NETWORK ANALYSIS OF SURVEY DATA FOR CHARACTERIZATION OF YIELD REDUCING FACTORS OF TROPICAL RICE ECOSYSTEMS IN SOUTH AND SOUTHEAST ASIA

SITH JAISONG

SUBMITTED TO THE FACULTY OF GRADUATE SCHOOL
UNIVERSITY OF THE PHILIPPINES LOS BAÑOS
IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE
DEGREE OF

DOCTOR OF PHILPSOPHY
(Plant Pathology)

JUNE 2016

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE **PAGE**

LIST OF TABLES

ABSTRACT

JAISONG, SITH. University of Philippines Los Baños, June 2016. **Network analysis of survey data for characterization of yield reducing factors of tropical rice ecosystem in South and Southeast Asia**

Major Professor: Dr. Ireneo B. Pangga
Dr. Adam H. Sparks

Rice injuries constrain yield production and may play a central role in global rice productivity. Detailed on-farm surveys conducted in five production environments (Central Plain; Thailand, Odisha; India, Red River Delta; Vietnam, Tamil Nadu; India, and West Java; Indonesia) are useful sources of data to help understand yield constraints in farmers' fields. These data also help us understand the interactions and importance of pests and the complex relationships within agroecosystems. Network analysis is a promising tool frequently used to describe the pairwise relationships of a large number of variables. Network analysis of the co-occurrence patterns of pest and disease incidence could offer new insight into pest management. In this study, Spearman's correlation was found to be the most suitable measure because of its robustness to noise, outliers, and ability to accurately predict the interactions. Resulting networks network links rice injuries (nodes) connected with co-occurrence relationships (edges). Based on node-wise properties, network can suggest the central injuries, which possibly are key indicators to monitor. The variations in co-occurrence activity of rice injuries in seasons, and yield levels were examined by using differential networks. They showed contrasting linkages among the injuries and identify injuries significantly responding to the conditions.

CHAPTER I

USING NETWORK ANALYSIS TO EXAMINE CO-OCCURRENCE PATTERNS OF ANIMAL INJURIES AND DISEASES IN FARMERS' FILEDS IN DIFFERENT PRODUCTION ENVIRONMENTS ACROSS SOUTH AND SOUTH EAST ASIA

INTRODUCTION

Agricultural crop plants are frequently injured, or infected by more than one species of pests and pathogens at the same time. Many of these injuries may affect yields. The combinations of injuries usually do not occur independently but as sets so-called "injury profiles", and there are strong statistical associations between these injury profiles and patterns of cropping practices (Savary et al., 2006). Co-occurrence patterns of injuries can provide important insight into these injury profiles, which possibly present co-occurring relationships among injuries. Uncovering these patterns is important to implications in plant disease epidemiology and management. It could be a difficult task since complex patterns of injury profiles are related to environmental conditions, cultural practices, and geography (Willocquet et al., 2008).

To address this issue, I used in-field surveys as a tool to develop ground-truth databases that can be used to identify the major yield reducing pests in irrigated lowland rice ecosystems. These sorts of databases provide an overview of the complex relationships between crop, cropping practices, pest injuries, and yields. Several studies Savary et al. (2000a,b); Dong et al. (2010) and Reddy et al. (2011) analyzed survey data in order to characterize injury profiles, production situations (a set of factors including cultural practices, weather condition, socioeconomics, *etc*.), and their relationships. These studies applied parametric and nonparametric multivariate analysis such as cluster analysis, correspondence analysis, or multiple correspondence analysis to characterize injury profiles in relation to production situations, and quantify yield losses due to the pests. In brief, their conclusions showed the strong link between patterns of injury profiles and yield levels and the relative importance of rice pests in specific locations, yield levels were associated with very distinct patterns of injury profiles.

Network has been widely used as a powerful tool in biology, mathematics, social science and computer science, to explore the interactions between entities or parameters (Kasari et al., 2011; Proulx et al., 2005; Barberán et al., 2012) and understand the behavior and function of the network system, even insight into a vast array of complex and previously poorly understood phenomena (Newman, 2003). Network analysis is the mapping and measuring of relationships and flows (edges) between entities (nodes), according to the mathematical, statistical and structural properties. For nodes, they are the fundamental units of a network, and for edges, they are the lines connecting the interacting nodes. According to Newman (2003) the theory of network primarily includes: finding out the statistical properties to suggest appropriate ways to measure the structure properties, creating network models, and understanding the meaning of these properties (network topologies). Network topologies can be used to determine the importance of entities of networks (*e.g.*, degree, betweenness, clustering coefficient), possibly identify the important entities within networks such as keystone species within an ecosystem (Lu et al., 2013; Borthagaray et al., 2014). Network analysis facilitates to explore and identify the co-occurring patterns of large and complex data that may be more difficult to detect or analyzed using traditional normalization methods. Therefore, in principle, network analysis could also be used in the crop health survey data to reflect the relationships between variables observed.

Co-occurrence patterns are ubiquitous and particularly important in understanding community structure, offering new insights into potential interaction in networks. Co-occurrence analysis and network theory have recently been used to reveal the patterns of co-occurrence between microorganisms in the complex environments ranging from human gut to ocean and soils (Faust et al., 2012; Ma et al., 2016). Recent reviews of network based approaches revealed that these tools have demonstrated previously unseen co-occurrence patterns, such as strong non-random association, topology based analysis of large networks has been proven powerful for studying the characteristics of co-occurrence pattern of the communities in ecological community (Williams et al., 2014; Barberán et al., 2012), or key actors in social networks (Crowston and Howison, 2006).

Till now, network analysis has not been applied to exploring co-occurrence patterns between rice injuries in farmers' fields based on crop health survey data, which untangles the structure of complex data among the various parameters of environment. With the analysis of network, it makes sense of co-occurring correlations of rice injuries. Moreover, the co-occurrence results the associations between injuries proposed by network analysis might help to characterize injury syndromes (the combinations of injuries) in these survey data. Therefore, this work provides a new method to analyze survey data, and directly visualize the correlations among co-occurring injuries under farmers' field levels.

MATERIALS AND METHODS

Crop health survey data

Crop health survey data were collected through 423 farmers’ fields over two production seasons, and three consecutive years (2013 to 2015) in the five main rice production environments, Central Plain (14^o 23’-14^o 53’N, 100^o 1’ - 100 textsuper-scripto 12’E); Thailand, Odisha (20^o 6’- 20^o 27’N, 85^o 31’ - 85^o 58’E);India, Red River Delta (20^o 28’-20^o 49’E, 106^o 13’ - 106^o 23’E); Vietnam, Tamil Nadu (10^o 54’-11^o 5’E, 79^o 19’ - 79^o 36’E); India, and West Java (6^o 9’- 6^o 19’S, 107^o 0’ - 107^o 32’E); Indonesia. The number of fields survey were summarized in Table. The survey procedure and the collection of field data were described in the previous chapter.

Network construction

I designed a statistical approach written in R version. 3.0.1 (R Core Team, 2015). All scripts necessary to replicate this analysis are included in the appendix. The mythology presented in this chapter was adopted from Williams et al. (2014) for constructing network models of co-occurrence patterns of rice injuries at different levels across cropping seasons (wet and dry season), and production environments (Central Plain; Thailand (CP), Odisha; India (OR), Red River Delta; Vietnam (RR), Tamil Nadu; India (TM), and West Java; Indonesia (WJ)).

Network construction was illustrated in Figure.I-1. The network is constructed in three steps. In step 1, an incidence matrix is obtained from the data set of rice injury occurrences. An incidence matrix lists each injury in the data set by row (farmer’s

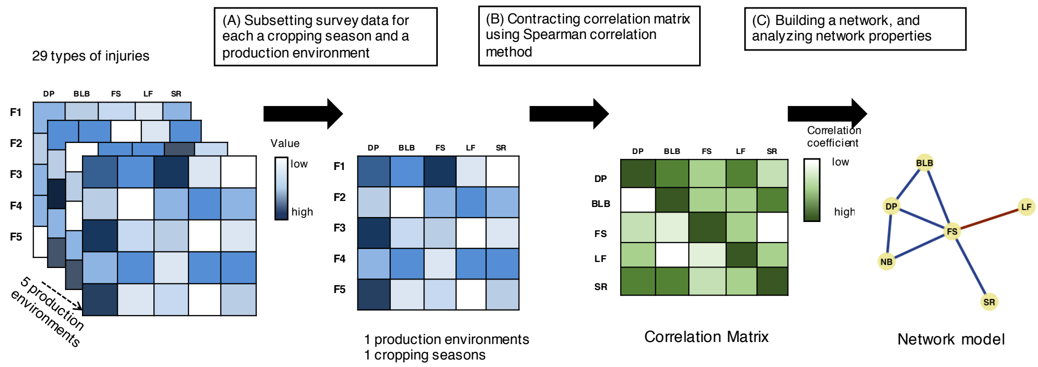


Figure I-1: Workflow used for constructing a network that represents the co-occurrence of rice injuries based on survey data. A) subsetting survey by season, and production environment, B) constructing correlation matrix using Spearman's correlation method, and C) building network models.

fields) with the columns corresponding to the injuries. Thus the incidence matrix shows the occurrence of injuries by rice fields. In step 2 an adjacency matrix is computed from the incidence matrix using Spearman's correlation method. Correlation matrix is pre X, denoted XT, has the same elements as X only the rows and columns are interchanged. In step 3, the network graph is drawn by connecting regions that have a non-zero entry in the adjacency matrix.

The co-occurrence network was inferred based on adjacency matrix, which is Spearman correlation matrix constructed with `R` function `cor.test` with parameter method 'Spearman' (package stats) was used for calculate Spearman's correlation coefficient (ρ) (R Core Team, 2015).

The adjacency matrix A of this network formally expresses injury occurrence, and is written in $A = [C_{ij}]$, which is

$$C_{ij} = \begin{cases} C_{ij} & \text{if } p\text{-value } < 0.05 \\ \\ 0 & \text{otherwise} \end{cases} \tag{I.1}$$

where C is rank correlations coefficient (ρ from the Spearman's correlation at p-value < 0.05) between pairs of injures.

$$A = \begin{pmatrix} 0 & C_{ij} \\ \\ C_{ji} & 0 \end{pmatrix} \tag{I.2}$$

where A is the adjacency (correlation) matrix, in which the rows and column are injuries. If I ordered first by injury ($1 \ldots n$) and second by grid cells ($j+1 \ldots n+j$), producing a square matrix with $i + j$ rows and $i + j$ columns.

From adjacency matrix, the networks were visualized with **igraph** package (Csardi and Nepusz, 2006) using indirected network and the Fruchterman–Reingold layout (Fruchterman and Reingold, 1991). Nodes in this network represent rice injuries and edges that connect these nodes represent correlations between injuries.

Topological feature analysis

I calculated the topological features of each network using **igraph** package. To describe the topology of the resulted networks, a set of measures (node degree, local clustering coefficient, and betweenness) were calculated (Newman, 2006). Node degree is measured by the number of the edges (connections) of a node has. Betweenness of a node is defined by the number of of shortest paths going through a node, and the local clustering coefficients of a node is the ratio of existing edges connecting a node's neighbors to each other to the maximum possible number of such edges. The network clustering coefficient measures the degree to which nodes of the network tend to cluster together and is a measure of the connectedness of the network and is indicative of the degree of relationships in the network.

Nodes were further classified by ranking all nodes according to three node features, partitioning this ranked list into three equally value of each node property. A node with high rank value in top third proportion of node degree, and betweenness is recognized as an indicator in co-occurrence network of rice injuries.

Community detection

Modularity reflects the degree to which a network is organized into a modular or community structure. Modules refer to a set of nodes with denser links among them but sparser links with the rest of the network (Newman, 2006). Detection and characterization of modular structure in rice injury co-occurrence can help us to identify groups of injuries that closely related and often occur together under same situation. Several optimization algorithms are currently available, each with different advantages (Brandes et al., 2008). Based on the identified community structure, nodes can be grouped in terms of their roles in maintaining intra or inter-module connectivity. In this chapter, the networks were detected community structures by maximizing the modularity measure over all possible partitions by using `cluster_optimal` function of **igraph** package. Injury nodes in the same group will be call as a syndrome, which is the combination of injuries that most likely to be observed together.

RESULTS

Prevalence of injuries across sites and seasons

Survey data were collected from farmers' fields in five production environments (Central Plain; Thailand (CP), Odisha; India (OD), Red River Delta; Vietnam (RR), Tamil Nadu; India (TM), and West Java; Indonesia (WJ)) across South and Southeast Asia, and recorded 29 injuries caused by animal pests and pathogens. The survey data used in this chapter were same as data analyzed in the previous chapter, which are summarized in Table **??** and Table.**??** Prevalences of injuries a across production environments and seasons were shown in Figure.I-2 to I-4.

The injuries caused by animal pests observed, and recorded during the survey period were deadheart (DH), panicle mite injury (PM), leaffolder (LF), rice hispa injury (RH), whorl maggot injury (WM), whitehead(WH), rat injury (RT) rice bug injury (RB) silver shoot (SS) rice thrip injury (RTH) rice leaf miner injury (LM). These injuries were not observed at all survey fields, cropping seasons, or production environments. DH, PM, LF, RH, WM, WH could be observed at all season and production environment. However, they had different levels of prevalence. For example, PM could be observed higher in RR than other production environments, and RT presented at all location too, but heavily at WJ. Some injuries were not reported in production environments. SS, RTH, and LM were not presented in RR, OD, and TM, respectively. RB were reported heavily in WJ, but not in other production environments.

Rice diseases recorded were bacterial leaf blight (BLB), bacterial leaf streak (BLS), brown spot (BS), leaf blast (LB), narrow brown spot (NBS), read stripe (RS), sheath blight (SHB), sheath rot (SR), false smut (FS), stem rot (SR). Diseases observed in this study were commonly found at all locations, but there were some diseases that could not be observed. DP seem to appear at all location, except in Odisha, this disease tended to occur in wet season, more than dry season in CP and RR. Conversely, in TM and WJ, dirty panicle prevailed higher in dry season than wet season. FS presented at all location. They are high prevalence especially in OR, and TM. BLS, LS were not reported in OD and TM. The diseases observed at all location were BS, NB, LB, SHB, SHR with different degree of prevalence. BS prevailed highly in CP, LB had high prevalence in OD, and SHR and NBS highly occurred in CP and WJ. RS were found heavily in CP, but a few in WJ and not reported in OD, RR, TM.

Systemic injuries in this survey include hopperburn (HB) caused by brown planthoppers, and white backed planthoppers, bugburn (BB) caused by rice black bugs, and three viral diseases; rice grassy stunt (GS), ragged stunt (RGS), and tungro (RTG). HB could be commonly found at all production environments. A few locations in CP, RR and WJ were observed BB. GS were reported in RR and WJ. RGS were not reported in OR and RR, and RTG were reported only in WJ.

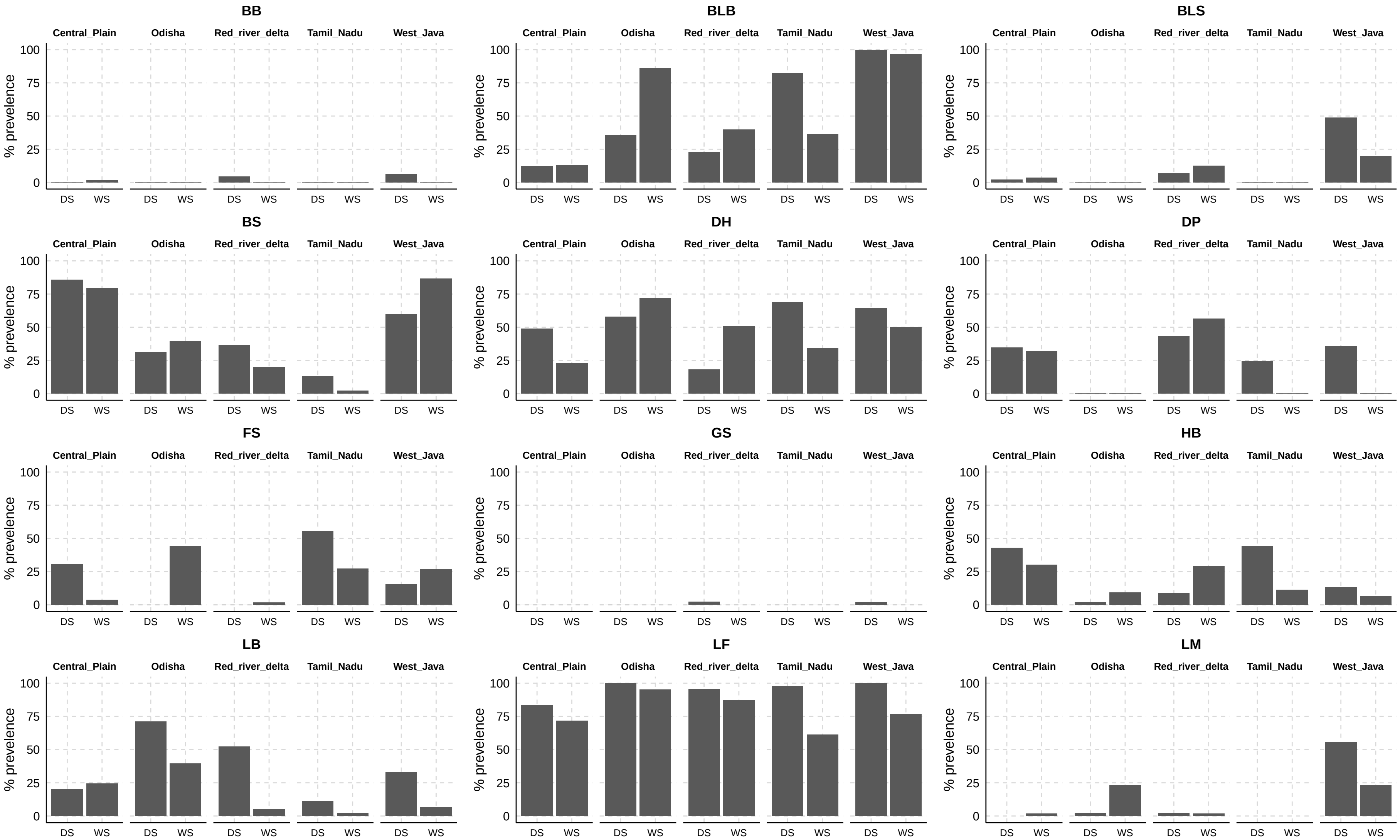


Figure I-2: Bar graphs showing prevalence of injuries a across production environments and seasons. BB: Bugburn, BLB: Bacterial leaf blight, BLS: Bacterial leaf streak, BS: Brown spot, DH: Deadheart, DP: Dirty panicle, FS: False smut, GS: Grassy stunt, HB: Hopperburn, LB: Leaf blast, LF: Leaffolder injury, LM: Leaf miner injury, LS: :Leaf scald, NB: Neck blast, NBS: Narrow brown spot, PM: Panicle mite injury, RB: Rice bug injuries, RGS: Ragged

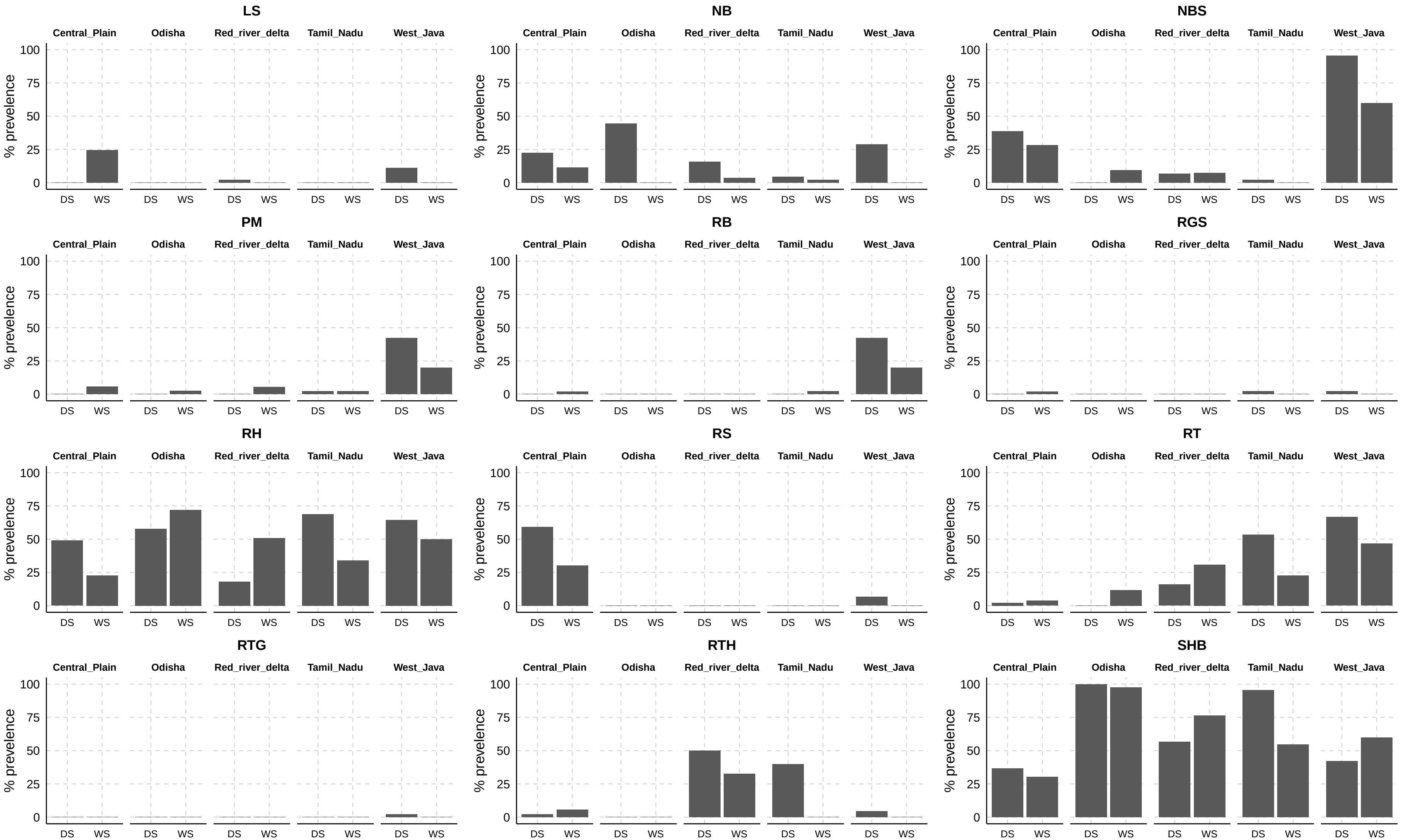
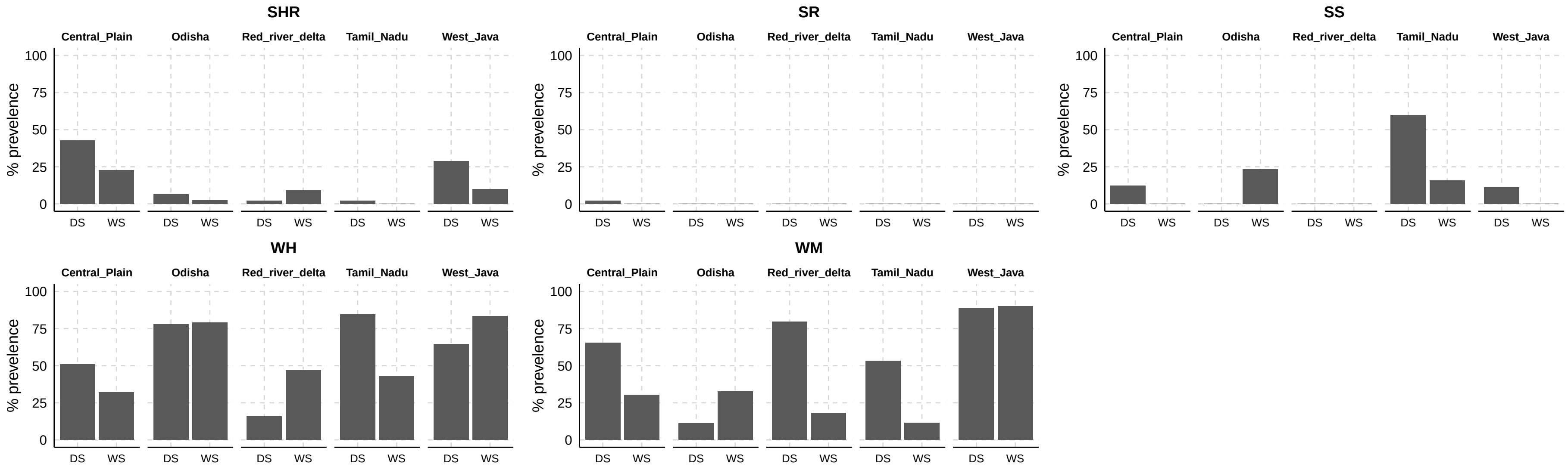


Figure I-3: (*Continue*) Bar graphs showing prevalence of injuries a across production environments and seasons.



centering

Figure I-4: (*Continue*) Bar graphs showing prevalence of rice injuries across production environments and seasons.

Correlation networks of rice juries

The correlation networks of rice injuires for crop health data are given in Figure. I-5 to Figure I-14. It considered co-occurrence to be positive rank correlations (Spearman's correlation) between pairs of injuries within each dataset with the strength of the relationship represented by the correlation coefficient. Negative correlations (indicative of competitive interactions) were also included in networks. The Fruchterman-Reingold algorithm applied in networks tends to co-locate injuries that are negatively correlated. Thus it may be possible to identify clusters with strong correlations, which is either positive or negative. Thus there is no clear delineation between groups of positively and negatively correlated injuries in the co-occurrence networks produced by the Fruchterman-Reingold algorithm.

Additionally, rice injuries in survey data show significant higher interaction strength among positive interactions. Moreover, higher frequency of positive strong interactions overall are found. All of the pairwise comparisons of the distribution of correlation coefficients were significantly different (Mann–Whitney U tests, all p-values < 0.05)

Structures, compositions, and communities of co-occurrence network of rice pest injuries

The co-occurrence networks for crop health data are given in Figure. I-17a to Figure I-26a. Nodes are co-occurring injuries. A connection stands for a positive (Spearman's $\rho > 0$) and significant ($p < 0.05$) correlation. To analyse co-occurrence network of rice injuries, I focus on the most prominent properties of nodes in a network node: node strength, betweenness, and clustering coefficient. Node degree is a measure of the number of connections a node has, weighted by Spearman's correlation coefficient. Betweenness measures how often a node lies on the shortest path between every combination of two other nodes, indicating how important the node is in the flow of information through the network (Opsahl et al., 2010). The local clustering coefficient is a measure of the degree to which nodes tend to cluster together. It is defined as how often a node forms a triangle with its direct neighbors, proportional to the number of potential triangles the relevant node can form with its direct neighbors Opsahl et al. (2010). These measures are indicative of the potential association activity through the network. As activated injuries can activate other injuries, a more densely connected network facilitates injury occurrence. Moreover, the community structure of the networks derived from the empirical data can be inspected to identify syndromes (clusters of injuries) that are especially highly associated.

Central Plain, Thailand

Dry season network (Figure I-17a) is comprised of 18 associated injuries and 60 associations (edges). The network show two groups of injury syndromes (the combination of injuries) based on the optimal clustering algorithm. Group1 (green) was more closely clustered than another group, according to node properties. The injuries in group1 (WH, SHR, SHB, DP, BS, RH, NB, DH, FS, HB, and RS) have high clustering coefficient This indicate that these injuries form complex co-occurrence relationships. Network properties (Figure I-17b) reveal WM and LF, BS, BLB, NBS are high-betweenness nodes. As opposed to other injuries, LB and BLS had low scores on at least two centrality measures. Apparently, LB occur less possibly (low between-

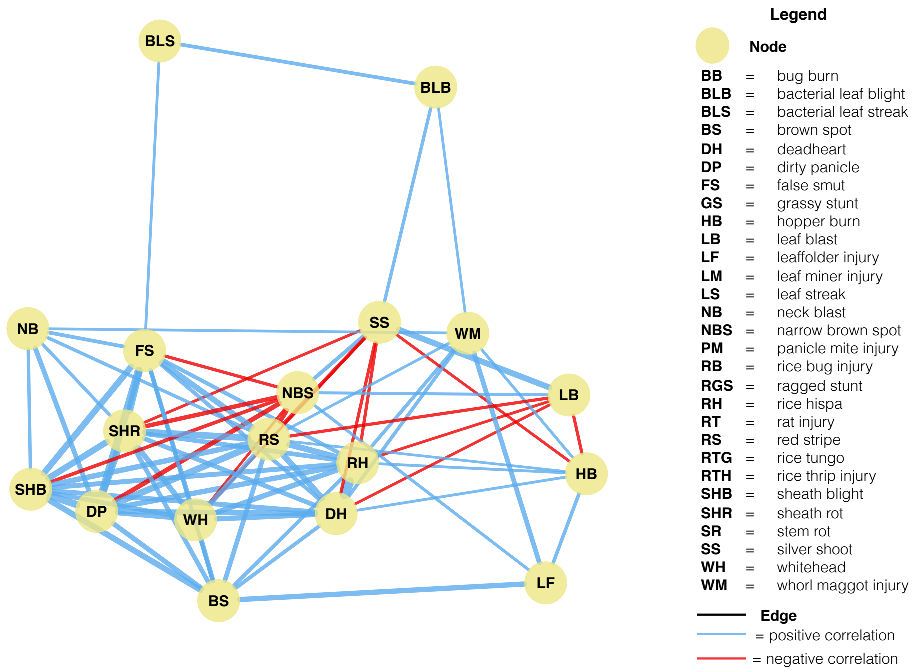


Figure I-5: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

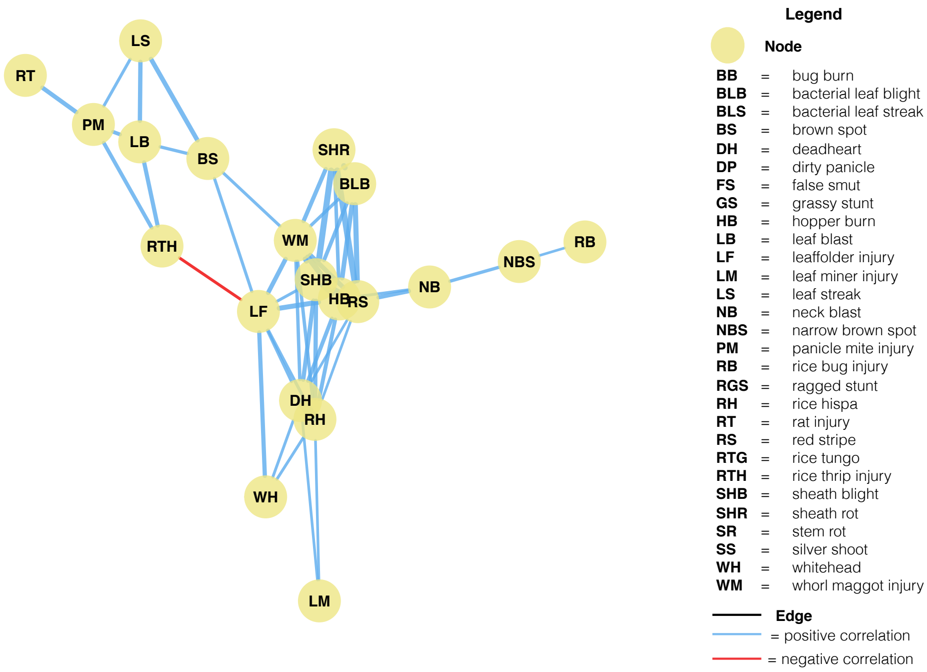


Figure I-6: Co-occurrence network of rice injuries in wet season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

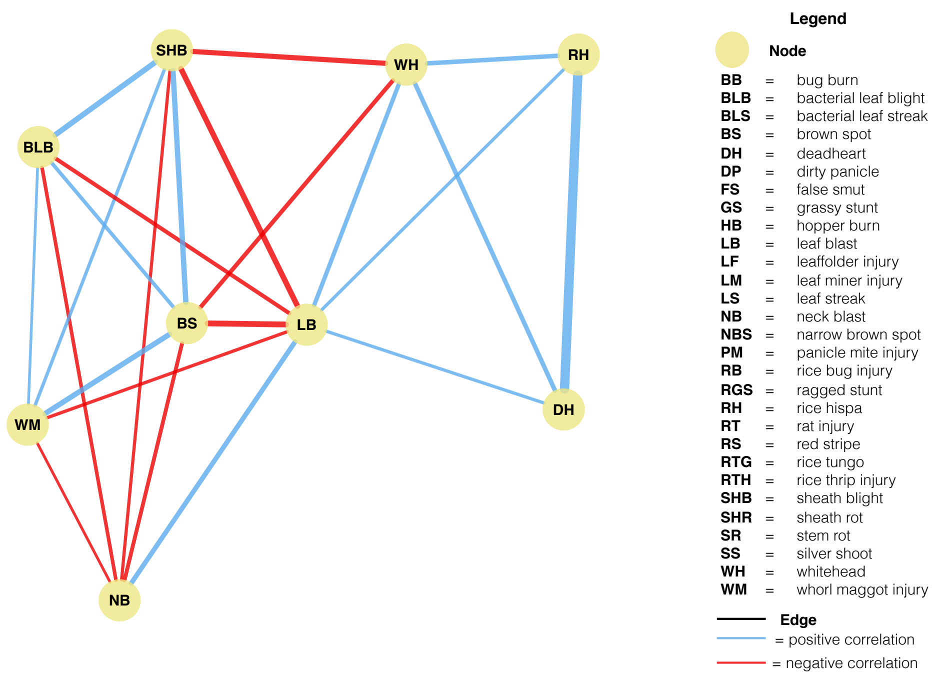


Figure I-7: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

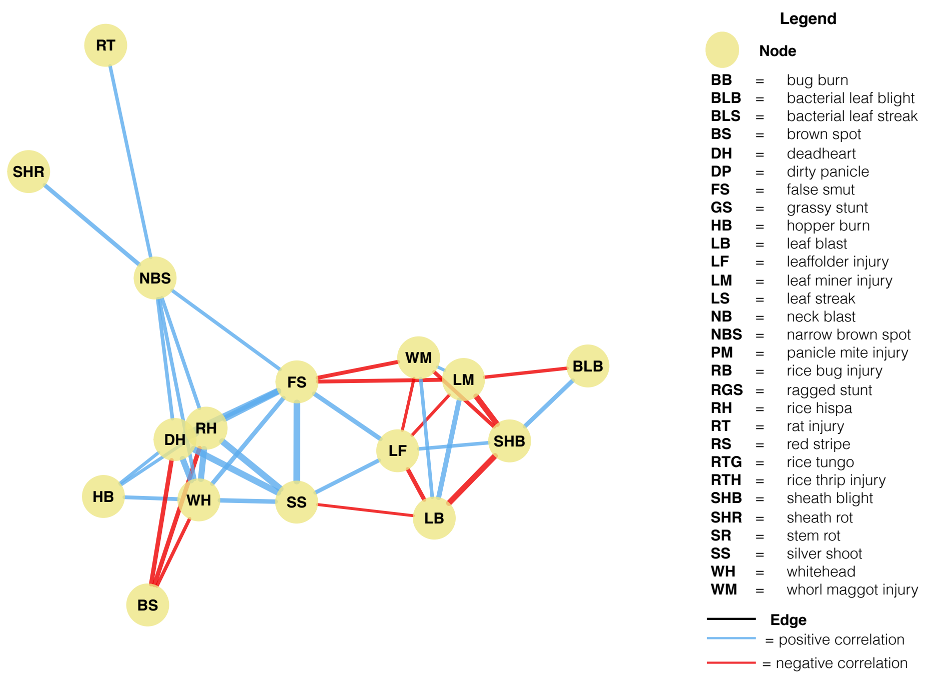


Figure I-8: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

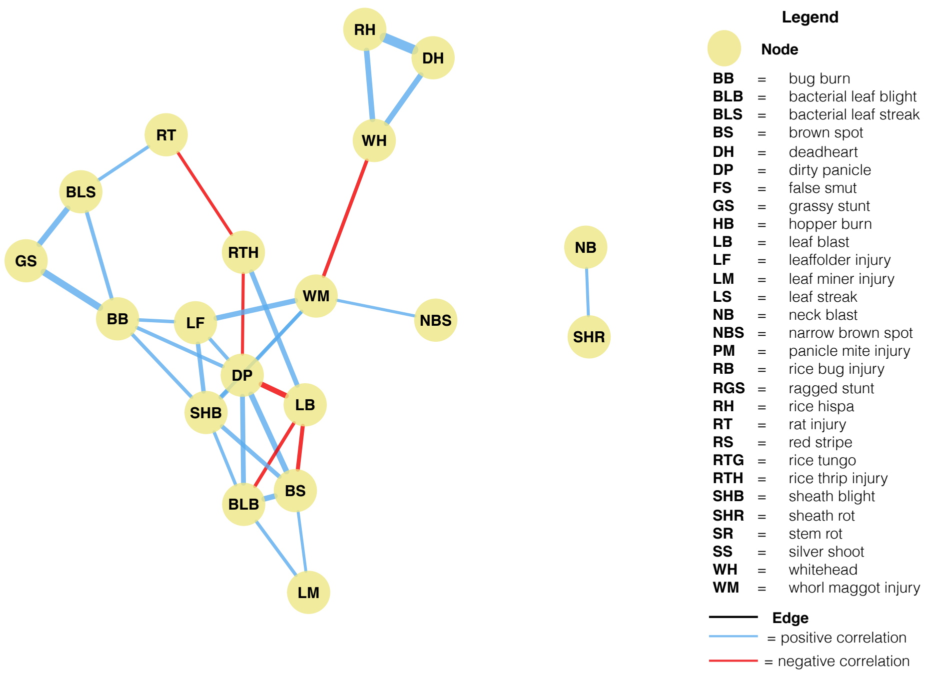


Figure I-9: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

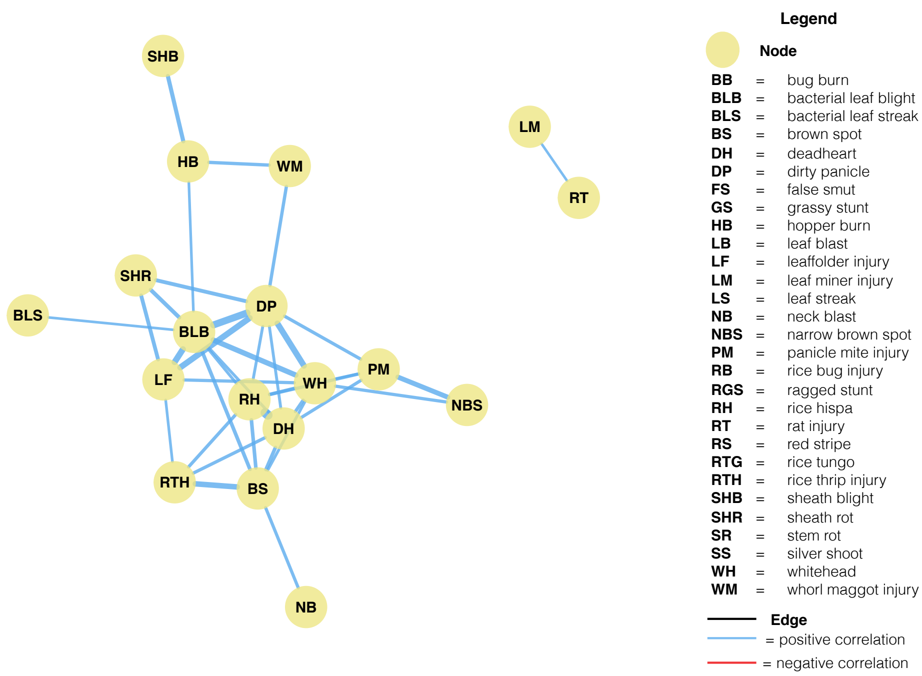


Figure I-10: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

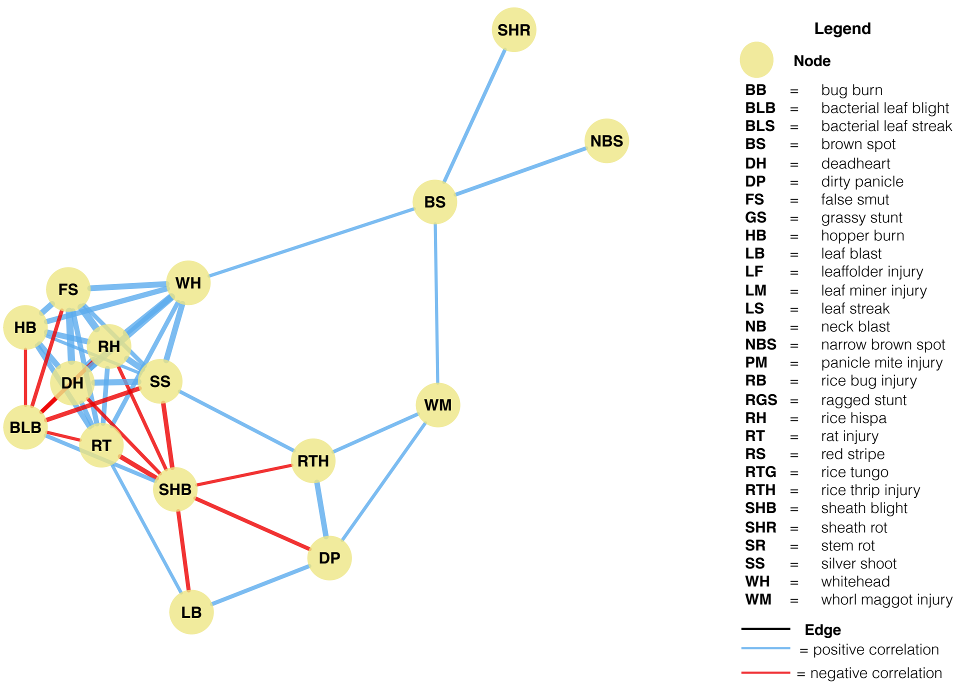


Figure I-11: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

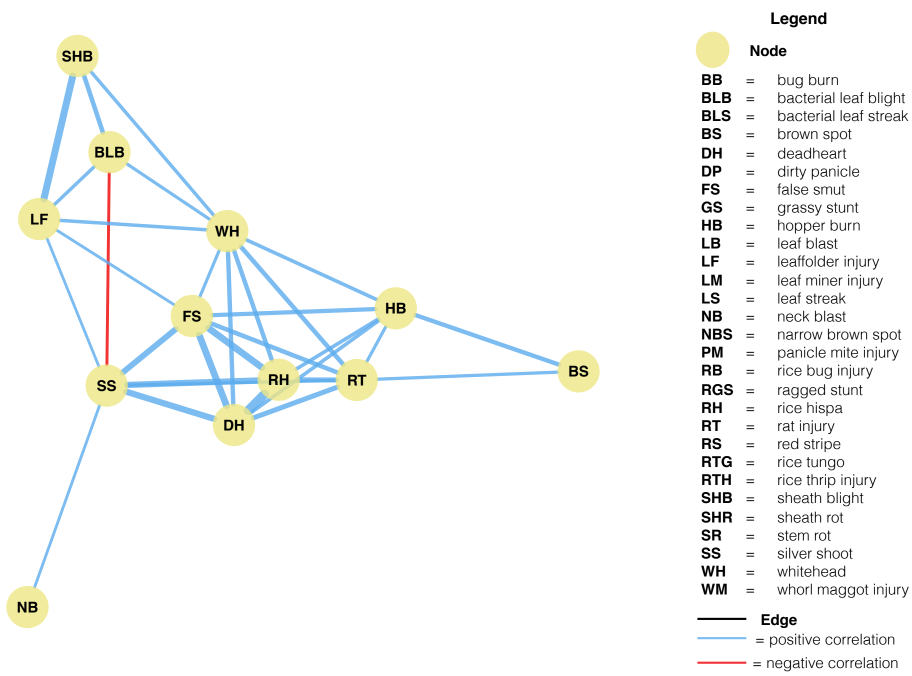


Figure I-12: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

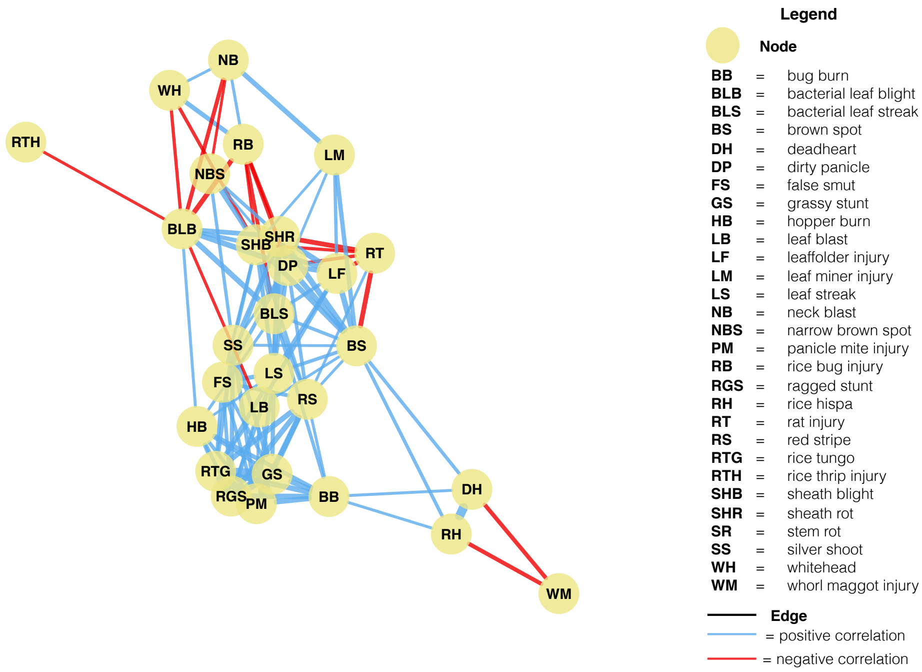


Figure I-13: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

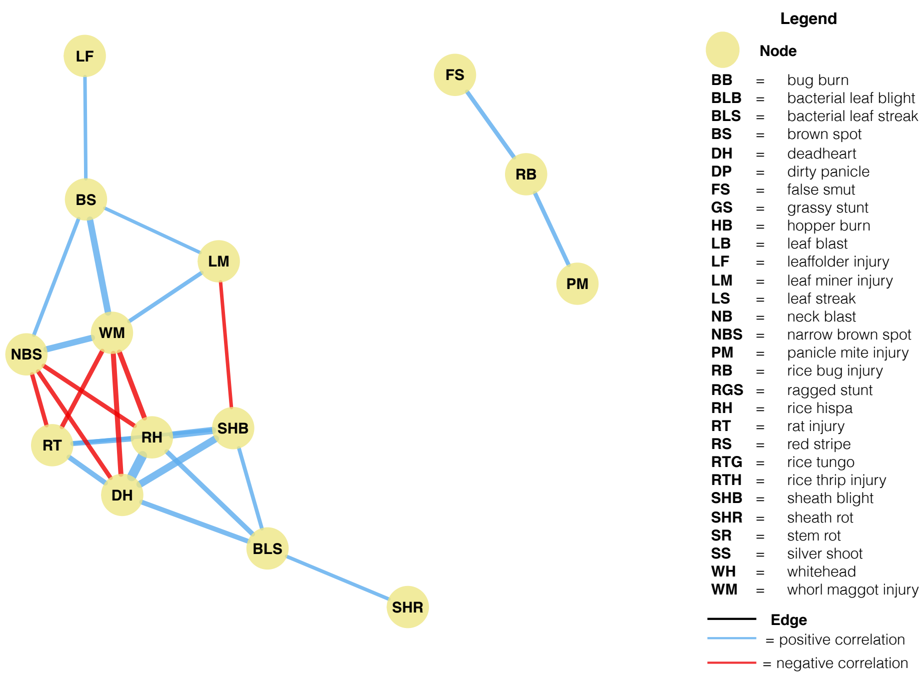
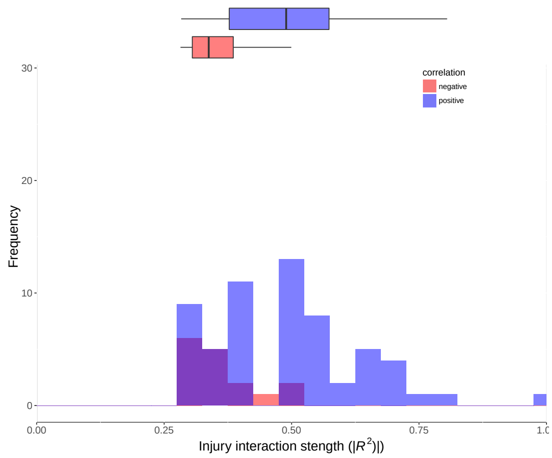
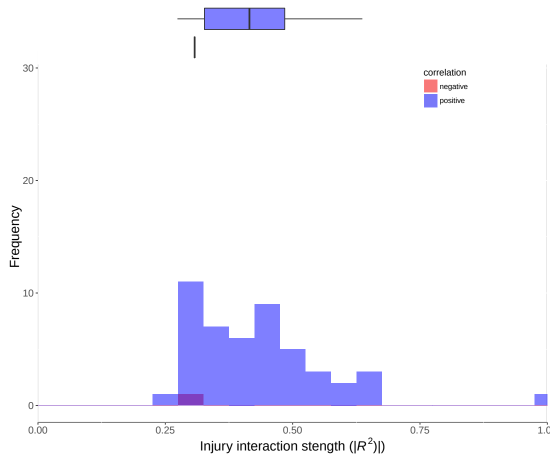


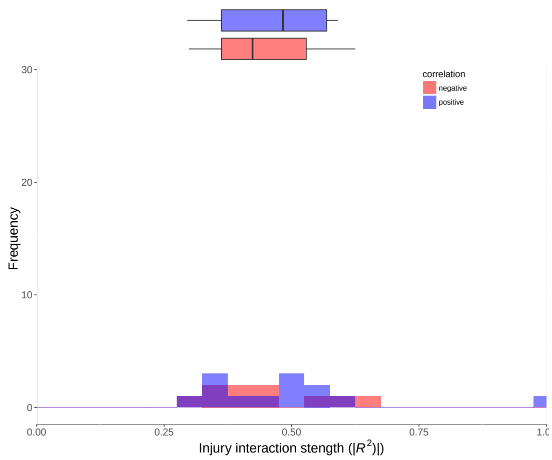
Figure I-14: Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



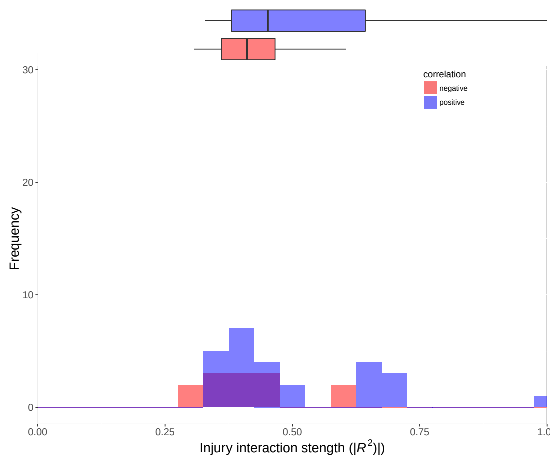
(a) Distribution of pairwise injury correlations in survey data in dry season at Central Plain



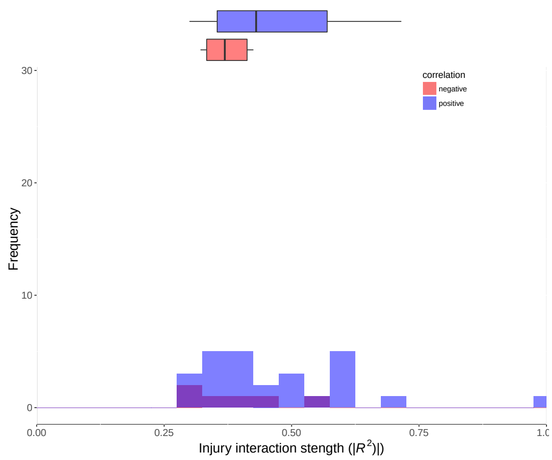
(b) Distribution of pairwise injury correlations in survey data in wet season at Central Plain



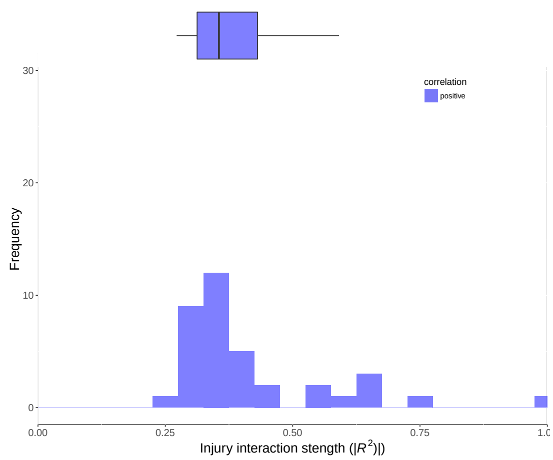
(c) istribution of pairwise injury correlations in survey data in dry season at Central Plain



(d) istribution of pairwise injury correlations in survey data in dry season at Central Plain

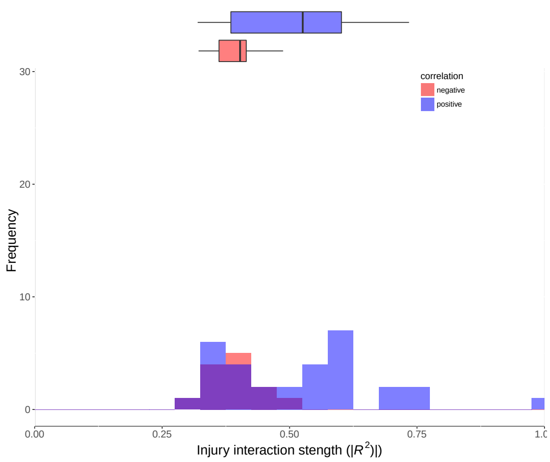


(e) istribution of pairwise injury correlations in survey data in dry season at Central Plain

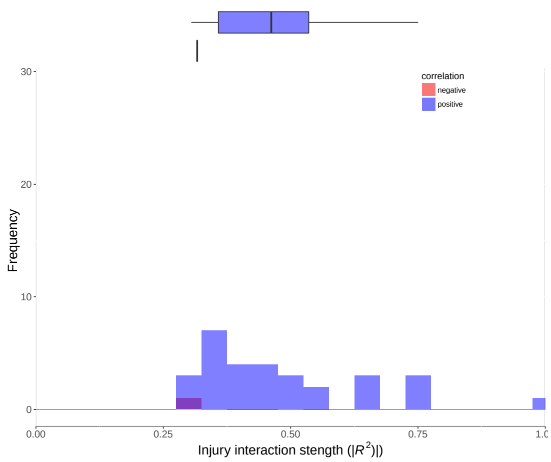


(f) Distribution of pairwise injury correlations in survey data in dry season at Central Plain

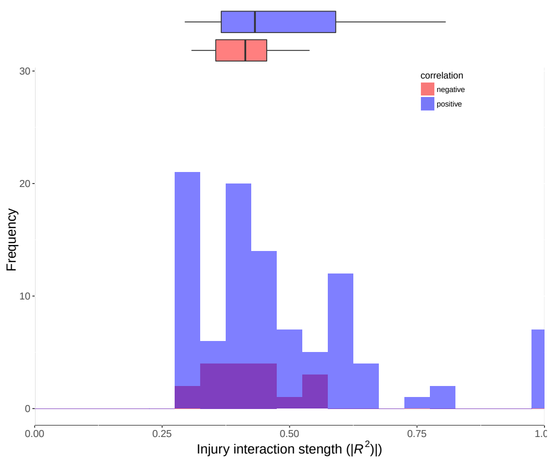
Figure I-15: .



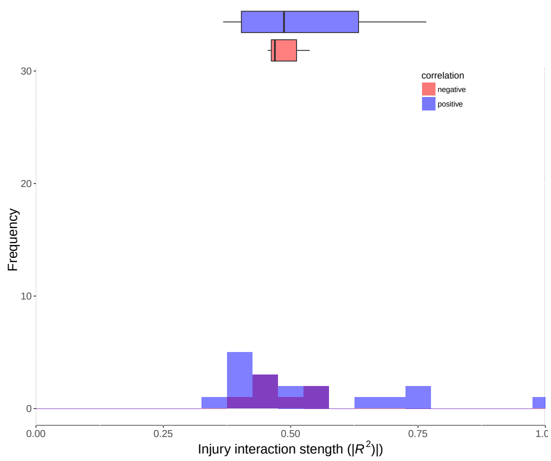
(a) Distribution of pairwise injury correlations in survey data in dry season at Central Plain



(b) Distribution of pairwise injury correlations in survey data in dry season at Central Plain



(c) Distribution of pairwise injury correlations in survey data in dry season at Central Plain



(d) Distribution of pairwise injury correlations in survey data in dry season at Central Plain

Figure I-16: .

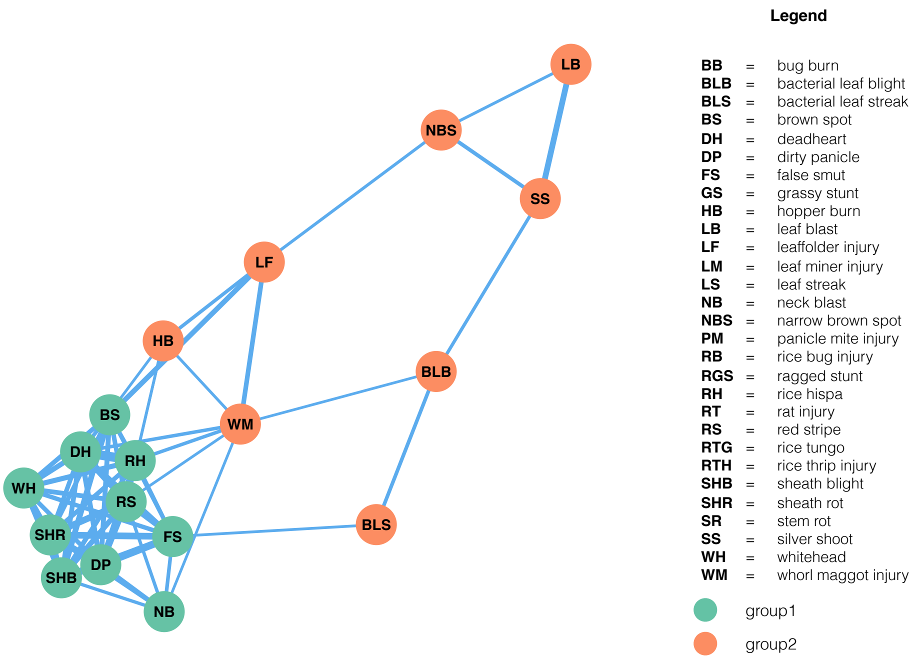
ness), and did not occur with other injuries (low degree and clustering coefficient). Because of high value of centrality, WM and BS can be indicators for monitoring the injury occurrence in each syndrome.

In wet season, the co-occurrence network of rice injuries (Figure I-18a) reveals 4 syndromes, 20 injuries, and 48 significant relationships (edges). Syndrome3 (purple) is comprised of BLS, RS, HB, SHB, SHR and WM. They were placed closer to each other than other syndromes based on the structure and clustering coefficient (Figure I-18b). This syndrome also links with syndrome1, 2 and 4. Based on network structure and betweenness, WM can be an indicator for monitoring pest and disease incidence in this season, and it also is the injury in syndrome3, which is the central syndrome of this network.

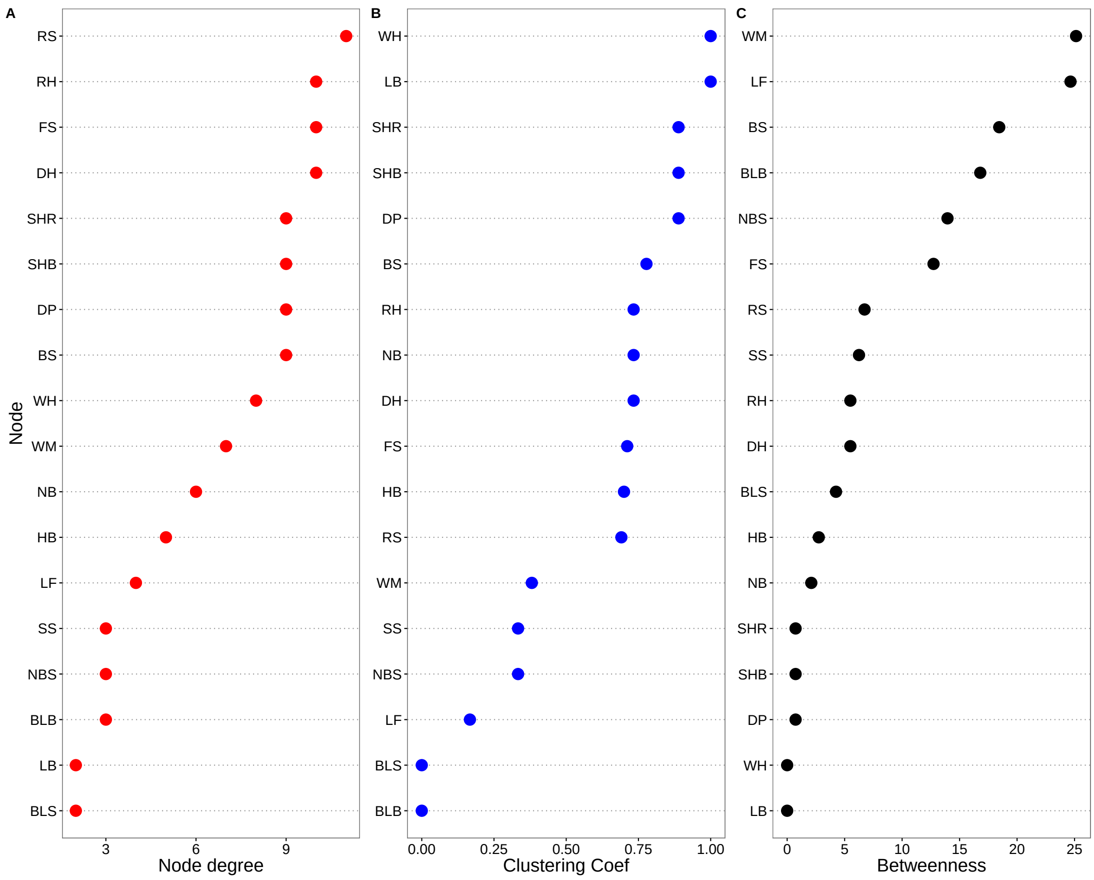
Odisha, India

Co-occurrence network of rice injuries in dry season (Figure I-19a) consists of of 9 associated injuries and 13 associations. The network shows two isolated injury syndromes. One was the combination of BS, BLB, WM and SHB. Another was RH, WH, DH, LB, NB. This network suggests indicators for monitoring such as LB for syndrome1, and WM, SHB, BS, and BLB for syndrome2 base on centrality (Figure I-19b).

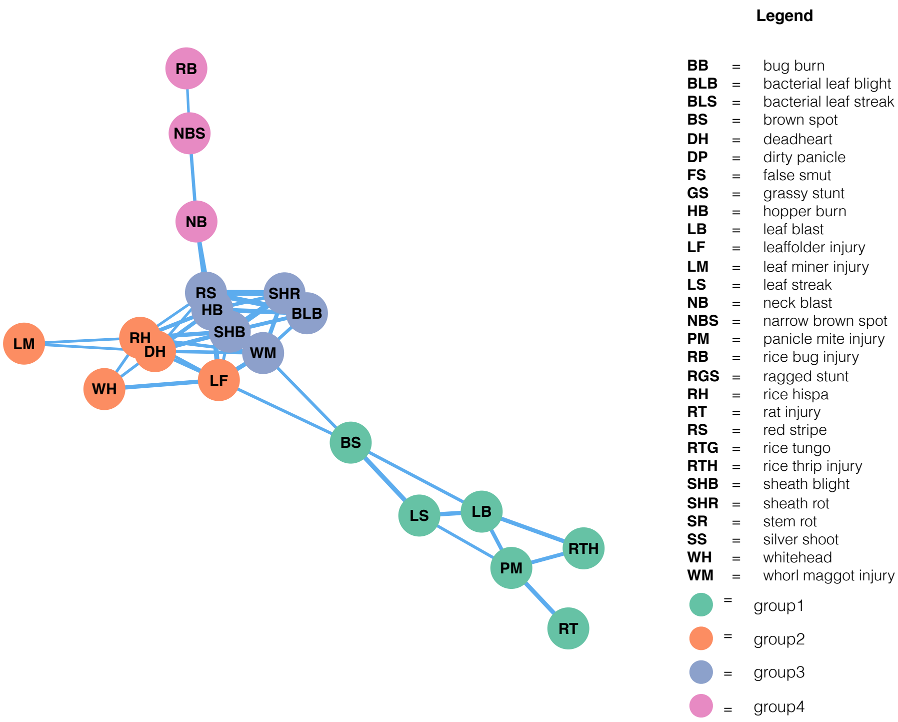
The network in wet season (Figure I-20a) is more complex than the network in dry season. It is comprised of 15 nodes with 26 edges. The network reveals four syndromes. Syndrome4, composed of LB, LM and WM, is isolated from the rest. Syndrome2 is placed in between syndrome1 and 3. The injuries in syndrome2 have high value of node degree and clustering coefficient (Figure I-20b), which mean they were



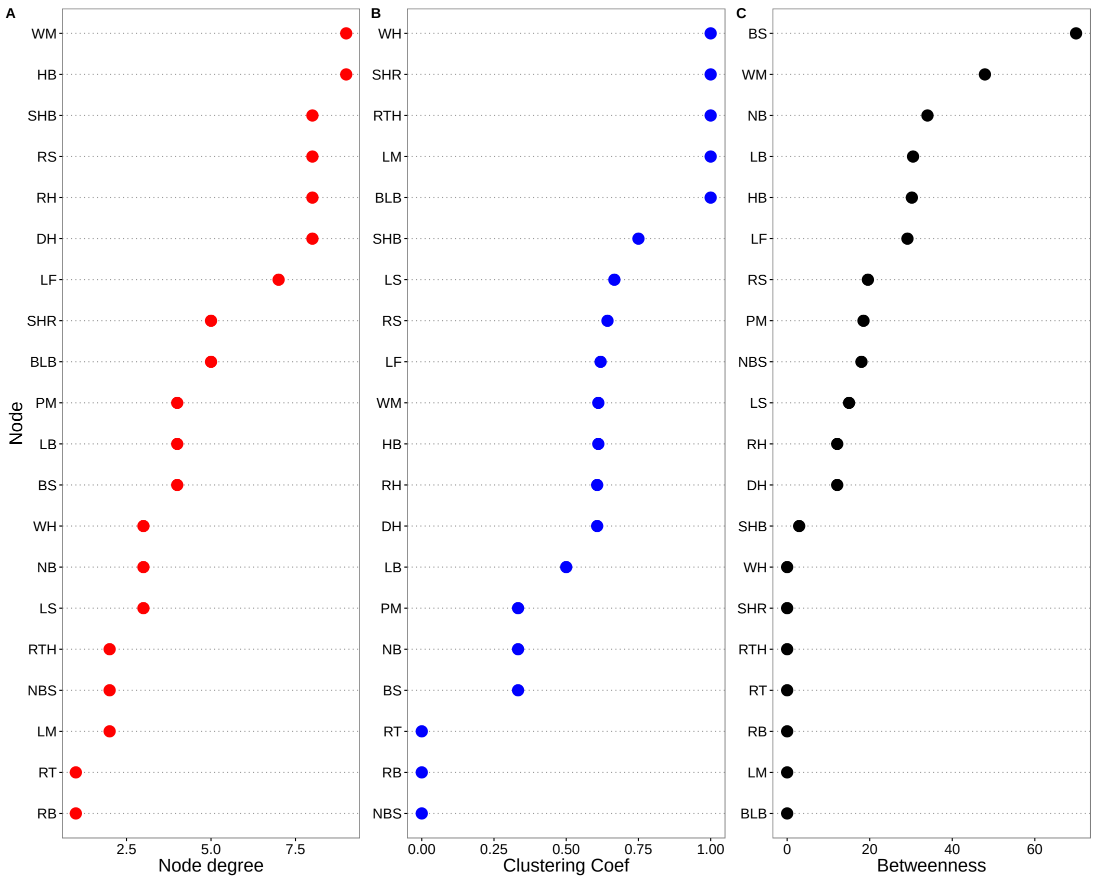
(a) Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.



(a) Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness connected to many injuries. NBS and LF, injuries in group1 and group2, respectively present high betweenness values and connected to injuries of group3. They are also good indicators for monitoring. This indicated that injuries of syndrome3 have high chance to occur together with group1 and 2, but not group4.

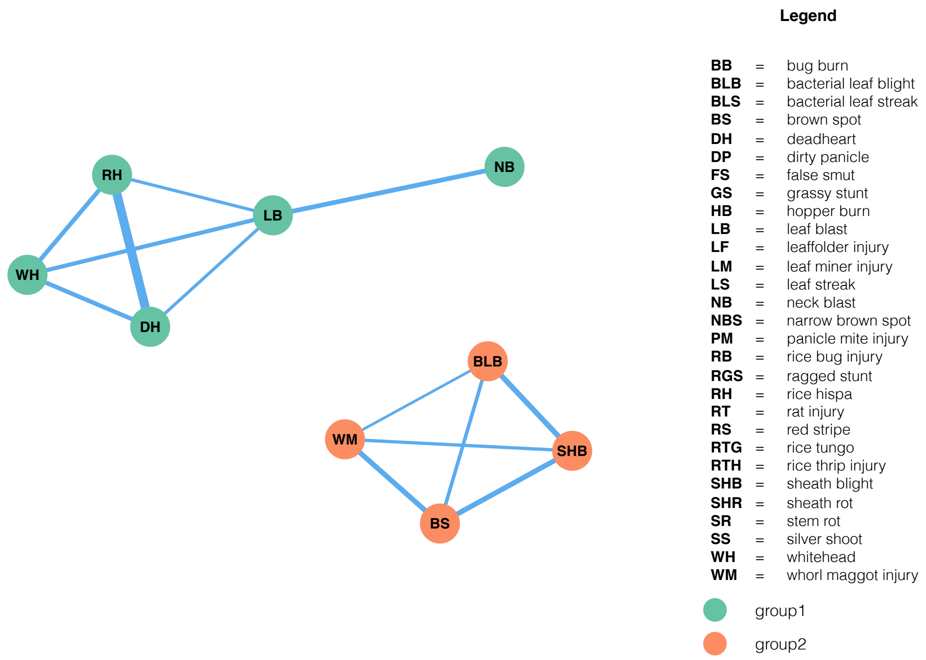
Red River Delta, Vietnam

Co-occurrence network of rice injuries in dry season (Fig. I-21a) is comprised of 19 nodes and 26 associations. The network reveals three isolated syndromes, and two connected syndromes. BB of syndrome1 and SHB of syndrome3 can be indicators because of high values of centrality measures (Figure I-21b).

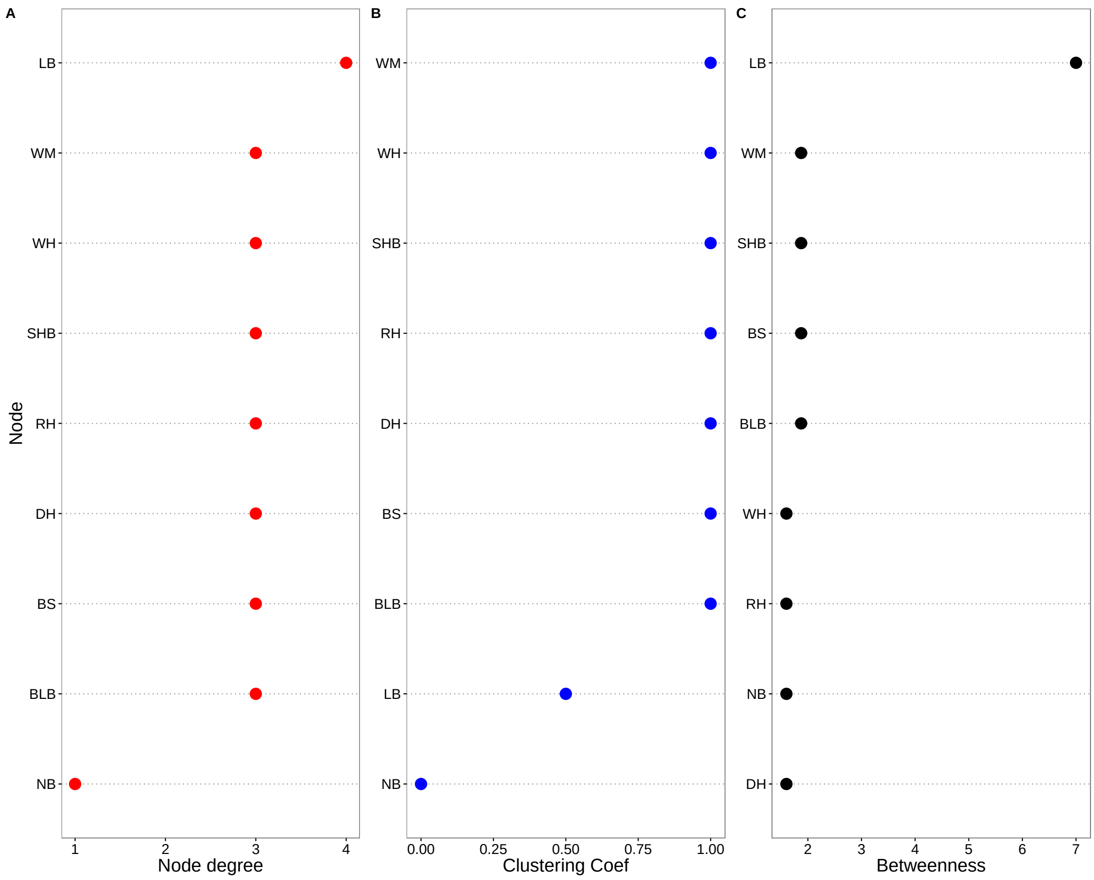
Wet season network (Figure I-22a)) consists of 18 injuries with 37 associations. It reveals 4 connected syndromes and an isolated syndrome. Group2 was located that could connect to Group3, 4, 5. According to Figure I-22b, BLB, DP could be good indicator, because they are likely to occur (high betweenness) and when they occurred, other injuries in other syndromes, except syndrome5 (high node degree) could be possibly observed.

Tamil Nadu, India

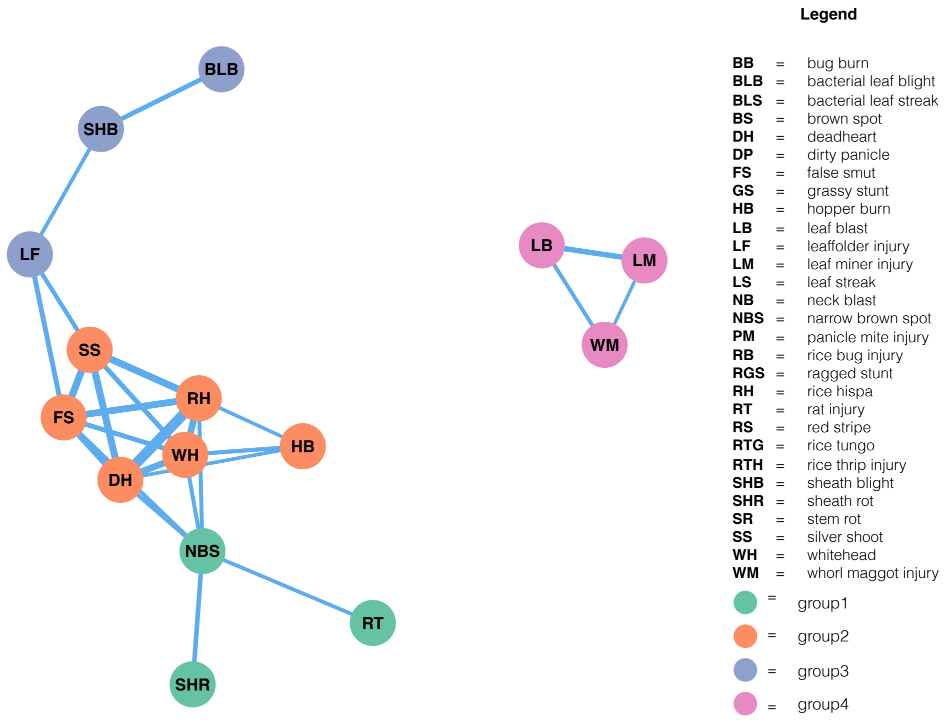
The dry season network (Fig I-23a) reveals three clustered groups of injury profiles. One of them is separated from other two. Syndrome1 is clustered tightly, which differ from group2 that injuries are placed further. SHB and BLB are disconnected to the rest. BS and WH highly tend to occur in this season (high betweenness) and are good indicators for monitoring in this season. Because clustering coefficient value of members in syndrome2 less than group1 (Fig I-23b), injuries in syndrome2 might occur together with one or two more injuries.



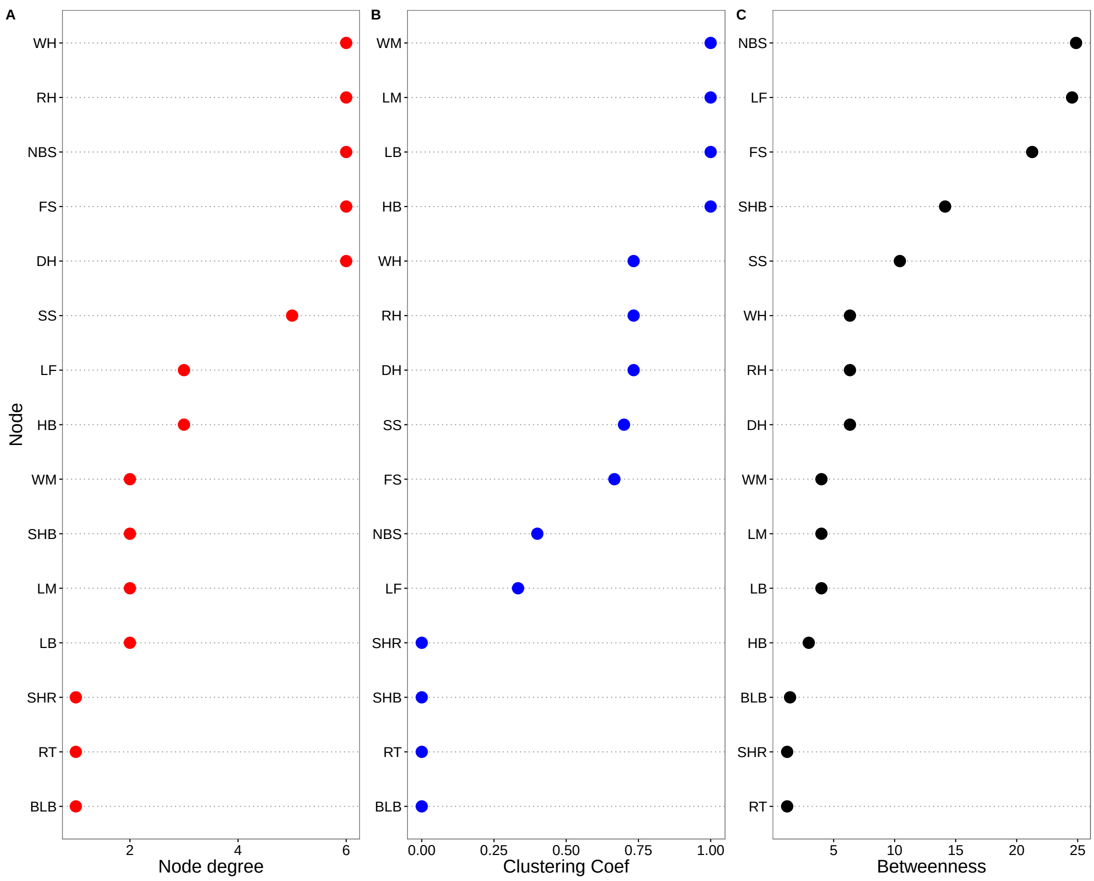
(a) Co-occurrence network of rice injuries in wet season at Odisha, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



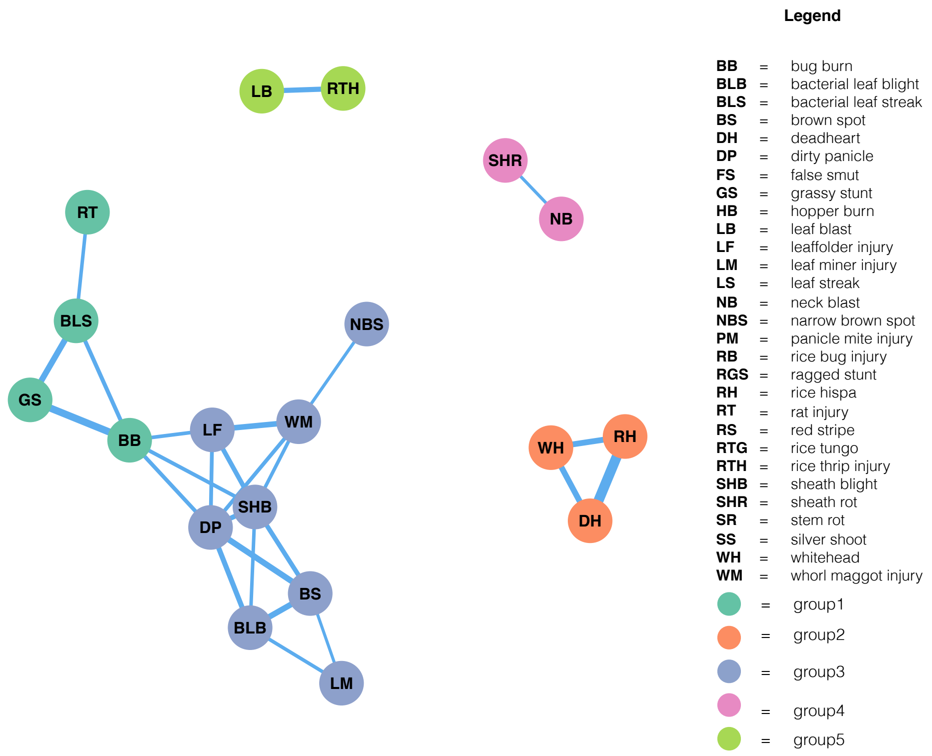
(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Odisha, India. A: node degree, B:clustering coefficient, and C:Betweenness.



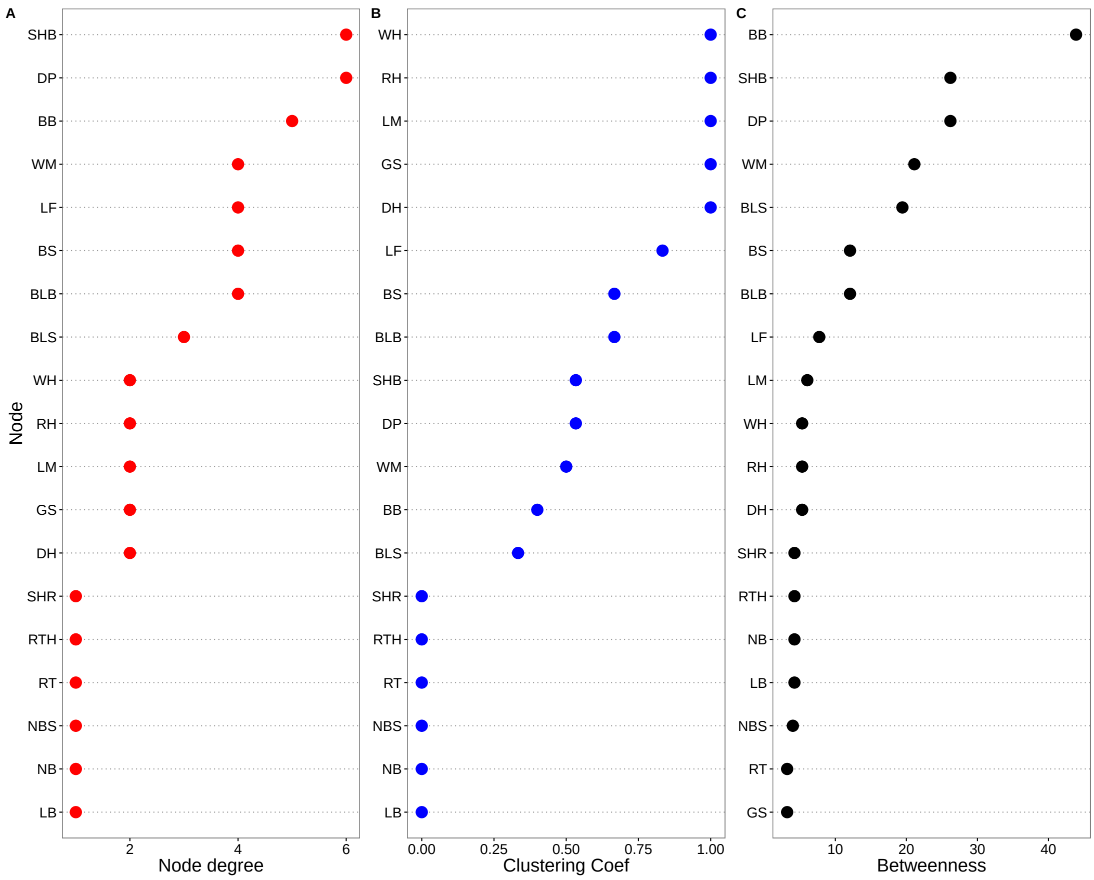
(a) Co-occurrence network of rice injuries in wet season at Odisha, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



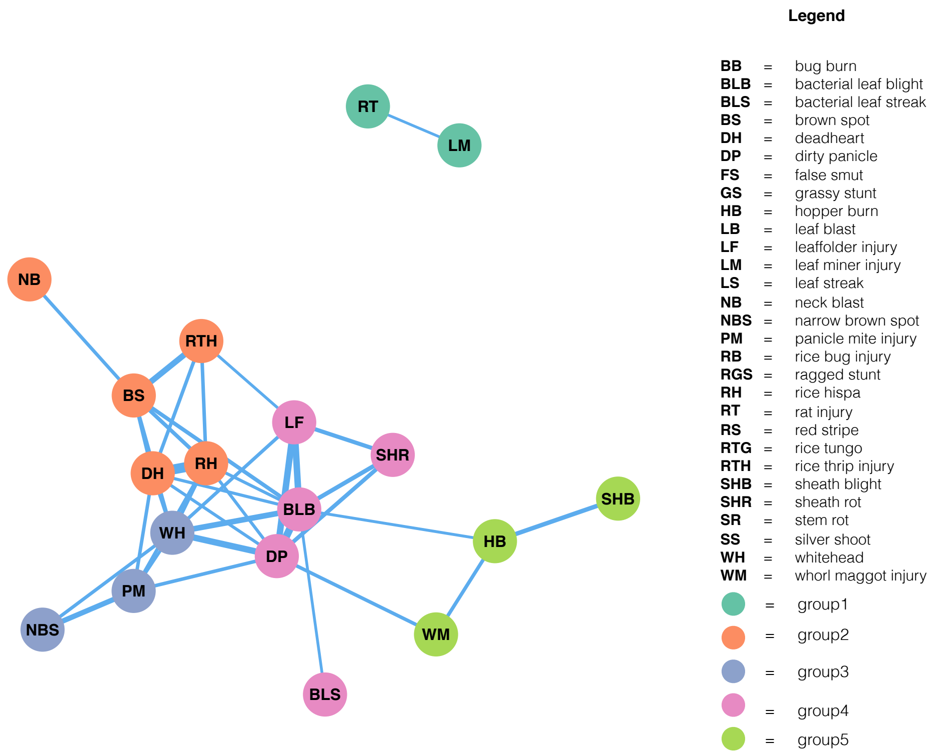
(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in wet season at Odisha, India. A: node degree, B:clustering coefficient, and C:Betweenness.



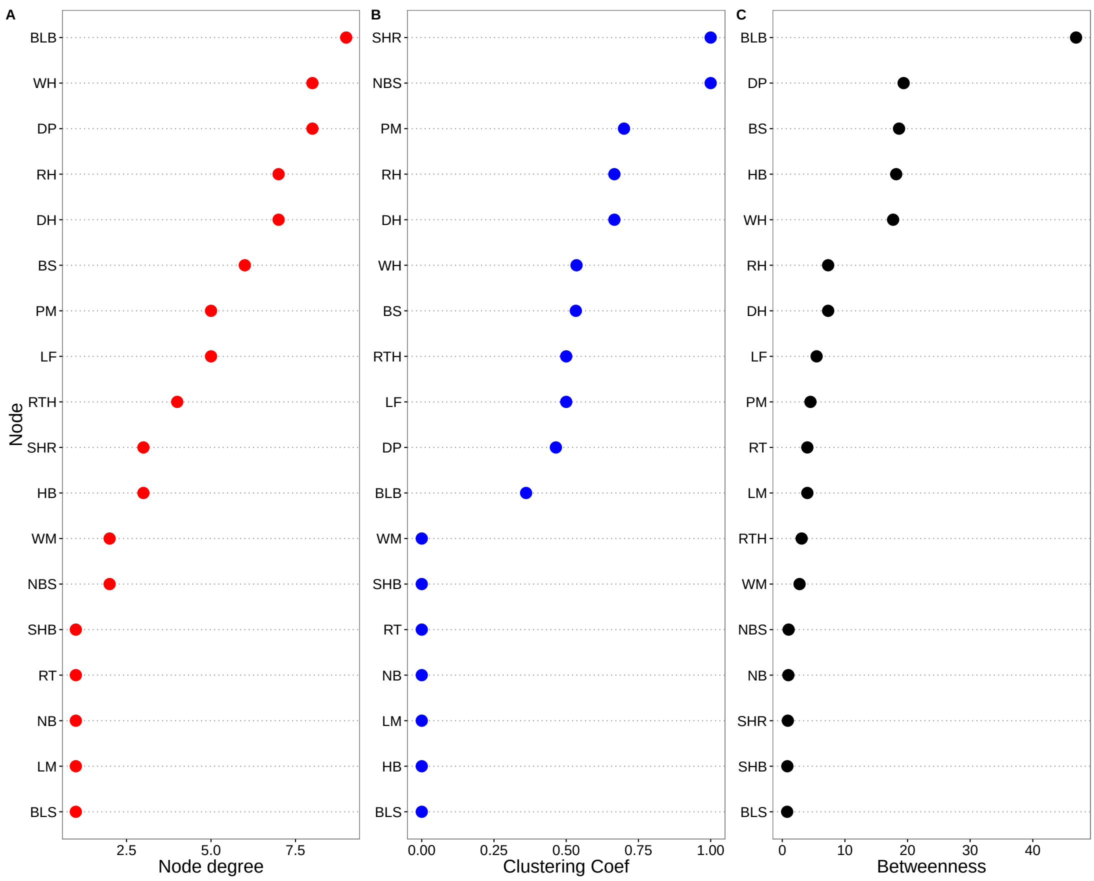
(a) Co-occurrence network of rice injuries in dry season at Red River Delta, Vietnam. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Red River Delta, Vietnam. A: node degree, B:clustering coefficient, and C:Betweenness



(a) Co-occurrence network of rice injuries in wet season at Red River Delta, Vietnam. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



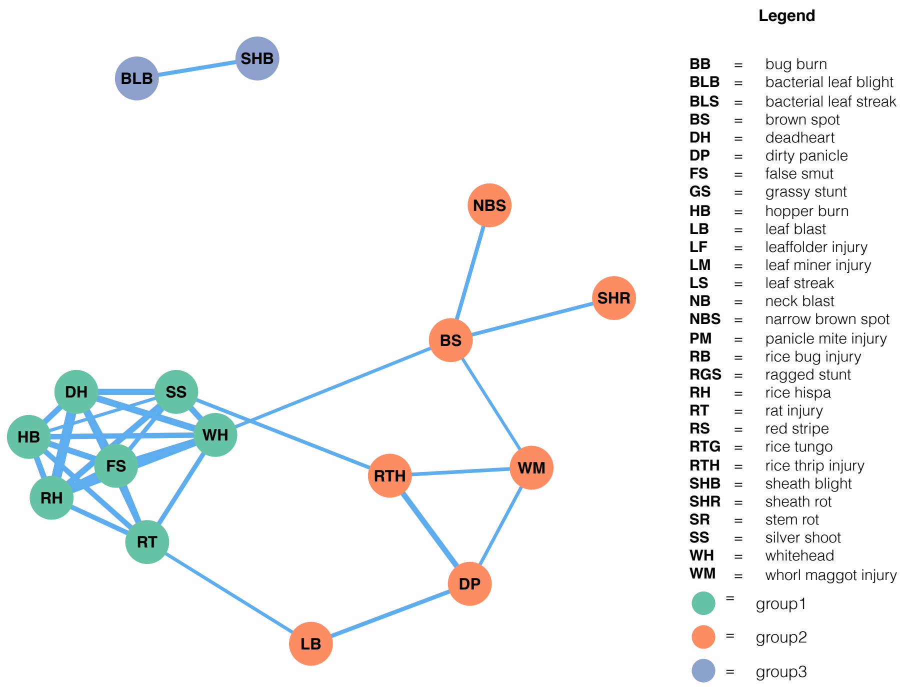
(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in wet season at Red River Delta, Vietnam. A: node degree, B:clustering coefficient, and C:Betweenness

Figure I-24a presents co-occurrence network of injury profiles in dry season. The network shows three syndromes related. Syndrome2 and syndrome 3 are linked with WH, and HB. NB is less possibly occur in this season because of low value of centrality measures. WH, RT can be good indicators because of high value of node degree and betweenness (Figure I-24b).

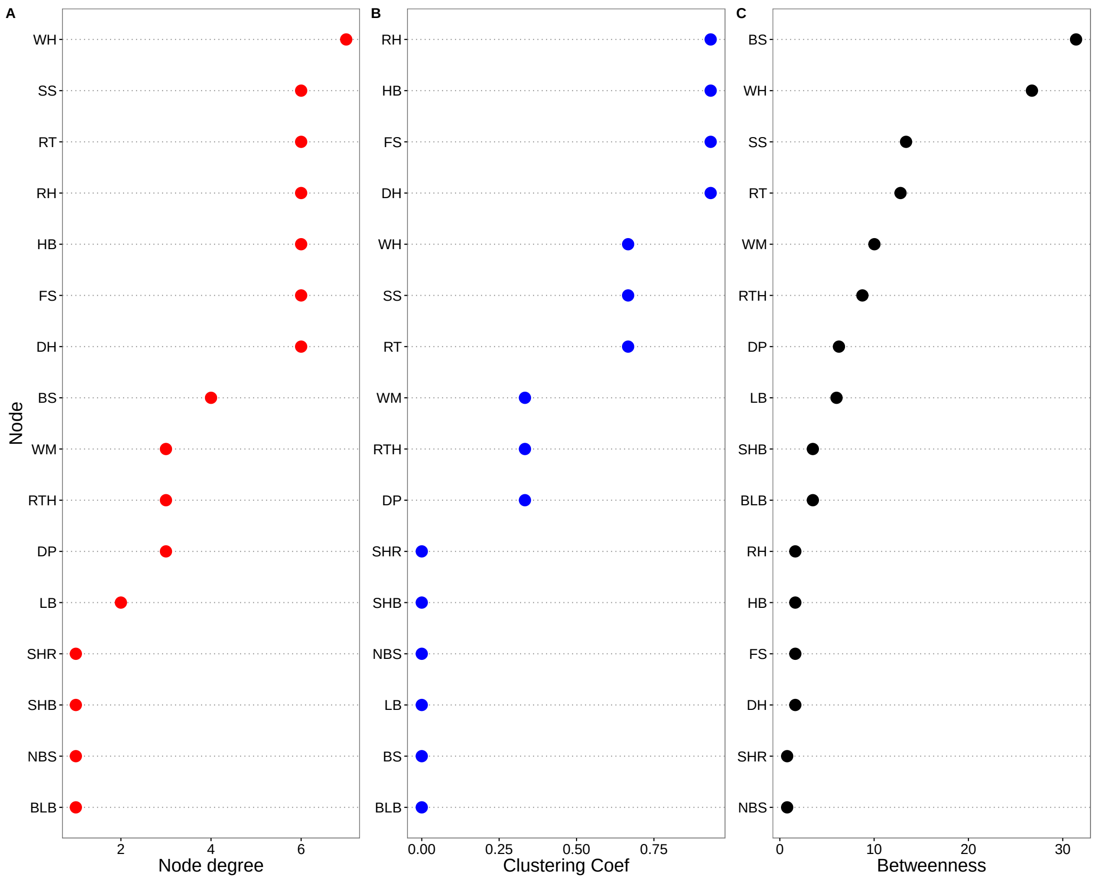
West Java, Indonesia

Co-occurrence network of injury profiles of dry season presented in Figure I-25a showing 26 injury nodes and 99 association. High number of pest injuries and disease could be observed in dry season. The network reveals the four syndromes of injury profiles. syndrome1 (green) and syndrome3 were close and syndrome2 and syndrome4 had less connection than others. Because of the structure and clustering coefficient, syndrome1 and syndrome3 are more likely to have chance to form complex association to each other. DH and RH of syndrome2 only related to BB of group1 and BS of syndrome3 but not to any of syndrome4. RT has smallest vales of all centrality measures. It indicated RT incidence is independent to other injuries. According to Figure I-26b, LM, BS and NB can be good indicators for monitoring pest and disease incidence in this season.

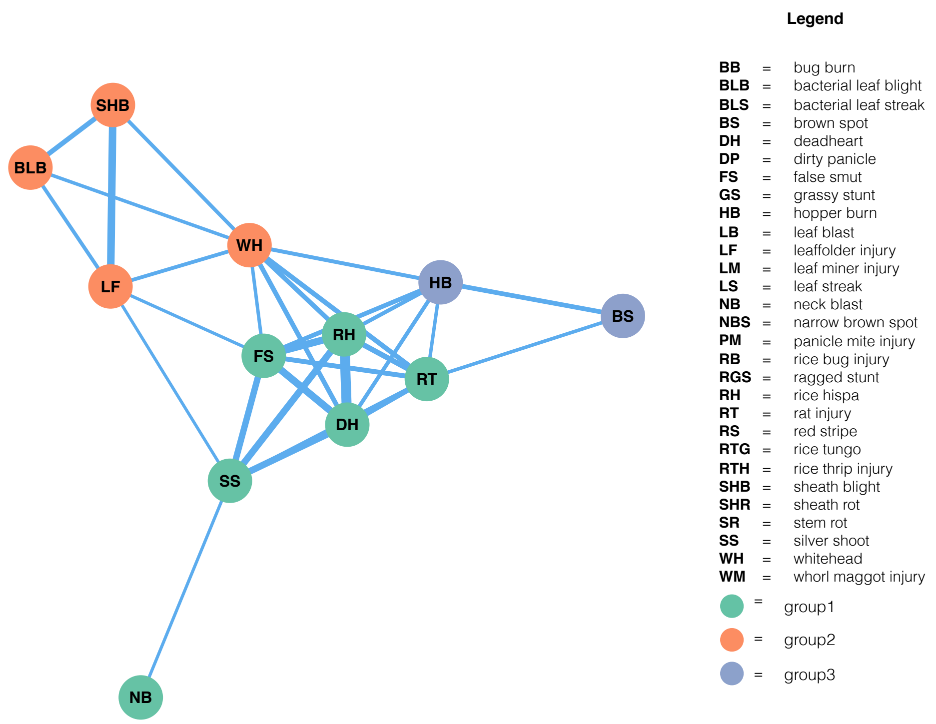
The co-occurrence network of injuries of wet season (Figure I-26a) shows 14 injuries and 18 associations. Compared to the network of dry season the numbers of pest injuries and diseases in wet season are less. The network structure reveals three types of injury syndromes. Syndrome1 (green) composed of SHB, RT, DH, RH, BLS, and SHR. Within this group, BLS and SHB seem to be good indicator (high betweenness and high node degree) according to Figure **??**. Syndrome2 (orange) seems to be



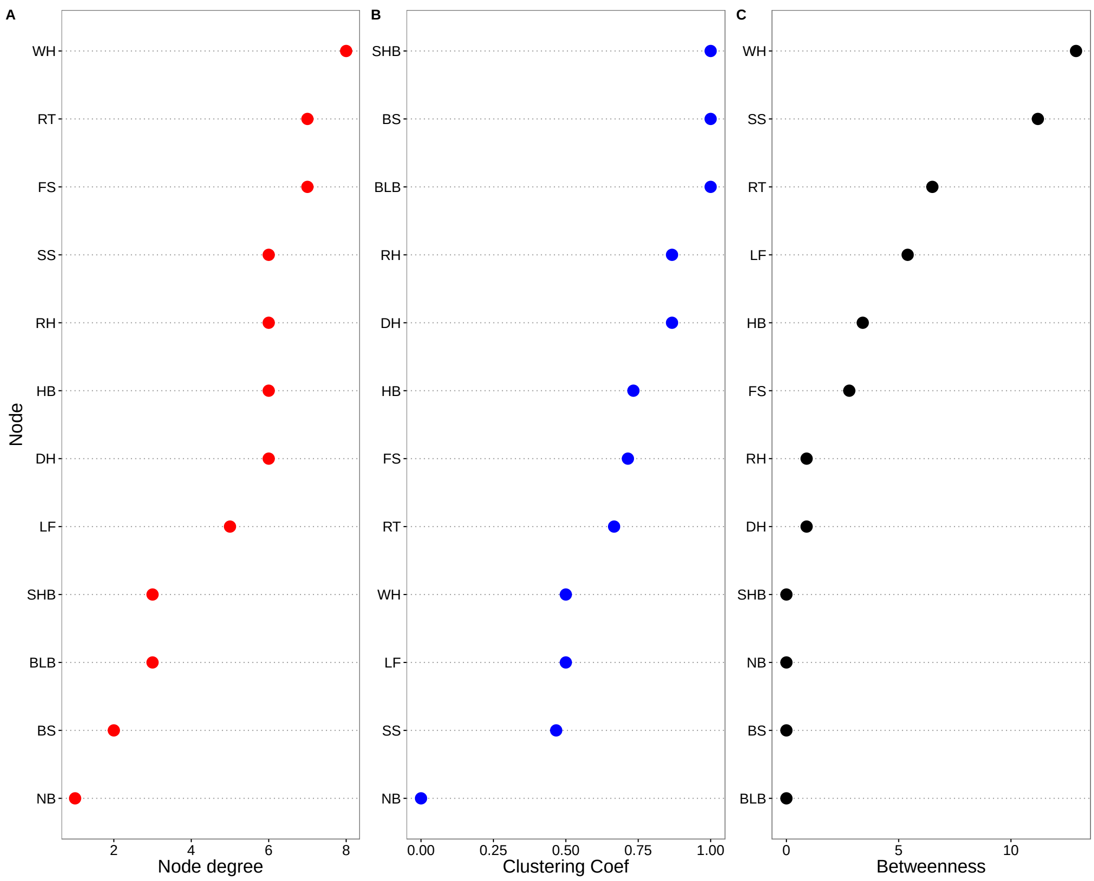
(a) Co-occurrence network of rice injuries in dry season at Tamil Nadu, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Tamil Nadu, India. A: node degree, B:clustering coefficient, and C:Betweenness.

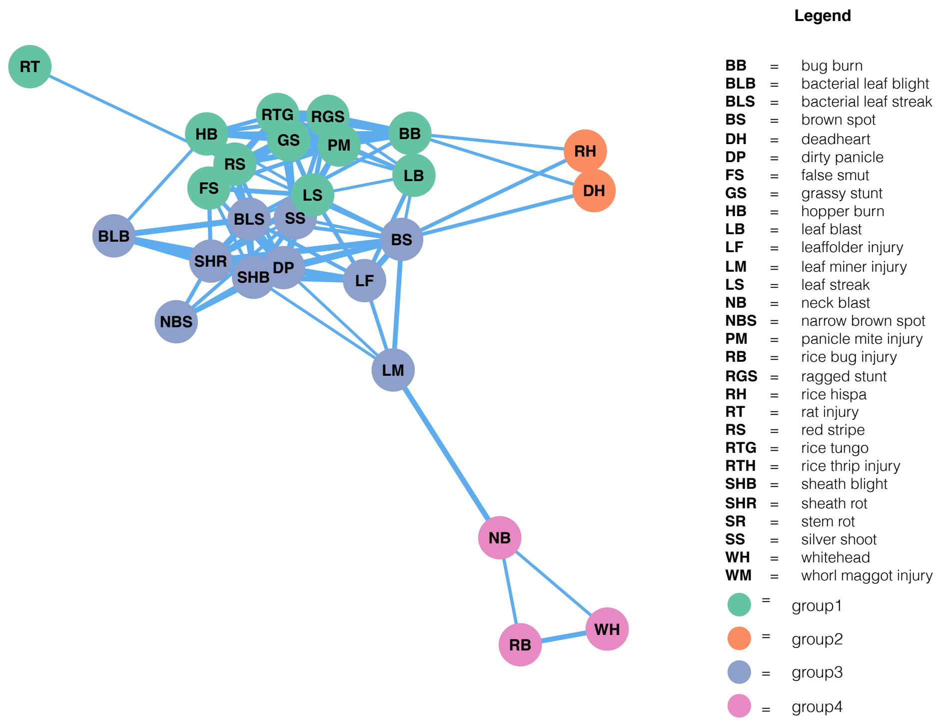


(a) Co-occurrence network of rice injuries in wet season at Tamil Nadu, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

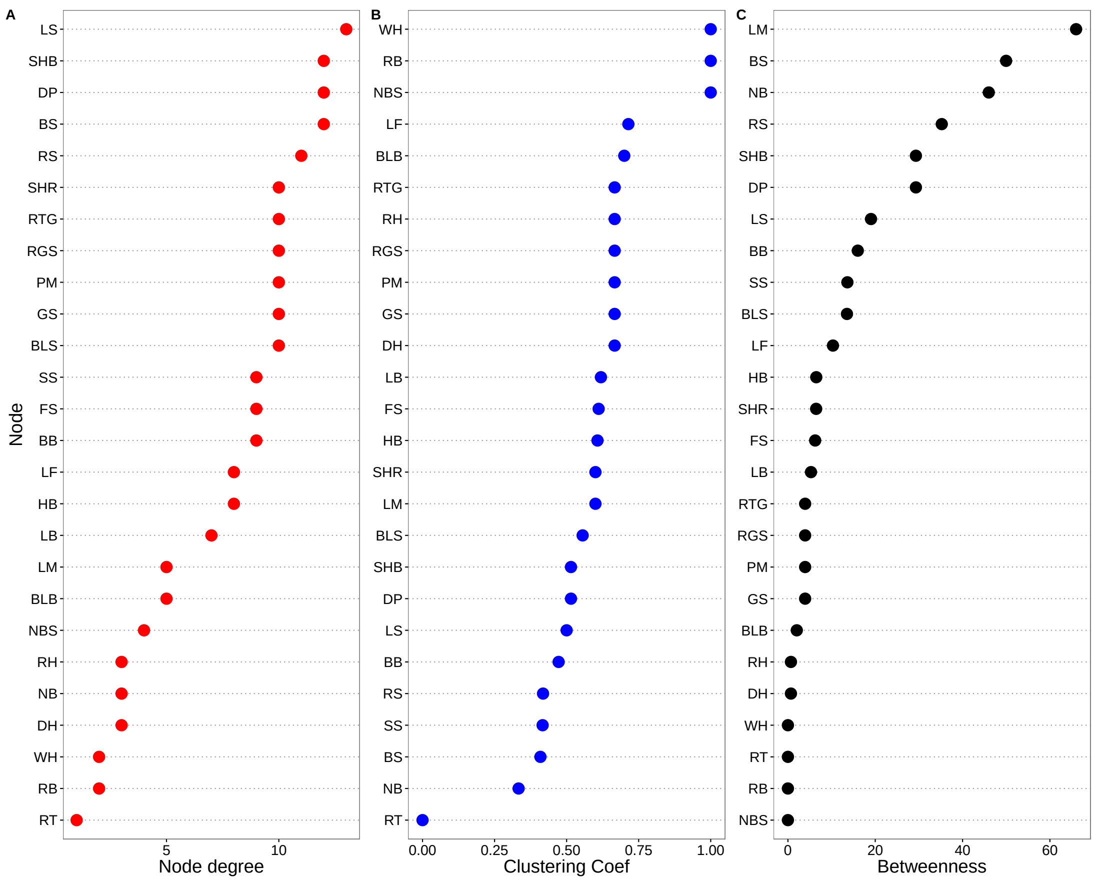


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in wet season at Tamil Nadu, India. A: node degree, B:clustering coefficient, and C:Betweenness.

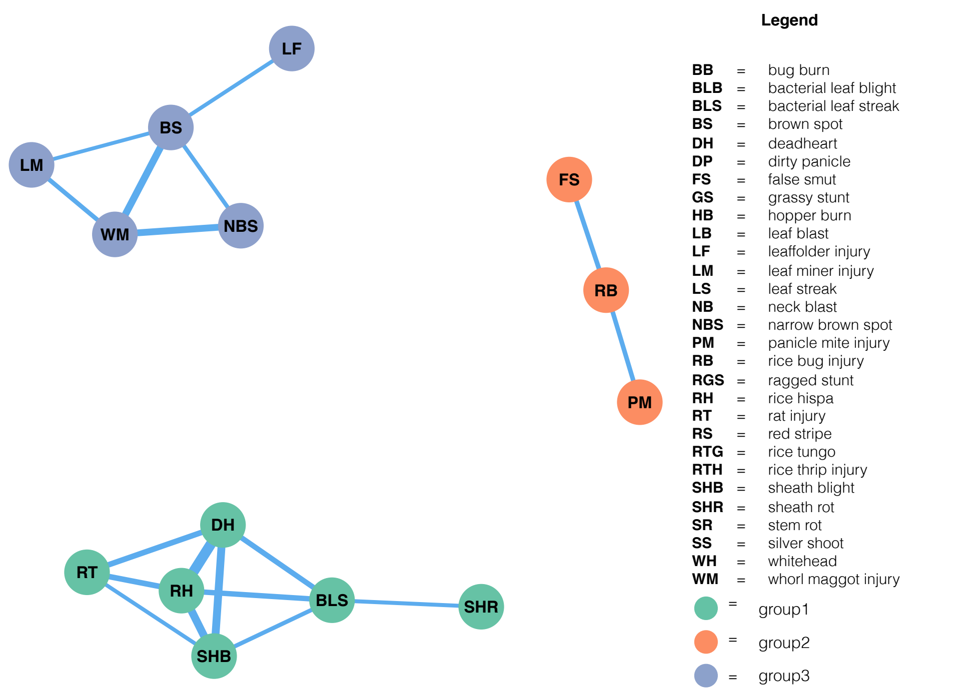
the panicle or tiller injury syndrome because there is the combination of PM, RB, and FS. Apparently, within this combination, BS is center of association, and early occur among other injuries. All of injuries in syndrome3 (purple), which is combination of BS, LF, WM, NBS, and LM are leaf injures.



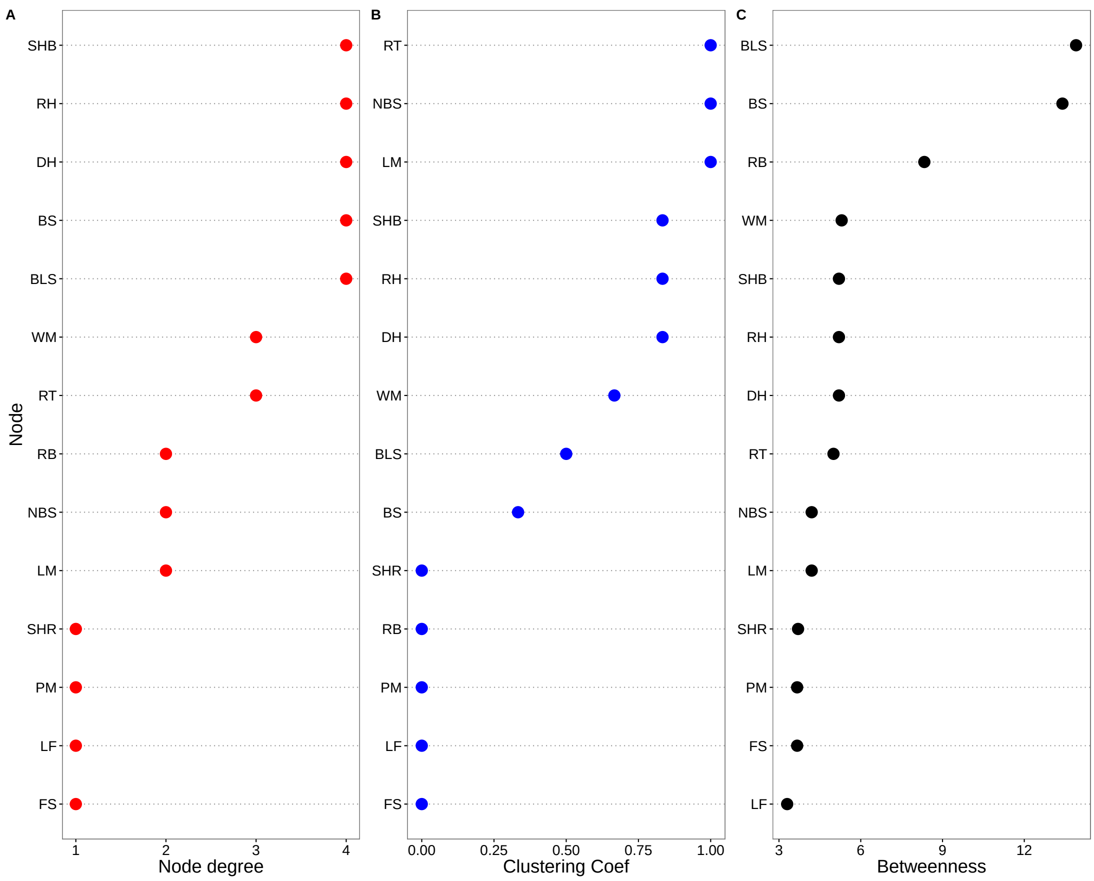
(a) Co-occurrence network of rice injuries in dry season at West Java, Indonesia. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at West Java, Indonesia. A: node degree, B:clustering coefficient, and C:Betweenness.



(a) Co-occurrence network of rice injuries in wet season at West Java, Indonesia. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in wet season at West Java, Indonesia. A: node degree, B:clustering coefficient, and C:Betweenness.

DISCUSSION

Rice injuries caused by animals, and pathogens were found commonly in South and South east Asia, but at different levels of incidence. Injuries indeed depended on locations or climate conditions favorable to develop (Savary et al., 2006). So they were not observed at all production environments or season during survey were conducting. For example, red stripe was often found in Central Plain, and West Java in dry season.

The co-occurrence correlations of rice injuries were explored using network inference based on strong and significant correlations through using non-parametric Spearman's rank coefficient. Usually, correlations were assessed using Pearson correlation. However, the use of the Pearson correlation coefficient is problematic because it requires the variables are applied with similar measure, and the variable values are normally distributed. Additionally, Pearson correlation can only capture linear relationships. Due to the fact that the assumptions of Pearson correlation are not fit with the survey data. The alternative is provided by using Spearman's rank correlation coefficient, which is also widely used in biological, and ecological studies

The exploration of co-occurrence networks is a useful method for determining interactions of co- occurring injuries. The centrality of nodes is considered further. The node centrality is the identification of which nodes are "central" than others (Barrat et al., 2004). Newman (2003) mentioned three measures of node centrality: node degree, clustering coefficient, and betweenness. The node degree is measured by the number of connections a node has. In co-occurrence networks of rice injuries, node degree of each injury was counted from the number of the positive relationships of injuries have with other injuries. The clustering coefficient measure a density of local connectivity. Higher clustering coefficient of an element, the higher is the relationship among their neighbors. In the context of theses co-occurrence networks, nodes with high clustering coefficients were located closely. This indicated that they strongly occur together; one increasingly occurred, the related one also occurred jointly. In biological network, betweenness is used for measuring has been of how central a node is in a network, because a node with high betweenness essentially play an important role as a bridge between different parts of the network (Proulx et al., 2005; Newman, 2010). In this study, nodes with high betweenness could be represented as indicators. Because theses nodes have many connections passed through, they are likely to be induced, and have higher chance to occur before other injuries associated.

Peripheral nodes or low-centrality nodes are also interesting. Ecological studies considered theses nodes as specialists that have a few links and link specially to curtain nodes (Lu et al., 2013; Borthagaray et al., 2014). In the context of co-occurrence networks of injuries, peripheral nodes have a few relationships within their groups in the network. They also slightly depend on other injuries. This may imply that these injuries are difficult to control because they may occur occasionally such as rat injury (RT) in dry season at West Java, Indonesia, neck blast (NB) in wet season, in Tamil Nadu, bacterial leaf streak (BLS) in Red River Delta, bacterial leaf blight (BLB), rat injury (RT) and sheath blight (SHB) in wet season at Odisha, India.

Nodes display high betweenness are suggested that these nodes have important roles in regulating network interactions such as key stone species in ecological network Wright et al. (2012). It can illustrate both the number of connections and how important those connections are to the overall network. Therefore, in the co-occurrence network, I identified injury indicators, which are the injuries are highly sensitive to the favorable conditions for the associated injuries in the networks. For example, in wet season at West Java, the network showed that brown spot is a good indictor of syndrome3, which is comprised of leaf miner injury, whorl maggot injury, narrow brown spot, and leaffolder injury. Compared to other injuries within this group, brown spot shared association to all the injuries, then brown spot potentially can be observed earlier under a curtain condition. When we first found brown spot, there is high chance that related injuries will occur.

It was found that the co-occurrence networks of rice injuries changed with seasons and production environments. The same injuries in the network would connect to different injuries as the change of seasons and production environments. In according with previous findings (Savary et al., 2000b; Avelino et al., 2006; Savary et al., 2012) , the pest and disease syndromes (combination of pest injuries and diseases) are strongly associated with climatic condition at regional scale. In Red River Delta, brown spot had relationships with sheath blight in dry season, but in wet season brown spot had association with bacterial leaf blight.

Community structures in networks can reveal hidden information that is maybe not easy to detect by simple observation. In social studies, community detection has been applied to search the groups of people who are interested in same topics. In this study, I detected node community based on the optimization of the modularity of a sub-network, which is a popular approach (Liu et al., 2014). Communities are groups that are densely connected among their members, and sparsely connected with the rest of the network. Community structure can reveal abundant hidden information about complex networks that is not easy to detect by simple observation. Communities in a co-occurrence network of rice injuries might represent the rice injury co-occurrence association under related conditions. For example, the network of dry season in Central Plain revealed that group2 (LB, WM, LF and BLB). According to Savary et al. (2000b), these injuries were in injuries profile group2 (IN2). They also mentioned that IN2 was related to production situation group2 (PR2), this type was also dominantly found in direct seeded rice fields. So injuries in group2 of the network in dry season in CP could co-occur favorably in rice fields where applied direct seedling method, which is the most common practices in Thailand (GRiSP, 2013).

CONCLUSTION

In order to establish priorities and strategies for pest management program, there is a need for characterization of multiple pests (Mew et al., 2004). I applied network analysis to characterize co-occurrence patterns of rice injuries from crop health survey data, which were collected in farmers' fields at five production environments (Central Plain; Thailand, Odisha; India, Red River Delta; Vietnam, Tamil Nadu; India, and West Java; Indonesia) across South and Southeast Asia for three consecutive years (2013-2015). The resulted networks depicted the co-occurrence patterns of rice injuries at different production seasons, and production environments. The networks revealed the different structure, which reflect to the co-occurrence patterns between injuries, in different production seasons and production environment. From the network structures, networks showed injury syndromes (groups of injuries) that are closely related in the network. Additionally, from three important of node centrality measures (node degree, clustering coefficient, and betweenness), I can identify indicators that are used for monitoring, and predict the trend that associated injuries to occur under a curtain condition that may be favorable for injury indicators. This information is useful to better understand the variation of rice injury occurrence, and to develop the more effective strategies of pest management specifically to seasons or production environments (locations).

CHAPTER II

DIFFERENTIAL NETWORKS REVEAL THE DYNAMICS OF ANIMAL INJURIES AND DISEASE CO-OCCURRENCE DIFFERENTIAL NETWORKS

INTRODUCTION

Rice (*Oryza sativa*) is the most important human food crop in the world, feeding more an extraordinarily high portion of the total planted area in South, Southeast Asia. The Food and Agriculture Organization of the United Nations (FAO) estimates that approximately 70 percent of total lowland rice area produces two rice crops each year, main crop in the wet season, while another is in the dry season. The important role of seasonal cropping in the temporal dynamics of animal pests and diseases has been studied under farmers field survey in South and Southeast Asia by the use of multivariate techniques Savary et al. (2000b); Willocquet et al. (2008). These studies showed that injuries profiles (the combination of injuries) differ from season to season due to weather patterns, which in dry season, crop losses were lower than in the wet season because of lower incidence and severity of pests and diseases (Litsinger, 1991). Additional a previous study based on surveys done in farmers' rice fields in the region of lowland rice were shown that injury profiles were strongly associated with season.

In the previous chapter, the co-occurrence networks showed co-occurrence networks, a methodological approach which has already proved fruitful in a variety of different applications. Plant injuries caused by pests maybe affect yield production. Therefore, in this chapter, I attempted to characterize the patterns of rice injuries by studying the changes in the co-occurrence patterns of rice injuries (*e.g* disease incidence, animal pest injury incidence) in different season and different yield levels.

Differential network analysis aims to compare the connectivity of two nodes at two different conditions. As demonstrated by several studies, differential networks can identify important nodes implicated in my fields, and also provide critical novel insights not obtainable using other approaches. In this work, I explore the the properties of network of a complex association of rice injuries at different yield levels. Elucidating the rice injuries association represents a key challenge, not only for achieving a deeper understanding of injury association (injury profiles) but also for identifying the unique association. Given that the injury association is governed by a complex network of injuries association, it seems natural to explore network properties which may help elucidate some of different association presenting in the different seasons.

In this chapter, I employ a differential network topology method to examine the co-occurrence relationships of rice injuries from survey data. I use graph theory methods to examine the topological feature dynamic of a co-occurrence network corresponding to different seasons, and production environments. The co-occurrence networks were built from differentially co-occurring injuries. I extract significantly differential co-occurring injuries from co-occurrence networks, which represent different seasons, to identify which injuries that may be involved specifically curtain season. I postulate that these selected injuries may contribute to the difference in the co-occurrence patterns in different season. Furthermore, I identified the injuries associated with yield from networks at different yield levels. Finally, I suggest key injuries that may contribute to yield reduction under a curtain production environment. The goal is to leverage insights to better understand the rice injury co-occurrence that may contribute to pest management development.

MATERIALS AND METHODS

Differential co-occurrence network construction

The survey data were pre-processed by using methods described in the previous chapter. Subsequently, I applied the method proposed by Fukushima (2013) to identify differentially co-occurrence links. The difference of co-occurrence of injury x and y between two conditions (A and B) was quantified by Fisher's z-test.

For the pair of x injury and y injury, I denoted the correlation coefficient based on Spearman's correlation coefficient by r_{xy}^{A} and r_{xy}^{A} in networks of condition A and condition B, respectively. To test whether the 2 correlation coefficients were significantly different, correlation coefficients for each of the 2 conditions, r_{xy}^{A} and r_{xy}^{B}, were transformed into Z_{xy}^{A} and Z_{xy}^{B}, respectively.

The Fisher's transformation of coefficient r_{xy}^{A} is defined by

$$Z_{xy} = \frac{1}{2} \log \left[\frac{1 + r_{xy}}{1 - r_{xy}} \right] \tag{II.1}$$

Next, The p-value of the difference in Z_{xy} values was calculated using the standard normal distribution.

$$p(Z \geq \left| \frac{Z_{xy}^{A} - Z_{xy}^{B}}{\sqrt{\frac{1}{N_A - 3} + \frac{1}{N_A - 3}}} \right| \tag{II.2}$$

Next, The p-value of the difference in Z values was calculated using the standard normal distribution

N_A and N_A represent the sample size for each of condition. The Z has an approximately Gaussian distribution under null hypotheses that the population correlations are equal. The pairwise correlation significants are considered at p-value < 0.05.

Differential co-occurrence network in different seasons

Consider any two injuries x and y in the survey data, let r_{xy}^{D} and r_{xy}^{W} be the Spearman's correlation coefficient calculated separately over the samples in dry and wet, respectively. I constructed differential co-occurrence networks that are specified by adjacency matrix $A^{diff} = (A_{xy}^{diff})$ where the entry A_{xy}^{diff} quantified by following:

$$A_{xy}^{diff} = \begin{cases} 1 & \text{when } r_{xy}^{D} > r_{xy}^{W} \text{ at } P_{z_{xy}}\text{-value} < 0.05 \\ 0 & \text{when } P_{z_{xy}}\text{-value} > 0.05 \\ -1 & \text{when } r_{xy}^{W} > r_{xy}^{D} \text{ at } P_{z_{xy}}\text{-value} < 0.05 \end{cases} \tag{II.3}$$

For this differential co-occurrence network,A_{xy}^{diff} equals 1 depending on whether any injury pairs show significantly higher correlation coefficient of co-occurrence in dry season than wet season, but -1 is vice versa, and if it equal 0, meaning that co-occurrence level of injury pairs were not different in dry and wet season. II-1 illustrated the differential co-occurrence network at different seasons.

Difference of co-occurrence network of rice injuries at different yield levels

Consider any two injuries x and y in the survey data, let r_{xy}^{L} and r_{xy}^{H} be the Spearman's correlation coefficient calculated separately over the samples in L and H yield level, respectively. I constructed differential co-occurrence networks that are specified by adjacency matrix $A^{diff} = (A_{xy}^{diff})$ where the entry A_{xy}^{diff} quantified by following:

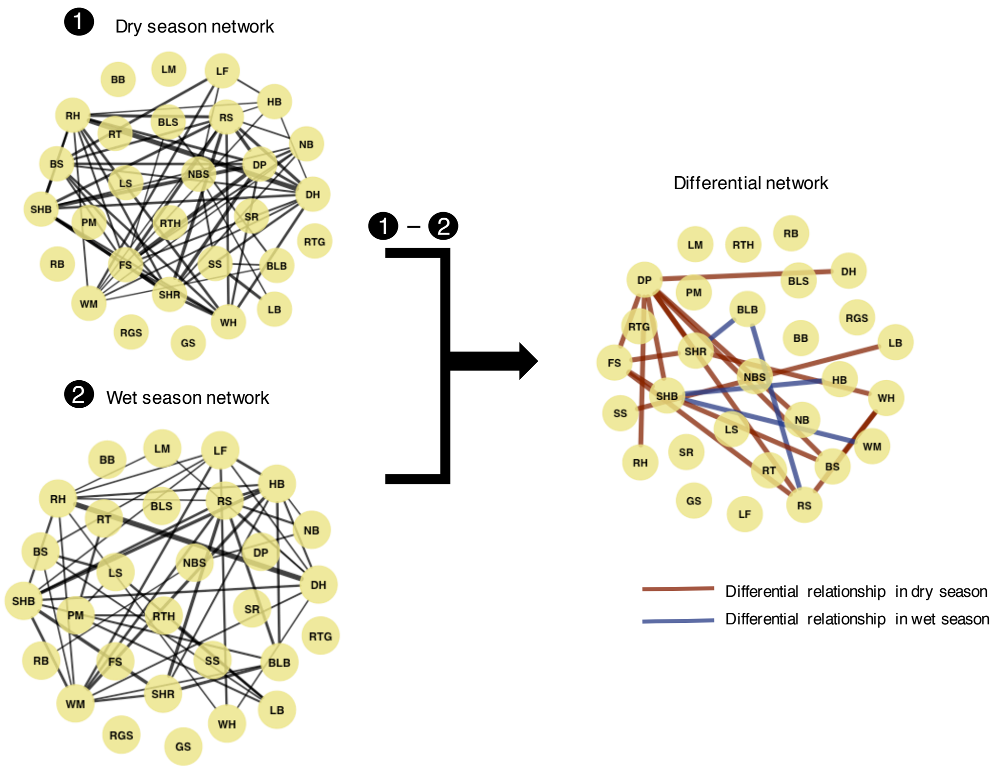


Figure II-1: Schematic showing differential analysis in seasons. Co-occurrence networks are measured in each of two seasons (left) resulting in interactions (black). Dry season is subtracted from wet season to create a differential co-occurrence network (right), in which the significant differential interactions are those that positive (red) or negative (blue) in score after the shift in conditions, which means differential in dry, and wet season, respectively.

$$A_{xy}^{diff} = \begin{cases} 1 & \text{when } r_{xy}^{L} > r_{xy}^{H} \text{ at } P_{z_{xy}}\text{-value} < 0.05 \\ 0 & \text{otherwise} \end{cases} \tag{II.4}$$

For this differential co-occurrence network, A_{xy}^{diff} equals 1 depending on whether any injury pairs show significantly higher co-occurrence level in low yield level than high yield state, and if it equal 0, meaning that co-occurrence level of injury pairs were not or lower different in low yield level state. II-2 illustrated the differential co-occurrence network at different seasons.

Topological properties To investigate the structural properties of differential

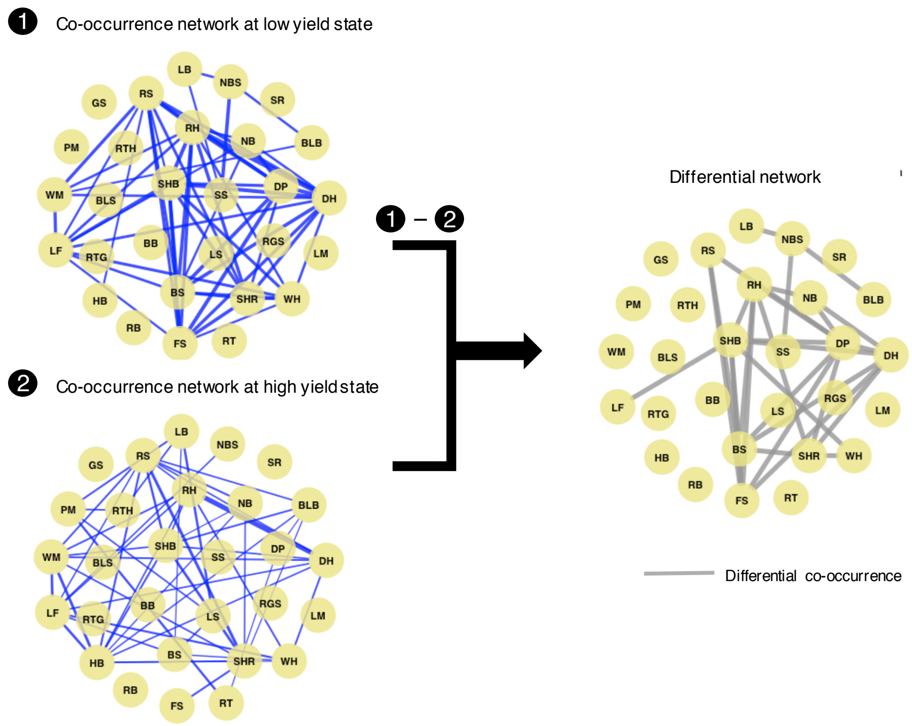


Figure II-2: Schematic showing differential analysis at different yield levels. Co-occurrence networks are measured in each of two different yield levels (left) resulting in interactions (blue). The network at low yield level network is subtracted from high yield level to create a differential co-occurrence network (right), in which the significant differential interactions are those that positive in score after the shift from low to high yield state.

networks, I calculated topological features for each node in the network with the **igraph** package. This feature set included node degree, clustering coefficient, and betweenness.

RESULTS

Differential network approach to constructing response networks enables easy comparison of the generated graphs. This, in turn allowed identification of the differences between the responses of co-occurrence relationships of rice injuries. In particular, in the differential network, I identified injuries and network components that are relevant for responsive condition as well as responsible for condition.

Construction of differential co-occurrence networks of rice injuries at different seasons

I determined differential co-occurrence patterns of rice injuries of survey data in dry and wet season. The Differential co-occurrence network in season (DCON-S), presenting pairs of injuries (nodes) connected with significantly different co-occurrence relationships (edges), are showed in Figure. II-3 to II-7. DCON-S at Central Plain (Figure. II-3) reveals SHB, SHR, and RS showing significantly different co-occurrence in both dry and wet season. They are likely to observed theses injuries. DP, SHB are high-betweenness that have both differential links passed through. This indicates that SHB can present in both dry and wet season, and can co-occur with injuries such as HB and WH in wet season. DP is interring node because it is high betweenness and node degree. DP shows differential co-occurrence only in dry season. Obviously, DP are more likely to be observed in dry than wet season because there are many differential links to increase it. DCON-S at Odisha **??** reveals three groups of injuries shared differential co-occurrences. WM, BS, and SHB present differential links in dry season.

These injuries and their associations may be observed higher in dry than wet season. RH, FS, SS, and DH only show differential relationships in wet season, and they have higher chances to co-occur in wet than dry season. LB different both in dry and wet season. It can co-occur with injuries in both season. LB may present in dry season, if HB occur, and if LM occur in wet season, LB may be found in wet season as well. DCONS at Red River Delta II-5 reveals that DP, BS, RTH, and LF show differential links in dry and wet season. They have different differential co-occurrence patterns. For example, DP may highly occur in wet season, if WH highly is observed in wet season, but in dry season, if BS is highly observed. DCON-S (Figure. II-6) at Tamil Nadu shows two injury pairs express significantly co-occurrence in dry season, which are WH-BS, and DP-RTH, and one injury pair (SHB-LF) apparently occur together in wet season. DCON-S at West Java (Figure.II-7 revealed that BS, NBS, SHB, and BLS form differential co-occurrence in both seasons. They can occur in dry and wet season, and form co-occurrence with some injuries. BLB, DP, LF occurred in dry season, then those injuries may occur. WM, DH, RH presented in wet season, then those injuries could be observed highly in wet season as well.

Differential co-occurrence networks of rice injuries at yield levels

Comparison of the injury co-occurrence networks differed by season and the generic stress response network induced by different yield levels revealed features unique to each graph. In particular, I not only detected the presence of injuries, but I was able to see how they interacted with other genes in their respective networks. The injuries showed in the differential networks are the injuries showed co-occurrence patterns represented by edges. Edges were determined as co-occurrence correlation

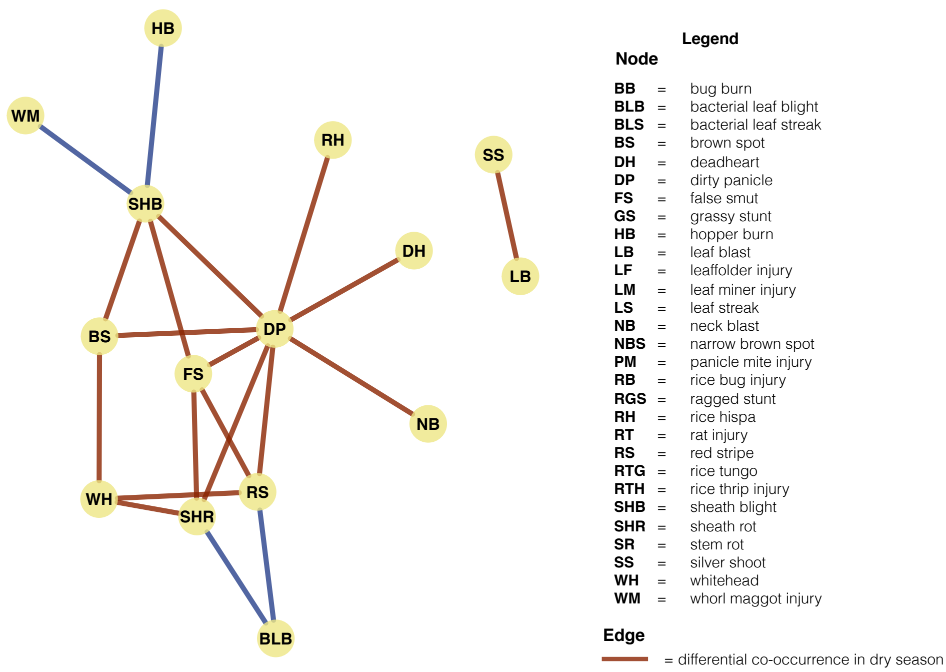


Figure II-3: Differential co-occurrence network of rice injuries in different seasons at Central Plain, Thailand

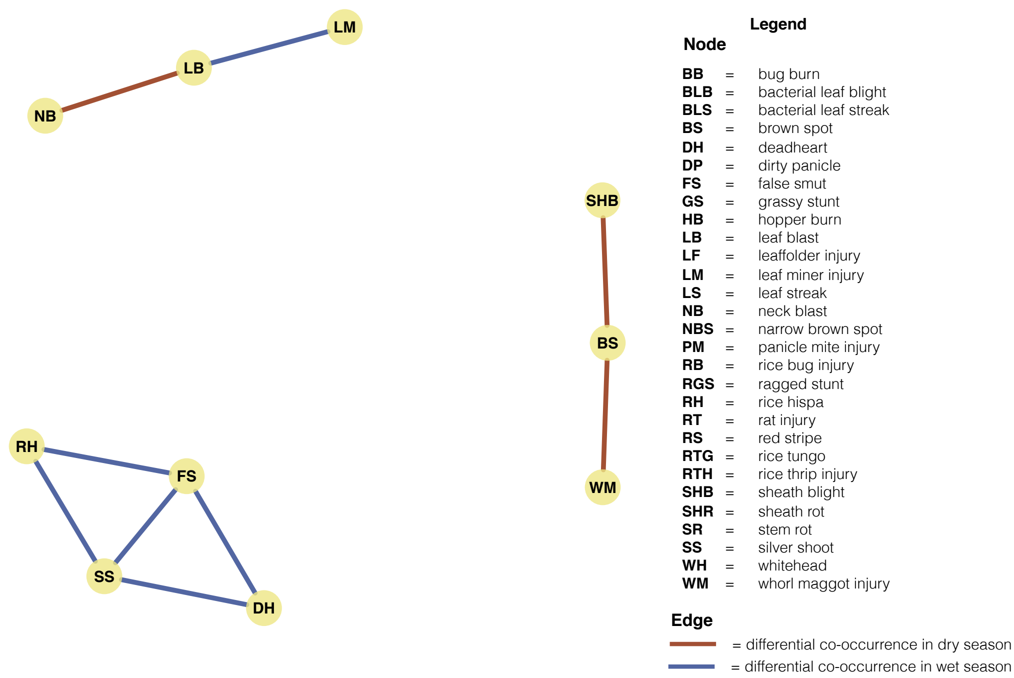


Figure II-4: Differential co-occurrence network of rice injuries in different seasons at Odisha, India

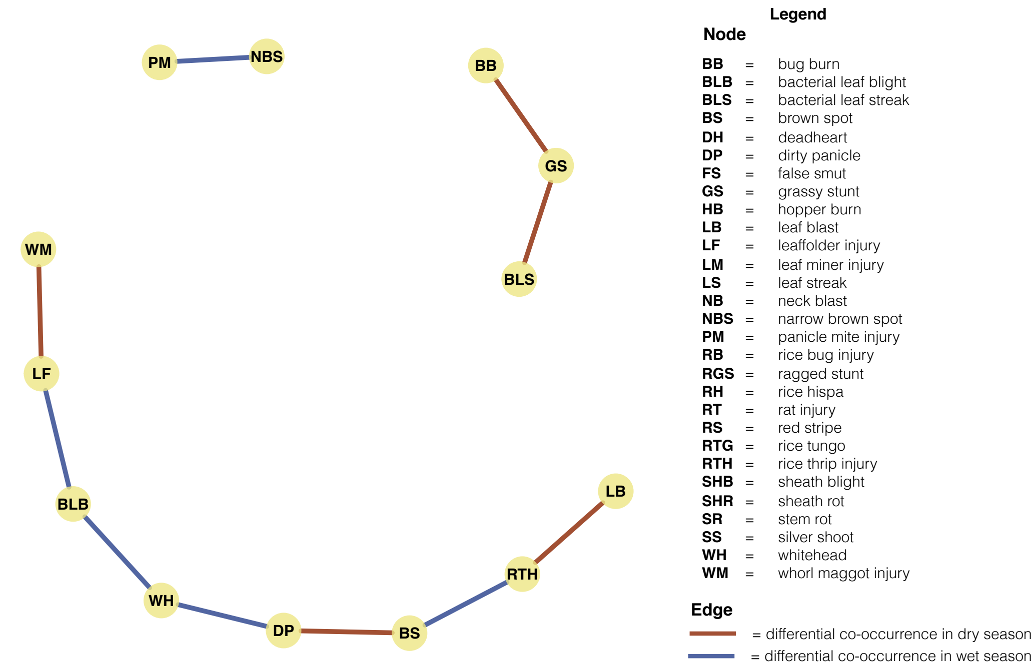


Figure II-5: Differential co-occurrence network of rice injuries in different seasons at Red River Delta, Vietnam

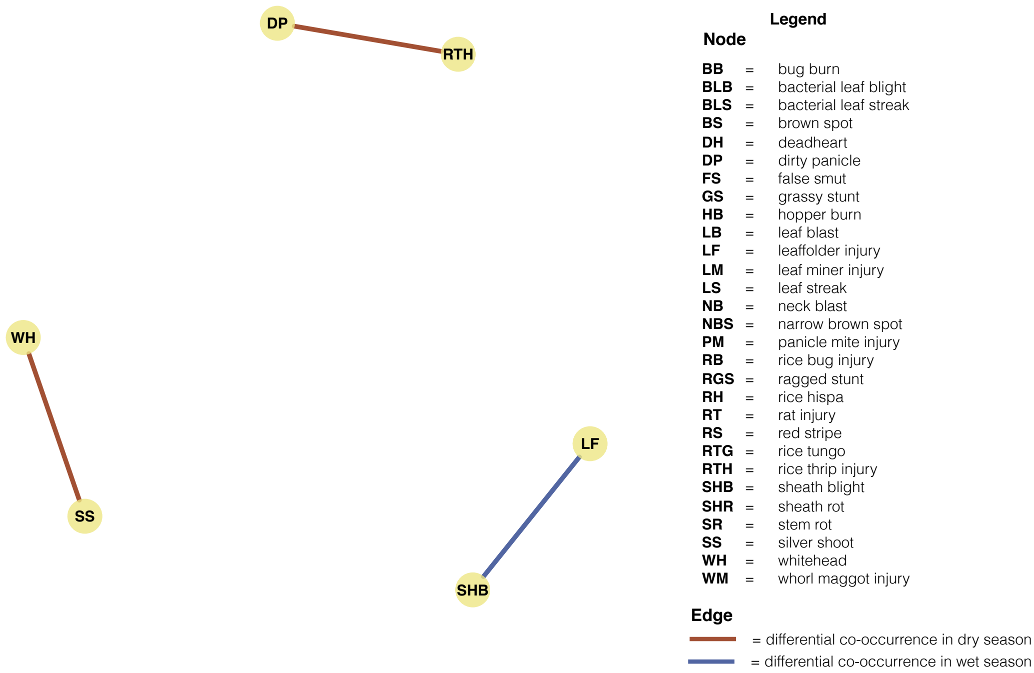


Figure II-6: Differential co-occurrence network of rice injuries in different seasons at Tamil Nadu, India

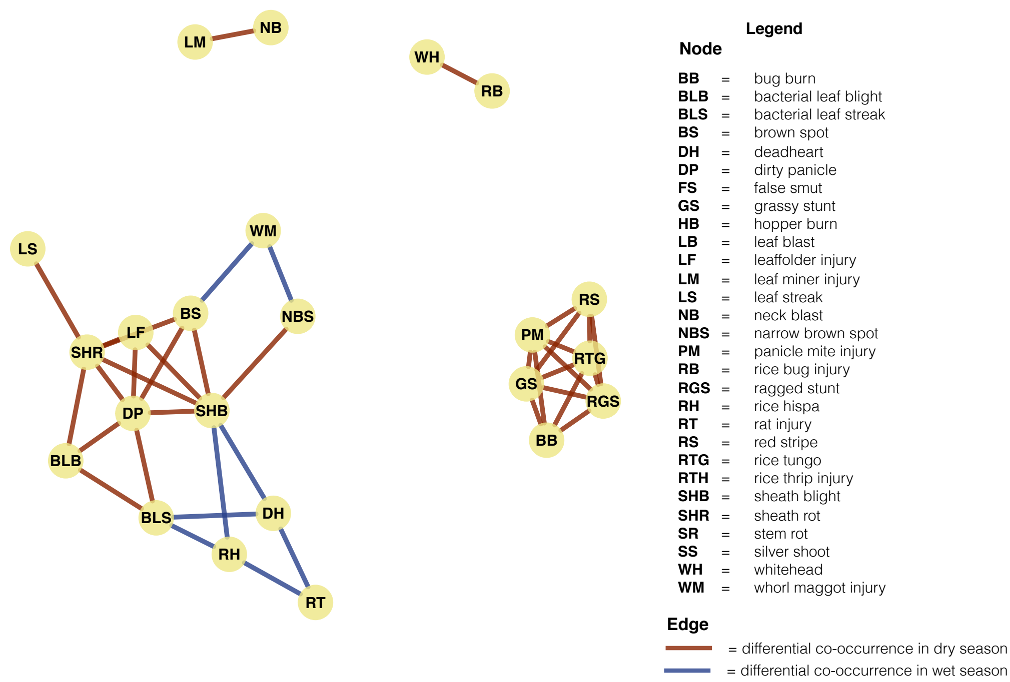


Figure II-7: Differential co-occurrence network of rice injuries in different seasons at West Java, Indonesia that express significantly higher in lower yield level.

In this study, three successive yield classes were defined, in order to enable a better description of actual yield, from low (< 4 ton/ha), medium (4 – 6 ton/ha), high (> 6 ton/ha) yield levels. Figure II-8 shows the number of farmers' fields surveyed classified in each season, and production environment.

Berry and Widder (2014) recommended that a co-occurrence network will be more reliable, it should be produced using a minimum of 25 samples or observations. From figure II-8, to be able compare the networks at different yield levels, I chose the data set, which are medium and high yield level of Central Plain, low and medium yield level of Odisha, medium and high yield level at Red River Delta, low and medium yield level of Tamil Nadu, and medium and high yield level of West Java.

The resulted networks, differential co-occurrence network in yield (DCON-Y), depicts the associations of injury pairs significantly expressing in lower yield state but absent in higher yield state. Figures.II-9a presents DCON-Y between medium and high yield level. It comprised of two groups of related injuries. Form node properties, DP, SHB, BS, RH, FS, and DH are high node degree and betweenness, which are likely to occur in this state because there are many connection, which can increase other injuries related, be induced by them as well. BS, SHB, are DP recorded in survey data showed these injuries were observed higher in medium than high yield data set(Fiure. II-9b). DCON-Y compared between low yield and medium yield level at Odisha (Figure II-10a) shows closely associated injures (DH, RH, FS, BS, WH). Their relationships were captured only in low yield state, and WH, BS, and FS are injuries connecting to all the injuries. Survey data also found that these injuries were observed in high frequency at

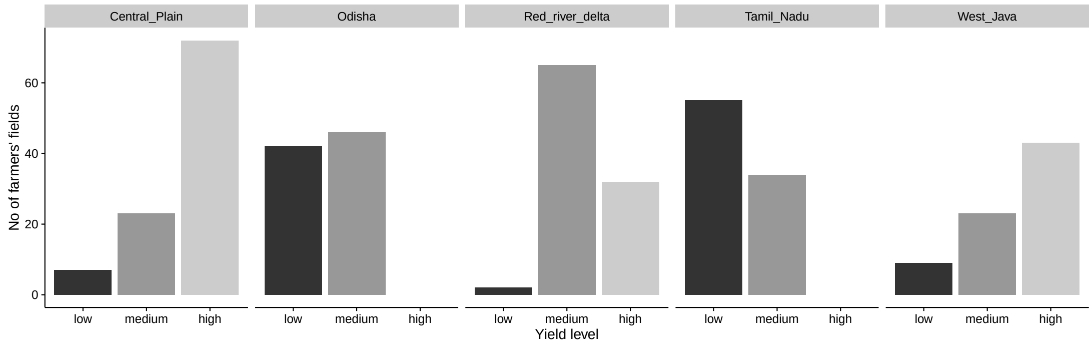
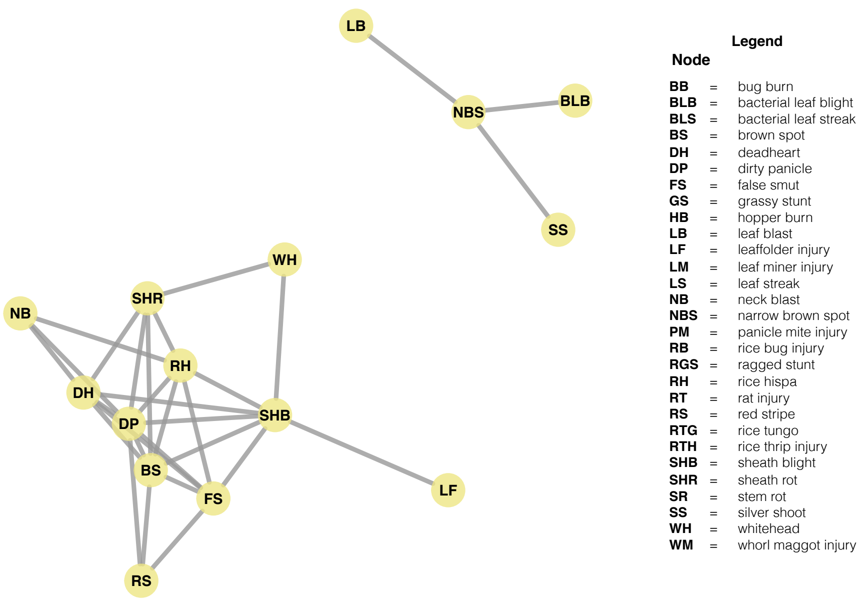
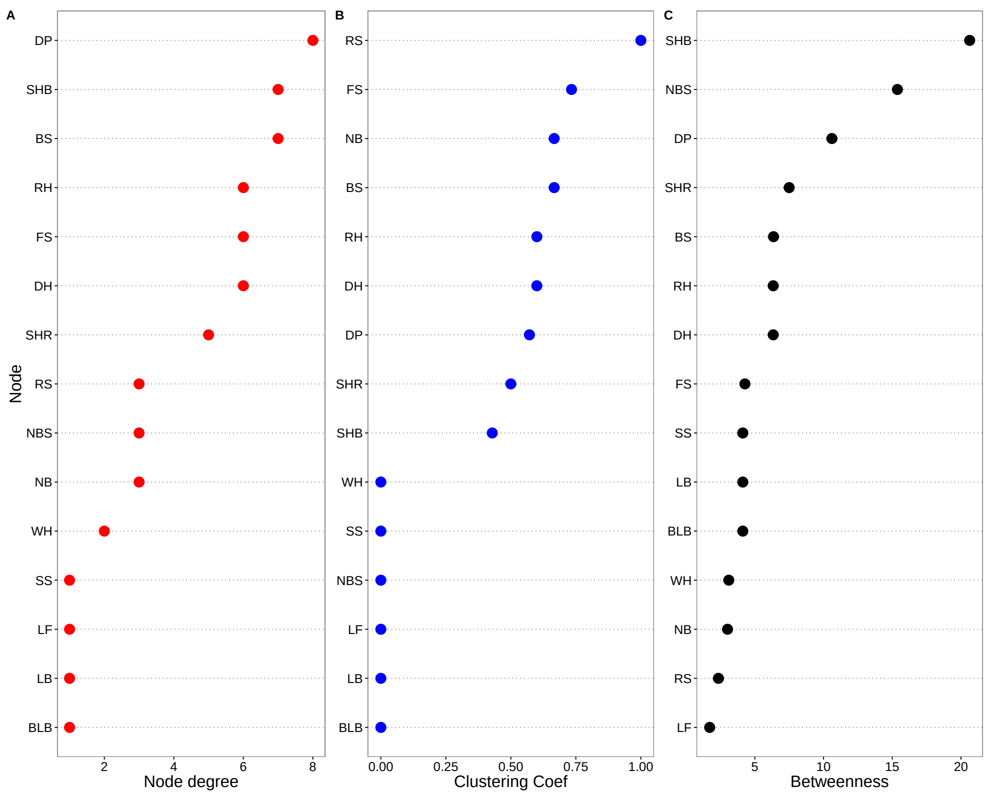


Figure II-8: Bar graphs showing number of farmers' fields classified by different yield levels in each production environment.

high level of incidence (Figure II-10b). In Red River delta, DCON-Y (Figure.II-11a) reveals three injury combinations (DP-WH, GS-BB, and WM-LB-RTH) presenting in medium yield level, not in high yield level. So their co-occurrences may affect to yield losses between medium to high yield levels. Based on the survey data, DP, WM, and WH were found more incidence in low yield than high yield level (Figure. II-11b). Apparently, in Tamil Nadu, Figure.II-12a showed the co-occurrence pattern between SHB and WH that was found only in low yield level state, even though, from survey data, WH were found higher incidence in medium yield level than low yield level, but SHB were observe high incidence in low yield level than medium yield level (Figure **??**). In West Java, DCONY (Figure. II-12a) reveals two groups of associated injuries that significantly express in medium yield level state. Form topological features of this DCON-Y, FS, RS, SHB, and LB are high betweenness nodes, which they are more likely to occur or be observed in this state because there are many pathway to increase them. SHB based on survey data also were highly found in medium than high yield level (Figure.II-13b).

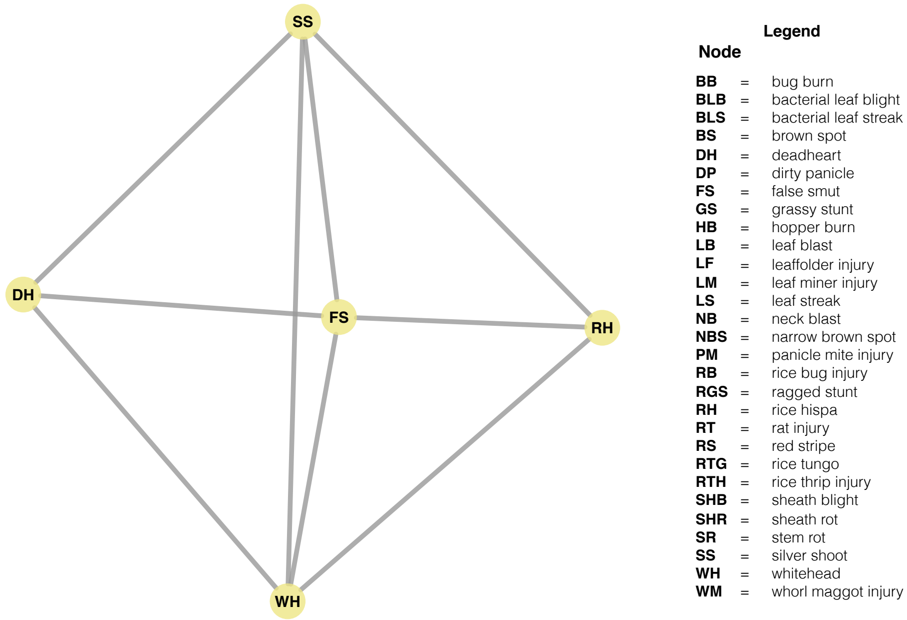


(a) Differential co-occurrence network of rice injuries in different yield levels at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

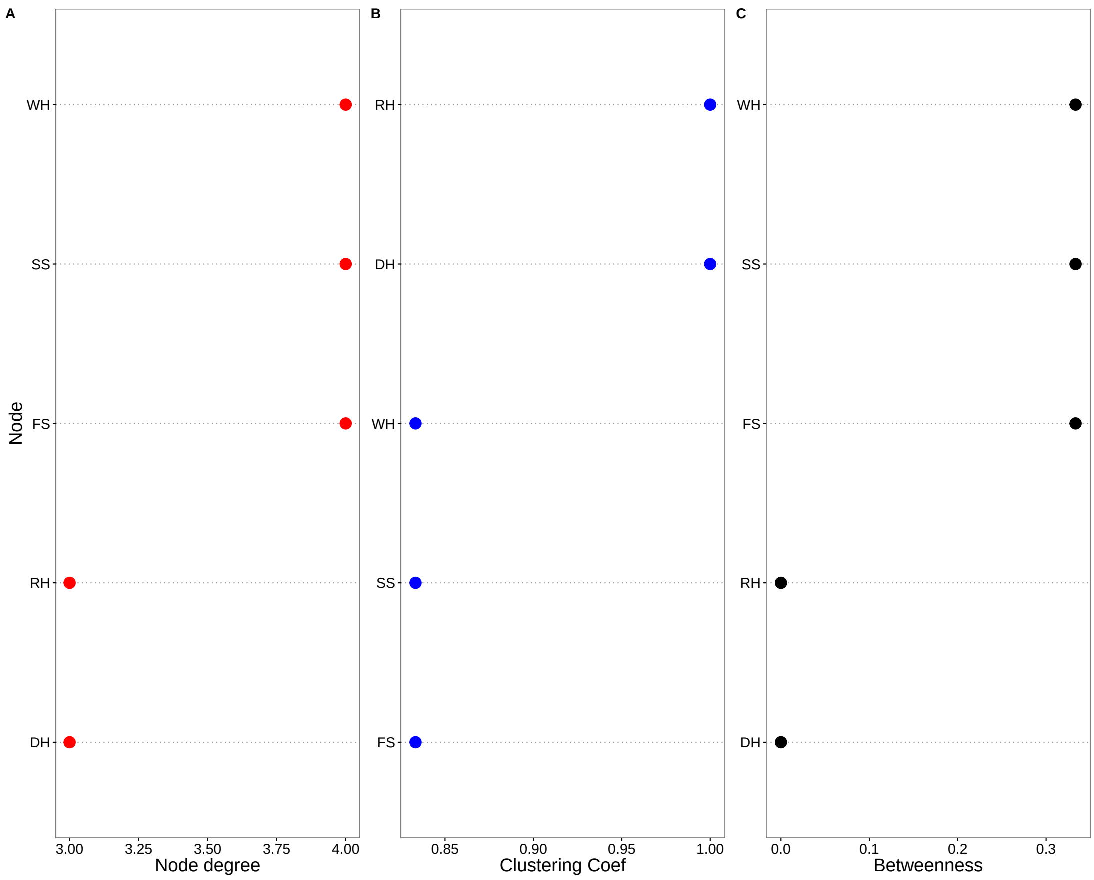


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.

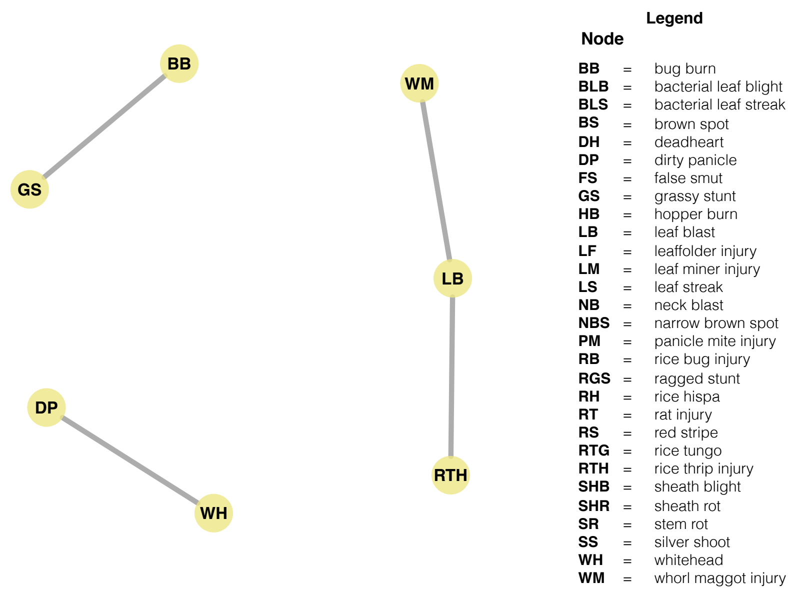
Figure II-9: Differential network analysis of survey data in different yield levels at Central Plain, Thailand.



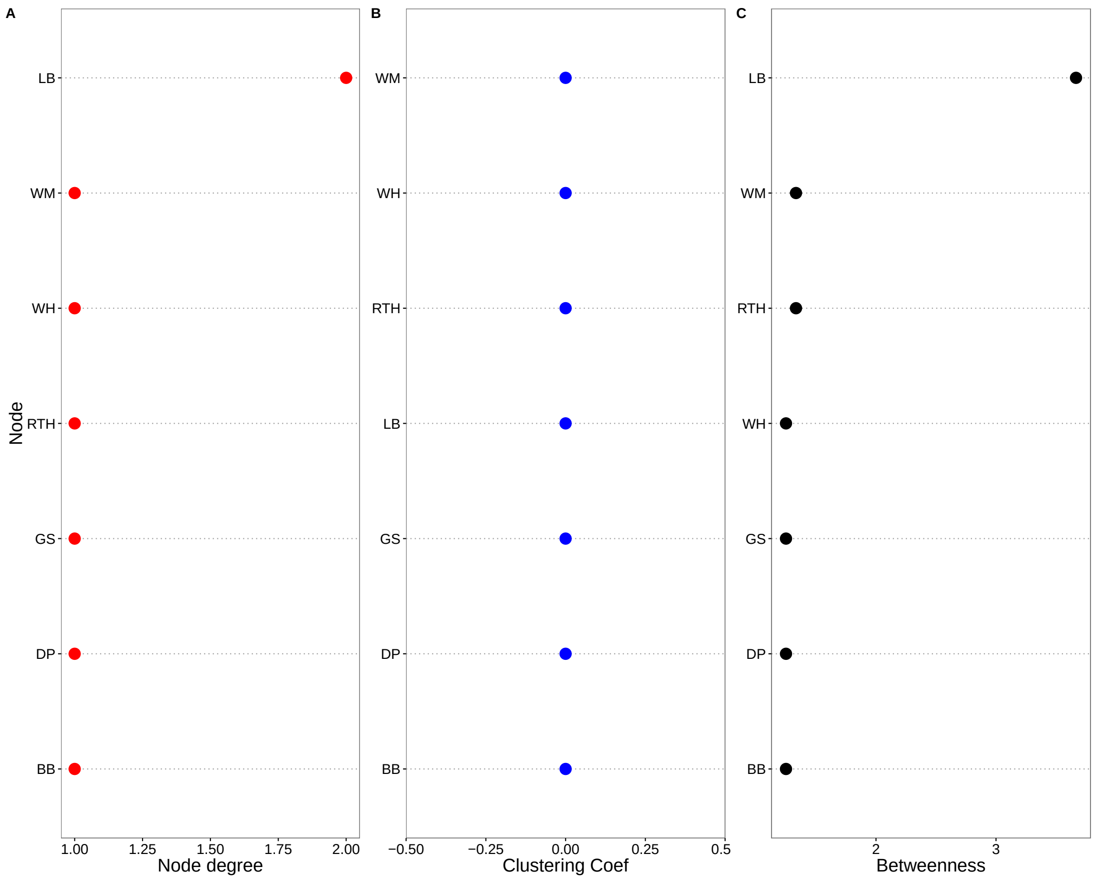
(a) Differential co-occurrence network of rice injuries in different yield levels at Odisha, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



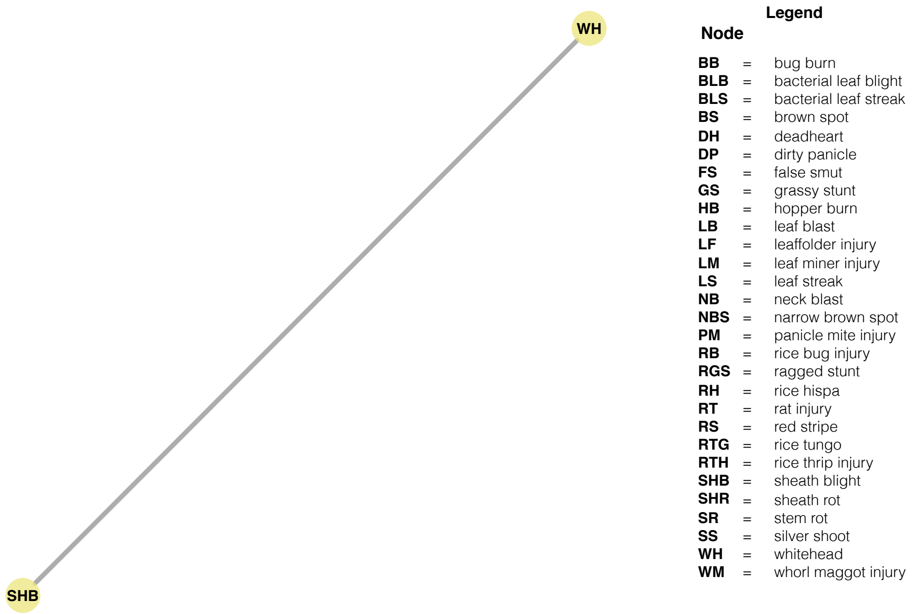
(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Odisha, India. A: node degree, B:clustering coefficient, and C:Betweenness.



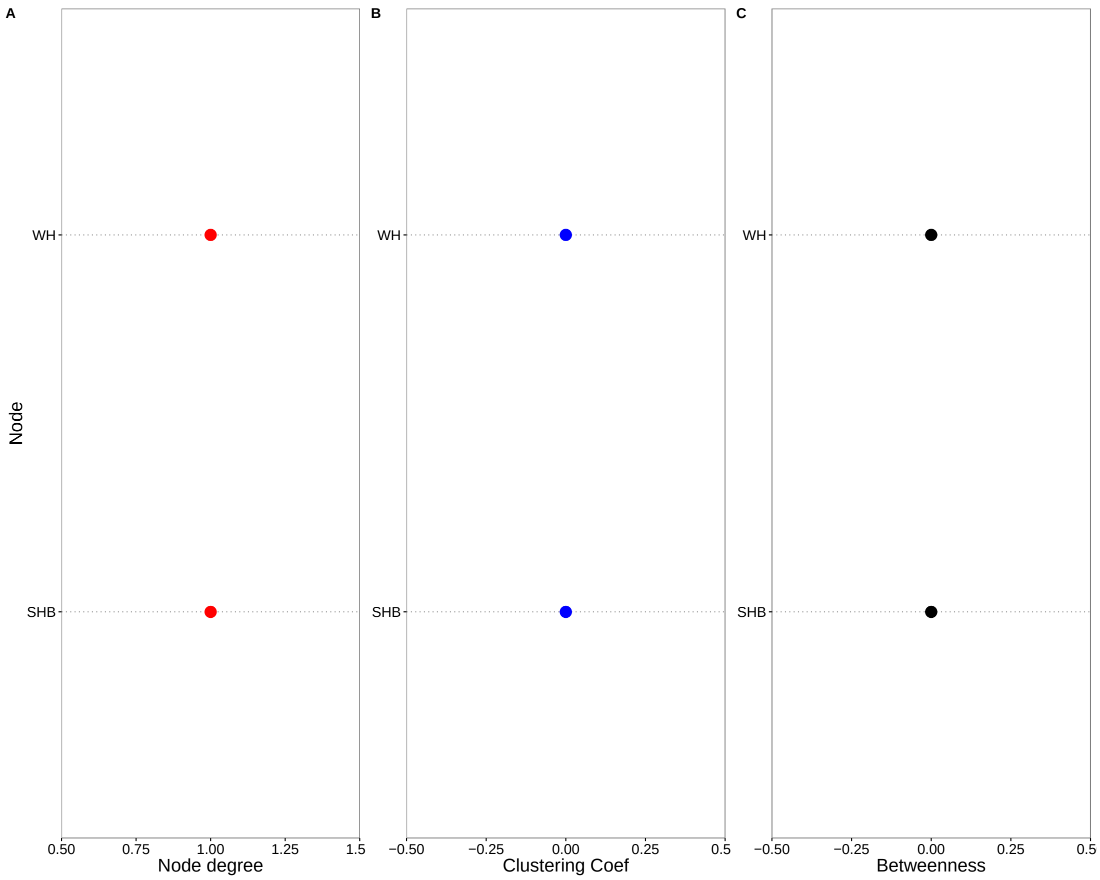
(a) Differential co-occurrence network of rice injuries in different yield levels at Red River Delta, Vietnam. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



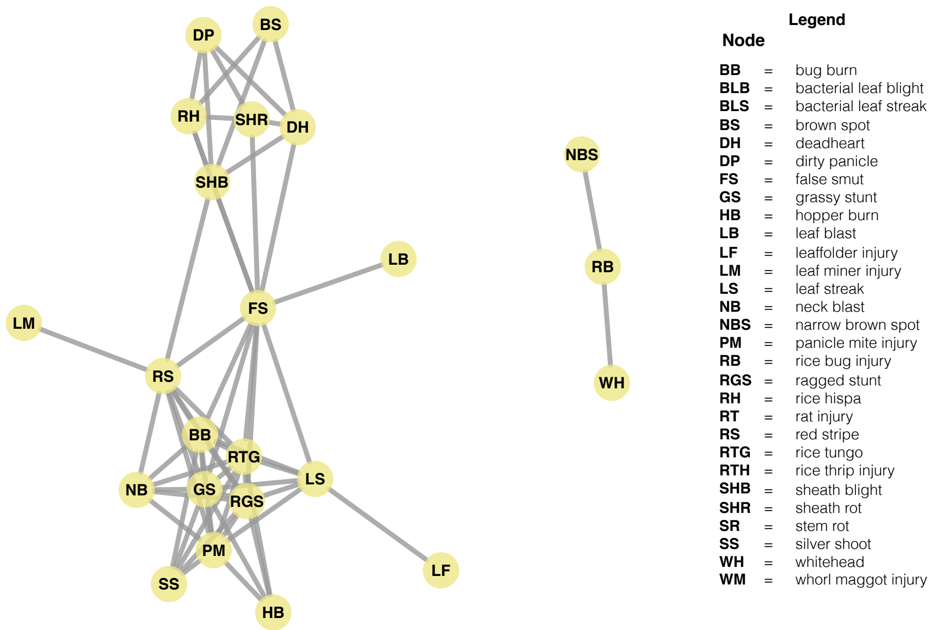
(b) Three centrality measures of the nodes in differential co-occurrence network of rice injuries in different yield levels at Red River Delta, Vietnam. A: node degree, B:clustering coefficient, and C:Betweenness.



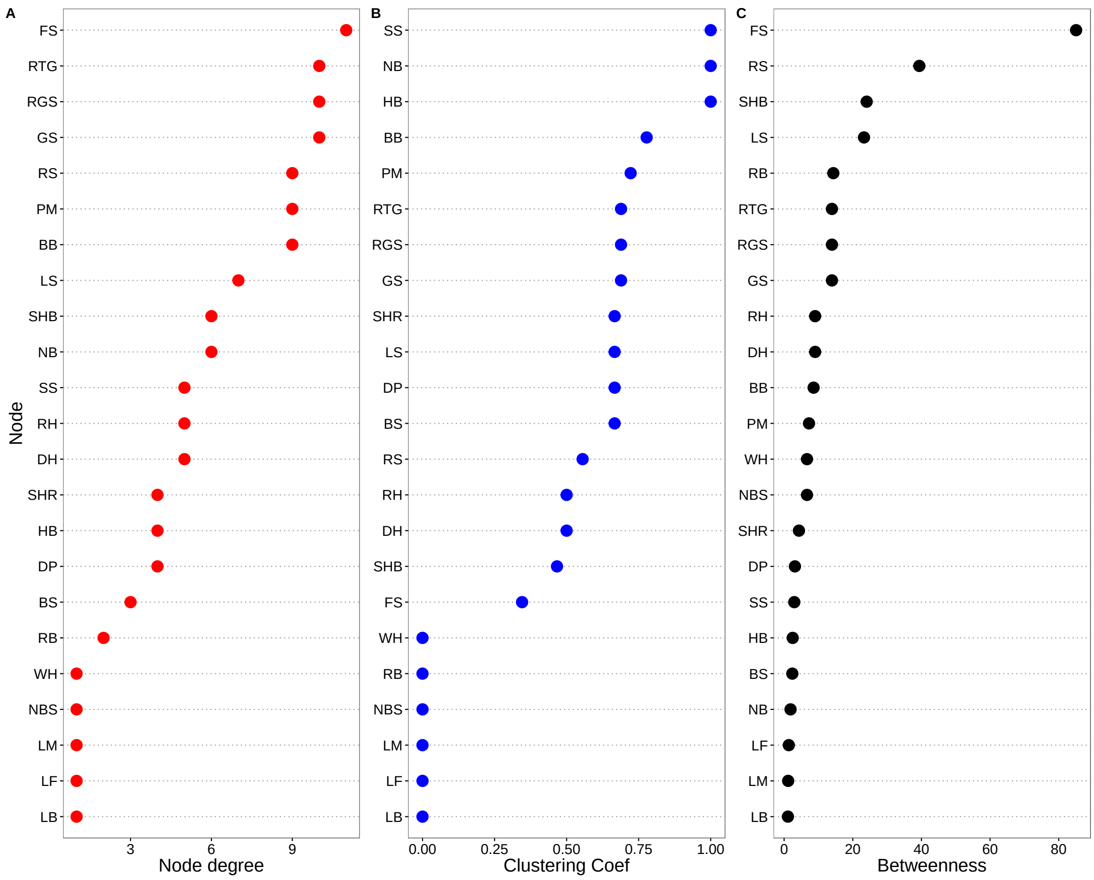
(a) Differential co-occurrence network of rice injuries in different yield levels at Tamil Nadu, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.



(a) Differential co-occurrence network of rice injuries in different yield levels at West Java, Indonesia. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.

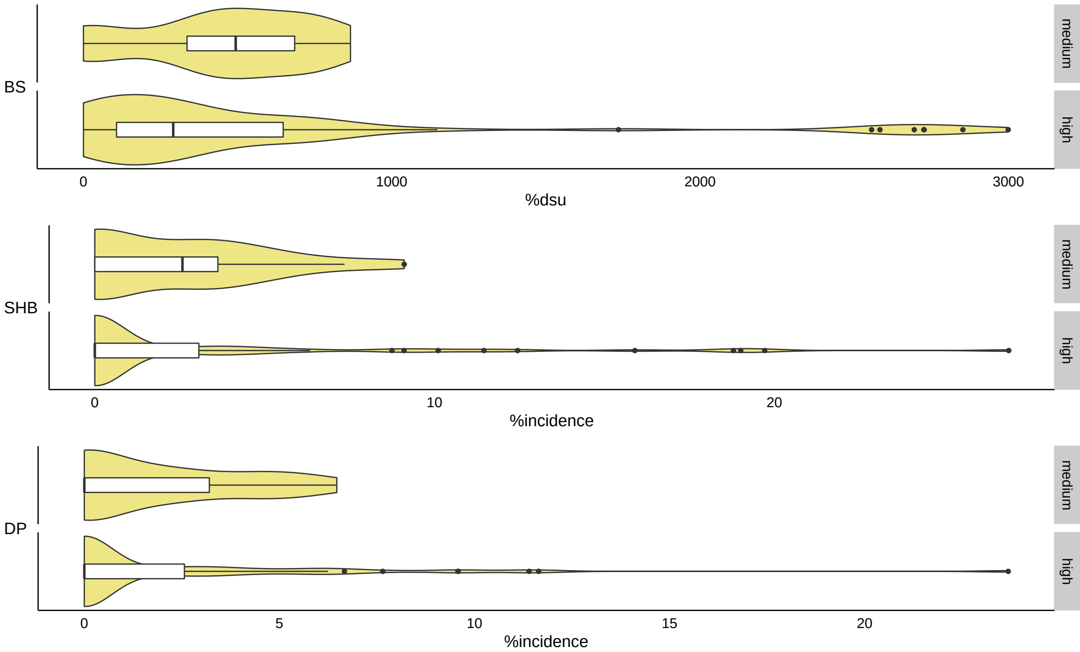


Figure II-14: .

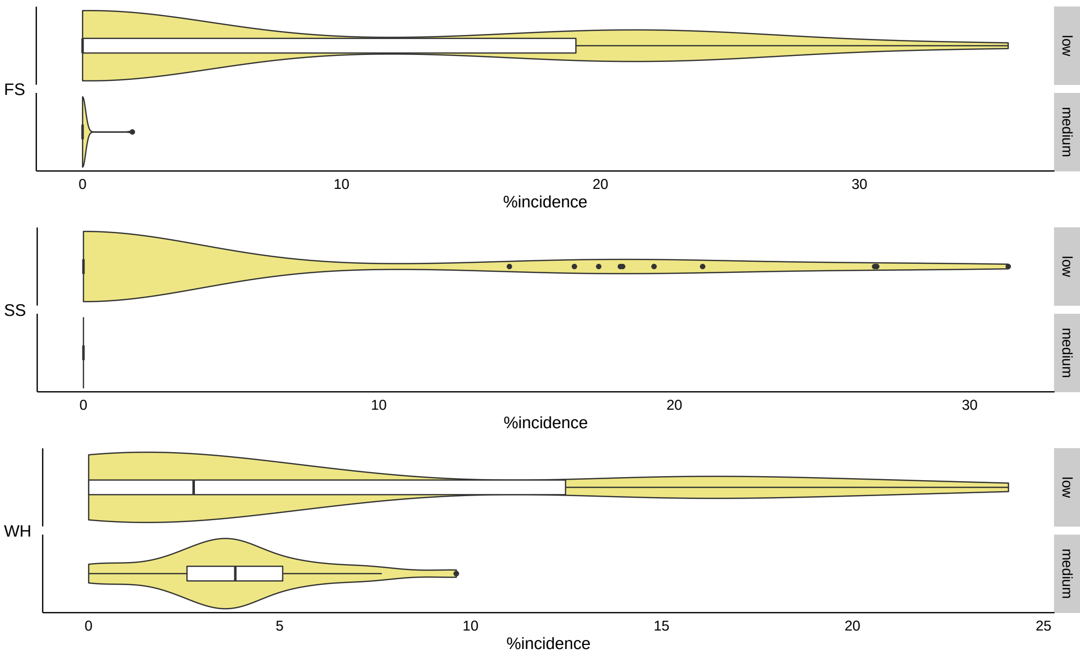


Figure II-15: .

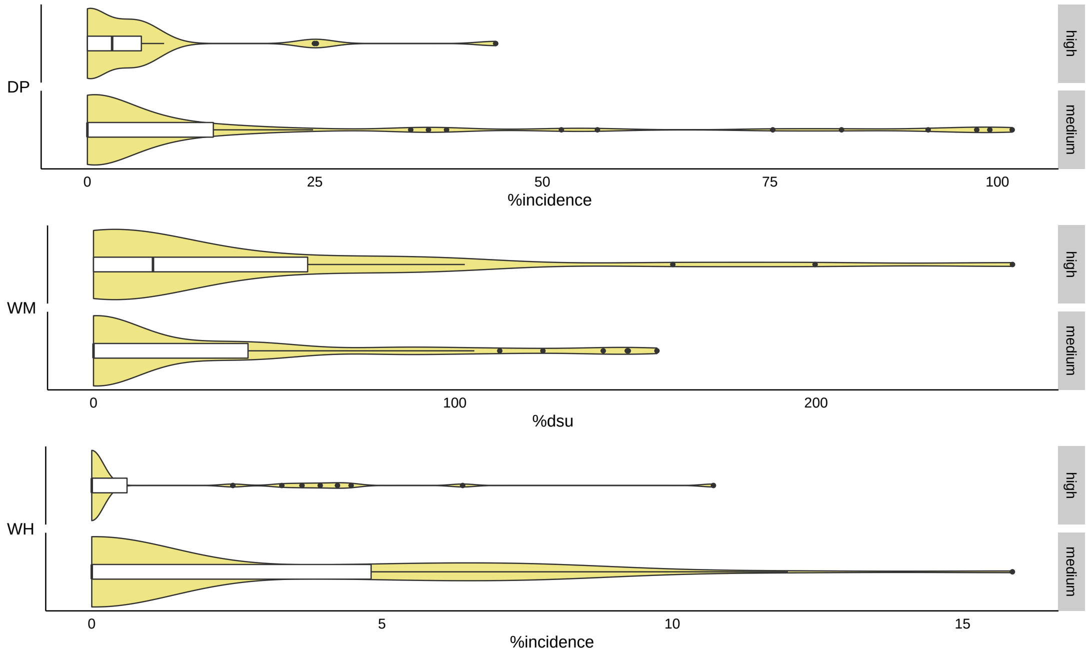


Figure II-16: .

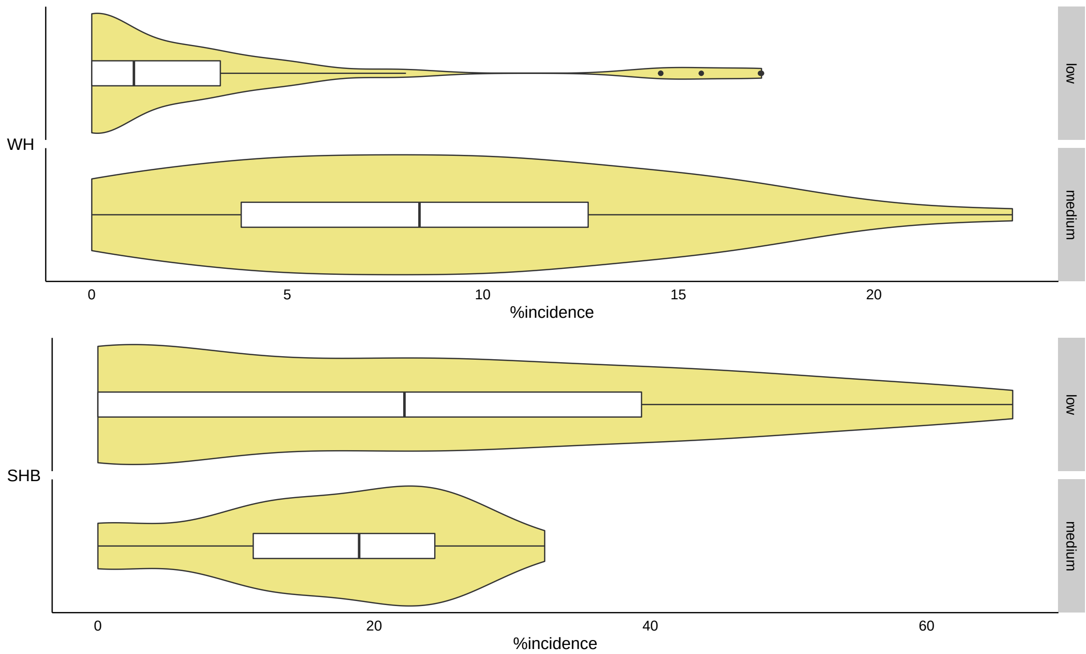
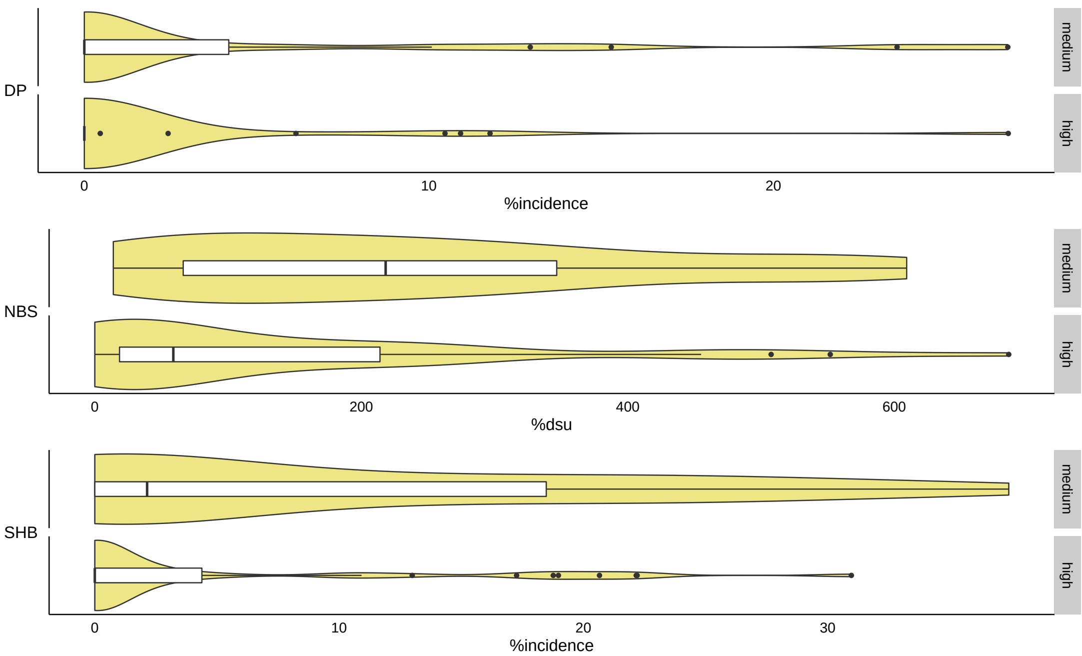


Figure II-17: .



DP
medium
high
0
10
20
%incidence
NBS
medium
high
0
200
400
600
%dsu
SHB
medium
high
0
10
20
30
%incidence

Figure II-18: .

DISCUSSION

From the previous chapter, static dry season and wet season network were dominated by associations of rice injuries that presented in farmers' fields based on survey data. The networks revealed that co-occurrence patterns were different across seasons. I applied differential network analysis based on quantifying the variability in the co-occurrence of rice injuries across seasons. Through network subtraction, however, the associations of injuries presenting in both seasons are removed, which allows sensitive detection of differentially represented conditions. Thus, network comparison reveal the whole patterns of co-occurrence particular tailored to the injury response to seasons.

Further investigation showed that many differential interaction hubs were already well known to function as key components of DNA repair pathways (Fig. 2), which led us to predict that the remaining hubs might encode this role. Two such differential interaction hubs encode Slt2 and Bck1, mitogen-activated protein kinases (MAPKs) that have been implicated in the maintenance of cell wall integrity but not yet linked to DNA repair (Fig. 2). We found that Slt2 is both up-regulated and translocated to the nucleus upon MMS treatment, and it is required for appropriate regulation of ribonucleotide reductase genes in response to DNA damage (fig. S6, A to D) (14). Furthermore, both MAPKs show strong genetic interactions with DNA damage checkpoint genes (fig. S6, E and F), which suggests that they may function in a parallel signaling pathway.

CHAPTER III

SUMMARY, LIMITATIONS AND FUTURE DIRECTIONS

SUMMARY

Paragraph 1

Network analysis has been used in a variety of fields to study relational data, but has yet, until now, to be used for characterizing rice injuries from crop health survey data. The big conceptual hurdle was how to represent the set of rice injuries activities in farmers' fields as a network. The dissertation considers two different approaches to overcome this hurdle. Due to the fact that the main objective is to capture the relationships of rice injuries, the relationships of rice injuries were captured. The present work is largely expository introducing network analysis and showing how it can be applied to possibly better understanding co-occurrence of rice injuries at farmers' fields. Results of network analysis may offer new insight into pest management with another tool to identify important injuries to be monitored or controlled.

This research was divided into two cases based on the approaches for constructing the network. The first case consisted of networks developed based on the co-occurrence relationships of rice injuries under each production seasons and production environments, and the second case consisted of networks developed based on the differential co-occurrence relationships of rice injuries in different production season and yield levels.

But before constructing a network, selecting the suitable correlation methods is important because the method that can identify injuries with true concordance support us to gain knowledge from co-occurrence analysis of survey data. The exploratory analysis of crop health survey data reveals that their values of injuries are not normally distributed, with different measurements. In this study, Spearman's correlation method is the most suitable correlation method because of it robustness to noise and outliers, and ability to accurately capture the interactions. Networks in this study are constructed in three steps. In step 1, selecting data for constructing are obtained. Next, co-occurrence matrix (adjacency matrix) is computed from the selected data using Spearman's correlation methods, then network graph is drawn by connecting injuries (nodes) that have a non-zero entry in the co-occurrence matrix.

The networks, in the first case present links rice injuries (nodes) and co-occurrence relationships (edges) with Fruchterman-Reingold layout, which placed nodes with stronger and/or more connections closer to each other. Next, the topology (structure) of the network was examined including degree, betweenness, and local clustering coefficient. Paths through the network are routes between nodes via the links. Node degree is defined as the number of links a node has. Betweenness quantify how many shortest paths go through each node. The local clustering coefficient is a measure of the degree to which nodes tend to cluster together. It is defined as how often a node forms a triangle with its direct neighbors, proportional to the number of potential triangles the relevant node can form with its direct neighbors. These measures can be as unified criteria for identifying the importance injuries to be monitored and controlled. For example, nodes with high node degree and betweenness are highly potential involved in the rice fields because they have many connections, and highly co-occur with many injuries, and they can be activated very easily, since a lot of co-occurrence activities flows through them (high betweenness). In this case, networks reveal syndromes ("communities" in term of network) is the group of injuries that relatively closely co-occur with each other. Networks shown in Figure to represent the injury syndromes in the same color.

The second case, differential networks are used for identifying a set of injuries whose significantly changes across two conditions. Therefore, theses network show injuries which are significantly co-occur in specific condition, but not in others. Differential co-occurrence networks of rice injuries in season (dry versus wet season) and in yield level (lower versus higher yield level) are constructed. The topological structures of these networks were again probed using various centrality measures in order to identify the most significant injuries in each network.

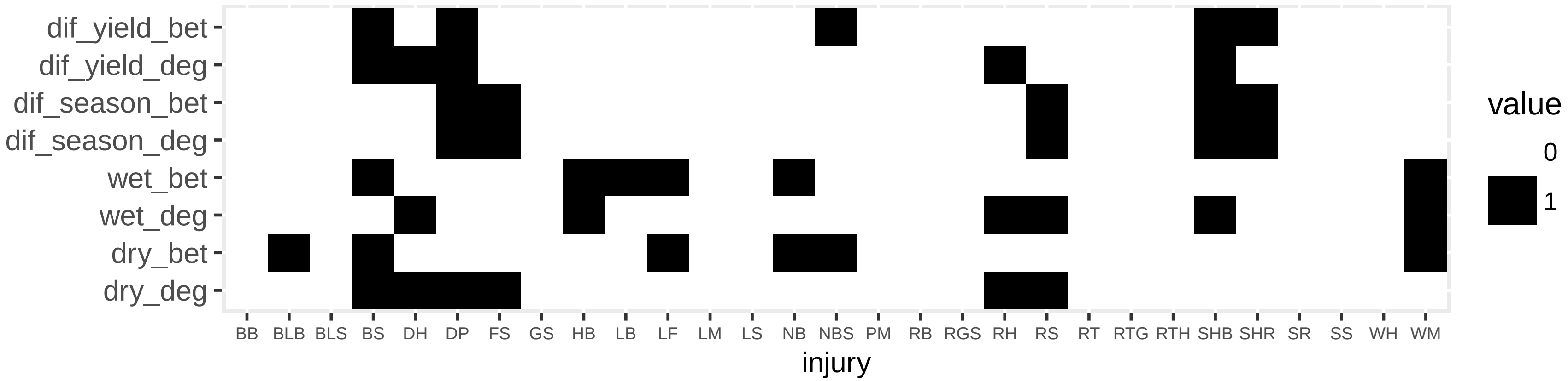


Figure III-1: Central Plain

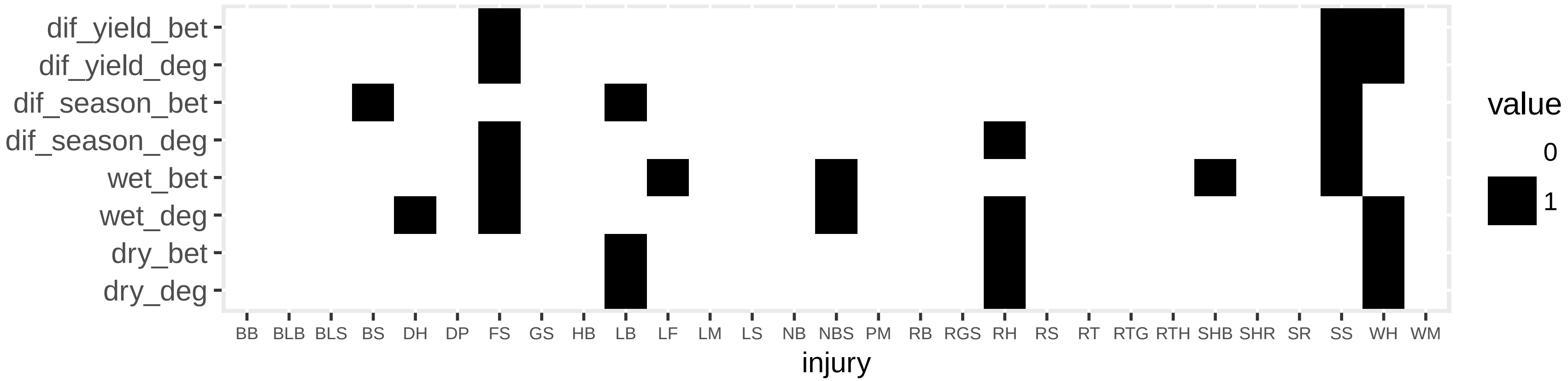


Figure III-2: Odisha

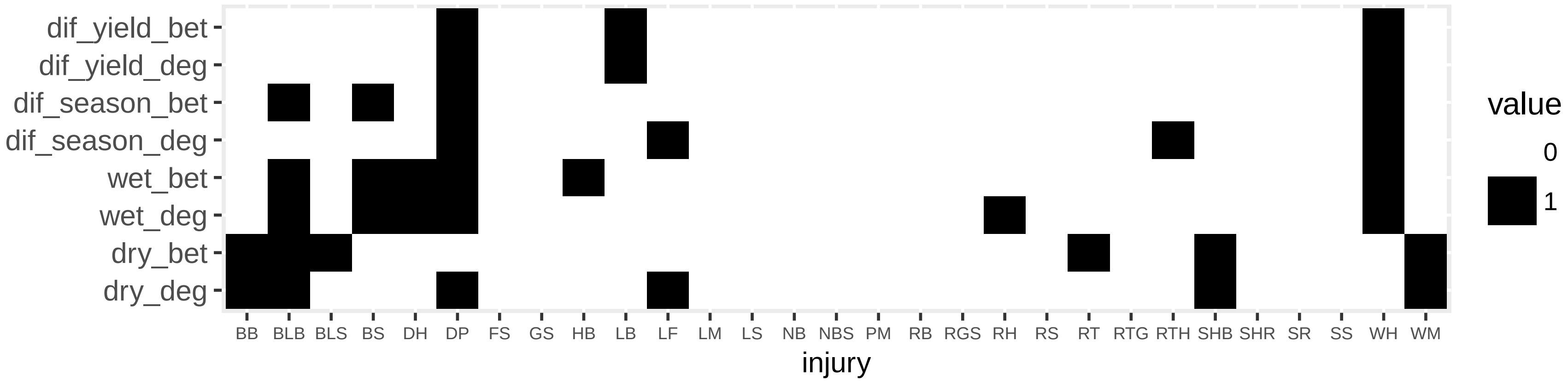


Figure III-3: Red River Delta

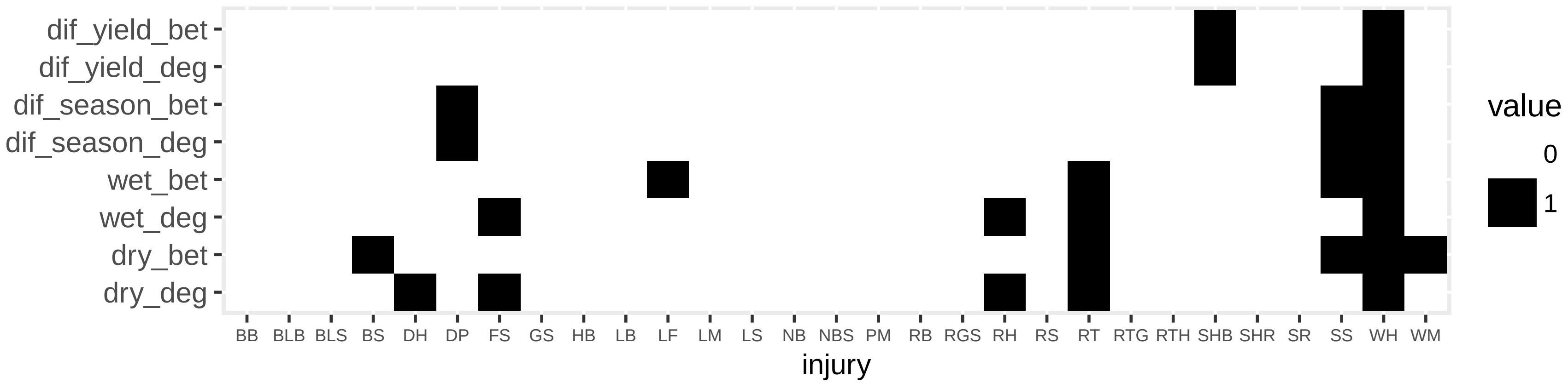


Figure III-4: Tamil Nadu

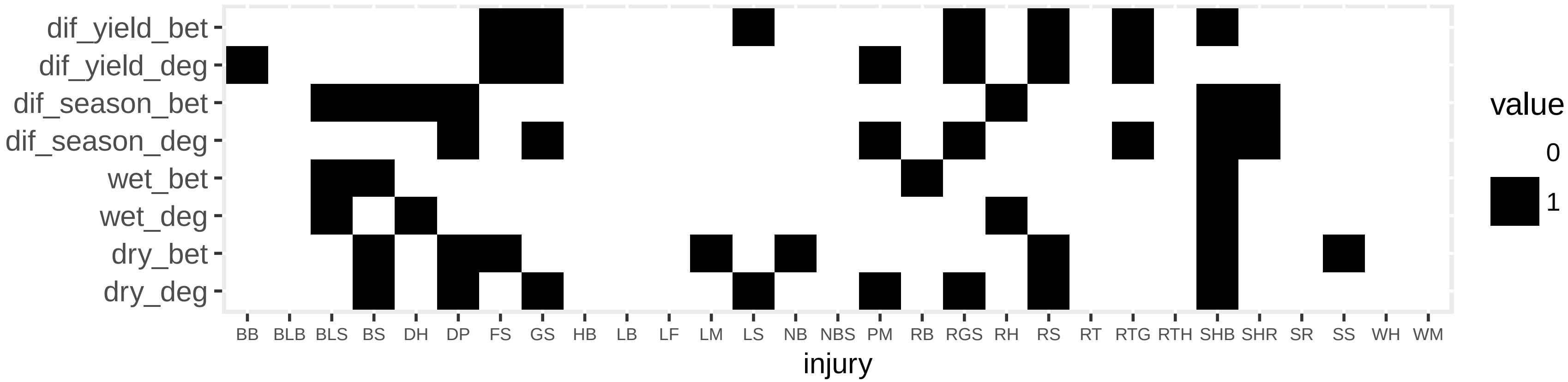


Figure III-5: West Java

LIMITATIONS AND FUTURE DIRECTION

This dissertation answered the question of how to create a network from a list of hurricane occurrences. However the two approaches taken resulted in undirected networks. That is, the links between nodes do not have a direction of relationship meaning the relationship is symmetric. Consider a hurricane that hits Florida before moving west and hitting Texas. In the above network analysis Texas and Florida are linked in an undirected way. However, since Florida gets hit first it might make sense to create a directed network where the link between Texas and Florida points in the direction of Texas. Similar analysis as was done on the undirected network could then be performed on the directed network with additional characteristics arising about the in and out-node linkages. However, the relatively small number of directed linkages may preclude an in-depth analysis using directed networks. One way around the relatively small sample size of 158 years is to include older historical data in the analysis. Recent work by Chenoweth (2006) to collate historical archives of tropical cyclones in the western North Atlantic basin back to 1700 is particularly relevant in this regard. Perhaps better still is to include data from paleotempestology studies (Liu and Fearn 1993; Donnelly and Woodruff 2007). Paleotempestology is the study of prehistoric storms from geological and biological evidence. Coastal wetlands and lakes are subject to overwash events during hurricanes, when barrier sand dunes are overtopped by storm surge. A sediment core taken from the bottom of a near-coastal lake or marsh will record these episodic events as sand layers between the organic peat. As Liu (2007) points out, each record serves as a type of climate station. Connecting these various stations together using network analysis could help better understand the relationship between hurricanes and climate (Elsner 2007). As mentioned in Chapter 4 two regions affected at least once over the period of record by the same hurricane results in a linkage. Information about how frequently this occurs is lost from the type of network analysis. Future research could consider links with various “strengths” depending on how often regions are both affected. The extra information provided by linkage strength would likely add to a better understanding of the hurricane network. Another direction would be to build prediction models of network structure based on pre–season climate conditions. One way this could be achieved is using a Bayesian network (or a belief network). A Bayesian network (or a belief network) is a probabilistic graphical model that represents a set of variables and their probabilistic independencies. Because a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries about them. For example, the network can be used to find out updated knowledge of the state of a subset of variables when other variables (the evidence variables) are observed. This process of computing the posterior distribution of variables given evidence is called probabilistic inference. The posterior gives a universal sufficient statistic for detection applications, when one wants to choose values for the variable subset which minimize some expected loss function, for instance the probability of decision error. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. In terms of the hurricane network given the landfall location, the network could be used to compute the probabilities of the same hurricane occurring in another location. For the visibility network given the visibility of one year, the network could be used to compute the probabilities of visibility in a future year. Finally is would be possible to construct visibility networks from the covariate data including the NAO, SST, and ENSO. The structure of these networks could then be compared with the structural properties of the hurricane network. Then each of those network structures could then be examined using local and global metrics. This investigation may provide further insight into how the indices are intra related in addition to being compared and/or conditioned against/on one another.

LITERATURE CITED

AVELINO, J., ZELAYA, H., MERLO, A., PINEDA, A., ORDOÑEZ, M., and SAVARY, S. 2006. The intensity of a coffee rust epidemic is dependent on production situations. Ecological Modelling, 197(3-4):431–447.

BARBERÁN, ALBERT, BATES, T, S., CASAMAYOR, E. O., and FIERER, N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME journal, 6(2):343–351.

BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R., and VESPIGNANI, A. 2004. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11):3747–3752.

BERRY, D. and WIDDER, S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol, 5(219):10–3389.

BORTHAGARAY, A. I., ARIM, M., and MARQUET, P. A. 2014. Inferring species roles in metacommunity structure from species co-occurrence networks. Proceedings of the Royal Society of London B: Biological Sciences, 281(1792):20141425.

BRANDES, U., DELLING, D., GAERTLER, M., GÖRKE, R., HOEFER, M., NIKOLOSKI, Z., and WAGNER, D. 2008. On modularity clustering. Knowledge and Data Engineering, IEEE Transactions on, 20(2):172–188.

CROWSTON, K. and HOWISON, J. 2006. Hierarchy and centralization in free and open source software team communications. Knowledge, Technology & Policy, 18(4):65–85.

CSARDI, G. and NEPUSZ, T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems:1695.

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(6):603–611.

FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J., and HUTTENHOWER, C. 2012. Microbial co-occurrence relationships in the human microbiome. PLoS Computational Biology, 8(7):e1002606–e1002606.

FRUCHTERMAN, T. M. and REINGOLD, E. M. 1991. Graph drawing by force-directed placement. Software: Practice and experience, 21(11):1129–1164.

FUKUSHIMA, A. 2013. Diffcorr: an r package to analyze and visualize differential correlations in biological networks. Gene, 518(1):209–214.

GRiSP 2013. Rice almanac. International Rice Research Institute (IRRI), 4 edition.

KASARI, C., LOCKE, J., GULSRUD, A., and ROTHERAM-FULLER, E. 2011. Social networks and friendships at school: comparing children with and without ASD. Journal of Autism and Developmental Disorders, 41(5):533–544.

LITSINGER, J. A. 1991. Crop loss assessment in rice. In Rice insects: management strategies, pages 1–65. Springer.

LIU, W., PELLEGRINI, M., and WANG, X. 2014. Detecting communities based on network topology. Scientific reports, 4.

LU, L., YIN, S., LIU, X., ZHANG, W., GU, T., SHEN, Q., and QIU, H. 2013. Fungal networks in yield-invigorating and -debilitating soils induced by prolonged potato monoculture. Soil Biology and Biochemistry, 65:186–194.

MA, B., WANG, H., DSOUZA, M., LOU, J., HE, Y., DAI, Z., BROOKES, P. C., XU, J., and GILBERT, J. A. 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. The ISME Journal.

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(2):103–127.

NEWMAN, M. 2010. Networks: an introduction. OUP Oxford.

NEWMAN, M. E. J. 2003. The structure and function of complex networks. SIAM review, 45(2):167–256.

NEWMAN, M. E. J. 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23):8577–8582.

OPSAHL, T., AGNEESSENS, F., and SKVORETZ, J. 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3):245–251.

PROULX, S., PROMISLOW, D., and PHILLIPS, P. 2005. Network thinking in ecology and evolution. Trends in Ecology & Evolution, 20(6):345–353.

R Core Team 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

REDDY, C., LAHA, G., PRASAD, M., KRISHNAVENI, D., CASTILLA, N., NELSON, A., and SAVARY, S. 2011. Characterizing multiple linkages between individual diseases crop health syndromes, germplasm deployment, and rice production situations in India. Field Crops Research, 120(2):241–253.

SAVARY, S., HORGAN, F., WILLOCQUET, L., and HEONG, K. L. 2012. A review of principles for sustainable pest management in rice. Crop Protection, 32:54–63.

SAVARY, S., TENG, P. S., WILLOCQUET, L., and NUTTER, F. W. 2006. Quantification and Modeling of Crop Losses: A Review of Purposes. Annu. Rev. Phytopathol., 44(1):89–112.

SAVARY, S., WILLOCQUET, L., ELAZEGUI, F. A., CASTILLA, N. P., and TENG, P. S. 2000a. Rice Pest Constraints in Tropical Asia: Quantification of Yield Losses Due to Rice Pests in a Range of Production Situations. Plant Disease, 84(3):357–369.

SAVARY, S., WILLOCQUET, L., ELAZEGUI, F. A., TENG, P. S., DU, P. V., ZHU, D., TANG, Q., HUANG, S., LIN, X., SINGH, H. M., and SRIVASTAVA, R. K. 2000b. Rice Pest Constraints in Tropical Asia: Characterization of Injury Profiles in Relation to Production Situations. Plant Disease, 84(3):341–356.

WILLIAMS, R. J., HOWE, A. C., and HOFMOCKEL, K. S. 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Frontiers in microbiology, 5:358.

WILLOCQUET, L., AUBERTOT, J., LEBARD, S., ROBERT, C., LANNOU, C., and SAVARY, S. 2008. Simulating multiple pest damage in varying winter wheat production situations. Field Crops Research, 107(1):12–28.

WRIGHT, J. J., KONWAR, K. M., and HALLAM, S. J. 2012. Microbial ecology of expanding oxygen minimum zones. Nature Reviews Microbiology, 10(6):381–394.