\subsection{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. Because of this co-occurrence in injuries, the idea of ``crop health" has been highlighted and implemented to manage the combination of injuries or so-called injury profiles \citep{Savary\_2006\_Quantification}. Co-occurrence patterns of injuries are beginning to provide important insight into these injury profiles, which possibly present co-occurring or anti-co-occurring (mutually exclusive) relationships between injury-injury. Uncovering these patterns is important to implications in plant disease epidemiology and management. However, there are only a few reports of injury–injury relationships in rice crop systems are currently unknown. This could be a difficult task since complex patterns of injury profiles are related to environmental conditions, cultural practices, and geography \citep{Willocquet\_2008\_Simulating}.

To address this issue, we used in-field surveys as a tool to develop ground-truth databases that allowed us 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, field management, pest injuries, and yields. Several previous studies \cite{Savary\_2000\_Characterization,Savary\_2000\_Quantification,Dong\_2010\_Characterization} and \cite{Reddy\_2011\_Characterizing} involved surveys that were used to characterize injury profiles in an individual production situation (a set of factors including cultural practices, weather condition, socioeconomics, \textit{etc}.) that determine agricultural production, and the injury profiles using nonparametric multivariate analysis such as cluster analysis, correspondence analysis, or multiple correspondence analysis. Their results led to the conclusions that injury profiles (the combination of disease and pest injury that may occur in a given farmer’s field) were found co-occurrence patterns across sites, which are associated at regional scale. For example, stem rot, sheath blight, planthopper, and rice whorl maggot injuries, are high incidence, with low incidence of brown spot, and absence of bacterial leaf blight, leaf blast, and neck blast are a common pattern in tropical Asia from the study of \citet{Savary\_2000\_Characterization}.

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 \citep{Faust\_2012\_Microbial\_co,Ma\_2016\_Geographic}. Co-occurrence patterns are ubiquitous and particularly important in understanding community structure, offering new insights into potential interaction network. Recent reviews of network based approaches reveal 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 \citep{Williams\_2014\_demonstrating, Barberan\_2012\_Network}, or key actors in social networks \citep{Crowston\_2006\_Hierarchy}. Here, we significantly advance this study by providing a comprehensive understanding of the topological shifts of animal pest injury and disease co-occurrence networks at regional scale.

South and Southeast Asia represent big bowl of rice for the world population. Comparing the topological properties of the node associated with occurrence in the different countries and examining network level topological features can provide us with insight into variation in the co-occurrence patterns of rice injuries in different countries. This approach helps contextualize the animal pests -- disease association by taking to account the complex network of potential association among animal pest and disease occurring in farmers's fields in theses countries. Specifically, we addressed the following questions:(i) How can the co-occurrence relationships of rice injuries be examined from the perspective of network analysis (ii) Which animal injuries and diseases are found commonly close co-occurrence patterns among other variables in order to target to control or monitor. To answer these questions, we performed crop health survey at the farmers's fields in two different seasons and five countries in South and Southeast Asia and implemented co-occurrence network analysis to examine the topological feature differences across countries. Our main objective was to characterize and better understand co-occurrence networks in the association of rice animal pests and diseases.

%As the size and extent of biological data sets grow, scientists turn to new techniques, such as network analysis, to understand biological complexity over large scales. For network analysis of microbial datasets, topological ‘co-occurrence’ networks are generated from correlative metrics, where nodes represent observed variables, and significant correlations are represented by the edges connecting them. We used an unprecedented

\subsection{Materials and Methods}

\subsubsection{Study sites, data collection}

We surveyed farmers' fields in 5 different locations;Tamil Nadu, India (TMN); West Java; Indonesia (WJV), Laguna, Philippines (LAG); Suphanburi, Thailand (SPB) and Mekong river delta, Vietnam (MKD) for consecutive years (2009 - 2013). These survey sites are in the major irrigated lowland rice growing areas, where rice is intensively cropped at least two seasons per year. The 412 individual fields were surveyed in both dry and wet seasons of 2010 to 2012 in India, Indonesia, Thailand, and Vietnam, except Philippines, which the surveys were only conducted in 2011. The number of survey were summerize in Table

\ref{table:Survey\_data}.

The survey procedure and data were based on a standardized protocol described in ``A survey portfolio to characterize yield-reducing factors in rice'' developed by \citet{Savary\_2009\_Survey}. Rice injuries were collected in this study including the injuries caused by animal pests, and pathogens, which are harmful to rice plants, and importantly considered to reduce yield productivity. They were evaluated at booting and ripening stage according to survey procedure. 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) 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), gall midge injuries (GM) were determined as a proportion of injured tillers or panicles. Insect pests including brown planthopper (BPH), white backed planthopper (WPH), green leafhopper (GLH), rice bug (RB) or stink bug were determined as number of insect found on the rice hill.

Before analysis, data were compacted over time during crop growth. Two types of data were computed, depending on the nature of injury \citet{Savary\_2009\_Survey}. 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) \citep{Campbell\_1990\_Introduction} were calculated by the mid-point method using the following equation:

\begin{equation}

AUIPC = \sum{\frac{1}{2(X\_{i} + X\_{i-1})(T\_{i} - T\_{i-1})}}

\end{equation}

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 $i$th 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 $i$th observation and $n$ is total number of observations.

\subsubsection{Network construction}

The co-occurrence network was inferred based on the Spearman correlation matrix constructed with \texttt{R} function \texttt{cor.test} with parameter method `Spearman' (package stats) was used for calculate Spearman's correlation coefficient ($\rho$), which is defined as the Pearson correlation coefficient between the ranked variables \cite{R\_2015}. The nodes in this network represent injuries and the edges that connect these nodes represent correlations between injuries. Based on correlation coefficients and $P$-values for correlation, we constructed co-occurrence networks.The cutoff of $P$-values was 0.05. Network properties were calculated with the \textbf{igraph} package \citep{Csardi\_2010\_igraph}. All farmers' fields were divided into groups by country and season. The impact of each sample group on the Spearman correlation value of each edge in the network was assessed by Spearman correlation value of these fields. The network of each of group was detected the community structures by maximizing the modularity measure over all possible partitions by using \texttt{cluster\\_optimal} function of \textbf{igraph} package \cite{Brandes\_2008\_Modularity}.

\subsubsection{Topological feature analysis}

We calculated the topological features for each network with the \textit{igraph} package. We measured two levels of network topologies. The node features that we focused on are node degree, betweenness, and clustering coefficient. 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.

Global features including network clustering coefficient, average path length, and diameter 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. In our analysis, both diameter and average path length are considered measures of the size of the network. Larger networks are less connected, meaning that the likelihood of a strong connection between any two randomly selected species is low.

The network clustering coefficient are considered measures of the complexity of the network. The networks are more complex, the network has higher clustering coefficient, and shorter average path length.

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

\newpage

\subsection{Results}

\subsubsection{Crop health survey data}

Prior to using the crop health survey data to construct networks, some exploratory plots of the data are presented and examined. The figure shows the distribution of the data. A total of 415 farmers' fields was surveyed in Tamil Nadu, India (TMN); West Java; Indonesia (WJV), Laguna, Philippines (LAG); Suphanburi, Thailand (SPB) and Mekong river delta, Vietnam (MKD) from 2009 to 2013. The data collected pertain to rice injuries caused by animal pests, and pathogens. The protocol gives more emphasis on the nature of injuries and not on the causal organism. The levels of injuries during a cropping season were computed or summarized according to the nature of injuries and thus varied in scale. For example, foliar injuries, diseases on leaves, and insect pests were summarized as area under the disease incidence progress curve, and tiller injuries and diseases as a maximum incidence.

The injuries caused by animal pests observed during the survey period were rat injury (RT), deadheart (DH) and whitehead (WH) caused by stem borers, whorl maggot injury (WM), leaffolder injury (LF), gall midge injury or silver shoot (GM). Rat injuries were observed at all 5 survey locations with low incidence ( less than 20\% incidence). We could observe 75\% incidence of the rat injury in MKD in dry season. They were also observed at WJV in dry season, and both season in TMN, LAG, SPB. Gall midge injuries during survey period were not observed in TMN and LAG, but it was found in SPB, MKD, but WJV at 25 \% incidence. Deadheart were observed all survey sites, and it was severe in dry season at SPB and MKD, but in WJV and LAG, it was severe in dry season. The trend of whitehead incidences observed was opposite the deadheart incidence, which whitehead incidences were more server in wet season at SPB and MKD, but less severe at WJV and LAG. Leaffolder injury was observed all survey locations. The leaffolder incidences were more severe in wet season the dry season at WJV, TMN, and LAG. As apposite to SPB and MKD, they were more severe in dry season than wet season. Whorl maggot injury was observed at all locations. Mostly, they were more severe in wet season than dry season at all surveyed locations, except LAG.

The disease we 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). Rice diseases we observed in this study were commonly found at all locations, but there were some diseases that could not find especially in TMN such as BLS, BS, NBS, DP, RS, SHR, and SR. BLB was observed all location. BLB incidence was higher in dry season than wet season in WJV and MKD except in TMN, LAG, SPB. Wet season was more favorable for BLS than dry season because the incidence was higher in wet season than dry season. As same as BLS, BS incidence was higher in wet season than dry season, and it was severe in SPB. Even though, LB is common disease in these survey locations, but there were some farmer’s fields observed LB. In WJV and MKD, the LB incidence was higher in dry than wet season season. There were many field in SPB found high level of NBS incidence. Like BLS and BS, NBS incidence was more severe in wet season than dry season. There were many fields found high RS incidences. The highest incidences of DP were found in MKD. FS were commonly found at all location. The high incidents were observed in SPB, and MKD. NB were observed at all location. They were found many observations in TMN. SHB commonly was found all location, and high incidence was in TMN, and LAG.

Brown planthoppers (BPH) were found all location of survey sites. During the survey period, they were found higher population in wet season than dry season in TMN and MKD, but there was some observation of BPH at other locations. White backed planthoppers (WPH) were commonly found at MKD in both of dry and wet season, and there are a few observations in WJV and LAG, but at TMN and SPB, they were not observed. Rice bug (RB) could be observed all survey location. They were highly found in dry season than wet season at WJV, LAG, and MKD. but in SPB, they were found only in wet season during the survey period. Green leafhoppers were observed all location. They were found in LAG higher than other locations.

A global network of microbial co-occurrence and mutual exclusion within and among body site niches of the human microbiome

%Global properties of the microbiome-wide network of microbial associations are summarized in Figures 2 and 3. A dominant characteristic of the network was its habitat-specific modularity. After grouping the 18 body sites into five broad areas (oral, skin, nasal, urogenital, and gut), the large majority of edges were found clustered within body areas (98.54%), and these clusters were sparsely connected through a minority of edges (1.46%). This is confirmed by the network's high modularity coefficient of 0.28 (as defined by [32]) and Markov clustering of the network (see Methods and Figure S2). It has long been observed that sites within the human microbiome are distinct in terms of microbial composition [33], and this proved to be true of microbial interactions as well: microbial relationships within each body area's community were largely unique (Table 2). The microstructure of interaction patterns - and thus in the underlying ecology - was different for different areas, however. For example, all vaginal sites within the urogenital area were interrelated in a single homogeneous community, whereas interactions within the oral cavity suggested microbial cross-talk among three distinct habitats [34]. This can be observed quantitatively based on the proportions of microbial interactions spanning body sites within each area, e.g. 69.57% among the vaginal sites and 53.19% among the oral sites, both exceeding the microbiome-wide baseline. The skin was further unique in that the large amount (57.65%) of its associations related microbes in corresponding left and right body sites (left and right antecubital fossae and retroauricular creases), reflecting consistent maintenance of bilateral symmetry in the skin microbiome.

\subsubsection\*{Co occurrence of injury profiles based on network analysis}

Co-occurrence correlations of injuries were explored using network inference based on significant correlations using non-parametric Spearman’s rank coefficient at $P$ < 0.05. The setup of Spearman's coefficient cutoff at significant correlation coefficient at at $P$ < 0.05 could efficiency reduce the sparse correlation and highlight the significant correlation between variables. The networks were determined for the co-occurrence analysis of rice injuries based on the survey data. Ten networks constructed were based on the survey data grouping different location (west java; Indonesia (WJA), Tamil Nadu; India (TMN), Laguna; Philippines (LAG), Suphan Buri; Thailand (SPB), and Mekong Rive Delta; Vietnam (MKD)) and seasons (dry, and wet season). Nodes represent rice injuries, and edges represent Spearman's correlation coefficients at $P$-value < 0.05. Once a network has been constructed, analytic tools and measures can be used to quantify by determining the structural properties.

\includegraphics{figures/simgraph.png}

\begin{table}

\centering

\caption{Node topology}

\begin{tabular}{lllll}

Node & Node degree & Clustering coefficient & Betweenness \\

A & 2 & 0 & 0 \\

B & 6 & 0 & 14 \\

C & 6 & 0.06 & 12 \\

D & 4 & 0.17 & 0 \\

E & 4 & 0.17 & 0 \\

F & 2 & 0 & 0

\end{tabular}

\end{table}

Measures of node properties of networks seek to identify the most important nodes in a network. Different measures indices the different contexts for the word ``importance''. In this study, we consider 3 features of node properties.

\begin{itemize}

\item \textbf{Node degree} defined as the number of connections of a node. Nodes with high degree can be implied that the node has many connections or relationships with other nodes. In co-occurrence network, the high degree node can be inferred that it was found together with occurs with connected nodes. For example, in fig(), node B, and node C are the highest-degree nodes in this network. Node B have the relationships with 3 nodes. We can consider node B as the important node of this network because If we remove this node, there will be two nodes (node A and F) would disappear from this network.

\item \textbf{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. In the context of this network, local clustering coefficients is the the probability that two neighbors of connected with each other in the network. These measures are indicative of the complex forming of co-occurrence patterns through the network. As presented nodes can be observed other nodes, a more closely connected network facilitates nodes co-occurrence. The higher clustering coefficient injury has high possibility to form complex. Node D and E have high clustering coefficient. They formed the complex relationships, which Node D connected with Node E connected to Node C, and Node C also connected with Node D. It indicated that they have close relationships. In the context of co-occurrence network, these three nodes usually occur together.

\item \textbf{Betweenness} is measure the number of paths, which a node is present on the shortest path between all other nodes. Nodes with high betweenness have been shown that there are many pair-wise relationships across a network pass through the node, which is implied that the nodes has high possibility to occur. We determine high-betweenness injuries as``indicator''.

\end{itemize}

Moreover, I inspected the community structure of the networks based on the optimal clustering algorithm. Such communities correspond to groups of injuries that are closely associated among themselves than other nodes in one others groups.

\subsubsection{Communities, Structures and compositions of co-occurrence network of rice pest injuries}

\paragraph{Tamil Nadu, India (TMN)}

Dry season network (Fig) was composed of 6 associated injuries (DH, NB, SR, FS, RB, and LF) and captured 7 associations. The network showed two groups of injury syndromes (the combination of injuries). The groups of the injury profiles corresponded to each community based on the optimal clustering algorithm. The first group is composed of DH, LF, RB and SR. Another group consists of FS and NB. Analysis of network properties revealed that LF and FS are high-betweenness nodes. They presented co-occurrence relationships within group 1 and group 2. SR and RB have high clustering coefficient. It indicated that these two injuries usually formed complex of co-occurrence relationships. As opposed to other injuries, NB and DH have low scores on all three centrality measure. Apparently, these injuries less possibly co-occur with other injuries (low betweenness), and do not have complex co-occurrences with other injuries (low degree and clustering coefficient).

Wet season network (Fig) was composed of 12 nodes (injury variables, DP SR, FS, BLB, NB, LF, GLH, RT, RB, DH, WM, BPH) with edges. Fig reveals three groups of injury profiles. DH, GLH, SR and BPH are in the group 1 (green). FS, NB, DP, BLB is in group 2 (orange). RT, LF, RB and WM are in group 3 (purple). Top four injuries with high betweenness, DP, SR, LF, and BLB, are the members of each of group. They possibly are found co-occurrence within the group and inter-groups of injury profiles. Considered in each group, DH and GLH have high clustering coefficient in the group 1. NB has high clustering coefficient in the group 2. RT has highest clustering coefficient comparing other injuries in the group 3. Group 2 has high clustering coefficients indicating that this group is closed clustered, and the injuries in this group formed co-occurrence. DP and SR are important as the linkage to occur with other groups of injuries profiles because of high betweenness and degree. WM, BPH have low value of the 3 local properties. It indicated that BPH and WM were less possible to occur, and present co-occurrence patterns, and when they were observed, they were also not able to relate to many injuries.

%=========================================

\paragraph{West Java, Indonesia (WJV)}

Dry season network (Fig.) composed of 20 nodes and 51 associations. The network reveals four groups of injury profiles. Group 1 (green color) include DH and RT. Group 2 (orange color) is NBS RB, FS, BLS, LB, DP, BLB, NB, BS. Group 3 (purple color) included GM, LF, BPH, SR, SHB, GLH, RS. Group 4 (pink color) include WM and WH. The second and third group are formed closely clusters. RT in group 1, BLB and BS in the group 2, LF of group 3, and WM in group 4 are high-betweenness nodes with intermediate clustering coefficient. NBS, NB, RB, BLS, FS, DP of group 2 have high clustering coefficients. Compared to group 2, the injuries in group 3 have relatively smaller than. It indicated that the injuries in the group 2 are tightly formed complex co-occurrence. DH, WH have low value of the three features, and located far from the center of the network. BPH and WM were less possible to occur, and present co-occurrence patterns, and when they were observed, they were also not able to relate to many injuries. connectivity.

Wet season network (Fig.) composed of 19 injuries (WM, LF, GLH, DH, BLS, WH, NBS, FS, DP, RS, LB, WPH, SR, SHR, SHB, NB, BS and BLB) with 25 associations. The network was loosely connected (low clustering coefficients). It reveals 3 connected groups and one isolated group of injury profiles. Group 1 (green) DH, WM, SHB, WH, GHL, LF, LB, and RS. Group 2 (orange) is composted of WPH, SR, GM, BLS and DP. Group 3 consisted of FS, NBS, SHB, BLB. Group 4, which is isolated, is composted of BS and NB. In group 1, WM is the injuries with high betweenness, and WH is the injury with high clustering coefficient. In group 2, BLS has high betweenness, and DP has high clustering coefficient. FS in group 3 node with high betweenness and high clustering coefficient. Group 1 appeared to form complex co-occurrence patterns because the average of clustering coefficient of injuries in this group are higher than other groups of injuries.

%=================================

\paragraph{Central Luzon, Philippines (LAG)}

The dry season network reveal three clustered groups of injury profiles. Group 1 composted of WM, RB, NB, SHB, and FS. WM and RB in this group have high rank of betweenness. Group 2 consist of LB and DP. DP in this group connected with BPH (high betweenness) from group 3, which have GLH, BPH and BLB. Interestingly, SHB in group1, and GLH in group 3 have high clustering coefficients. NB, LB, FS, and BLB featured low in three of centrality.

Figure revealed co-occurrence network of injury profiles in dry season at Laguna, the Philippines. Network resulted four groups of injury profiles, which there are three connected and one isolated. Group 1 (green) has SHR, RS, RT, which is isolated. Group 2 (orange) has NB, RB, GLH, BLS, and DP. This group has more more complex combination than others because the clustering coefficients of injuries in this group are higher. Group 3 (purple) has LB, SHB, LB. This group is between group1 and group 4 (pink), which has NBS, LF, BS, and BLB. The position of SHB, LB and WM in the group with high betweenness, but they do not form the the complex combination of the injuries. They tend to link the co-occurrence of the first and the fourth group of injury profiles, which connected to NB in the first group and LF in the fourth group. So they are potentially good target to be monitored too. For example, when LF presented without the present of LB, SHB, WM, it is less likely NB would present, and other injuries in the first group would less likely to present neither.

\paragraph{Suphan Buri, Thailand (SUP)}

Dry season network shows 8 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH) showing 12 significant relationships. The network revealed three closed cluster of co-occurrence patterns of injuries. Group 1 (green) is composted of NBS, DH, BS, and SHB. Group 2 (orange) is SHB and GLH, and group 3 (purple) is SR and WH. Group 1 is more complex than other two groups because three out of four nodes in the groups presenting high clustering coefficient. NBS, DH, SHB, and SR seem to be clustered together. BS present high betweenness in group 1, which is associated with the other two groups. SHB has high moderate degree and high betweenness, whereas it has low betweenness. It can form complex association with the injuries in group 1 and group 2. WH and GLH have low betweenness and node degree. Apparently, it is less easily found with other injuries, do not tend to complex combination (low clustering coefficient), or less possible for expressing co-occurrence through the network (low betweenness).

Network of co-occurrence patterns of injury profiles wet season at Suphan Buri, Thailand revealed 18 associated injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH), and 79 associations. Network analysis resulted three closely clustered groups. Group 1 (green) composted of NBS, FS, RT, SHB, LB, WM, RS and RB. Group 2 (orange) consist of GLH, BS, GM, SR, BLB, BLS, LF and DP. Group 3 consisted of SHR and NB. The group 1 are bigger and 3 were relatively high clustering coefficients than other two groups of injuries. RB, NBS, LF, and BS were highly associated in wet season. Even though they were not in the same group, but group 1 and 2 were close, which is indicated by the positions in the network. Interestingly, SHR in group 3 has high betweenness, which is in-between the association of injuries of group 1 and group 3. So SHB is more likely to present in wet season, and form complex association with other injuries because it connected to the high clustering coefficient groups.

\newpage

\paragraph{Makong Rive Delta, Vietnam (MKD)}

Co-occurrence network of injury profiles of dry season in Mekong Rive Delta, Vietnam presented 20 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH) with 61 associations. The network reveals the three groups of injury profiles. The group 1 (green) consisted of LB, DP, RB, DH, BS, and FS. The group 2 (orange) was composed of GLH, WPH, SR, WH, BPH, RS, WM. And the group 3 (purple) consisted of RT, NB, SHB, BLB, LF, and NBS. RB in the group 1 has high rank of betweenness and clustering coefficients. The betweenness of injuries in group 2 ranged low to intermediate, but their clustering coefficients are relatively high. This indicated that they are more likely to form complex association within group than between groups. BLB in group 3 and BS in group 1 have high rank of betweenness and they were associated. The average clustering coefficient of group 1 and 3 are smaller than the group 3. So group 1 and 3 are more likely to have chance to form association between the groups.

Co-occurrence network of injury profiles in wet season at Mekong River delta, Vietnam presented 21 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH), and 54 associations. From the structure of this network (Fig), it seems to have three clustered groups base on optimal clustering algorithm. The group 1 (green) is composed of NB, SR, BS, WPH, RS, WM, DH, GLH, FS, RT. The second group (orange) has BLB, BPH, LB, NBS, SHB, SHR, DP. The third group is the smallest groups, which has RB, LF and WH. The members within the first groups are relatively close following the layout, and have similar level of clustering coefficients. SHB and RB, RT, LB have high node degree and betweenness, which are inferred that they have possibility to occur in wet season because they shared many co-occurrence patterns with others. Even though, SHB have high level of betweenness and node degree, but intermediate clustering coefficient. It connected to the high-betweenness injuries such as RB and RT, where are in different groups. SHB also acted like a ``bridge'', which link the injuries of the group 1 and group 2.%\input{Network-analysis-/table\_survey\_data.tex}

%\input{Network-analysis-/table\_rice\_production\_season.tex}

\begin{figure}[p!]

\centerline{

\includegraphics[width = 1\textwidth]{figures/idn\_ds\_net.png}

}

\caption{A picture of the same gull looking the other way!}

\end{figure}

\afterpage{\clearpage}

\begin{figure}[p!]

\centering

\includegraphics[width = 1\textwidth]{figures/nodepropidn\_ds.pdf}

\end{figure}

%=== Here is the temp files

\subsection{Results}

\subsubsection{Crop health survey data}

Prior to using the crop health survey data to construct networks, some exploratory plots of the data are presented and examined. The figure shows the distribution of the data. A total of 415 farmers' fields was surveyed in Tamil Nadu, India (TMN); West Java; Indonesia (WJV), Laguna, Philippines (LAG); Suphanburi, Thailand (SPB) and Mekong river delta, Vietnam (MKD) from 2009 to 2013. The data collected pertain to rice injuries caused by animal pests, and pathogens. The protocol gives more emphasis on the nature of injuries and not on the causal organism. The levels of injuries during a cropping season were computed or summarized according to the nature of injuries and thus varied in scale. For example, foliar injuries, diseases on leaves, and insect pests were summarized as area under the disease incidence progress curve, and tiller injuries and diseases as a maximum incidence.

The injuries caused by animal pests observed during the survey period were rat injury (RT), deadheart (DH) and whitehead (WH) caused by stem borers, whorl maggot injury (WM), leaffolder injury (LF), gall midge injury or silver shoot (GM). Rat injuries were observed at all 5 survey locations with low incidence ( less than 20\% incidence). We could observe 75\% incidence of the rat injury in MKD in dry season. They were also observed at WJV in dry season, and both season in TMN, LAG, SPB. Gall midge injuries during survey period were not observed in TMN and LAG, but it was found in SPB, MKD, but WJV at 25 \% incidence. Deadheart were observed all survey sites, and it was severe in dry season at SPB and MKD, but in WJV and LAG, it was severe in dry season. The trend of whitehead incidences observed was opposite the deadheart incidence, which whitehead incidences were more server in wet season at SPB and MKD, but less severe at WJV and LAG. Leaffolder injury was observed all survey locations. The leaffolder incidences were more severe in wet season the dry season at WJV, TMN, and LAG. As apposite to SPB and MKD, they were more severe in dry season than wet season. Whorl maggot injury was observed at all locations. Mostly, they were more severe in wet season than dry season at all surveyed locations, except LAG.

The disease we 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). Rice diseases we observed in this study were commonly found at all locations, but there were some diseases that could not find especially in TMN such as BLS, BS, NBS, DP, RS, SHR, and SR. BLB was observed all location. BLB incidence was higher in dry season than wet season in WJV and MKD except in TMN, LAG, SPB. Wet season was more favorable for BLS than dry season because the incidence was higher in wet season than dry season. As same as BLS, BS incidence was higher in wet season than dry season, and it was severe in SPB. Even though, LB is common disease in these survey locations, but there were some farmer’s fields observed LB. In WJV and MKD, the LB incidence was higher in dry than wet season season. There were many field in SPB found high level of NBS incidence. Like BLS and BS, NBS incidence was more severe in wet season than dry season. There were many fields found high RS incidences. The highest incidences of DP were found in MKD. FS were commonly found at all location. The high incidents were observed in SPB, and MKD. NB were observed at all location. They were found many observations in TMN. SHB commonly was found all location, and high incidence was in TMN, and LAG.

Brown planthoppers (BPH) were found all location of survey sites. During the survey period, they were found higher population in wet season than dry season in TMN and MKD, but there was some observation of BPH at other locations. White backed planthoppers (WPH) were commonly found at MKD in both of dry and wet season, and there are a few observations in WJV and LAG, but at TMN and SPB, they were not observed. Rice bug (RB) could be observed all survey location. They were highly found in dry season than wet season at WJV, LAG, and MKD. but in SPB, they were found only in wet season during the survey period. Green leafhoppers were observed all location. They were found in LAG higher than other locations.

A global network of microbial co-occurrence and mutual exclusion within and among body site niches of the human microbiome

%Global properties of the microbiome-wide network of microbial associations are summarized in Figures 2 and 3. A dominant characteristic of the network was its habitat-specific modularity. After grouping the 18 body sites into five broad areas (oral, skin, nasal, urogenital, and gut), the large majority of edges were found clustered within body areas (98.54%), and these clusters were sparsely connected through a minority of edges (1.46%). This is confirmed by the network's high modularity coefficient of 0.28 (as defined by [32]) and Markov clustering of the network (see Methods and Figure S2). It has long been observed that sites within the human microbiome are distinct in terms of microbial composition [33], and this proved to be true of microbial interactions as well: microbial relationships within each body area's community were largely unique (Table 2). The microstructure of interaction patterns - and thus in the underlying ecology - was different for different areas, however. For example, all vaginal sites within the urogenital area were interrelated in a single homogeneous community, whereas interactions within the oral cavity suggested microbial cross-talk among three distinct habitats [34]. This can be observed quantitatively based on the proportions of microbial interactions spanning body sites within each area, e.g. 69.57% among the vaginal sites and 53.19% among the oral sites, both exceeding the microbiome-wide baseline. The skin was further unique in that the large amount (57.65%) of its associations related microbes in corresponding left and right body sites (left and right antecubital fossae and retroauricular creases), reflecting consistent maintenance of bilateral symmetry in the skin microbiome.

\subsubsection\*{Co occurrence of injury profiles based on network analysis}

Co-occurrence correlations of injuries were explored using network inference based on significant correlations using non-parametric Spearman’s rank coefficient at $P$ < 0.05. The setup of Spearman's coefficient cutoff at significant correlation coefficient at at $P$ < 0.05 could efficiency reduce the sparse correlation and highlight the significant correlation between variables. The networks were determined for the co-occurrence analysis of rice injuries based on the survey data. Ten networks constructed were based on the survey data grouping different location (west java; Indonesia (WJA), Tamil Nadu; India (TMN), Laguna; Philippines (LAG), Suphan Buri; Thailand (SPB), and Mekong Rive Delta; Vietnam (MKD)) and seasons (dry, and wet season). Nodes represent rice injuries, and edges represent Spearman's correlation coefficients at $P$-value < 0.05. Once a network has been constructed, analytic tools and measures can be used to quantify by determining the structural properties.

\includegraphics{figures/simgraph.png}

\begin{table}

\centering

\caption{Node topology}

\begin{tabular}{lllll}

Node & Node degree & Clustering coefficient & Betweenness \\

A & 2 & 0 & 0 \\

B & 6 & 0 & 14 \\

C & 6 & 0.06 & 12 \\

D & 4 & 0.17 & 0 \\

E & 4 & 0.17 & 0 \\

F & 2 & 0 & 0

\end{tabular}

\end{table}

Measures of node properties of networks seek to identify the most important nodes in a network. Different measures indices the different contexts for the word ``importance''. In this study, we consider 3 features of node properties.

\begin{itemize}

\item \textbf{Node degree} defined as the number of connections of a node. Nodes with high degree can be implied that the node has many connections or relationships with other nodes. In co-occurrence network, the high degree node can be inferred that it was found together with occurs with connected nodes. For example, in fig(), node B, and node C are the highest-degree nodes in this network. Node B have the relationships with 3 nodes. We can consider node B as the important node of this network because If we remove this node, there will be two nodes (node A and F) would disappear from this network.

\item \textbf{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. In the context of this network, local clustering coefficients is the the probability that two neighbors of connected with each other in the network. These measures are indicative of the complex forming of co-occurrence patterns through the network. As presented nodes can be observed other nodes, a more closely connected network facilitates nodes co-occurrence. The higher clustering coefficient injury has high possibility to form complex. Node D and E have high clustering coefficient. They formed the complex relationships, which Node D connected with Node E connected to Node C, and Node C also connected with Node D. It indicated that they have close relationships. In the context of co-occurrence network, these three nodes usually occur together.

\item \textbf{Betweenness} is measure the number of paths, which a node is present on the shortest path between all other nodes. Nodes with high betweenness have been shown that there are many pair-wise relationships across a network pass through the node, which is implied that the nodes has high possibility to occur. We determine high-betweenness injuries as``indicator''.

\end{itemize}

Moreover, I inspected the community structure of the networks based on the optimal clustering algorithm. Such communities correspond to groups of injuries that are closely associated among themselves than other nodes in one others groups.

\subsubsection{Communities, Structures and compositions of co-occurrence network of rice pest injuries}

\paragraph{Tamil Nadu, India (TMN)}

Dry season network (Fig) was composed of 6 associated injuries (DH, NB, SR, FS, RB, and LF) and captured 7 associations. The network showed two groups of injury syndromes (the combination of injuries). The groups of the injury profiles corresponded to each community based on the optimal clustering algorithm. The first group is composed of DH, LF, RB and SR. Another group consists of FS and NB. Analysis of network properties revealed that LF and FS are high-betweenness nodes. They presented co-occurrence relationships within group 1 and group 2. SR and RB have high clustering coefficient. It indicated that these two injuries usually formed complex of co-occurrence relationships. As opposed to other injuries, NB and DH have low scores on all three centrality measure. Apparently, these injuries less possibly co-occur with other injuries (low betweenness), and do not have complex co-occurrences with other injuries (low degree and clustering coefficient).

Wet season network (Fig) was composed of 12 nodes (injury variables, DP SR, FS, BLB, NB, LF, GLH, RT, RB, DH, WM, BPH) with edges. Fig reveals three groups of injury profiles. DH, GLH, SR and BPH are in the group 1 (green). FS, NB, DP, BLB is in group 2 (orange). RT, LF, RB and WM are in group 3 (purple). Top four injuries with high betweenness, DP, SR, LF, and BLB, are the members of each of group. They possibly are found co-occurrence within the group and inter-groups of injury profiles. Considered in each group, DH and GLH have high clustering coefficient in the group 1. NB has high clustering coefficient in the group 2. RT has highest clustering coefficient comparing other injuries in the group 3. Group 2 has high clustering coefficients indicating that this group is closed clustered, and the injuries in this group formed co-occurrence. DP and SR are important as the linkage to occur with other groups of injuries profiles because of high betweenness and degree. WM, BPH have low value of the 3 local properties. It indicated that BPH and WM were less possible to occur, and present co-occurrence patterns, and when they were observed, they were also not able to relate to many injuries.

%=========================================

\paragraph{West Java, Indonesia (WJV)}

Dry season network (Fig.) composed of 20 nodes and 51 associations. The network reveals four groups of injury profiles. Group 1 (green color) include DH and RT. Group 2 (orange color) is NBS RB, FS, BLS, LB, DP, BLB, NB, BS. Group 3 (purple color) included GM, LF, BPH, SR, SHB, GLH, RS. Group 4 (pink color) include WM and WH. The second and third group are formed closely clusters. RT in group 1, BLB and BS in the group 2, LF of group 3, and WM in group 4 are high-betweenness nodes with intermediate clustering coefficient. NBS, NB, RB, BLS, FS, DP of group 2 have high clustering coefficients. Compared to group 2, the injuries in group 3 have relatively smaller than. It indicated that the injuries in the group 2 are tightly formed complex co-occurrence. DH, WH have low value of the three features, and located far from the center of the network. BPH and WM were less possible to occur, and present co-occurrence patterns, and when they were observed, they were also not able to relate to many injuries. connectivity.

Wet season network (Fig.) composed of 19 injuries (WM, LF, GLH, DH, BLS, WH, NBS, FS, DP, RS, LB, WPH, SR, SHR, SHB, NB, BS and BLB) with 25 associations. The network was loosely connected (low clustering coefficients). It reveals 3 connected groups and one isolated group of injury profiles. Group 1 (green) DH, WM, SHB, WH, GHL, LF, LB, and RS. Group 2 (orange) is composted of WPH, SR, GM, BLS and DP. Group 3 consisted of FS, NBS, SHB, BLB. Group 4, which is isolated, is composed of BS and NB. In group 1, WM is the injuries with high betweenness, and WH is the injury with high clustering coefficient. In group 2, BLS has high betweenness, and DP has high clustering coefficient. FS in group 3 node with high betweenness and high clustering coefficient. Group 1 appeared to form complex co-occurrence patterns because the average of clustering coefficient of injuries in this group are higher than other groups of injuries.

%=================================

\paragraph{Central Luzon, Philippines (LAG)}

The dry season network reveal three clustered groups of injury profiles. Group 1 composted of WM, RB, NB, SHB, and FS. WM and RB in this group have high rank of betweenness. Group 2 consist of LB and DP. DP in this group connected with BPH (high betweenness) from group 3, which have GLH, BPH and BLB. Interestingly, SHB in group1, and GLH in group 3 have high clustering coefficients. NB, LB, FS, and BLB featured low in three of centrality.

Figure revealed co-occurrence network of injury profiles in dry season at Laguna, the Philippines. Network resulted four groups of injury profiles, which there are three connected and one isolated. Group 1 (green) has SHR, RS, RT, which is isolated. Group 2 (orange) has NB, RB, GLH, BLS, and DP. This group has more more complex combination than others because the clustering coefficients of injuries in this group are higher. Group 3 (purple) has LB, SHB, LB. This group is between group1 and group 4 (pink), which has NBS, LF, BS, and BLB. The position of SHB, LB and WM in the group with high betweenness, but they do not form the the complex combination of the injuries. They tend to link the co-occurrence of the first and the fourth group of injury profiles, which connected to NB in the first group and LF in the fourth group. So they are potentially good target to be monitored too. For example, when LF presented without the present of LB, SHB, WM, it is less likely NB would present, and other injuries in the first group would less likely to present neither.

\paragraph{Suphan Buri, Thailand (SUP)}

Dry season network shows 8 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH) showing 12 significant relationships. The network revealed three closed cluster of co-occurrence patterns of injuries. Group 1 (green) is composed of NBS, DH, BS, and SHB. Group 2 (orange) is SHB and GLH, and group 3 (purple) is SR and WH. Group 1 is more complex than other two groups because three out of four nodes in the groups presenting high clustering coefficient. NBS, DH, SHB, and SR seem to be clustered together. BS present high betweenness in group 1, which is associated with the other two groups. SHB has high moderate degree and high betweenness, whereas it has low betweenness. It can form complex association with the injuries in group 1 and group 2. WH and GLH have low betweenness and node degree. Apparently, it is less easily found with other injuries, do not tend to complex combination (low clustering coefficient), or less possible for expressing co-occurrence through the network (low betweenness).

Network of co-occurrence patterns of injury profiles wet season at Suphan Buri, Thailand revealed 18 associated injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH), and 79 associations. Network analysis resulted three closely clustered groups. Group 1 (green) composted of NBS, FS, RT, SHB, LB, WM, RS and RB. Group 2 (orange) consist of GLH, BS, GM, SR, BLB, BLS, LF and DP. Group 3 consisted of SHR and NB. The group 1 are bigger and 3 were relatively high clustering coefficients than other two groups of injuries. RB, NBS, LF, and BS were highly associated in wet season. Even though they were not in the same group, but group 1 and 2 were close, which is indicated by the positions in the network. Interestingly, SHR in group 3 has high betweenness, which is in-between the association of injuries of group 1 and group 3. So SHB is more likely to present in wet season, and form complex association with other injuries because it connected to the high clustering coefficient groups.

\newpage

\paragraph{Makong River Delta, Vietnam (MKD)}

Co-occurrence network of injury profiles of dry season in Mekong River Delta, Vietnam presented 20 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH) with 61 associations. The network reveals the three groups of injury profiles. The group 1 (green) consisted of LB, DP, RB, DH, BS, and FS. The group 2 (orange) was composed of GLH, WPH, SR, WH, BPH, RS, WM. And the group 3 (purple) consisted of RT, NB, SHB, BLB, LF, and NBS. RB in the group 1 has high rank of betweenness and clustering coefficients. The betweenness of injuries in group 2 ranged low to intermediate, but their clustering coefficients are relatively high. This indicated that they are more likely to form complex association within group than between groups. BLB in group 3 and BS in group 1 have high rank of betweenness and they were associated. The average clustering coefficient of group 1 and 3 are smaller than the group 3. So group 1 and 3 are more likely to have chance to form association between the groups.

Co-occurrence network of injury profiles in wet season at Mekong River delta, Vietnam presented 21 injuries (DH, BS, SR, SHB, NBS, SHR, GLH, and WH), and 54 associations. From the structure of this network (Fig), it seems to have three clustered groups base on optimal clustering algorithm. The group 1 (green) is composed of NB, SR, BS, WPH, RS, WM, DH, GLH, FS, RT. The second group (orange) has BLB, BPH, LB, NBS, SHB, SHR, DP. The third group is the smallest groups, which has RB, LF and WH. The members within the first groups are relatively close following the layout, and have similar level of clustering coefficients. SHB and RB, RT, LB have high node degree and betweenness, which are inferred that they have possibility to occur in wet season because they shared many co-occurrence patterns with others. Even though, SHB have high level of betweenness and node degree, but intermediate clustering coefficient. It connected to the high-betweenness injuries such as RB and RT, where are in different groups. SHB also acted like a ``bridge'', which link the injuries of the group 1 and group 2.%\input{Network-analysis-/table\_survey\_data.tex}

%\input{Network-analysis-/table\_rice\_production\_season.tex}

\begin{figure}[p!]

\centerline{

\includegraphics[width = 1\textwidth]{figures/idn\_ds\_net.png}

}

\caption{A picture of the same gull looking the other way!}

\end{figure}

\afterpage{\clearpage}

\begin{figure}[p!]

\centering

\includegraphics[width = 1\textwidth]{figures/nodepropidn\_ds.pdf}

\end{figure}

%=== Here is the temp files

\subsection{Discussion}

Rice injuries were found commonly in South and South east Asia, but at different levels of incidence.

Some injuries of this study is relatively low prevalence of areas that have been reported such as leaf blast, brown spot. It could be implied that these injuries strongly depended on locations or climatic conduction to develop, so they were not observed at all locations or seasons during survey were conducted. Another reason is that there are some factors such as the utility of resistance verities in the farmer’s fields surveyed, so we could observe those injuries at low level of incidence. The similar reasoning could explain to many injuries such as BLB, RB, NB, which widely occur in rice growing areas.

From the survey data of rice injuries observed in farmers’ fields, I analyzed the interaction and build the network based on that data. The methods applied for building a co-occurrence network of rice injuries were adapted from ecological studies. Usually, relationships 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 ecological studies.

The exploration of co-occurrence networks is a useful method for determining interactions of co- occurring injuries. Network analysis has also suggested important injuries in networks. The important injuries were selected from the node features such as node degree, clustering coefficient, and betweenness. The betweenness represents the importance of the control potential that an injury exerts over the associations of other injuries in the network. The clustering coefficients are indicative of the potential spreading of the incidence of injuries through the network. As activated injury can activate other injuries, a more densely connected network facilitates injury activation \cite{Williams\_2014\_demonstrating}. In the network of dry season in West Java, Indonesia, BLB and SR can be the targets to be monitored because they have high betweenness, which indicated that they are more likely to present then others injuries.

It is good attempt to detect communities in a network because it can reveal information about the networks that is maybe not easy to detect by simple observation. The communities are groups of nodes that are densely connected among their node members, and slightly connected with the rest of the network. In this study, we detected node community based on the optimization of the modularity of a sub-network, which is an approach that widely applied in many fields \cite{Liu\_2014\_Detecting}. Even though, the groups of injury profiles from this study are different from seasons and countries, but they are reasonable because some rice injuries present in some seasons (seasonal occurrence) such as gall midge injury \cite{Krishnaiah\_2004\_Rice}. The result also showed similar groups of injuries to the patterns of injuries profile from the study of \cite{Savary\_2000\_Characterization}.