

Network Analysis of Cropping Practices and Injury Profiles in Irrigated Rice Agroecosystems

Sith Jaisong

October 15, 2015

Abstract

Here is the abstract.....

Introduction

Every year farmers loss their rice yields around 37 percent because of pests and diseases. injuries

The use of in-field surveys is a useful tool to develop ground-truth databases that allow one identify actual constraints due to pests in an agricultural productions system. These sorts of databases provide an overview of the complex relationships between the crop, its management, pest injuries, yields. Understanding theses relationships may lead to better management, and guide researchers the new research hypotheses .

Several previous studies involved surveys that have been used to identify relationships in an individual production situation (a set of factors that determine agricultural production) and the injury profiles (combination of disease and pest injuries that may occur in a given farmer's field) using nonparametric multivariate analysis such as cluster analysis, correspondence analysis, multiple correspondence analysis. Performing correspondence analysis , they characterized the relationships between categorized levels of variables: actual yield, production situations, and injuries profiles. Their results led to the conclusions that observed injuries profiles were strongly associated with production situations and the level of actual yields.

this is the para three

Here, we use a network approach to infer the relationships species roles in the metacommunity structure of a desert ecosystem in the Atacama Desert, Chile. Starting with the incidence matrix of species across local communities, we built a unipartite species network such that any two species that co-occur, more frequently than expected by chance, in local communities were linked. Using this positive co-occurrence network, we: (i) carried out a modularity analysis and classified species into different topological roles; and (ii) tested for the importance of trophic status, body size, numerical abundance and incidence of the species as the biological attributes that could determine these roles. The result herein this report reported identify modularity as a main component of species co-occurrences networks and, body size and trophic position as chief determinants of the ecological role of species within metacommunities.

Material & Methods

Study sites, sampling and data collection

We conducted the surveys located in the South and South East Asia, Kerala, India (Lat , Long), Indonesia (Lat , Long), Philippines (Lat , Long), Central Plain, Thailand (Lat , Long), and Mekong Delta Vietnam (Lat , Long). These are the important rice growing areas, where use irrigated lowland rice ecosystem. intensive condition, which grow twice per year. We sampled in , under same standardized protocol described in the IRRI publication, "A survey portfolio to characterize yield-reducing factors in rice", was used for data collection (Savary and Castilla, 2009).e conducted the surveys located in the South and South East Asia, Kerala, India (Lat , Long), Indonesia (Lat , Long), Philippines (Lat , Long), Central Plain, Thailand (Lat , Long), and Mekong Delta Vietnam (Lat , Long). These are the important rice growing areas, where use irrigated lowland rice ecosystem. intensive condition, which grow twice per year. We sampled in , under same standardized protocol described in the IRRI publication, "A survey portfolio to characterize yield-reducing factors in rice", was used for data collection.

We collected the injury variables, which are consist of the groups of injuries caused by insect pests, pathogens, and weeds. The injuries were grouped in the field assessment procedure according to their nature: leaf injuries caused by insect pest or pathogens (BLB, BS, LB: proportion of injured leaves) tiller or panicle injuries (SR, SHB, DH: proportion of injured tillers; panicles: SHR, WH: proportion of injured panicles), systemic injuries (viral diseases)

Crop health survey data were collected through surveys comprising 420 farmers' fields from 2010 to 2012 for wet and dry seasons in different production environments across South and South East Asia. The survey protocol described in the IRRI publication, "A survey portfolio to characterize yield-reducing factors in rice", was used for data collection (Savary and Castilla, 2009). The variables collected included patterns of cropping practices, crop growth measurement and crop management status assessments, measurements of levels of injuries caused by pests, and direct measurements of actual yields from crop cuts. The data collected can be classified into three groups: cropping practices, injuries, and actual yield measurements.

Co-occurrence analysis

A total of 420 fields. We designed a statistical approach written in R v. 3.0.1 . The analysis presented in this paper is designed to test for differences in co-occurrence patterns at the community level across ecosystems, identify modules of co-occurring microorganisms within communities, and identify pairwise co-occurrence patterns within modules that are consistent across ecosystems (summarized). We considered co-occurrence to be positive rank correlations (Spearman's correlation) between pairs of microbes within each dataset with the strength of the relationship represented by the correlation coefficient (Figure 1B). Negative correlations (indicative of either competitive interactions or non-overlapping niches between microbes; Faust and Raes, 2012) were also included in

The first step in the analysis is identify outlier sample using absolute hi-

erarchical cluster analysis. After removing the outliers for analysis we would construct a weight networking weight

Co-occurrence networks were produced by applying an association metric or correlation coefficient to the simulated abundance data in a pair-wise manner. Statistically significant aggregation or avoidance was determined by generating a null distribution for each species pair by shuffling the site-abundance of one of the species and re-calculating the association metric. This resampling was performed 1000 times and the resulting distribution was used to generate p-values for observed association metric. P-values were corrected for multiple comparisons using the method of Benjamini and p-values less than $p = 0.05$ were considered to be statistically significant edges in the network. The sparCC program (Friedman and Alm, 2012), which was used for treatment of relative abundance data, uses a similar approach based on matrix permutation and null distribution generation.

Network analysis

All calculations concerning the social network analysis were carried out using the Python module Network. Global parameters describing networks

Global descriptors of the modules were obtained using Cytoscape. The neighborhood of a given node n is the set of its neighbors. The connectivity is the size of its neighborhood. The average number of neighbors indicates the average connectivity of a node in the network. A normalized version of this parameter is the network density. Density ranges between 0 and 1. It shows how densely the network is populated with edges, A network which contains no edges and solely isolated nodes has a density of 0. In contrast, the density of a clique is 1. Another related parameter is the network centralization [43]. Networks whose topologies resemble a star have a centralization close to 1, whereas decentralized networks are characterized by having a centralization close to 0. In undirected networks, the clustering coefficient C_n of a node n is defined as $C_n = 2e_n / (k_n(k_n - 1))$, where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of the network [44,45]. The clustering coefficient of a node is always a number between 0 and 1. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network. Nodes with less than two neighbors are assumed to have a clustering coefficient of 0.

We chose ranked correlations based on absolute abundance data to compare between rather different types of ecosystems to avoid spurious correlations, which may inflate networks and reduce specificity (Friedman and Alm 2012). A recent analysis of modeled metacommunities has shown that co-occurrence networks based on Spearman and Pearson correlation coefficients outperform other metrics such as similarity based metrics (Bray-Curtis, Sorensen) in terms of sensitivity and specificity (Berry and Widder 2014).

The matrix can be viewed as an adjacency matrix of a weighted network. The matrix contains the correlation coefficient between each node (i.e., the variable). Thus, the matrix can be thought of as the population average of the network structure. Because we are looking at several specific links, we control for multiple testing by controlling the False Discovery Rate (FDR method) at 5%. The generated network structure can be visualized through the R package qgraph (?). Only connections that surpass the significance threshold are shown

in the visual representation.

0.0.1 Network topology

To evaluate the topological properties of both the interaction and the co-occurrence network, we used the package `igraph` and `qgraph` in The R environment. Particularly we were interested in properties potentially relevant for community roles and functioning as previously hypothesized in and reference therein in there are :

- Mean degree $\langle k \rangle$: the degree of a node counts the number of edges it has. The mean degree of nodes calculate over all nodes in the network
- Degree distribution: the frequency of node vs. their (increasing) degree.
- Average shortest path length, $\langle l \rangle$: the shortage path between any two nodes is the single path with fewest links between them. Alternative paths are feasible. The average shortest path length is the mean over all shortest oaths between any two nodes in the network.
- Mean clustering coefficients: a cluster of nodes a triangle of nodes. The clustering coefficient calculates the fraction of observed vs possible triangles for each mode. The mean is subsequently determined from all nodes in the network,
- Betweenness centrality:
- Closeness centrality:

Important information about a network can be gained by analyzing its global structure, for example by looking at the relative centrality of different nodes. In a centrality analysis, nodes are ordered in terms of the degree to which they occupy a central place in the network. Global descriptors of the modules were obtained using package `qgraph` in R. The neighborhood of a given node n is the set of its neighbors. The connectivity is the size of its neighborhood. The average number of neighbors indicates the average connectivity of a node in the network. A normalized version of this parameter is the network density. Density ranges between 0 and 1. It shows how densely the network is populated with edges. A network, which contains no edges and solely isolated nodes has a density of 0.

In correlation (undirected) networks, the clustering coefficient is the number of connected pairs between all neighbors of the network. The clustering coefficient of a node is always a number between 0 and 1. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network. Nodes with less than two neighbors are assumed to have a clustering coefficient of 0. We then determined network centralities on the modules obtained from network analysis. Centralities were assessed using `qgraph` package in R. We calculated Degree centrality and Betweenness centrality.

Key factor analysis

Note that the absolute values of the eigenvector derived from R match the values derived from Netdraws computation of eigenvector centrality, while the results from igraph are simply normalized values of the eigenvector. The residual calculation will be used in the key actor analysis discussed below.

One way to identify key actors in a network is to compare relative values of centrality such as eigenvector centrality and betweenness. It is apparent that many measures of centrality are correlated. If we assume a linear relationship between eigenvector centrality and betweenness and regress betweenness on eigenvector centrality, the residuals can be used to identify key players (Conway, 2009). A vertex or individual with higher levels of betweenness and lower EV centrality may be a critical gatekeeper or an individual that is central to the functioning of the network. Someone with lower levels of betweenness and higher EV centrality may have unique access to other individuals that are key to the functioning of the network.

The limitation of each measure are different assumptions and different patterns. In this study, we evaluated four association methods, Pearson, Spearman rank correlation, Kendall, Biweight midcorrelation to be suitable for the properties of the survey data (*i.e.* type of variables, pattern of distribution, normality) based on their assumptions. The other evaluation is to examine which methods can associate those variables that are known to present the biological relationships.

Next, we inferred correlation network from surveys comprising five countries (India, Indonesia, Philippines, Thailand, and Vietnam), 420 lowland farmers' fields. We determine the correlation patterns among the incidence of injuries caused by animal pests and diseases and the cropping practices, potentially indicative of their occurrence relations. We then constructed the network from these pairwise correlations.

Network analysis

Important information about a network can be gained by analyzing its global structure, for example by looking at the relative centrality of different nodes. In a centrality analysis, nodes are ordered in terms of the degree to which they occupy a central place in the network. Global descriptors of the modules were obtained using package qgraph in R. The neighborhood of a given node n is the set of its neighbors. The connectivity is the size of its neighborhood. The average number of neighbors indicates the average connectivity of a node in the network. A normalized version of this parameter is the network density. Density ranges between 0 and 1. It shows how densely the network is populated with edges. A network, which contains no edges and solely isolated nodes has a density of 0.

In correlation (undirected) networks, the clustering coefficient is the number of connected pairs between all neighbors of the network. The clustering coefficient of a node is always a number between 0 and 1. The network clustering coefficient is the average of the clustering coefficients for all nodes in the network. Nodes with less than two neighbors are assumed to have a clustering coefficient of 0. We then determined network centralities on the modules obtained from network analysis. Centralities were assessed using qgraph package

in R. We calculated Degree centrality and Betweenness centrality.

Results

Discussion

References

Savary, S. and Castilla, N. (2009). A survey portfolio to characterize yield-reducing factors in rice. *IRRI Discussion Paper No 18*, page 32.