

# Machine Learning a Probabilistic Network of Ecological Interactions

Alireza Tamaddoni-Nezhad<sup>1</sup>, David Bohan<sup>2,3</sup>,  
Alan Raybould<sup>4</sup>, and Stephen H. Muggleton<sup>1</sup>

<sup>1</sup> Department of Computing, Imperial College London, London, SW7 2AZ, UK  
`{a.tamaddoni-nezhad,s.muggleton}@imperial.ac.uk`

<sup>2</sup> Rothamsted Research, West Common, Harpenden, Herts, AL5 2JQ, UK  
`david.bohan@rothamsted.ac.uk`

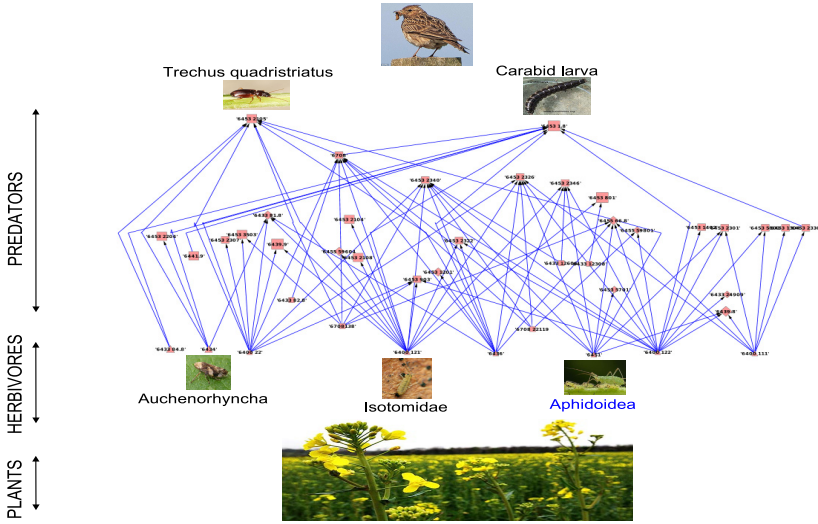
<sup>3</sup> INRA, UMR 1210 Biologie et Gestion des Adventices, BP 86510, Dijon, France

<sup>4</sup> Syngenta Ltd., Bracknell, Berkshire, RG42 6EY, UK  
`alan.raybould@syngenta.com`

**Abstract.** In this paper we demonstrate that machine learning (using Abductive ILP) can generate plausible and testable food webs from ecological data. In this approach, unlike previous applications of Abductive ILP, the abductive predicate ‘eats’ is entirely undefined before the start of the learning. We also explore a new approach, called Hypothesis Frequency Estimation (HFE), for estimating probabilities for hypothetical ‘eats’ facts based on their frequency of occurrence when randomly sampling the hypothesis space. The results of cross-validation tests suggest that the trophic networks with probabilities have higher predictive accuracies compared to the networks without probabilities. The proposed trophic networks have been examined by domain experts and comparison with the literature shows that many of the links are corroborated by the literature. In particular, links ascribed with high frequency are shown to correspond well with those having multiple references in the literature. In some cases novel high frequency links are suggested, which could be tested.

## 1 Introduction

Machine Learning has the potential to address many challenging problems in ecological sciences [5]. Discovery of trophic links (food chains) which describe the flow of energy/biomass between species is one of these problems. Networks of trophic links (food webs) are important for explaining ecosystem structure and dynamics [3]. However, relatively few ecosystems have been studied through detailed food webs because finding out the predation relationships between the many hundreds of species in an ecosystem is difficult and expensive. Hence, any technique which can automate the discovery of trophic links from ecological data is highly desirable. Similar problems of network construction have been tackled in other complex systems, such as metabolic networks (e.g. [16]). In this paper we demonstrate that Abductive ILP can generate plausible and testable food webs from ecological data. In this approach the abductive predicate ‘eats’ is entirely undefined before the start of the learning process. This contrasts with



**Fig. 1.** A farmland food web. Networks of trophic links (food web) are the primary routes for translocation of energy/biomass between species at different levels and are important for explaining ecosystem structure and dynamics.

previous applications of Abductive ILP where partial, non-empty, definitions exist and the gaps are filled by abduced hypotheses. In this paper we also explore a new approach, called Hypothesis Frequency Estimation (HFE), for estimating probabilities for hypothetical ‘eats’ facts based on their frequency of occurrence when random permutations of the training data (and hence different seeds for defining the hypothesis space) are considered. We empirically evaluate the hypothetical trophic networks using leave-one-out cross-validation tests on the observable data. The results of cross-validation tests for the networks with and without probabilities are presented. The proposed trophic networks have been examined by domain experts and the results of comparison with the literature are presented.

This paper is organised as follows. In Section 2 we describe the ecological problem and data. In Section 3 we explain how Abductive ILP has been used to learn trophic networks from ecological data. Hypothesis Frequency Estimation (HFE), a method which we used for estimating probabilities, is described in Section 4. Empirical and ecological evaluations of the hypothetical trophic networks are presented and discussed in Sections 5 and 6 respectively. Related and further work are discussed in Section 7. Section 8 concludes the paper.

## 2 Ecological Problem and Data

Ecosystems are structured by flows of energy/biomass between primary producer plants (autotrophs) and consumers (heterotrophs), such as invertebrates,

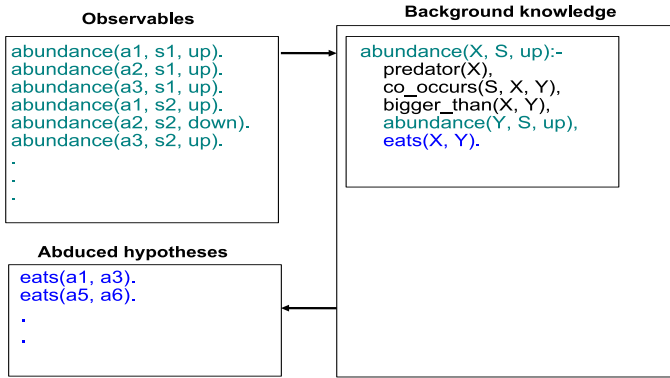
mammals and birds. Networks of trophic links (food web) are the primary routes for translocation of energy/biomass between species at different levels. For example, Figure 1 shows a food web in farmland in which the food chains link plants, herbivores and predators; herbivores eat plants, predators eat herbivores or other predators and so on. There is much concern about anthropogenic and natural degradation of ecosystems worldwide, and the knock-on consequences for ecosystem functioning. Much research that seeks to identify mechanisms of ecosystem change and devise methods for mitigating its effects is being hampered by an inability to construct food webs efficiently and effectively. Put simply, it takes considerable time and effort to establish trophic relationships between the many hundreds of species in an ecosystem - this means that only a relatively few systems have been studied to any depth making it difficult to produce general theories about ecosystem change. A method that yields plausible and testable food webs from already collected field data would be a major step forward in the identification of rules about how ecosystems react when perturbed.

In this paper we try to answer the following question. Can machine learning be used to construct food webs from ecological data? We think the answer is yes and in fact the food web shown in Figure 1 has been learned from ecological data. In this paper we show how this food web has been constructed using Abductive ILP.

The training data we use comes from arable farmland where anthropogenic disturbance and farm management has led to great increases in crop productivity, but often at cost to biodiversity. The data set was sampled from 257 fields across the UK in the Farm Scale Evaluations (FSE) of GM, herbicide tolerant (GMHT) crops. This national-scale experiment evaluated the change in weed plants and invertebrates between the current, conventional herbicide management of spring-sown Maize, Beet and Oilseed Rape and winter-sown Oilseed Rape, and the herbicide management of GMHT varieties of the same crops using a split-field design. We use data from the Vortis suction sampling protocol for epigeal invertebrates [7,1] to calculate a treatment effect ratio. The counts from each conventional and GMHT half-field pair were converted to multiplicative treatment ratio,  $R$ , and as in [7,1] treatment ratio values of  $R < 0.67$  and  $R > 1.5$  were regarded as important changes in count with direction of down (decreased) and up (increased), respectively. This information on up and down abundances is regarded as our primary observational data for the learning.

### 3 Machine Learning of Trophic Links Using Abductive ILP

The main role of abductive reasoning in machine learning and its use in the development of scientific theories [6] is to provide hypothetical explanations of the empirical observations. Then based on these explanations we try to inject back into the current scientific theory, new information that helps complete the theory. This process of generating abductive explanations and then updating the theory with them can be repeated several times when new observational data is



**Fig. 2.** Learning trophic links from ecological data using Abductive ILP

made available. In many implementations of abductive reasoning, such as that of Progol 5 [11], which is used in this paper, the approach taken is to choose an explanation that best generalises under some form of inductive reasoning (e.g. simplest explanation approximated by compressibility). We refer to this approach as Abductive ILP (A/ILP). We believe that ecological data in this study fulfil the conditions for the use of A/ILP: firstly, the given background knowledge is incomplete; and secondly, the problem requires learning in the circumstance in which the hypothesis language is disjoint from the observation language. In our problem, the set of observable data can be represented by predicate *abundance*( $X, S, up$ ) (or *abundance*( $X, S, down$ )) expressing the fact that the abundance of  $X$  at site  $S$  is *up* (or *down*). This information is compiled from FSE data as described in Section 2. The knowledge gap that we initially aim to fill is a predation relationship between species. Thus, we declare abducible predicate *eats*( $X, Y$ ) capturing the hypothesis that species  $X$  eats species  $Y$ . In order to use abduction, we also need to provide the rules which describe the observable predicate in terms of the abducible predicate. An example of such a rule is shown below.

```

abundance(X, S, up):-
  predator(X),
  co_occurs(S, X, Y),
  bigger_than(X, Y),
  abundance(Y, S, up),
  eats(X, Y).
  
```

Similarly, a rule for *abundance*( $X, S, down$ ) can be defined. This Prolog rule expresses the inference that following a perturbation in the ecosystem (caused by the management), the increased (or decreased) abundance of species  $X$  at site  $S$  can be explained by the fact that  $X$  eats species  $Y$  which is further down in the food chain and the abundance of species  $Y$  is increased (or decreased). It also includes additional conditions to constraint the search for abducible predicate *eats*( $X, Y$ ), i.e.  $X$  should be a predator,  $X$  and  $Y$  should co-occur and that

$X$  should be bigger than  $Y$ . Predicates  $predator(X)$ ,  $co\_occurs(S, X, Y)$  and  $bigger\_than(X, Y)$  are provided as part of the background knowledge. Given this model and the observable data, Progol 5 generates a set of ground abductive hypotheses in the form of ‘eats’ relations between species as shown in Figure 2. These abductive hypotheses are automatically generated by matching the given information to the rule in order to abduce a fact which explains the observations. In this example, given the inputs, abduction will generate the hypotheses that a particular species  $a$  eats a particular species  $b$ . In general, many choices for matching could be made, leading to a variety of alternative hypotheses and a preference is imposed by Progol 5 using an information-theoretic criterion known as compression [11]. Here,  $compression = p - n - h$ , where  $p$  is the number of observations correctly explained by the hypothesis,  $n$  is the number incorrectly explained and  $h$  is the length of the hypothesis (e.g. 1 for a single fact). The set of ground hypotheses can be visualised as a network of trophic links (food webs) as shown in Figure 3. In this network a ground fact  $eats(a, b)$  is represented by a trophic link from species  $b$  to species  $a$ .

## 4 Hypothesis Frequency Estimation (HFE)

In order to get probability estimates for hypotheses, we use a technique which is based on direct sampling from the hypothesis space. In some ILP systems, including Progol, training examples also act as seeds to define the hypothesis space (e.g. a most specific clause is built from the next positive example). Hence, different permutations of the training examples define different parts of the hypothesis space. We use this property to sample from the hypothesis space by random permutations of the training data. Probability of ground hypotheses can be estimated based on the frequency of occurrence when random permutations of the training data (and hence different seeds for defining the hypothesis space) are considered. Using this technique, the thickness of trophic links in Figure 3 represent probabilities which are estimated based on the frequency of occurrence from 10 random permutations of the training data. As shown in Figure 4, the probabilistic trophic network can be also represented using standard probabilistic representations in ILP such as SLPs [12] or ProbLog [13]. For this we can use relative frequencies in the same way probabilities are used in probabilistic ILP. We can then use the probabilistic inferences based on these representations to estimate probabilities. For example, the probability  $p(abundance(a, s, up))$  can be estimated by relative frequency of hypotheses which imply  $a$  at site  $s$  is  $up$ . Similarly,  $p(abundance(a, s, down))$  can be estimated and by comparing these probabilities we can decide to predict whether the abundance is up or down. We have used this method in leave-one-out experiments in Section 5 to measure the predictive accuracies of probabilistic trophic networks.

## 5 Empirical Evaluation

The purpose of the experiments in this section is to empirically evaluate hypothetical trophic networks constructed from real ecological data using the methods



HFE		Probabilistic Representation	Probabilistic Inference
<u>abduced trophic link</u>	<u>freq.</u>		
eats('6453 801','6433 12308').	10	1.0:: eats('6453 801', '6433 12308').	
eats('6453 1.8','6455 94').	9	0.9:: eats('6453 1.8', '6455 94').	
eats('6453 2330','6400 111').	6	0.6:: eats('6453 2330', '6400 111').	
eats('6453 2346','6433 12606').	10	1.0:: eats('6453 2346', '6433 12606').	
eats('6453 2346','6433 82.8').	9	0.9:: eats('6453 2346', '6433 82.8').	
eats('6455 66.8','6436').	1	0.1:: eats('6455 66.8', '6436').	
eats('6453 2204','6400 111').	1	0.1:: eats('6453 2204', '6400 111').	
.		.	
.		.	
.		.	
		abundance(X, S, up) :-	$p(\text{abundance}(a,s,up))$
		predator(X),	relative frequency of hypotheses
		co_occurs(S, X, Y),	which imply <i>a</i> at site <i>s</i> is up
		bigger_than(X, Y),	.
		abundance(Y, S, up),	.
		eats(X, Y).	.
		.	
		.	

**Fig. 4.** Hypothesis Frequency Estimation (HFE) and probabilistic representation and inference

described in the previous sections. In these experiments, firstly, we determine if trophic networks can be learned from real ecological data such that their predictive accuracies on unseen data are higher than the default accuracy of the majority class. Secondly, we examine if using probabilities estimated by the permutation based HFE method (see Section 4) leads to increased predictive accuracy.

### 5.1 Experiment 1

In this experiment we test the following null hypothesis:

**Null Hypothesis 1:** A trophic network with predictive accuracy greater than the default accuracy cannot be constructed from ecological data using Abductive ILP.

**Materials and Methods.** In this experiment Progol 5.0 <sup>1</sup> is used to generate abductive hypotheses (i.e. ‘eats’ relations between species) from observable data (i.e. up/down abundance of species at different sites). The observable data has been compiled from FSE data as described in Section 2. The up/down abundance of species at different sites are then encoded as predicates *abundance(X, S, up)*

<sup>1</sup> Available from: <http://www.doc.ic.ac.uk/~shm/Software/progol5.0/>

---

```

for j in [1..300] do
  Tsj = test example; abundance of species-site j
  Trj = E - Tsj
  for k in (25,50,75,100) do
    for i in [1..10] do
      Trji = ith permutation of Trj
      Trjik = training examples; k% random sample from Trji
    end
  end
end
for j in [1..300] do
  for k in (25,50,75,100) do
    for i in [1..10] do
      Hjik = learned hypotheses using the training set Trjik
      Atji = predictive accuracy of Hjik = on the test set Tsj
    end
  end
end
for k in (25,50,75,100) do
  Plot average of Ajik versus k (j ∈ [1..300] and i ∈ [1..10])

```

---

**Fig. 5.** Experimental method using a leave-one-out test strategy.  $E$  is the set of abundance data, variable  $j \in [1..300]$  represents all species-sites combination, variable  $i \in [1..10]$  represents 10 random permutation and variable  $k = (25, 50, 75, 100)$  represents the varying size of training data.

and  $abundance(X, S, down)$ . The background knowledge includes information about sites and species and Prolog rules for  $abundance$  as well as predicates  $predator(X)$ ,  $co\_occurs(S, X, Y)$  and  $bigger\_than(X, Y)$  as described in Section 3. In order to empirically evaluate the hypothetical trophic networks, we use a leave-one-out cross-validation test on the observable data for species in the network. This cross-validation test was done by leaving out the abundance of each predator at each site and then trying to predict whether the abundance is up or down, given the trophic network generated from the rest of the data. The experimental method is detailed in Figure 5.

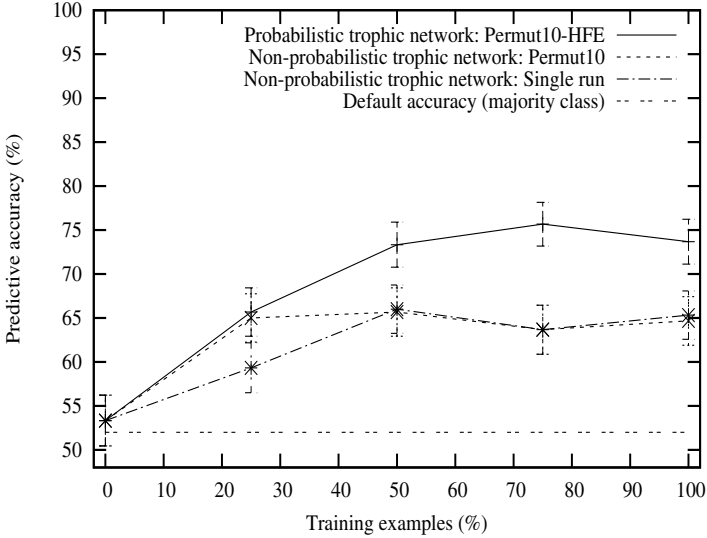
**Results and Discussion.** Figure 6 compares the predictive accuracy of non-probabilistic networks, i.e. networks generated from a single run or from 10 random permutations as well as probabilistic networks. In all cases the predictive accuracies are significantly higher than the default accuracy of the majority class (i.e. *down* for 53% of all abundance data). We can thus refute null hypothesis 1.

## 5.2 Experiment 2

In this experiment we test the following null hypothesis:

**Null Hypothesis 2:** Using the permutation based HFE method for estimating probabilities of trophic links does not lead to increased predictive accuracies compared to the non-probabilistic trophic network.





**Fig. 6.** Predictive accuracies of probabilistic trophic network vs. non-probabilistic networks from leave-one-out cross-validation tests

**Materials and Methods.** As in Experiment 1, we use a leave-one-out cross-validation test on the observable data for species in the network, i.e. leaving out the abundance of each predator at each site and trying to predict whether the abundance is up or down, given the trophic network generated from the rest of the data. For the trophic network with probabilities, we first need to calculate the relative frequencies of hypotheses which imply that the abundance of the test example is *up* or *down*. Let  $p(\text{abundance}(a, s, \text{up}))$  be the relative frequency of hypotheses which imply the abundance of  $a$  at site  $s$  is *up* and  $p(\text{abundance}(a, s, \text{down}))$  is defined analogously. If  $p(\text{abundance}(a, s, \text{up})) > p(\text{abundance}(a, s, \text{down}))$  then we predict that the abundance of the test example  $a$  is *up* and otherwise it is *down*.

**Results and Discussion.** As shown in Figure 6 the predictive accuracies for the non-probabilistic networks are significantly lower than the probabilistic networks when more than 50% of the training data are provided. Hence, we can reject null hypothesis 2.

## 6 Ecological Evaluation

The trophic network in Figure 3 has been examined by the domain experts and corroboration of many of the links in the literature have been found. A detailed analysis of this hypothetical trophic network is presented in [2]. Table 7 is a

	Anthrenorhynchus nemorum	Bembidion lampros	Bembidion lunulatum	Bembidion obtusum	Cimicidae nymphs	Curculionidae	Entomobryidae	Isotomidae	Lepthyphantes tenuis	Licoon tripustulatus	Miridae nymphs	Ortus vicinus	Poduridae	Scoloposlethus affinis	Sminthuridae
Agonum dorsale							9 [13]	4	5 [6]						10 [13]
Bembidion aeneum							10 [11]		9 [6]				9 [11]		
Bembidion biguttatum															10 [11]
Bembidion guttula							7 [11]	10 [11]	9 [6]						
Bembidion lampros					9		10 [11]	10 [11]	10 [6]				10 [11]	9	10 [11]
Bembidion obtusum					9		10 [11]	10 [11]	10 [6]			10			10 [11]
Bembidion quadrimaculatum					9		10 [11]	10 [11]	9 [6]			10			9 [11]
Bradycellus verbasci					8										
Carabid larvae	9	10	3			9	10 [2]	10 [2]	9		10		9 [2]		10 [2]
Clivina fossor							7 [12,9]		2						
Coccinellid larvae		9			9		10 [14,16]	9 [14,16]			1		10 [14,16]		9 [14,16]
Coccinella septempunctata								10 [14,16]							
Dromius linearis							10 [1]	7					3		
Loxocera pilicornis										9					
Nabidae nymphs				3				10	7		2 [10,7]				10 [9]
Nebria brevicollis	10							10 [15]							
Notophilus biguttatus							10 [4,11,3]	10 [4,11,3]							10 [4,11,3]
Propylea quatuordecimpunctata							10 [14,16]		10 [16]						
Salidula saltatoria							10 [8]								
Trechus quadristriatus		9		9	9		9 [15,3]	4 [15,3]	9 [6]		2				10 [15,3]
Trechus secalis				2				8 [15,3]							

**Fig. 7.** Tabulated trophic links for some prey (columns) and predator (rows) species combination in Figure 3. Each pairwise hypothesised link has a strength (i.e. frequency between 1 to 10) followed by references (in square brackets) in the literature (listed in Appendix) supporting the link. Multiple references are indicated by yellow and green circles and potential novel hypotheses by dashed red circles.

tabular representation for some prey (columns) and predator (rows) species combination in Figure 3. Each pairwise hypothesised link has a strength (i.e. frequency between 1 to 10) followed by references (in square brackets) in the literature (listed in Appendix) supporting the link. In this table, only prey/predators are shown which have at least one link with strength more than or equal to 7. This

table shows that many of the links, suggested by the model, are corroborated by the literature. In particular, links in the model ascribed with high frequency are shown to correspond well with those having multiple references in the literature. For example, there are 15 links with more than two references and 8 of these are with frequency 10 and from these all the three links with three references (marked by green circles) have frequency 10. But there are also highly frequent links with no references in the literature which could potentially be novel hypotheses. It should be noted that the corroboration of the links in the literature was started by the domain experts based on a network from single runs (without probabilities) and the probabilities were added to the table after the references were assigned. Hence, the corroboration of the links was done independently and could have not been affected by the frequency assignments. Figure 8a shows the correspondence between the frequencies of hypothetical trophic links, the number of multiple references and the total number of references in the literature. We use the numbers in this table to test the following null hypothesis.

**Null Hypothesis 3:** The frequency of hypothetical trophic links in the learned trophic network are not correlated with the number of references in the literature for those trophic links.

Figure 8b shows the correlation between frequencies and the total number of references. If we use Spearman’s correlation between the frequencies and the total number of references then the  $\rho$  value and the  $p$ -value are 0.77 and 0.009 respectively. Hence, hypothesis 3 is refuted.

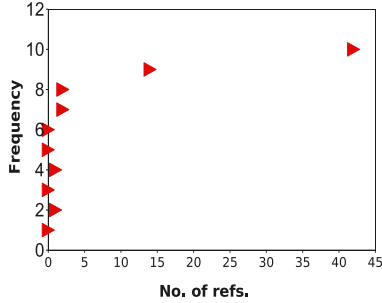
As mentioned before, according to Figure 7 in some cases novel high frequency links (with no references in the literature) are suggested, which could be tested. New experimental studies are needed to test these potential novel hypotheses.

## 7 Discussion and Related Work

In this section we first try to answer the following questions. Can the synthesis of trophic networks, in the way described in this paper, be regarded as machine learning? and if so, how this can be compared to other related work? In order to answer these questions, first we note that abduction is normally defined as an inference based on formulation of explanations. The abductive inference (i.e. A/ILP) described in this paper has been used to construct a graph given a relational background knowledge. This is also related to other forms of graph learning in ILP which consider graphs as relations (e.g. [8]). It is therefore important to note that in this approach we use first-order background knowledge in the construction of relations. Moreover, according to the definition of machine learning, a system learns if it improves on prediction performance when provided with observational data [10]. In this paper we have demonstrated that the constructed trophic networks improve predictive accuracy on “out of sample” test data. Therefore it follows from definition that the construction of trophic networks, in the way described in this paper, can be regarded as machine learning.

Freq.	1 Ref.	2 Ref.	3 Ref.	Total refs.
1	0	0	0	0
2	0	1	0	1
3	0	0	0	0
4	0	1	0	1
5	0	0	0	0
6	0	0	0	0
7	1	1	0	2
8	0	1	0	2
9	8	3	0	14
10	23	5	3	42

(a)



(b)

**Fig. 8. a)** Correspondence between the frequencies, the number of multiple references and the total number of references in the literature. **b)** Correlation between frequencies and the total number of references. Spearman's correlation  $\rho$  value is 0.77 with  $p$ -value 0.009.

This work is closely related to previous applications of A/ILP for the analysis of biological networks at the molecular and cellular scale. In particular, the logical models for up/down abundances and the flow of biomass at species level can be compared with the logical models for up/down concentration of metabolites and the bio-chemical reactions in metabolic networks (e.g. [16]). However, there are major differences and in particular two aspects of the use of A/ILP in this paper are novel. Firstly, unlike previous applications of A/ILP, the abductive predicate 'eats' is entirely undefined before the start of the learning process. The second novel aspect of the approach relates to the Hypothesis Frequency Estimation (HFE) method for the assignment of probabilities to hypothetical 'eats' facts based on their frequency of occurrence when randomly sampling the hypothesis space. This approach helps to separate those trophic links with low probabilities, which represent unstable artefacts, possibly of example ordering, from those with high probabilities that can be viewed as stable and reliable hypotheses. The resulting probabilistic network is a compact summary of the hypothesis space with a posterior distribution which could be viewed as a Bayes predictor.

The permutation based Hypothesis Frequency Estimation (HFE) is comparable to the probabilistic learning methods such as [15] and [9] which use an EM algorithm in an abductive learning setting and [14] which can learn ground acyclic

ProbLog programs by transforming them into Bayesian networks. Even though these methods provide advanced parameter learning algorithms, the advantage of our simple frequency based method (which directly samples the hypothesis space) is that the structure and the parameters of the network can be learned at the same time in an incremental learning approach.

In this paper we have only reported the predictive accuracies for binary classification. However, we have also used expected utilities implemented as Decision-Theoretic Logic Programs (DTLPs) [4] for estimating  $R$  values (treatment effect ratio as described in Section 2). Initial results suggest that using probabilities leads to reduced mean square errors when estimating  $R$  values in cross-validation tests. The probabilistic trophic network together with the expected utility approach can be viewed as a Decision-Theoretic representation which we call an Acyclic Expectation Network (AEN). We intend to study different aspects of this representation in a follow up paper.

## 8 Conclusions

We have shown that machine learning, using A/ILP, can produce a convincing food web from sample ecological data. We have also demonstrated a new approach, called Hypothesis Frequency Estimation (HFE), for estimating probabilities for hypothetical trophic links based on their frequency of occurrence when randomly sampling the hypothesis space. The results of cross-validation tests suggest that the trophic networks with probabilities have higher predictive accuracies compared to the networks without probabilities. Many of the abduced trophic links, especially those with high frequencies, are supported either by the literature or the expert knowledge of agricultural ecologists. The food web representing probabilistic interactions between species can readily be interpreted by ecologists and the logical framework for learning trophic links can be openly discussed, *a priori*, and the hypothesised links are not an abstract, statistical product of the data.

This is to our knowledge the first time that machine learning (using ILP) has been used to generate plausible and testable food webs from ecological data. We believe that automated discovery of food webs from ecological data could lead to important advances in ecological theory and improved management of ecosystems under environmental change. The ecological aspects of this work and a more detailed analysis of the learned trophic links are discussed in [2].

**Acknowledgements.** We would like to thank S. Dunbar for his encouragement and support. We also thank G. Caron-Lormier, J. Chen, V. Lesk and V. Santos Costa for helpful discussions and the anonymous referees for useful comments. We acknowledge the help provided by Agricultural Ecologists and Entomologists at Rothamsted Research. This work was funded by Syngenta University Innovation Centre (UIC) at Imperial College London.

## References

1. Bohan, D.A., Boffey, C.W.H., Brooks, D.R., Clark, S.J., Dewar, A.M., Firbank, L.G., Haughton, A.J., Hawes, C., Heard, M.S., May, M.J., et al.: Effects on weed and invertebrate abundance and diversity of herbicide management in genetically modified herbicide-tolerant winter-sown oilseed rape. *Proceedings of the Royal Society B: Biological Sciences* 272(1562), 463 (2005)
2. Bohan, D.A., Caron-Lormier, G., Muggleton, S., Raybould, A., Tamaddoni-Nezhad, A.: Automated Discovery of Food Webs from Ecological Data Using Logic-Based Machine Learning. *PLoS ONE* 6(12), e29028 (2011)
3. Caron-Lormier, G., Bohan, D.A., Hawes, C., Raybould, A., Haughton, A.J., Humphry, R.W.: How might we model an ecosystem? *Ecological Modelling* 220(17), 1935–1949 (2009)
4. Chen, J., Muggleton, S.: Decision-theoretic logic programs. Poster Presentation *Proceedings of 19th International Conference on Inductive Logic Programming* (2010)
5. Dietterich, T.G.: Machine learning in ecosystem informatics and sustainability. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI, Pasadena, Calif.*, pp. 8–13 (2009)
6. Flach, P.A., Kakas, A.C.: *Abduction and Induction: Essays on their relation and integration*. Springer, Netherlands (2000)
7. Haughton, A.J., Champion, G.T., Hawes, C., Heard, M.S., Brooks, D.R., Bohan, D.A., Clark, S.J., Dewar, A.M., Firbank, L.G., Osborne, J.L., et al.: Invertebrate responses to the management of genetically modified herbicide-tolerant and conventional spring crops. ii. within-field epigeal and aerial arthropods. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358(1439), 1863 (2003)
8. Horváth, T., Bringmann, B., De Raedt, L.: Frequent Hypergraph Mining. In: Muggleton, S.H., Otero, R., Tamaddoni-Nezhad, A. (eds.) *ILP 2006. LNCS (LNAI)*, vol. 4455, pp. 244–259. Springer, Heidelberg (2007)
9. Inoue, K., Sato, T., Ishihata, M., Kameya, Y., Nabeshima, H.: Evaluating abductive hypotheses using an em algorithm on bdds. In: *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pp. 810–815. Morgan Kaufmann Publishers Inc. (2009)
10. Mitchell, T.M.: *Machine learning*, p. 368. Mac Graw Hill (1997)
11. Muggleton, S.H., Bryant, C.H.: Theory Completion Using Inverse Entailment. In: Cussens, J., Frisch, A.M. (eds.) *ILP 2000. LNCS (LNAI)*, vol. 1866, pp. 130–146. Springer, Heidelberg (2000)
12. Muggleton, S.H.: Stochastic logic programs. In: De Raedt, L. (ed.) *Advances in Inductive Logic Programming*, pp. 254–264. IOS Press (1996)
13. De Raedt, L., Kimmig, A., Toivonen, H.: Problog: A probabilistic prolog and its applications in link discovery. In: Lopez de Mantaras, R., Veloso, M.M. (eds.) *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 804–809 (2007)
14. Riguzzi, F.: Learning ground problog programs from interpretations. In: *6th International Workshop on Multi-relational Data Mining*, p. 105 (2007)
15. Sato, T., Kameya, Y.: Prism: a language for symbolic-statistical modeling. In: *International Joint Conference on Artificial Intelligence*, vol. 15, pp. 1330–1339. Citeseer (1997)

16. Tamaddoni-Nezhad, A., Chaleil, R., Kakas, A., Muggleton, S.: Application of abductive ILP to learning metabolic network inhibition from temporal data. *Machine Learning* 64(1), 209–230 (2006)

## Appendix: References Used for Corroboration in Figure 7

1. Alexander, K.N.A.: The invertebrates of living and decaying timber in Britain and Ireland—a provisional annotated checklist. *English Nature Research Reports* 467, 1–142 (2002)
2. Bauer, T.: Prey-capture in a ground-beetle larva. *Animal Behaviour* 30(1), 203–208 (1982)
3. Bell, J.R., Andrew King, R., Bohan, D.A., Symondson, W.O.C.: Spatial co-occurrence networks predict the feeding histories of polyphagous arthropod predators at field scales. *Ecography* 33(1), 64–72 (2010)
4. Berg, K.: The role of detrital subsidies for biological control by generalist predators evaluated by molecular gut content analysis. PhD thesis, Universitäts- und Landesbibliothek Darmstadt (2007)
5. Desender, K., Pollet, M.: Ecological data on *clivina fossor* (coleoptera, carabidae) from a pasture ecosystem. ii. reproduction, biometry, biomass, wing polymorphism and feeding ecology. *Rev. Ecol. Biol. Sol.* 22(2), 233–246 (1985)
6. Dinter, A.: Intraguild predation between erigonid spiders, lacewing larvae and carabids. *Journal of Applied Entomology* 122(1-5), 163–167 (1998)
7. Lattin, J.D.: Bionomics of the nabidae. *Annual Review of Entomology* 34(1), 383–400 (1989)
8. Lindsey, J.: Ecology of *Commanster*, <http://www.commanster.eu/commanster/insects/bugs/spbugs/saldula.saltatoria.html>
9. Pons, X., Lumbierres, B., Albajes, R.: Heteropterans as aphid predators in intermountain alfalfa. *European Journal of Entomology* 106(3), 369–378 (2009)
10. Schaefer, C.W., Panizzi, A.R.: Heteroptera of economic importance. CRC (2000)
11. Sunderland, K.D.: The diet of some predatory arthropods in cereal crops. *Journal of Applied Ecology* 12(2), 507–515 (1975)
12. Sunderland, K.D., Lovei, G.L., Fenlon, J.: Diets and reproductive phenologies of the introduced ground beetles *harpalus-affinis* and *clivina-australasiae* (coleoptera, carabidae) in new-zealand. *Australian Journal of Zoology* 43(1), 39–50 (1995)
13. Toft, S.: The quality of aphids as food for generalist predators: implications for natural control of aphids. *European Journal of Entomology* 102(3), 371 (2005)
14. Turner, B.D.: Predation pressure on the arboreal epiphytic herbivores of larch trees in southern england. *Ecological Entomology* 9(1), 91–100 (1984)
15. Warner, D.J., Allen-Williams, L.J., Warrington, S., Ferguson, A.W., Williams, I.H.: Mapping, characterisation, and comparison of the spatio-temporal distributions of cabbage stem ea beetle (*psylliodes chrysocephala*), carabids, and collembolan in a crop of winter oilseed rape (*brassica napus*). *Entomologia Experimentalis et applicata* 109(3), 225–234 (2003)
16. Weber, D.C., Lundgren, J.G.: Assessing the trophic ecology of the coccinellidae: their roles as predators and as prey. *Biological Control* 51(2), 199–214 (2009)