Modeling and Simulating of Spatial Spread
of Cross-Boundary Crop Diseases

Jiaogen Zhou [2], Xu Chen[1,*], Jingyin Zhao[1,*], and Dongsheng Wang[1]

[1] Center of Information Technology in Agriculture, Shanghai Academy of Agricultural
Sciences, Shanghai 201403, China
[2] Institute of Subtropical Agriculture, The Chinese Academy of Science,
Changsha 410125, China
`{chenxu,zjy}@saas.sh.cn`

Abstract. Understanding of the timing, orientation and transmission mode of
crop disease spread is very important in prevention and treatment of crop
diseases, especially for cross-border crop diseases. In this paper, we introduced
the complex network approach to model trans-boundary spread of diseases. In
the model, the process of crop disease spread is characterized as a network,
where patches of crop planting are abstracted into nodes, and the diseases
spread from a node to neighboring nodes. The edge between two nodes is
connected, only if the Euclidean distance between two nodes does not exceed
the radius of disease spread. We obtained wheat-growing distribution data of
Beijing in 2007 by image classification, and further abstracted patches of wheat
growing into network nodes, and finally constructed the disease spread network
to analyze its topology characteristics. The experimental results show the heavy
influence of spatial constraint on disease spread.

Keywords: complex network, spatial model, crop epidemics.

1 Introduction

Occurrence and spread of crop diseases is of great harm to crop yield and quality.
Forecasting of occurrence and spread of crop diseases is very significant for diseases
control. Existing forecasting methods can be divided into two categories: 1) linear or
nonlinear statistical models; 2) multi-spectral remote sensing prediction.

In the linear or nonlinear statistical models, there is a close association between the
occurrence and spread of crop pests and diseases with environmental factors of
growth and reproduction habitat (such as light, heat, water, terrain, etc.).So, based on
history data extensively collected and surveyed on the disease bio-statistical
information (such as spores), disease occur strength data as well as the habitat factor
data during disease occurrence, a statistical model underlying relationship between
disease bio-statistical data or intensity data occurs and environmental factors can
be constructed to predict the range, intensity or time of future disease occurrence [1].

---

[*] Corresponding author.

For example, environmental data of average temperature, precipitation and relative humidity are used to determine spreading areas of wheat scab epidemic in the irrigation areas, and ground meteorological data, upper air circulation and North Pacific sea surface temperature field factor can be effectively used in the Shanghai area of wheat scab occurrence degree forecast[2]. Statistical regression model is relatively simple and accurate, and still it is time-consuming and laborious, due to the requirement of fine and enough information of disease, environmental factors and their relationships.

In environmental conditions and physiological cycle, the reflectance spectrum of the crop follows certain distribution law. However, when subjected to disease invasion, the change of crop physiology and structure will result in the change if its reflection spectrum [3].Variation of its reflectance spectra during the disease occurrence can be used as indicators to predict or identify crop diseases For example, it demonstrates that crop spectral changes before and after the study onset are used to effectively identify the corn borer pest, wheat embroidery disease and the cotton wilt disease. Overall, the multi-spectral remote sensing prediction methods can accurately monitor the scope and area of crop pests and diseases, but not be able to predict earlier occurrence of crop diseases. After the occurrence of crop diseases, the changes of its physiological and ecological structure and reflectance spectra are not synchronized, but the slow responds. In fact, the predicted results of the multi-spectral remote sensing prediction methods show the late extent of disease occurrence, but not for early prediction in advance.

In essence, statistical models or multi-spectral remote sensing prediction methods the region to be predicted as an isolated, static unit of non-interaction with neighboring areas, and cannot give a reasonable explanation of driving mechanism that the spatial spread of disease runs from infected areas spread to neighboring areas.

In this paper, we first briefly survey the advantage and disadvantage of two traditional methods of disease occurrence forecasting, and focus on the natural essence of the cross-boundary diseases, and propose a complex network-based model to simulate the spatial spread of a new disease from a early source area of its out breaking to adjacent ones, and finally throughout the whole region. Our contribution is that the experimental results show the influence of spatial constraint on disease spread using our proposed model.

The following of the paper is organized as follows. Section 2 gives an emphasis of the spatial transition of cross-boundary diseases and propose a complex network-based model to simulate its spatial spread when a new and unknown disease outbreaks. Section 3 presents experimental results on real data, and finally, Section 4 makes a conclusion of the paper.

2   Modeling of Cross-Boundary Disease Spread

In previous section, the goodness and limits of two traditional methods are briefly surveyed. This section will focus on the characteristics of cross-boundary disease transmit ion , and further give a hypothesis to model and simulate the process of its spatial spread when a new and unknown disease outbreak.

2.1   Model Hypothesis

Agricultural cross-border disease usually first occurs in a planting area (the source of the disease area), and then gradually spread to the surrounding adjacent planting areas, finally throughout the region. There is a significant difference between cross-border agricultural diseases and conventional ones, which disease transmission only from the source of the disease area spread to nearby areas that are not infected. Additionally, considering that radius of disease spread is limited, whether the disease occurrence in non-infected areas appears or not depends on its spatial distance to the source of the disease. So for any new and unknown disease, a hypothesis is proposed, that is :(1) its propagation path is only from the pathogens areas to the adjacent non-epidemic areas, and whether the outbreak of the new epidemics in crops of adjacent non-epidemic areas depends on spatial spread radium of the epidemic; (2) crops are certainly infected within a radius of the epidemic spread, regardless of environmental factors, while not outside the radius;(3)subjectively, epidemic spread is not allowed to occur between adjacent non-epidemic areas to reduce the complexity of the network.

Based on this, the spatial spread of trans-boundary diseases is considered as network evolution problem, the complex network theory method is used to solve the problem. Given study area, crop plaques are abstracted into a graph node, the onset of disease from one planting unit spread to neighboring units, even edge between two nodes, in order to build a crop epidemic communication network.
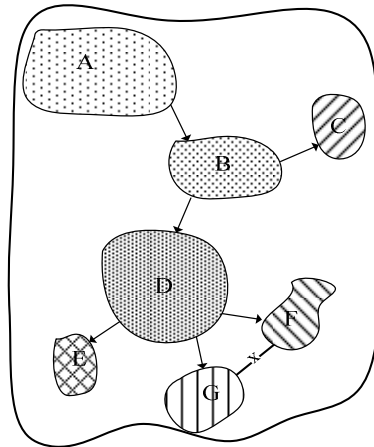

Fig. 1. The process of crop disease occurrence and spread. Polygons means different neighboring planting plaques.

To further introduce the process of construction of epidemic spread network in detail, an example of disease spread is given. Figure 1 indicates a crop planting area consisting of A~G seven polygon blocks. In the blocks, different/same crops grow, and are susceptible to the infection of a new epidemic. Supposing a new epidemic first appears in the A block, and then the epidemic spreads to the B block close to A
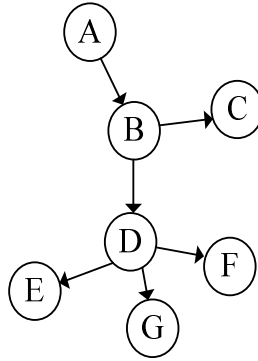
Fig. 2. Directed graph mapped into spatial spread of crop disease. the circles stand for nodes, while the arrows for the edges.

and within the spread radium, finally appear in all other blocks. By the limit of the model assumption, epidemic spread doesn't appear between the blocks of F and G, which are non-epidemic areas. The epidemic spread is abstracted to the network, which its nodes stands for the blocks and directed edge for the interaction of two nodes, and then a simple network consisting of seven nodes and six edges appears (shown in figure 2).

2.2   Evaluation Index

There is a great influence on network topology on the spread of the disease [4-6].Investigation of network topology characteristics helps to understand the disease propagation. The network structures, which meet random, uniform, small world or free scale distribution law are generated to investigate the robustness and vulnerability of different types of network on disease spread [6-8]. Network degree distribution exponent, clustering coefficient, node / edge betweenness, average path length are used to indicate the network topology characteristic [9-11]. Here two indexes of clustering coefficient and average path length are considered as the indicators of epidemic spread.

  A clustering coefficient is a measure of degree to which nodes in a graph tend to cluster together. Average path length is a concept in network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network. Giving a disease spread network consisting of a set of vertices V and a set of edges E between them. The degree k_i for a vertex V_i is defined as the number of its immediately connected vertexes, E_i as the number of its immediately connected edges and local clustering coefficient (C_i) for V_i is then given by the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them. So for all vertexes, the average clustering coefficient (CC) can be obtained from the equation (2). Let d (v_i, vj) denote the shortest path between v_i and v_j. The average path length is calculated by the equation (3).

$$C_i = \frac{E_i}{k_i(k_i - 1)} \tag{1}$$

$$C_i = \frac{\sum C_i}{n} \tag{2}$$

$$APL = \frac{1}{n(n-1)} \sum_{i,j} d(v_i, v_j) \tag{3}$$

3   Experiment

3.1   Data Preparation

In this paper, all algorithms are encoded using vb.net language, and run inWin7 operation platform. The Study area of Beijing is located at latitude 39 ° 28 '~ 41 ° 05', longitude 115 ° 24 '~ 117 ° 30'   with Plains of southeast, western and northern mountains, has a temperate continental climate.

In the study area, there are major land use types of grain field, woodland, orchard, and residents and facilities for agricultural land (vegetable).The land use of wheat data is abstracted from classified agricultural land in Beijing, and used to construct to the spread network of diseases, and simulate the response of diseases spread network of environmental elements, spatial elements and scale changes.

Classification of the land use of wheat in the suburb of Beijing are with the following data: the remote sensing data (4 m spatial-resolution panchromatic image of Landsat TM in March, 2007 and 100m spatial-resolution Beijing-1 satellite image in March, 2007), basic GIS data (1:10000 basic farmland of 1996, 1:100000 land use data, as well as 1:10000 Beijing administration, roads, water systems, residential and other data).To identify the land use type of wheat n the suburb of Beijing through remote image classification, data preprocessing is first performed. The process of data preprocessing includes projection transformation and convert geographic coordinates, geometric correction and image fusion. Projective transformation and conversion of geographical coordinates are done for all data, to convert to a variety of reference system to the uniform Beijing local coordinate system.

The Image geometric correction process is as follows: first crudely correct Beijing-1 satellite image and TM image geometry using1:50000 topographic map of the Beijing, and then further precisely correct them with truth ground GPS control points, in order to ensure the geometric correction accuracy of better than 1 pixel. The Beijing-1 satellite panchromatic image is high clear, but subject to the limitations of the single-band image, and so it is very difficult to distinguish its approximate surface features. TM image has multi-spectral bands characteristics of surface features, but the low resolution. Therefore, after fusion of the Beijing-1 satellite image and TM image, a high-resolution and multispectral image is obtained.

Based on the reference image (the fusion image), visual interpretation of surface features begins. In interpretation process, the following data of 1:1000000 basic farmland data of 1996, 1:100000 land use data of 2004 as well as 1:10000 Beijing administration, roads, water systems, residential and other data are taken as the priori knowledge. Clearly, the land use types of non-concern, such as cities, villages, water, roads, etc., are directly excluded, and but the type of wheat is outlined and specified attributes. The land use types not easy to determine are further confirmed through aerial photos and field investigation. Finally, the land use data of wheat (shown in figure 3) is used to map into the network which is characteristic of spatial spread of disease.
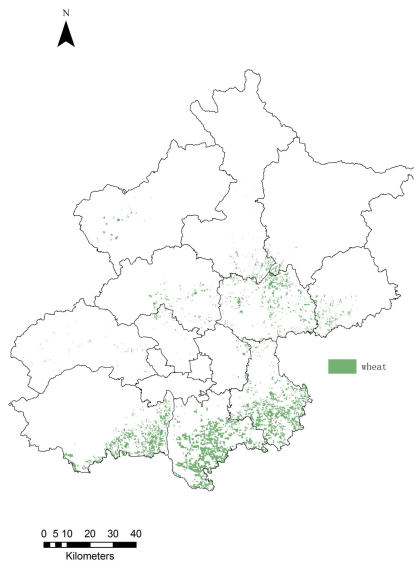

Fig. 3. The land use type of wheat

3.2   Topology Characteristics of Disease Spread

The land use of wheat extracted from remote sensing image of Beijing is mapped to an epidemic spread network. To test the influence of topology structure on epidemic spread, a new epidemic first randomly outbreak in a node, and then gradually transfers from the node to others. Considering epidemic spread radium heavily affect network structure, a variety of epidemic spread radium are set to generate the corresponding networks. For every generated network, its clustering coefficient and average path length are calculated (shown figure 4(a-b)). The results show that the network is not connected with the small radius of disease spread, and disease spread occurs only in a small amount of local connected subnets. With increasing of disease transmission radius, the network's clustering coefficient and average path length gradually show the topology features of a small world network that means when

spread radium exceeds a certain range, spatial segmentation is no longer a factor to inhibit epidemic spread. So an interesting conclusion can inferred from the above results, which spatial isolation can deter epidemic spread with small spread radium, while with the radium more than a certain range, the epidemic will quickly spread the whole region.
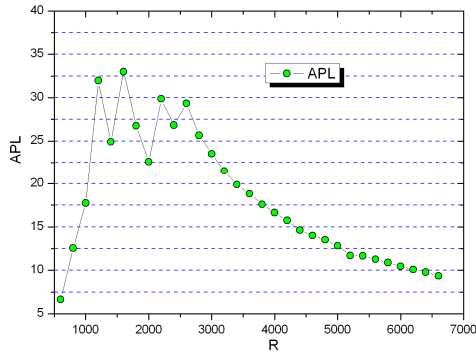
(a)
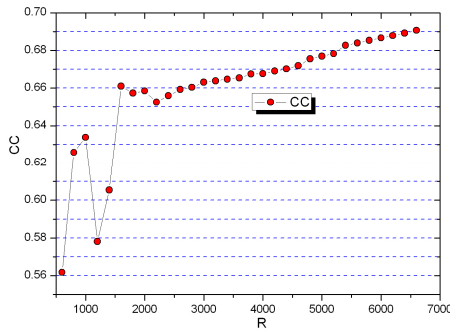
(b)

Fig. 4. The results of avarage path length (APL) and clustering coefficient of disease spread network, respectively

4   Conclusion

In this paper, we have introduced and discussed the problem of a crossing boundary crop disease spread, and construct a complex network-based model to simulate crop disease spread. We took wheat-growing distribution data of Beijing in 2007 by image classification, and further abstracted patches of wheat growing into network nodes, and finally constructed the disease spread network to analyze its topology characteristics. The results demonstrate that the network gradually shows the topology features of a small world network with epidemic spread radium increasing. Based on that, an interesting conclusion can be inferred that spatial segmentation may be a

reasonable treatment in control of a new epidemic spread with a small spread radium, while when the radium exceeds a certain threshold, spatial segmentation will not so work that the epidemic will quickly spread all over.

References

1. Nutter, F.W., Guan, J.: Quantifying alfalfa yield losses caused by foliar diseases in Iowa, Ohio, Wisconsin and Vermont. Plant Disease 86, 269–277 (2002)
2. Guan, J., Nutter, F.W.: Relationships between defoliation, spot pathosystem. Computers and Electronics in Agriculture 37, 97–112 (2002)
3. Crooks, W.T., Archer, D.J.: SAR observations of dry land moisture—towards monitoring outbreak areas of the brown locust in South Africa. In: Proceedings of IEEE International Symposium on Geosciences and Remote Sensing 2002, vol. 4, pp. 1994–1996 (2002)
4. Vittoria, C., Alessandro, V.: Epidemic modeling in metapopulation sysyems with hetergeneous coupling pattern: Theory and simulations. Journal of Theoretical Biology 251, 450–467 (2008)
5. Ball, F., Mollison, D., Scalia, T.G.: Epidemics with two levels of mixing. Ann. Appl. Proba. 7, 46–89 (1997)
6. Barthelemy, B., Bartar, M., Vespignani, A.: Dynamical patterns of epidemic outbreaks in complex heterogeous networks. Journal of Theoretical Biology 235, 275–288 (2005)
7. Cooper, B.S., Pitman, R.J., Edmunds, W.J., Gay, N.J.: Delaying the international spread of pandemic influenza. PLoS Med. 3, e12 (2006)
8. Colizza, V., Pastor-Satorras, R., Vespignani, A.: Reaction-diffusion process and metapopulation models in heterogeous networks. Nat. Phys. 3, 276–282 (2007b)
9. Newman, E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Physical Reviw E 64, 118 (2001)
10. Bodin, O., Norberg, J.: A network approach for analyzing spatially structured populations in fragmented landscape. Landscape Ecology 22, 31–44 (2007)
11. Albert, R., Barabasi, A.L.: Statistical mechanics of comlex network. Review of Modern Physics 74, 47–91 (2002)