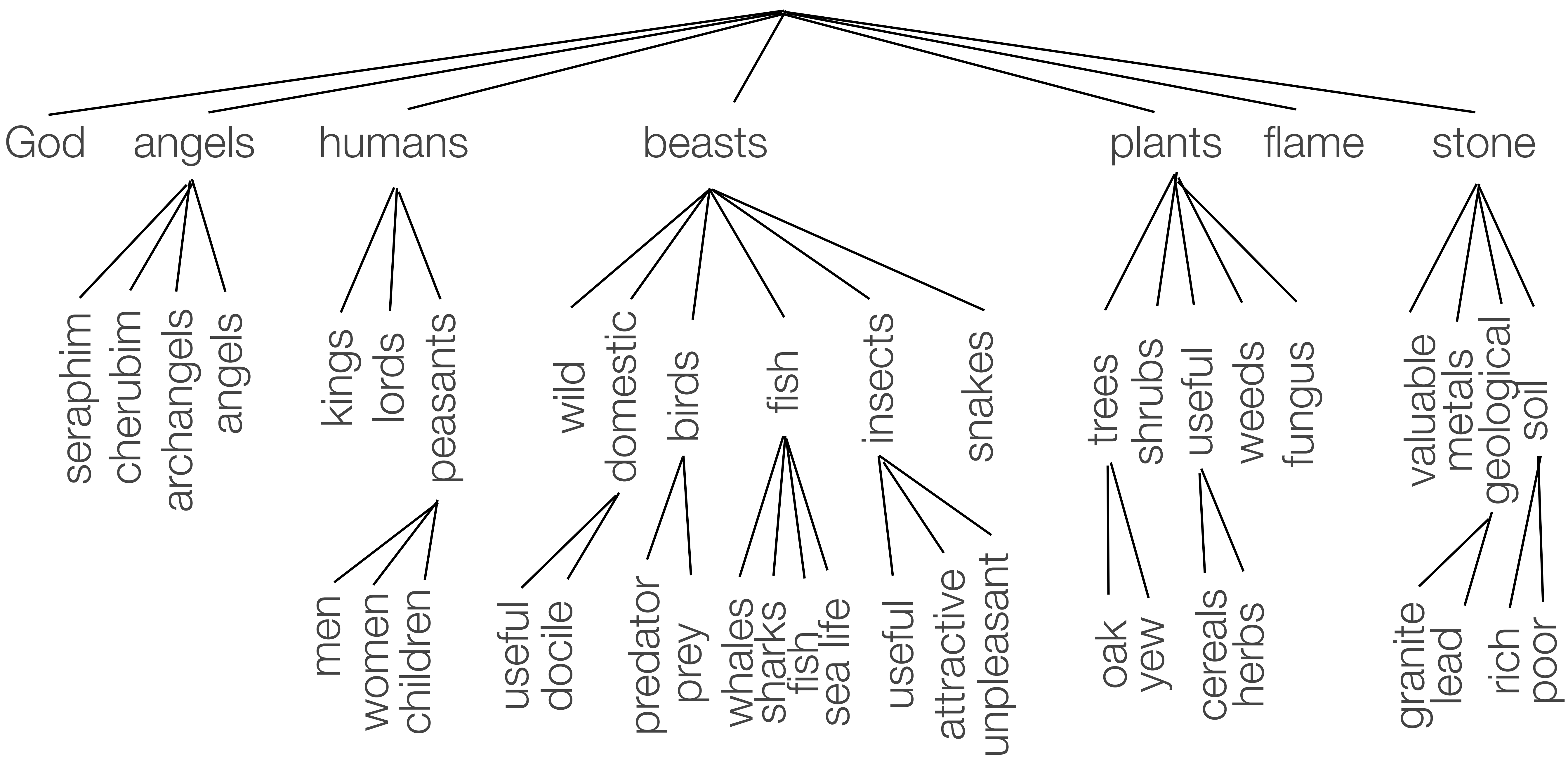
There are exceptions

God — angels — humans ——— beasts ——————— plants — flame — stone



Structure in different domains: biology

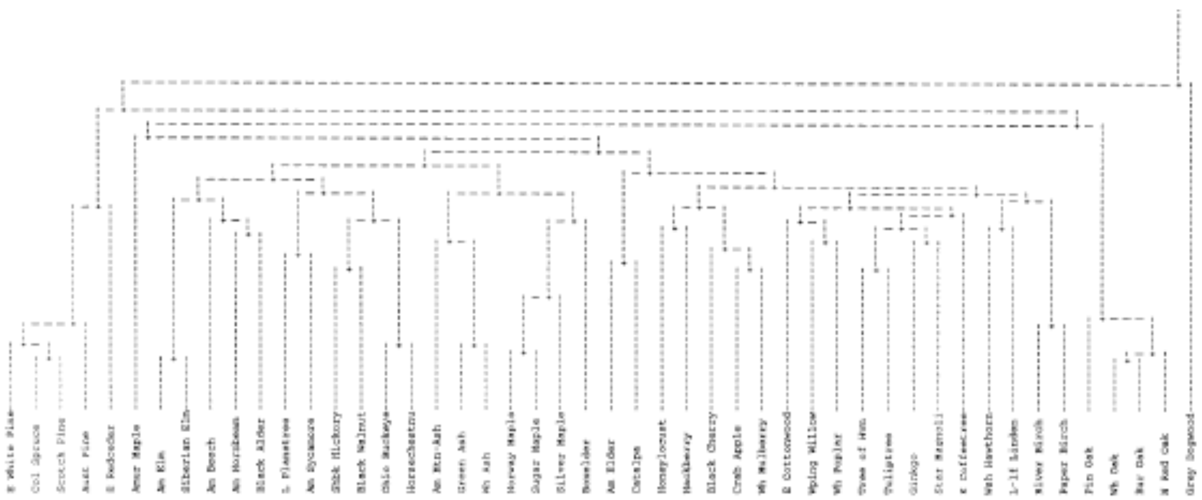
There are exceptions.. but they are very rare



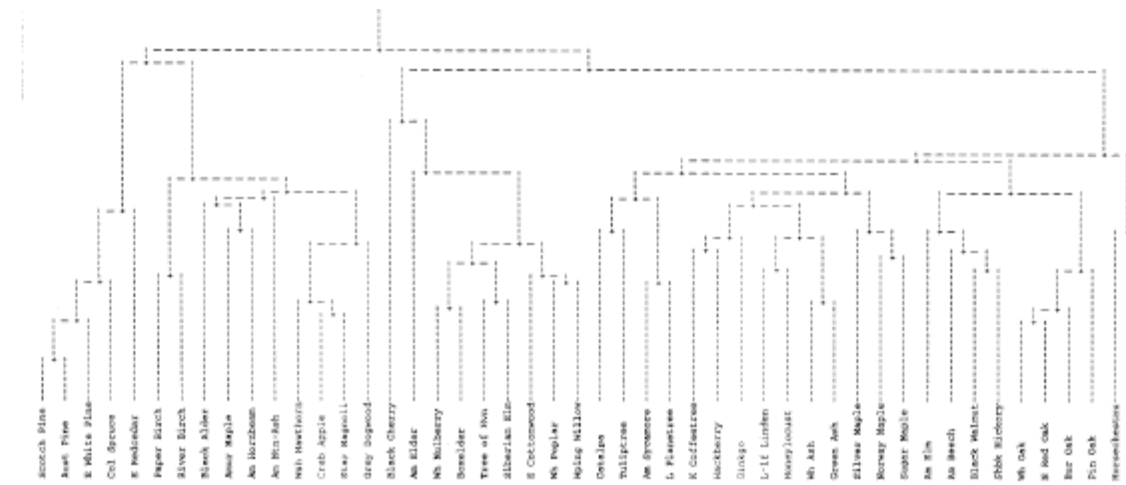
Structure in different domains: biology

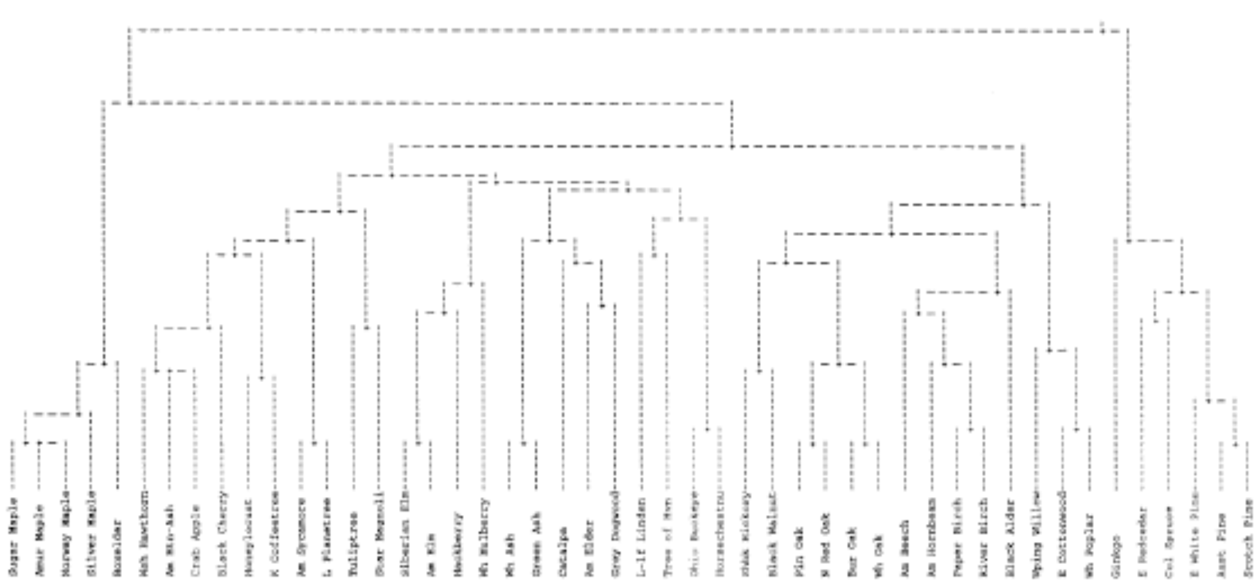
The same thing occurs for plants as well!

Maintenance workers



Landscapers

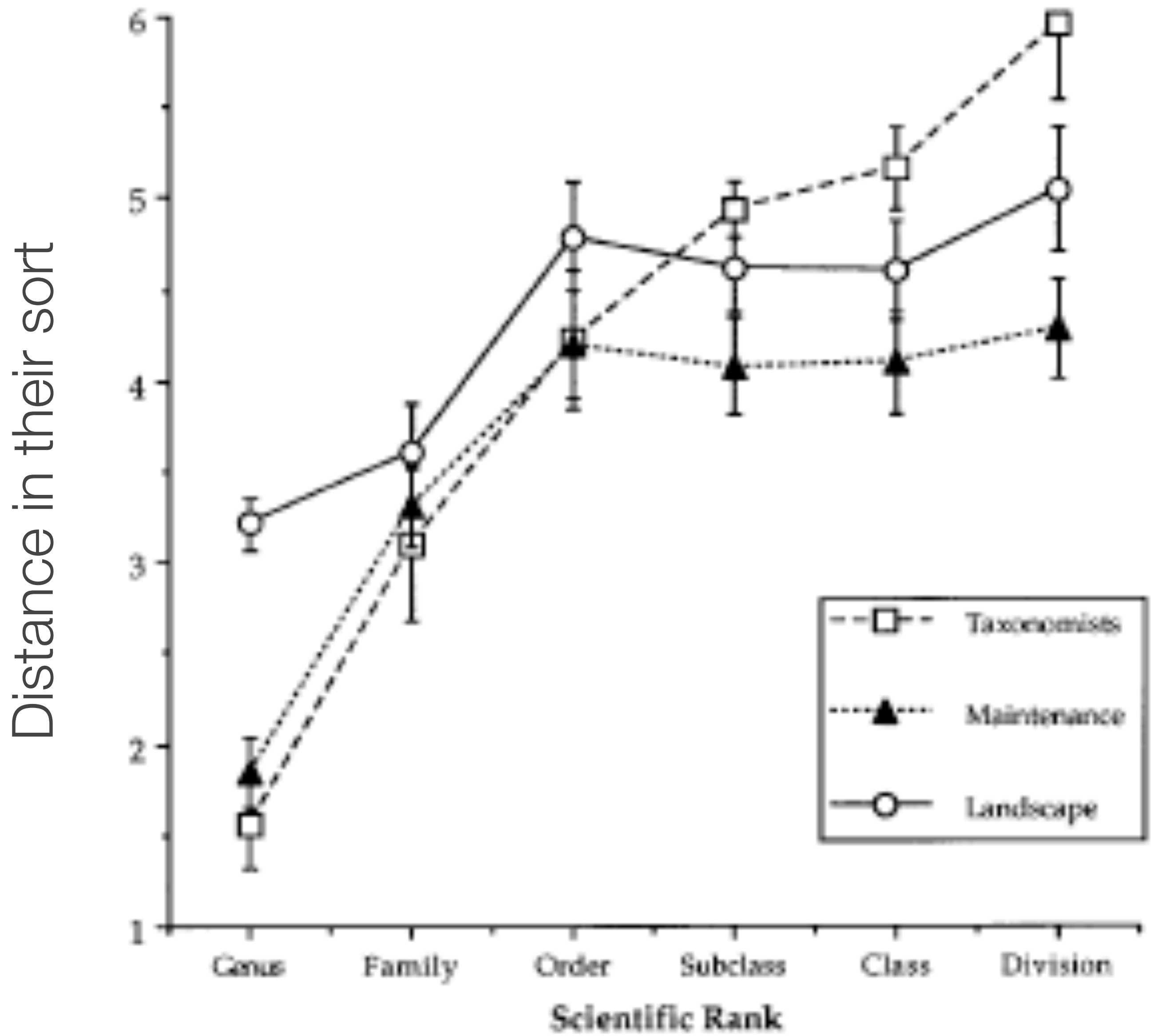




Taxonomists

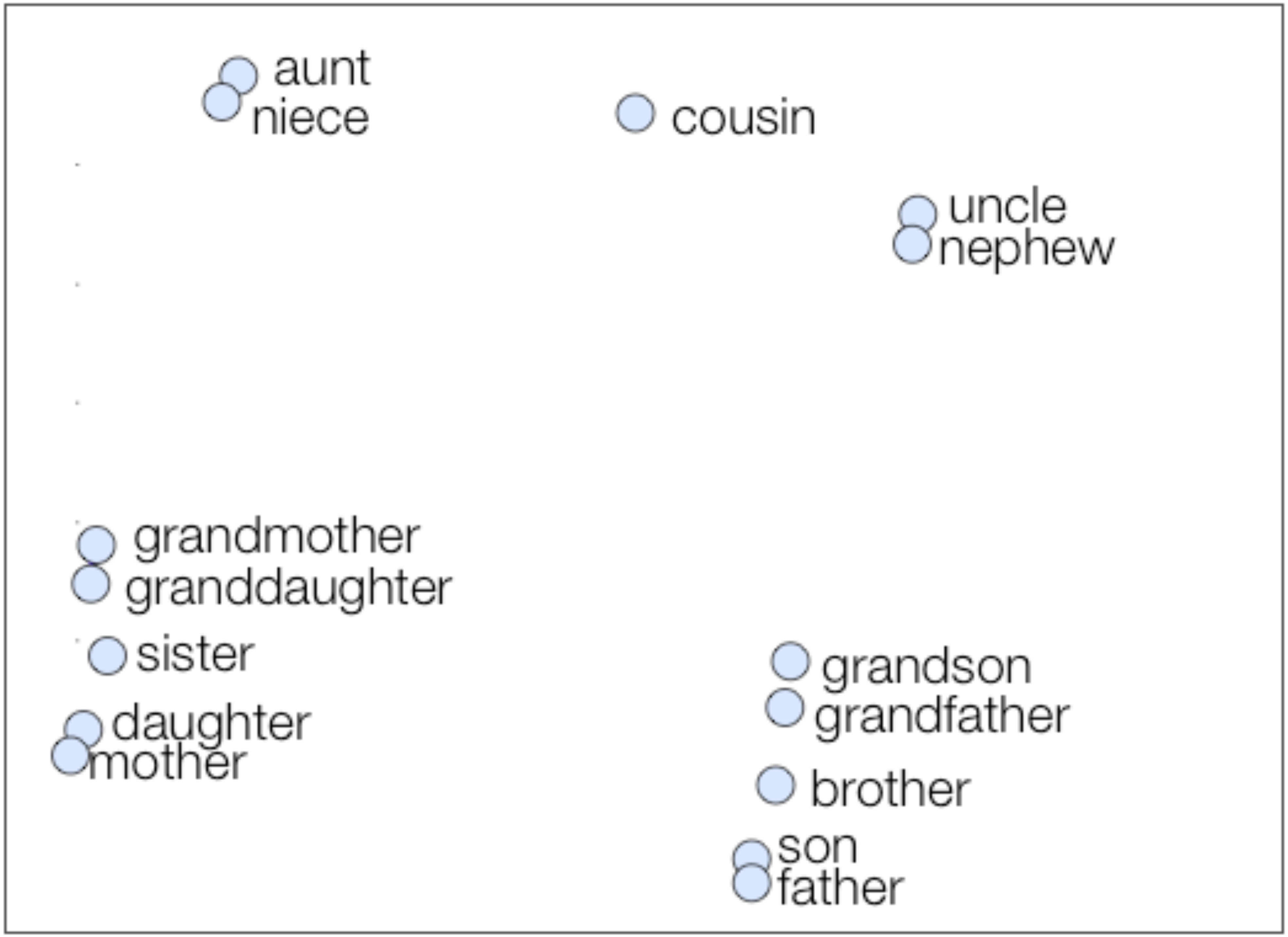
Structure in different domains: biology

The same thing occurs for plants as well!



Differences between the three reflected differences in their reliance on the taxonomy (although all of them generally followed it)

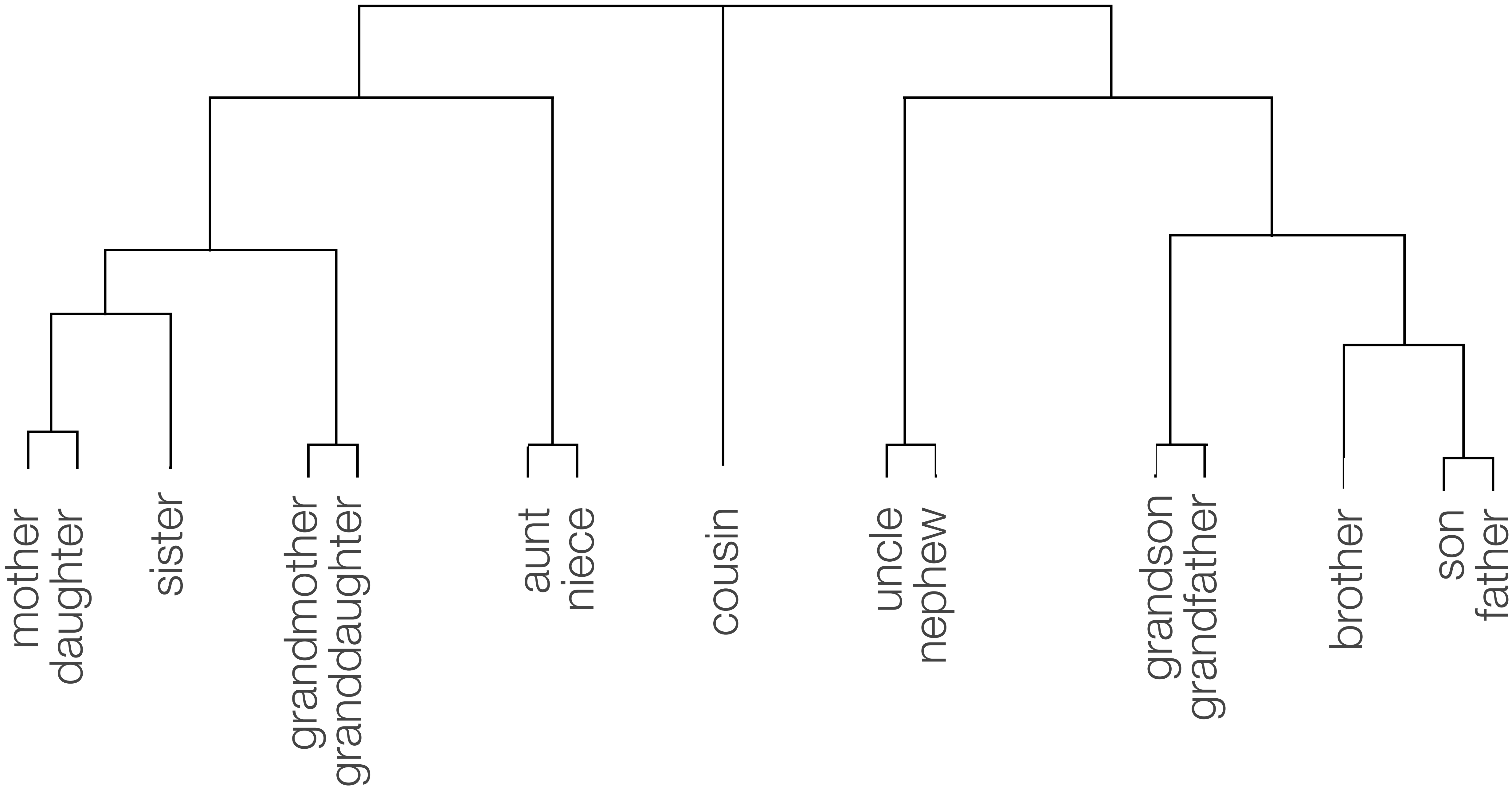
Structure in different domains: kinship



This "clumping" strongly suggests the true structure is not a space...

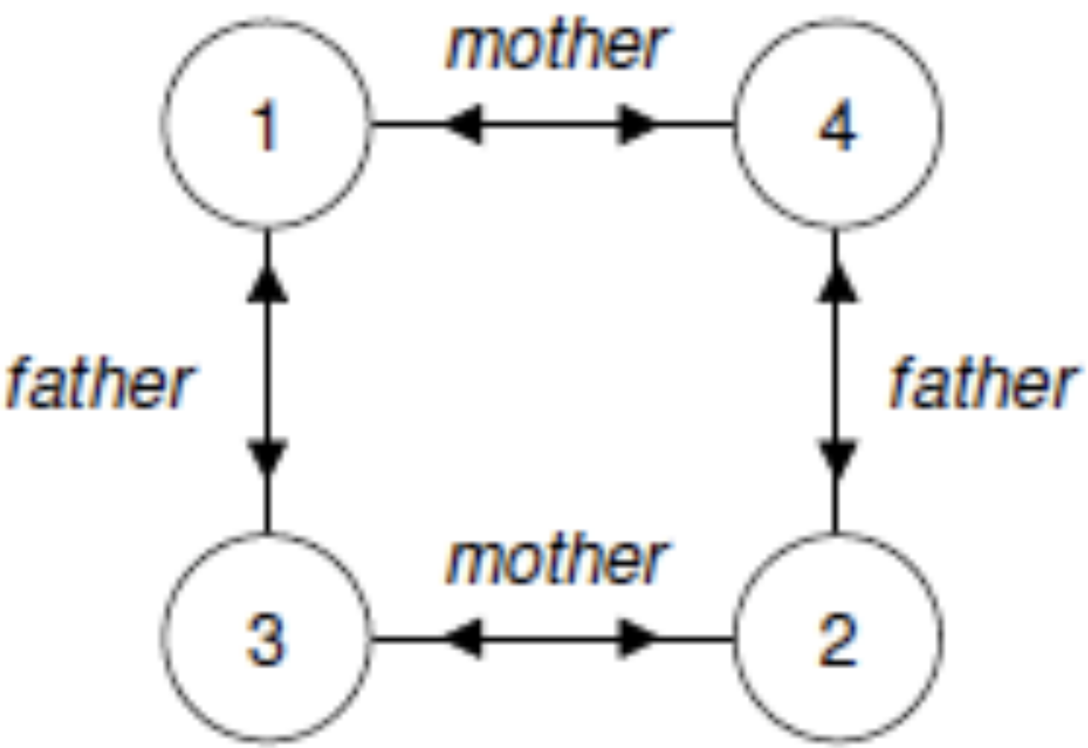
Structure in different domains: kinship

...but rather something more like this



Structure in different domains: kinship

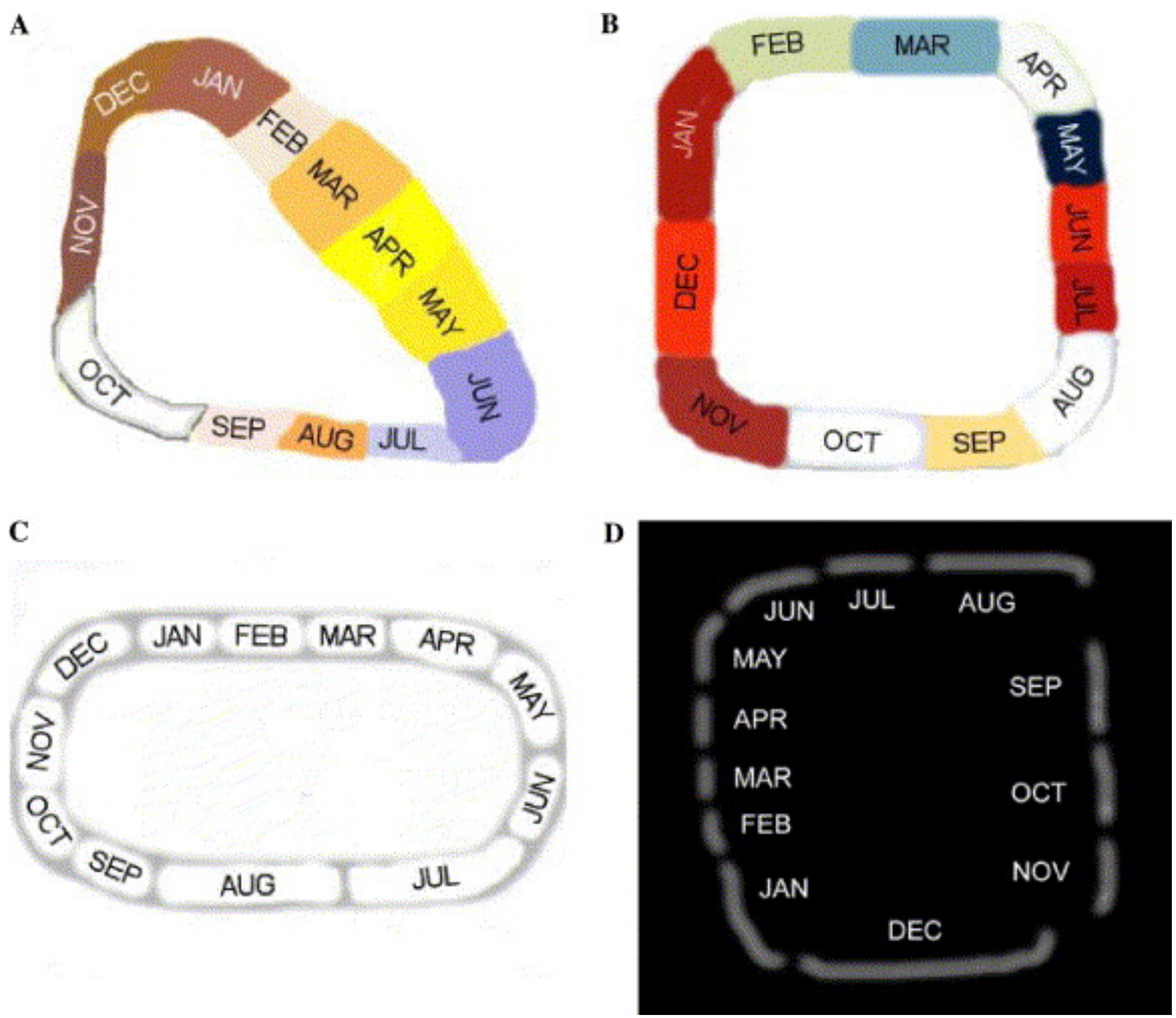
There is also some cultural differentiation!

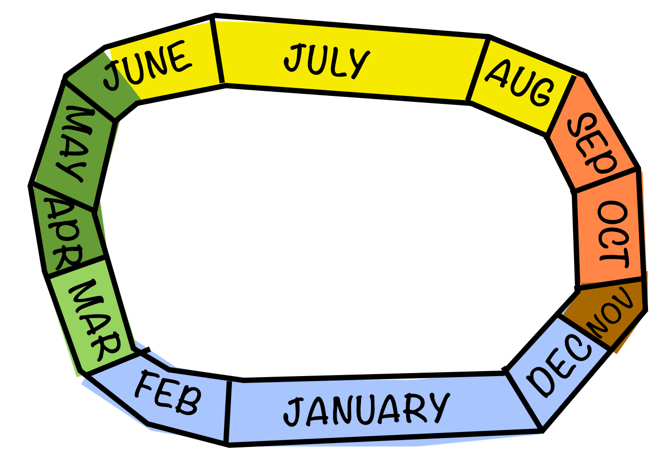


This structure is derived from the kinship terms used for each other by 104 Alyawarra tribe members (studied by an anthropologist named Denham).

People are classified into sections. Someone in section 4 has a mother in section 1 and a father in section 2

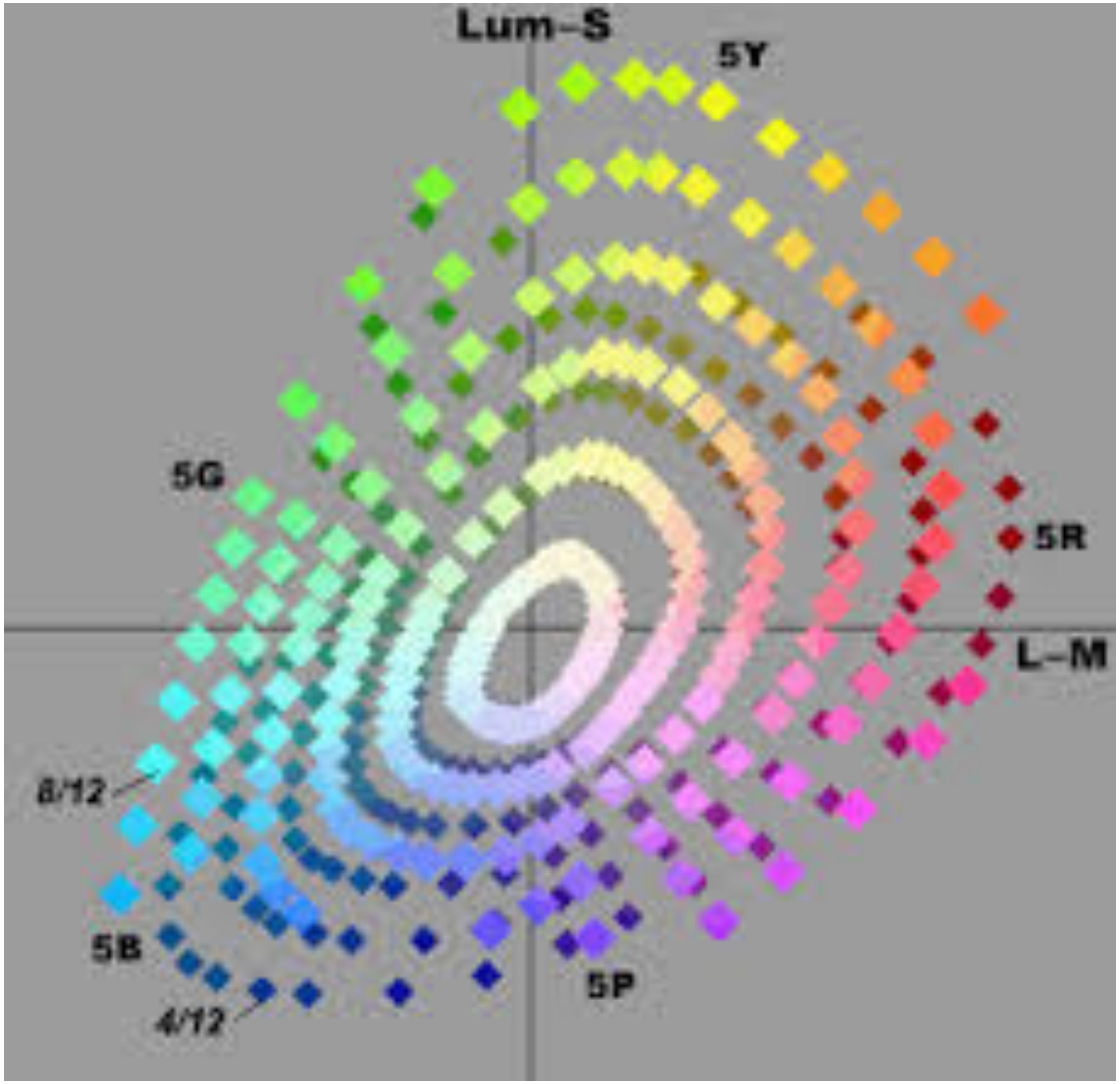
Structure in different domains: time





images from peope with time-space synesthesia

Structure in different domains: colour



perceptual space (based on similarities reported by people)

Structure in different domains: non-humans

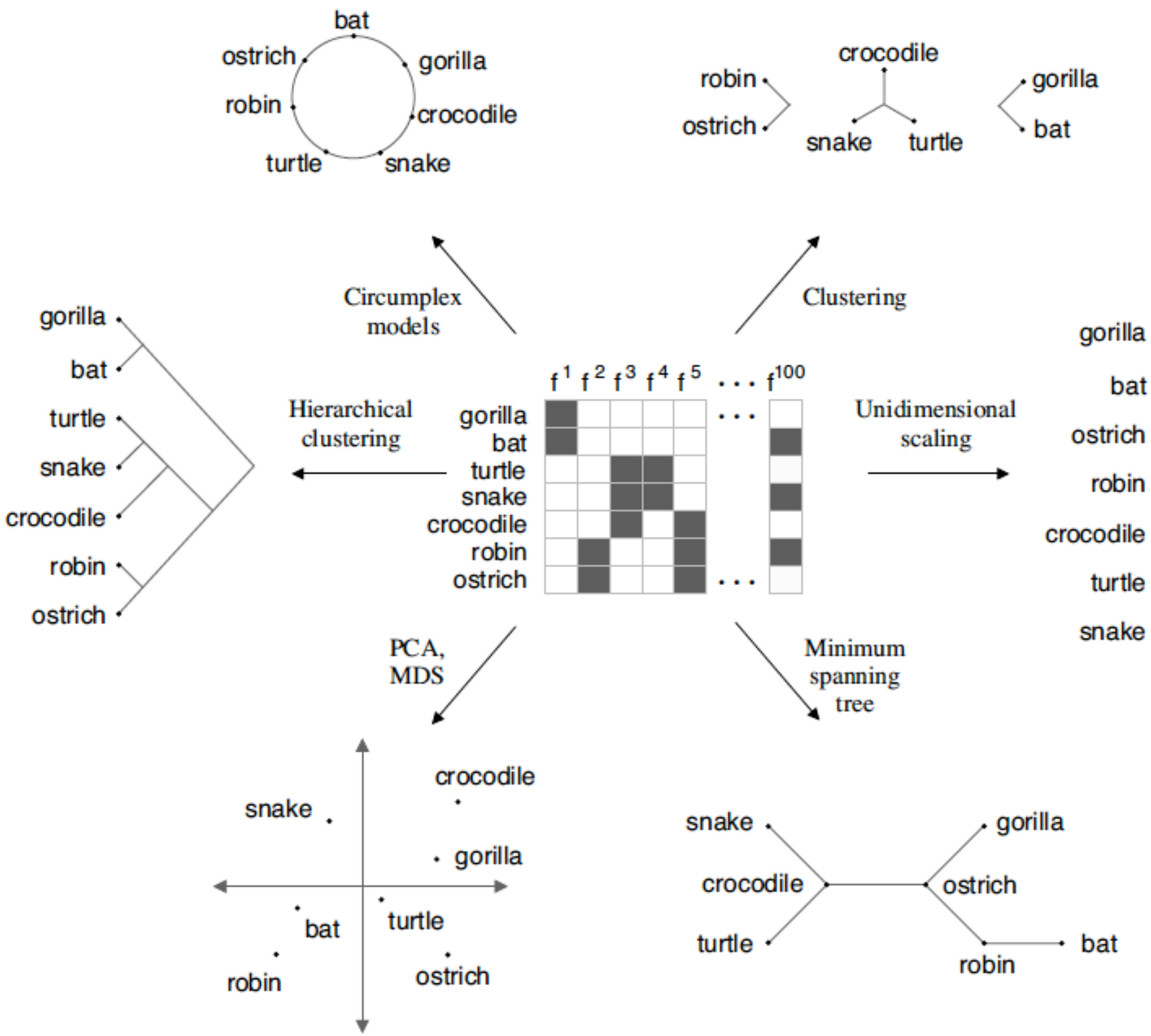
alpha

beta

low-
ranked

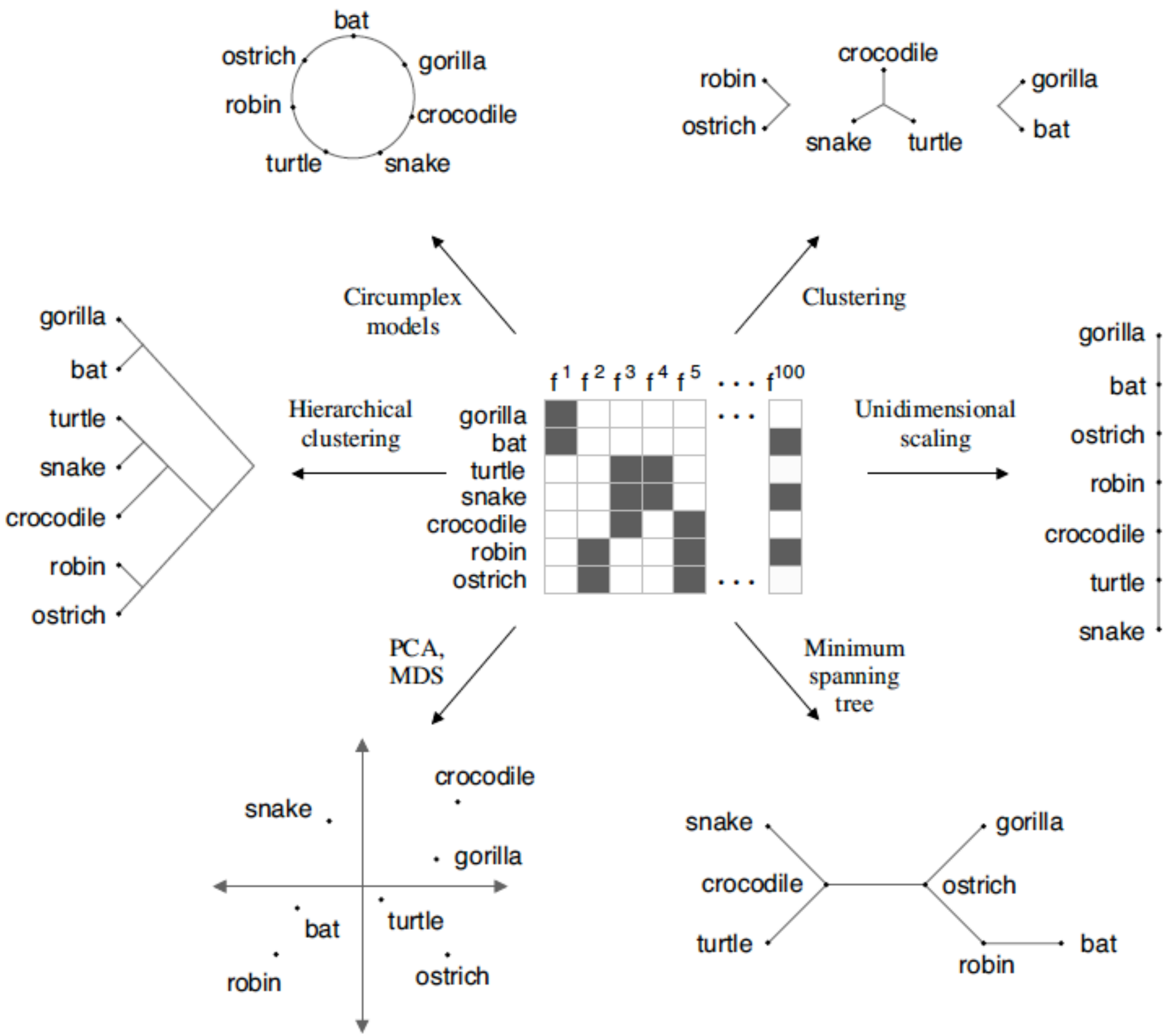
even primates
have dominance
hierarchies that
they are clearly
sensitive to!

Learning structure



We have different methods for deriving different structures given the same data...

Learning structure



...but how would a learner know what method to use?

More generally, we want to be able to learn *which structure is appropriate*

Structure in different domains: the questions

What kind of general-purpose learner could acquire *different* kinds of structures, without being told which ones were appropriate?

What is the computational problem being solved when doing this sort of structure learning?

Lecture outline (next three lectures)

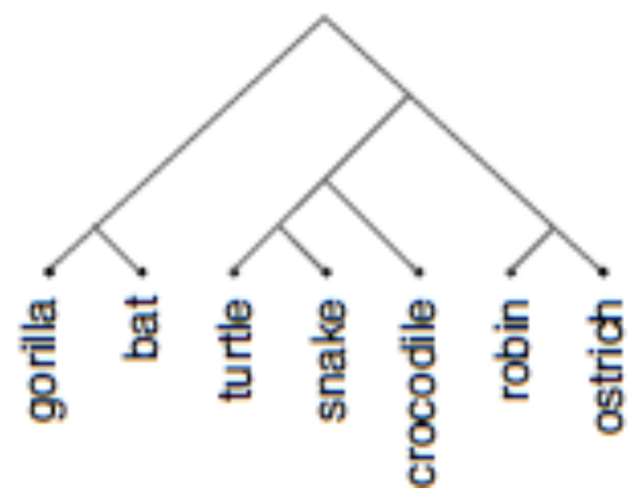
- Lecture 11: Learning about category variability
  - This kind of learning in children and adults
  - A model for this kind of learning
  - Limitations of this model
- Last time: Learning about distributions of categories
  - This kind of learning in adults
  - Failure of current models
  - A model for this kind of learning
- ➡ Today: Learning about category structure
  - This kind of learning in people
  - ➡ A model for this kind of learning

A hierarchical model of conceptual structure

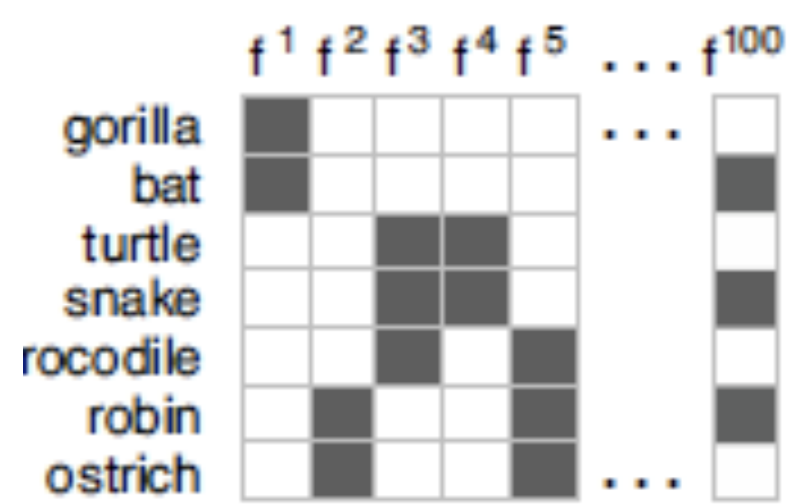
Structure "form" F

Hierarchy

Structure S



Object-feature matrix D



(also works with similarity data, but feature data is more intuitive to explain)

A hierarchical model of conceptual structure

Structure
"form" F

$\downarrow$

Structure S

$\downarrow$

Object-feature
matrix D

Given D, choose S and F
that maximise

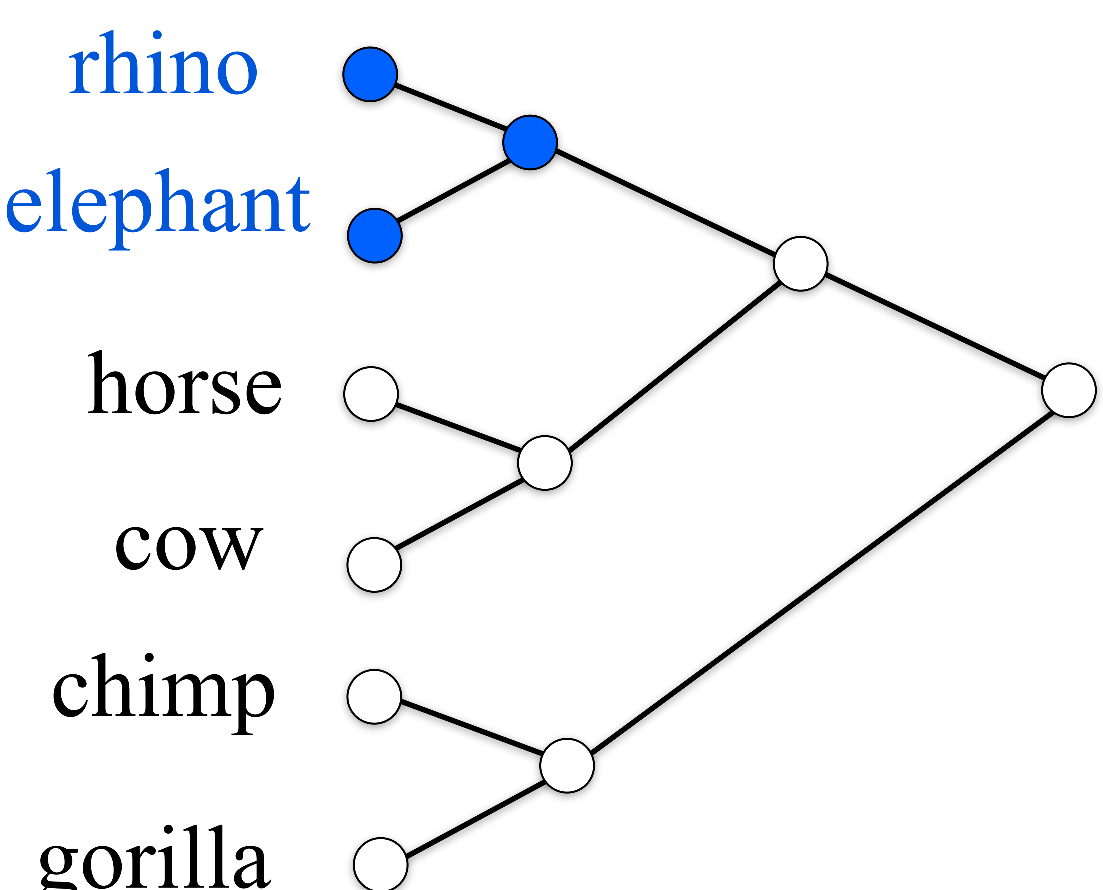
$$p(S, F|D) \propto p(D|S)p(S|F)p(F)$$

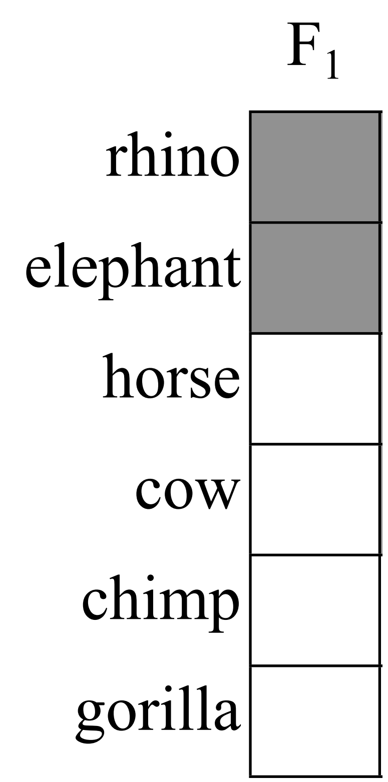
Questions

- How do you pick a structure that “fits” some data well? (in other words, how is data generated from a structure?)
- How do we prevent the model from simply picking the most complex structures possible? (in other words, what prior is placed on structures, to prefer simple ones?)
- Where do all these structures come from? (in other words, how is a “structure form” chosen?)
- How well does this model do at coming up with the correct structures based on object-feature data?

Fitting the data to a structure: Intuition

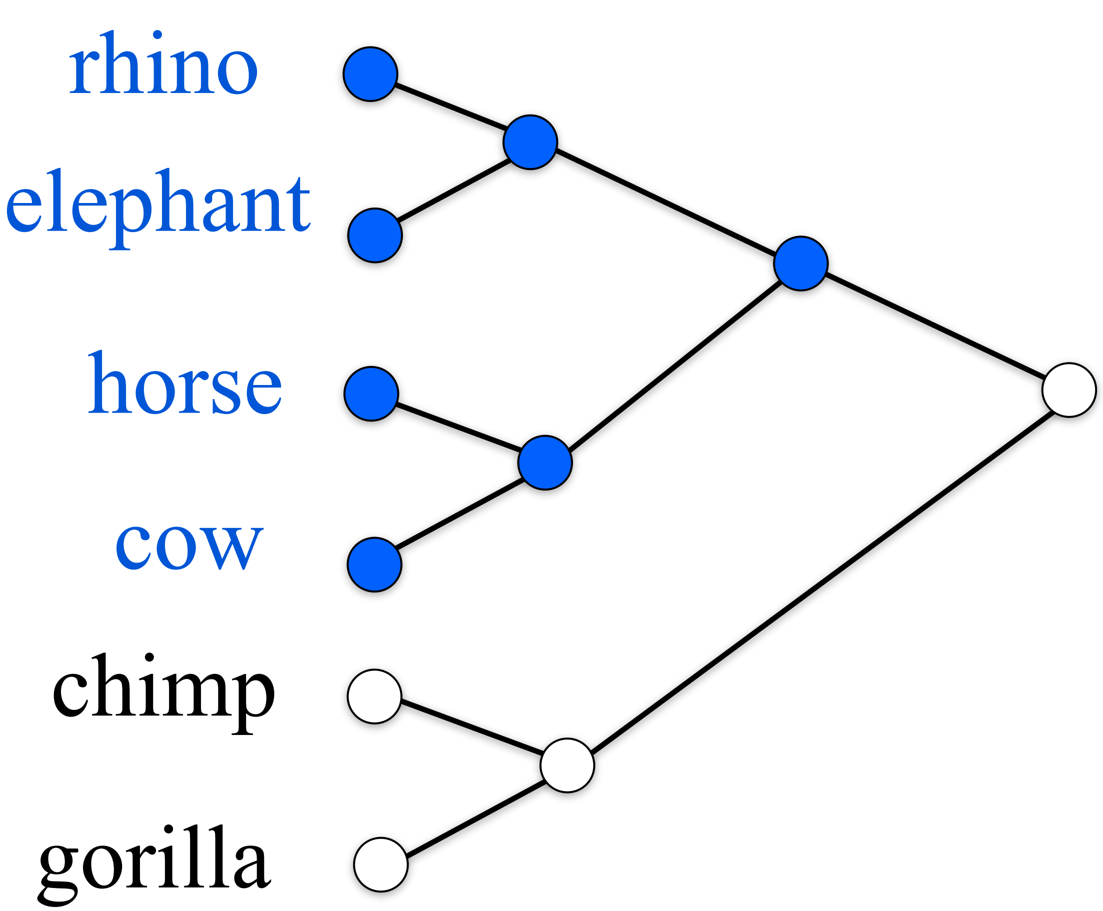
Some features “track” an underlying structure, and others do not

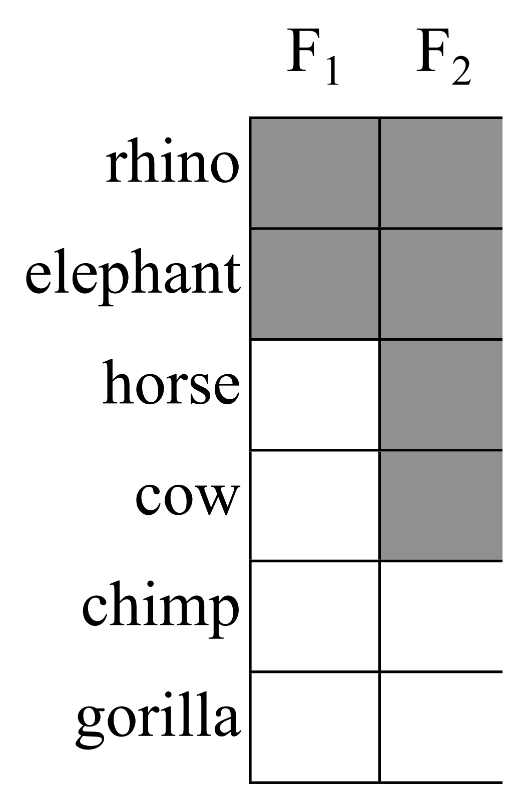




Fitting the data to a structure: Intuition

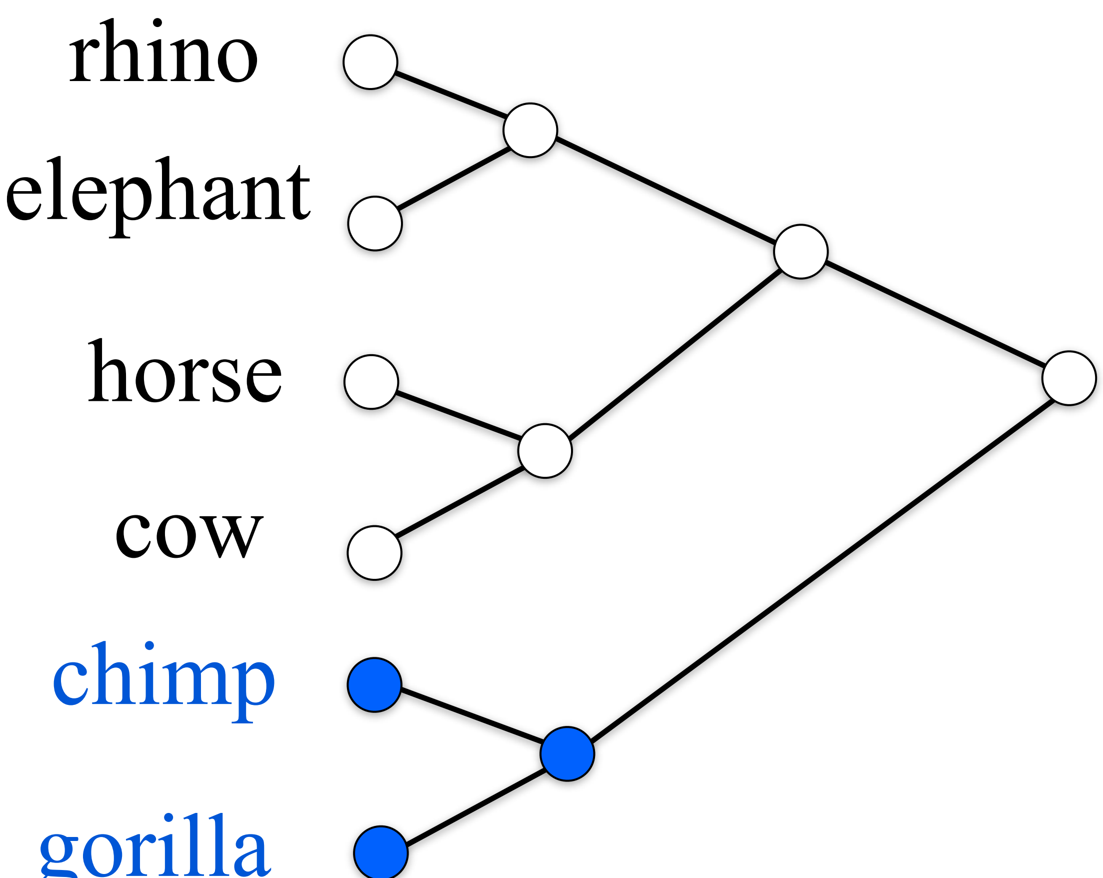
Some features “track” an underlying structure, and others do not

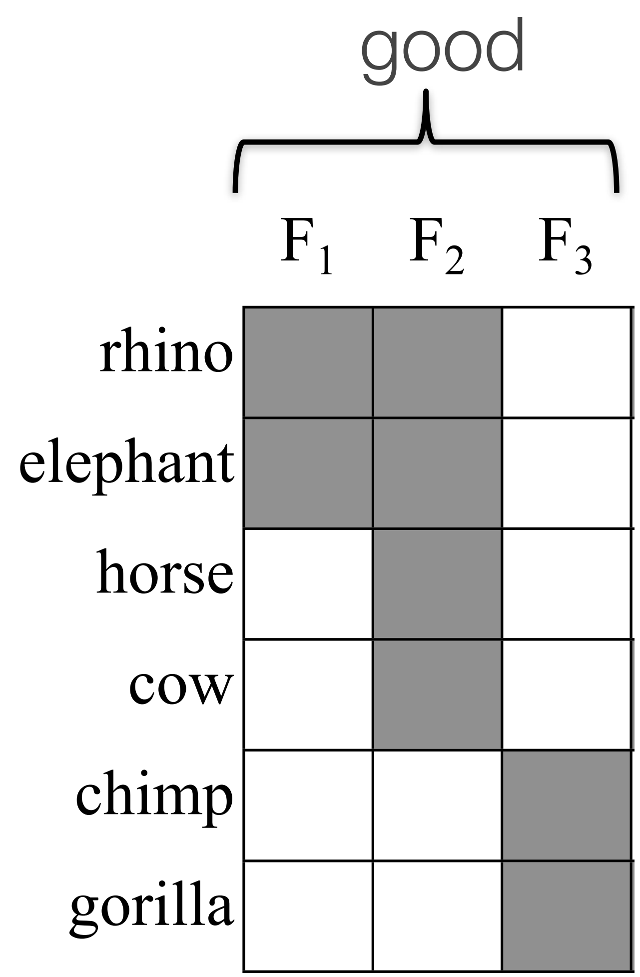




Fitting the data to a structure: Intuition

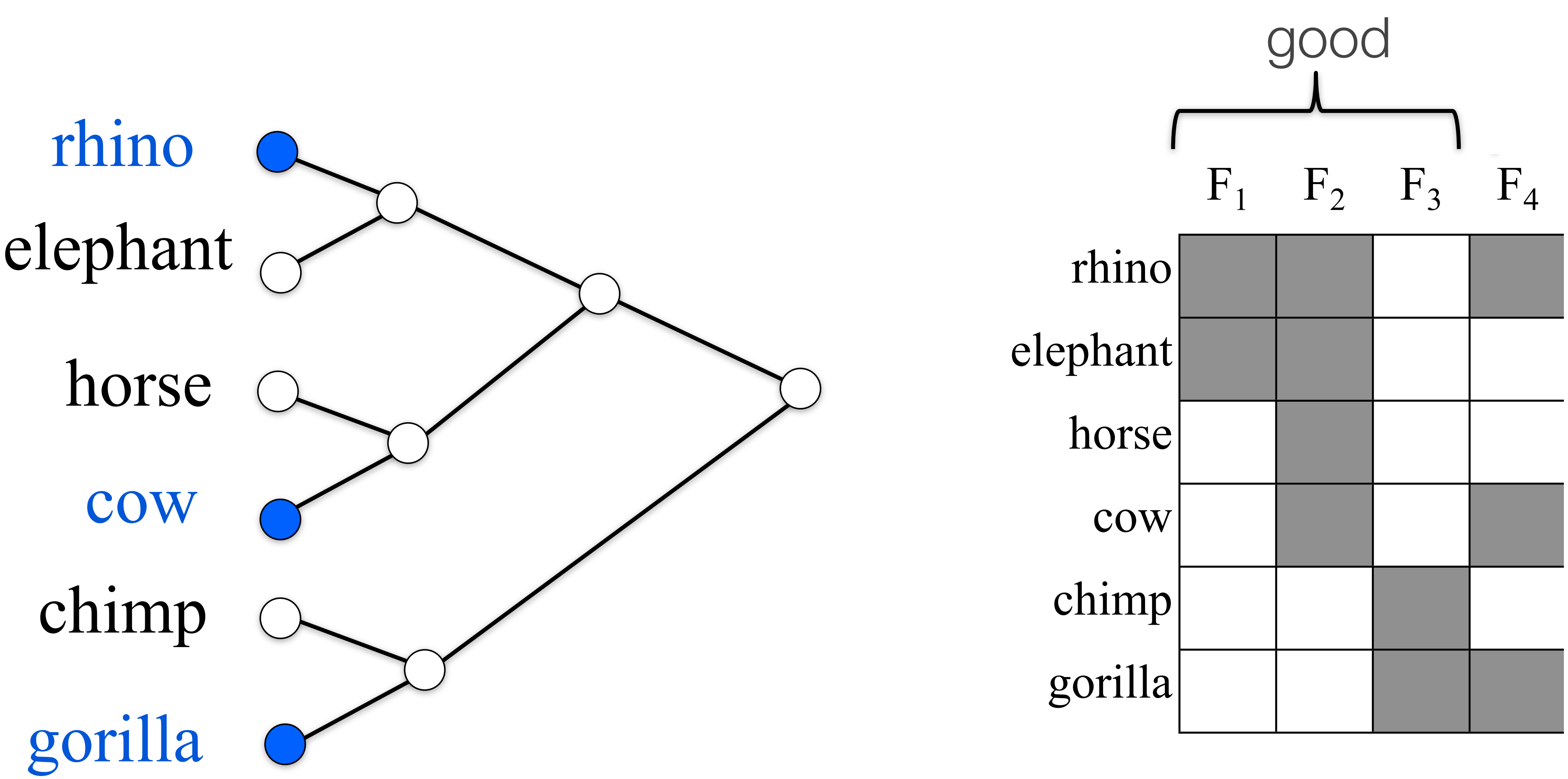
Some features "track" an underlying structure, and others do not





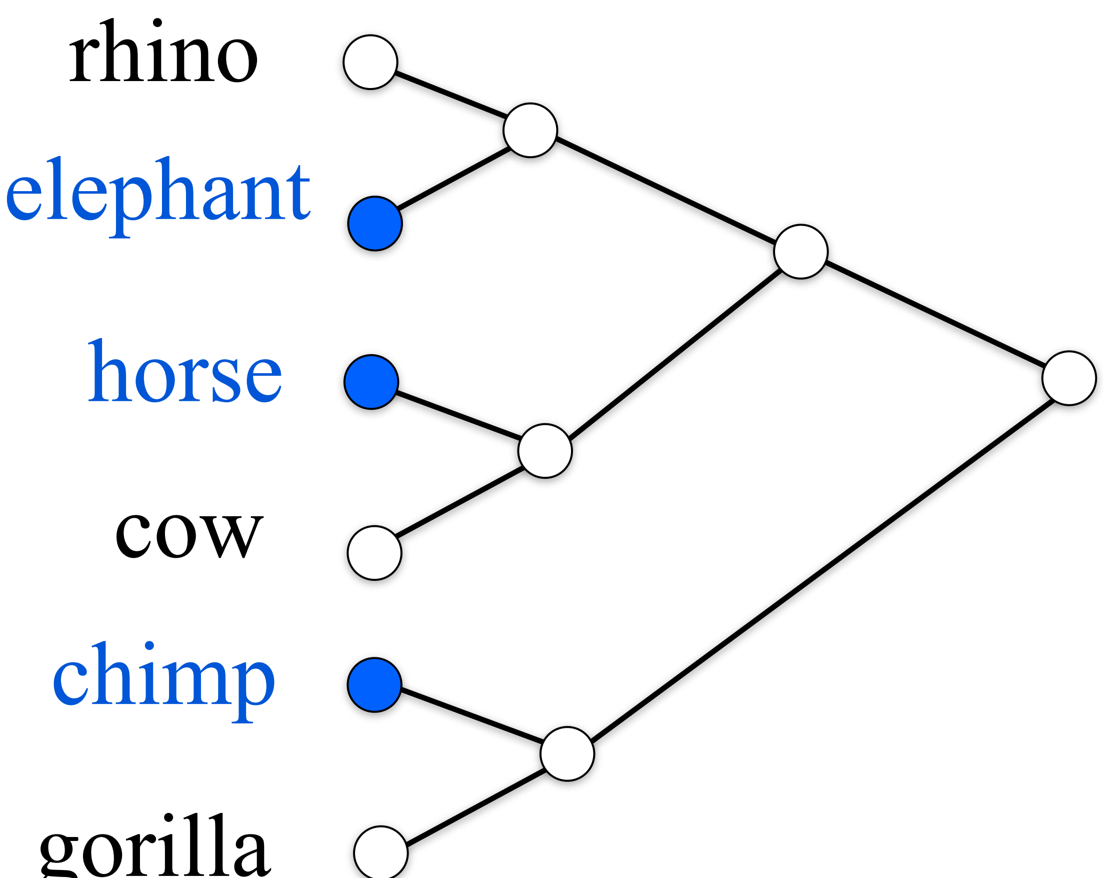
Fitting the data to a structure: Intuition

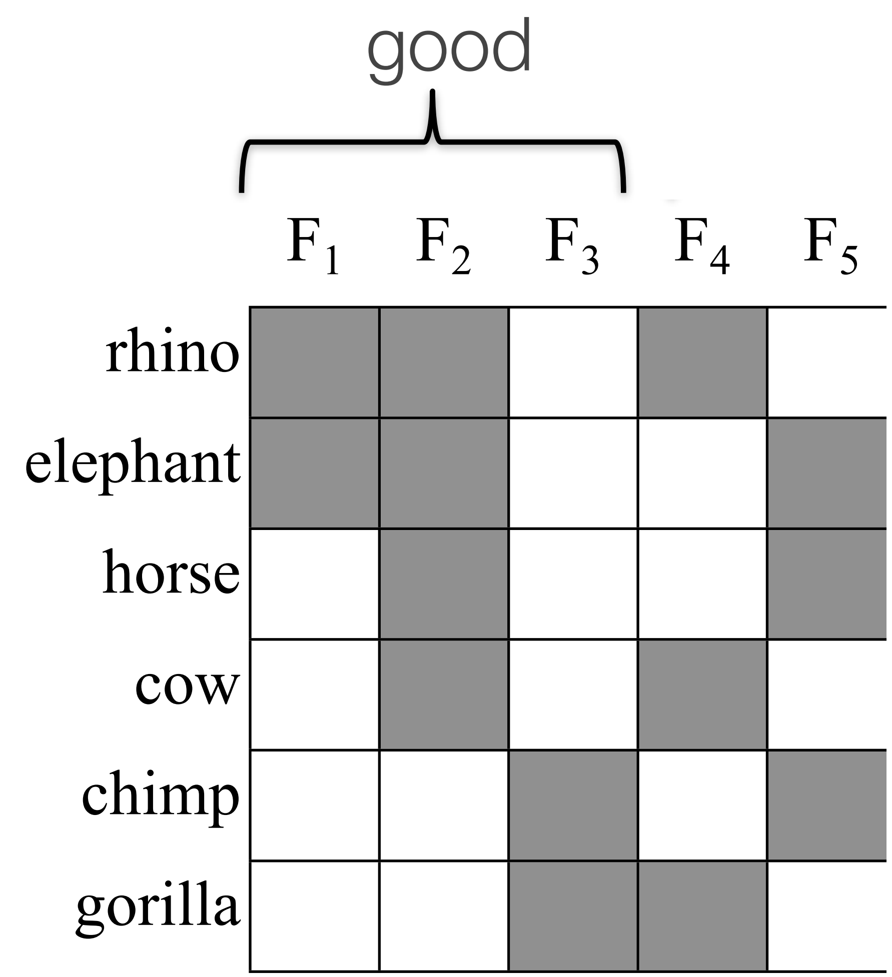
Some features “track” an underlying structure, and others do not



Fitting the data to a structure: Intuition

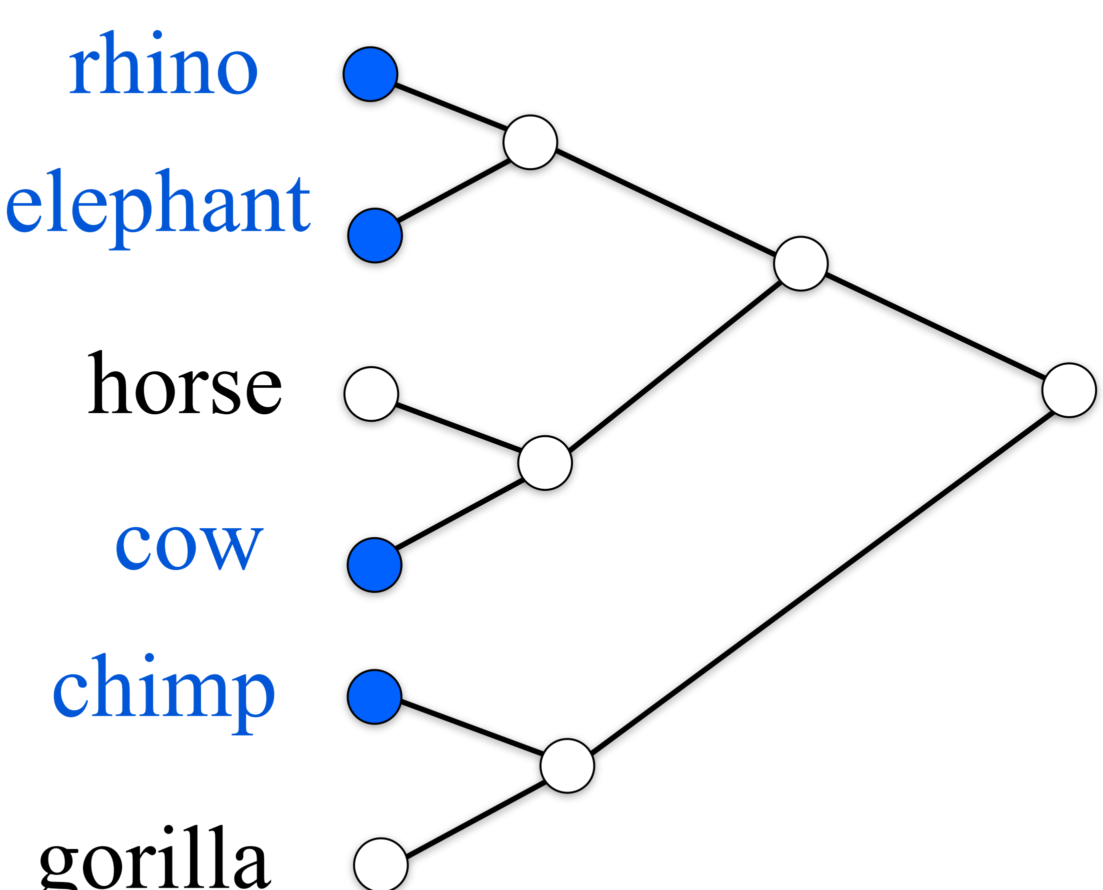
Some features "track" an underlying structure, and others do not

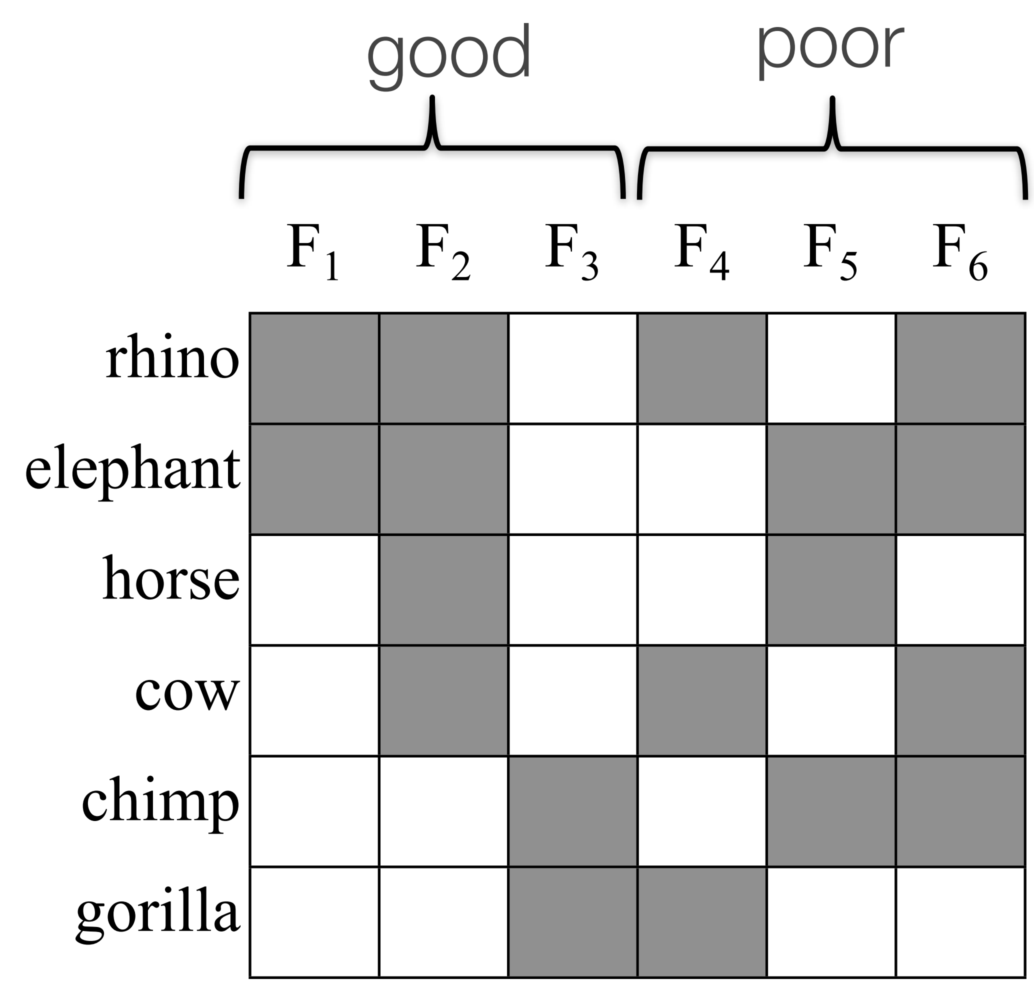




Fitting the data to a structure: Intuition

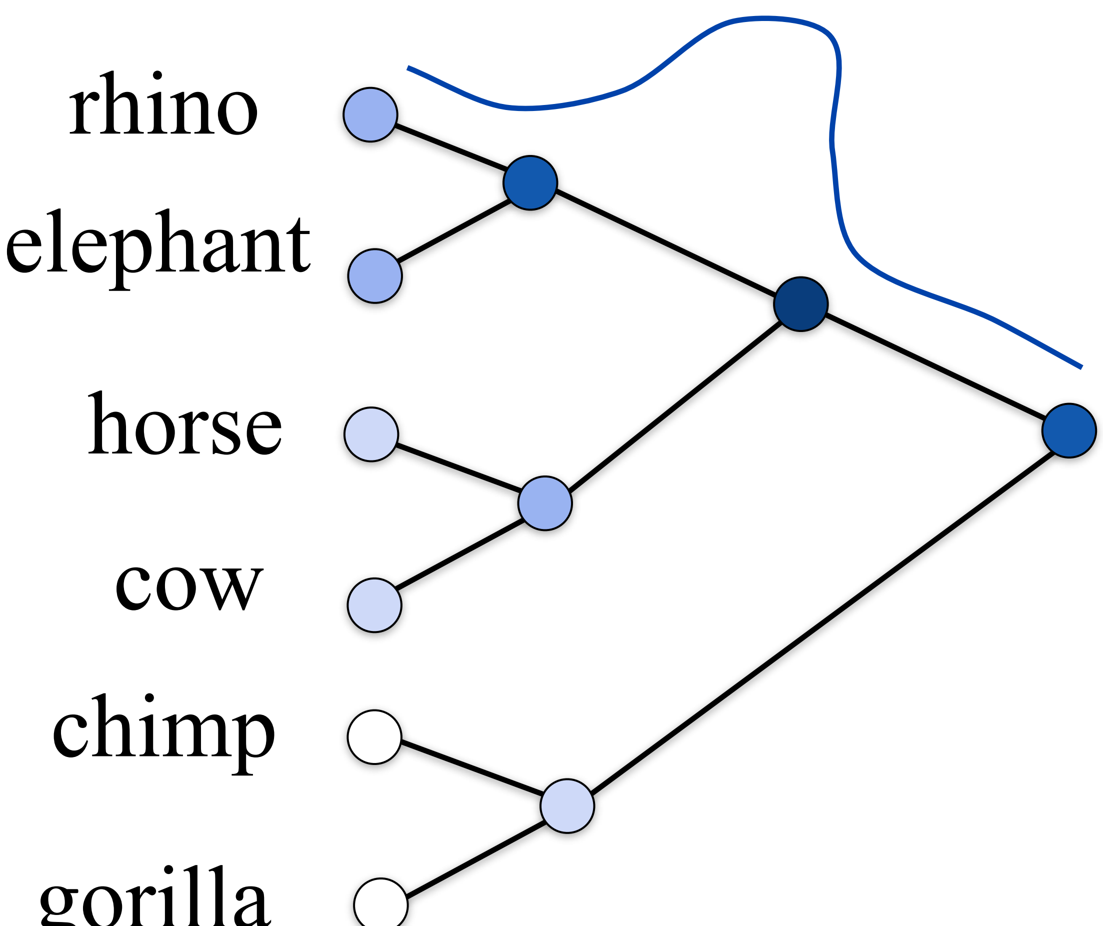
Some features “track” an underlying structure, and others do not





Fitting the data to a structure: Formalisation

Assume that features are independently generated from a Gaussian distribution* over the graph



W is a weight matrix, where $w_{ij} = 1/e_{ij}$ if nodes i and j are joined by an edge of length e_{ij} and $w_{ij}=0$ otherwise

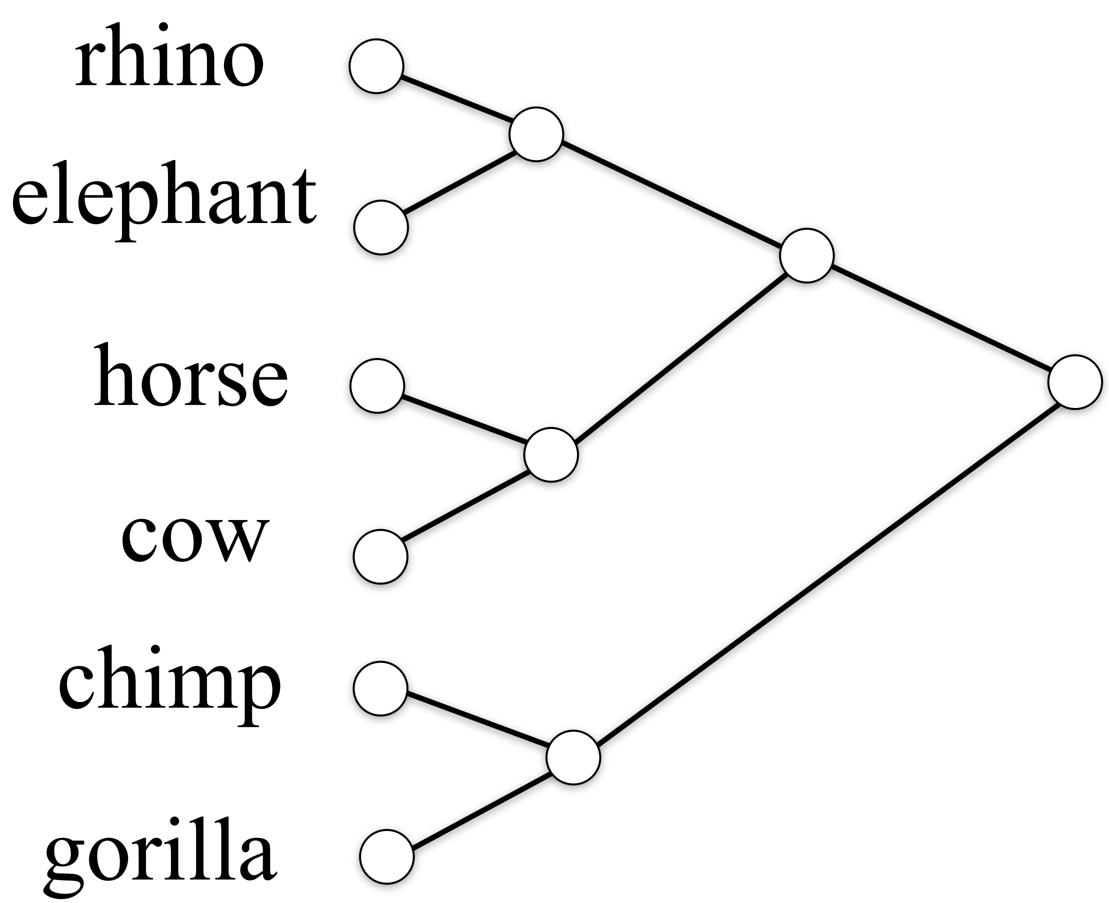
$$P(f|W) \propto \exp\left(-\frac{1}{4}\sum_{i,j} w_{ij}(f_i - f_j)^2\right)$$

This penalises a feature vector if $f_i \neq f_j$ and i and j are adjacent in the graph. The penalty increases if the edge between them is shorter.

* Need to also make assumptions about the variance of the Gaussian for the prior to be proper.

Fitting the data to a structure: Formalisation

Assume that features are independently generated from a Gaussian distribution* over the graph



Favours shorter branch lengths and Gaussians with shorter variance by putting a prior over both:

$$\sigma | \beta \sim \text{Exponential}(\beta)$$

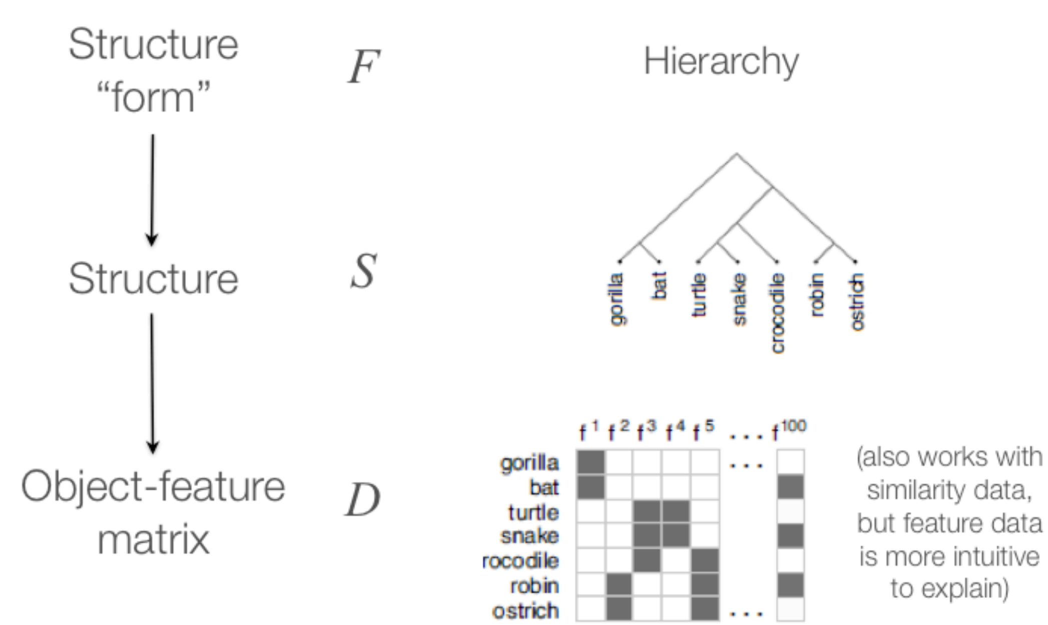
$$e_{ij} | \beta, S \sim \text{Exponential}(\beta) \text{ if } s_{ij} = 1$$

Fitting the data to a structure: Formalisation

Since the thing we actually care about is the structure itself, we integrate out the variances and edge weights

$$p(D|S) = \int p(D|S, W, \sigma^2)p(W|S)p(\sigma^2)dW\,d\sigma^2$$

$$p(S, F|D) \propto p(D|S)p(S|F)p(F)$$



Questions

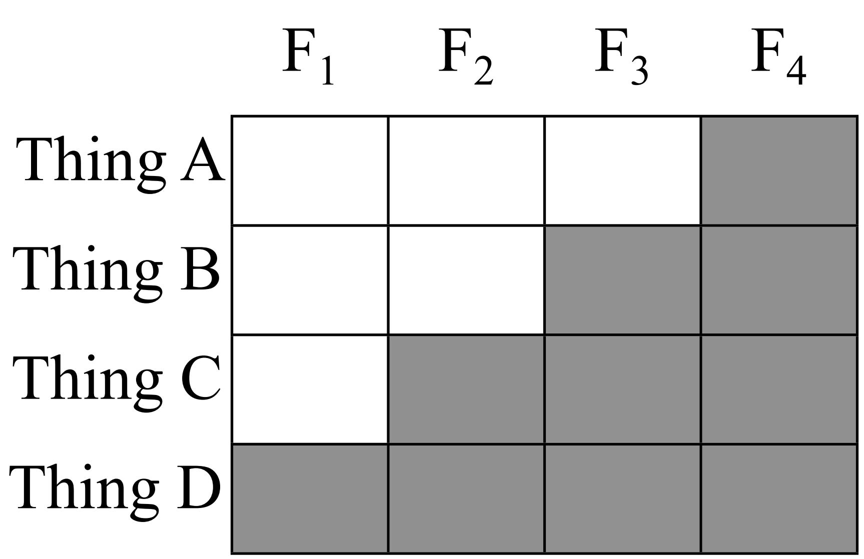
- How do you pick a structure that "fits" some data well? (in other words, how is data generated from a structure?)

➡ How do we prevent the model from simply picking the most complex structures possible? (in other words, what prior is placed on structures, to prefer simple ones?)

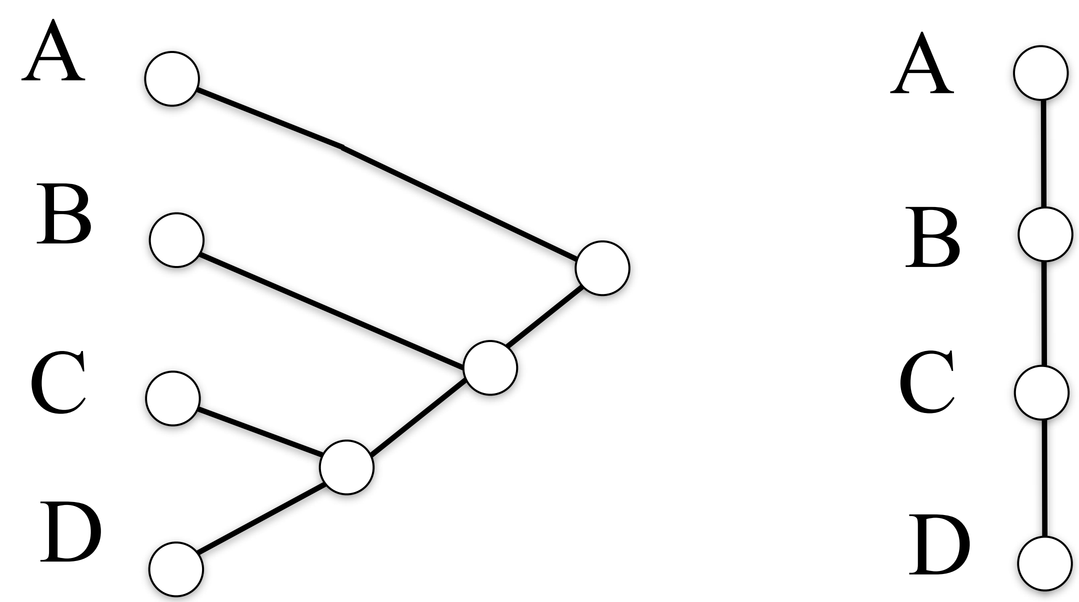
- Where do all these structures come from? (in other words, how is a "structure form" chosen?)
- How well does this model do at coming up with the correct structures based on object-feature data?

Favouring simpler structures: The issue

Suppose you saw this data:



It is consistent with both of these options



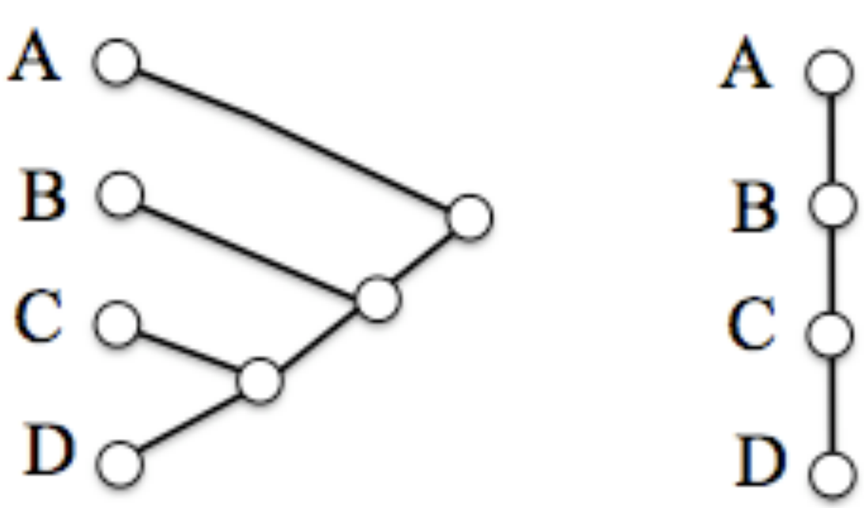
Intuitively, we want to favour the chain, because it seems simpler

Favouring simpler structures: The solution

Set a prior that favours structures with fewer nodes

$$P(S|F) \propto \begin{cases} 0 & \text{if } S \text{ is incompatible with } F \\ \theta^{|S|} & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1$, and $|S|$ is the number of nodes in S



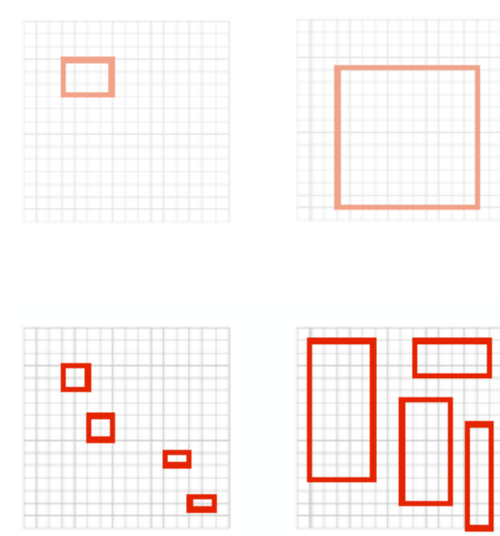
The chain is therefore favoured *a priori*, since it has only 4 nodes and the hierarchy has 7

Favouring simpler structures: One complexity

$$P(S|F) \propto \begin{cases} 0 & \text{if } S \text{ is incompatible with } F \\ \theta^{|S|} & \text{otherwise,} \end{cases}$$

The normalising constant for this is going to be different depending on what the form is (hierarchy, chain, etc), because there are more possible ways to make a hierarchy than a chain.

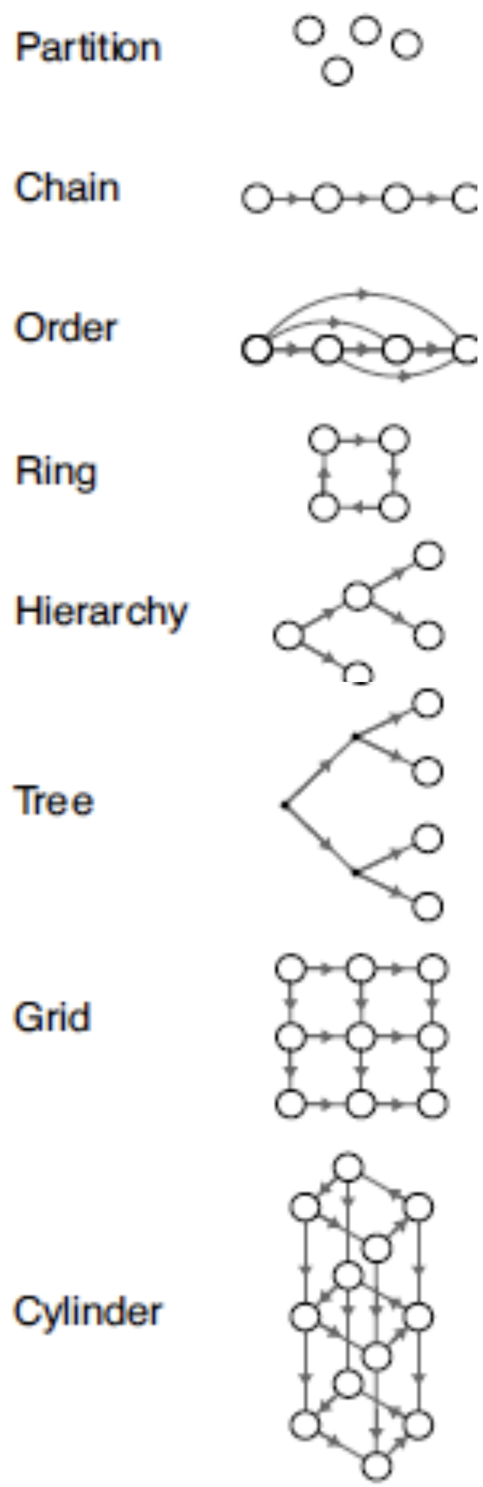
this is another way the model favours simpler structures - for the very same reason we favoured fewer rectangles in the rectangle world: there are more things to spread the same probability mass over



Questions

- How do you pick a structure that “fits” some data well? (in other words, how is data generated from a structure?)
- How do we prevent the model from simply picking the most complex structures possible? (in other words, what prior is placed on structures, to prefer simple ones?)
- ➡ Where do all these structures come from? (in other words, how is a “structure form” chosen?)
- How well does this model do at coming up with the correct structures based on object-feature data?

What forms are there?

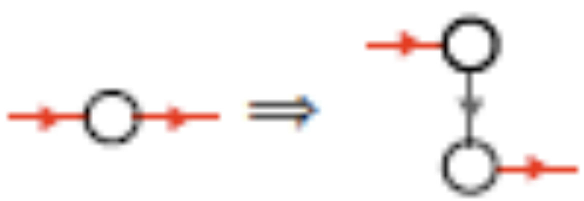
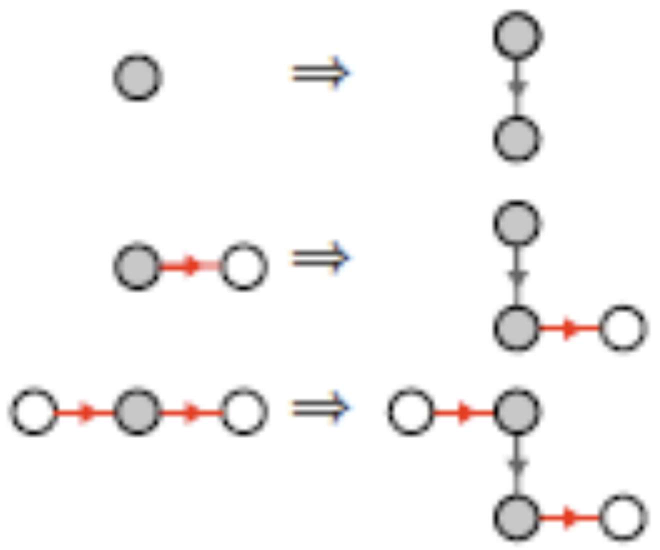


Form *F*	**# of possible forms with *k* nodes**
Partition	1
Directed chain	$k!$
Undirected chain	$k!/2$
Order	$k!$
Directed ring	$(k-1)!$
Undirected ring	$(k-1)!/2$
Directed hierarchy	k^{k-1}
Undirected hierarchy	k^{k-2}
Tree	$(2k-5)!!$

This follows from a generative model

It is a model for structures given specific forms

The idea is that each form defines a graph grammar which allows you to “grow” any specific structure of that form

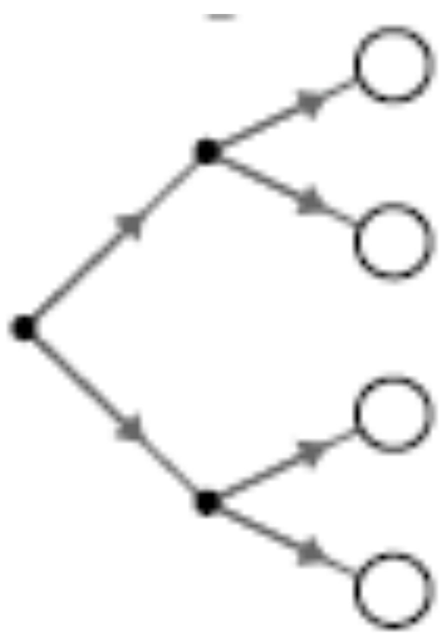
Chain	Graph grammar “rule”	Example derivations
		

This follows from a generative model

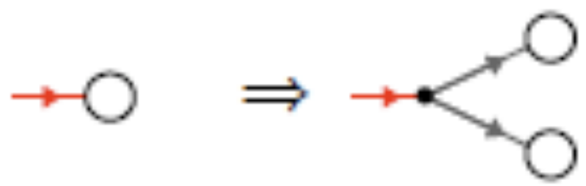
It is a model for structures given specific forms

The idea is that each form defines a graph grammar which allows you to “grow” any specific structure of that form

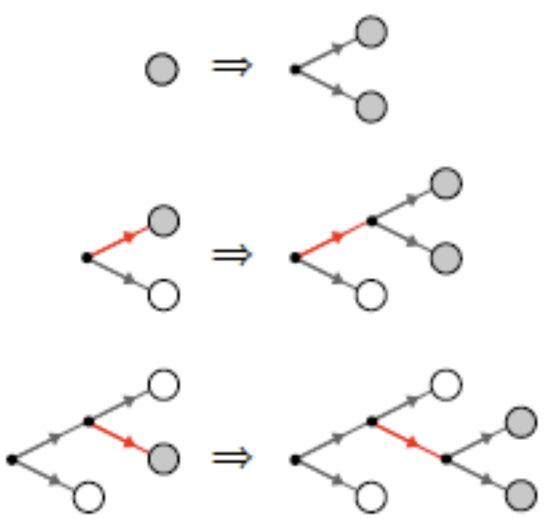
Tree



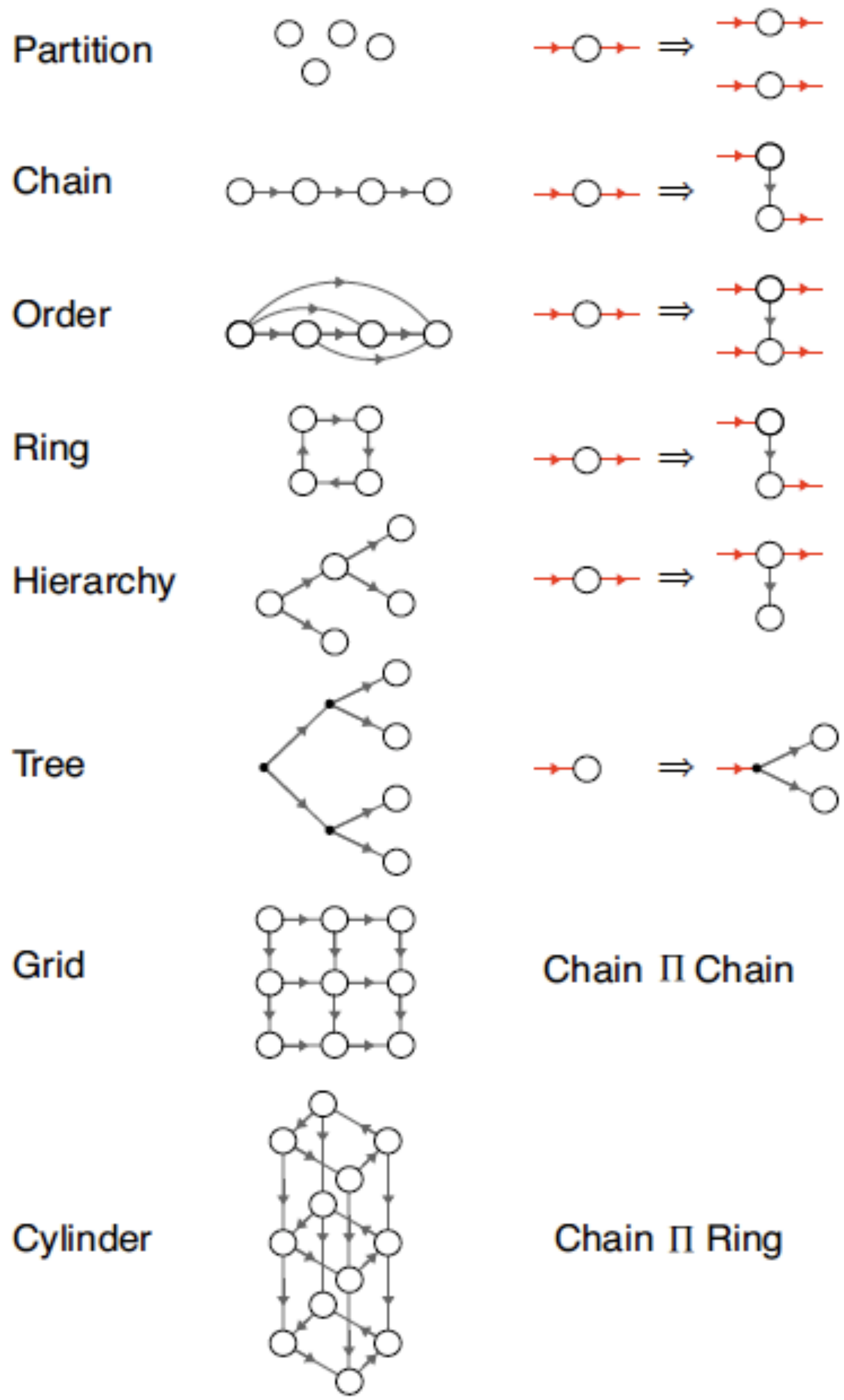
Graph grammar “rule”



Example derivations



Each form is defined by a graph grammar



This means that only are structures with fewer nodes favoured, but simpler forms are too!

This is for the same Bayesian Ockham's Razor reasons that we saw in the rectangle world: the more complex forms can fit more data, so if a simpler form will do, then we prefer that

So far, then...

$$p(S, F|D) \propto p(D|S)p(S|F)p(F)$$

Favours structures for which the data more closely "aligns"

Favours simpler structures and forms

Set to uniform

Questions

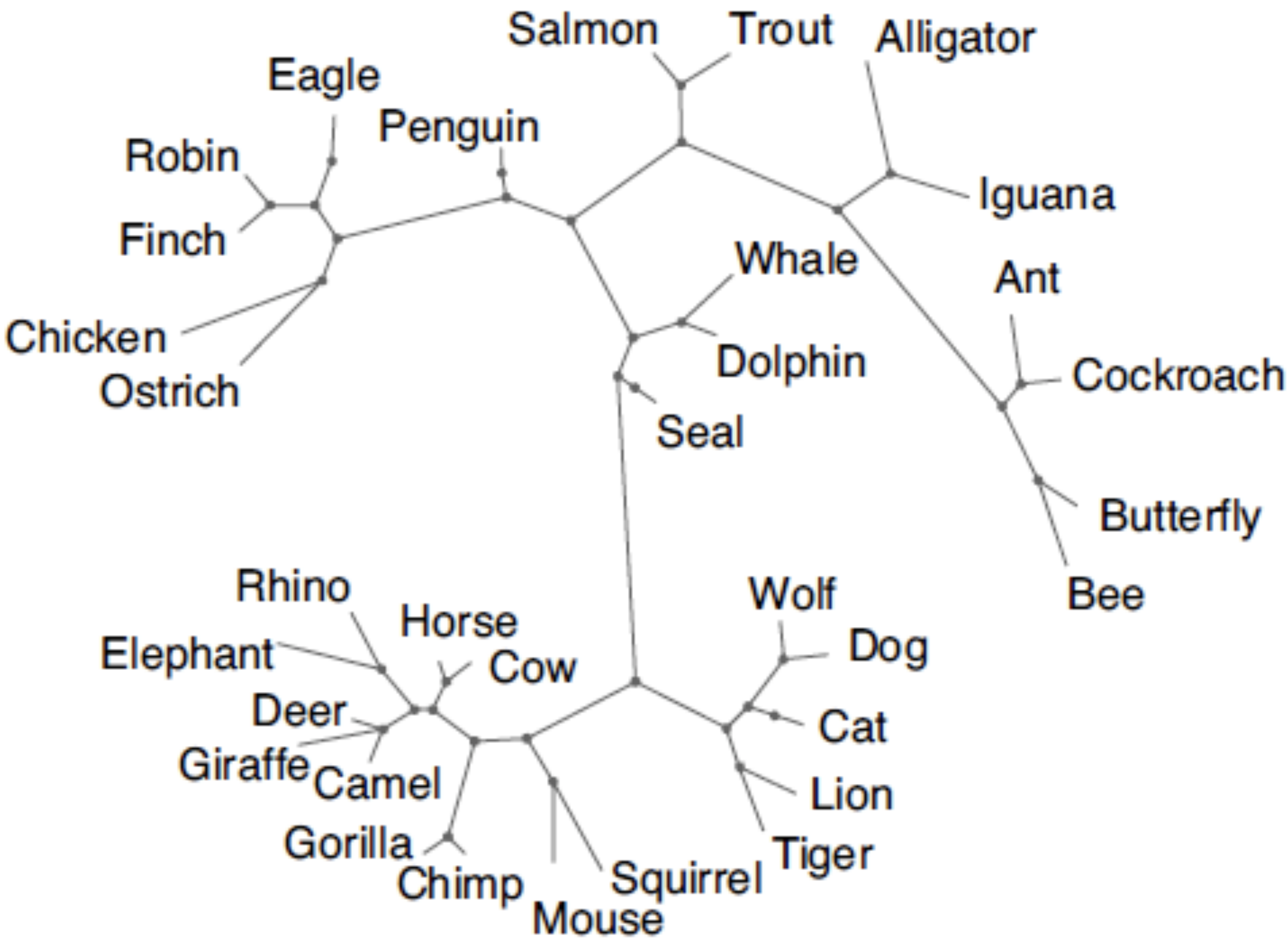
- How do you pick a structure that "fits" some data well? (in other words, how is data generated from a structure?)
- How do we prevent the model from simply picking the most complex structures possible? (in other words, what prior is placed on structures, to prefer simple ones?)
- Where do all these structures come from? (in other words, how is a "structure form" chosen?)

➡ How well does this model do at coming up with the correct structures based on object-feature data?

Dataset 1: Animals

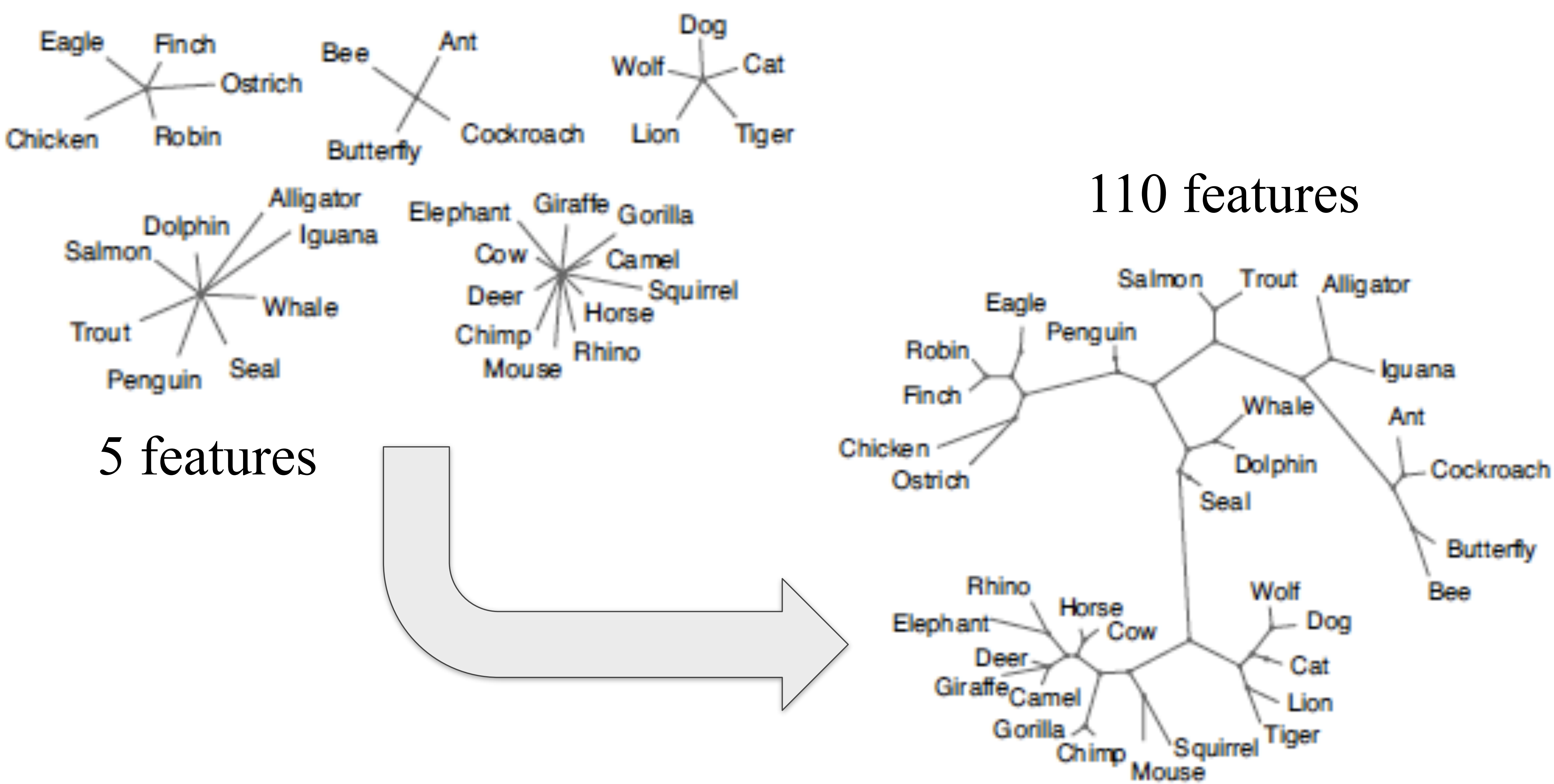
Object-feature lists generated by people

Dataset 1: Animals



Dataset 1: Animals

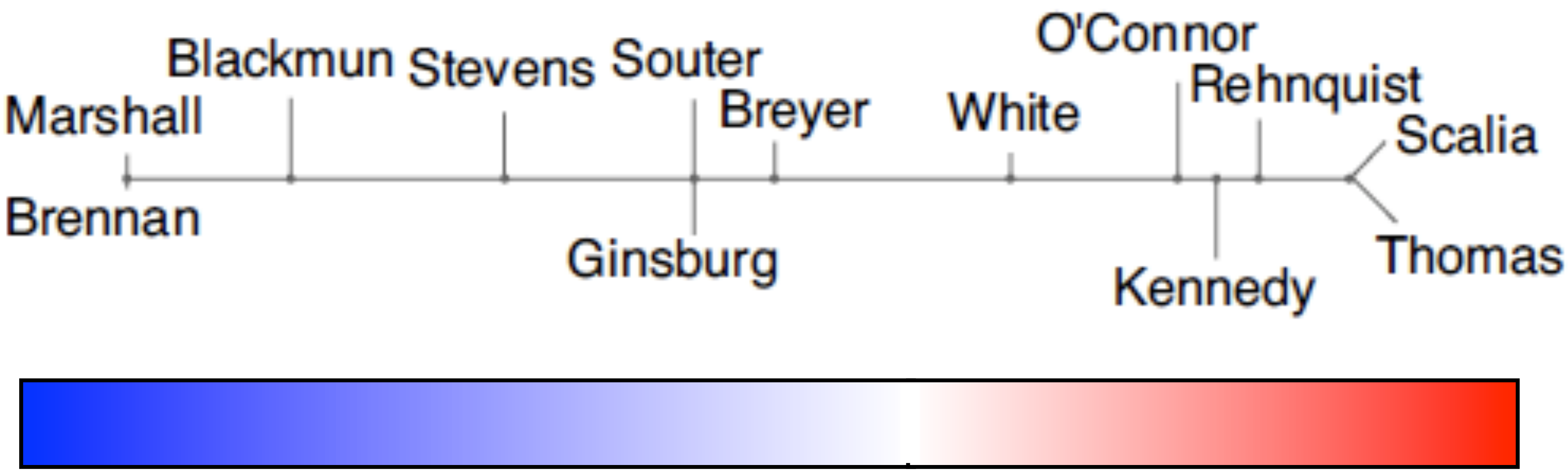
Simpler structures are preferred with less data



Dataset 2: Supreme court votes

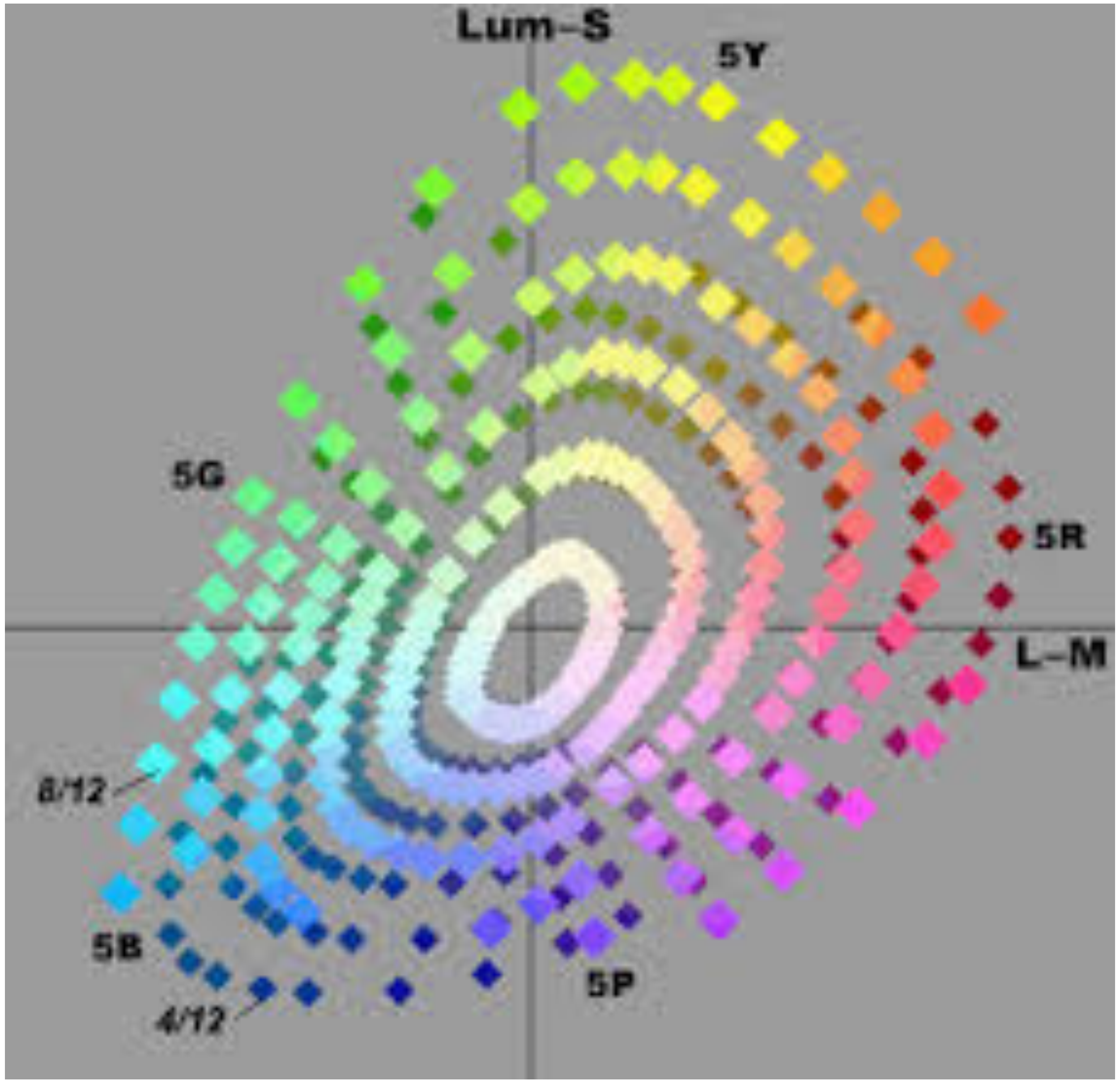
objects = cases,   features = votes

Dataset 2: Supreme court votes

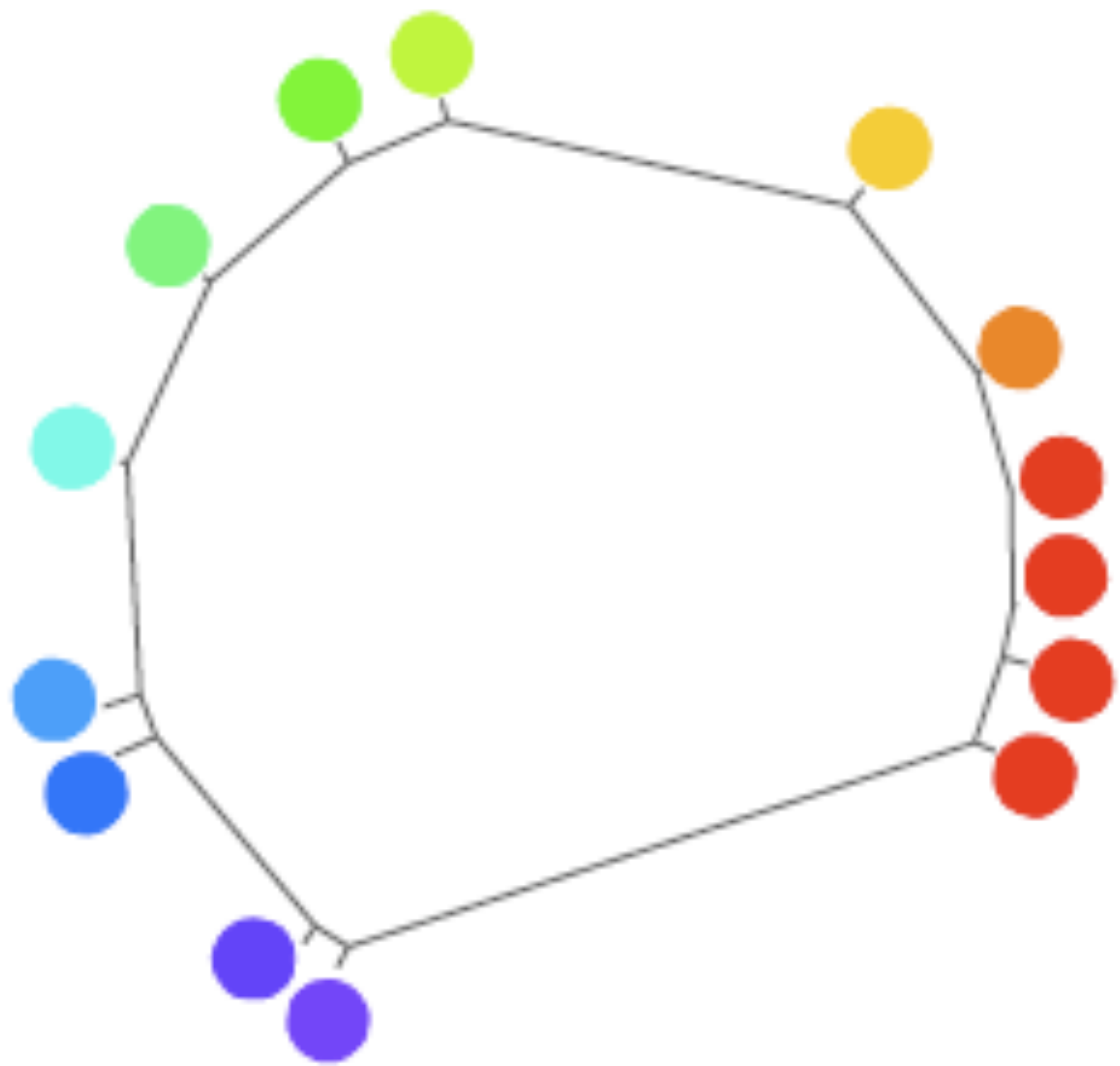


Dataset 3: Colours

similarity judgments based on wavelengths



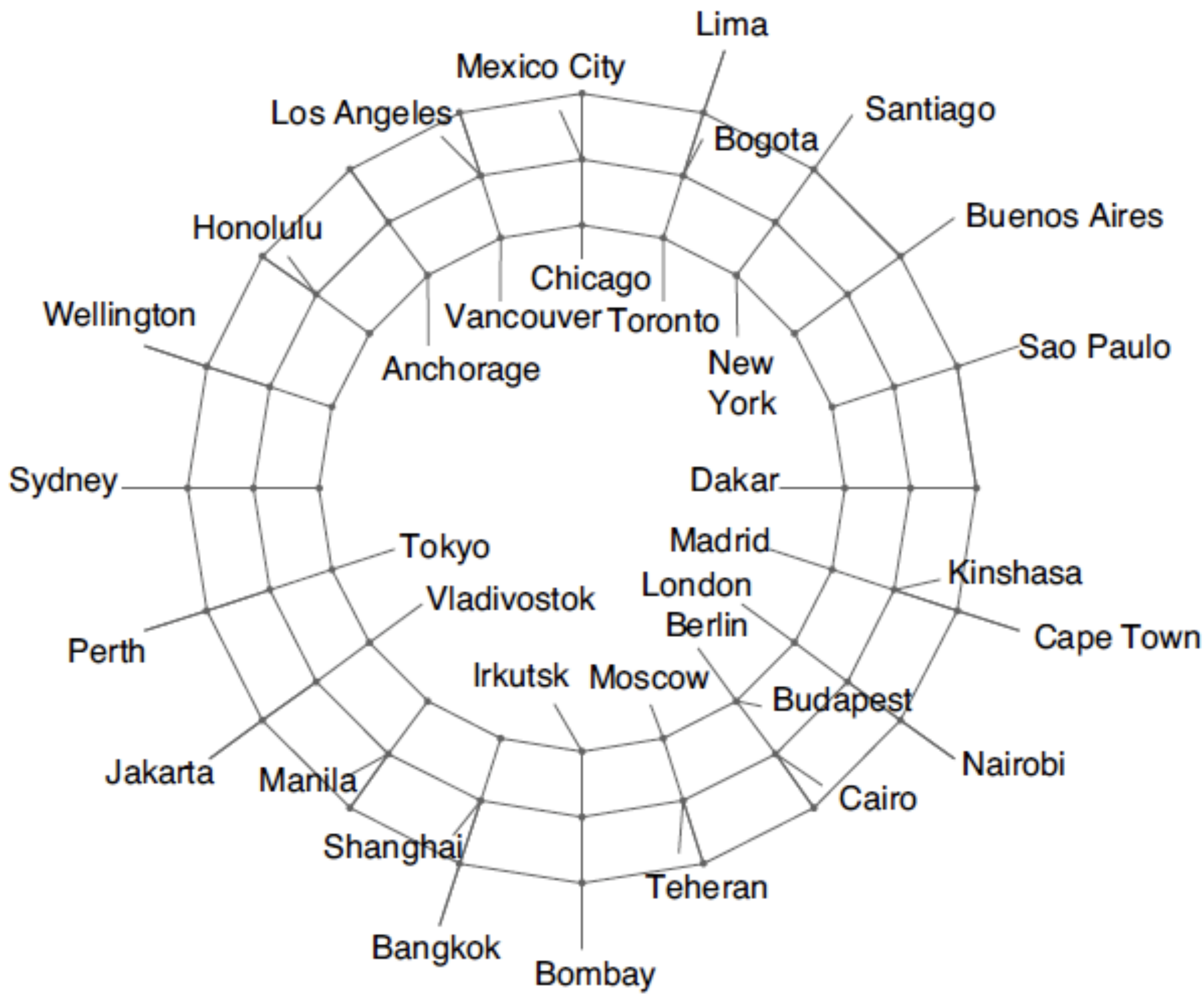
Dataset 3: Colours



Dataset 4: World cities

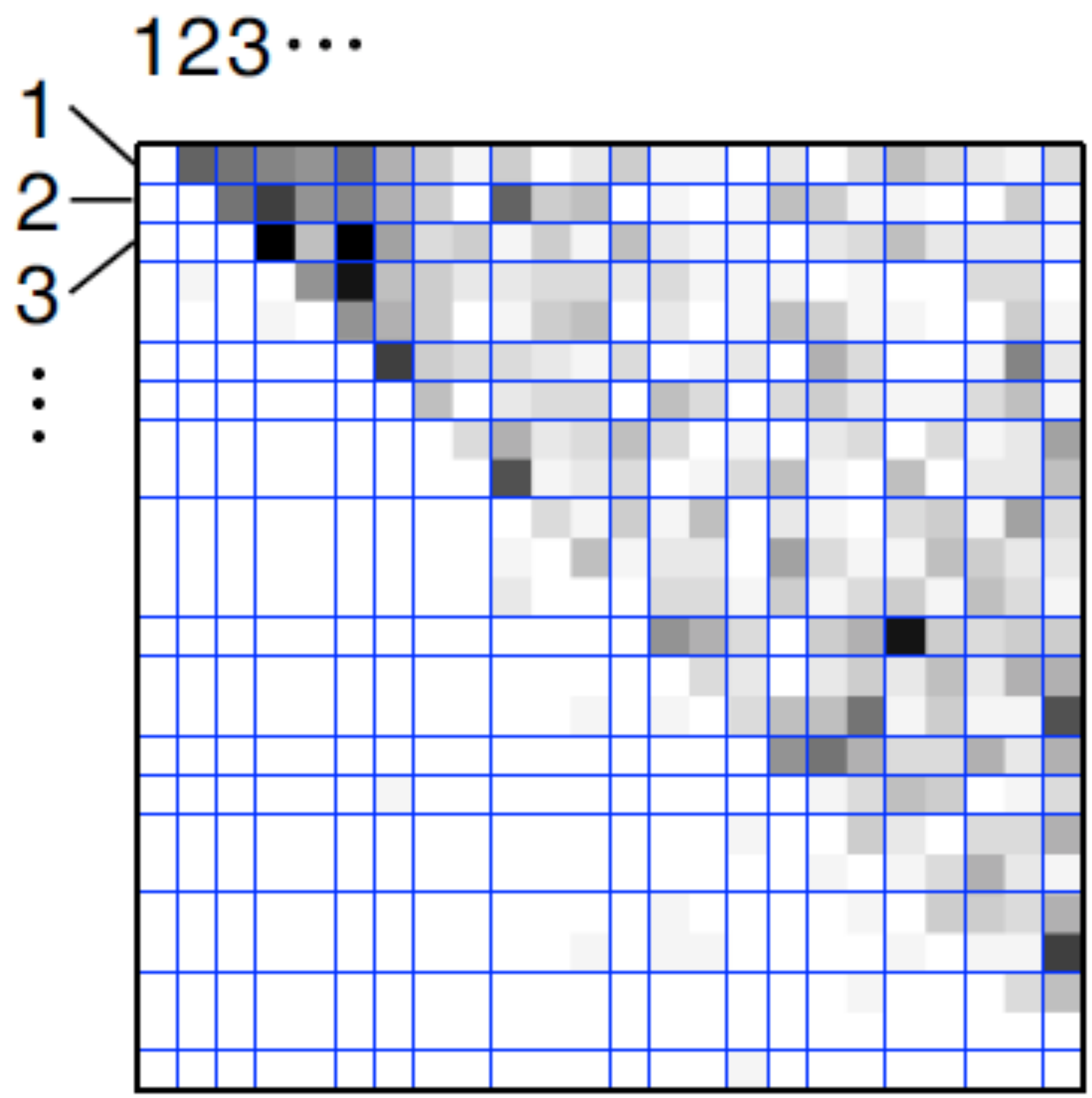
similarities derived from distances

Dataset 4: World cities



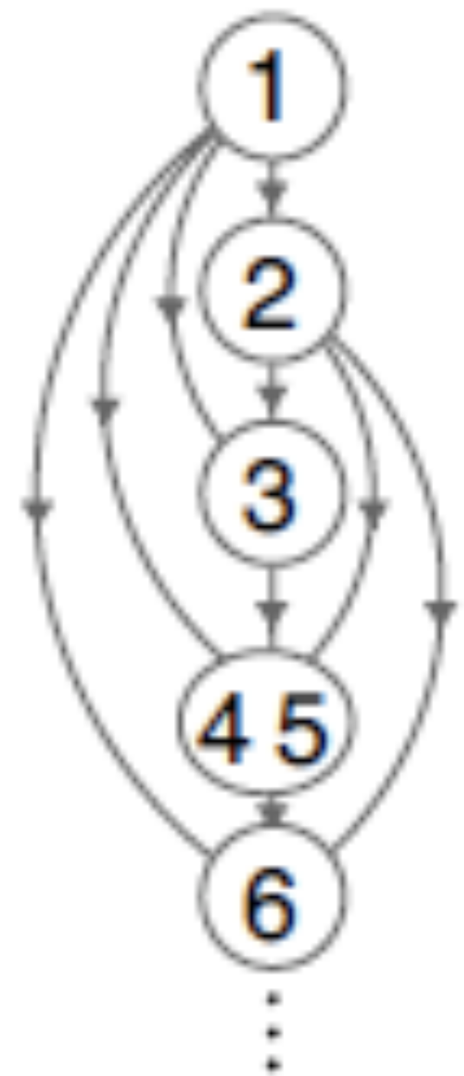
Dataset 5: Dominance hierarchies

Troop of sooty mangabees (object x object matrix, where objects are each individual, features = who hit who)



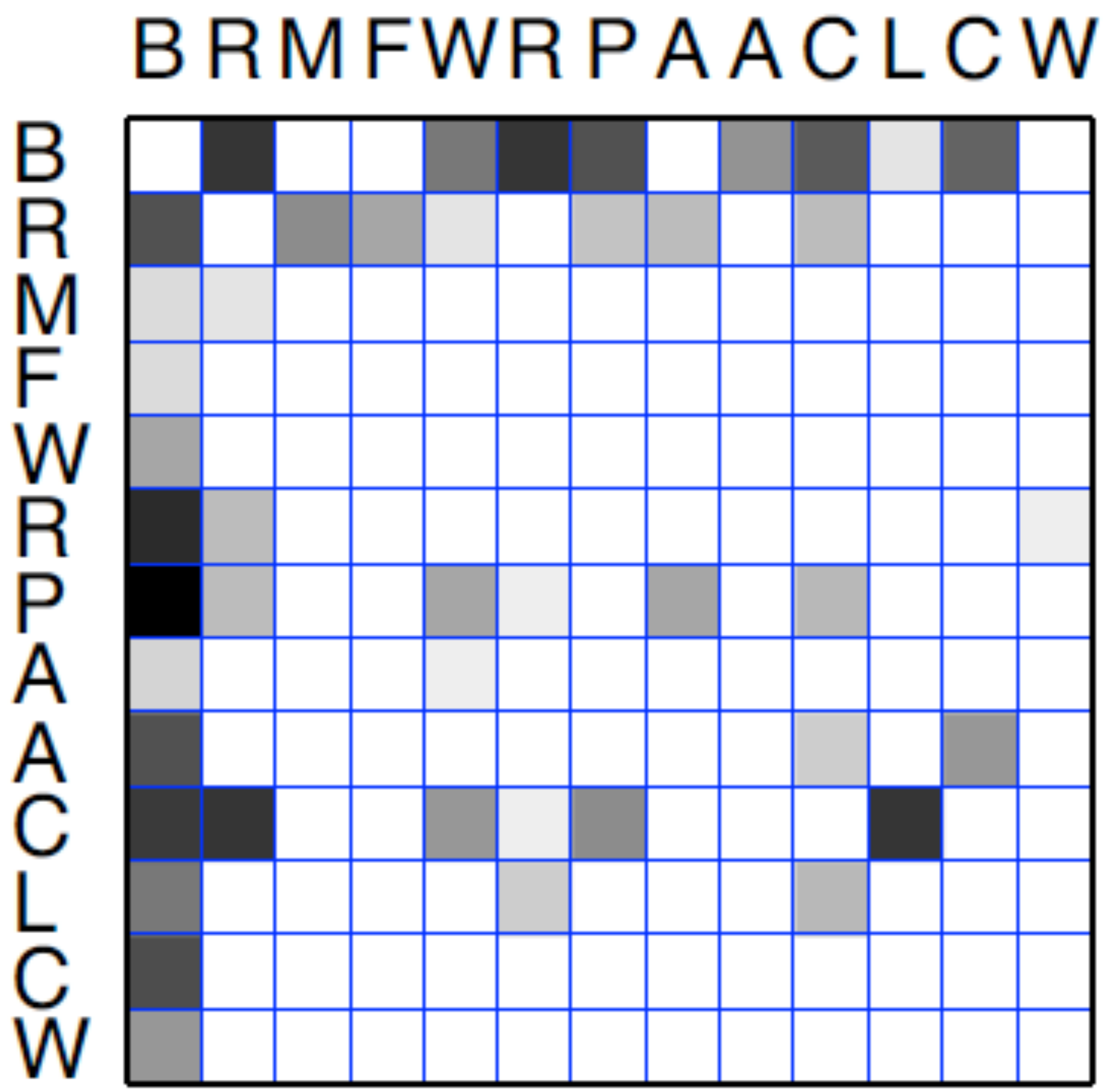
Dataset 5: Dominance hierarchies

Troop of sooty mangabees (object x object matrix, where objects are each individual, features = who hit who)

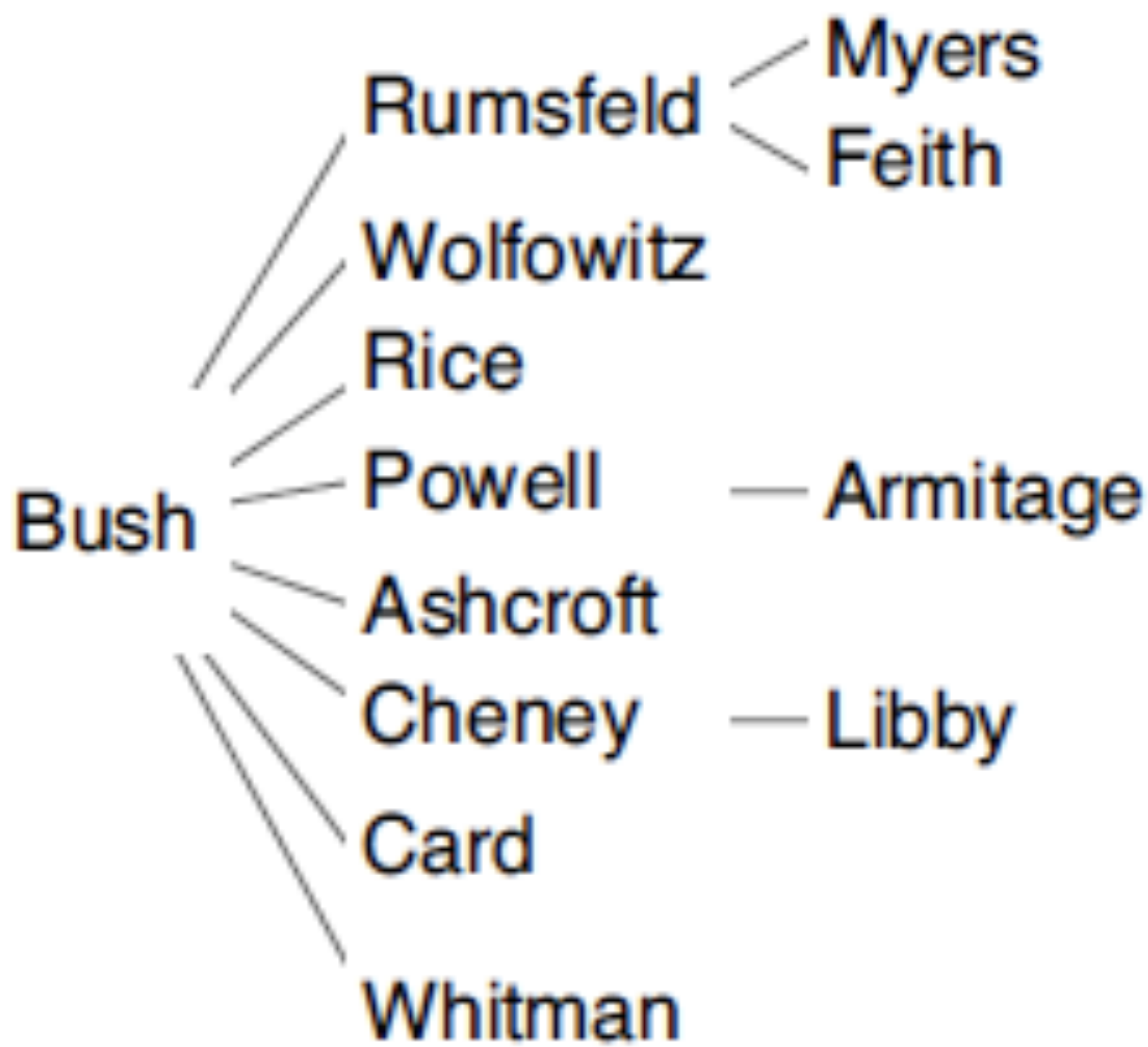


Dataset 6: Dominance hierarchies

Members of the Bush administration (features = interactions)

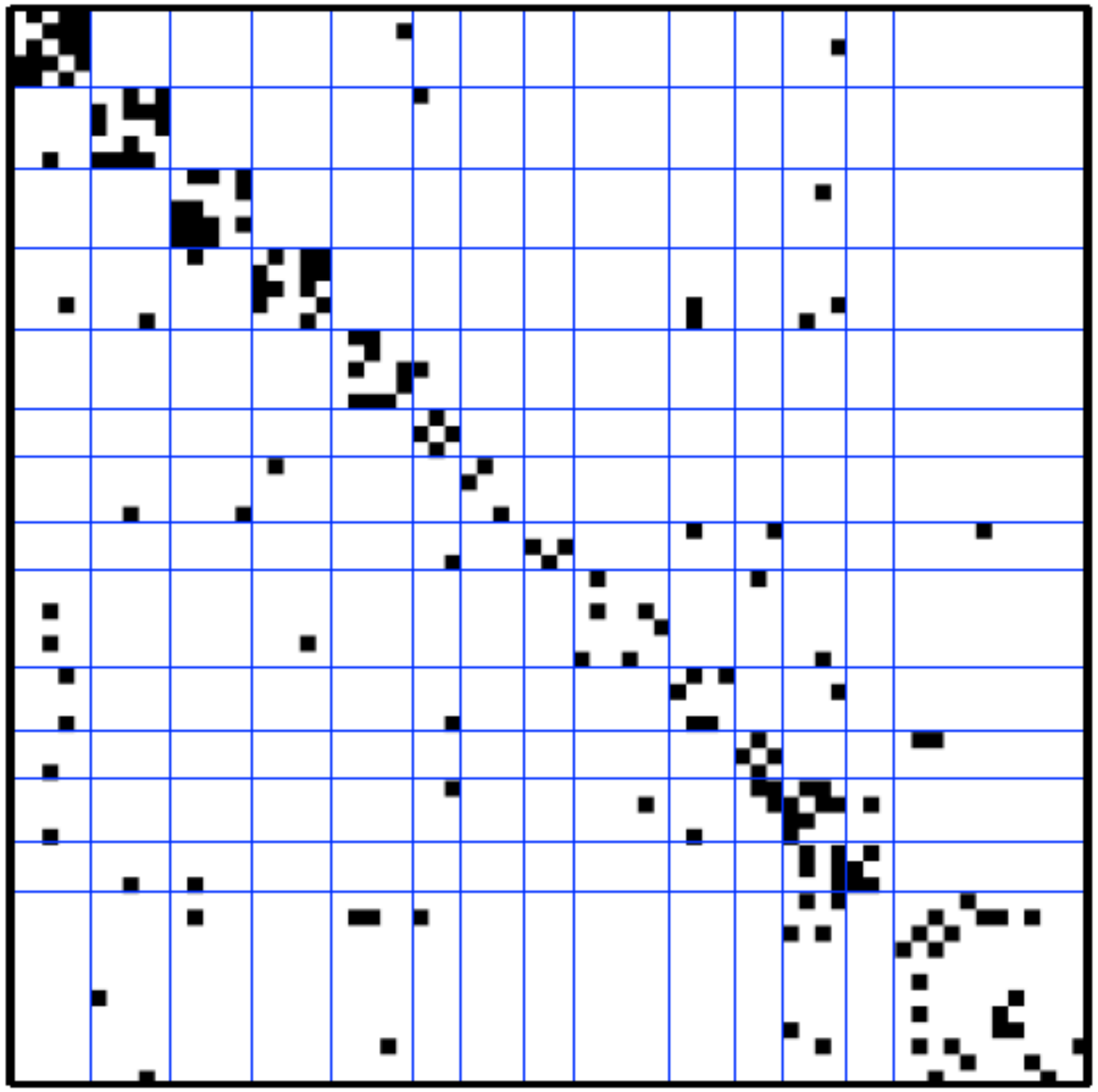


Dataset 6: Dominance hierarchies



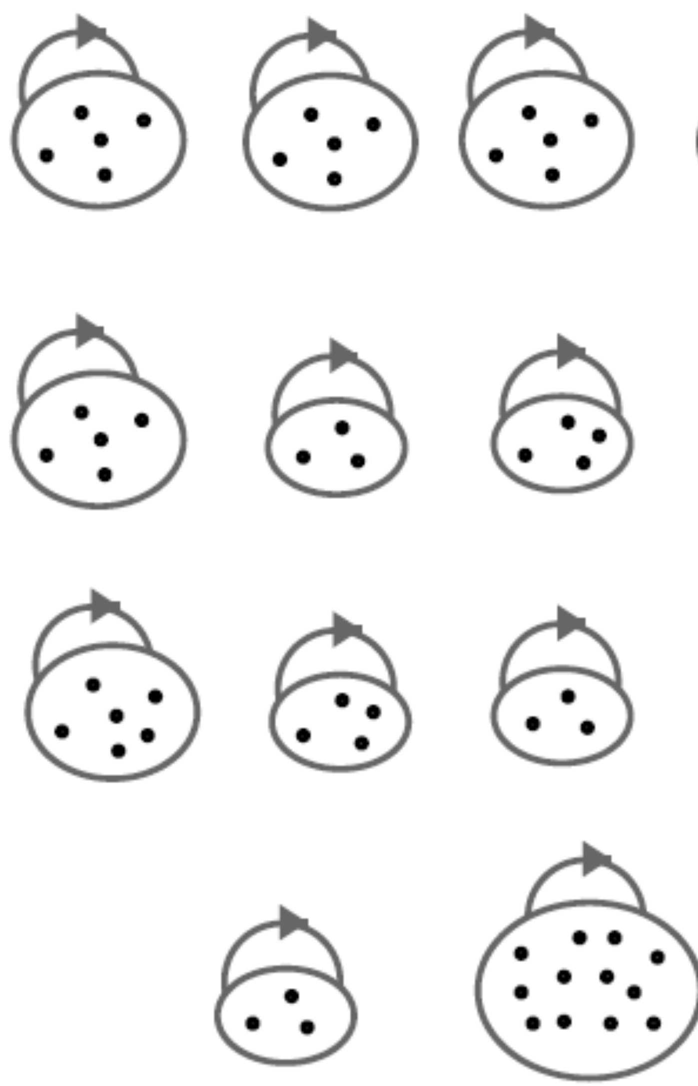
Dataset 7: Social structures

Cliques between prisoners (objects are prisoners, features are who they said they were friends with)

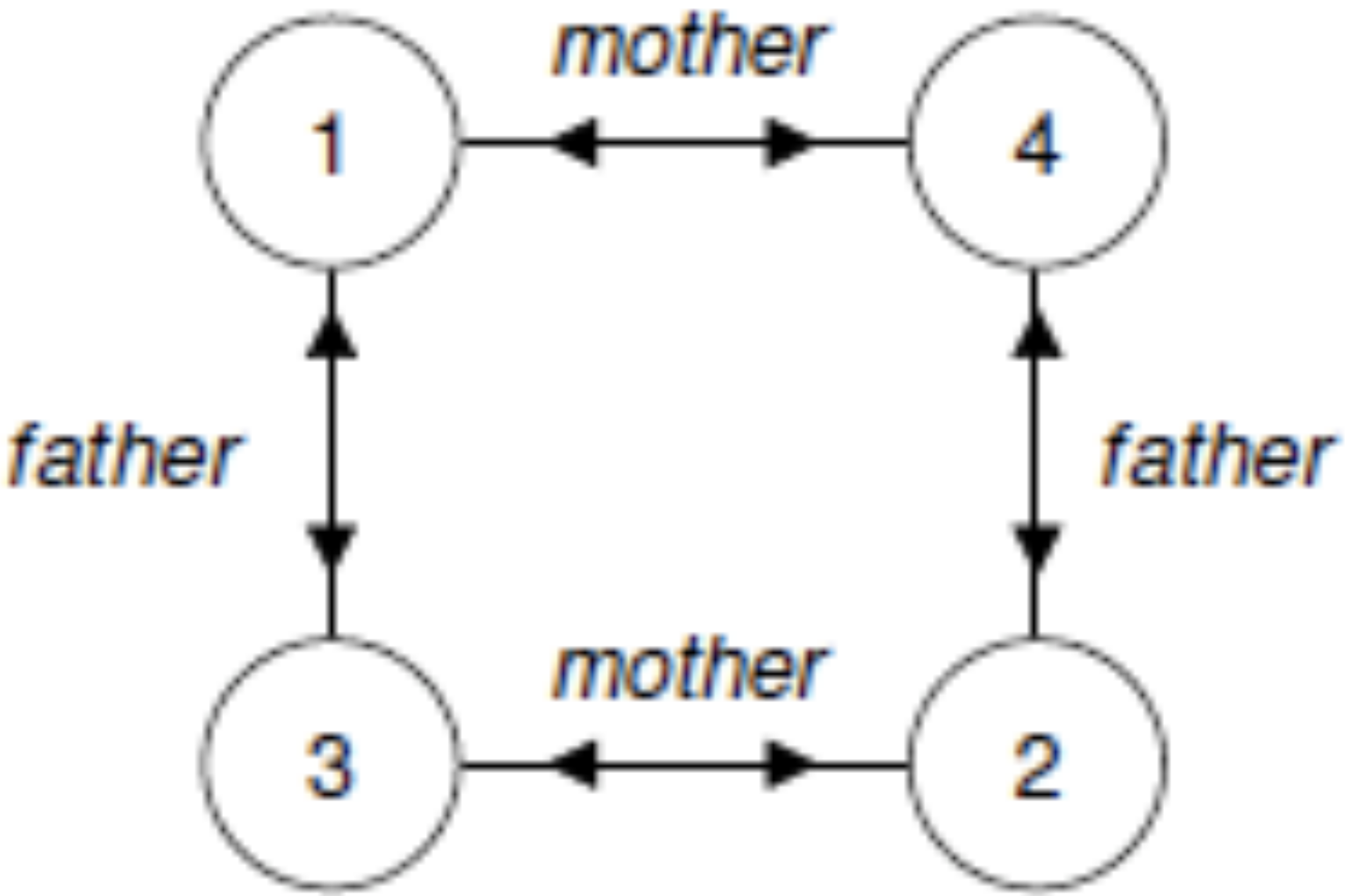


Dataset 7: Social structures

Cliques between prisoners (objects are prisoners, features are who they said they were friends with)

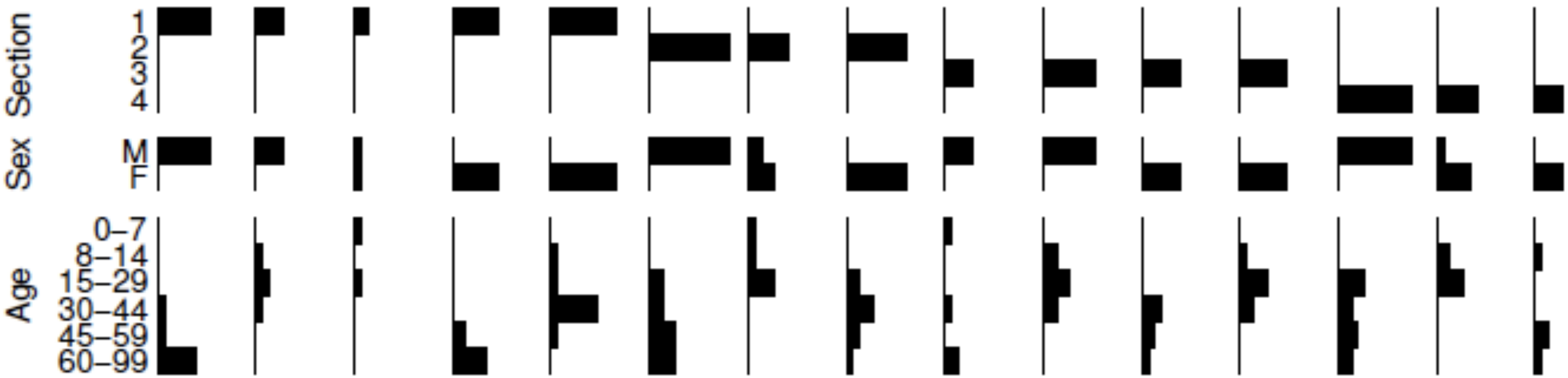


Dataset 8: Alyawarra kinship terms

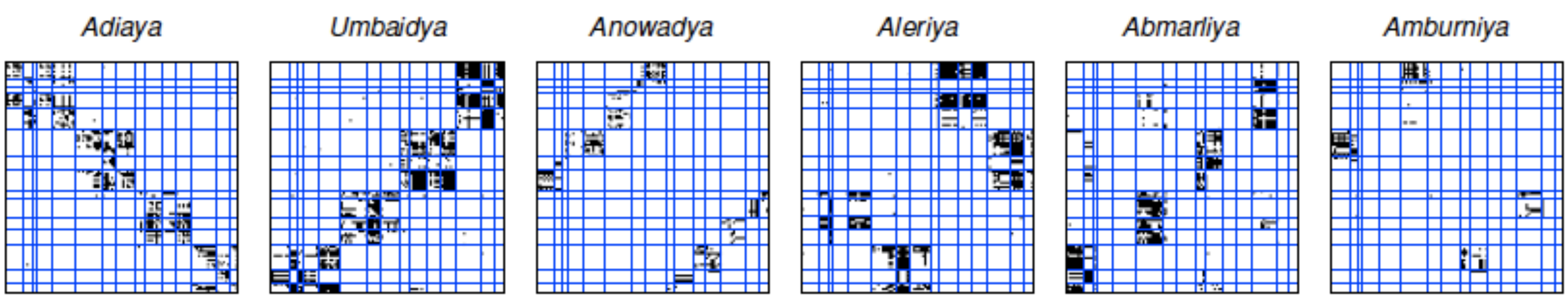


Dataset 8: Alyawarra kinship terms

15 different clusters (of the 104 individuals) found by the model

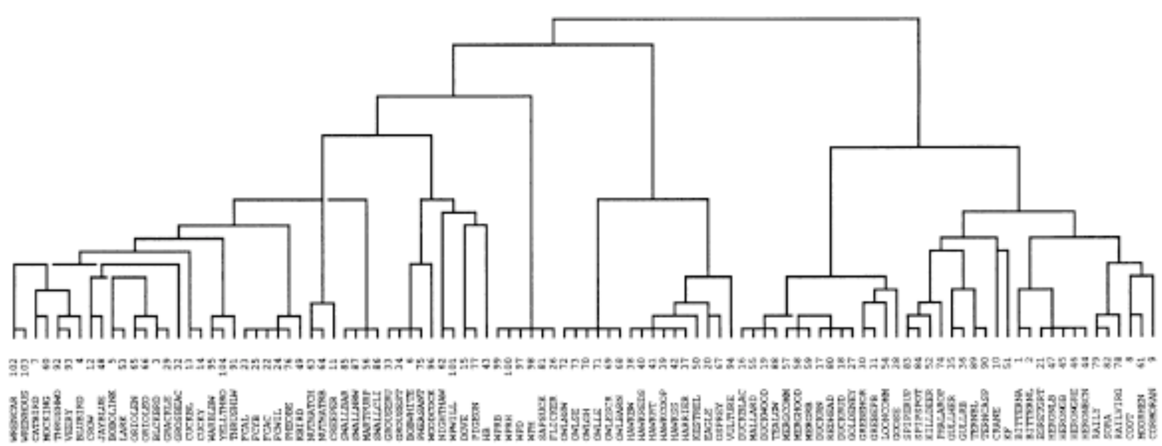


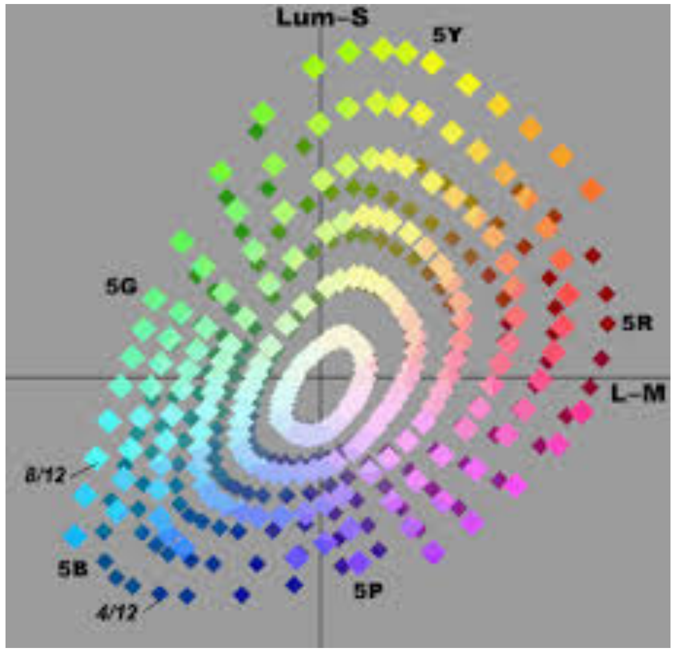
Some of the individual kinship terms

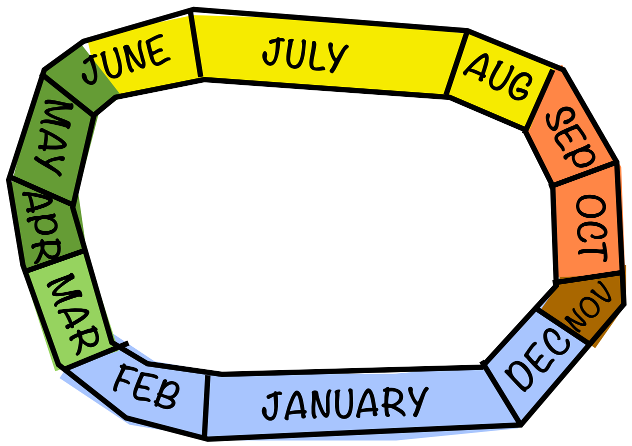


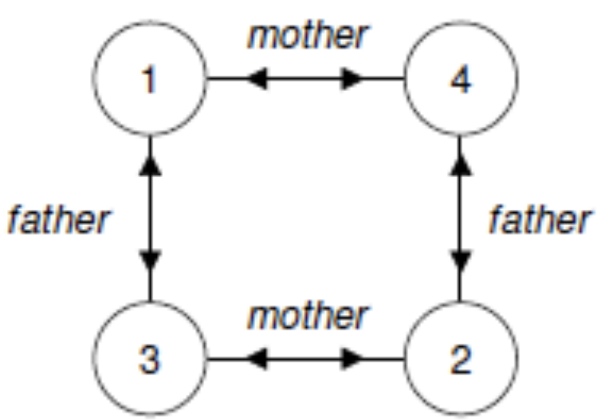
Summary

- There is a lot of evidence that people use and infer different structures in different domains



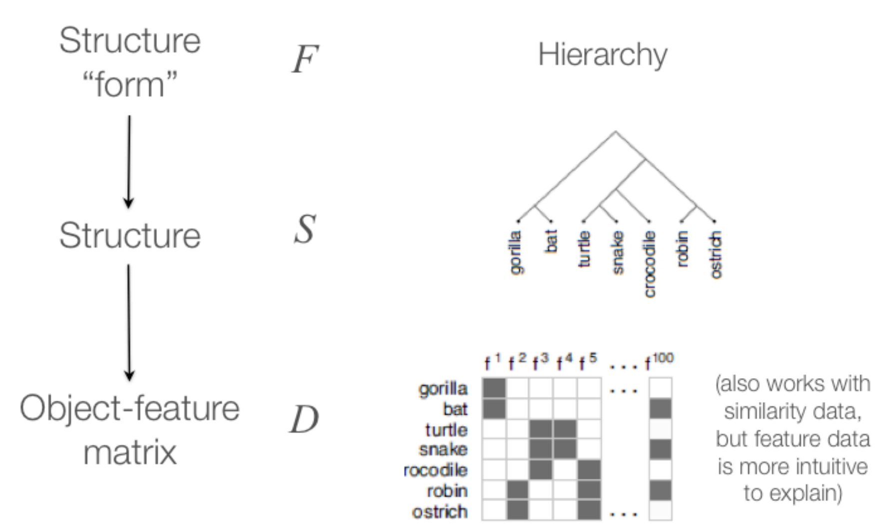






Summary

- There is a lot of evidence that people use and infer different structures in different domains
- Presented a model which can take raw data (object-feature or object-object matrix) and figure out which structure fits it best
  - Trades off between structures that fit the data well, and structures that are simpler (fewer nodes, simpler forms)



Summary

- There is a lot of evidence that people use and infer different structures in different domains
- Presented a model which can take raw data (object-feature or object-object matrix) and figure out which structure fits it best
  - Trades off between structures that fit the data well, and structures that are simpler (fewer nodes, simpler forms)
- This is another kind of *hierarchical* or *overhypothesis* learning, which people excel at

Summary

- There is a lot of evidence that people use and infer different structures in different domains
- Presented a model which can take raw data (object-feature or object-object matrix) and figure out which structure fits it best
  - Trades off between structures that fit the data well, and structures that are simpler (fewer nodes, simpler forms)
- This is another kind of *hierarchical* or *overhypothesis* learning, which people excel at
- Next lectures: Learning structure over time as well as space

Additional references (not required)

Human structure learning

- Bailenson, J., Shum, M., Atran, S., Medin, D., & Coley, J. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition* 84*: 1-53.*
- Medin, D.,Lynch, E., and Coley, J. (1997). Categorisation and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology 32*: 49-96

Models of structure learning

- Kemp, C. & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science 336*(6084):1049-1054
- Kemp, C., & Tenenbaum, B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences 105*(31): 10687-10692
- Kemp, C., Tenenbaum, B., Griffiths, T., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. *Proceedings of the 21st National Conference on Artificial Intelligence*