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

First Author [1,*], **Co-Author** [2] **and Co-Author** [2]

[1] *Laboratory X, Institute X, Department X, Organization X, City X , State XX (only USA, Canada and Australia), Country X*

[2] *Laboratory X, Institute X, Department X, Organization X, City X , State XX (only USA, Canada and Australia), Country X*

Correspondence*:
Corresponding Author
Laboratory X, Institute X, Department X, Organization X, Street X, City X , State XX (only USA, Canada and Australia), Zip Code, X Country X, email@uni.edu

ABSTRACT

For full guidelines regarding your manuscript please refer to Author Guidelines or **Table 1** for a summary according to article type.

Keywords: Text Text Text Text Text Text Text Text

1 INTRODUCTION

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 (Mew et al., 2004; Savary et al., 2006).

Several previous studies (**??**Savary et al., 2005; Dong et al., 2010; **?**) 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.

The components of production situation and injury profiles are biologically related. For example, the excess amount of fertilizers applied in the rice files increases the susceptibility of rice to blast and directed seeded flooded rice fields with high seedling rate is the favorable condition for sheath blight (Ou, 1985). The relationships will be more complex when the number of their components increased. A way to systemically model and intuitively interpret such relationships is the depiction as a graph or network. This approach has been widely used and proven very useful in biological studies (Moslonka-Lefebvre et al., 2011). Networks typically consist of nodes, usually representing components, while links between the nodes depict their interactions (Proulx et al., 2005). A correlation network is a type of network in which two nodes are connected if their respective correlation lies above a certain threshold. The construction of this network is obtained from pairwise correlation methods (Toubiana et al., 2013). By using appropriate correlation measure, correlation networks can capture biologically meaningful relationships, and discover valuable information in crop health surveys.

The main objective of this study is to apply network theory to the rice crop health survey data with the proposed methods for network construction. Selecting the suitable measure is important because the method should be able to capture the relationships with true concordance often determined the type and amount of knowledge we can gain from survey data, moreover it will affect the topological structure of network (the patterns of pairwise relationships between variables).

2 MATERIAL & METHODS

In the first step of the methods, we compute partial correlation coefficients between each edge and a behavioral measure of interest independently, by taking other behavioral measures as covariates. Either Pearson or Spearman correlation coefficients can be employed, though we employed Spearman rank correlation coefficient since the distribution of the brain connectivity data is unknown and often distribution not satisfy the normality condition. (copy from the)

Next, we inferred association 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.

Survey datasets

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.

Evaluation

One: Data exploration

There are three main properties to be determined before deciding the appropriate correlation measure for use in constructing the network.

2.0.0.1 Check data distribution

This test can be achieved by significance test and visual methods. Each variable in survey dataset was tested normality using the Shapiro-Wilk test (**?**). The Shapiro-Wilk test is based on the correlation between the data and the corresponding normal scores.

H_0: sample distribution is normal.

H_a: sample distribution is not normal.

Thus, if the p-value is less than the chosen alpha level, the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population. In other words, the data are not normal. On the contrary, if the p-value is greater than the chosen alpha level, then the null hypothesis that the data came from a normally distributed sample cannot be rejected. However, for small sample sizes, normality tests have little power to reject the null hypothesis, so a QQ (quantile–quantile plot) plot and the frequency distribution (histogram) are required for verification, in addition, to check normality visually.

The `R` function for Shapiro-Wilk Normality test is `shapiro.test` (package stats), whihc is (**?**).

Network Construction

Correlation network construction

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 (Epskamp et al., 2012). Only connections that surpass the significance threshold are shown in the visual representation.

Evaluating Network properties

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 m theres are :

- Mean degree ¡k¿: 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. item Average shortest path length,l.: 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 ¡c¿: 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 ¡CB¿
- Closeness centrality ¡CC¿

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.

Table 1. Maximum size of the Manuscript

	Abstract max. legth (incl. spaces)	Figures or tables	Manuscript max. length
Clinical Case Study Clinical Trial Hypothesis and Theory Methods Original Research Review Technology Report	2000 characters	15	12000 words
Focused Review	2000 characters	5	5000 words
CPC	1250 characters	6	2500 words
Perspective Mini Review	1250 characters	2	3000 words
Data Report	None	2	3000 words
Classification	1250 characters	10	2000 words
Editorial	None	None	1000 words
Frontiers Commentary General Commentary Book review	None	1	1000 words
Opinion Specialty Grand Challenge Field Grand Challenge	None	1	2000 words

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.

Please note that large tables covering several pages cannot be included in the final PDF for formatting reasons. These tables will be published as supplementary material on the online article abstract page at the time of acceptance. The author will notified during the typesetting of the final article if this is the case. A link in the final PDF will direct to the online material.

2.1 Original Research Articles, Clinical Trial Articles, and Technology Reports

For Original Research Articles, Clinical Trial Articles, and Technology Reports the section headings should be those appropriate for your field and the research itself. It is recommended to organize your manuscript in the following sections or their equivalents for your field:

Table 2. Resolution Requirements for the figures

Image Type	Description	Format	Color Mode	Resolution
Line Art	An image composed of lines and text, which does not contain tonal or shaded areas.	TIFF, JPEG	RGB, Bitmap	900 - 1200 dpi
Halftone	A continuous tone photograph, which contains no text.	TIFF, EPS, JPEG	RGB, Grayscale	300 dpi
Combination	Image contains halftone + text or line art elements.	TIFF, JPEG	RGB,Grayscale	600 - 900 dpi

- Introduction: Succinct, with no subheadings.
- Materials and Methods: This section may be divided by subheadings. This section should contain sufficient detail so that when read in conjunction with cited references, all procedures can be repeated.
- Results: This section may be divided by subheadings. Footnotes should not be used and have to be transferred into the main text.
- Discussion: This section may be divided by subheadings. Discussions should cover the key findings of the study: discuss any prior art related to the subject so to place the novelty of the discovery in the appropriate context; discuss the potential short-comings and limitations on their interpretations; discuss their integration into the current understanding of the problem and how this advances the current views; speculate on the future direction of the research and freely postulate theories that could be tested in the future.

Please note that the Material and Methods section can be placed in any of the following ways: before Results, before Discussion or after Discussion.

2.2 Clinical Case Studies

For Clinical Case Studies the following sections are mandatory:

- Introduction: Include symptoms at presentation, physical exams and lab results.
- Background: This section may be divided by subheadings. Include history and review of similar cases.
- Results: This section may be divided by subheadings. Include diagnosis and treatment.
- Concluding Remarks

3 RESULTS

3.1 Figures

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript. Figures will then be automatically embedded at the bottom of the submitted manuscript. Kindly ensure that each table and figure is mentioned in the text and in numerical order. Permission must be obtained for use of copyrighted material from other sources (including the web). Please note that it is compulsory to follow figure instructions. Figures which are not according to the guidelines will cause substantial delay during the production process.

Table 2 shows the resolution requirements for the figures. The figures must be legible:

1. The smallest visible text is no less than 8 points in height, when viewed at actual size.
2. Solid lines are not broken up.

3. Image areas are not pixelated or stair stepped.
4. Text is legible and of high quality.
5. Any lines in the graphic are no smaller than 2 points width.
6. The actual size of the figure must be of at least 8.5 cm.

3.2 Nomenclature

- The use of abbreviations should be kept to a minimum. Non-standard abbreviations should be avoided unless they appear at least four times, and defined upon first use in the main text. Consider also giving a list of non-standard abbreviations at the end, immediately before the Acknowledgments.
- Gene symbols should be italicized; protein products are not italicized.
- Chemical compounds and biomolecules should be referred to using systematic nomenclature, preferably using the recommendations by IUPAC.
- We encourage the use of Standard International Units in all manuscripts.
- To take part in the Resource Identification Initiative, please cite antibodies, genetically modified organisms, software tools, data, databases and services using the corresponding catalog number and RRID in your current manuscript. For more information about the project and for steps on how to search for an RRID, please click here.

$$\sum x + y = Z \tag{1}$$

4 DISCUSSION

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Additional Requirements:

4.1 Corrections

If you need to communicate important changes to a published article please submit a General Commentary. Submit the article with the title Corrigendum: Original Title of Article.

4.2 Commentaries on Articles

At the beginning of your Commentary, please provide the citation of the article commented on. Rebuttals may be submitted in response to Commentaries; our limit in place is one commentary and one response. Rebuttals should also be submitted as General Commentary articles.

4.3 Human Search and Animal Research

All experiments on live vertebrates or higher invertebrates must be performed in accordance with relevant institutional and national guidelines and regulations. In the manuscript, authors must identify the committee approving the experiments and must confirm that all experiments conform to the relevant regulatory standards. For manuscripts reporting experiments on human subjects, authors must identify the committee approving the experiments and must also include a statement confirming that informed consent was obtained from all subjects. In Original Research Articles and Clinical Trial Articles these statements should appear in the Materials and Methods section.

4.4 Clinical Trial Registration

Clinical trials should be registered in a public trials registry in order to become the object of a publication at Frontiers. Trials must be registered at or before the start of patient enrollment. A clinical trial is defined as ”any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes.”(www.who.int/ictrp/en). A list of acceptable registries can be found at www.who.int/ictrp/en and www.icmje.org.

4.5 Inclusion of Proteomics Data

Authors should provide relevant information relating to how the peptide/protein matches were undertaken, including methods used to process and analyze data, false discovery rates (FDR) for large-scale studies and threshold or cut-off rates for peptide and protein matches. Further information could include software used, mass spectrometer type, sequence database and version, number of sequences in database, processing methods, mass tolerances used for matching, variable/fixed modifications, allowable missed cleavages, etc.

Authors should provide as supplementary material information used to identify proteins and/or peptides. This should include information such as accession numbers, observed mass (m/z), charge, delta mass, matched mass, peptide/protein scores, peptide modification, miscleavages, peptide sequence, match rank, matched species (for cross species matching), number of peptide matches, ambiguous protein/peptide matches should be indicated, etc. For quantitative proteomics analyses authors should provide information to justify the statistical significance including biological replicates, statistical methods, estimates of uncertainty and the methods used for calculating error.

For peptide matches with biologically relevant post-translational modifications (PTM) and for any protein match that has occurred using a single mass spectrum, authors should include this information as raw data, annotated spectra or submit data to an online repository (recommended option). Authors are encouraged to submit raw or matched data and 2-DE images to public proteomics repositories. Submission codes and/or links to data should be provided within the manuscript.

4.6 Data Sharing

Frontiers supports the policy of data sharing, and authors are advised to make freely available any materials and information described in their article, and any data relevant to the article (while not compromising confidentiality in the context of human-subject research) that may be reasonably requested by others for the purpose of academic and non-commercial research. In regards to deposition of data and data sharing through databases, Frontiers urges authors to comply with the current best practices within their discipline.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

The statement about the authors and contributors can be up to several sentences long, describing the tasks of individual authors referred to by their initials and should be included at the end of the manuscript before the References section.

ACKNOWLEDGMENTS

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text. Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.

Funding: Text Text Text Text Text Text Text Text.

SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder

Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.

REFERENCES

Dong, K., Chen, B., Li, Z., Dong, Y., and Wang, H. (2010). A characterization of rice pests and quantification of yield losses in the japonica rice zone of yunnan, china. *Crop Protection* 29, 603–611

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., and Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* 48, 1–18

Mew, T. W., Leung, H., Savary, S., Vera Cruz, C. M., and Leach, J. E. (2004). Looking ahead in rice disease research and management. *Critical Reviews in Plant Sciences* 23, 103–127

Moslonka-Lefebvre, M., Finley, A., Dorigatti, I., Dehnen-Schmutz, K., Harwood, T., Jeger, M. J., et al. (2011). Networks in plant epidemiology: from genes to landscapes, countries, and continents. *Phytopathology* 101, 392–403

Ou, S. H. (1985). *Rice diseases* (International Rice Research Institute (IRRI))

Proulx, S., Promislow, D., and Phillips, P. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution* 20, 345–353

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

Savary, S., Castilla, N. P., Elazegui, F., and Teng, P. S. (2005). Multiple effects of two drivers of agricultural change, labour shortage and water scarcity, on rice pest profiles in tropical asia. *Field Crops Research* 91, 263–271

Savary, S., Teng, P. S., Willocquet, L., and Nutter, F. W. (2006). Quantification and modeling of crop losses: a review of purposes. *Annual Review of Phytopathology* 44, 89–112

Toubiana, D., Fernie, A. R., Nikoloski, Z., and Fait, A. (2013). Network analysis: tackling complex data to study plant metabolism. *Trends in Biotechnology* 31, 29–36

FIGURES

Figure 1. Enter the caption for your figure here. Repeat as necessary for each of your figures