

Network structure and risk-based surveillance algorithms for live shrimp movements in Thailand.

Short title: Live shrimp movements in Thailand

Saleetid, Nattakan ¹ & Green, Darren Michael ²

¹ **Department of Fisheries, Kasetsart University Campus, Chatuchak, Bangkok, 10900 Thailand.**

² **Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK.**

Correspondence

Darren Green, Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK.

Email: darren.green@stir.ac.uk

Abstract

Live shrimp movements pose a potential route for site-to-site transmission of acute hepatopancreatic necrosis disease (AHPND) and other shrimp diseases. We present the first application of network theory to study shrimp epizootiology, providing quantitative information about the live shrimp movement network of Thailand (LSMN), and supporting practical and policy implementations of disease surveillance and control measures. We examined the LSMN over a 13-month period from March 2013 to March 2014, with data obtained from the Thailand Department of Fisheries. The LSMN had a mixture of characteristics both limiting and facilitating disease spread. Importantly, the LSMN exhibited power-law distributions of *in* and *out* degrees with exponents of 2.87 and 2.17, respectively. This characteristic indicates that the LSMN behaves like a scale-free network and suggests that an effective strategy to control disease spread in the Thai shrimp farming sector can be achieved by removing a small number of targeted inter-site connections (arcs between nodes). Specifically, a disease-control algorithm based on betweenness centrality (defined as the number of shortest paths between node pairs that traverse a given arc) is proposed here to prioritise targets for disease surveillance and control measures.

Keywords: AHPND; graph theory; risk-based surveillance

1 Introduction

Shrimp farming has helped drive the socio-economic growth of Thailand since the 1980s (Szuster, 2006). Nevertheless, invasion of infectious diseases presents a major barrier to success in the sector. Looking at the last 30 years of production, microorganisms, i.e. viruses and bacteria, are a major cause of Asian farmed shrimp deaths, and represented a substantial economic loss over US\$ 15 billion during the period 1990-2005 (Flegel, 2012). To minimise these losses, the World Organisation for Animal Health (OIE) requires all member countries to develop disease surveillance and control measures for aquatic animals.

For shrimp, long-distance transmission, largely by movements of live shrimp, poses a high risk for site-to-site disease spread. A recent example, acute hepatopancreatic necrosis disease (AHPND, earlier known as EMS) has hit Thai shrimp farming since late 2011, and has been transmitted across Thai regions via the movements of infectious live shrimp (FAO, 2013; OIE, 2013). This epizootic disease reduced Thai shrimp production by an estimated 500 000 tonnes of shrimp during the 3-year period from 2011 to 2014 (Songsanjinda, 2015). Although there have been many worldwide efforts to stop the spread of AHPND, such as movement restrictions, biofloc technology, genetic improvements and enhanced breeding techniques (Hong *et al.*, 2016; Pakingking *et al.*, 2016), risk-based surveillance and control approaches are needed.

In recent years, contact network analysis has played a growing role in epidemiology as increasingly rich data sets become available. At large spatial scales, epidemiological networks typically consist of a set of *nodes* (sites) connected by directionless *edges* or directed *arcs*. Network approaches can mathematically study disease dynamics and persistence (Kurvers *et al.*, 2014; Meyers *et al.*, 2003), as well as the impact of potential control measures. Centrality measures can be used to identify high-risk network features whose removal from the network minimises potential disease spread. As suggested by Woolhouse *et al.* (1997), for heterogeneous networks, it may be that a small

fraction of infectious nodes correspond to most of the transmission. Thus, surveillance and control targeted at the core group of nodes can be the most effective strategy (Woolhouse *et al.*, 2005).

Appropriate algorithms for identifying influential nodes or contacts appear to be network specific (Newman and Park, 2003). Degree centrality is one means to demonstrate the most important nodes in networks (Bohm *et al.*, 2009; Christley *et al.*, 2005; Zhang *et al.*, 2010). Algorithms based on betweenness centrality were the most efficient way to reduce potential epizootic size for the network of cattle movements in France (Rautureau *et al.*, 2011), as measured through component structure; and for the network of pig movements in Germany (Lentz *et al.*, 2016). Betweenness centrality was also effective in reducing potential epizootic size in models of the live salmonid movements in Scotland (Green *et al.*, 2009). For comparison, other measures of centrality include closeness centrality (Fournie *et al.*, 2013) and eigenvector centrality (Herrera *et al.*, 2016), but which centrality measures will be effective cannot be assumed without studying the specific system of interest and its structure.

In this paper, we perform the first analysis of the structure of the live shrimp movement network (LSMN) for Thailand, with several aims. First, we explore its connectedness and how that may relate to potential disease transmission; second, we model optimal disease-control algorithms for the Thai shrimp farming system at large spatial scales. This is possible following on from the aquatic animal trade regulation of Thailand, B.E.2553 (2010) that requires farmers to report most farmed-shrimp movements to the authorities. The outcome is quantitative information concerning potential epizootic dynamics via live shrimp movements, which can form part of the process of implementing mitigation measures for management areas, either in real time during an epizootic, or in peacetime for developing disease surveillance and control programmes.

2 Material and Methods

2.1 Data sources

The live shrimp movement network (LSMN) data—a Thai government electronic database—was provided by the Thailand Department of Fisheries. The sample contained some 99 000 records collected from daily batch movements consisting of: (1) the farm registration number of source and destination sites, which also indicates province and status as an ongrowing or seed (postlarvae for ongrowing) producing site (2) details on site location i.e. house number, sub-district and district, (3) the date of movement and (4) the seed quantity. We omitted any records containing missing data in one or more fields. Multiple records between a pair of sites on a single day were combined. The data included the 13-month period from March 2013 to March 2014 during which AHPND continued to spread around the country (NACA, 2017; Putth & Polchana, 2016), and covered up to three production cycles of farmed shrimp (Flaherty *et al.*, 2000). This is the first time series of shrimp movements were readily available for Thailand, due to the advent of computerised recording and centralised data collection. We discuss the likely completeness of the data set below.

2.2 Characterisation of the LSMN

We used a selection of network metrics to quantify and describe the LSMN, as well as visualise it at provincial level. This allows us to identify network features that limit or facilitate disease spread. Inter- and intra- provincial movements for 37 provinces were visualised using Pajek Software (Mrvar and Batagelj, 1996).

Quantitative analysis of the LSMN structure was performed in the *R* environment (R Core Team, 2018), where a number of helpful packages are available, detailed below. We performed both a qualitative and quantitative analysis of the network by representing the LSMN by both weighted (w) and unweighted (a) adjacency matrices. In the absence of knowledge of how batch size of shrimp might relate to disease risk, we chose to weight by frequency of contact, in terms of

numbers of days with movement events. A weight w_{ij} implies the frequency of shrimp batches moved from March 2013 to March 2014 as directed arcs between a node pair (i, j) . Element a_{ij} took the value 1 if a directed arc existed from site i to site j and the value 0 otherwise. About 300 self-loop arcs ($a_{ii} = 1$) were removed from the analysis because these self-loops were assumed not to contribute to further spread of diseases inter-farm (Britton *et al.*, 2011; Draief *et al.*, 2008), and likely represent intra-site operations.

A node's degree is its number of contacts, the most basic indicator of network structure and infection risk posed. Nodes with high degree will be more at risk of becoming infected or causing infection, depending on the direction of the arcs. Examining the local network structure around a node, *in* and *out* degrees (k_i^{in} and k_i^{out} for node i), and undirected degrees (k_i^{und}) were calculated as in in Barrat *et al.* (2004) and Green *et al.* (2009). All summation limits are taken as $1 \dots N$ where N is the total number of nodes.

$$k_i^{und} = k_i^{in} + k_i^{out} - \sum_j a_{ji}a_{ij}$$

Additionally, weighted degree measures were calculated for the quantitative network analysis, which requires a slightly different formulation for undirected degree.

$$k_i^{und} = k_i^{in} + k_i^{out}$$

As higher variance in degree distribution reduces the threshold for disease spread (Pastor-Satorras and Vespignani, 2002a), the degree distributions of the LSMN (the weighted network) were statistically analysed for power-law characteristics and exponents were estimated, using the Kolmogorov–Smirnov test with the *igraph* software package (Csardi and Nepusz, 2006). Its implementation uses the method of Clauset *et al.* (2009) and Newman (2005); the null hypothesis is that the LSMN is generated from a power-law distribution. Other basic statistics of the degrees, i.e. mean, maximum and minimum, and coefficient of variation, were also derived.

Short path lengths (number of arcs between pairs of nodes across the network)—as found in "small world"-type networks—imply the potential for fast disease spread through the network (Moore and Newman, 2000). Therefore, we examined the connectivity of the LSMN to determine the length and existence of these paths. The path-length matrix for non-weighted directed networks was calculated by using Dijkstra's algorithm (Csardi and Nepusz, 2006; Dijkstra, 1959). Weighted directed networks require more care, however. For these, the path length matrix was computed by using the *tnet* package from Opsahl (2009): first, the matrix L_{ij} was calculated by using Dijkstra's algorithm, then divided by the mean arc weight of the network ($\sum_{ij} w_{ij} / \sum_{ij} a_{ij}$). Weighted and unweighted path length would therefore be equivalent where all weights are equal. The mean shortest path length $\langle L \rangle$ was computed following Mao and Zhang (2013) :

$$\langle L \rangle = \sum_{i \neq j} L_{ij} / \sum_{i \neq j} [L_{ij} \neq \infty]$$

Here, we use Iverson bracket notation to show the average is taken only over those paths (i, j) where a route is possible (defined as non-infinite) within the network.

Clustering refers to the presence of non-random mixing with preferential connection amongst subgroups (Shirley and Rushton, 2005). In an epizootiological context, high clustering tends to lower the spread of disease across networks as a whole due to sharing of contacts amongst pairs of connected nodes (Keeling, 2005). The widely used global clustering coefficient C was calculated as a ratio of the number of triangles to the number of triples. Ahnert and Fink (2008) defined a triangle as a set of three nodes with arcs $\{i \rightarrow j, j \rightarrow k, i \rightarrow k\}$, meaning that both direct and indirect routes for $i \rightarrow k$ exist, whereas a triple is merely three connected nodes $\{i \rightarrow j, j \rightarrow k\}$.

Assortativity measures the degree to which nodes contact like nodes, usually based on degree. Increased assortativity leads to increased likelihood of epidemic spread (Kiss *et al.*, 2008). We computed the assortativity coefficient (r) to represent the assortative mixing by degree in the LSMN. Foster *et al.* (2010) state that the assortativity of directed networks can be represented by

four distinct measures: $r(out, in)$, $r(in, out)$, $r(out, out)$ and $r(in, in)$. Among these four measures, however, the most interesting for epizootiological studies is probably $r(in, out)$ —pertaining to directed arcs joining nodes with a high *in* degree to nodes with high *out* degree. This is because this motif produces arcs that can easily propagate infection from high-risk node to high-risk node. Its equation is shown in Green *et al.* (2009).

Heesterbeek and Dietz (1996) define the basic reproduction number R_0 as the expected number of secondary cases generated by a typical case over the whole infectious period in a particular group of susceptible individuals. The disease percolation threshold is in general $R_0 > 1$ (Heesterbeek and Dietz, 1996). In order to estimate an upper limit for R_0 (in the absence of disease-specific transmission rates, and a single equation for R_0 that applies to all networks) based on network topology, one estimate is based on node degree (Green *et al.*, 2009):

$$R_0 \sim \frac{\langle k^{in} k^{out} \rangle}{\langle k^{out} \rangle}$$

Here, $\langle \cdot \rangle$ represents an average. In addition, the largest eigenvalue (λ) of the LSMN's adjacency matrix w was calculated, since this also relates closely to the epizootic transmission rate in the network (Becker and Hall, 1996; Chakrabarti *et al.*, 2008; Prakash *et al.*, 2010). With few closed cycles in the network, however, this measure could easily be zero, or highly non-representative of the whole network. Thus, a simple adjustment was made, following Green *et al.* (2009), i.e. adding a constant number ($K = 0.5$) to all arcs.

$$w_{ij}^{adjust} = w_{ij} + \frac{K}{N} \langle w \rangle$$

Then, the largest eigenvalue λ was computed with the *eigen_centrality* function in the *igraph* package (Csardi and Nepusz, 2006). This result was obtained by solving the equation $w^{adjust} = V\lambda$, where V is the corresponding eigenvector (Restrepo *et al.*, 2007).

To evaluate the significance of the clustering coefficient and the mean shortest path length, the network was compared with randomly rewired networks, which preserved the number of nodes and degree distributions of the original (Evans, 2007; Kiss and Green, 2008; Maslov and Sneppen, 2002; Noldus and Van Mieghem, 2013). One thousand rewired networks were developed using the *igraph* software package (Csardi and Nepusz, 2006), where in such rewired networks the probability of rewiring was set at one, resulting in all arcs in the LSMN being swapped. For example, two arcs $i \rightarrow j$ and $u \rightarrow v$ are replaced by the simulated arcs $i \rightarrow v$ and $u \rightarrow j$. A non-parametric two-tailed test was used to compare $\langle L \rangle$ in the original network with the distribution of rewired networks. The null hypothesis was rejected when $\langle L \rangle$ lay below the 2.5th percentile, or above the 97.5th percentile for the same parameter in 1000 rewired networks.

2.3 Disease-control algorithms for risk-based disease surveillance and control

To find optimal disease control targets for preventing disease spread in the LSMN, disease-control algorithms were developed in *R* to locate high-risk arcs $i \rightarrow j$, whose removal from the network reduced the potential transmission of disease.

Risk-based surveillance can be visualised as focussed on those nodes or arcs that are most likely to pose a risk of disease transmission, either as sources (outbound arcs only) or sinks (inward arcs only). Thus, here, we focus on arc removal rather than node removal, representing a scenario where prevention or control measures are focused at the level of shipment, rather than all shipments from a particular site. A static network representation is used here, so the smallest possible removal from the network is a whole arc. The algorithms used in this research were developed from those described by Green *et al.* (2012), which described four risk-based approaches based on betweenness centrality, arc weight, eigenvector centrality, and subnet crossing to identify targets for removal. As a control, a random non-targeted approach was used. For efficient disease control, relatively few removals should markedly reduce susceptibility of the network to epizootic spread.

The algorithms each begin with the whole network. Then, the arcs $i \rightarrow j$ in the network are ranked according to one of the following criteria. Where two or more arcs have equal rank, their order of removal is randomised for each repeat of the algorithm.

Betweenness Arc betweenness was calculated for each arc in the network. It was defined by Girvan and Newman (2002) as the number of shortest paths between all possible pairs of nodes that utilise an arc (using the unweighted network). These arcs were identified using the *edge-betweenness* function in the R *igraph* package (Brandes, 2001; Csardi and Nepusz, 2006).

Arc weight Nodes with the highest numbers of days with movement events can be considered more important for disease transmission (Christley *et al.*, 2005; Tanaka *et al.*, 2014). We chose arcs $i \rightarrow j$ with the highest value of w_{ij} for removal from the network.

Eigenvector centrality Eigenvector centrality can be a useful measure in disassortative networks (Bonacich, 2007). We can think of the eigenvector as a measure of the relative likelihood of infection of a node in the case of a rare disease propagating through the network. For the LSMN, the high-risk nodes, with the highest eigenvector centrality, were detected using an eigenvector criterion within the *evcent* function in the *igraph* package (Csardi and Nepusz, 2006). Then randomisation was used to choose outbound arcs for removal.

Subnet crossing We define a subnet here as a set of interconnected nodes that are relatively poorly connected to other such sets. Arcs joining such subnets are therefore of interest in maintaining connectivity of the network as a whole. The *fastgreedy.community* algorithm in the *igraph* package was used (Clauset *et al.*, 2004; Csardi and Nepusz, 2006), and arcs linking subnets found with the *crossing* function (Clauset *et al.*, 2004; Csardi and Nepusz, 2006).

Non-targeted (control) Arcs were chosen completely at random, representing the case of non risk-based control

To measure impact of arc removal, node reach R_i —the number of nodes reachable from a node i and therefore at risk of infection—was calculated according to Green *et al.*, (2012) using the path length matrix. As used by Green *et al.*, (2012), the maximum value of R_i across all nodes served as an estimate of the worst-case epizootic size, and the arithmetic mean of R_i across all nodes serves as an estimate of typical epizootic size in terms of the number of sites affected.

Each algorithm was repeated for 1 000 arc removals using the R environment. At each iteration of the algorithm, the highest-ranked arc is identified according to the criterion under consideration, and permanently removed from the network. As removing an arc can have a substantial impact on the network as a whole, rerouting shortest paths, etc., network properties were recalculated for each iteration—each new removal.

As this is computationally intensive, and there is no guarantee that such "greedy" algorithms (which make stepwise, locally optimal choices rather than identify a globally best solution, which here would involve evaluating a mind-boggling $P(N, 1000)$ possible combinations) work best with a step size of one, larger numbers of removals prior to recomputation of network properties were also explored, varying from 1 to 500 removals, e.g. 1 000 total removals in 100 batches of the 10 highest-ranking arcs each.

As the algorithms contain a stochastic element, we present averages of 1 000 replicates. Furthermore, the properties of removed arcs were investigated in terms of arc length. Straight-line distance between sources and destinations were estimated using the coordinates of the Thai sub-districts available from Google Earth (2015) according to Dubé *et al.* (2008).

3 Results

3.1 The live shrimp movement network at provincial level

Figure 1 displays the general characteristics of the LSMN. A total of 13 801 shrimp farming sites were located in 37 provinces of five regions, i.e. south (5 665 sites; 41 % of the total), east (4 874 sites; 35 %), central (1 949 sites; 14 %), west (1 312 sites; 9 %) and one site in the northeast. Overall, there were 33 720 site-to-site arcs (Figure 2).

Those arcs contained 74 462 repeated movements that included 57 281 arcs entailing inter-province movements (77 % of the total repeated arcs), peaking at 5 831 arcs in March 2014. The remaining movements were intra-province (23 %), peaking at 1 814 arcs in September 2013. Mean straight-line distances were 24 and 192 km, respectively, for intra- and inter-province movements. A total of 252 movement records were removed due to incompleteness.

The network is visualised in Figure 3 at provincial level. The most active inter-province arcs are Chonburi (CBI) → Chanthaburi (CTI), Chonburi (CBI) → Chachoengsao (CCO), Chachoengsao (CCO) → Chanthaburi (CTI), Trat (TRT) → Chanthaburi (CTI) and Chumphon (CPN) → Suratthani (SNI), whereas the most arcs intra-province occur in Chachoengsao (CCO), Nakhonsithammarat (NRT), Phuket (PKT), Songkhla (SKA) and Prachuapkhirikhan (PKN).

3.2 The node-level network

The contrast in descriptive statistics for the weighted and non-weighted node-level networks are shown in Table 1. For the non-weighted network, there was noticeably more variability in *out* degree (mean 2.4, coefficient of variation 9.2) than *in* degree (mean 2.4, c.v. 0.9) or undirected degree (mean 4.9, c.v. 4.6). This reflects the narrower range in *in* degree compared with *out* (zero to 30 versus 932). The Pearson correlation between the node degrees k^{in} and k^{out} was weakly positive with a value of 0.03 ($P < 0.01$). This weakly positive correlation indicated that nodes with

a high risk of becoming infected posed a low risk for onward transmission of the disease (assuming solely network spread) (Kiss *et al.*, 2006).

The same general pattern was visible in the weighted network albeit the node degrees are about twice as high, with more variability in *out* degree (mean 5.4, c.v. 13.4, max 5839) than *in* degree (mean 5.4, c.v. 1.7, max 232) and *undirected* (mean 10.8, c.v. 6.9). The significant in–out degree correlation remained and was stronger ($r = 0.24$; $P < 0.01$). This suggests that studying the non-weighted network would neglect this correlation, and underestimate potential epizootic spread.

The weighted network obeyed a power-law, scale-free degree distribution, $P(k) \sim k^{-\gamma}$ for both *in* and *out* degree distributions with exponents $\gamma = 2.87$ and 2.17 , respectively (Figure 4). Exponents in the range 2–3 are similar to many scale-free networks, as proposed in Goh *et al.* (2002), and, the Kolmogorov–Smirnov test accepted a power-law as a plausible model (P-values for $k^{in} = 0.7$ and $k^{out} = 0.54$), consistent with scale-free topology.

The upper-limit R_0 estimate from the weighted degree-based calculation was high (34.5), compared to the largest eigenvalue (16.2). This difference may reflect the bipartite structure of the network, with two types of nodes: seed-producing sites and ongrowing sites. To understand this better, Table 2 accounts for the arcs between the two node types in the weighted LSMN. The bipartite structure is also evident in the non-weighted LSMN (Table 3). The majority of arcs (>80%) join sites of the two types. The remaining arcs join sites within the same type of site, of which most of them terminate the arcs at seed-producing sites (14.5%), illustrating the nursery system for shrimp in Thailand.

The in–out degree correlation for the weighted network was $r = -0.09$. This low preference for high in-degree nodes to have outbound arcs to nodes with high out-degree is a feature that lowers transmission risk. A key property of small-world networks, short average path length, is shown by both the non-weighted and weighted LSMN. Focusing on the non-weighted network, the value of

the mean path length $\langle L \rangle$ was small (3.47), but with a small fraction (0.5 %) of potential total paths $N(N - 1)$ possible. For the weighted LSMN, $\langle L \rangle$ was equal to 2.99, with 0.14 % of potential total paths. The relatively small fraction of possible paths also reflects the hierarchical nature of the network with seed-producer and ongrowing sites, reducing potential for loops, as well as its nature as a sparse network.

The distribution of weighted path lengths is plotted in Figure 5, clearly demonstrating many short paths and few long paths. The same network features also gave rise to a very low clustering coefficient (weighted, $C = 0.1$; non-weighted, $C = 0.00051$) due to the hierarchical structure and paucity of triangles within the LSMN.

We explored the significance of these aspects of network structure using a comparison against partly randomised networks using a rewiring algorithm. For 1 000 randomly rewired networks with a rewiring probabilities of one (conserving the original LSMN's weighted degree distribution and the number of nodes N), the average C was 0.06 (s.d. 0.008), and the average of $\langle L \rangle$ was 3.84 (s.d. 2.12), with 0.5 % of potential paths existing. These values are close to those of the original network.

3.3 Risk-based removal in the live shrimp movement network

To evaluate effective disease surveillance strategies, the results of the five disease-control algorithms with unit step size are shown in Figure 6. The algorithms demonstrate different abilities to reduce maximum and mean reach in the LSMN. After removal of 1 000 arcs, the betweenness-based algorithm performed well for both network measures. The maximum and mean R_i were reduced by 50 % after 400 removals from the network. The other algorithms (arc weight-, eigenvector-, subnet crossing- and random-based) performed relatively poorly. The betweenness-based algorithm still performed well when a step size of more than one removal was applied. For example, after 400 removals from the network, a step size of 10 (one of removal tests), also shows the 50 % reduction in maximum and mean R_i .

The targeted arcs from the disease-control algorithms above are denoted as potential high-risk arcs for disease spread in the LSMN. Analysing these 1 000 target arcs from the betweenness algorithm, we found that the geographic distances of the targeted arcs were (on average) longer (mean 271 km), compared to the mean arc length in the whole arcs of the original LSMN (mean 200 km). In addition, Table 4 shows that, based on the betweenness algorithm, most targeted arcs join nodes of the same type, i.e. seed-producing sites.

4 Discussion

Network analysis of 13-months of live shrimp movement data for Thailand found epizootiological properties both facilitating and limiting the spread of diseases. Because the movements of live shrimp play a crucial role in infectious disease transmission from site to site, particularly in respect to the recent outbreak of AHPND and other known shrimp diseases, these results are a step towards designing an effective disease surveillance and control programme for Thai shrimp farming.

Overall, we found a hierarchical network structure with relatively few nodes as sources for many arcs, in common with certain other aqua- and agricultural industries. This structure naturally leads to short average path lengths and low clustering. Limiting transmission, the LSMN showed a weak correlation between node degrees and a low fraction of possible paths between nodes. Repeated connections between nodes are frequent because there are relatively few shrimp seed producers in Thailand. Facilitating transmission, the LSMN showed a small characteristic path length (L), particularly for the weighted network, and a low clustering coefficient. Shirley and Rushton (2005) note that a network with short average path lengths is likely to have a fast rate of infection during epizootics. Fast decision-making is therefore required to prevent disease epizootics in Thai shrimp farming. The low transitivity clustering coefficient (C) is a very local measure of network clustering. That L and C resembled those in the partly-rewired networks suggests a lack of structure; however not all clustering coefficients are necessarily that local, and our analysis provides evidence of clustering at larger spatial scales with local trading being common in the LSMN (i.e. within province). Furthermore, certain network characteristics were conserved during this rewiring.

The structure of the LSMN was found to be consistent with scale-free behaviour (exhibiting power-law *in*- and *out*-degree distributions) with exponents less than three. Pastor-Satorras and Vespignani (2002a) showed that the variances of power-law distributions with exponent $\gamma < 3$ are infinite, removing the threshold transmission rate for epizootic spread even at low transmission rate. To prevent disease spread in such power-law networks, therefore, control strategies targeting

highly connected nodes are more effective (Dezső and Barabási, 2002). Nevertheless, we should not overplay these theoretical considerations: Any real dataset—as opposed to a theoretical distribution—must have a finite mean and variance.

With the scale-free properties in the LSMN, random arc removal tends to become highly inefficient and costly as a control strategy (Dezső and Barabási, 2002; Eubank *et al.*, 2004; May and Lloyd, 2001; Pastor-Satorras and Vespignani, 2002b). Hence, the effectiveness of five disease-control algorithms was evaluated on the LSMN. As surveillance resources are limited, each algorithm was specified to allow a maximum of 1 000 removals, accounting for 3 % of all node-to-node arcs in the LSMN. Algorithms were evaluated according to reduction in theoretical epizootic size as measured by maximum and mean reach R_i . Maximum R_i compromises between the giant strongly connected component size (the largest number of nodes in a network that are interconnected by directed arcs) and the giant weakly connected component size (the largest number of nodes that are interconnected by undirected arcs), but effectively captures the tree-like nature of chains of infection amongst nodes.

In the LSMN, the betweenness-based algorithm performed well in reducing the susceptibility of the network to a disease epizootic with few removals. This is evaluated by using the total number of sites affected, which represents the overall epidemic size at country level, more practical in terms of regulating disease preventive and control measures than using number of shrimp infected. Removal of 3% of arcs led to a decrease in maximum R_i in excess of 68%. This is network structure dependent, and other centrality measures may become more effective for different networks. In other words, the performance of the betweenness-based algorithm is network dependent and dependent on the outcome selected (e.g. sites affected, shrimp affected, or cost). Nevertheless, betweenness centrality was linked to disease risk for bovine diarrhoea virus in Scotland (Gates *et al.*, 2014) and for the cattle network of France (Rautureau *et al.*, 2011). Most high-betweenness arcs were long distance (mean 271 km), between seed-producing sites, and constitute a very small

proportion of total arcs. These arcs play an important role in distributing shrimp seed (with their pathogens) to many ongrowing sites in the network. Onward connections can occur within a short time window, due to the life cycle of shrimp (Quispe *et al.*, 2016). For example, nursery sites rear seed from nauplius until postlarva over 20 days before selling on to ongrowing sites (FAO, 2014). Given this rate of turnover, disease surveillance and control measures applied to high-betweenness arcs (e.g. diagnostic testing) needs to be fast to achieve early detection. In this paper, arc-based surveillance is considered, rather than node-based, though most of the metrics used above to apply to arcs can be applied in a related manner to nodes. We do not intend to make recommendations on the form of surveillance measures here, only identify algorithms that can intelligently identify targets where to deploy them.

The LSMN as described excludes unreported movements and 320 self-loop arcs. Unreported movements are typically generated from non-commercial farming with low productivity and for breeding improvement purposes, whereas self-loops are caused by nodes that act as both seed-producing and ongrowing sites, but hold the same farm registration number. As of 2014 there were 21 071 registered shrimp farms in Thailand (either active but also including inactive sites), thus the majority of these are represented in our network. Production destined for export (the large majority of production) must comply with application of movement documents, however this is optional for the domestic market. However, companies producing for the domestic market will frequently also be producing for the export market and reporting movements, increasing coverage of the movement documents.

We examine here a relatively short time period of 13 months of data. Future analyses will allow for seasonal variation in network structure to be categorised, and its consistency over time. Though the data relate to an outbreak period, and thus must be treated with caution in terms of their relevance during "peace time", it is noteworthy that Thai shrimp production and annual losses were broadly similar throughout the longer time period of 2013-2015 (Putth & Polchana, 2016).

Nevertheless, 2013 marks however a notable drop in production mass compared with 2012, largely due to AHPND.

Additionally, our analysis takes place using the single network layer of live shrimp movements. Different results might arise if, for example, direct (unreported) local contact or water-borne contact were included. This extra complexity could be explored using simulation models, potentially supplemented by GIS techniques and spatial data. The implications of incompleteness of the network data due to the relatively small proportion of non-reported movements could also be explored, and whether that incompleteness is missing at random across the network, or not, introducing potential biases.

In summary, our network analysis describes the scope of potential disease transmission among the Thai shrimp farming sites via live shrimp movements. The LSMN is characterised by important epizootiological properties that both facilitate and limit disease transmission. Because scale-free properties are found in the LSMN, we emphasise that optimal targeted disease surveillance and control can potentially reduce the spread of epizootics in Thai shrimp farming, and note that arc betweenness was the most effective algorithm to identify important arcs within the network.

Acknowledgements

The authors thank to the Thailand Department of Fisheries for providing the data concerning live shrimp movements. With thanks to Dr Francis Murray.

Funding information

This research was funded by the Agricultural Research Development Agency (Public Organisation) of Thailand.

Conflict of interest

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Orcid

Darren Green ID <http://orcid.org/0000-0001-9026-5675>

References

- Ahnert, S. E., & Fink, T. (2008). Clustering signatures classify directed networks. *Physical Review E*, 78(3), 036112.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., & 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. doi:10.1073/pnas.0400087101
- Becker, N. G., & Hall, R. (1996). Immunization levels for preventing epidemics in a community of households made up of individuals of various types. *Mathematical Biosciences*, 132(2), 205-216. doi:10.1016/0025-5564(95)00080-1
- Bohm, M., Hutchings, M. R., & White, P. C. (2009). Contact networks in a wildlife-livestock host community: Identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. *PloS One*, 4(4), e5016.
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555-564. doi:10.1016/j.socnet.2007.04.002
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2), 163-177.
- Britton, T., Deijfen, M., & Liljeros, F. (2011). A weighted configuration model and inhomogeneous epidemics. *Journal of Statistical Physics*, 145(5), 1368-1384.
- Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., & Faloutsos, C. (2008). Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, 10(4), 1.
- Christley, R. M., Pinchbeck, G. L., Bowers, R. G., Clancy, D., French, N. P., Bennett, R., & Turner, J. (2005). Infection in social networks: Using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, 162(10), 1024-1031. doi:10.1093/aje/kwi308
- Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6)
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4), 661-703.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. Available at: <http://igraph.org>
- Dezső, Z., & Barabási, A. (2002). Halting viruses in scale-free networks. *Physical Review E*, 65(5), 055103.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1), 269-271.
- Draief, M., Ganesh, A., & Massoulié, L. (2008). Thresholds for virus spread on networks. *The Annals of Applied Probability*, 18(2), 359-378.

- Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988), 180-184.
- Evans, T. (2007). Exact solutions for network rewiring models. *The European Physical Journal B*, 56(1), 65-69.
- FAO. (2013). Report of the FAO/MARD technical workshop on early mortality syndrome (EMS) or acute hepatopancreatic necrosis syndrome (AHPNS) of cultured shrimp (under TCP/VIE/3304). (No. FAO Fisheries and Aquaculture Report No. 1053). Rome: FAO.
- FAO. (2014). Cultured aquatic species information programme *Penaeus vannamei* (Boone, 1931). Retrieved, 2014, Retrieved from http://www.fao.org/fishery/culturedspecies/Litopenaeus_vannamei/en
- Flaherty, M., Szuster, B., & Miller, P. (2000). Low salinity inland shrimp farming in Thailand. *Ambio*, 29(3), 174-179.
- Flegel, T. W. (2012). Historic emergence, impact and current status of shrimp pathogens in Asia. *Journal of Invertebrate Pathology*, 110(2), 166-173. doi:10.1016/j.jip.2012.03.004
- Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24), 10815-10820. doi:10.1073/pnas.0912671107
- Fournie, G., Guitian, J., Desvaux, S., Cuong, V. C., Dung do, H., Pfeiffer, D. U., Ghani, A. C. (2013). Interventions for avian influenza A (H5N1) risk management in live bird market networks. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), 9177-9182. doi:10.1073/pnas.1220815110
- Gates, M.C., Humphry, R.W., Gunn, G.J., & Woolhouse, M.E.J. (2014). Not all cows are epidemiologically equal: quantifying the risks of bovine viral diarrhoea virus (BVDV) transmission through cattle movements. *Veterinary Research*, 45, 110. doi: 10.1186/s13567-014-0110-y
- Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821-7826. doi:10.1073/pnas.122653799
- Goh, K. I., Oh, E., Jeong, H., Kahng, B., & Kim, D. (2002). Classification of scale-free networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), 12583-12588. doi:10.1073/pnas.202301299
- Green, D. M., Werkman, M., & Munro, L. A. (2012). The potential for targeted surveillance of live fish movements in Scotland. *Journal of Fish Diseases*, 35(1), 29-37.
- Green, D. M., Gregory, A., & Munro, L. A. (2009). Small- and large-scale network structure of live fish movements in Scotland. *Preventive Veterinary Medicine*, 91(2-4), 261-269. doi:10.1016/j.prevetmed.2009.05.031
- Heesterbeek, J. A. P., & Dietz, K. (1996). The concept of R_0 in epidemic theory. *Statistica Neerlandica*, 50(1), 89-110.

- Herrera, J. L., Srinivasan, R., Brownstein, J. S., Galvani, A. P., & Meyers, L. A. (2016). Disease surveillance on complex social networks. *PLoS Comput Biol*, 12(7), e1004928.
- Hong, X., Lu, L., & Xu, D. (2016). Progress in research on acute hepatopancreatic necrosis disease (AHPND). *Aquacult Int*, 24, 577-593.
- Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology* 67, 1-8.
- Kiss, I. Z., & Green, D. M. (2008). Comment on "Properties of highly clustered networks". *Physical Review E*, 78(4), 048101.
- Kiss, I. Z., Green, D. M., & Kao, R. R. (2006). The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *Journal of the Royal Society, Interface*, 3(10), 669-677. doi:10.1098/rsif.2006.0129
- Kiss, I. Z., Green, D. M., & Kao, R. R. (2008). The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. *Journal of the Royal Society Interface*, 5, 791-799. doi:10.1098/rsif.2007.1272
- Kurvers, R. H. J. M., Krause, J., Croft, D. P., Wilson, A. D. M., & Wolf, M. (2014). The evolutionary and ecological consequences of animal social networks: Emerging issues. *Trends in Ecology & Evolution*, 29(6), 326-335. doi:10.1016/j.tree.2014.04.002
- Lentz, H. H., Koher, A., Hövel, P., Gethmann, J., Sauter-Louis, C., Selhorst, T., & Conraths, F. J. (2016). Disease spread through animal movements: A static and temporal network analysis of pig trade in Germany. *PloS One*, 11(5), e0155196.
- Mao, G., & Zhang, N. (2013). Analysis of average shortest-path length of scale-free network. *Journal of Applied Mathematics*, 2013, 865643. doi:10.1155/2013/865643
- Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science (New York, N.Y.)*, 296(5569), 910-913. doi:10.1126/science.1065103
- May, R. M., & Lloyd, A. L. (2001). Infection dynamics on scale-free networks. *Physical Review E*, 64(6), 066112.
- Meyers, L. A., Newman, M., Martin, M., & Schrag, S. (2003). Applying network theory to epidemics: Control measures for *Mycoplasma pneumoniae* outbreaks. *Emerging Infectious Diseases*, 9(2), 204-210.
- Moore, C., & Newman, M. E. J. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, 61(5B), 5678-5682.
- Mrvar, A., & Batagelj, V. (1996). Pajek: Analysis and visualisation of large networks. Retrieved from <http://pajek.imfm.si>
- NACA. (2017). Quarterly aquatic animal disease report (Asia and Pacific region) 1998-2016. Thailand: NACA.
- Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical Review E*, 66(1), 016128.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.

Newman, M. E., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(3), 036122.

Noldus, R., & Van Mieghem, P. (2013). Effect of degree-preserving, assortative rewiring on OSPF router configuration. *Teletraffic Congress (ITC), 2013 25th International*, 1-4.

OIE. (2013). Acute hepatopancreatic necrosis disease, aetiology epidemiology diagnosis prevention and control references. Retrieved, 2014, Retrieved from http://www.oie.int/fileadmin/Home/eng/International_Standard_Setting/docs/pdf/Aquatic_Commission/AHPND_DEC_2013.pdf

Opsahl, T. (2009). *Structure and Evolution of Weighted Networks*. University of London (Queen Mary College), London, UK, pp. 104-122. Available at <http://toreopsahl.com/publications/thesis/> <http://toreopsahl.com/tnet/>

Pakingking Jr, R. V., de Jesus-Ayson, Evelyn Grace T, & Acosta, B. O. (2016). Addressing acute hepatopancreatic necrosis disease (AHPND) and other transboundary diseases for improved aquatic animal health in Southeast Asia. *Proceedings of the ASEAN Regional Technical Consultation on EMS/AHPND and Other Transboundary Diseases for Improved Aquatic Animal Health in Southeast Asia*, Makati City, Philippines.

Pastor-Satorras, R., & Vespignani, A. (2002a). Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3), 035108.

Pastor-Satorras, R., & Vespignani, A. (2002b). Immunization of complex networks. *Physical Review E*, 65(3), 036104.

Prakash, B. A., Chakrabarti, D., Faloutsos, M., Valler, N., & Faloutsos, C. (2010). Got the flu (or mumps)? check the eigenvalue! *arXiv:1004.0060*

Putth, S., & Polchana, J. (2016). Current status and impact of early mortality syndrome (EMS)/acute hepatopancreatic necrosis disease (AHPND) and hepatopancreatic microsporidiosis (HPM) outbreaks on Thailand's shrimp farming. *Proceedings of the ASEAN Regional Technical Consultation on EMS/AHPND and Other Transboundary Diseases for Improved Aquatic Animal Health in Southeast Asia*, Makati City, Philippines. Quispe, R. L., Justino, E. B., Vieira, F. N., Jaramillo, M. L., Rosa, R. D., & Perazzolo, L. M. (2016). Transcriptional profiling of immune-related genes in Pacific white shrimp (*Litopenaeus vannamei*) during ontogenesis. *Fish & Shellfish Immunology*, 58, 103-107.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Restrepo, J. G., Ott, E., & Hunt, B. R. (2007). Approximating the largest eigenvalue of network adjacency matrices. *Physical Review E*, 76(5), 056119.

Shirley, M. D. F., & Rushton, S. P. (2005). The impacts of network topology on disease spread. *Ecological Complexity*, 2(3), 287-299. doi:10.1016/j.ecocom.2005.04.005

Songsanjinda, P. (2015, September). *Crisis of Thai shrimp production*. PowerPoint presentation at the meeting of the National Research Council of Thailand, Bangkok.

Szuster, B. (2006). Coastal shrimp farming in Thailand: Searching for sustainability. In T. H. Chu (Ed.), *Environment and livelihoods in tropical coastal zones: Managing agriculture-fishery-aquaculture conflicts* (pp. 86-97). Norfolk, UK: CAB International.

Tanaka, G., Urabe, C., & Aihara, K. (2014). Random and targeted interventions for epidemic control in metapopulation models. *Scientific Reports*, 4, 5522. doi:10.1038/srep05522.

Woolhouse, M. E., Dye, C., Etard, J. F., Smith, T., Charlwood, J. D., Garnett, G. P., Anderson, R. M. (1997). Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America*, 94(1), 338-342.

Woolhouse, M. E., Shaw, D. J., Matthews, L., Liu, W. C., Mellor, D. J., & Thomas, M. R. (2005). Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule. *Biology Letters*, 1(3), 350-352. doi:10.1098/rsbl.2005.0331

Zhang, H., Zhang, J., Zhou, C., Small, M., & Wang, B. (2010). Hub nodes inhibit the outbreak of epidemic under voluntary vaccination. *New Journal of Physics*, 12(2), 023015.

Tables

Table 1. Degree properties of the live shrimp movement network of Thailand (LSMN) for the weighted and non-weighted networks of 13 801 nodes.

Property	Non-weighted network	Weighted network
Total degrees		
<i>in</i> degree = <i>out</i> degree	33 720	74 462
<i>undirected</i> degree	67 414	148 924
Mean degree		
<i>in</i> degree = <i>out</i> degree	2.4	5.4
<i>undirected</i> degree	4.9	10.8
Coefficient of variation		
<i>in</i> degree	0.9	1.7
<i>out</i> degree	9.2	13.4
<i>undirected</i> degree	4.6	6.9
In–out degree correlation (Pearson)	0.03	0.24

Table 2. Number of arcs (and percentage of total) between seed-producing sites and ongrowing sites based on the weighted degree for the live shrimp movement network.

		Destination	
		Seed-producing site	Ongrowing site
Source	Seed-producing site	10 775 (14.5 %)	63 596 (85.4 %)
	Ongrowing site	7 (<0.1 %)	84 (0.1 %)

Table 3. Number of arcs (and percentage of total) between seed-producing sites and ongrowing sites based on the non-weighted degree for the live shrimp movement network.

		Destination	
		Seed-producing site	Ongrowing site
Source	Seed-producing site	2 047 (6.1 %)	31 589 (93.7 %)
	Ongrowing site	7 (<0.1 %)	77 (0.2 %)

Table 4. Source and destination site types of the top 1 000 removals identified by the betweenness-based algorithm. The percentage of total arcs is shown in parentheses.

		Destination	
		Seed-producing site	Ongrowing site
Source	Seed-producing site	99.3 % (2.9 %)	0 %
	Ongrowing site	7 % (<0.1 %)	0 %

Figures

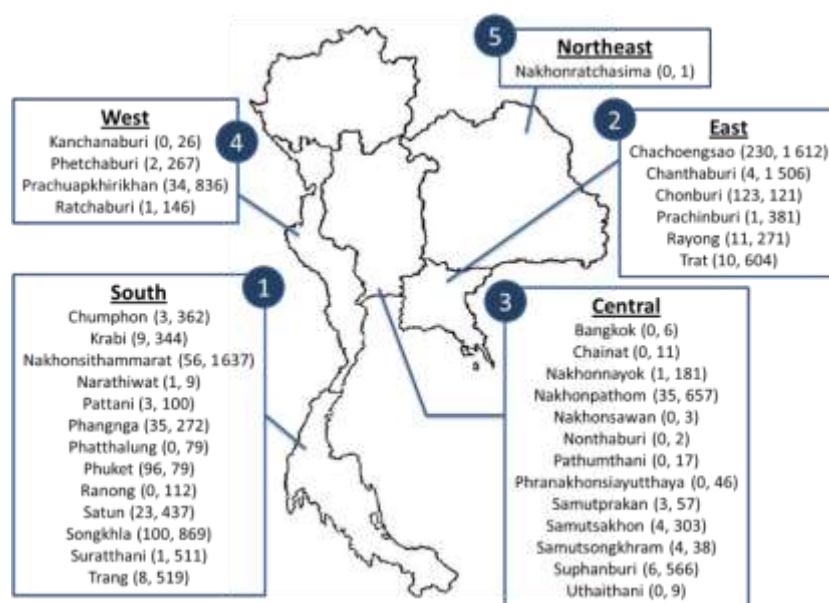


Figure 1. Distribution of the c. 13 800 shrimp farming sites across the five regions and 37 provinces of Thailand. Values in parentheses are the number of seed-producing sites and on-growing sites, respectively. Data correspond to sources and destinations of live shrimp movements from March 2013 to March 2014. Map royalty free from www.dreamstime.com.

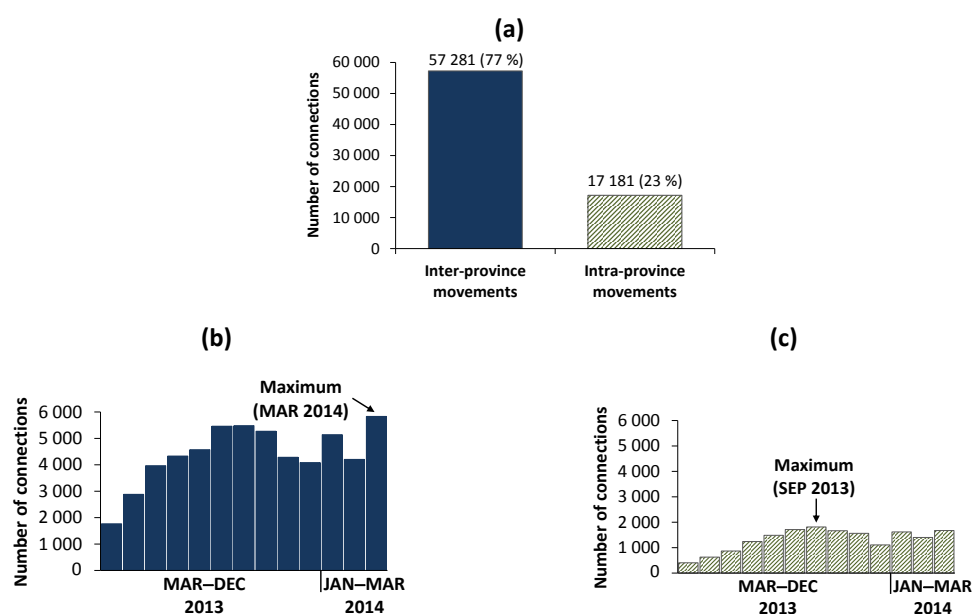


Figure 2. Distribution of the number of repeated live shrimp movements in Thailand, according to: (a) inter- and intra-province movements; (b) Calendar month for inter-province movements. (c) Calendar month for intra-province movements.

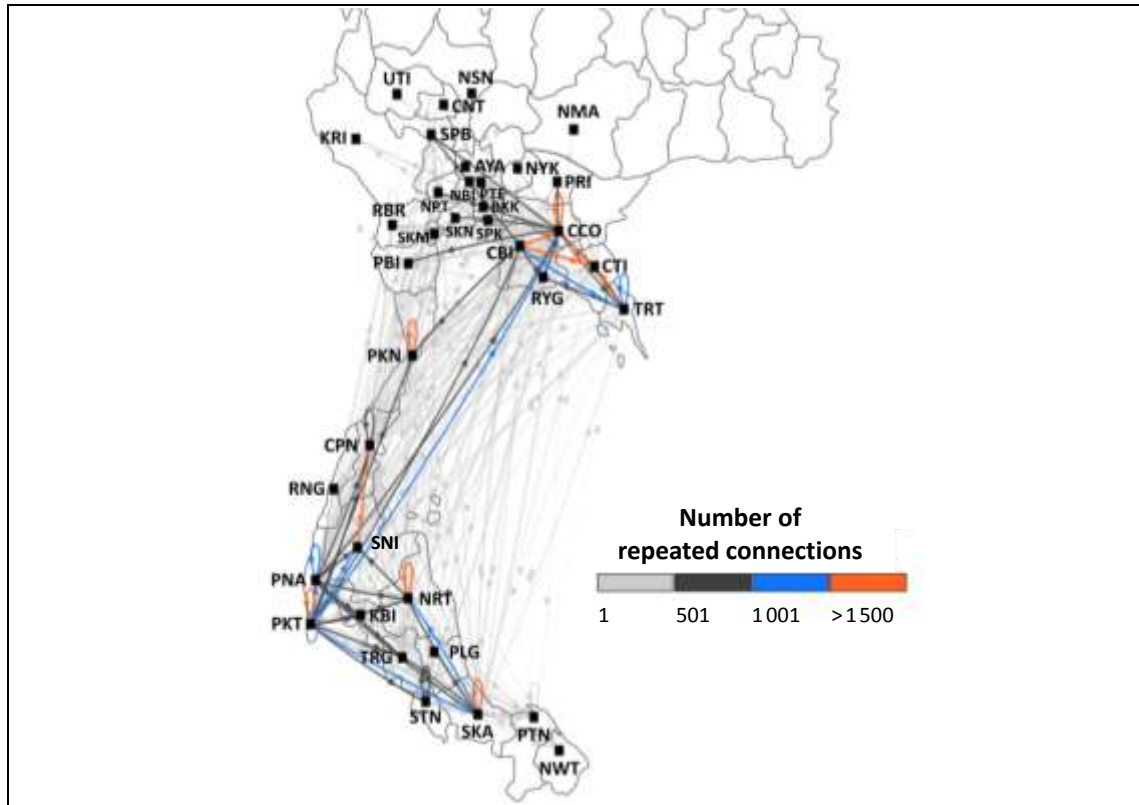


Figure 3. The provincial network structure of shrimp movements in Thailand over a 13-month period (March 2013–March 2014), at province ($n=37$) level. The line width is plotted on a log scale according to the number of repeated arcs. The abbreviation list of national provincial centres is shown in Appendix.

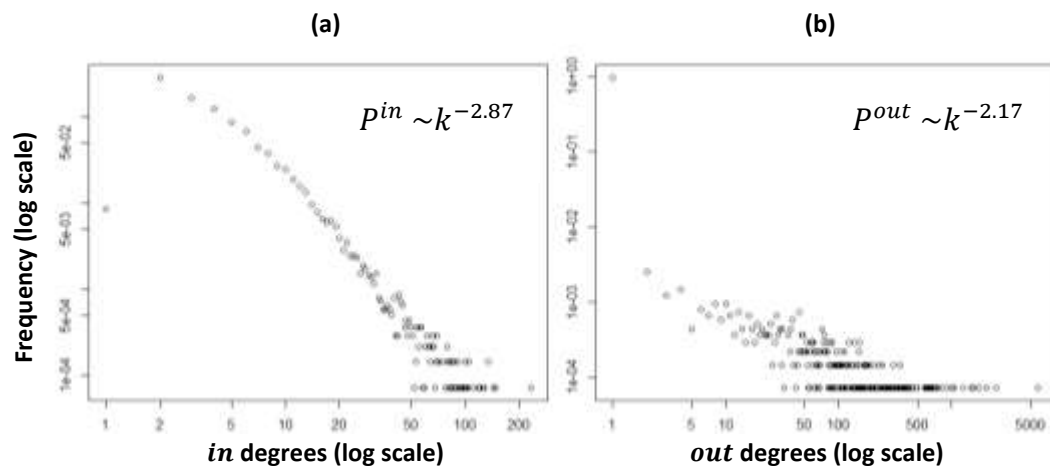


Figure 4. Weighted degree distributions for the live shrimp movement network plotted on a log–log scale.

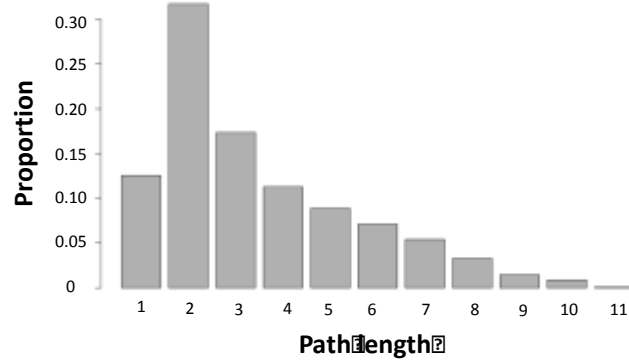


Figure 5. Distribution of weighted path lengths in the live shrimp movement network of Thailand, as a fraction of total possible paths. (Only a fraction of possible paths exists).

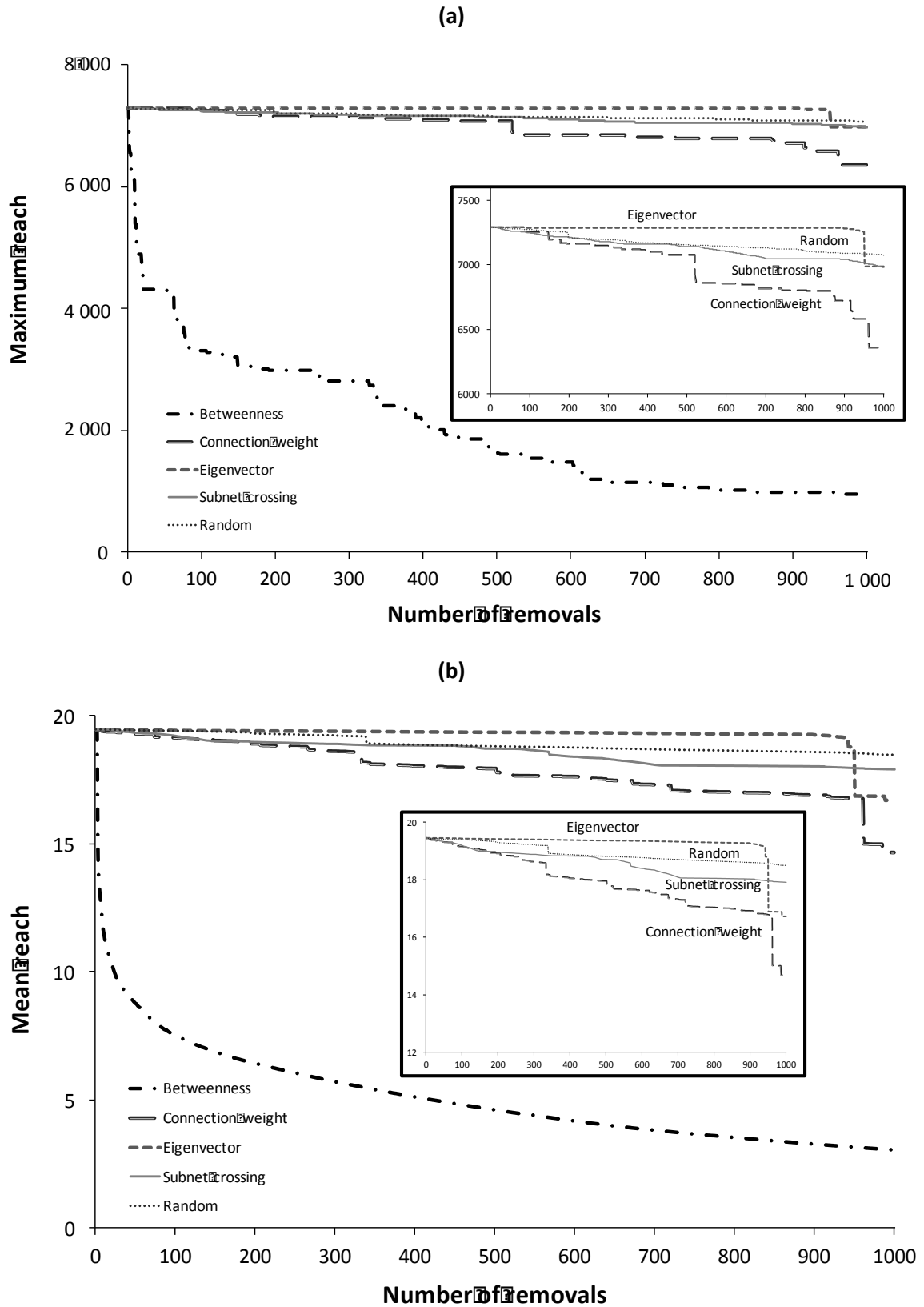


Figure 6. Evaluating reduction in network reachability using several disease-control algorithms. The betweenness algorithm performs well for both measures: (a) maximum reach and (b) mean reach.

Appendix

National provincial centres and abbreviation

National provincial centre	Abbreviation	National provincial centre	Abbreviation
Samutprakan	SPK	Nakhonratchasima	NMA
Chainat	CNT	Prachuapkhirikhan	PKN
Nakhonpathom	NPT	Ratchaburi	RBR
Suphanburi	SPB	Kanchanaburi	KRI
Uthaithani	UTI	Phetchaburi	PBI
Samutsakhon	SKN	Songkhla	SKA
Pathumthani	PTE	Pattani	PTN
Phranakhonsiayutthaya	AYA	Nakhonsithammarat	NRT
Nakhonnayok	NYK	Suratthani	SNI
Samutsongkhram	SKM	Chumphon	CPN
Bangkok	BKK	Phangnga	PNA
Nakhonsawan	NSN	Krabi	KBI
Nonthaburi	NBI	Phuket	PKT
Trat	TRT	Narathiwat	NWT
Chachoengsao	CCO	Satun	STN
Rayong	RYG	Ranong	RNG
Chonburi	CBI	Trang	TRG
Chanthaburi	CTI	Phatthalung	PLG
Prachinburi	PRI		