

Computer inference of network of ecological interactions from sampling data

Wenjun Zhang

Received: 11 December 2005 / Accepted: 17 February 2006 / Published online: 7 September 2006
© Springer Science + Business Media B.V. 2007

Abstract Both direct and indirect ecological interactions may occur in an ecosystem with large numbers of taxa. Traditional food web technique is a popular tool to measure the quality and health of the environment. Much of works must be done before constructing a food web for an ecosystem especially with many taxa. This food web is generally specific for some ecological interactions and fixed for a set of given species. It is therefore not an effective method for dynamic and prompt assessment of environment. Ecological interactions and their interactive intensity may be detected by sampling biological taxa in the field and by detecting various between-taxa distances or similarities. Network may clearly exhibit the complex interactions among biological taxa. Statistic tests on various distance or similarity measures and computer designs are required to infer the network.

We develop an algorithm and software to infer the network of direct or indirect ecological interactions in ecosystem. It is a prompt and effective tool in monitoring and assessment of the environment. A redundant network may be inferred and drawn by computer based on the statistic tests on sampling data or the pathway information given in HTML file. Dominant taxa

may be found in the network. In total of 16 distance and similarity measures, including Euclidean distance, Manhattan distance, Pearson correlation, partial correlation, point correlation, linkage coefficients, Jaccard coefficient etc., are provided to detect taxa pairs with significant parametric or nonparametric similarities, based on randomization tests and ordinary statistic tests. Criteria to use distance and similarity measures are discussed.

Keywords Network · Ecological interactions · Computer inference · Distance and similarity measures · Randomization test

1. Introduction

Ecological interactions refer to these interactions as prey-predator interaction, parasite-parasitoid interaction, symbiosis, inter-specific facilitation, or other direct or indirect interactions between two biological taxa (Cohen *et al.*, 1994; Cardinale *et al.*, 2002). Both direct and indirect interactions may exist in an ecosystem with large numbers of biological taxa. Traditional food web technique is a common method to measure the health of ecological environment. However, the large number of data and references, in-depth one-by-one biological investigations, various experimental methods, and longer time are needed to construct a food web. Moreover, such food web is generally specific for some ecological interactions and

W. Zhang (✉)
Research Institute of Entomology and State Key
Laboratory of Biocontrol, School of Life Sciences,
Zhongshan University, Guangzhou 510275, P.R. China
e-mail: LS71@zsu.edu.cn, zhangwenjun@scientist.com

fixed for a set of given species. It is not an effective method for dynamic and prompt assessment of environment.

Ecological interactions and interactive intensity may be detected by sampling the habitat, and by testing the distances or similarities among biological taxa based on the sampling data. Ordinary statistic methods may be used to test distances or similarities if stricter statistic assumptions on sampling data are met. These assumptions include, for example, are the individuals randomly sampled from the population of interest? Do the two samples tested share the equal population standard deviations or means? Do the values coincide with a normal distribution, or other known distributions? (Manly, 1997). These assumptions are not exactly met in ecological studies, especially in a dynamic ecosystem with many biological taxa. On the other hand, these statistics may not be conducted if various distance or similarity measures are used. Randomization test, always with less statistic restrictions, is an effective nonparametric method for testing statistic significance of ecological data and has been used in ecological studies (Manly, 1997; Qi, 2003c; Qi and Zhang, 2003; Zhang and Schoenly, 2001; Zhang *et al.*, 2002a, c, 2004a). It may reasonably be used to various distance and similarity measures.

In a habitat, there are always dozens, hundreds, or even thousands of biological taxa that interact with each other (Schoenly *et al.*, 1996). In its nature a complex network of direct or indirect interactions between biological taxa exists in the habitat. The biological taxa may be species, families, or even orders, etc., dependent on the research objectives. Assume that 60 biological taxa are found in a habitat and they interact with each other, then there will be a maximum of 1770 interaction pathways in the network. Obviously it is impractical to artificially infer the network. The objective of this study is to develop an algorithm and software to automatically infer and draw network of direct or indirect ecological interactions in ecosystem. Network can be inferred and drawn using computer from the sampling data or the pathway information contained in HTML file. Sixteen distance and similarity measures are provided in this algorithm to detect taxa pairs with significant parametric or nonparametric similarities, based on randomization tests (Manly, 1997; Qi and Zhang, 2003; Zhang and Schoenly, 2001) and ordinary statistic tests (Zhang and Fang, 1982).

2. Inference methods

Two choices are provided to infer network pathways of ecological interactions. The first is to infer the pathways by the algorithm for computing significant between-taxa interactions, the second is simply to include the pathways information in the HTML file.

2.1. Pathways information resided in HTML file

If pathways information is to be included in the HTML file, the following syntax, for instance, should be included inside the HTML tags `<applet>...</applet>`, which is a complete string of a comma-delimited list of all the taxa pairs or habitat-taxa pairs:

```
<param name=lines value="Rice Field-Coleoptera/150, Rice Field-Diptera/150, Rice Field-Odonata/150, Rice Field-Hemiptera/150, Rice Field-Hymenoptera/150, Rice Field-Araneae/150, Rice Field-Ephemeroptera/150, Rice Field-Lepidoptera/150, Rice Field-Cyprida/150, Rice Field-Cyclopoida/150, Rice Field-Arthropleona/150, Rice Field-Orthoptera/150, Rice Field-Acari/150, Rice Field-Mesogastropoda/150, Rice Field-Symphyleona/150, Rice Field-Thysanoptera/150, Rice Field-Blattodea/150, Coleoptera-Cyclopoida/400, Diptera-Hemiptera/400, Diptera-Cyclopoida/400, Hemiptera-Symphyleona/400, Hymenoptera-Acari/400, Arthropleona-Acari/400">
```

In this HTML syntax, "Rice Field" is the name of ecological habitat and it should be linked to every biological taxa. The pathway lengths in the network are 150 pixels and 400 pixels for different kinds of pathways.

2.2. Pathways information produced by algorithm

2.2.1. Distance and similarity measures used

A delimited text file should be prepared for algorithm use. In this data file the first row are sample ID numbers and the first column are names of biological taxa. The values are the numbers of individuals. Distance and similarity measures, including measures for interval values: Euclidean distance, Manhattan distance, Chebyshev distance, correlation coefficient, and angular cosine; measures for nominal values, linkage

coefficient, co-linkage coefficients 1-3; and measures for binary values, point correlation coefficient, quadratic correlation coefficient, angular cosines 1-2, and Jacarrd coefficient are available to be chosen by users (Krebs, 1989; Zhang and Fang, 1982; Qi, 2003c; Qi and Zhang, 2003; Zhang *et al.*, 2004a). The following are distance and similarity measures used in present paper:

Assume that the sampling data for *i*th taxa and *j*th sample is a_{ij} , $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Mahalanobis distances (n_{ij}), Euclidean distances (e_{ij}), Manhattan distances (m_{ij}), and Chebyshov distances (c_{ij}) are as follows

$$n_{ij} = ((a_{i1} \ a_{j1}, a_{i2} \ a_{j2}, \dots, a_{in} \ a_{jn})^T V^{-1} \times (a_{i1} \ a_{j1}, a_{i2} \ a_{j2}, \dots, a_{in} \ a_{jn}))^{1/2}$$

$$e_{ij} = \left(\sum_{k=1}^n (a_{ik} - a_{jk})^2 / n \right)^{1/2}$$

$$m_{ij} = \sum_{k=1}^n |a_{ik} - a_{jk}| / n$$

$$c_{ij} = \max_k |a_{ik} - a_{jk}| \quad i, j = 1, 2, \dots, m$$

where *V* is the covariance matrix for sampling data.

Pearson correlations (co_{ij}), angular cosines (a_{ij}), and partial correlations (p_{ij}) between taxa are as follows

$$co_{ij} = \frac{\sum_{k=1}^n ((a_{ik} - a_{ibar})(a_{jk} - a_{jbar}))}{\left(\sum_{k=1}^n (a_{ik} - a_{ibar})^2 \sum_{k=1}^n (a_{jk} - a_{jbar})^2 \right)^{1/2}}$$

$$a_{ij} = \frac{\sum_{k=1}^n (a_{ik} a_{jk})}{\left(\sum_{k=1}^n a_{ik}^2 \sum_{k=1}^n a_{jk}^2 \right)^{1/2}}$$

$$p_{ij} = co_{ij}^1 / (co_{ii}^1 co_{jj}^1)^{1/2} \quad i, j = 1, 2, \dots, m$$

where co_{ij}^1 is the element in reverse matrix of matrix (co_{ij}), a_{ibar} and a_{jbar} are means of a_{ik} 's and a_{jk} 's.

Linkage coefficients (l_{ij}) and co-linkage coefficients ($l1_{ij}$, $l2_{ij}$, $l3_{ij}$) are as follows

$$l_{ij} = (x^2 / (x^2 + n..))^{1/2}$$

$$l1_{ij} = (x^2 / (n.. \max(p - 1, q - 1)))^{1/2}$$

$$l2_{ij} = (x^2 / (n.. \min(p - 1, q - 1)))^{1/2}$$

$$l3_{ij} = (x^2 / (n.. ((p - 1)(q - 1))^{1/2}))^{1/2}$$

$$i, j = 1, 2, \dots, m$$

where there are *p* nominal values, i.e., t_1, t_2, \dots, t_p , for taxa *i* and *q* nominal values, i.e., r_1, r_2, \dots, r_q , for taxa *j*. Assume that n_{kl} is the number of taxa *i* takes value t_k and taxa *j* takes value r_l , $k = 1, 2, \dots, p$; $l = 1, 2, \dots, q$, and

$$x^2 = n.. \left(\sum_{i=1}^p \sum_{j=1}^q n_{ij}^2 / (n_{i.} n_{.j}) - 1 \right)$$

$$n.. = \sum_{i=1}^p n_{i.}, \quad n_{i.} = \sum_{j=1}^q n_{ij}, \quad n_{.j} = \sum_{i=1}^p n_{ij}$$

Point correlations (po_{ij}), quadratic correlations (q_{ij}), two types of angular cosines ($\cos 1_{ij}$, $\cos 2_{ij}$), and Jaccard coefficients (d_{ij}) are as follows

$$po_{ij} = (ad - bc) / ((a + b)(c + d)(a + c) \times (b + d))^{1/2}$$

$$q_{ij} = \sin((a + d - (b + c)) / (a + b + c + d) \times 3.1415926 / 2)$$

$$\cos 1_{ij} = (a - b) / ((a + b)(a + c))^{1/2}$$

$$\cos 2_{ij} = (a - a - d - d) / ((a + b)(a + c)(b + d) \times (c + d))^{1/2}$$

$$d_{ij} = 1 - (b_i + b_j) / (c_i + c_j + e) \quad i, j = 1, 2, \dots, m$$

where both taxa *i* and taxa *j* take values 0 or 1. *a* is number of both taxa *i* and taxa *j* take value 0, *b* is number of taxa *i* takes 0 and taxa *j* takes 1, *c* is number of taxa *i* takes 1 and taxa *j* takes 0, and *d* is number of both taxa *i* and taxa *j* take value 1. *b_i* is the non zero number present in taxa *i* but not in taxa *j*, *b_j* is the non zero number present in taxa *j* but not in taxa *i*, *c_i* and *c_j* is the non zero number in taxa *i* and taxa *j* respectively, and *e* is non zero number shared by taxa *i* and taxa *j*.

2.2.2. Randomization test procedure

If $\min a_{ij} < 0$, then let $a_{ij} = a_{ij} - \min a_{ij}$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Suppose z_{ij} is the decimal numbers of a_{ij} if sampling data contain the decimal value a_{ij} , and calculate $c_{ij} = 10^{Z_{ij}}$. Let $a_{ij} = a_{ij} \max c_{kl}$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Through these transformations all of the values in sampling data become integers which are equivalent to numbers of individuals.

The randomization test used in present study is based on the idea that, if no difference exists, then the distribution of individuals in taxa i and j will be a result of allocating the mixed taxa values at random into two taxa of size equal to those of the original taxa (Manly, 1997; Zhang and Schoenly, 2001; Zhang *et al.*, 2004a). The randomization test procedure is described as the follows. Assume that the two taxa to be tested are i and j , $i = 1, 2, \dots, m - 1$; $j > i$, which contain $\sum_{k=1}^n a_{ik}$ and $\sum_{k=1}^n a_{jk}$ individuals respectively. The $\sum_{k=1}^n a_{ik} + \sum_{k=1}^n a_{jk}$ individuals of the combined taxa are randomly reallocated into two randomized taxa with $\sum_{k=1}^n a_{ik}$ and $\sum_{k=1}^n a_{jk}$ labeled individuals. Calculate the expected absolute distance (similarity) between the two randomized taxa and compare whether it is not less than the absolute distance (similarity) between the true taxa i and j . Repeat the simulation many times, calculate the number of the expected are not less than the absolute distance (similarity) between the taxa i and j , take the percentage as the p value. The p value is used to make statistical test. The threshold p value for test may be defined as 0.05, 0.01, etc. If the calculated p value is larger than (in the case of similarity, less than) p threshold, then the (direct or indirect, the same in the following) interaction between taxa i and j is statistically significant, i.e., there is an interaction between the two taxa with respect to the distance or similarity measure used.

The taxa pairs with significant interactions will be calculated by the algorithm and used to construct network information. The Pearson correlation and partial correlation are tested using t -test, and the other distance and similarity measures are tested with the randomization test.

The algorithm may be featured with several characteristics, such as: (1) the randomization method is used to conduct statistic test and thus various distance and similarity measures that short of strict test methods can be statistically tested; (2) various types of distance

and similarity measures are provided for different uses.

3. Computer implementation

This software was developed using JDK 1.1.8. It is a network computation tool and consists of several Java classes and an HTML file:

- (1) EcoIntNet.class. This is the main computation class that loads other classes to jointly complete the entire task. The mathematical algorithm mentioned is completely coded in this class. When loading into web browser, an interface is displayed to receive parameter inputs and choices. The parameters and choices include the available types for inferring pathways, the distance or similarity measures, the number of biological taxa, the number of samples, the significance level (e.g., 0.01, means the 99% confidence degree), the name of ecological habitat, and loading sampling data file.
- (2) ParaInput.class. If distance or similarity measure other than Pearson correlation and partial correlation coefficients is chosen, the window for this class will be launched to allow for the input of the number of randomizations (e.g., 1000, or 10000).
- (3) ResultShow.class. This class was developed to display results, i.e., pathways information, and to save these results (Fig. 1). When it is loaded by EcoIntNet.class, a window will be displayed on which the elapsed time and time progress of computation and the final results will be displayed. Results can be saved by clicking a button in this window and another window for saving results will be launched.
- (4) GraphicsFrame.class. The class to produce an applet container for network.
- (5) NetLine.class. A class to find node taxa of the pathway.
- (6) NetNode.class. An intermediate class.
- (7) NetPanel.class. The class to draw network nodes and pathways.
- (8) NetMap.class. The class to produce component layout for the network.
- (9) Hint.class. A window for this class will be launched when the button "Hint" is clicked. The algorithm and references will be displayed in this window.

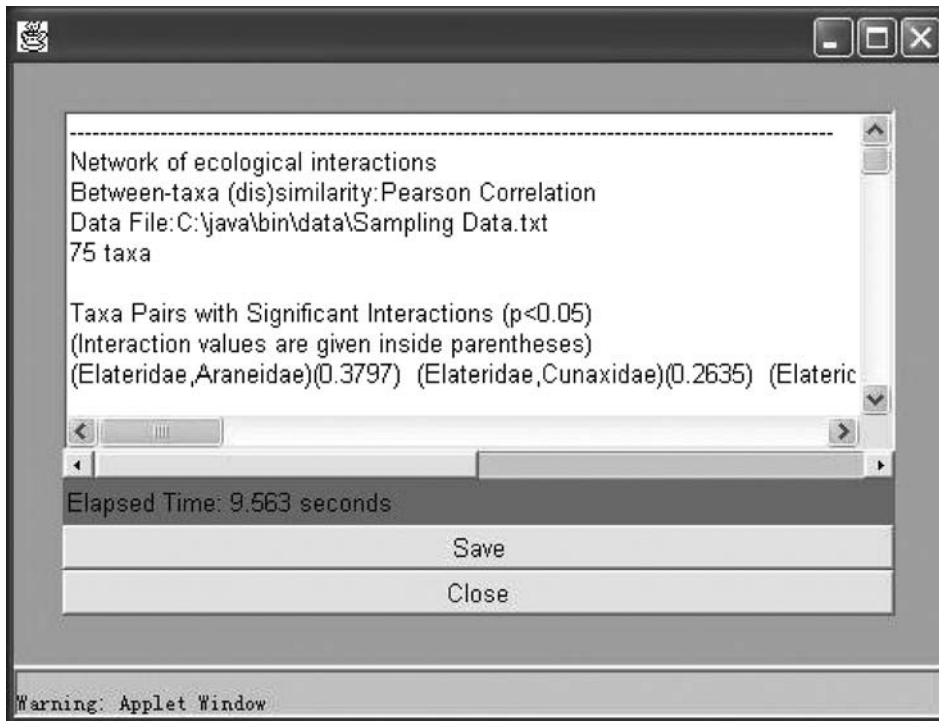


Fig. 1 Result output of ecological interactions

- (10) WarningShow.class. This class is loaded by EcoIntNet.class and ResultShow.class. A window will be launched and the messages of errors that occurred in running program and saving result will be displayed on it.
- (11) EcoIntNet.html. This file is used to load the EcoIntNet.class to Java-enabled web browser and within which the parameter values such as the size of windows and components can be adjusted, and pathways information may be included.

In the network window, all interactions linked to any of the taxa can be found by dragging the box of this taxon with mouse.

This web application is featured with the following advantages (Qi, 2003a-c; Zhang *et al.*, 2002a-c):

- (1) Program codes reside in web server and can be load and run on various java-enabled browsers such as HotJava and IE. Users only need to download Java bytecodes from the Internet and run them on their own machines;
- (2) It is a platform-independent software which can be run on different Java-enabled operational systems, computers and workstations;
- (3) It is easily updated and no version problems and,
- (4) Higher safety

and strong robustness of the software guarantee safe running on client machine.

4. Network of rice invertebrate ecological interactions

In the following example, we use a set of invertebrates data sampled in a rice field (Zhang *et al.*, 2004b; Table 1). There are in total of 75 invertebrate families ($m = 75$) and 60 samples ($n = 60$) in this data file. We use Pearson correlation as the similarity measure and the significance level is specified as 0.05. Some of the results are given as follows:

- (Elateridae, Araneidae)(0.3797) (Elateridae, Cunaxidae)(0.2635) (Elateridae, Coccinellidae)(0.3087) (Elateridae, Veliidae)(0.2762) (Elateridae, Phalacridae)(0.3375) (Elateridae, Empidae)(0.3501) (Elateridae, Anthicidae)(0.2808)
- (Culicidae, Gryllidae)(0.3031) (Culicidae, Araneidae) (0.2921) (Culicidae, Cicadellidae)(0.5054) (Culicidae, Miridae)(0.5467) (Culicidae, Linyphiidae) (0.3467) (Culicidae, Dryinidae)(0.3239) (Culicidae, Carabidae)(0.4095) (Culicidae, Veliidae)(0.3721)

Table 1 Format of sampling data of rice invertebrates. As an example, the data, i.e., numbers of individuals, for 20 samples and 30 families (taxa) are listed in the table

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| Elateridae | 2 | 0 | 4 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 1 | 0 |
| Culicidae | 3 | 4 | 0 | 0 | 1 | 0 | 6 | 4 | 3 | 0 | 5 | 5 | 5 | 1 | 0 | 2 | 4 | 1 | 10 | 0 |
| Coenagrionidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| Aleyrodidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mymaridae | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| Chloropidae | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| Gryllidae | 1 | 0 | 4 | 5 | 2 | 6 | 0 | 2 | 4 | 2 | 4 | 3 | 1 | 0 | 2 | 3 | 5 | 0 | 0 | 2 |
| Araneidae | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Theridiidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Baetidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Hydrophilidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Blattellidae | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| Braconidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| Cicadellidae | 2 | 1 | 0 | 5 | 0 | 2 | 1 | 0 | 5 | 1 | 1 | 1 | 0 | 1 | 3 | 1 | 0 | 1 | 2 | 0 |
| Miridae | 3 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 |
| Tettigoniidae | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| Linyphiidae | 2 | 1 | 1 | 0 | 3 | 0 | 3 | 1 | 0 | 4 | 1 | 1 | 3 | 0 | 3 | 2 | 2 | 1 | 2 | 4 |
| Ceratopogonidae | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Chironomidae | 0 | 2 | 0 | 0 | 0 | 3 | 1 | 1 | 2 | 2 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 4 |
| Encyrtidae | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Cunaxidae | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 |
| Drosophilidae | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Dryinidae | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tetragnathidae | 2 | 0 | 0 | 2 | 1 | 3 | 4 | 3 | 0 | 1 | 4 | 1 | 0 | 1 | 0 | 1 | 1 | 3 | 3 | 2 |
| Dytiscidae | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Carabidae | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Entomobyidae | 2 | 1 | 2 | 2 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 7 | 0 | 0 | 0 | 1 |
| Hydrometridae | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Hydraenidae | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |

(Culicidae, Oribatellidae)(0.4888) (Culicidae, Phalacridae)(0.2549) (Culicidae, Pentatomidae)(0.2586) (Culicidae, Alydidae)(0.6706) (Culicidae, Cyclopiidae)(0.6706)

.....

Total number of eco-interactions = 224
 Total No. eco-interactions/No. taxa = 2.98666

Totally 224 direct or indirect ecological interactions of invertebrate families are found in this rice field. All of the biological taxa are linked to the rice field, but they don't belong to the category of ecological interactions. Values of between-taxa Pearson correlation are included in the parentheses following taxa pairs. Network inferred is shown in Fig. 2. In a sense, dominant families may be visually found through their numbers of pathways linked to other families.

Partial correlation ($p < 0.05$) is also specified as the similarity measure and the network of direct interactions of invertebrate orders is inferred from sampling data (Fig. 3):

(Coleoptera, Hemiptera)(0.7608) (Coleoptera, Orthoptera)(0.32) (Coleoptera, Araneae)(0.3464) (Coleoptera, Ephemeroptera)(0.3243) (Coleoptera, Blattodea)(0.4459) (Diptera, Hemiptera)(0.4409) (Diptera, Hymenoptera)(0.4064) (Diptera, Araneae)(0.6923) (Diptera, Acari)(0.5975) (Diptera, Arthropleona) (0.3423) (Diptera, Lepidoptera)(0.3356) (Diptera, Thysanoptera)(0.3267) (Diptera, Strepsiptera) (0.555) (Diptera, Dermaptera)(0.4257) (Odonata, Isoptera)(0.3268)

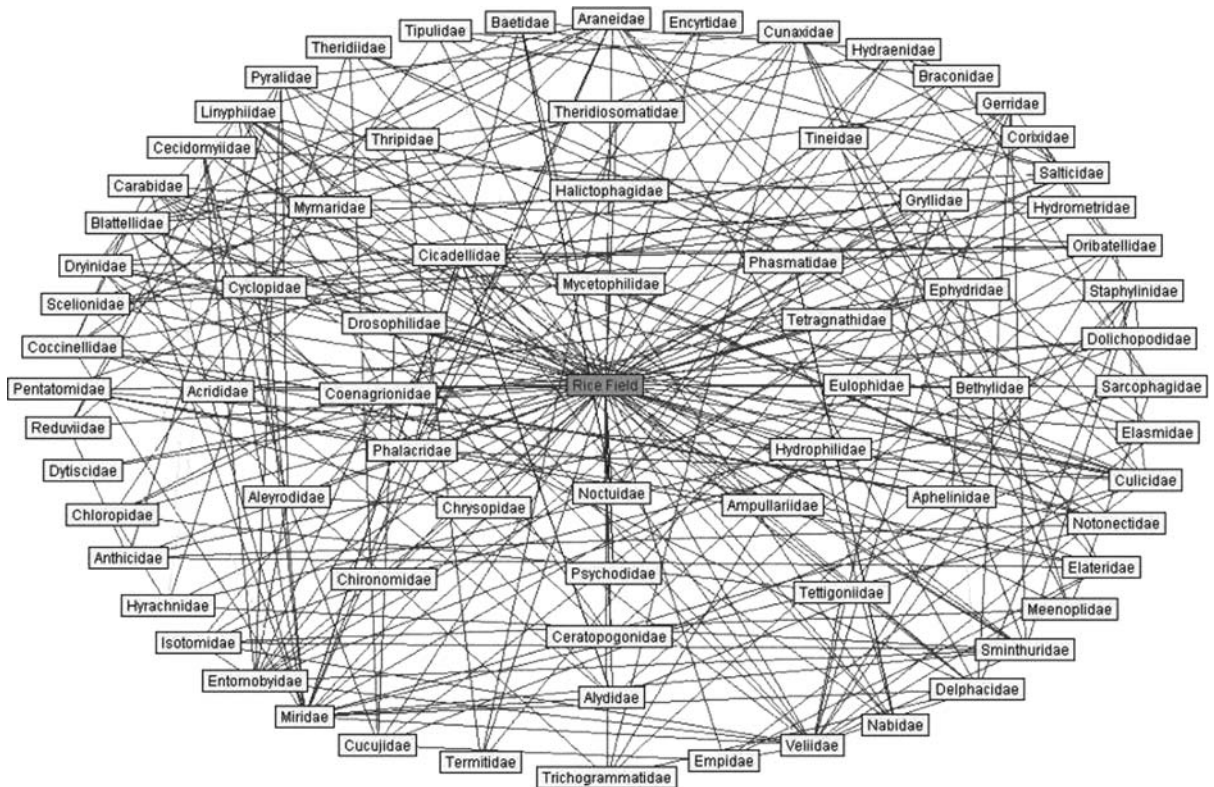


Fig. 2 Network of direct or indirect ecological interactions of invertebrate families in a rice field. Similarity measure: Pearson correlation ($p < 0.05$)

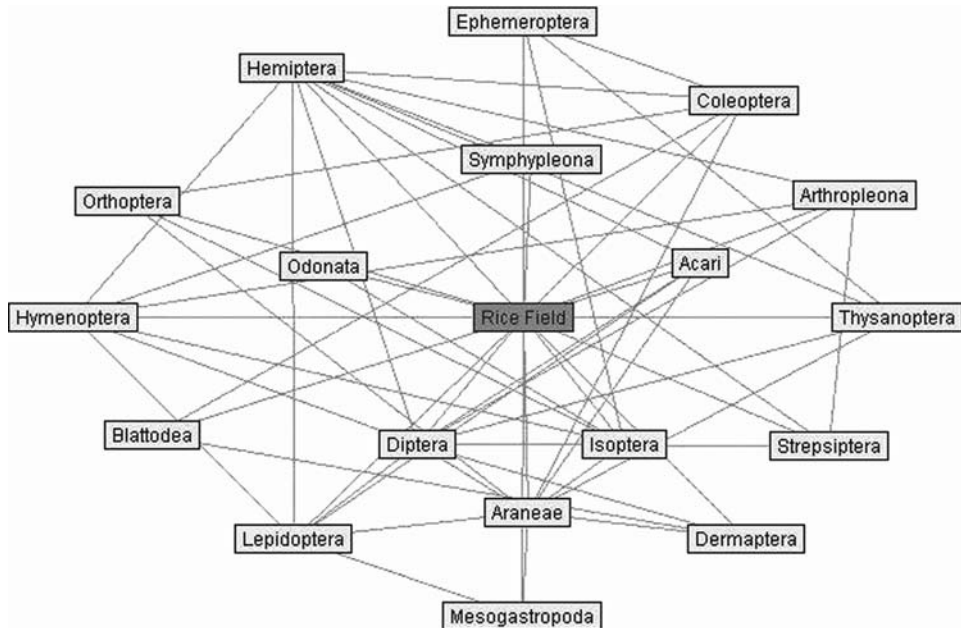


Fig. 3 Network of direct ecological interactions of invertebrate orders in a rice field. Similarity measure: Partial correlation ($p < 0.05$). In total of 40 direction interactions are inferred based on sampling data

(Hemiptera, Hymenoptera)(0.7089) (Hemiptera, Acari)(0.3812) (Hemiptera, Arthropleona) (0.6752) (Hemiptera, Lepidoptera)(0.3401) (Hemiptera, Symphypleona)(0.6955) (Hemiptera, Thysanoptera)(0.2997) (Hemiptera, Strepsiptera) (0.7303) (Hymenoptera, Arthropleona)(0.3869) (Hymenoptera, Lepidoptera)(0.3774) (Hymenoptera, Symphypleona)(0.3313) (Hymenoptera, Isoptera)(0.425) (Orthoptera, Araneae)(0.363) (Orthoptera, Isoptera)(0.653) (Araneae, Acari)(0.4992) (Araneae, Lepidoptera) (0.6573) (Araneae, Mesogastropoda)(0.3046) (Araneae, Isoptera)(0.3705) (Araneae, Thysanoptera)(0.3808) (Araneae, Dermaptera)(0.3154) (Ephemeroptera, Isoptera)(0.3799) (Ephemeroptera, Thysanoptera)(0.3187) (Blattodea, Dermaptera)(0.2975) (Acari, Lepidoptera)(0.3414) (Arthropleona, Strepsiptera)(0.432) (Lepidoptera, Mesogastropoda)(0.3563)

Total number of eco-interactions = 40

Total No. eco-interactions/No. taxa = 2.2222

Dozens of sampling data sets are tested against the software and it is found that various network are yielded if different distance measures, similarity measures and significance levels are selected for these data sets.

5. Discussion and conclusions

If sampling data contain the biological taxa with the same type of values, i.e., all of the taxa have interval values (or binary values, or nominal values), then we may choose the corresponding distance or similarity measures provided in the algorithm.

Computation can also be conducted based on the data with this type of taxa mixture, i.e., some taxa have interval values (or nominal values) and some taxa have binary values. In that case, if we choose the binary measures such as point correlation, Jaccard coefficient, etc., the taxa with interval values (or nominal values) will be transformed into the taxa with binary values, i.e., non-zero values are all transformed into 1's. However, if the distance or similarity measures for interval values (or nominal values), such as Pearson correlation, Euclidean distance, etc. (or linkage coefficient, co-linkage coefficients), then sampling

data will be directly used by the algorithm without any transformation. Transformation need not be conducted when the sampling data contain both the taxa with interval values and the taxa with nominal values.

The randomization test will yield results with little differences between different platforms and machines, because the mechanisms of generations of random seeds are different. Overall these differences in results may be largely eliminated through increasing the number of randomizations.

Distance and similarity measures should be chosen for different uses. The measures such as Euclidean distance, Manhattan distance, and Chebyshev distance, etc., may be used when the between-sample difference of total individual numbers is the focus of consideration. However, the measures such as Pearson correlation, partial correlation, angular cosine, Jaccard coefficient, etc., should be used if the taxa composition and community structure are major considerations. Furthermore, if partial correlation is chosen, then we can infer a network with direct interactions. Generally the choice of distance and similarity measures is dependent upon the types of sampling data and objectives of the researches.

It may be concluded that the network of ecological interactions can be completely inferred and drawn by using the algorithm and software in this paper, once the distance or similarity measure has been specified according to specific requirements. This method does not require in-depth one-by-one biological investigations and wide knowledge on ecological interactions. The network structure may change with the change of ecological environment. It is therefore a prompt and dynamic method for monitoring and assessment of environment. It should be noted that in most of the cases this method would infer a redundant network of ecological interactions, i.e., various direct and indirect ecological interactions are included in the network. Finally, the quality of sampling data is also important in inferring a reasonable network.

Acknowledgment This project was supported by the National Natural Science Foundation of China, through 30170184; Foundation of Scientific Research for Personnel from Abroad, through Ministry of Education of China 2000, and Asian Development Bank, through RETA 5711. I thanks for all participants of rice invertebrates investigation, identification, and data sheet processing in Philippines and China. I also thanks for reviewer's comments and suggestions. Contact author (zhangwenjun@scientist.com) for free software program.

References

- Cardinale, B.J., Palmer, M.A., & Collins, S.L. (2002). Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature*, *415*, 426–429.
- Cohen, J.E., Schoenly, K.G., Heong, K.L., Justo, H.D., Arida, G., Barrion, A.T., & Litsinger, J.A. (1994). A food web approach to evaluating the effect of insecticide spraying on insect pest population dynamics in a Philippine irrigated rice ecosystem. *Journal of Applied Ecology*, *31*, 747–763.
- Manly, B.F.J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology* (Second Edition), London, UK: Chapman & Hall.
- Krebs, C.J. (1989). *Ecological methodology*. New York: Happer Collins Publishers Inc.
- Qi, Y.H., & Zhang, W.J. (2003). CorreDetector: a network sharing software for correlation analysis of information and data. *Journal of the China Society for Scientific and Technical Information*, *22*(Suppl.), 266–268.
- Qi, Y.H. (2003a). A web based computation software and its application in evaluation analysis of books and journals. *Journal of Information*, *22*(1), 43–45.
- Qi, Y.H. (2003b). A web based computation software of fuzzy clustering analysis for webmetrics. *Information Science*, *21*(10), 1069–1071, 1079.
- Qi, Y.H. (2003c). *Habitat Heterogeneity and Pest Diffusion in Different Habitats: Models and Algorithms*. MSc Thesis, China: Zhongshan University.
- Schoenly, K.G., Cohen, J.E., Heong, K.L., Litsinger, J.A., Aquino, G.B., Barrion, A.T., & Arida, G. (1996). Food web dynamics of irrigated rice fields at five elevations in Luzon, Philippines. *Bulletin of Entomological Research*, *86*, 451–466.
- Zhang, W.J., & Qi, Y.H. (2002a). Statistic test and network software on statistic significance of biological diversity and evenness. *Modern Computer*, *145*, 6–9.
- Zhang, W.J., Qi, Y.H., & Schoenly, K.G. (2002b). Functional link artificial neural network and agri-biodiversity analysis. *Biodiversity Science*, *10*(3), 345–350.
- Zhang, W.J., Qi, Y.H., & Schoenly, K. G. (2002c). Randomization tests and computational software on significance of community biodiversity and evenness. *Biodiversity Science*, *10*(4), 431–437.
- Zhang, Y.Q., & Fang, K.T. (1982). *Introduction to multivariate statistics*. Beijing: Science Press.
- Zhang, W.J., & Schoenly, K.G. (2001). A randomization test and software to compare ecological communities. *International Rice Research Notes*, *26*(2), 48–49.
- Zhang, W.J., Qi Y.H., & Wu, H.X. (2004). A generator of web map of biological interactions for rice invertebrates. *International Rice Research Notes*, *29*(2), 45–47.
- Zhang, W.J., Feng, Y.J., & Schoenly, K. G. (2004b). Lumping and correlation analyses of invertebrate taxa in tropical irrigated rice field. *International Rice Research Notes*, *29*(1), 41–43.