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

CHAPTER **PAGE**

.1 Evaluation of correlation methods for co-occurrence network construction of rice crop health survey data 1

Introduction 1

Materials and methods................................. 2

Result 7

Discussion... 12

Conclusions 14

I **USING NETWORK ANALYSIS TO EXAMINE CO-OCCURRENCE PATTERNS OF ANIMAL INJURIES AND DISEASES IN FARMERS' FIELDS IN DIFFERENT PRODUCTION ENVIRONMENTS ACROSS SOUTH AND SOUTH EAST ASIA** 15

Introduction 15

Materials and methods................................. 17

Result 21

Discussion... 38

Conclusion 43

II **DIFFERENTIAL NETWORKS REVEAL THE DYNAMICS OF ANIMAL PESTS AND DISEASE CO-OCCURRENCES** 45

Introduction 45

Materials and Methods 47

II.1Results 50

Construction of differential co-occurrence networks of rice injuries at different seasons 50

Construction of differential co-occurrence networks of rice injuries at yield level 52

Discussion 55

LIST OF FIGURES

FIGURE PAGE

-1 Correlation methods selected for constructing network model. The plots on the left show sample survey data taken over 29 rice injuries at 5 production environments. The next step is to compute pair-wise correlations to obtain a cross-correlation matrix (the middle plot, rows and columns represent the nodes). Then, the matrix will be used for network analysis. 3

-2 Histograms showing the distribution of rice injury values in crop health survey data. BB:Bug burn, BLB:Bacterial leaf blight, BLS: Bacterial leaf streak, BS:Brown spot, DH:Deadheart, DP: Dirty panicle, FS:False smut, GS:Gressy stunt, HB:Hopper burn, LB: Leaf blast, LF: Leaffolder injury, LM: Leaf miner injury, LS: :Leaf scald, NB:Neck blast, NBS:Nerrow brown spot, PM: Panicle mite injury, RB: Rice bug injuries, RGS:Ragged stunt, RH:Rice hispa injury, RS: Red stripe, RT:Rat damage, RTG: Tungro, RTH:Rice thrip injury, SHB:Sheath blight, SHR:Sheath rot, SNL:Snail damage, SR:Stem rot, SS:Silver shoot, WH:White head, WM:Whorl maggot injury. 8

-3 Heatmap and dendrogram showing hierarchical unsupervised clustering analysis correlation measures of survey data 11

I-1 Network construction 18

I-2 Bar graphs showing prevalence of injuries a across production environments and seasons. BB: Bug burn, BLB: Bacterial leaf blight, BLS: Bacterial leaf streak, BS: Brown spot, DH: Deadheart, DP: Dirty panicle, FS: False smut, GS: Grassy stunt, HB: Hopper burn, 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 stunt, RH: Rice hispa injury, RS: Red stripe, RT: Rat damage, RTG: Tungro, RTH: Rice thrip injury, SHB: Sheath blight, SHR: Sheath rot, SR: Stem rot, SS: Silver shoot, WH: White head, WM: Whorl maggot injury. 23

I-3 Bar graphs showing prevalence of injuries a across production environments and seasons (continue) 24

I-4 Bar graphs showing prevalence of injuries a across production environments and seasons (continue) 25

I-5 Rice injuries in dry season in Central Plain, Thailand 28

I-6 Injuries in Central Plain, Thailand 29

I-7 Injuries in dry season at Odisha, India 31

I-8 Injuries in wet season at Odisha, India 32

I-9 Rice injuries in dry season in Red River Delta, Vietnam 33

I-10 Rice injuries in wet season in Red River Delta, Vietnam 34

I-11 Rice injuries in dry season in Tamil Nadu, India 36

I-12 Rice injuries in wet season in Tamil Nadu, India 37

I-13 Rice injuries in dry season in West Java, Indonesia 39

I-14 Rice injuries in wet season in West Java, Indonesia 40

II-1 Schematic showing principle of 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. . . 49

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. 50

II-3 Differential co-occurrence network of rice injuries in different seasons at Central Plain, Thailand 51

II-4 Differential co-occurrence network of rice injuries in different seasons at Odisha, India 52

II-5 Differential co-occurrence network of rice injuries in different seasons at Red River Delta, Vietnam 53

II-6 Differential co-occurrence network of rice injuries in different seasons at Tamil Nadu, India 54

II-7 Differential co-occurrence network of rice injuries in different seasons at West Java, Indonesia 54

II-8 55

II-9 Rice injuries in dry season in Central Plain, Thailand 56

II-10Rice injuries in dry season in Central Plain, Thailand 57

II-11Rice injuries in dry season in Central Plain, Thailand 58

II-12. 59

II-13. 60

II-14. 60

II-15. 61

II-16. 61

II-17. 62

.1 Evaluation of correlation methods for co-occurrence network construction of rice crop health survey data

Introduction

Rice is not threatened by one, but by many pests in a season. A combination of injuries caused by diseases and rice pests can be thought of as a crop health syndrome. The development a crop health syndrome depend on the production situation (*i.e.*, the cultural practices, inputs used to produce a rice crop) as a range of agroecosystem (Savary et al., 2006).

A characterization study based on survey data collected in South and South East Asia (Savary et al., 2000b) showed the patterns of crop health syndromes were common and different across sites. The study indicated that sheath blight, brown spot and leaf blast are the most important diseases and were commonly found in some sites, causing yield loss between 1 to 10%. Among insect injuries, stem borer caused yield losses of 2.3%. (Savary et al., 2000b) characterized patterns of injury profiles into five groups. For example, injury profile group1 (IN1) was characterized by comparatively high incidence of stem rot, sheath blight, plant hopper, and whorl maggot injuries, but low brown spot, and absence of bacterial leaf blight, leaf blast, and neck blast Asia.

Networks are ubiquitous systems in nature, technology and society (Newman, 2003). A network is defined as one or more sets of nodes connected by links in various ways. A node can represent the individual units depending on the context. Links or edges are the connections between nodes, which may be directed or undirected. Network models are now becoming increasingly interesting and useful in social science, biology, and ecology. The network applications are also relevant in plant pathology were also increasingly studied (Moslonka-Lefebvre et al., 2011). And more citation.

Network analysis is a promising tool frequently used to describe the pairwise relationships of a large number of variables. For example, association networks or correlation networks were represented by their association or correlation (adjacency) matrices, which rows and columns denote nodes, and matrix entries denote links. They were widely applied in biological studies (Toubiana et al., 2013; Barabási and Oltvai, 2004)

Selecting the most suitable correlation methods for correlation network construction is important since different correlation measures lead to different network structure and provide different information. In this chapter, four correlation methods, including Pearson, Spearman rank correlation, Kendall correlation (Prokhorov, 2001) and Biweight mid-correlation, to associate rice injuries. **To do** (??)

Materials and methods

The method for evaluating and selecting correlation methods to construct correlation network of rice injuries in crop health survey data was showed in Fig. Finally, correlation metrics were used for the network analysis in next chapter.

Survey data

Survey data was collected from 450 farmers' fields growing rice on irrigated lowland areas across South and Southeast Asia Tamil Nadu, India (TM); Odisha, India, West Java; Indonesia; Central Plain, Thailand, and Red River Delta, Vietnam were collected from 2013 to 2016. The number of survey are summarized in Table 1.

The survey procedure and data were based on a standardized protocol described

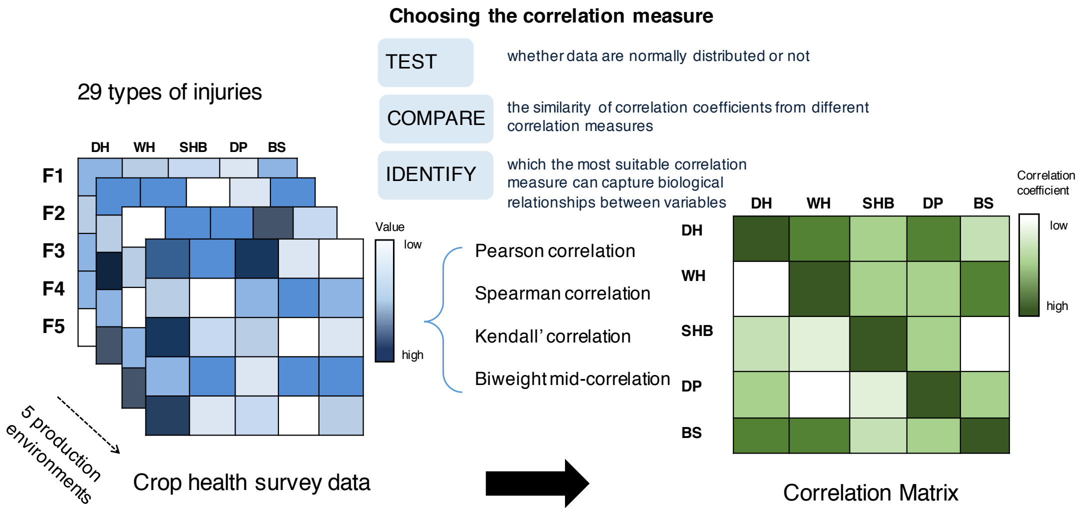

Choosing the correlation measure
TEST
whether data are normally distributed or not
COMPARE
the similarity of correlation coefficients from different correlation measures
IDENTIFY
which the most suitable correlation measure can capture biological relationships between variables
29 types of injuries
DH
WH
SHB
DP
BS
F1
F2
F3
F4
F5
Value
low
high
5 production environments
Crop health survey data
Pearson correlation
Spearman correlation
Kendall' correlation
Biweight mid-correlation
Correlation coefficient
low
high
Correlation Matrix

Figure -1: Correlation methods selected for constructing network model. The plots on the left show sample survey data taken over 29 rice injuries at 5 production environments. The next step is to compute pair-wise correlations to obtain a cross-correlation matrix (the middle plot, rows and columns represent the nodes). Then, the matrix will be used for network analysis.

Production environment	year						Total
	2013		2014		2015		
	DS	WS	DS	WS	DS	WS	
Central Plain, Thailand	20	20	14	21	15	12	102
Odisha, India	15	12	15	16	15	15	88
Red River Delta, Vietnam	15	15	15	15	15	15	90
Tamil Nadu, India	15	15	15	14	15	15	89
West Java, Indonesia	15	15	14	15	15		74
Total	80	77	73	81	75	57	443

Table 1: Number of farmers' fields surveyed by production environment and year

in "A survey portfolio to characterize yield-reducing factors in rice" developed by Savary and Castilla (2009). Twenty-nine rice injuries were collected including the injuries caused by animal pests, and pathogens, which are harmful to rice plants, and importantly considered to reduce yield productivity. The injuries were evaluated at booting and ripening stage according to survey procedure. They were found on different organs of rice plants depending on their natures.

Injuries on leaves such as whorl maggot injury (WM), leaffolder injury (LF), bacterial leaf blight (BLB), bacterial leaf streak (BLS), leaf blast (LB), brown spot (BS), leaf miner injuries (LM), leaf scald (LS), narrow brown spot (NBS), rice hispa injury

(RH), red stripe (RS), rice thrip injury (RTH) were determined as a proportion of injured leaves. Injuries on tillers or hills such as stem rot (SR), sheath rot (SHR), sheath blight (SHB), whitehead (WH), deadheart (DH), silver shoot (SS), false smut (FS), Neck blast (NB), Panicle mite injury (PM), Rice bug injury (RB), rat injury (RT) were determined as a proportion of injured tillers or panicles. Systemic injuries such as Bug burn (BB), grassy stunt (GS), hopper burn(HB) , ragged stunt (RGS), tungo (RTG) were determined as the percentage of area affected. The rice injury lists are showed in Table 2.

Table 2: Variables describing individual fields in surveys of rice injuries in five production environments in tropical Asia

Injury variables	Acronym	Description	Unit
Bug burn	BB	maximum percentage of bugburn in a one-sqm area	%
Bacterial leaf blight	BLB	area under the progress curve of the mean percentage of leaves with bacterial leaf blight	% dsu
Bacterial leaf blight	BLS	area under the progress curve of the mean percentage of leaves with bacterial leaf streak	% dsu
Brown spot	BS	area under the progress curve of the mean percentage of leaves with brown spot	% dsu
Deadheart	DH	maximum percentage of tillers with deadheart	%
Dirty panicle	DP	maximum percentage of panicles with dirty panicle	%
False smut	FS	maximum percentage of panicles with false smut	%
Grassy stunt	GS	maximum percentage of grassy stunt disease in a one-sqm area	%
Hopper burn	HB	maximum percentage of hopperburn in a one-sqm area	%
Leaf blast	LB	area under the progress curve of the mean percentage of leaves with leaf blast	% dsu
Leaffolder	LF	area under the progress curve of the mean percentage of leaves with leaffolder injury	% dsu
Leafminer	LM	area under the progress curve of the mean percentage of leaves with leaf miner injury	% dsu
Leaf scald	LS	area under the progress curve of mean percentage of leaves with leaf scald	% dsu
Neck blast	NB	maximum percentage of panicles with neck blast	%
Narrow brown spot	NBS	area under the progress curve of the mean percentage of leaves with narrow brown spot	% dsu
Panicle mite injury	PM	maximum percentage of tillers with panicle mite injury	%
Rice bug injury	RB	maximum percentage of panicles with rice bug injury	%
Ragged stunt	RGS	maximum percentage of grassy stunt disease in a one-sqm area	
Rice hispa	RH	area under the progress curve of the mean percentage of leaves with rice hispa injury	% dsu
Rat injury	RT	maximum percentage of tillers with rat injury	%
Red stripe	RS	area under the progress curve of mean percentage of leaves with red stripe	% dsu
Rice tungro	RTG	maximum percentage of tungro in a one-sqm area	%
Rice thrip injury	RTH	area under the progress curve of the mean percentage of leaves with rice thrip injury	% dsu
Sheath blight	SHB	maximum percentage of tillers with sheath blight	%
Sheath rot	SHR	maximum percentage of tillers with sheath rot	%
Stem rot	SR	maximum percentage of tillers with stem rot	%
Silver shoot	SS	maximum percentage of tillers with silvershoot	%
Whitehead	WH	maximum percentage of panicles with whitehead	%
Whorl maggot injury	WM	area under the progress curve of the mean percentage of leaves with whorl maggot injury	% dsu

Before analysis, data were compacted over time during crop growth. Two types of data were computed, depending on the natures of injuries as defined by (Savary and Castilla, 2009). One is an area under injury progress curve (AUIPC) used for injury variables, which present on the leaves, and for weed infestation. Another is the maximum level at any of the two observations used for injury variables that can be observed on tillers, panicles, and hills, and insect pest count. The area under injury progress curve (AUIPC) (Campbell et al., 1990) were calculated by the mid-point method using the following equation:

$$AUIPC = \sum \frac{1}{2(X_i + X_{i-1})(T_i - T_{i-1})} \tag{1}$$

where X_i is percentage (%) of leaves, tillers or panicles injured due to rice pests (e.g., leaf blast, leaf folder), or number of insects (e.g., plant hoppers, leaf hoppers) per quadrat, or percentage (%) of weed infestation (ground coverage) at the ith observation, T_i is time in rice development stage units (dsu) on a 0 to 100 scale (10: seedling, 20: tillering, 30: stem elongation, 40: booting, 50: heading, 60: flowering, 70: milk, 80: dough, 90: ripening, 100: fully mature) at the ith observation and n is total number of observations.

Evaluation In this study, correlation measures including, Pearson, Spearman, Kendal and Biweight mid-correlation (Wilcox, 2012) were evaluated to discover the true functionally related variables in crop health survey data. The data will have to follow the assumption of correlation measures. The correlation measures will also be able to effectively capture biological relationships that are well published. I proposed three steps for correlation methods selection:

- **Testing** whether or not the data are normally distributed by visual assessments and statistical tests. The distribution of values of rice injuries in crop health survey data was examined and tested the hypothesis hypothesis that the sample comes from a population which has a normal distribution by performed Shapiro-Wilk test.

- **Comparing** correlation measures by testing similarity of correlation coefficients. I evaluated the similarity of correlation coefficients of different correlation measures by using the Euclidean distance, and perform clustering analysis.

- **Identifying** the most suitable correlation measure that can capture biological relationships between variables confirm with the published relationships.

Result

Checking and testing the distribution of crop health survey data

To determine normality of the survey data, I presented the histograms (Fig -2) showing distribution of value of rice injuries, calculated summary statistics, and performed Shapiro-Wilk test. The histograms depict the distribution of values of injuries. The histograms showed that values of injuries are skewed to the left. Common values of the injuries were 0. A few farmers' fields presenting in high values of injuries were relatively low. The distribution of most injuries are described power low or long tails. ?? the power low or long tailed.

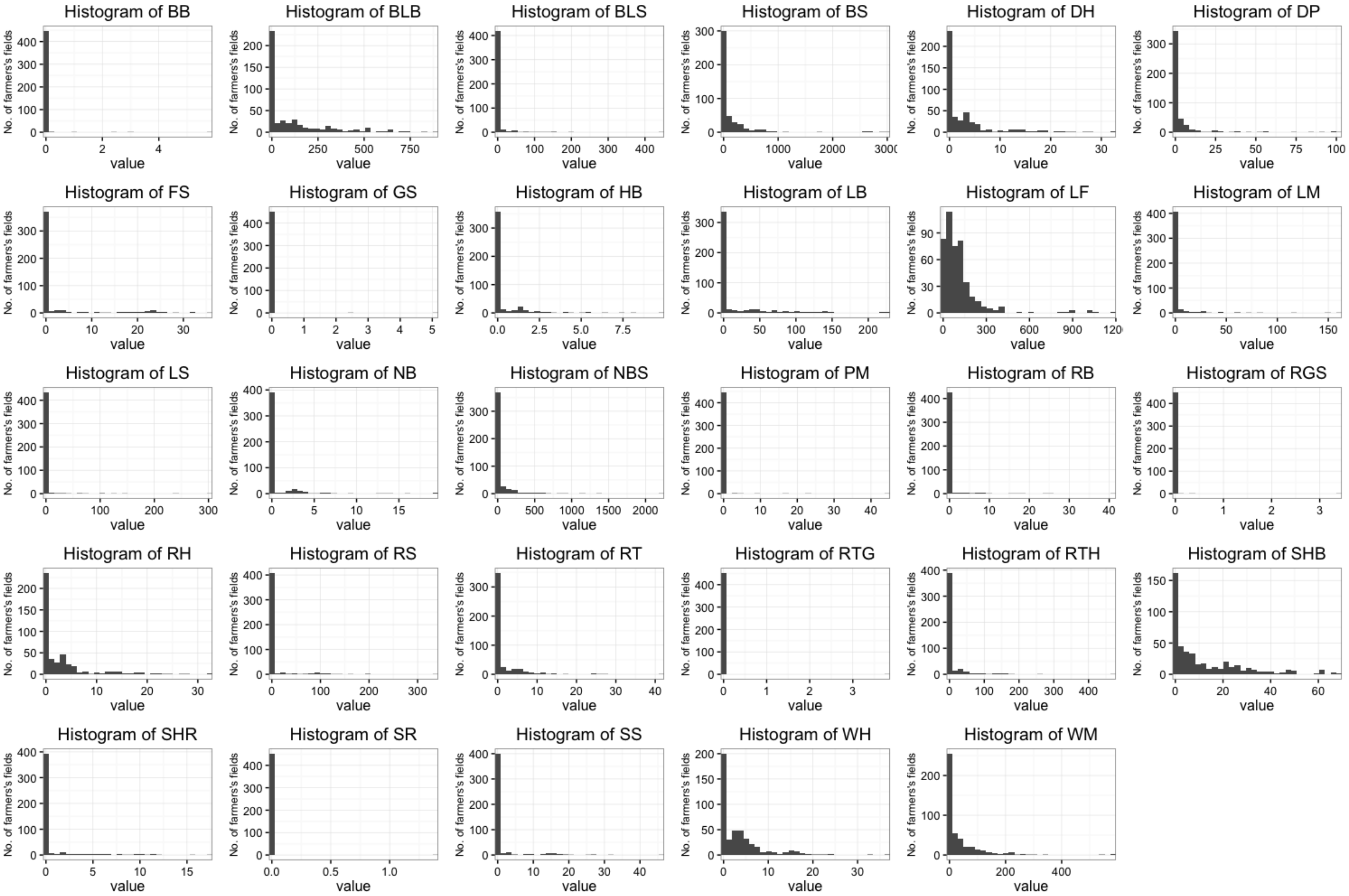


Figure -2: Histograms showing the distribution of rice injury values in crop health survey data. BB:Bug burn, BLB:Bacterial leaf blight, BLS: Bacterial leaf streak, BS:Brown spot, DH:Deadheart, DP: Dirty panicle, FS:False smut, GS:Gressy stunt, HB:Hopper burn, LB: Leaf blast, LF: Leaffolder injury, LM: Leaf miner injury, LS: :Leaf scald, NB:Neck blast, NBS:Nerrow brown spot, PM: Panicle mite injury, RB: Rice bug injuries, RGS:Ragged stunt, RH:Rice hispa injury, RS: Red stripe, RT:Rat damage, RTG: Tungro, RTH:Rice thrip injury, SHB:Sheath blight, SHR:Sheath rot, SNL:Snail damage, SR:Stem rot, SS:Silver shoot, WH:White head, WM:Whorl maggot injury.

Table 3: Summary statistics of rice injuries in crop health survey data

Injuries	Mean	SD[1]	Median	Min	Max	Skewness	Kurtosis	SE[2]	Shapiro-Wilk test[3]
BB	0.03	0.33	0.00	0.00	5.80	14.43	228.44	0.02	**
BLB	113.64	178.34	0.00	0.00	886.67	1.86	2.93	8.38	**
BLS	4.94	28.26	0.00	0.00	444.48	10.35	138.48	1.33	**
BS	147.54	386.19	0.00	0.00	2999.42	5.09	30.22	18.14	**
DH	3.18	5.62	0.00	0.00	32.33	2.49	6.50	0.26	**
DP	3.54	12.39	0.00	0.00	101.62	5.70	36.17	0.58	**
FS	2.41	6.61	0.00	0.00	35.74	2.90	7.35	0.31	**
GS	0.02	0.26	0.00	0.00	5.10	17.38	317.16	0.01	**
HB	0.41	1.13	0.00	0.00	9.80	4.18	22.02	0.05	**
LB	17.06	38.50	0.00	0.00	226.21	2.79	8.43	1.81	**
LF	114.48	156.86	76.74	0.00	1180.29	3.94	18.95	7.37	**
LM	2.93	14.33	0.00	0.00	160.47	7.63	68.27	0.67	**
LS	3.13	22.06	0.00	0.00	302.08	9.82	110.91	1.04	**
NB	0.60	2.14	0.00	0.00	19.32	5.69	38.72	0.10	**
NBS	53.98	178.50	0.00	0.00	2213.54	6.49	58.95	8.39	**
PM	0.23	2.51	0.00	0.00	44.42	14.35	229.31	0.12	**
RB	0.53	3.08	0.00	0.00	40.98	8.52	87.20	0.14	**
RGS	0.01	0.16	0.00	0.00	3.40	21.02	445.24	0.01	**
RH	3.18	5.62	0.00	0.00	32.33	2.49	6.50	0.26	**
RS	8.14	31.13	0.00	0.00	336.71	5.32	37.10	1.46	**
RT	1.47	4.02	0.00	0.00	41.56	4.90	32.92	0.19	**
RTG	0.01	0.18	0.00	0.00	3.80	21.28	453.00	0.01	**
RTH	9.01	34.69	0.00	0.00	470.55	7.52	79.09	1.63	**
SHB	11.67	15.49	4.72	0.00	68.65	1.60	2.08	0.73	**
SHR	0.66	2.23	0.00	0.00	17.41	4.24	19.76	0.10	**
SR	0.00	0.07	0.00	0.00	1.39	21.28	453.00	0.00	**
SS	1.40	4.91	0.00	0.00	46.01	4.47	24.33	0.23	**
WH	3.53	5.18	1.67	0.00	36.64	2.36	7.45	0.24	**
WM	36.72	71.63	0.00	0.00	583.41	3.84	21.16	3.37	**

[1] Stardard variation

The summary statistics, and the result of Shapiro-Wick test of each injury were calculated and summarized in Table 3. As can be seen, from the previous observations that rice injuries histograms tend to be positively skewed, the median values of injuries were considered due to their insensitivity to outliers. Median of injuries were almost 0, except LF. According to Doane and Seward (2011), skewness and kurtosis values were more than 0, which mean that the values of injuries were asymmetrically distributed with a long tail to the right. The normality is defined as p value 0.01 in Shapiro-Wilk testing. That test indicates that values of injuries were not normally distributed.

Comparing correlation coefficients of rice injuries from four correlation methods

I performed pair-wise analysis between each of injuries using all four correlation methods (Pearson, Spearman, Kendall correlation, and Biweight mid-correlation). I examined the similarity of correlation coefficients and clustered according to hierarchical clustering using Euclidean distance. The result is shown in Figure -3. Two groups of correlation methods can be distinguished: (i) parametric correlation measures (Pearson correlation and Biweight mid-correlation) and (ii) nonparametric correlation measure (Spearman correlation and Kendall correlation).

Identify the most suitable correlation measure

Although examination of correlation coefficients from correlation methods, it was also interested in learning the efficiency of each method if the output of each method were cut off by a threshold p-value. Since the resultant p-values from different methods can be different, I obtain the a higher number of significant injury pairs and a small number of significant injury pairs when implementing the same cut-off p-value threshold (e.g. p-value, 0.05) on different methods, making it difficult to compare the efficiency

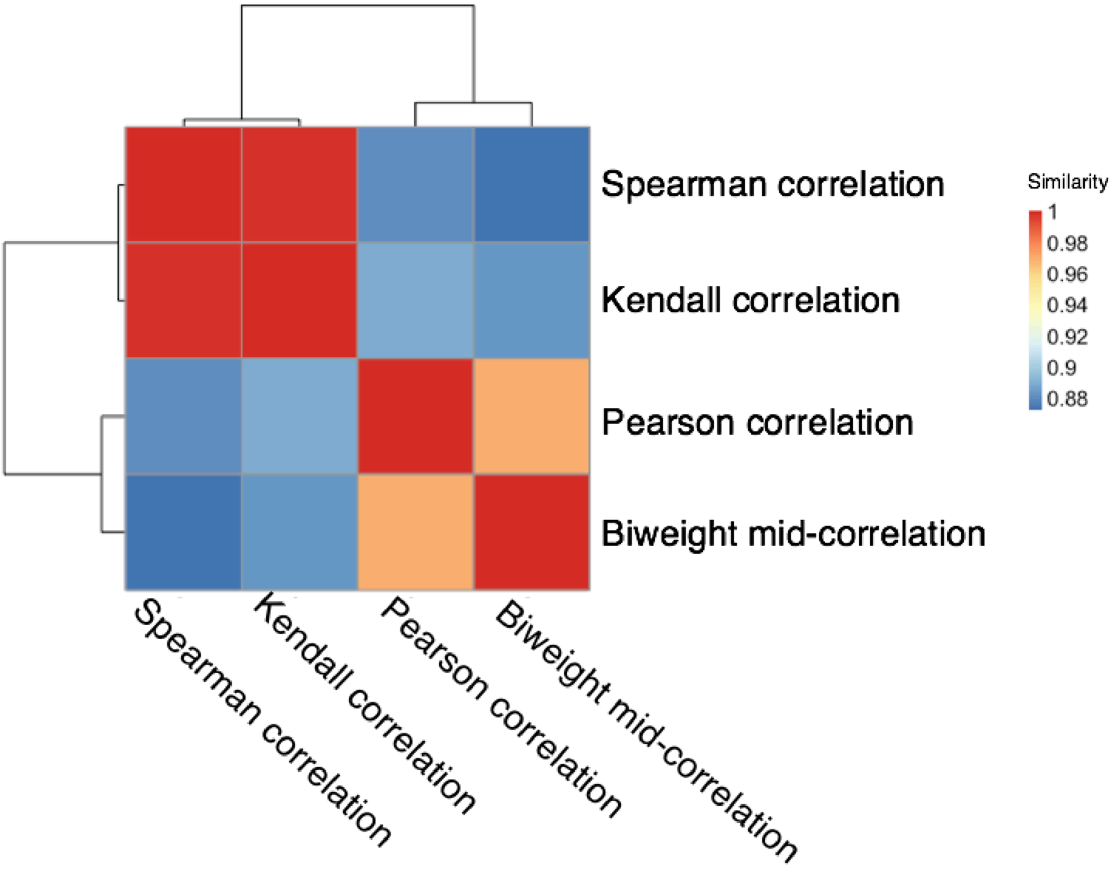


Figure -3: Heatmap and dendrogram showing hierarchical unsupervised clustering analysis correlation measures of survey data

Table 4: The percentage of injury pairs in survey data at p-value thresholds at 0.01 and 0.05 were applied to cut off correlated lists of injury pairs resulting from four methods

Methods p-value	<0.01	<0.05	>0.05
Biwweight	23.89% (97)	31.03%(126)	68.97% (280)
Kendall	13.79% (56)	21.18% (86)	83.25% (338)
Pearson	21.18% (86)	29.56% (120)	70.44% (286)
Spearman	33.25% (135)	44.83% (182)	55.17% (224)

of different methods (see Table 4). Outputs for each method was sorted by p-values in ascending order and cut off at p-value < 0.05. Spearman method could capture 182 pairwise relationships, following with Biweight-mid correlation Pearson method captures and Kendall with respectively, which could captured 120, 126, and 86 significant pairwise relationships. In a series of different cut-off p-values, it was generally high number of injury pairs resulting from Spearman appeared to be higher than Biweight, Pearson, and Kendall.

According to the previous results, the group of parametric correlation measures was selected out because these measures required the data normally distributed. So I would consider the correlation measures in the group of rank based methods that do not required normality. Compared between Spearman and Kendall correlation, the pair-wise relationships of rice injuries were captured differently. One of many relationships captured by Spearman correlation method,but not by Kendall methods is the relationship between dirty panicle and brown spot (Table). This relationship has been reported in many studies (Ou, 1985; Barnwal et al., 2013).

Discussion

An important criteria to select the suitable correlation measures is to check normality of the variables analysed, because a vital assumption in Pearson's contribution is the normality of the variables analysed, which could be true only for quantitative variables. Pearson's correlation coefficient is a measure of the strength of the linear relationship between two such variables. Thus, it is worth to check and test this assumption. Based on visual assessment of the histograms, all the variables show a skewness to the left skewed clearly. Skewness and kurtosis values were higher than zero, which indicated that the populations were not non-normal distribution according to Doane and Seward (2011). Visual inspection of the distribution may be used for assessing normality, although this approach is usually unreliable and does not guarantee that the distribution is normal (Ghasemi et al., 2012). So normality tests were suggested such as Kolmogorov-Smirnov test, Shapiro-Wilk test. Some researchers recommend the Shapiro-Wilk test as the best choice for testing the normality of data (Peat and Barton, 2005). Shapiro-Wilk test showed that these results were in accord with skewness and kurtosis values.

Evaluation of four correlation measures, including Pearson, Spearman, Kendall correlation, and Biweight mid-correlation by determining similarity of correlation coefficients from each method. The results showed two groups clustered according to hierarchical clustering using Euclidean distance. The Spearman and Kendall correlation were grouped in rank-base methods, and another group is non-rank-based correlations including Pearson correlation and Biweight mid-correlation. From the previous result of testing normality of data, it suggested that the data did not meet the assumption of parametric correlation methods, such as Pearson' method. However, Pearson correlation coefficient is sensitive to outliers (citation). Biweight midcorrelation is considered to be a good alternative to Pearson correlation since it is more robust to outliers(Wilcox, 2012). Among the four correlation methods, Spearman and Kendall are nonparametric rank-based methods. The rank-base methods are nonparametric (distribution-free) statistics, which uses ranks for correlation and therefore provides a robust measure of a monotonic relationship between two continuous random variables. For this reason, they are particularly suitable for identifying the injuries that increase or decline in monotonic trends in survey data collected during a biological processes or developmental stages.

Although we can opt for a method based on its principle of statistical operation without paying attention to the biological models in a given data set, this may not lead to a coordination network that will reveal biological knowledge (Kumari et al., 2012). The appropriate correlation measures for studied data should closely associate to the prior knowledge of biological correlation. For this study, relationship between dirty panicle and brow spot was detected by Spearman correlation but not by other correlation methods.

Conclusions

The analyses I have performed clearly demonstrate the distinct and common performance of four correlation methods. Pearson has been widely used in correlation analyses (Zhang and Horvath, 2005). However, Pearson correlation is limited to be suitable to normally distributed data, and is only able to capture the linear relationships. Biweight mid-correlation is more robust to take outlier into account compared to Pearson method, but this is not seem relevant for this survey data, as the outputs between the two methods were not different (in same cluster). The Spearman and Kendall method performed ?? and can capture many relationships in this survey data. Chosen between these methods, Spearman could capture and identify biologically or functionally associated injuries.

CHAPTER I

Using network analysis to examine co-occurrence patterns of animal injuries and diseases in farmers' fields 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 links 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 key- stone 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.

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 between injuries proposed by network analysis might help to distinguish the natural or anthropogenic source of metals in these sediments. 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 textsuperscripto 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

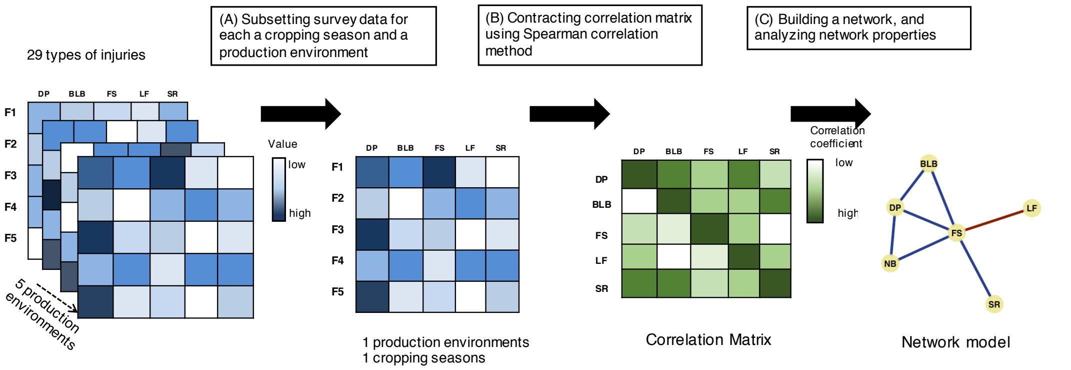

29 types of injuries
(A) Subsetting survey data for each a cropping season and a production environment
(B) Contracting correlation matrix using Spearman correlation method
(C) Building a network, and analyzing network properties
DP
BLB
FS
LF
SR
F1
F2
F3
F4
F5
5 production environments
Value
low
high
1 production environments
1 cropping seasons
Correlation coefficient
low
high
Correlation Matrix
BLB
DP
LF
FS
NB
SR
Network model

Figure I-1: Network construction

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 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 occurrences, and is written in $A = [C_{ij}]$, which is

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

where C is positive 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 injuries and the 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 resulting networks, a set of measures (node degree, betweenness, local clustering coefficient, average clustering coefficient, and average path length) 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. Average clustering coefficient, and average path length were measured for each network. 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. Average path length is the average number of steps along the shortest paths for all possible pairs of network nodes, and diameter is the greatest distance between any pair of nodes.

The average clustering coefficient is defined as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triplets of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triplets of vertices}}. \tag{I.3}$$

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 (but not always) 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.

Result

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.

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 hopper burn (HB) caused by brown plant hoppers, and white backed plant hoppers, bug burn (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.

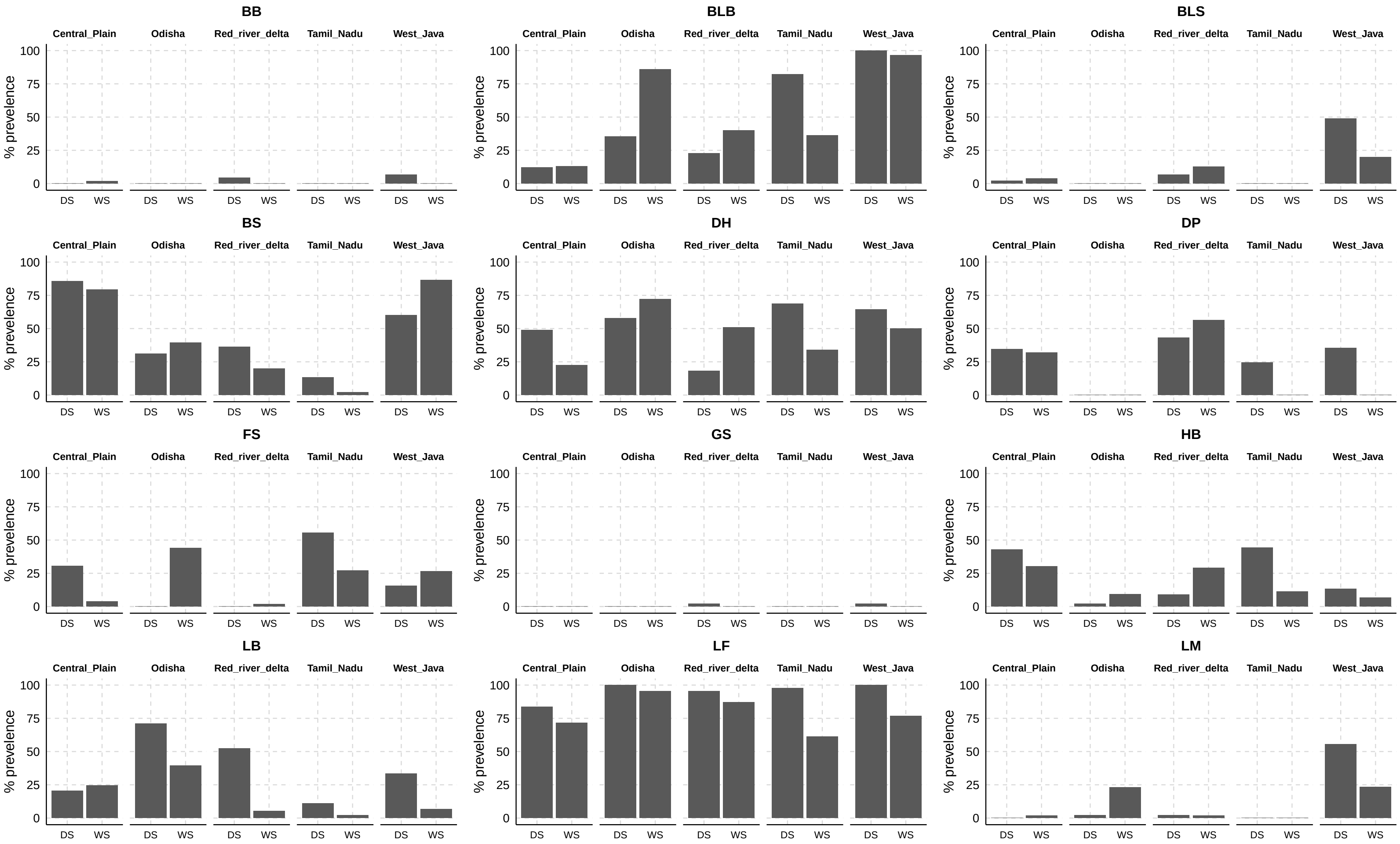


Figure I-2: Bar graphs showing prevalence of injuries a across production environments and seasons. BB: Bug burn, BLB: Bacterial leaf blight, BLS: Bacterial leaf streak, BS: Brown spot, DH: Deadheart, DP: Dirty panicle, FS: False smut, GS: Grassy stunt, HB: Hopper burn, LB: Leaf blast, LF: Leaffolder 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 stunt, RH: Rice hispa injury, RS: Red stripe, RT: Rat damage, RTG: Tungro, RTH: Rice thrip injury, SHB: Sheath blight, SHR: Sheath rot, SR: Stem rot, SS: Silver shoot, WH: White head, WM: Whorl maggot injury.

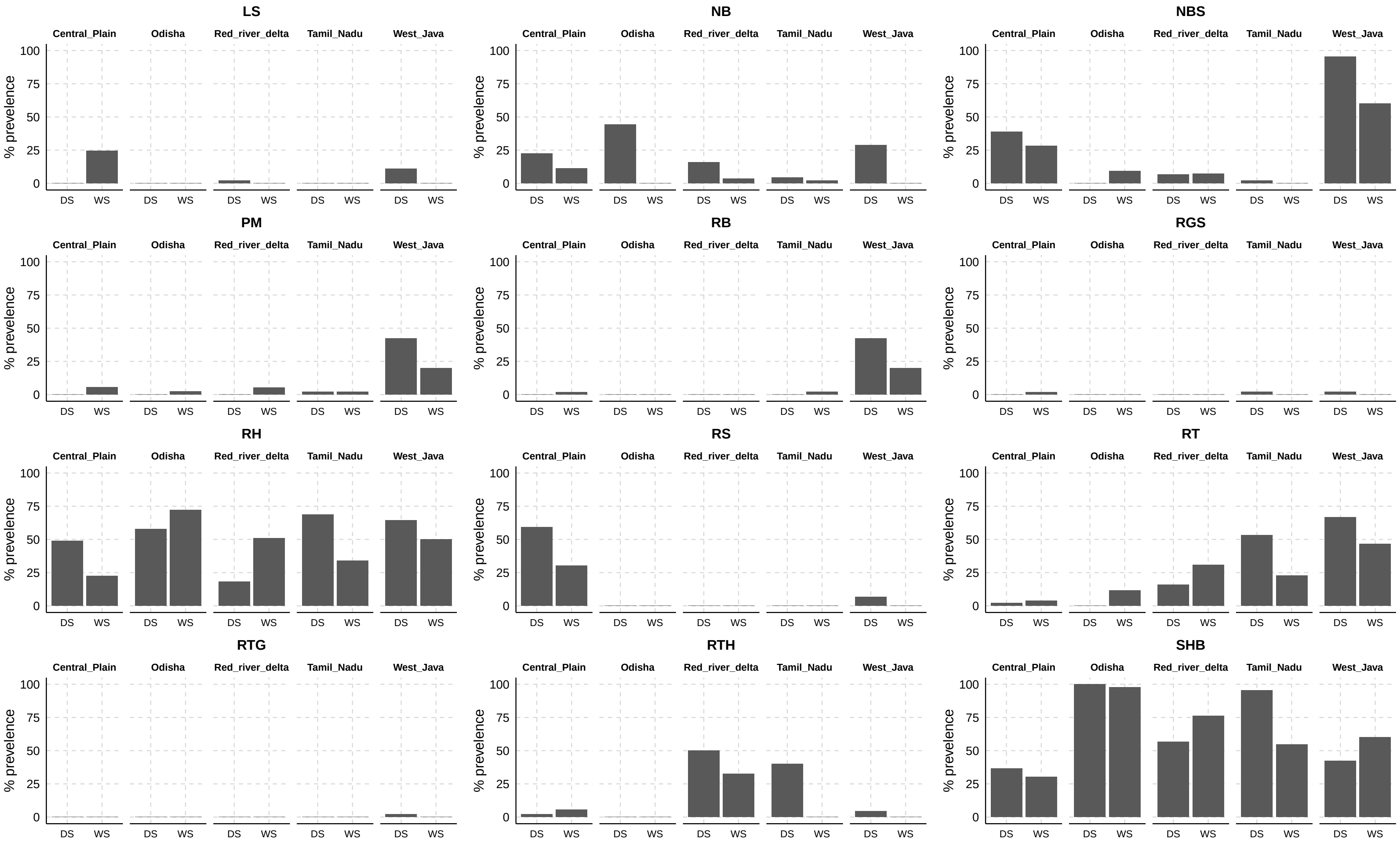


Figure I-3: Bar graphs showing prevalence of injuries a across production environments and seasons (continue)

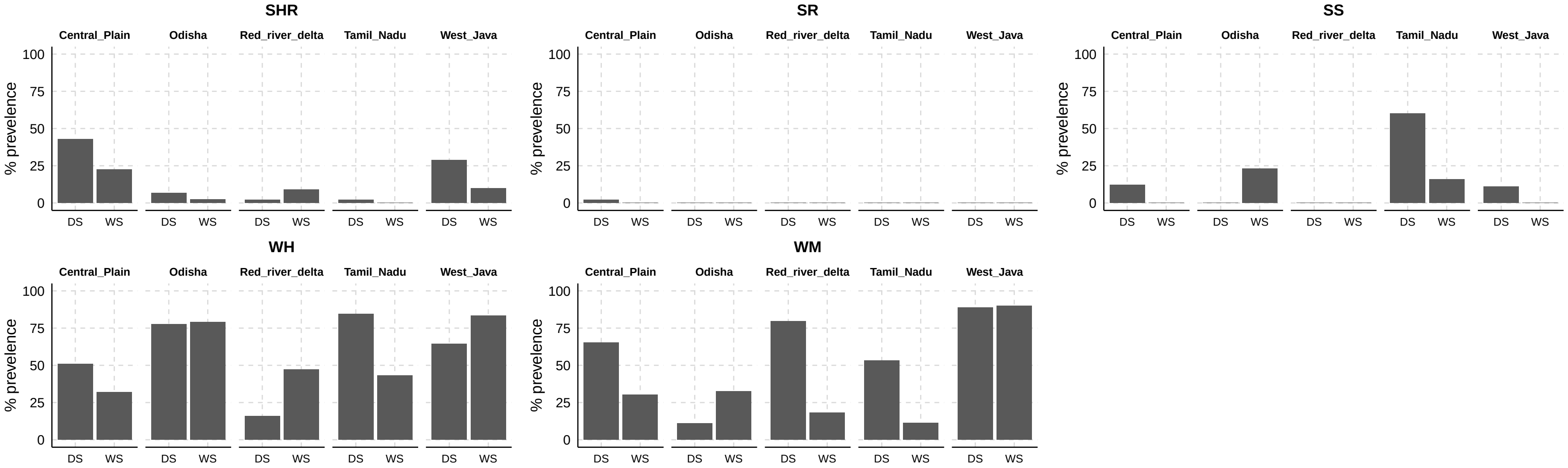


Figure I-4: Bar graphs showing prevalence of injuries a across production environments and seasons (continue)

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

The co-occurrence networks for crop health data are given in Figure. II-11 to Figure I-14. To analyze the rice injury network, I focus on the most prominent properties of nodes in a network node: node strength, betweenness, and clustering coefficient. Node strength 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

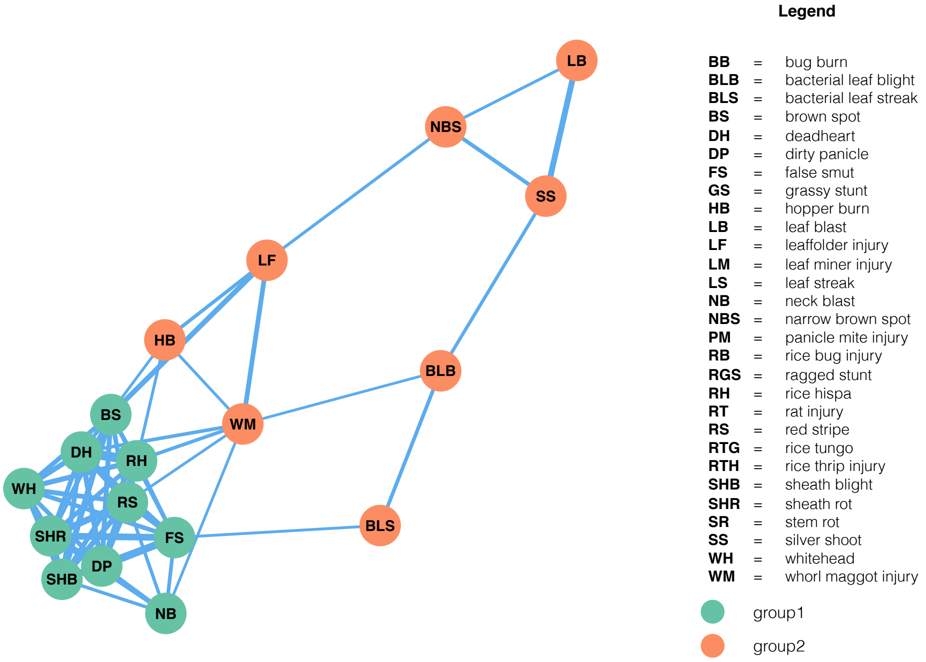
Dry season network (Figure II-9a) is comprised of 18 associated injuries and 60 associations (edges). The network show two groups of injury syndromes (the combination of injuries) based on the optimal clustering algorithm. Group1 (green) was more closely clustered than another group, according to node 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 II-9b) 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 betweenness), 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-6a) 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-6b). 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.

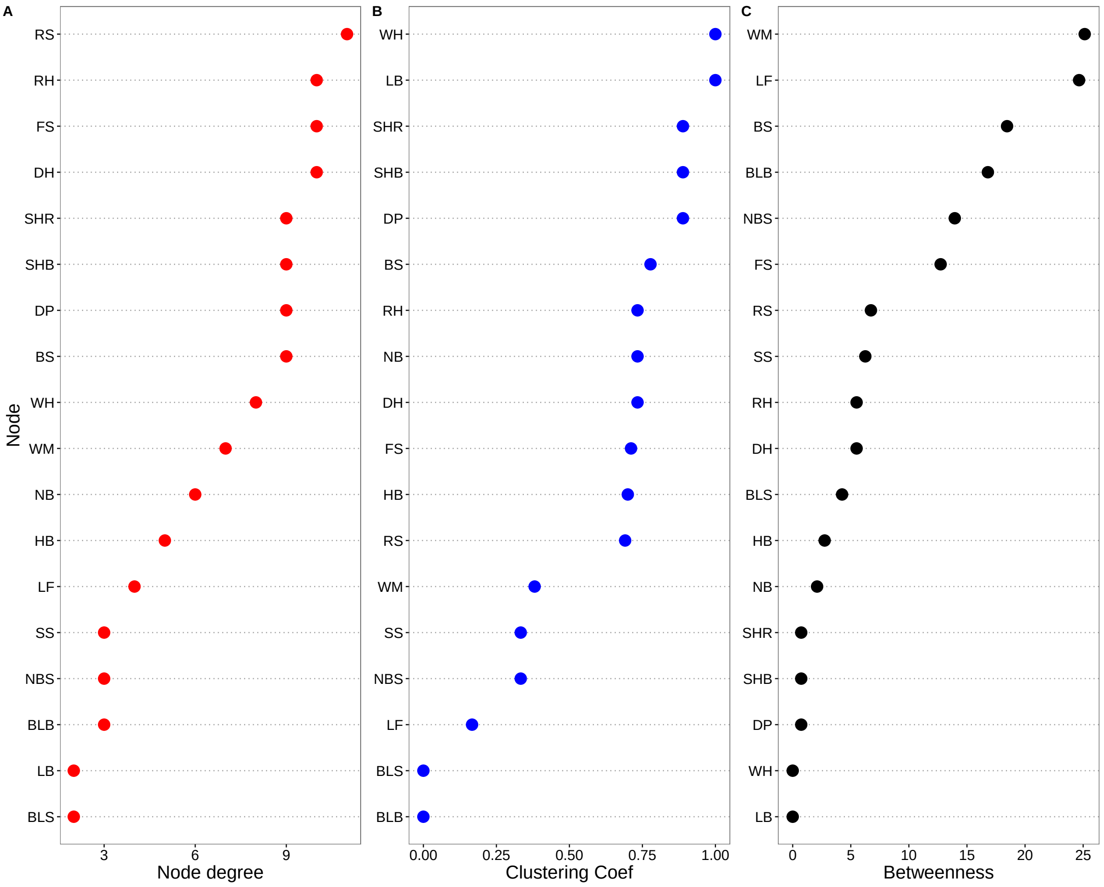
Odisha, India

Co-occurrence network of rice injuries in dry season (Figure I-7a) 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-7b).

The network in wet season (Figure I-8a) 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. Syn-

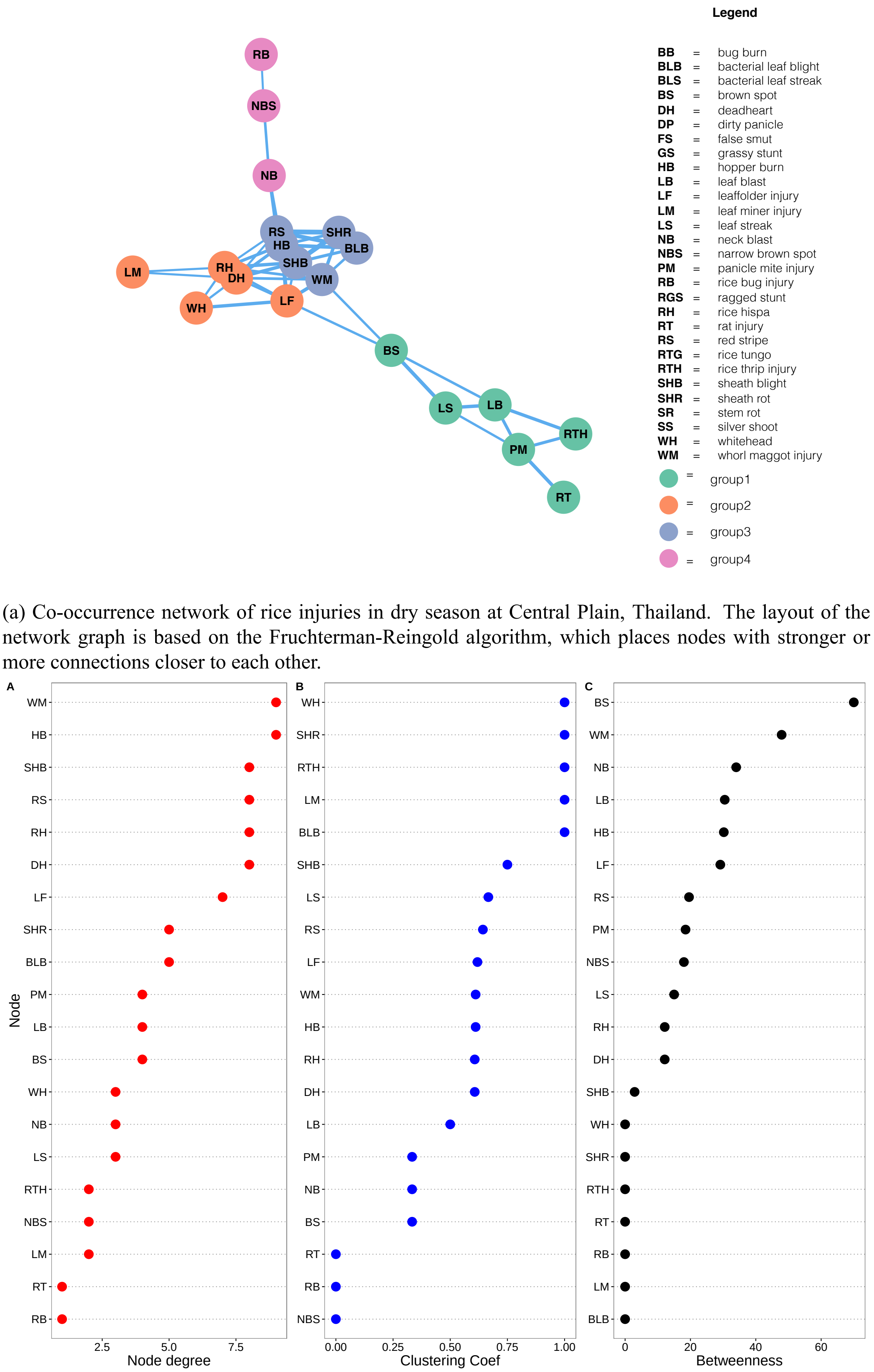


(a) Co-occurrence network of rice injuries in dry season at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.



(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.

Figure I-5: Rice injuries in dry season in Central Plain, Thailand



(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

Figure I-6: Injuries in Central Plain, Thailand

drome2 is placed in between syndrome1 and 3. The injuries in syndrome2 have high value of node degree and clustering coefficient (Figure I-8b), which mean they were 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.

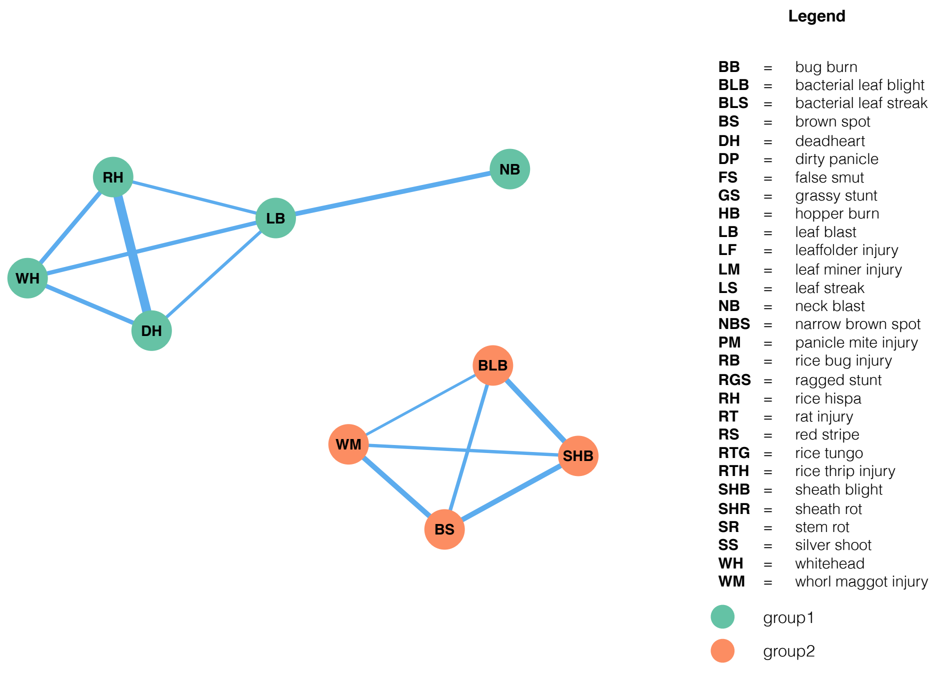
Red River Delta, Vietnam

Co-occurrence network of rice injuries in dry season (Fig. I-9a) 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-9b).

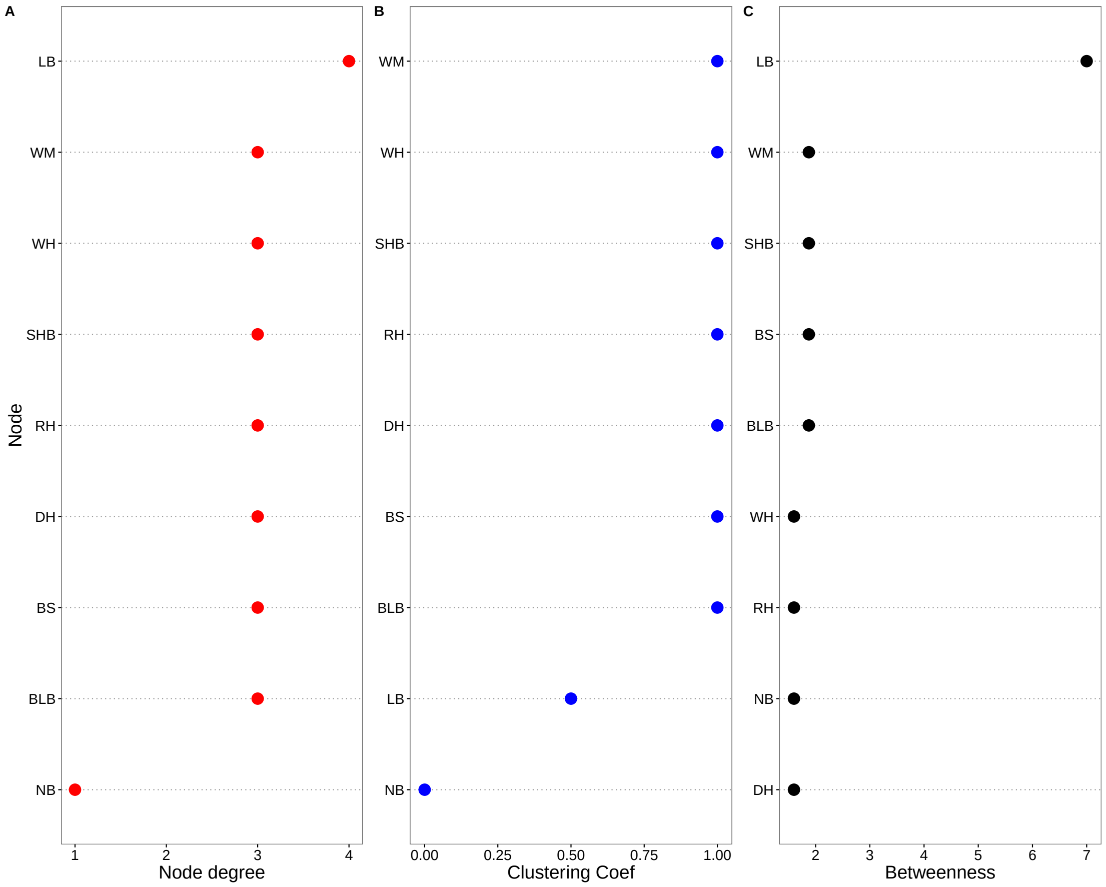
Wet season network (Figure I-10a)) consists of 18 injuries with 37 associations. It reveals 4 connected syndromes and an isolated syndrome. Group2 was located that could connect to Group3, 4, 5. According to Figure I-10b, BLB, DP could be good indicator, because they are likely to occur (high betweenness) 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-11a) 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-11b), injuries in syndrome2 might occur

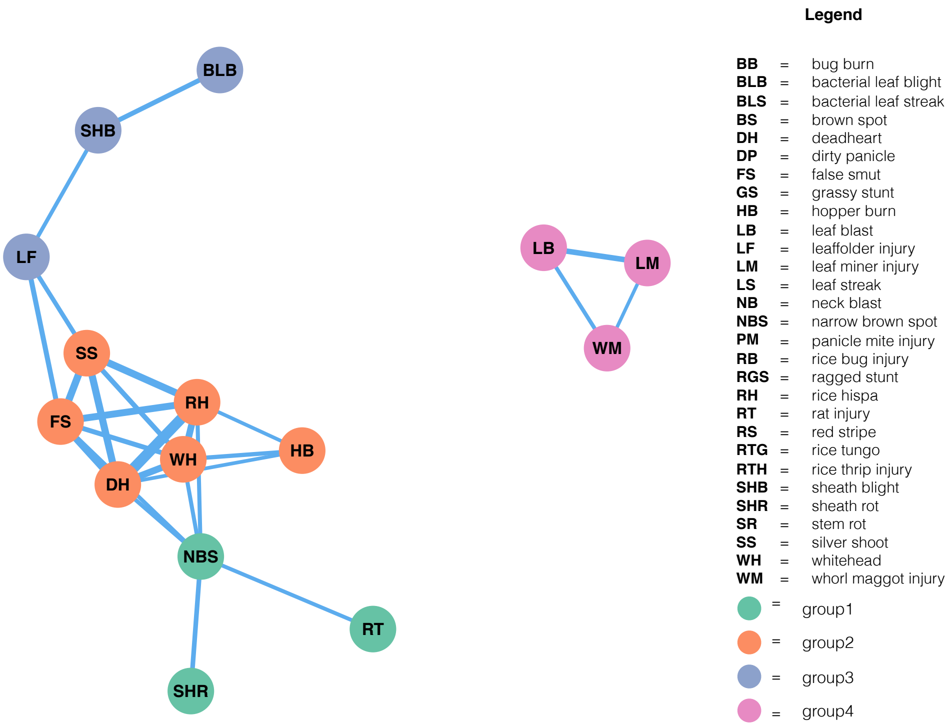


(a) Co-occurrence network of rice injuries in wet season at Odisha, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

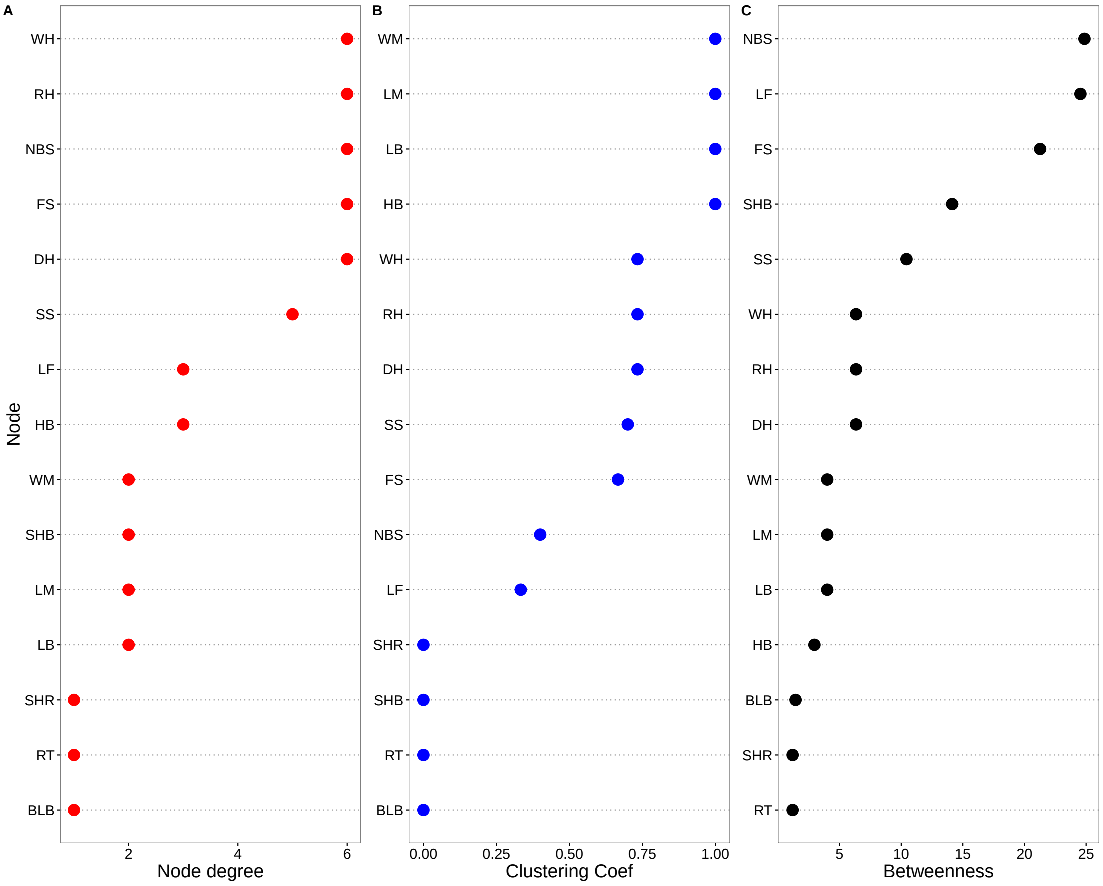


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Odisha, India. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure I-7: Injuries in dry season at Odisha, India

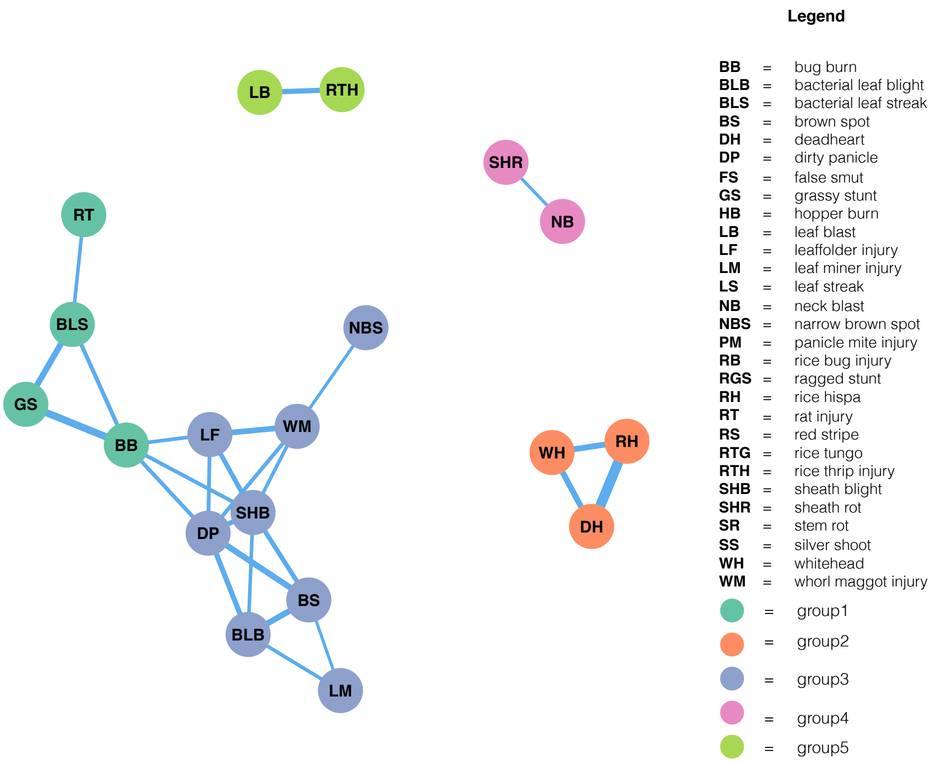


(a) Co-occurrence network of rice injuries in wet season at Odisha, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

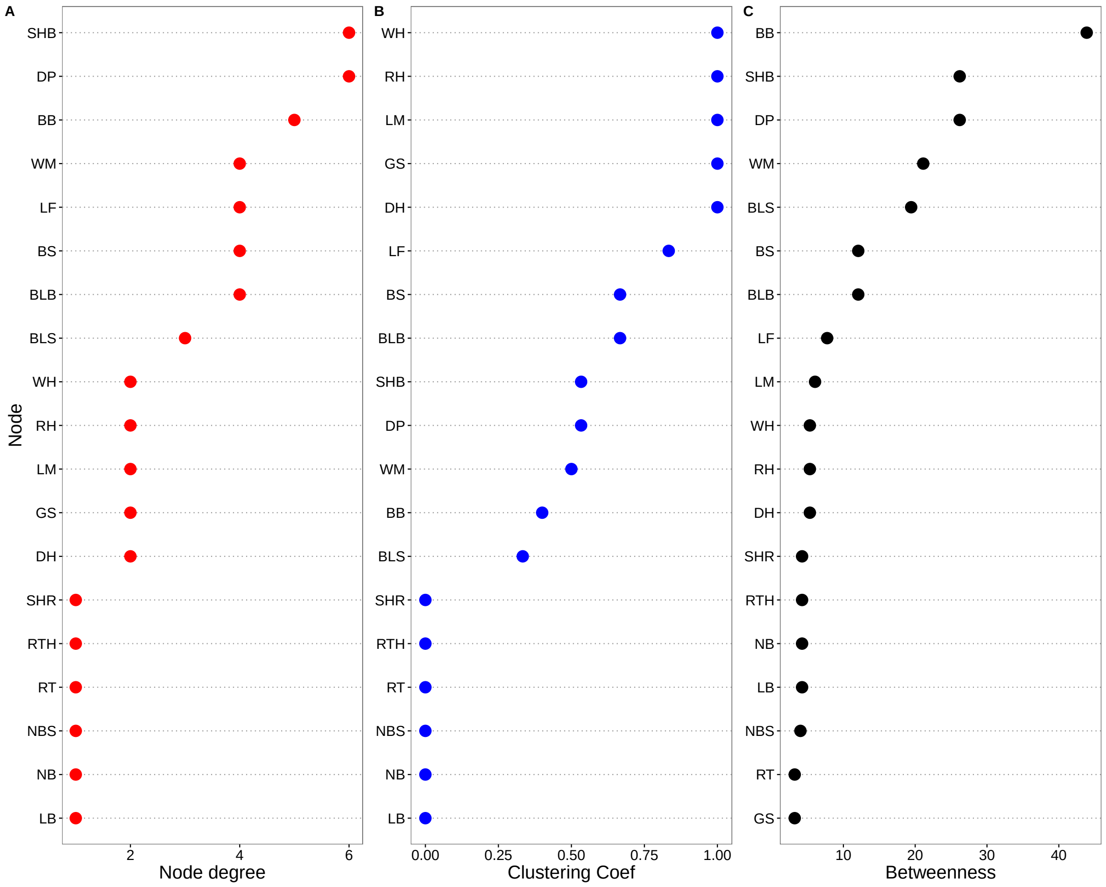


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in wet season at Odisha, India. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure I-8: Injuries in wet season at Odisha, India

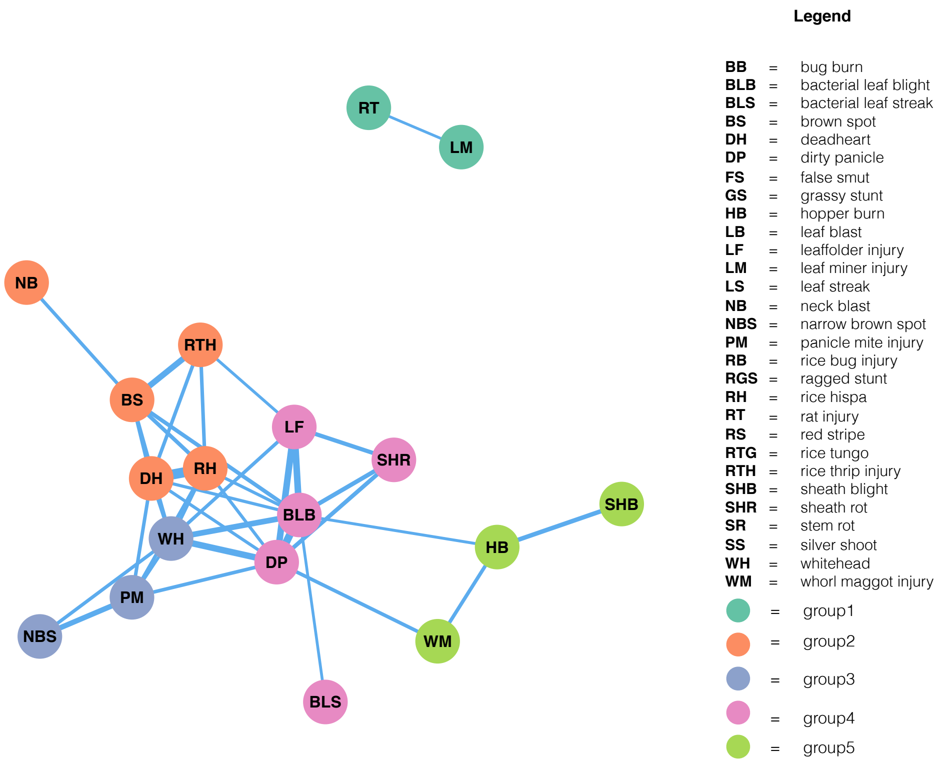


(a) Co-occurrence network of rice injuries in dry season at Red River Delta, Vietnam. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

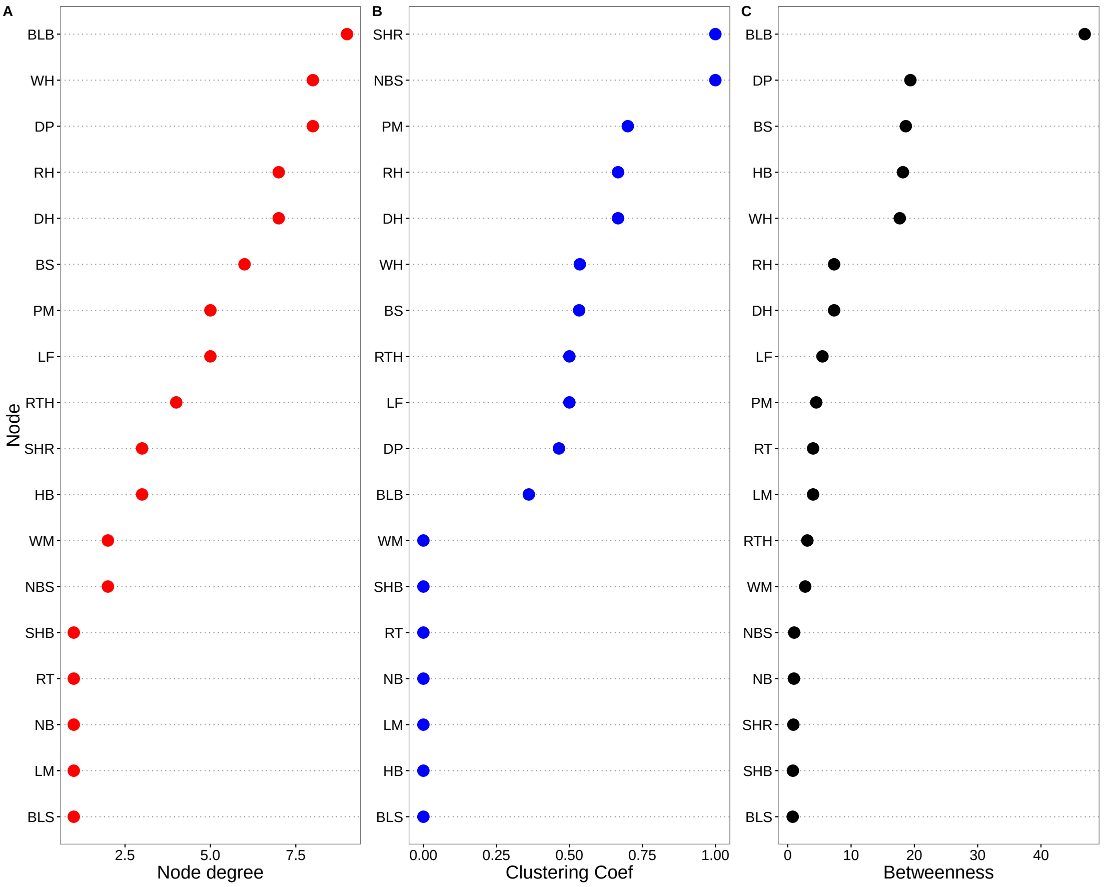


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Red River Delta, Vietnam. A: node degree, B:clustering coefficient, and C:Betweenness

Figure I-9: Rice injuries in dry season in Red River Delta, Vietnam



(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-10: Rice injuries in wet season in Red River Delta, Vietnam

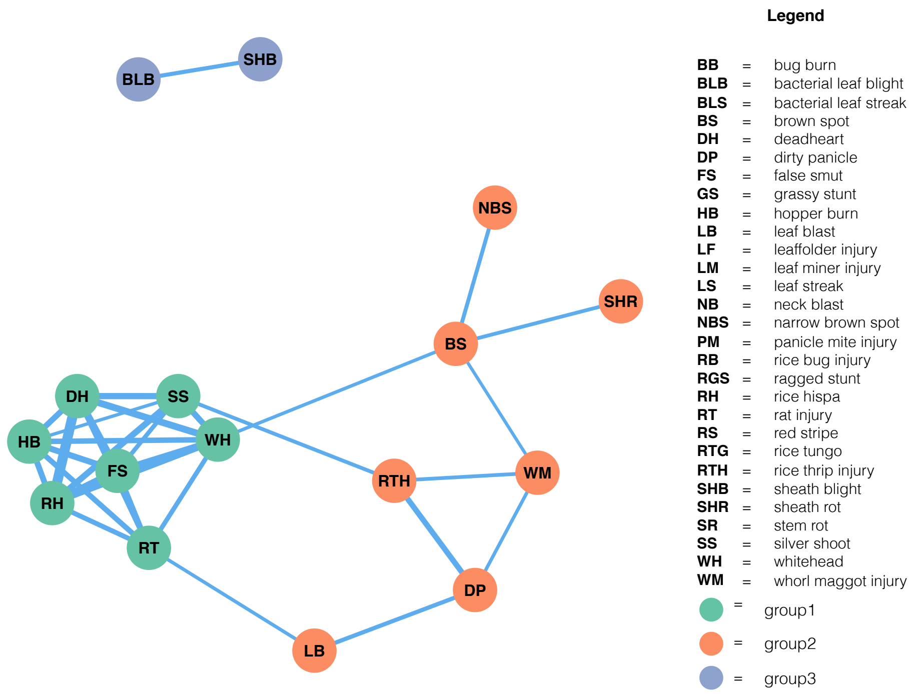
together with one or two more injuries.

Figure I-12a 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-12b).

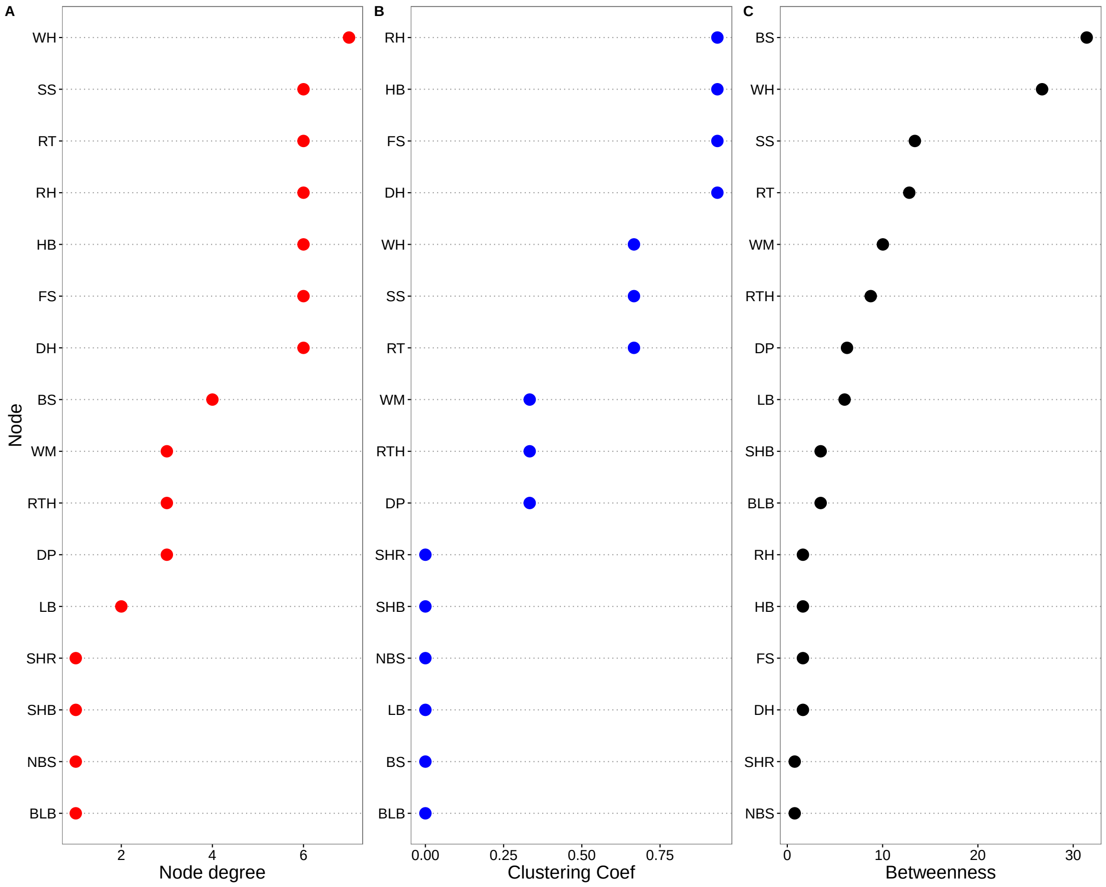
West Java, Indonesia

Co-occurrence network of injury profiles of dry season presented in Figure I-13a 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-14b, 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-14a) 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 the

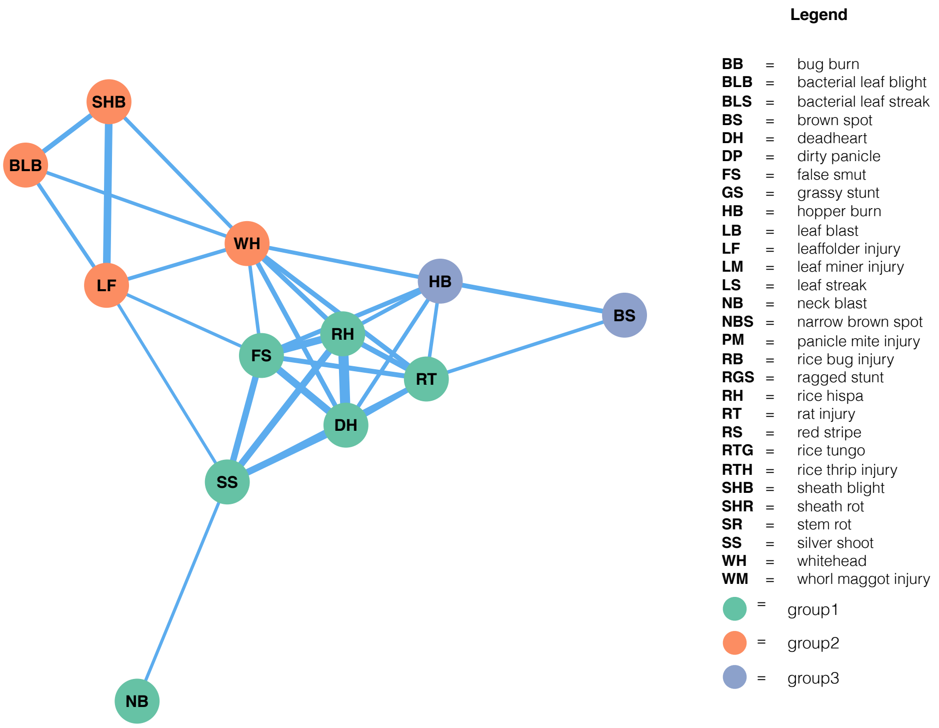


(a) Co-occurrence network of rice injuries in dry season at Tamil Nadu, India. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

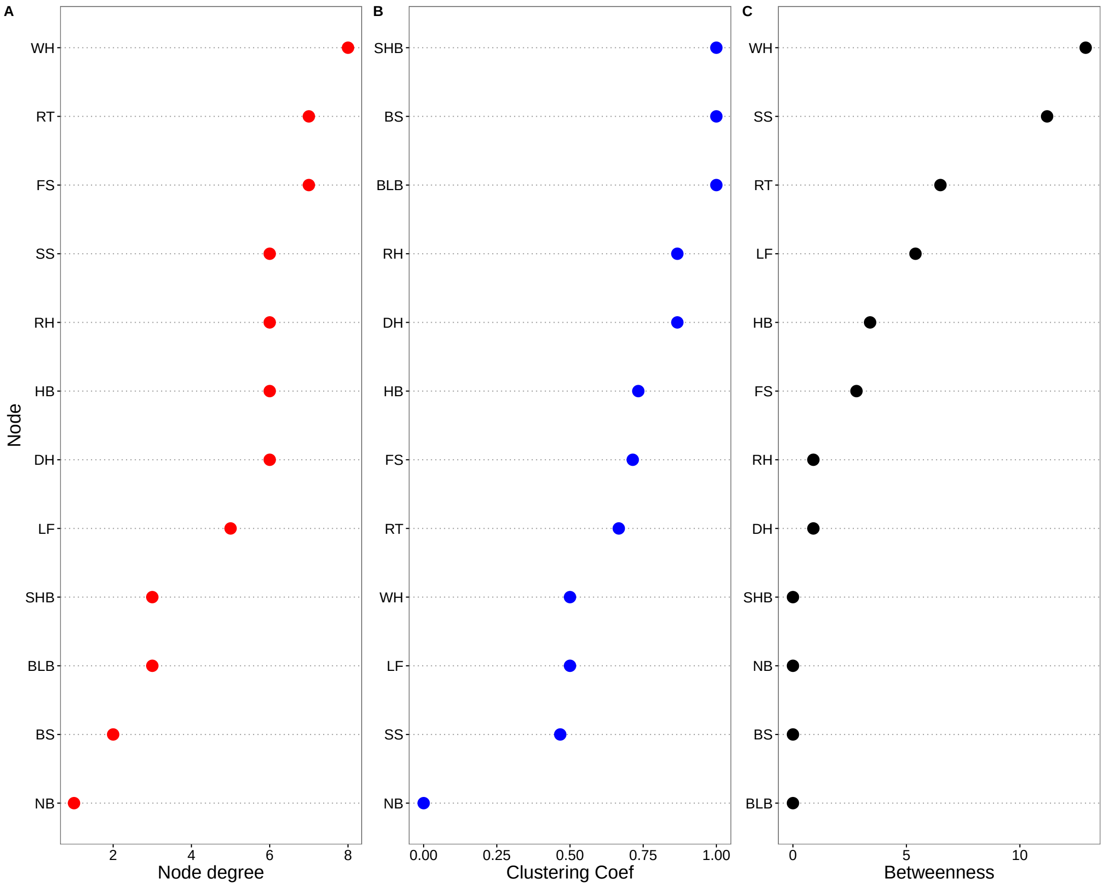


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Tamil Nadu, India. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure I-11: Rice injuries in dry season in Tamil Nadu, India



(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.

Figure I-12: Rice injuries in wet season in Tamil Nadu, India

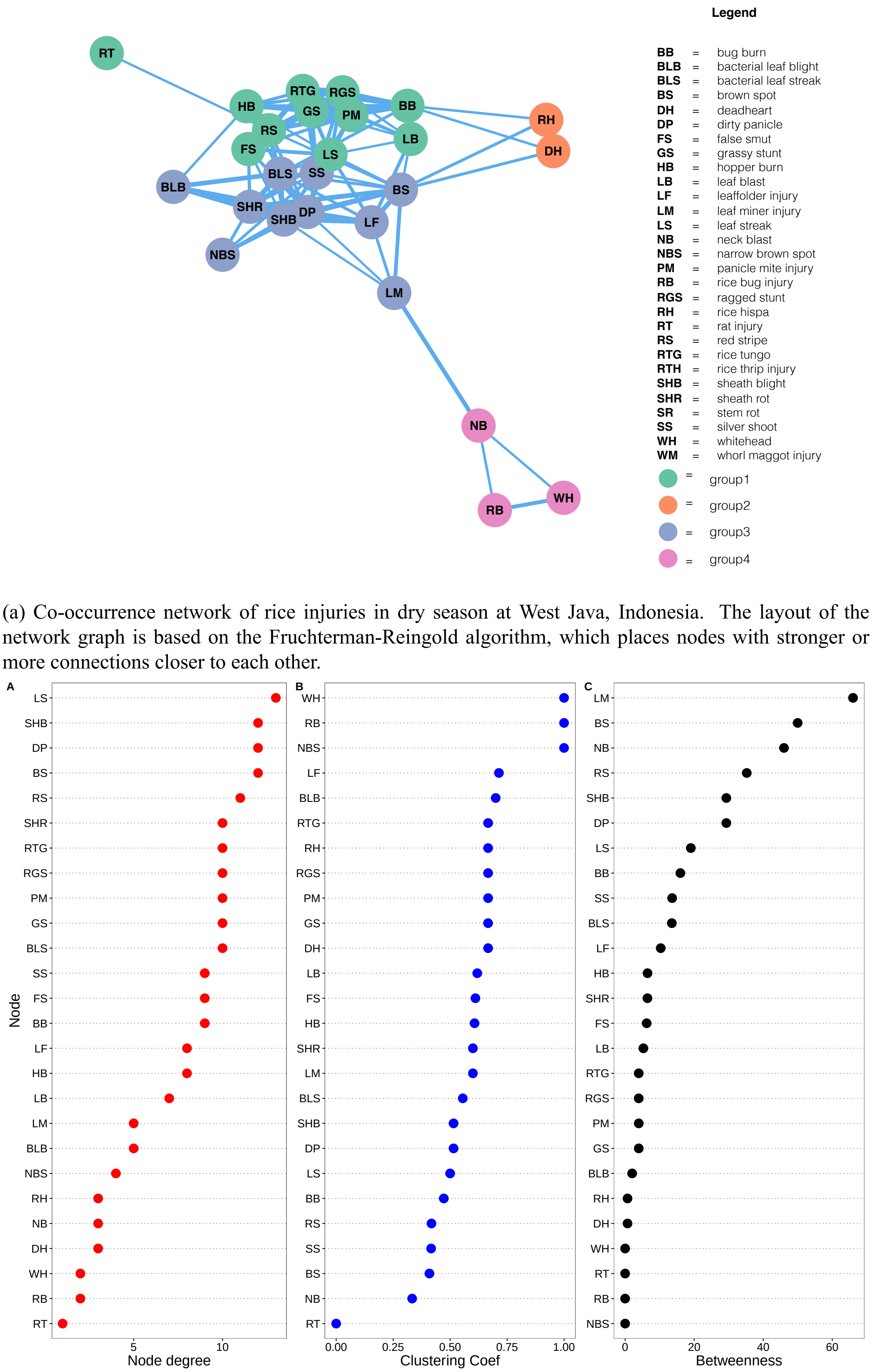
panicle or tiller injury syndrome because there is the combination of PM, RB, and FS. Apparently, within this combination, BS is center of association, and early occur among other injuries. All of injuries in syndrome3 (purple), which is combination of BS, LF, WM, NBS, and LM are leaf injures.

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

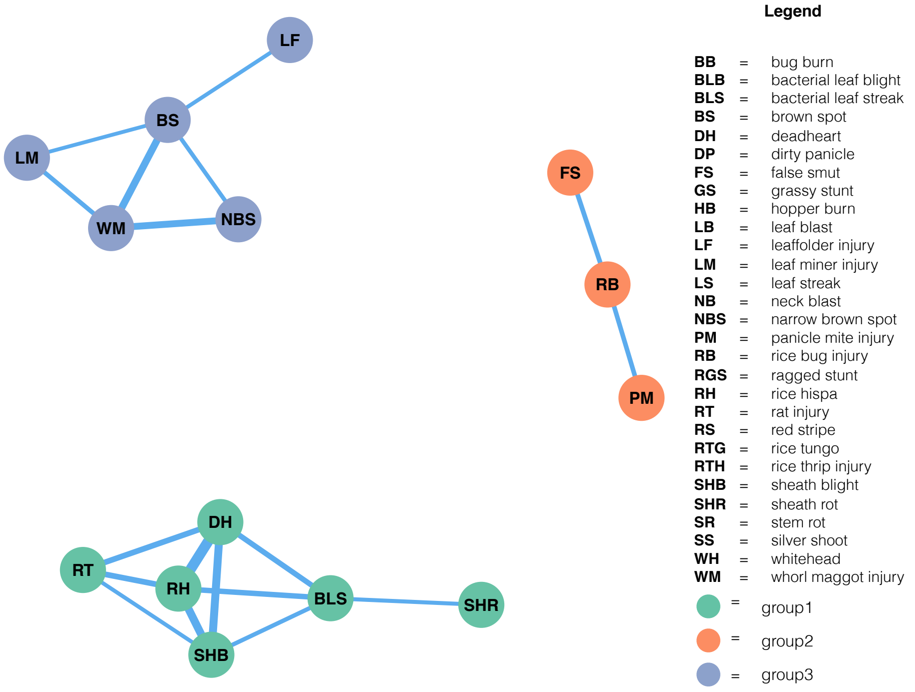
The exploration of co-occurrence networks is a useful method for determining interactions of co- occurring injuries. The centrality of nodes is considered further. The node centrality is the identification of which nodes are "central" than others (Barrat et al., 2004). Newman (2003) mentioned three measures of node centrality: node degree,



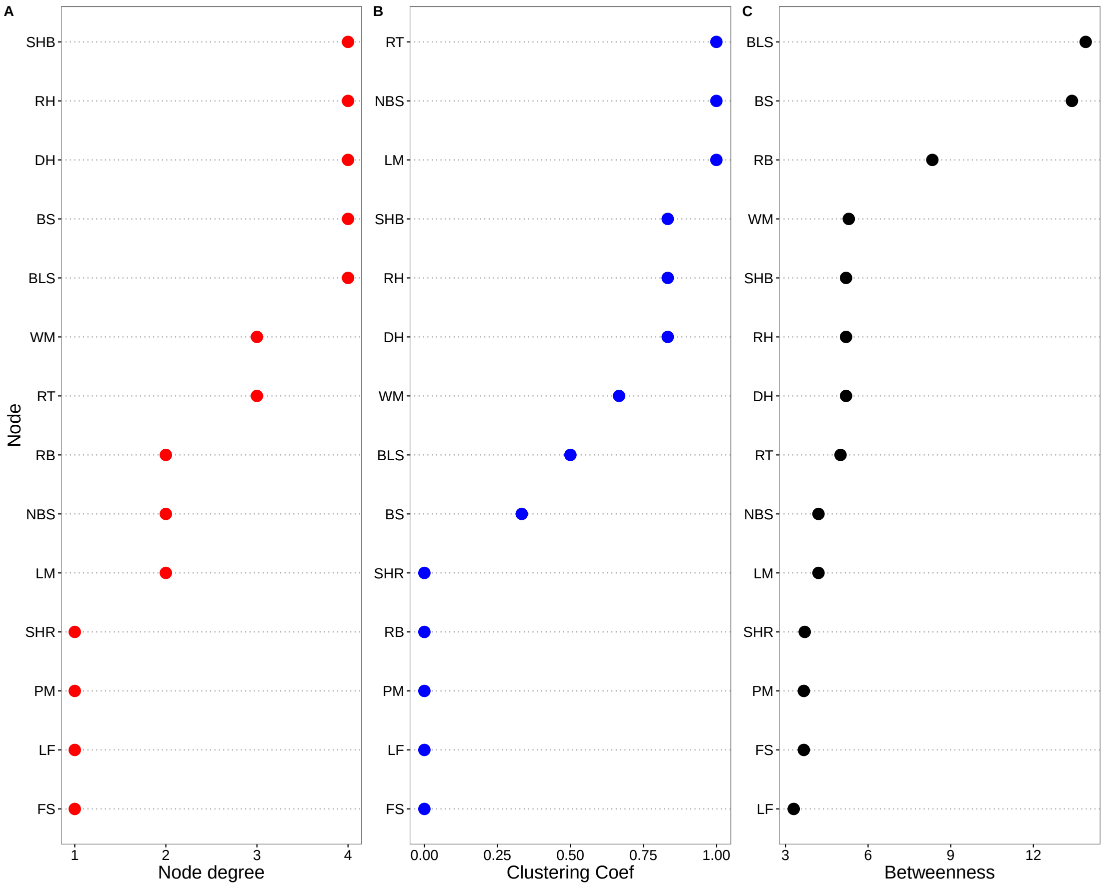
(a) Co-occurrence network of rice injuries in dry season at West Java, Indonesia. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

(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.

Figure I-13: Rice injuries in dry season in West Java, Indonesia



(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.

Figure I-14: Rice injuries in wet season in West Java, Indonesia

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).

Conclusion

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 injury syndrome from crop health survey data, which were collected in farmers' fields at 5 production environments () across South and Southeast Asia for 3 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. Different production seasons, and production environment contributed to different structure of networks. From the network structures, networks present 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), the networks can reveal 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, to develop the more effective strategies of pest management specifically to seasons or production environments (locations).

Important challenges, however; remain in order to further rice injury characterization:

- What are the rice injuries that occurred differently in season;
- at different yield levels, the patters of rice injury occurrence;
- differentiate the patterns for injuries at deferent production environments.

CHAPTER II

Differential networks reveal the dynamics of animal pests and disease co-occurrences

Introduction

Rice (*Oryza sativa*) is a major crop in South and Southeast Asian. Generally, rice farmers cultivate 2 rice crops per year, with the typical seasonal crop cycles or rotations being rice-rice-fallow or rice-rice-secondary crops (corn, soybean, peanut). The Food and Agriculture Organization of the United Nations (FAO) estimates that approximately 70 percent of total lowland rice area produces 2 rice crops each year. The first crop is cultivated 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). The previous studies showed that Injuries profiles (the combination of injuries) differ from season to season in term of weather pattern. In the dry season, crop losses were lower than in the wet season. A previous study based on surveys done in farmers' rice fields in the region of lowland rice were shown to be strongly associated with injury profiles.

In the previous chapter, the co-occurrence networks to yield 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) at different yield levels.

Differential network analysis aims to compare the connectivity of two nodes at 2 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^D_{xy} and r^W_{xy} 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 $A^{diff} = (A^{diff}_{xy})$ where the entry A^{diff}_{xy} quantified by following:

$$A^{diff}_{xy} = \begin{cases} 1 & \text{when } r^D_{xy} > r^W_{xy} \text{ at } P_{z_{xy}}\text{-value} < 0.05 \\ 0 & \text{when } P_{z_{xy}}\text{-value} > 0.05 \\ -1 & \text{when } r^W_{xy} > r^D_{xy} \text{ at } P_{z_{xy}}\text{-value} < 0.05 \end{cases} \tag{II.3}$$

For this differential co-occurrence network,A^{diff}_{xy} equals 1 depending on whether any injury pairs show significantly higher co-occurrence level 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^L_{xy} and r^H_{xy} 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 $A^{diff} = (A^{diff}_{xy})$ where the entry A^{diff}_{xy} quantified by following:

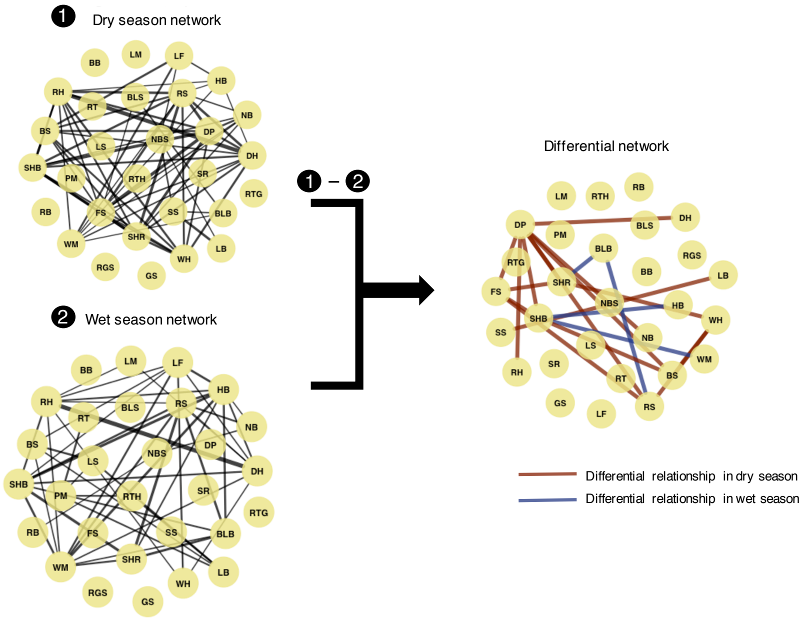


Figure II-1: Schematic showing principle of differential analysis in seasons. Co-occurrence networks are measured in each of two seasons (left) resulting in interactions (black). Dry season is subtracted from wet season to create a differential co-occurrence network (right), in which the significant 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 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.

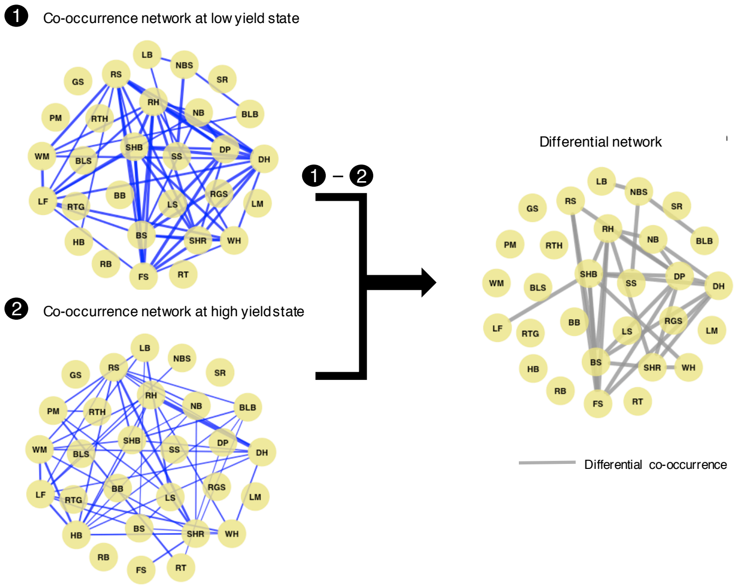

❶ Co-occurrence network at low yield state
❷ Co-occurrence network at high yield state
❶ – ❷
Differential network
Differential co-occurrence
LB
NBS
SR
GS
RS
RH
NB
BLB
PM
RTH
WM
BLS
SHB
SS
DP
DH
LF
RTG
BB
LS
RGS
LM
HB
BS
SHR
WH
RB
FS
RT

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.

II.1 Results

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

I determined which co-occurrence patterns of rice injuires of survey data were differentially expressed between dry and wet season. The resulted networks are showed in II-3) to Figure.II-7). Differential co-occurrence network (DCON) in seasons at CP (Figure. II-3) reveals SHB, RS, SHB showing significantly different co-occurrence level in both dry and wet season. DCON in season at OD **??** reveals LB different both in dry and wet season. DCON in season at RR II-5 reveals that DP, BS, RTH, LF. DCON in season at WJ (Figure.II-7 revealed that BS, NBS, SHB BLS. II-6

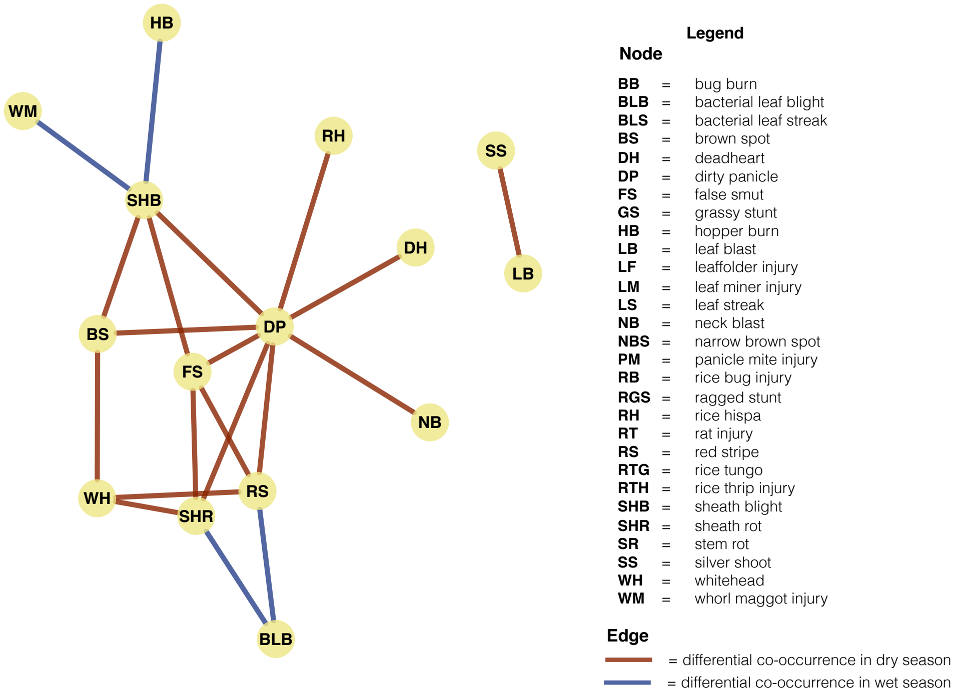


Figure II-3: Differential co-occurrence network of rice injuries in different seasons at Central Plain, Thailand

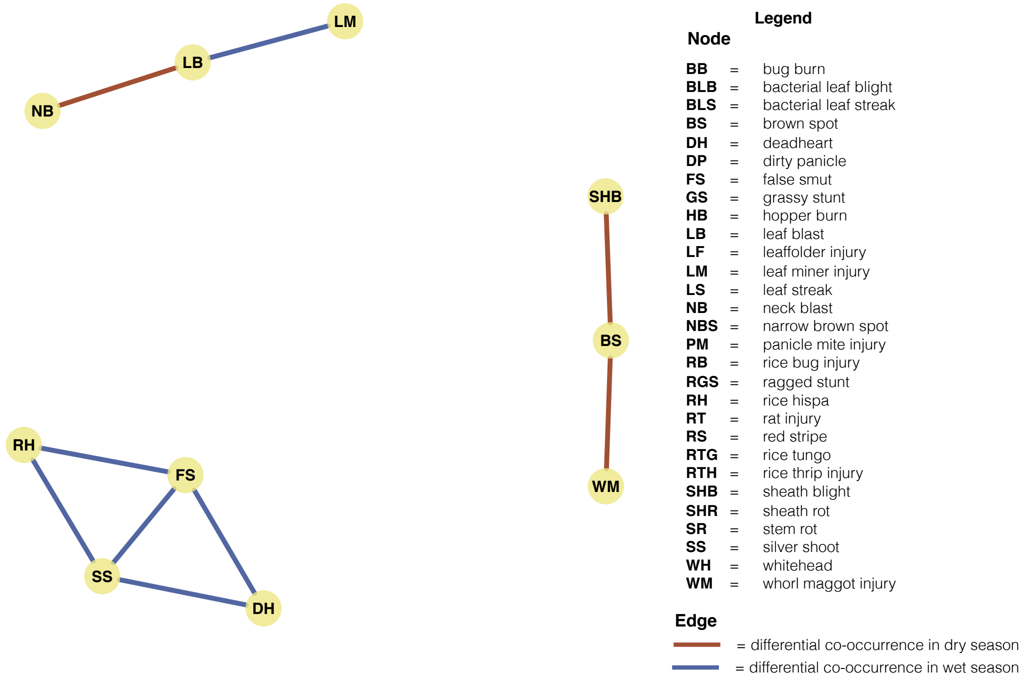


Figure II-4: Differential co-occurrence network of rice injuries in different seasons at Odisha, India

Construction of differential co-occurrence networks of rice injuries at 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 CP, low and medium yield level of TM, medium and high yield level at RR, low and medium yield level of TM, and

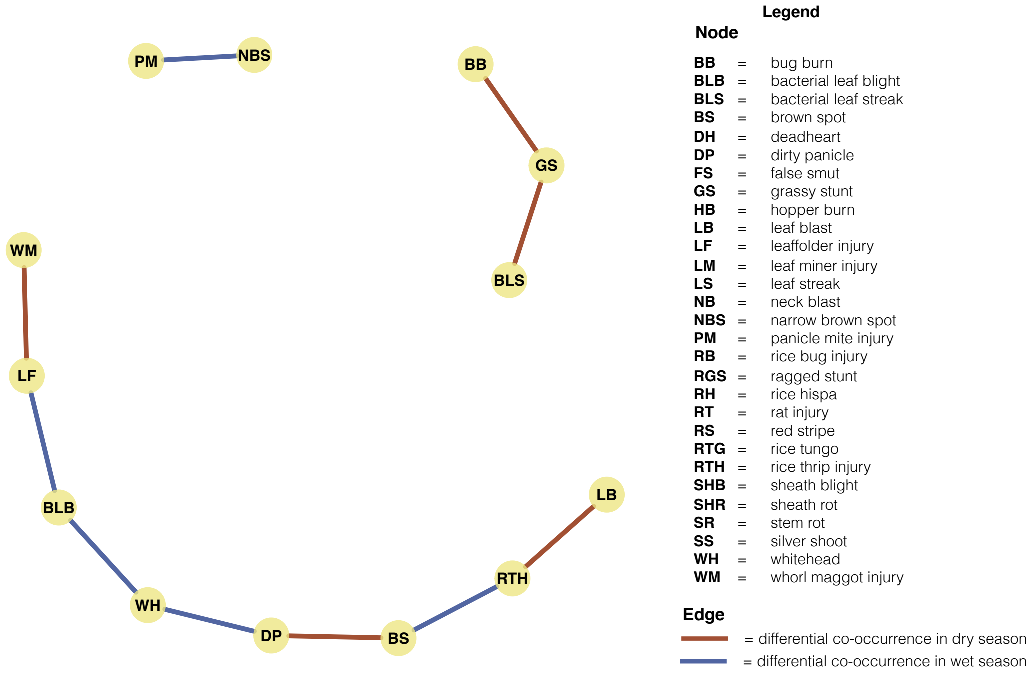


Figure II-5: Differential co-occurrence network of rice injuries in different seasons at Red River Delta, Vietnam

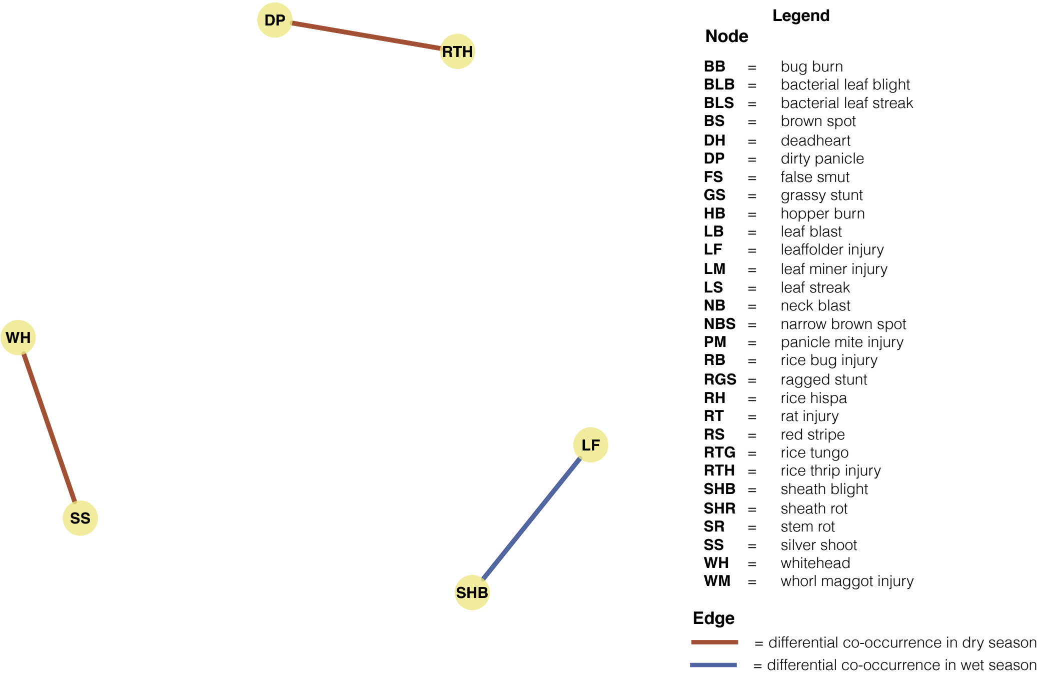


Figure II-6: Differential co-occurrence network of rice injuries in different seasons at Tamil Nadu, India

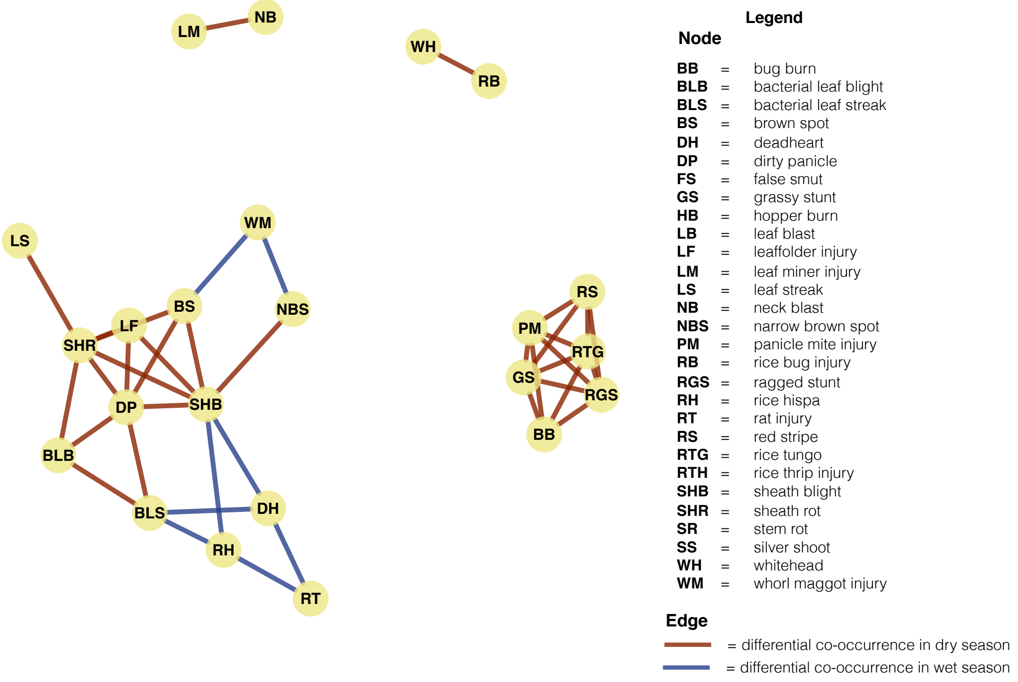


Figure II-7: Differential co-occurrence network of rice injuries in different seasons at West Java, Indonesia

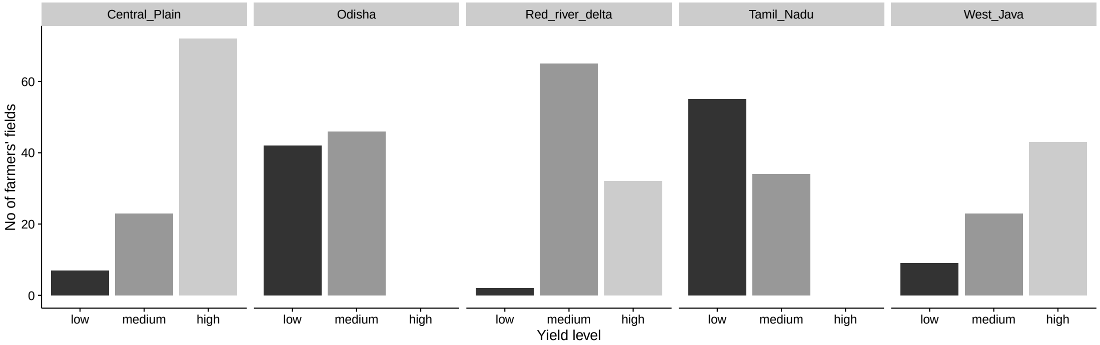
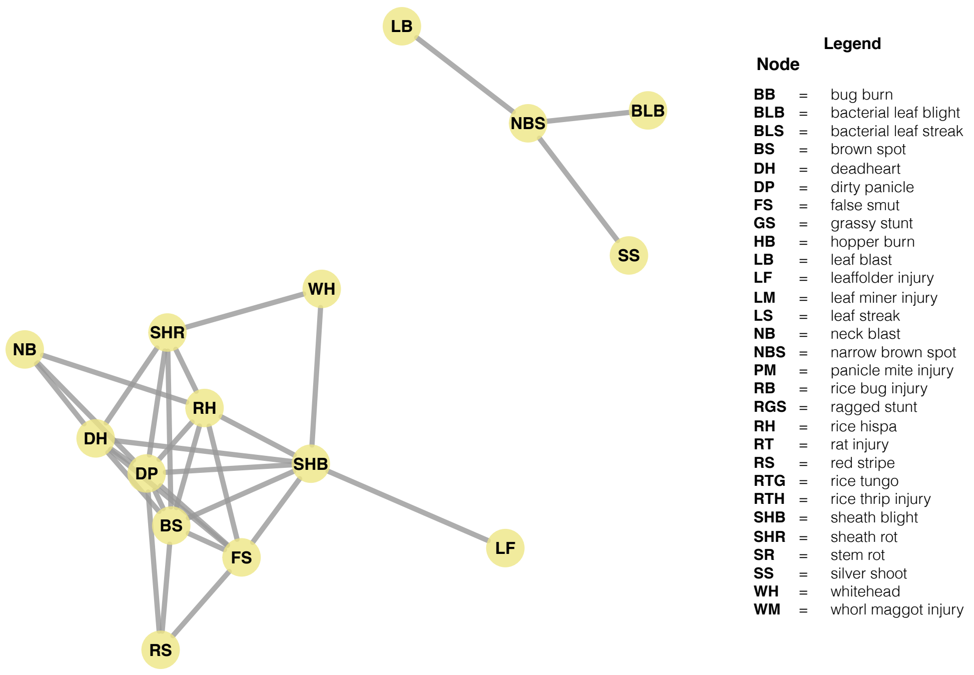


Figure II-8: .

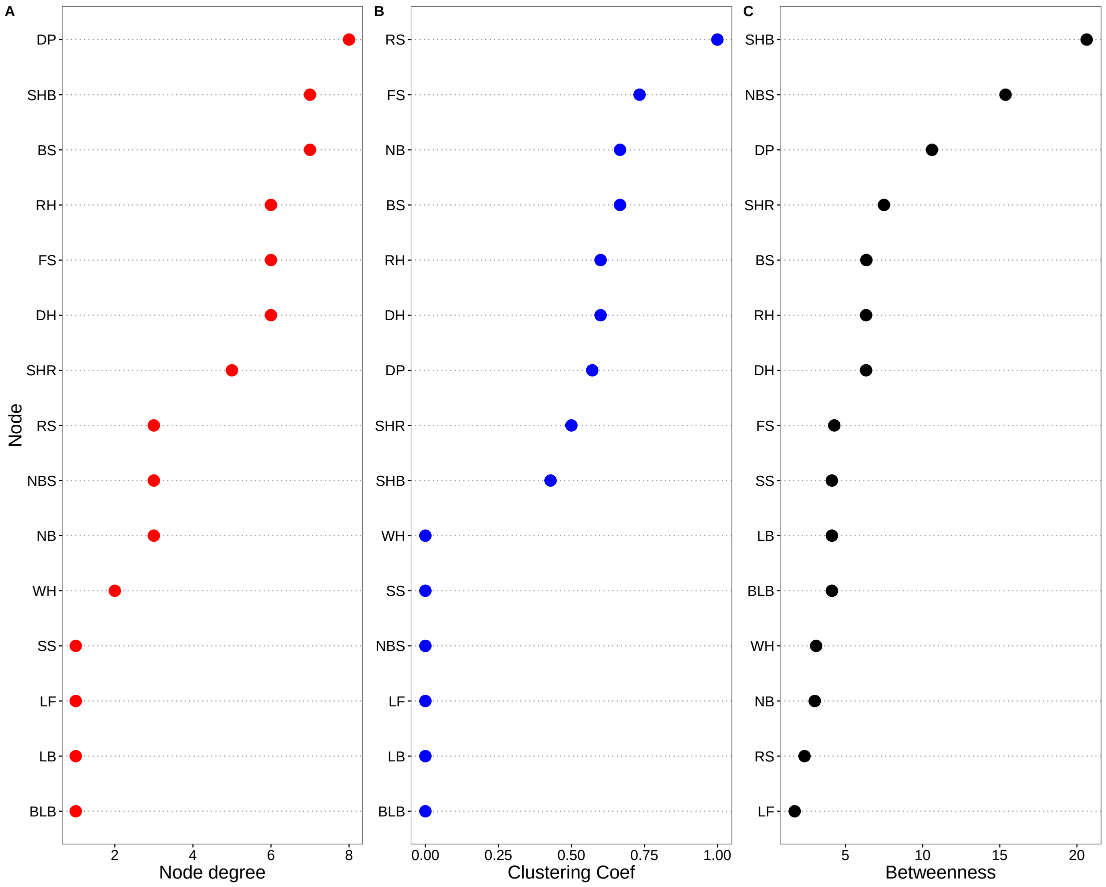
medium and high yield level of WJ.

In this study, an differential co-occurrence network (DCON) in yield depicts the associations of injury pairs presenting in lower yield state but absent in higher yield state (Figures. **??**, II-10, II-11a, **??**, **??**). DCON

Discussion

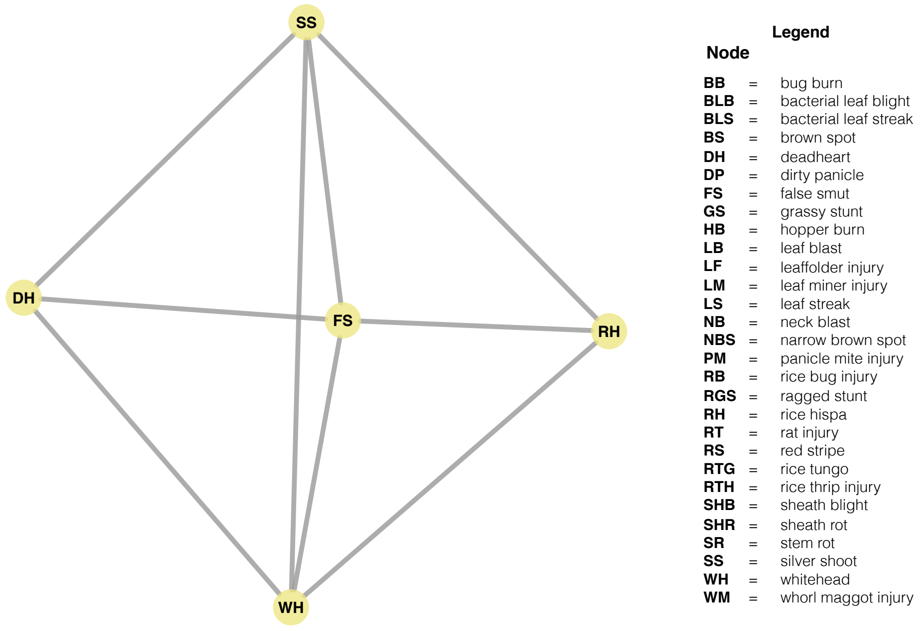


(a) Differential co-occurrence network of rice injuries in different yield levels at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

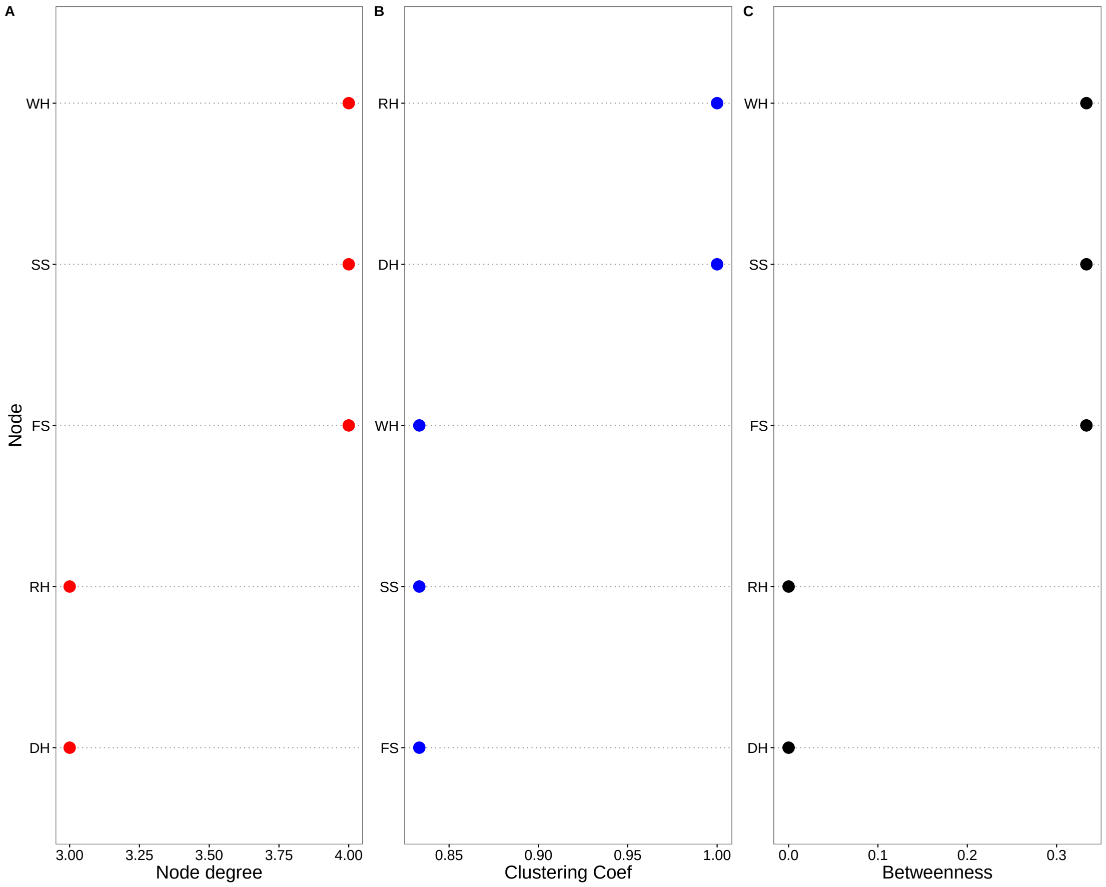


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure II-9: Rice injuries in dry season in Central Plain, Thailand

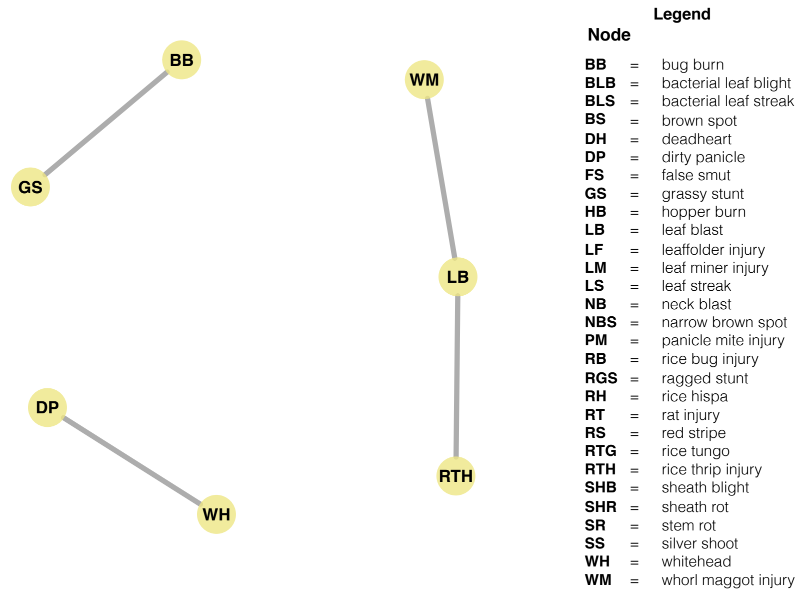


(a) Differential co-occurrence network of rice injuries in different yield levels at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

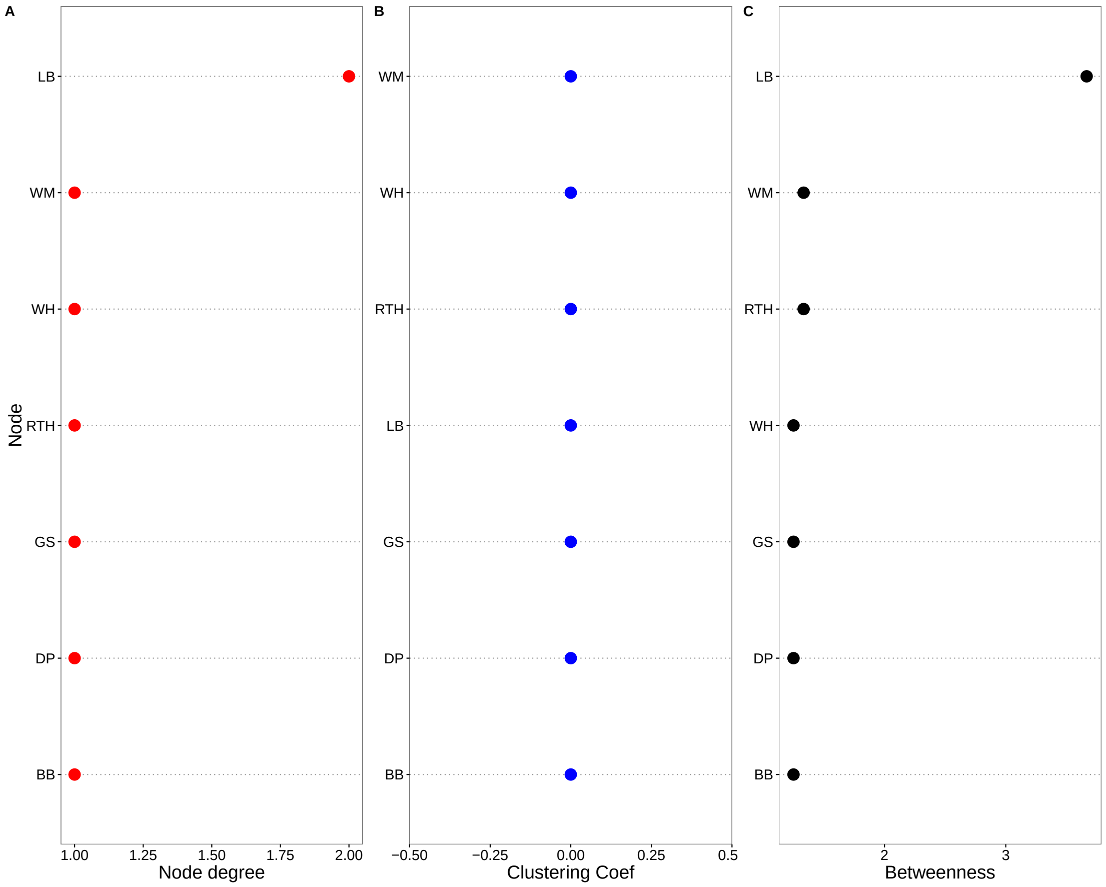


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure II-10: Rice injuries in dry season in Central Plain, Thailand

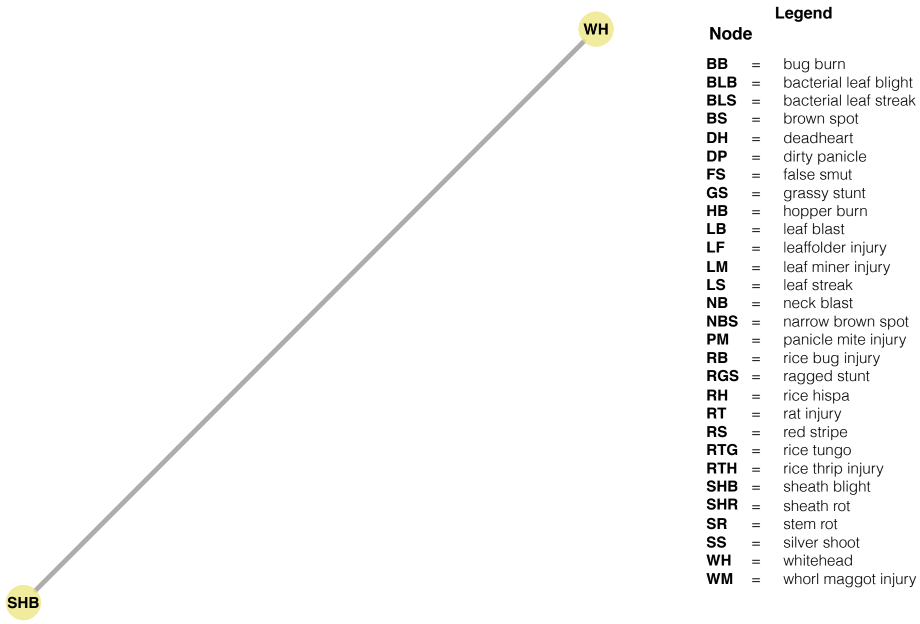


(a) Differential co-occurrence network of rice injuries in different yield levels at Central Plain, Thailand. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other.

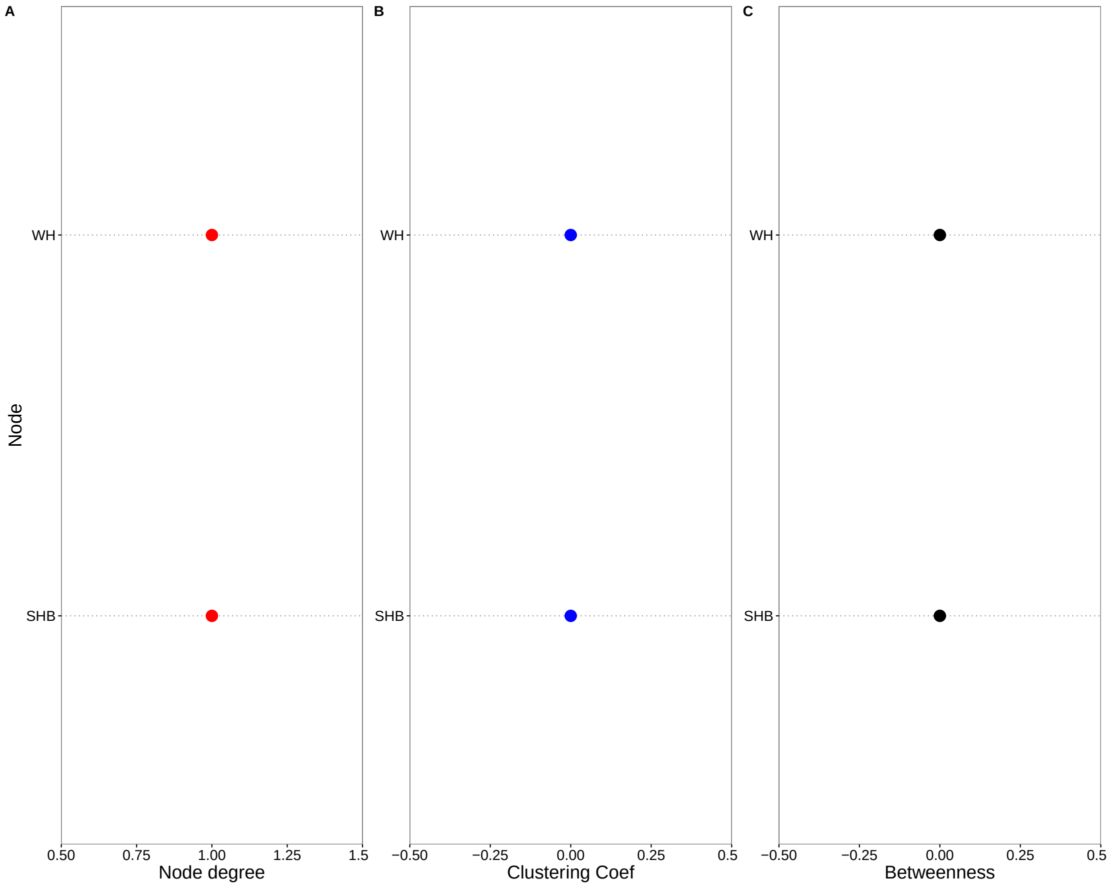


(b) Three centrality measures of the nodes in co-occurrence network of rice injuries in dry season at Central Plain. A: node degree, B:clustering coefficient, and C:Betweenness.

Figure II-11: Rice injuries in dry season in Central Plain, Thailand



(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-12: .

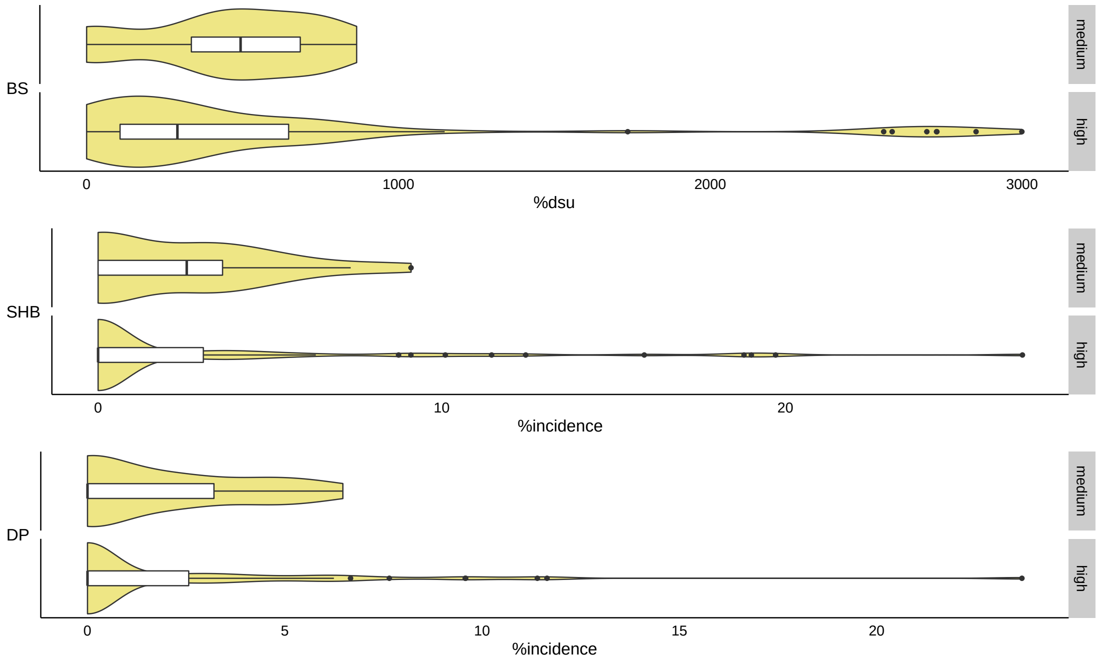


Figure II-13: .

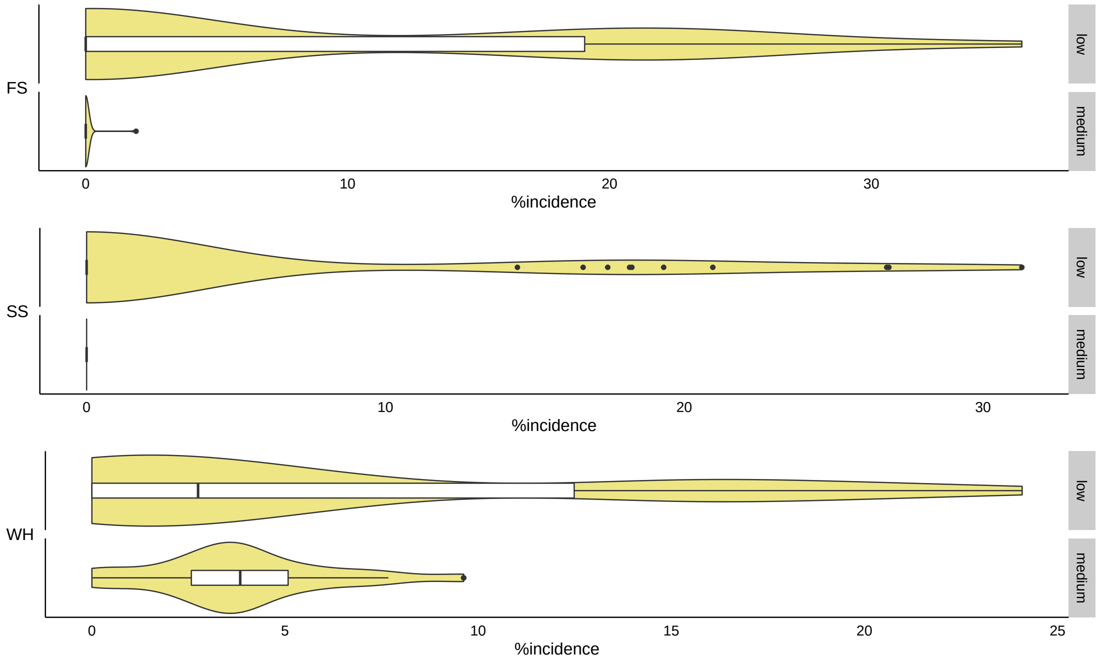


Figure II-14: .

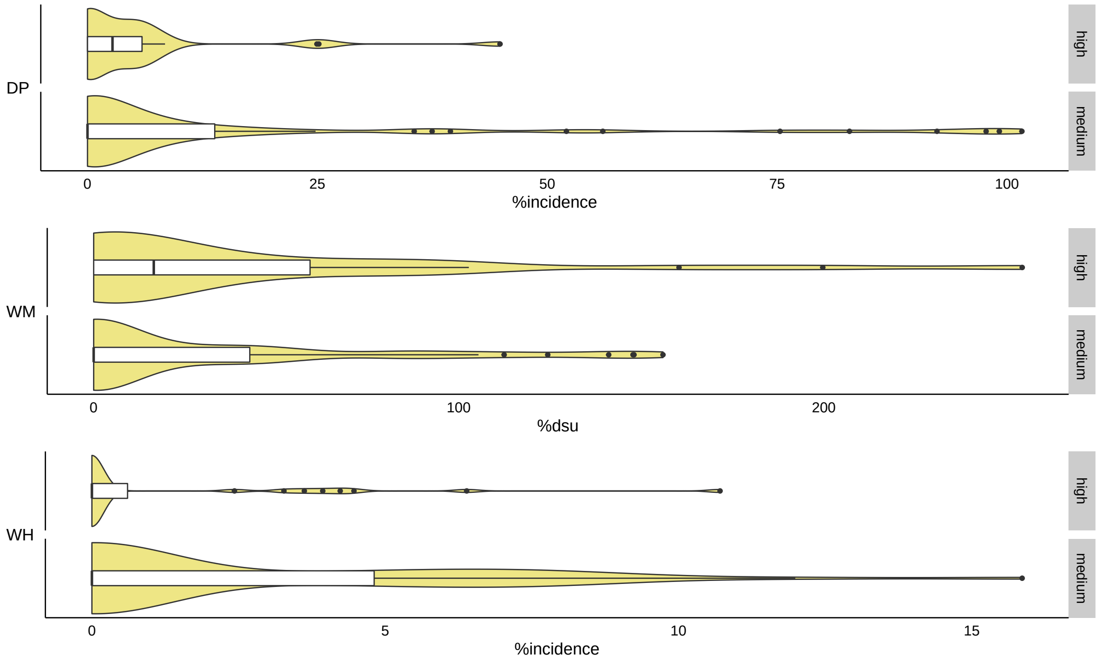


Figure II-15: .

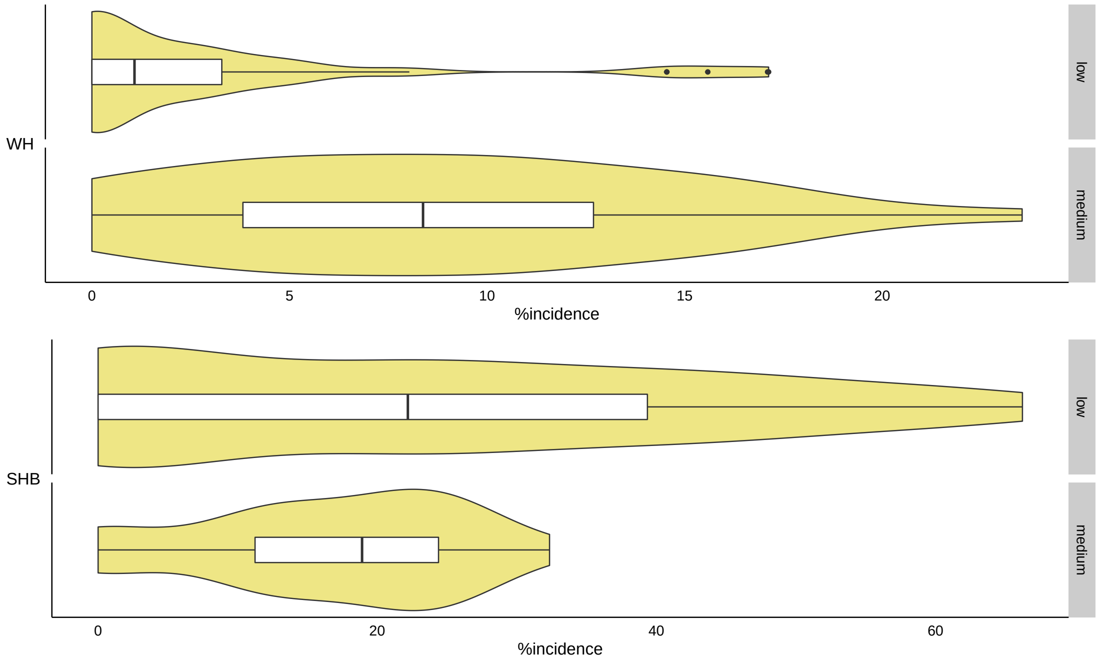


Figure II-16: .

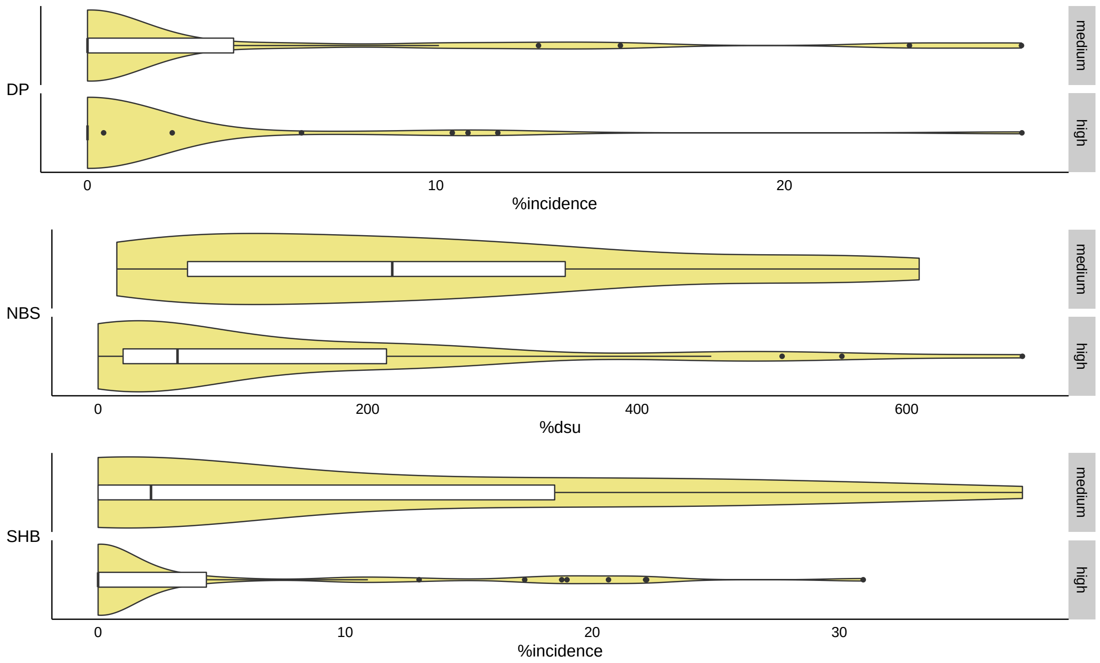

DP
medium
high
0
10
20
%incidence
NBS
medium
high
0
200
400
600
%dsu
SHB
medium
high
0
10
20
30
%incidence

Figure II-17: .

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.

BARABÁSI, A.-L. and OLTVAI, Z. N. 2004. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5(2):101–113.

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.

BARNWAL, M., KOTASTHANE, A., MAGCULIA, N., MUKHERJEE, P., SAVARY, S., SHARMA, A., SINGH, H., SINGH, U., SPARKS, A., VARIAR, M., ET AL. 2013. A review on crop losses, epidemiology and disease management of rice brown spot to identify research priorities and knowledge gaps. European Journal of Plant Pathology, 136(3):443–457.

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.

CAMPBELL, C. L., MADDEN, L. V., ET AL. 1990. Introduction to plant disease epidemiology. John Wiley & Sons.

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

DOANE, D. P. and SEWARD, L. E. 2011. Measuring skewness: a forgotten statistic. Journal of Statistics Education, 19(2):1–18.

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.

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.

GHASEMI, A., ZAHEDIASL, S., ET AL. 2012. Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2):486–489.

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.

KUMARI, S., NIE, J., CHEN, H.-S., MA, H., STEWART, RON, L., XIANG, LU, M.-Z., TAYLOR, W. M., and WEI, H. 2012. Evaluation of gene association methods for coexpression Network construction and biological knowledge discovery. PLoS ONE, 7(11).

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.

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.

MOSLONKA-LEFEBVRE, M., FINLEY, A., DORIGATTI, I., DEHNEN-SCHMUTZ, K., HARWOOD, T., JEGER, M. J., XU, X., HOLDENRIEDER, O., and PAUTASSO, M. 2011. Networks in Plant Epidemiology: From Genes to Landscapes Countries, and Continents. Phytopathology, 101(4):392–403.

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.

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

PEAT, J. and BARTON, B. 2005. A guide to data analysis and critical appraisal. Wiley Online Library.

PROKHOROV, A. 2001. Kendall coefficient of rank correlation. Online Encyclopedia of Mathematics.

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. and CASTILLA, N. 2009. A survey portfolio to characterize yield-reducing factors in rice. IRRI Discussion Paper No 18.

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.

TOUBIANA, D., FERNIE, A. R., NIKOLOSKI, Z., and FAIT, A. 2013. Network analysis: tackling complex data to study plant metabolism. Trends in Biotechnology, 31(1):29–36.

WILCOX, R. R. 2012. Introduction to robust estimation and hypothesis testing. Academic Press.

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.

ZHANG, B. and HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology, 4(1).