Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2004-050
AIM-2004-019

July 22, 2004

Discovering Latent Classes in Relational Data

Charles Kemp, Thomas L. Griffiths, and Joshua B. Tenenbaum

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

We present a framework for learning abstract relational knowledge with the aim of explaining how people acquire intuitive theories of physical, biological, or social systems. Our approach is based on a generative relational model with latent classes, and simultaneously determines the kinds of entities that exist in a domain, the number of these latent classes, and the relations between classes that are possible or likely. This model goes beyond previous psychological models of category learning, which consider attributes associated with individual categories but not relationships between categories. We apply this domain-general framework to two specific problems: learning the structure of kinship systems and learning causal theories.

1 Introduction

Imagine a hotel employee serving drinks at the general convention of the Episcopal church. All the delegates are in casual clothes, and at first he finds it difficult to identify the people who hold the most influential positions within the church. Eventually he notices that a group of delegates is treated with deference by everyone — call them the archbishops. Another group is treated with deference by everyone except the archbishops — call them the bishops. Just by observing the delegates mingle, he might be able to guess the seniority of the office each person holds.

Imagine now a child who stumbles across a set of identical-looking metal bars: magnets, although she does not know it. As she plays with the bars she notices that some some surfaces attract each other and some repel each other. She should realize that there are two types of surfaces — call them North poles and South poles. Each surface repels others of the same type and attracts surfaces of the opposite type.

As these examples show, learning to reason about social and physical systems can require the discovery of latent structure in relational data. Both examples highlight latent classes (bishops and archbishops, North and South poles) which influence the relations (deference, attraction) that hold among a set of objects (guests, metal surfaces). These latent classes provide the building blocks of our intuitive domain theories, allowing us to understand and predict the interactions between novel objects in each domain.

Throughout this paper we distinguish between relations and attributes. Attributes are properties of individual objects, but relations specify how pairs of objects interact. In many cases latent classes are associated with predictable patterns of attributes. For instance, archbishops might tend to look a little older than bishops, and some sets of magnets have the North poles painted red and the South poles painted black. Real-world learning problems come in at least three forms — problems involving attributes alone, relations alone, or some combination of the two. Psychologists have paid most attention to the first problem, and have developed successful computational models [1, 2] that explain how people discover classes given attribute data. Our approach handles all three problems, although it reduces to Anderson's rational model of categorization [1] when given only attribute data. In real-world settings it is perhaps most usual for both types of data to be available. We focus here on the purely relational case, since this provides the greatest contrast with previous work on category learning.

This paper gives a rational account [1] of the discovery of latent classes in relational data. We define a generative model in which a particular relation holds between a pair of objects with some probability that depends only on the classes of those objects. Statisticians and sociologists have used a model of this kind, called the *stochastic blockmodel* [3, 4], to analyze social networks. The stochastic blockmodel, however, assumes a fixed, finite number of classes. When discovering the latent structure of a domain, people learn the number of classes at the same time as they learn the class assignments. Our model, the *infinite blockmodel*, allows an unbounded number of classes. We provide an algorithm that simultaneously discovers the number of classes and the class assignments, and use this model to explain human inference in two settings: learning kinship systems and learning causal theories.

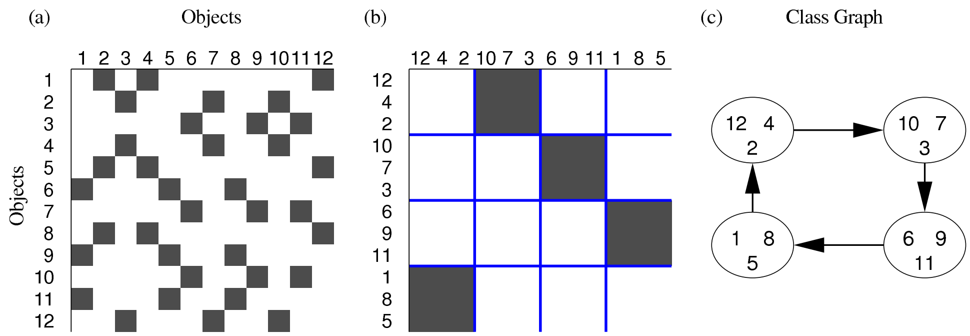


Figure 1: (a) An adjacency matrix representing a relation over a domain with twelve objects (b) The adjacency matrix permuted according to z, a vector of class assignments. The blue lines separate the four classes. (c) A class graph showing relations between the four latent classes. Each class sends links to one class and receives links from another class.

2 A generative model for relational data

Suppose we are interested in a system with a single directed relation R (multiple relations will be considered later). The relation can be represented as a graph G with labeled edges, where edge g_{ij} between objects i and j has value 1 if R holds between i and j and value 0 otherwise. Our goal is to identify the latent classes z of the objects using the information contained in G. For example, suppose that G is the graph represented as an adjacency matrix in Figure 1(a). Our goal is to find a partition of the 12 objects into classes. The best partition is represented in Figure 1(c), and Figure 1(b) shows the matrix G sorted according to this four-class solution.

To find the best assignment of objects to classes, we define a process by which z and G are generated and use Bayesian inference to infer z for an observed graph G. We specify this generative model in two stages, first showing how G is generated given z, and then describing the process by which z is generated.

2.1 Generating relations from classes

Assume that each potential relation between two objects is generated independently, and $p(g_{ij} = 1)$, the probability that the relation holds between i and j, depends only on z_i and z_j, the classes of i and j. Given a set of assignments z, the probability of G is

$$p(G|z, \eta) = \prod_{a,b} (\eta_{ab})^{m_{ab}} (1 - \eta_{ab})^{\bar{m}_{ab}} \tag{1}$$

where a and b range over all classes, η_{ab} is the probability that the relation holds between a member of class a and a member of class b, m_{ab} is the number of edges

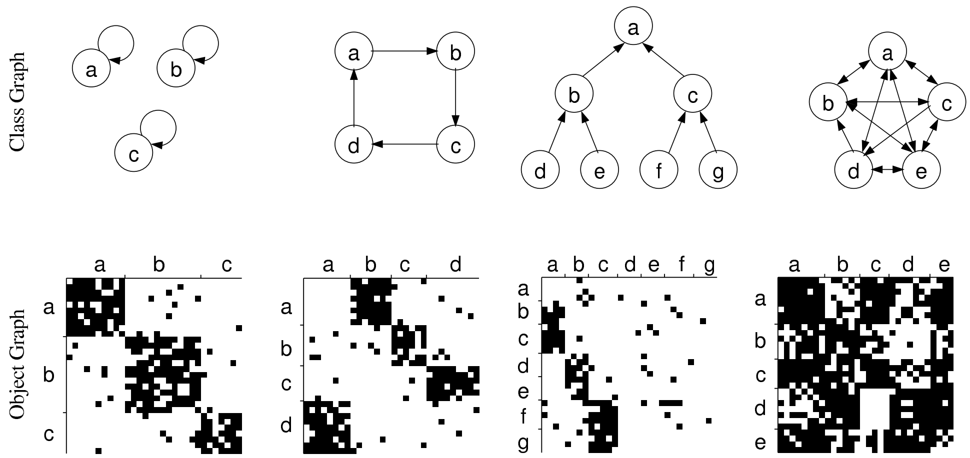


Figure 2: Bottom row: adjacency matrices representing relations in four domains, each with thirty objects. The objects are sorted into classes (indicated by labels a through g). Top row: class graphs corresponding to the object graphs below. Only the edges in the class graph with large weights are shown. The leftmost class graph, for example, indicates a case where objects in one class tend to link only to other objects in the same class.

between a and b with label equal to 1, and $\bar{m}_{ab}$ is the number of edges between a and b with label equal to zero.

The matrix η specifies a *class graph*: a graph over the classes where the edge between class a and class b has weight η_{ab}. Figure 2 shows several examples of class graphs η and corresponding object graphs G. Despite the simplicity of the model it can express a rich class of relational structures. Figure 2 shows just four examples: a graph with community structure, a ring, a hierarchy and a fully connected graph.

Multiple relations can be handled by assuming that each relation is conditionally independent of the others given class assignments z. Attribute information can be incorporated similarly if we assume that each attribute or relation is conditionally independent of all other attributes and relations given a set of class assignments.

2.2 Generating classes

Statisticians and sociologists have defined a relational model which uses Equation 1 and assumes that the z_i are drawn from a fixed multinomial distribution over a finite number of classes [4]. This model (the *stochastic blockmodel*) does not capture one of the most important aspects of human learning: the discovery that a domain includes a certain number of latent classes.

We use an alternative method for generating the z_i which does not require the number of classes to be specified in advance. The key intuition is that the the number of classes should be allowed to grow as more objects are added to the system. Given one object, we need only a single class. As each object is added, we randomly decide whether that object belongs to the same class as some object we have seen before, or whether it represents a new class. A Chinese restaurant process (CRP)[1] requires that the probability that a new object belongs to an existing class is directly proportional to the number of objects of that class seen previously. Under the CRP, the probability distribution over classes for the ith object, conditioned on the classes of the previous objects $1, \ldots, i-1$ is

$$p(z_i = a | z_1, \ldots, z_{i-1}) = \begin{cases} \frac{n_a}{i-1+\alpha} & n_a > 0 \\ \frac{\alpha}{i-1+\alpha} & a \text{ is a new class} \end{cases} \tag{2}$$

where n_a is the number of objects already assigned to class a, and α is a parameter of the distribution.

The CRP prior on z can generate partitions with as many classes as objects, and will potentially create a countably infinite set of classes given a countably infinite set of objects. We therefore call the model in which z is generated according to Equation 2 and G is generated according to Equation 1 the *infinite blockmodel*. Other infinite models have been proposed by machine learning researchers [5, 6] using a similar construction.

2.3 Model inference

Having defined a generative model for G and z, we use Bayesian inference to compute a posterior distribution over z given G:

$$p(z|G) \propto p(G|z)p(z) \tag{3}$$

where $p(G|z)$ can be derived from Equation 1, and $p(z)$ follows from Equation 2. For the finite stochastic blockmodel, Snijders and Nowicki [4] describe a Gibbs sampler in which η and the distribution over classes are explicitly represented. We will define a Gibbs sampler for the infinite blockmodel, integrating out η and using the CRP to sample z.

Gibbs sampling is a form of Markov chain Monte Carlo, a standard statistical tool for Bayesian inference with otherwise intractable distributions [7]. A Gibbs sampler is a Markov chain in which the state corresponds to the variables of interest, in our case z, and transitions result from drawing each variable from its distribution when conditioned on all other variables, in our case the conditional probability of z_i given all other assignments z_{-i}, $p(z_i|z_{-i})$. It follows from Equation 3 that

$$p(z_i|z_{-i}, G) \propto p(G|z)p(z_i|z_{-i}). \tag{4}$$

To compute the first term on the right hand side, we integrate out the parameters η in Equation 1 using a symmetric Beta prior over every η_{ab}:

$$p(G|z) = \prod_{a,b} \frac{\text{Beta}(m_{ab} + \beta, \bar{m}_{ab} + \beta)}{\text{Beta}(\beta, \beta)} \tag{5}$$

[1]The process was named out of respect for the apparently infinite capacity of Chinese restaurants. The same process is also known less colorfully as a Polya urn scheme.

where β is a hyperparameter. The second term follows from the fact that the CRP is exchangeable, meaning that the indices of the z_i can be permuted without affecting the probability of z. As a consequence we can treat z_i as the last object to be drawn from the CRP. The resulting conditional distribution follows directly from Equation 2.

To facilitate mixing, we supplement our Gibbs sampler with two Metropolis-Hastings updates. First, we consider proposals that attempt to split a class into two or merge two existing classes [8]. Split-merge proposals allow sudden large-scale changes to the current state rather than the incremental changes characteristic of Gibbs sampling. Second, we run a Metropolis-coupled Markov chain Monte Carlo simulation: we run several Markov chains at different temperatures and regularly consider swaps between the chains. If the coldest chain becomes trapped in a mode of the posterior distribution, the chains at higher temperatures are free to wander the state space and find other regions of high probability if they exist. To avoid free parameters, we sample the hyperparameters α and β using a Gaussian proposal distribution and an (improper) uniform prior over each.

Even though η is integrated out, it is simple to recover the class graph given z. The maximum likelihood value of η_{ab} given z is $\frac{m_{ab}+\beta}{\bar{m}_{ab}+m_{ab}+2\beta}$. Predictions about missing edges are also simple to compute. The probability that an unobserved edge between objects i and j has value 1 is $p(g_{ij} = 1) = \frac{m_{z_i z_j}+\beta}{\bar{m}_{z_i z_j}+m_{z_i z_j}+2\beta}$. If some edges in graph G are missing at random, we can ignore them and maintain counts m_{ab} and $\bar{m}_{ab}$ over only the observed part of the graph.

We do not claim that the MCMC simulations used to fit our model are representative of cognitive processing. The infinite block model addresses the question of what people know about relational systems and our simulations will show that this knowledge can be acquired from data, but we do not consider the process by which this knowledge is acquired.

3 Relational and attribute models on artificial data

We ran the infinite blockmodel on the structures shown in Figure 2, which are simple versions of some the relational structures found in the real world. Our algorithm solves each of these cases perfectly, finding the correct number of classes and the correct assignment of objects to classes.

To further explore our model's ability to recover the true number of classes we gave it graphs based on randomly-generated η matrices of different dimensions. The entries in each matrix were drawn independently from a symmetric Beta prior with hyperparameter β. When β is small, the average connectivity between blocks is usually very high or very low. As β increases, the blocks of objects are no longer so cleanly distinguished. Figure 3 shows that the model makes almost no mistakes when the β is small but recovers the true number of classes less often as β increases.

For comparison, we also evaluated the performance of a model defined on attributes rather than relations. The analogous model for attributes uses a feature matrix F where f_{ik} is 1 if object i possesses attribute k and 0 otherwise. Assuming that attributes are generated independently and that $p(f_{ik} = 1)$, the probability that object i has attribute

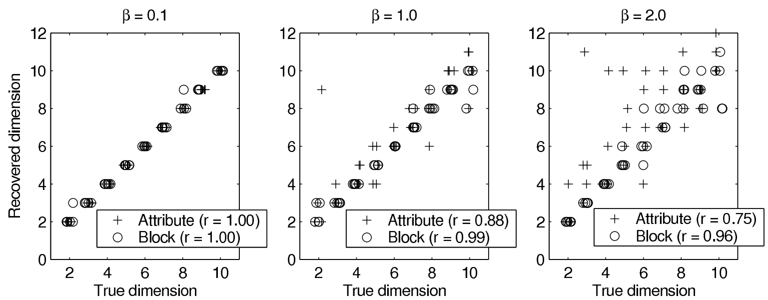


Figure 3: Success at recovering the true number of latent classes from artificially generated data. The infinite blockmodel performs better than the infinite mixture model as the hyperparameter β increases.

k, depends only on z_i, we have

$$p(F|z,\theta) = \prod_{a,k} \theta_{ak}^{n_{ak}} (1-\theta_{ak})^{\bar{n}_{ak}}$$

where the product is over all classes a and features k, n_{ak} is the number of objects i for which $z_i = a$ and $f_{ik} = 1$, $\bar{n}_{ak}$ is the number of objects i for which $z_i = a$ and $f_{ik} = 0$, and θ_{ak} is the probability that feature k takes value 1 for class a. Using a CRP prior on z, we can apply Gibbs sampling as in Equation 4 to infer $P(z|F)$, except we now use

$$p(F|z) = \prod_{a,k} \frac{\text{Beta}(n_{ak}+\beta, \bar{n}_{ak}+\beta)}{\text{Beta}(\beta,\beta)}$$

in place of Equation 5. This model is an infinite mixture model [6] and is equivalent to Anderson's rational model of categorization [1].

The infinite mixture model can be applied to relational data if we convert the relational graph G into an attribute matrix F. We flattened each n by n adjacency matrix into an attribute matrix with $k = 2n$ features, one for each row and column of the matrix. For example, a matrix for the social relation "defers to" is flattened into an attribute matrix with two features corresponding to each person P: "defers to P" and "is deferred to by P". This model does well when β is small, but its performance falls off more sharply than that of the blockmodel as β increases.

4 Kinship Systems

Australian tribes are renowned among anthropologists for the complex relational structure of their kinship systems. In a mathematical appendix to a work by Levi-Strauss, André Weil notes that several Australian kinship systems are isomorphic to the dihedral group of order eight [9]. Even trained field workers find these systems difficult to

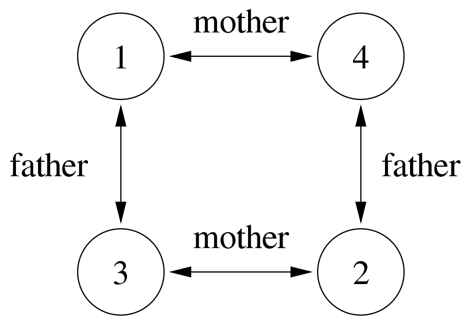


Figure 4: Relations between the four sections of a Kariera kinship system. Members of section 4 have mothers in section 1 and fathers in section 2. This simple model is consistent with the relational data shown in Figure 5.

understand [10] which raises an intriguing question about cognitive development: how do children discover the social structure of their tribe? The learning problem is particularly interesting since some communities appear to have no explicit representations of kinship rules, let alone cultural transmission of such rules. [2] We focus here on the Alyawarra, a Central Australian tribe studied extensively by Denham [11, 12, 13]. Using Denham's data[3], we show that our model is able to discover some of the properties of the Alyawarra kinship system.

The Alyawarra have four kinship sections and the Alyawarra language contains a name for each section. To a first approximation, Alyawarra kinship is captured by the Kariera system shown in Figure 4. The system shows how the kinship sections of individuals are related to the kinship sections of their parents. For example, every member of section 1 has a mother in section 4 and a father in section 3. It follows from the system that marriages are only permitted between sections 1 and 2 and between sections 3 and 4. These marriage restrictions are one example of the important behavioral consequences of the Alyawarra kinship system. Denham, McDaniel and Atkins [12] have suggested that Alyawarra kinship is better captured by the Aranda system, an eight-class system which can be derived by splitting each of the four sections in two[4]. For our purposes, however, the simpler four class model is enough to give a flavor for the structure of the Alyawarra kinship system.

As part of his fieldwork with the Alyawarra, Denham took photographs of 225 people and asked 104 to provide a single kinship term for the subject of each photograph in the collection. We analyze the 104 by 104 square submatrix of the full 104 by 225

[2]Findler describes a case where the "extremely forceful injunction against a male person having sexual relations with his mother-in-law" could only be expressed by naming the pairs who could and could not engage in this act [10]

[3]Available online at the Alyawarra Ethnographic Archive: `http://www.alc.edu/denham/Alyawarra/`

[4]Denham and colleagues have also emphasized that Alyawarra practice departs from both the Kariera and Aranda systems[12, 13]. Both these normative systems rule out marriages that are sometimes seen in practice.

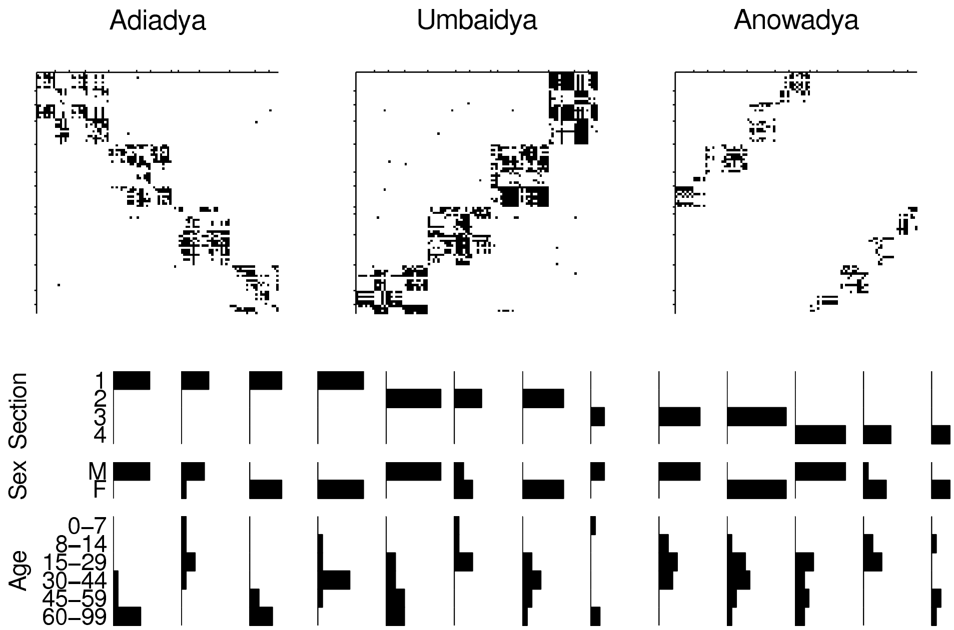


Figure 5: Top: Object graphs for three Alyawarra kinship terms described in the text. The 104 individuals are sorted by the 13 classes found by the infinite blockmodel (tick marks separate latent classes). Bottom: Breakdown of the 13 classes by kinship section, gender and age.

matrix of relations.[5] Figure 5 shows three of the 26 different kinship terms recorded, and the complete set of kinship terms is shown in Appendix A. For each term, the (i, j) cell in the corresponding matrix is shaded if person i used that term to refer to person j. The four kinship sections are clearly visible in the first two matrices. 'Adiadya' refers to a classificatory younger brother or sister: that is, to a younger person in one's own section, even if he or she is not a biological sibling. 'Umbaidya' is used by female speakers to refer to a classificatory son or daughter, and by male speakers to refer to the child of a classificatory sister. We see from the matrix that women in section 1 have children in section 4, and vice versa. 'Anowadya' refers to a preferred marriage partner. The eight rough blocks indicate that men must marry women, that members of section 1 are expected to marry members of section 2, and that members of section 3 are expected to marry members of section 4.

[5]We also ran simulations on the 104 by 225 matrix, ignoring the values missing from the full 225 by 225 matrix as described in Section 2.3. The results are similar but not as clean as for the 104 by 104 case, perhaps because the data missing from the full matrix are not missing at random as our algorithm assumes. For instance, no very young children appear among the 104 informants but there are many among the set of 225.

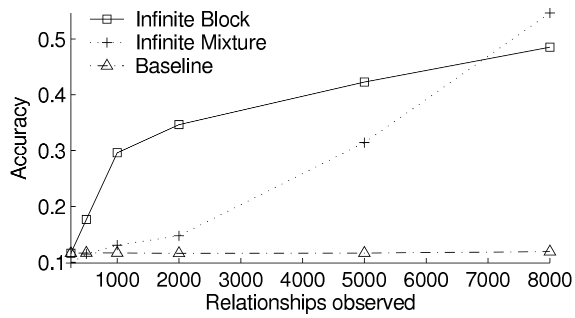


Figure 6: Accuracy at predicting unobserved kinship terms. The baseline model always chooses the most common term in the training set

We fit the infinite blockmodel to all 26 kinterm matrices simultaneously, treating each matrix as conditionally independent of all the others given an assignment of objects to classes. The maximum likelihood solution is represented in Figure 5. Denham recorded the age, gender and kinship section of each of his informants, and Figure 5 shows the composition of each class along each of these dimensions. The six age categories were chosen by Denham, and are based in part on his knowledge of Alyawarra terms for age groupings [11].

The blockmodel finds 13 classes, each of which includes members of just one kinship section. Section 1 is split into four classes corresponding to older men, older women, younger men and younger women. Section 3 is split into three classes: younger men, older men, and women. The remaining sections have one class for the younger people, and a class each for older men and older women. Note that none of the demographic data was used to fit the model — the 13 classes were discovered purely from the relational kinship data.

When given the same data, the maximum likelihood partition found by the purely attribute-based infinite mixture model is qualitatively worse. It includes only 5 classes: one for each of three kin sections, and two for Section 1 (split into older and younger people). We might expect that the true class structure has at least 16 classes (4 sections by 2 genders by 2 age categories) and probably more, since the age dimension might be broken into more than two categories. Neither model achieves a solution quite like this, but the blockmodel comes closer than the infinite mixture model.

To further explore the difference between the two models, we used them both for link prediction. We created sparse versions of the 104 by 104 matrix where only some of the entries were observed, and asked the algorithms to fill in the missing values. Figure 6 shows that the infinite blockmodel performs substantially better than the infinite mixture model when the number of examples is small. On very large training sets, however, the infinite mixture model has the advantage since it has more parameters to capture subtle trends in the data. A blockmodel with 13 classes has 169 parameters (one for each pair of classes): the mixture model with 5 classes has 1040 parameters, one for each feature-class pair. Both algorithms perform much better than a simple baseline that always chooses the most common kinship term in the training set.

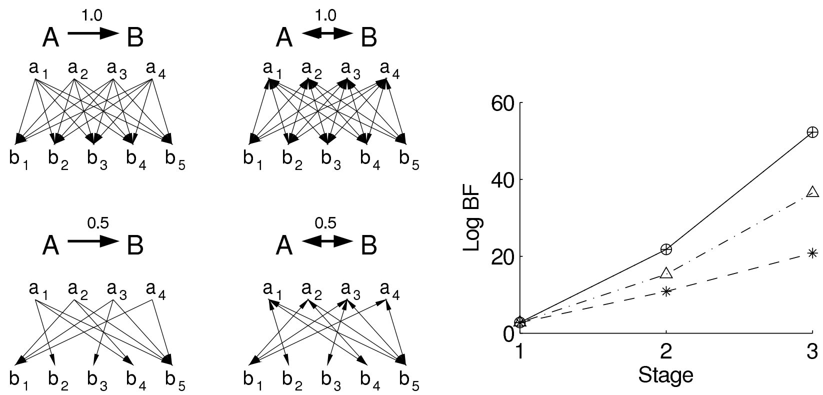


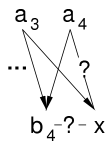
Figure 7: Left: Graphs representing the four conditions in the experiments of Tenenbaum and Niyogi. Right: Bayes factors (y-axis) for the first three stages (x-axis) of each condition comparing the infinite blockmodel to a null hypothesis where each object is placed in its own class. ($\circ : A \rightarrow B$, $+ : A \leftrightarrow B$, $\triangle$: $A \overset{0.5}{\rightarrow} B$, $*$: $A \overset{0.5}{\leftrightarrow} B$)

A limitation of both models is that neither is able to represent higher-order relationships (hierarchical or factorial) between the classes. We are currently exploring such extensions. A hierarchical model should learn, for example, that the first four classes found by our blockmodel can be grouped together into a higher-level class.

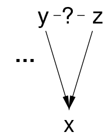
5 Causal Theories

Tenenbaum and Niyogi (2003) studied people's ability to learn simple causal theories in situations similar to the magnetism example mentioned earlier. This section uses the infinite blockmodel to explain some of their findings. The experiments considered here require subjects to interact with a computer-generated world where identical-looking objects can be moved around on screen. Some objects "activate" other objects whenever they touch. If x activates y (denoted $x \rightarrow y$), then y lights up and beeps whenever x and y touch. In some worlds, activation is symmetric (denoted $x \leftrightarrow y$): both x and y light up and beep. Unknown to the subjects, each object belongs to one of two classes, A or B. Figure 7 shows class graphs and object graphs for four experimental conditions. In the first world ($A \rightarrow B$), every A activates every B but activations are asymmetric. In the second world ($A \leftrightarrow B$), activations are symmetric: every A activates every B and every B activates every A. In the remaining two worlds ($A \overset{0.5}{\rightarrow} B$ and $A \overset{0.5}{\leftrightarrow} B$) , each A activates (asymmetrically or symmetrically) a random subset (on average, 50%) of B's.

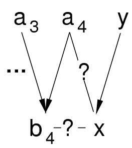
Tenenbaum and Niyogi (2003) examined whether subjects could discover these simple theories after interacting with some subset of the objects. Their experiments had seven phases and three new objects were added to the screen during each phase (see [14] for details). As new objects were added, subjects made predictions about how these objects would interact with old objects or with each other. At the end of the



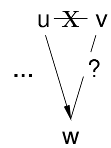
Query	True	P	B	A
$a_4 \to x$	1	0.91	0.98	0.95
$a_4 \leftarrow x$	0	0.06	0.01	0.05
$b_4 \to x$	0	0.09	0.05	0.29
$b_4 \to x$	0	0.15	0.06	0.29



Query	True	B	A
$y \to z$	0	0.08	0.5
$y \leftarrow z$	0	0.08	0.5



Query	True	P	B	A
$a_4 \to x$	1	0.88	0.92	0.61
$a_4 \leftarrow x$	0	0.15	0.03	0.39
$b_4 \leftarrow x$	0	0.15	0.13	0.39
$b_4 \to x$	0	0.21	0.09	0.40



Query	True	B	A
$v \to w$	1	0.94	0.50
$v \leftarrow w$	0	0.06	0.34

Figure 8: Predictions about new objects (v, w, y, z), after seeing old objects from the theory $A \to B$. Edges with question marks show activation relations to be predicted. The cross on the edge between u and v indicates that u and v have been observed not to activate each other. Tables show predictions of experimental subjects (P), of the infinite blockmodel (B) and of the infinite mixture model (A).

experiment, subjects gave a written description of how the objects worked. No mention of classes was made during the instructions, so inferring the existence of two classes and the relation between them constitutes a genuine discovery.

We consider two aspects of these experiments: the relative difficulty of learning the four theories shown in Figure 7, and the specific predictions that people make about relations involving new objects after they have learned one of these theories. Given experience with 18 objects, people had no difficulty learning the two deterministic theories $A \to B$ and $A \leftrightarrow B$. The asymmetric nondeterministic structure, $A \overset{0.5}{\to} B$, was much more difficult; only about half of 18 subjects succeeded on this task. The symmetric nondeterministic structure, $A \overset{0.5}{\leftrightarrow} B$, was the most difficult; only two out of 18 subjects attained even partial success.

These findings are consistent with the behavior of a Bayesian learner inferring the theory that best explains the observed relations. The weight of the evidence that the world respects a block structure can be expressed as the marginal likelihood of the observed relational data under the infinite block model. We computed these likelihoods by enumerating then summing over all possible class assignments z for up to 9 objects. Figure 7 plots Bayes factors (log ratio of evidence terms) for the infinite blockmodel relative to a “null hypothesis” where each object belongs to its own class. The Bayes factors increase in all cases as more objects and relations are observed, but the rate of increase varies across the four theories in accordance with their relative ease of learning.

Learning the correct causal theory based on a set of observed relations should allow people to infer the unobserved causal relations that will hold for a new object x in the same domain, provided they observe sufficient data to infer the class membership of x. Figure 8 shows several kinds of relational prediction that human learners can perform. All of these examples assume a learner who has observed the objects and relations in Figure 7 generated by the $A \to B$ theory. Given a new object x which has just been activated by an old A object, a learner with the correct theory should classify x as a

B, and predict that another A will activate x but that nothing will happen between x and a B. Analogous predictions can be made if x is observed only to be activated by a new object y. Figure 8 shows that people make these predictions correctly after learning the theory [14], as does the infinite blockmodel. The infinite mixture model performs poorly on these tasks (Figure 8) as a consequence of treating relations like attributes. Under this model, learning about relations between new objects is identical to learning about entirely new features, and *none* of the learner's previous experience is relevant. Only the blockmodel thus accounts for a principle function of intuitive theories: generalization from previous experience to wholly new systems in the same domain.

6 Conclusions and future directions

We have presented an infinite generative model for representing abstract relational knowledge and discovering the latent classes generating those relations. This analysis hardly begins to approach the richness and flexibility of people's intuitive domain theories, but may provide some of the critical building blocks. It may also be useful in other fields. Our framework for discovering latent classes is an extension of relational models previously proposed in sociology (stochastic block models [4]) and machine learning (probabilistic relational models [15]). We are also exploring the cognitive relevance of other relational structures discussed in these fields, such as the overlapping class model of Kubica et al. [16], and structures where groups are defined by regular equivalence [17]. Developing a framework that can form spontaneous and flexible combinations of these structures remains a formidable open task.

Acknowledgments

We thank Woodrow Denham for supplying the Alyawarra data in machine readable form, and Whitman Richards for helpful comments on an early version of this paper. This work was supported by NTT Communication Sciences Lab, the DARPA CALO program, and the Paul E. Newton Chair (JBT).

References

[1] J. R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429, 1991.

[2] R. M. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57, 1986.

[3] S. Wasserman and C. Anderson. Stochastic a posteriori blockmodels: construction and assessment. 9:1–36, 1987.

[4] K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[5] R. M. Neal. Bayesian mixture modeling by Monte Carlo simulation. Technical Report 91-2, University of Toronto, 1991.

[6] C. E. Rasmussen. The infinite Gaussian mixture model. In *NIPS*, volume 13, 2002.

[7] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, Suffolk, 1996.

[8] S. Jain and R. M. Neal. A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. Technical report, University of Toronto, 2000.

[9] C. Lévi-Strauss. *The Elementary Structures of Kinship*. Beacon Press, Cambridge, MA, 1969.

[10] Nicholas Findler. Automatic rule discovery for field work in anthropology. *Computers and the Humanities*, 25:295–392, 1992.

[11] Woodrow Denham. *The detection of patterns in Alyawarra nonverbal behavior*. PhD thesis, University of Washington, 1973.

[12] Woodrow W. Denham, Chad K. McDaniel, and John R. Atkins. Aranda and Alyawarra kinship: A quantitative argument for a double helix model. *American Ethnologist*, 6(1):1–24, 1979.

[13] Woodrow W. Denham and Douglas R. White. Multiple measures of Alyawarra kinship. *Field Methods*. To appear.

[14] J. B. Tenenbaum and S. Niyogi. Learning causal laws. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. 2003.

[15] B. Taskar, E. Segal, and D. Koller. Probabilistic classification and clustering in relational data. In *IJCAI*, volume 15, 2001.

[16] J. Schneider J. Kubica, A.Moore and Y. Yang. Stochastic link and group detection. In *IJCAI*, 2002.

[17] S. Wasserman and K. Faust. *Social network analysis: Methods and applications*. Cambridge University Press, 1994.

[18] Woodrow W. Denham. Alyawarra ethnographic database: Numerical data documentation file, version 7. 2001. Available at `http://www.alc.edu/denham/Alyawarra/`.

A Kinship Data

Arengiya	FF/FFZ, SS/SD (ms), BSS/BSD (fs)
Anyaina	MM/MMB, MMBSS/MMBSD, ZDS/ZDD (ms), DS/DD (fs)
Aidmeniya	MMBSS/MMBSD, ZDS/ZDD (ms), DS/DD (fs)
Aburliya	FM/FMB, FMBSD/FMBSS, ZSS/ZSD (ms), SS/SD (fs)
Adardiya	MF/MFZ, DS/DD, BDS/BDD (fs)
Agniya	F
Aweniya	FZ, FMZD
Amaidya	M, SW (ms)
Abmarliya	MB, SWB (ms)
Awaadya	EB
Anguriya	EZ
Adiadya	YB/YZ
Angeliya	FZS/MBS
Algyeliya	FZD/MBD
Adniadya	MBS
Aleriya	S/D (ms), BS/BD (fs)
Umbaidya	S/D (fs), ZS/ZD (ms), FMBS/FMBD
Anowadya	W/MMBDD (ms), H/MFZDS (fs)
Muriya	MMBD/MMBS, WM/WMB (ms), ZDH/ZDHZ (ms)
Agenduriya	ZS/ZD (ms), rare term for biological sister's child
Amburniya	WB/ZH
Andungiya	HZ/BW (fs)
Aneriya	BWM/DHZ (fs)
Aiyenga	"Myself"
Undyaidya	WZ (ms), rare term used as reciprocal for amburniya
Gnaldena	YZ, rare term for biological younger sister

Table 1: Glosses given by Denham [18] for the 26 kinship terms in Figure 9. F=father, M=mother, B=brother, Z=sister, S=son, D=daughter, H=husband, W=wife, E=elder, Y=younger, fs=female speaker, ms=male speaker. For example, adiadya refers to a classificatory younger brother (YB) or younger sister (YZ).

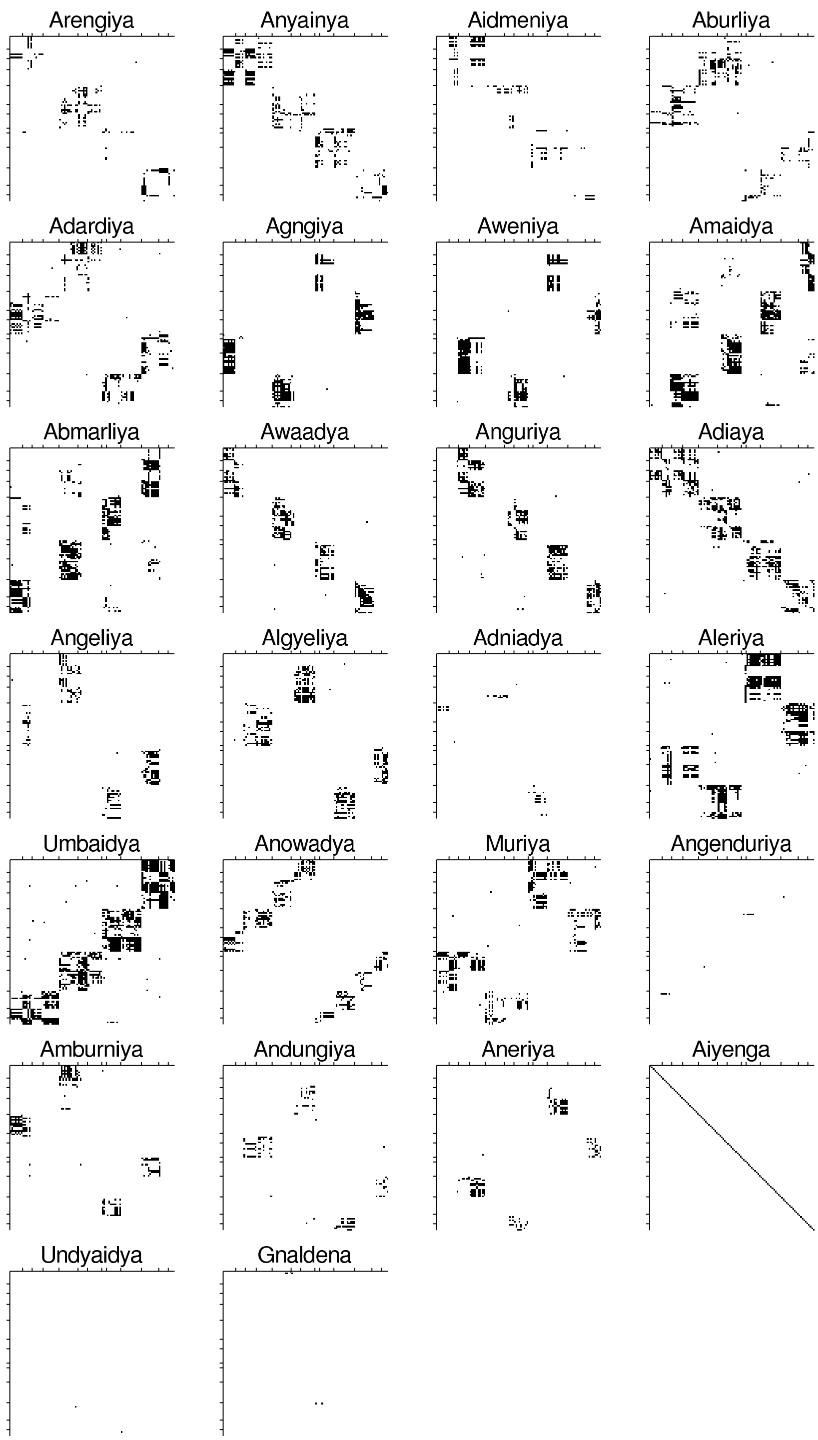


Figure 9: Matrices for the 26 kinterms recorded by Denham. The 104 individuals are sorted according to the 13 classes found by the infinite blockmodel. Glosses for the terms are shown in Table 1.