

# Spread Patterns of Epidemics: Survey

Siddharth Satish  
Department of ISE,  
R. V. College of Engineering  
Bangalore, India

Smitha G. R.  
Assistant Professor,  
R. V. College of Engineering  
Bangalore, India

## ABSTRACT

The main purpose of this paper is to bring out a distinct strategy to promote and predict the dispersion behaviors of epidemic diseases before they actually happen. With innumerable cases of epidemic outbreaks being recorded in various parts of the world, techniques such as those discussed in this paper, if widely used, under supervision can help to circumvent such occurrences even before they actually happen with accurate predictions. This would also ensure a better coping mechanism is provided to study the spread of such infections and adequate control mechanisms are put in place to prevent loss of human life. Through the course of the paper, the hope is to develop a well-defined prediction methodology that can predict the likeliness of an individual being affected by a particular epidemic through assessing of early symptoms and also predicting future instances in their infancy.

## General Terms

Big Data Analytics, Predictive Analytics, Data Science, Disease Forecasting.

## Keywords

Epidemics, Diseases, Forecasting, Pathogen, Detection, Bio-surveillance.

## 1. INTRODUCTION

With the recent growth of tourism and global nature of modern life, diverse societies are evolving to form new communities of people in different parts of the world. With this form of vibrancy, people are placing an increased importance on health and the threat of diseases [1]. The health officials are doing everything they can to obtain timely information on epidemic outbreaks. They are putting several mechanisms in place to identify the first signs of an epidemic and to contain them as soon as possible. The conventional methods that are being utilized to obtain data from practitioners and aggregate them is a slow process. Officials and governments are working in unison to find more efficient means to predict the outbreak of an epidemic and to control its impact.

The timeline of epidemic infections and outbreaks is usually divided into several stages. The first stage is called as the Susceptible state. A transition happens when a person who is infected and a person who is in Susceptible state come in contact with one another [2]. When a host organism becomes infected, the communicable agent begins to reproduce inside of the host, to enable the infected host to spread the epidemic onto another Susceptible organism. This time between which a host is infected and is capable of spreading the infection onto another host is called as the Latency state. Figure 1 shows the timeline of an Epidemic Infection.

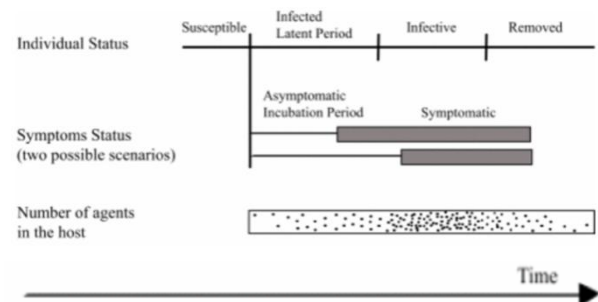


Fig.1.1 Epidemic Infection Timeline

The period before which the infected host starts to show early signs of the epidemic is called as the Incubation state. It is possible for a host to be infectious and spread the epidemic onto other organisms, but show no signs of the epidemic itself. This maybe as a result of an immunity that the individual may possess.

Specific symptoms can then be tracked based on the type and degree of the outbreak. Particular conditions can also be tracked by examining permutations of these symptoms [4]. We attempt to predict the spread pattern of the epidemic based on the number of cases that are initially reported in a sporadic manner and also by utilizing social media symptom data.

## 2. METHODOLOGY

Several models are used to represent the symptom data that is obtained from various sources. Some of the most used data models are represented below

### 2.1 Risk Assessment Model

It correlates specifics of a particular place – such as humidity, rainfall, weather which are the risk factors to calculate the risk of the epidemic.

### 2.2 Epidemic Prediction Model

It assigns a probability to each epidemic based on specific data inputs related to the epidemic and other variables. It provides a prediction of whether an epidemic is likely to occur under the current circumstances.

### 2.3 Spatial Model

It predicts the outbreak of the epidemic in that particular geographical location based on the relationship between the outbreak and the geographic factors. They keep changing as time progresses as the geographic factors of a place might vary considerably over time.

### 2.4 Dynamic Model

It illustrates how a particular epidemic is likely to surge through a given population. It includes finite parameters such as restriction of movements, presence of an immunity. It can also be used to find out when the pathogenicity of an infectious disease may change.

## 2.5 Event Detection Model

It tries to identify the outbreak of an epidemic by analysis of real time clinical data and social media symptomatic data. Through this it tries to identify likely causes of a certain epidemic or first signs of a particular infectious disease [4].

All the data that is collected at the initial stage of an epidemic or in situations where an outbreak of an epidemic is expected is aggregated and is subjected to the above data modelling process. The modelled data is then used to classify the epidemic on various aspects and identify sections of populations that are most likely to be affected by such an event of outbreak.

The Bayesian Model was also incorporated in order to further model and analyse the data. Bayesian Model is a type of Spatial Model that is mainly utilised to predict the portions of the population that are likely to be affected next. This model has seen large amounts of success after being applied on other infectious diseases such as Pneumonia and Influenza Virus [3]. The Bayesian Model was able to make accurate predictions related to percentage of people likely to be affected in the next window frame.

The first step that is performed is the sampling of data. For this purpose, a weighted adjacency matrix is constructed for each set of sampled data, in which the nodes represent areas affected by the epidemic and the edges represent the epidemic spread pattern. An edge  $a_{ij}$  is said to be non-zero if there is a confirmed movement of the epidemic from a place  $i$  to place  $j$  [5].

- a) The movement of epidemic first has to be sampled. For this purpose, we list the recorded movements of the epidemic and then randomly sample them [8].

$$A^{rebuild} = W^{original} A^{depleted} / W^{depleted} \quad (Eq\ 2.1)$$

The depleted network of epidemic spread patterns is then built from the remaining movements and the resultant network that is obtained, is rescaled such that the total weight of the rebuilt network is equal to that of the original network.

- b) The next step after sampling the movement is to sample each node and each edge attached to it [7]. A node is known as ‘captured’, if it connected to a node that is sampled, but is not sampled. For this purpose, we consider a network of  $N$  nodes. Let  $N_c$  indicate the captured nodes and  $N_s$  indicate the sampled nodes [5].

We first estimate the total weight of the network, using –

$$W^{original} = (\omega^{sampled})N/2 \quad (Eq\ 2.2)$$

Then we estimate the average weighted degree of the sampled nodes using the formula –

$$(\omega^{sampled}) = 1 / N_s \sum_{i \in S} \sum_j (a_{ij} + a_{ji}) \quad (Eq\ 2.3)$$

We then rescale the network using Eq 2.1, but by using the estimate for the total weight from Eq 2.2.

- c) Once these steps are followed, we have effectively reduced the size of the network that has to be scanned. Now a stochastic susceptible–infectious–recovered (SIR) model is used in order to test the predictions that are made on the smaller network to the reality of the situation in the actual network [3].

The probability of a particular area ( $\lambda_i$ ) being infected by an epidemic is calculated using,

$$\lambda_i = 1 - \exp \left\{ -\beta \sum_j a_{ij} I_j \right\} \quad (Eq\ 2.4)$$

Here,  $I_j = 1$  if an area is infected by an epidemic and  $\beta$  is the rate at which the epidemic spreads in a given area. Infected areas recover within a time  $T$ , after which they cannot be affected by the same epidemic again.

We calculate the probability by assuming different values for  $T = 7, 14, 21, 28$  days. Based on the values obtained for the probability, the spread pattern is estimated [3].

## 3. RESULTS

Once the data modelling of the data was complete, it was revealed that close to 84% of all transmission of infectious diseases causing epidemics is either through direct contact or through vector-borne means. Water and other means of spread of diseases accounted for the rest [1].

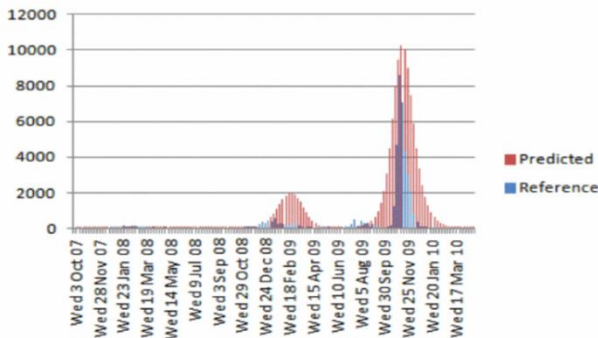
The Bayesian Model was used to make weekly and monthly predictions on several infectious diseases which have the potential to finally lead to an epidemic [3]. The communicable diseases that were chosen for this experiment were Influenza Virus, RS- Virus, Rhinovirus and Pneumonia [1].

These results are shown through Table 3.1 and Fig.3.1

**Table 3.1 Root Mean Squared Predictions**

Disease	Weekly	Monthly
Influenza A	0.0721	0.1713
RS-virus	0.1299	0.2437
Rhinovirus	0.1498	0.2545
Norovirus	0.1243	0.1863
Mycoplasma pneumoniae	0.1424	0.1735

The above table shows the predictions for the occurrences of each of the epidemic diseases listed above.



**Fig.3.1 Prediction of Influenza Virus using Bayesian Model**

The results showed that the Bayesian Model was closer in its weekly predictions but wavered further away during the monthly predictions [6].

#### 4. CONCLUSION

The various data modelling techniques to classify large scale data that is obtained from various sources have been discussed over the course of this paper. Each of the data is modelled and is assigned a particular level of likeliness based on the spatial features and other parameters. Also, a well-defined and well-trusted means for predicting the likeliness of an epidemic has been well defined and documented. The Bayesian Prediction Model takes into account likelihood and other important factors as it tries to assess the weekly outbreak percentages of the epidemic.

The main contributions of this paper are three-fold. Firstly, a neoteric technique to collect large amounts of symptomatic data from various sources and model it easily has been defined. Secondly, a generalised epidemic prediction approach based on the current activity of the epidemic has been developed. Finally, an attempt to forecast future activity and spread patterns of the outbreak based on patterns viewed from the Bayesian Model body is made.

The data collected here, has been reverse-engineered to predict the likeliness of forecasting the epidemic if it were available before-hand. The future work in relation to this paper involves, working on current data and trends regarding similar issues and predicting the occurrence of an epidemic in a certain area. The success or accuracy of the prediction is yet

to be fully validated and will be confirmed based on the current data and whether or not it is successful in predicting and avoiding a future epidemic outbreak.

#### 5. ACKNOWLEDGMENTS

This research was carried out in collaboration with the Information Science and Engineering Department of R.V. College of Engineering.

#### 6. REFERENCES

- [1] Daphne Lopez, M. Gunasekaran, B. Senthil Murugan, Spatial big data analytics of influenza epidemic in Vellore, India, IEEE International Conference on Big Data, 10.1109/BigData.2014.7004422, 2014.
- [2] Nadra Guizani, Arif Ghafoor., Modelling and evaluation of disease spread behaviours, Wireless Communications, and Mobile Computing Conference, 10.1109/IWCMC.2014.6906491, 2014.
- [3] Konstantinos P. Exarchos et al, "Prediction of coronary atherosclerosis progression using dynamic Bayesian networks", 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2013.
- [4] Courtney D. Corley et al, "Disease Prediction Models and Operational Readiness", in PMC3960139, 2014.
- [5] Nicholas Then, et al, "DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response", in plos - /journal.pone.0155417, 2016.
- [6] João Andrade, Artur Arsenio, Epidemic Spreading Over Social Networks Using Large-scale Biosensors: A Survey, Volume 5, Pages 922-931, Procedia Technology, 2012.
- [7] Lauren N. Carroll, et al, "Visualization and analytics tools for infectious disease epidemiology: A systematic review", Volume 51, Pages 287-298, Journal of Biomedical Informatics, 2014.
- [8] Baroba.si Albert-Laszlo, Reka Albert, "Emergence of Scaling in Random Networks", Science, vol. 286, pp. 509-512, Oct.1999.