Outbreak of Infectious Diseases through the Weighted Random Connection Model

M. Bradonjić*

Bell Labs, Alcatel-Lucent, 600 Mountain Avenue 2C-318, Murray Hill, NJ 07974, USA

Abstract. When modeling the spread of infectious diseases, it is important to incorporate risk behavior of individuals in a considered population. Not only risk behavior, but also the network structure created by the relationships among these individuals as well as the dynamical rules that convey the spread of the disease are the key elements in predicting and better understanding the spread. In this work we propose the weighted random connection model, where each individual of the population is characterized by two parameters: its position and risk behavior. A goal is to model the effect that the probability of transmissions among individuals increases in the individual risk factors, and decays in their Euclidean distance. Moreover, the model incorporates a combined risk behavior function for every pair of the individuals, through which the spread can be directly modeled or controlled. We derive conditions for the existence of an outbreak of infectious diseases in this model. Our main result is the almost sure existence of an infinite component in the weighted random connection model. We use results on the random connection model and site percolation in $\mathbb{Z}^2$.

Keywords and phrases: epidemic modeling, Poisson point process, random connection model, infinite component, percolation

Mathematics Subject Classification: 35Q53, 34B20, 35G31

1. Introduction

The spread of infectious diseases is a network dynamic process, whose prediction could improve our life quality. Addressing this challenge, which is both theoretical and computational, requires better understanding of the underlying network and behavior of interacting individuals in the network.

Concretely, the spread of infectious diseases has been vastly studied in the area of mathematical epidemiology. The study mainly includes: modeling [3], parameter estimation [5], simulations, and sensitivity analysis [4]. Mathematical epidemiology is usually concerned with partitioning the entire population of individuals into compartments, for example, S, I, R, E, and M, which respectively correspond to susceptible, infectious, recovered, exposed, and young individuals with maternal antibodies.

Originally, many classic compartmental epidemic models have been developed such as: SI, SIS, SIR, SEIS, SEIR, MSIR, MSEIR; for an excellent reference see [1]. All of these compartmental models aimed to describe, through a system of differential equations, the number of individuals that belong to a certain

---

*Corresponding author. E-mail: `milan@research.bell-labs.com`

compartment. It is assumed that the individuals are uniformly distributed in space. Moreover the notion of geometry does not exist in these compartmental models.

In this work we propose a weighted version of the random connection model [8], in order to address some population structure properties that appear in infectious diseases. Each individual of the population is characterized by two parameters: its position and risk behavior. Our goal is to model the effect that the probability of transmissions among individuals increases in the individual risk factors, and decays in their Euclidean distance. Moreover, our model incorporates a combined risk behavior function for every pair of individuals through which the spread can be directly modeled or controlled. We are interested in conditions for the existence of an outbreak of infectious diseases in this model. The main result is the almost sure existence of an infinite component in the weighted random connection model. We use results on the random connection model and site percolation in $\mathbb{Z}^2$.

2. Model

The random connection model has been well understood, see the monograph [8]. The probability of the connection between any two points is given as a non-increasing function that depends only upon the distance between the points. One easily finds that those connection probabilities are dependent since the (Euclidean) distances between pairs of any three points are mutually dependent. In the sequel we formally state this model.

Random Connection Model. Consider the Poisson point process $\mathcal{X}$ on $\mathbb{R}^2$ with intensity $\lambda > 0$. Two points $x_i, x_j \in \mathcal{X}$ are connected with probability $g(x_i - x_j)$, where the *random connection* function $g : \mathbb{R}^2 \to [0, 1]$ is a non-increasing and depends only on the Euclidean distance $|x_i - x_j|$ between the pair of points. It is assumed that the integrability condition is satisfied

$$I := \int_{\mathbb{R}^2} g(x)dx \in (0, \infty) , \tag{2.1}$$

otherwise: (i) when $I = \infty$ then a.s. the degree of the point x_0 at the origin $\deg(x_0) = \infty$, and (ii) when $I = 0$ then a.s. $\mathbb{P}(\deg(x_0) = 0) = 1$, see [6, 8].

The goal of this work is to model the spread of infectious diseases in unoriented discrete networks. To the best of our knowledge the random connection model has not been extensively used in the epidemics literature. When modeling the epidemic spread, it is important to incorporate risk behavior of the individuals. However, the random connection model does not include risk behavior. Not only risk behavior, but also the network structure and dynamical rules are the key elements that control the spread of infectious diseases.

In this work we propose a node weighted modification of the random connection model. Concretely, each individual i of the population is characterized by its two parameters: position x_i given by the Euclidean coordinates, and risk behavior parameter Z_i. The goal is to model the effect that the probability of the transmission from i to j increases in the individual risk factors both z_i and z_j, and decays in their Euclidean distance $|x_i - x_j|$. We concentrate on the two-dimensional space $\mathbb{R}^2$. The following proofs will be the adoption of the ones on the random connection model [6]. We now formally state the weighted random connection model.

Weighted Random Connection Model. Consider the Poisson point process $\mathcal{X}$ on $\mathbb{R}^2$ with intensity $\lambda > 0$. For each point of the process $x_i \in \mathcal{X}$, where $i = 0, 1, 2, \ldots$, associate risk behavior parameter $Z_i \geq 0$, which is a random variable that combines both infectious and susceptibility of the individual i. Assume that Z_i are i.i.d. random variables drawn from a probability density function $f(z)$ on the support $[0, +\infty)$. Let $g : \mathbb{R}^2 \to [0, 1]$ be a non-increasing function that depends only upon the Euclidean distance $|x_i - x_j|$ between the pair of points, and satisfies the integrability condition $0 < I < \infty$, see (2.1). Let $\sigma : \mathbb{R}^2 \to [0, 1]$ be a symmetric function, increasing in both z_i and z_j, which models a combined risk behavior of two individuals. Finally, two points $x_i, x_j \in \mathcal{X}$ with the risk factors z_i, z_j, are connected with probability

$$g(x_i - x_j)\sigma(z_i, z_j). \tag{2.2}$$

Let us now comment on the combined risk behavior function $\sigma(z_i, z_j)$. The function $\sigma(z_i, z_j)$ needs only to have positive partial derivatives and to be a mapping into the interval $[0, 1]$. If so, then the spread of infectious diseases can be designed through the function $\sigma(z_i, z_j)$. For example: (i) $\sigma(0.01, 0.9) = \sigma(0.9, 0.01) \approx 0.0$ indicates small infectious probabilities among the individuals one of whom is with low risk factor, (ii) $\sigma(0.01, 0.9) < \sigma(0.1, 0.9)$ indicates the monotonic property of $\sigma(z_i, z_j)$, (iii) $\sigma(0.9, 0.9) \approx 1.0$ indicates large infectious probabilities among the individuals with high risk factors.

Now, consider the node degree distribution of point $x \in \mathcal{X}$. One immediate observation is that node degree of an individual x with risk factor z follows the Poisson probability distribution with parameter

$$\mu(z) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^2} \lambda \sigma(z, z') g(x) dz' dx = \lambda \int_{\mathbb{R}^+} \sigma(z, z') dz' \int_{\mathbb{R}^2} g(x) dx = \lambda I \int_{\mathbb{R}^+} \sigma(z, z') dz' . \tag{2.3}$$

That is, the degree distribution of x given $Z = z$, for $\ell \geq 0$, is given by

$$\mathbb{P}\left(\deg(x) = \ell : Z = z\right) = e^{-\mu(z)} \frac{\mu(z)^\ell}{\ell!}. \tag{2.4}$$

Taking the expectation over Z, we obtain that the degree distribution of x follows

$$\mathbb{P}\left(\deg(x) = \ell\right) = \int_{\mathbb{R}^+} f(z) e^{-\mu(z)} \frac{\mu(z)^\ell}{\ell!} dz. \tag{2.5}$$

3. Main Result

In this section, we concentrate on the almost sure existence of an infinite component in the weighted random connection model. Given the symmetry, without loss of generality, consider the origin denoted by x_0, and the component containing x_0 denoted by $C(x_0)$. First, define the probability of the point at the origin x_0 being contained in the infinite component by

$$\theta(\lambda) = \mathbb{P}_\lambda(|C(x_0)| = \infty). \tag{3.1}$$

Second, notice that the realizations of risk behavior Z_i are independent of the Poisson point process $\mathcal{X}$. We now provide an argument that $\theta(\lambda)$ is monotone (non-decreasing) in the density λ of the process $\mathcal{X}$.

Consider a graph G generated by the process $\mathcal{X}$ with parameter λ and its realizations of risk behavior Z. Let G' be obtained by thinning G with probability $1 - \lambda'/\lambda$ where $0 < \lambda' < \lambda$. In other words, each point x_i of the process $\mathcal{X}$ as well as its risk factor z_i remove independently with probability $1 - \lambda'/\lambda$ (that is, keep x_i and z_i with probability λ'/λ). The obtained graph G' is a subgraph of G. Moreover G' is a realization of the original model with parameter $\lambda' < \lambda$. Hence the existence of an infinite component containing x_0 in G' implies the existence of one in G. Finally we conclude that $\theta(\lambda)$ is monotone, that is, for every $\lambda' < \lambda$ the following holds

$$\mathbb{P}_{\lambda'}(|C(x_0)| = \infty) \leq \mathbb{P}_\lambda(|C(x_0)| = \infty). \tag{3.2}$$

Our main result is given by the following theorem which shows the almost sure existence of the infinite component that contains the origin. More precisely, we show that: (i) for any sufficiently small $\lambda > 0$ the probability that x_0 belongs to an infinite component is zero, and (ii) an infinite component exists with positive probability when the density λ of the process is greater than some positive real number. From the fact that $\theta(\lambda)$ is monotonic in λ, the existence of a unique critical λ_c will follow.

Theorem 3.1. *There exists a finite critical value $0 < \lambda_c < \infty$, such that $\theta(\lambda) = 0$ for $\lambda < \lambda_c$, and $\theta(\lambda) > 0$ for $\lambda > \lambda_c$.*

Proof. The proof consists of two parts.

Part 1. As in [6], we construct a process which will be stochastically dominated by a tree whose probability of extinction is one. Thus consider the following process starting from the origin x_0.

The nodes will be sequentially labeled x_t for $t = 1, 2, \ldots$ as created by the following independent Poisson processes. We will inductively iterate over all levels and over all points given at a considered level. Create all neighbors of x_0 given by the Poisson point process with intensity $\lambda g(x - x_0)\sigma(x, x_0)$. Notice that $\lambda g(x - x_0)\sigma(x, x_0) \leq \lambda g(x - x_0)$. Next, we create the progeny of x_1, that is, we create a new independent Poisson point process with intensity $\lambda g(x - x_1)\sigma(x, x_1)(1 - g(x - x_0)\sigma(x, x_0)) \leq \lambda g(x - x_1)$. The points of this process are all connected to x_1 but not to x_0. Similarly, we create the progeny of x_2, that is, we create a new independent Poisson point process with intensity $\lambda g(x - x_2)\sigma(x, x_2)(1 - g(x - x_0)\sigma(x, x_0))(1 - g(x - x_1)\sigma(x, x_2)) \leq \lambda g(x - x_2)$. The points of this process are all connected to x_2 but not to x_1 and not to x_0. Now, iterate over all nodes at the same level and repeat. So inductively, iterating over all levels and over all points at the considered level, we realize the Poisson point process from some point x_s whose intensity is $g(x - x_s)(1 - g(x - x_{s-1})\sigma(x, x_{s-1}))$ times the intensity which corresponds to the Poisson process generated from point x_{s-1}.

Denote by G the resulting graph and consider any two points $x_u, x_v \in G$. From the previous construction if x_u and x_v are connected to the origin x_0 in the weighted random connection model, then x_u and x_v are connected to x_0 in G as well. Also notice that for any x_s in the construction, the size of its progeny is stochastically bounded above by the size of a progeny of a tree with the expectation $\lambda \int_{\mathbb{R}^2} g(x - x_s)dx = \lambda \int_{\mathbb{R}^2} g(x)dx = I$. Given $I < \infty$ there exists λ sufficiently small such that $I \leq 1$. For such λ the probability of extinction of the constructed tree is one. Hence we conclude the first part of the proof that there exists sufficiently small $\lambda > 0$ for which the probability that x_0 belongs to an infinite component is zero, that is, $\theta(\lambda) = 0$.

Part 2. For the second part, tessellate the plane $\mathbb{R}^2$ into squares of length $\delta > 0$, where δ is to be determined later. Let the tessellation be such that x_0 is the crossing of diagonals of one of the squares. We call two squares neighboring if they share a side. The maximal distance between any two points that belong to the same square or two neighboring squares is $\delta\sqrt{5}$. Hence the probability that these two points are connected is $g(x_i, x_j)\sigma(z_i, z_j)$, which is at least $g((2\delta, \delta))\sigma(z_i, z_j)$. The combined risk function $\sigma(z, z) \in [0, 1]$ is increasing and $\mathbb{P}(Z > z)$ is decreasing. Let z^* be the value at which the product $\mathbb{P}(Z > z)\sigma(z, z)$ is maximized

$$z^* = \text{argmax}_{z \geq 0}\mathbb{P}(Z > z)\sigma(z, z). \tag{3.3}$$

Note that z^* is finite since Z does not contain (by definition) a point mass at infinity, i.e., $\lim_{z^* \uparrow \infty} \mathbb{P}(Z > z^*) = 0$, and the combined risk function is bounded $0 \leq \sigma(z, z) \leq 1$ for every z. Moreover $\mathbb{P}(Z \geq z^*) > 0$, otherwise the maximum of the right hand side of (3.3) would be equal to 0.

We now introduce a new function $\alpha(\delta)$, which depends on the side of a square in the tessellation

$$\alpha(\delta) := 1 - \mathbb{P}(Z \geq z^*)\sigma(z^*, z^*)g((2\delta, \delta)). \tag{3.4}$$

The function $\alpha(\delta)$ is related to the probability that two points at the distance at most $\delta\sqrt{5}$, both of which have risk factors at least z^*, are connected. Consider the probability that a square contains at least $k \geq 1$ points from $\mathcal{X}$, for some positive integer k to be discussed and optimized later. By the Poisson tail bound [9], the number of points in any square of the tessellation satisfies

$$\mathbb{P}\left(\text{Poiss}(\lambda\delta^2) \geq k\right) \geq 1 - \exp\left(-\lambda\delta^2 H(k/\lambda\delta^2)\right), \tag{3.5}$$

where $H(x) = x \log x - x + 1$ for $x > 0$ and $H(0) = 0$. Hence for any given $\delta, \varepsilon > 0$, given $\lambda > 0$, there exists sufficiently large k, such that the tail bound (3.5) is sufficiently close to one, i.e,

$$\exp\left(-\lambda\delta^2 H(k/\lambda\delta^2)\right) \leq \varepsilon. \tag{3.6}$$

Hence conditioning on x_0 at the origin, where Z_0 is its risk factor, the probability that x_0 belongs to an infinite component satisfies

$$\mathbb{P}(x_0 \text{ belongs to infinite component}) \geq \mathbb{P}(Z_0 \geq z^*)\mathbb{P}(x_0 \text{ belongs to infinite component} : Z_0 \geq z^*). \tag{3.7}$$

Since $\mathbb{P}(Z_0 \geq z^*)$ is positive, the probability $\mathbb{P}(x_0$ belongs to infinite component) will be strictly positive if $\mathbb{P}(x_0$ belongs to infinite component : $Z_0 \geq z^*)$ is positive.

To conclude the proof, we will reduce the problem to site percolation in $\mathbb{Z}^2$, whose threshold for the existence of the infinite component is $p_c \approx 0.592746$, see [7]. We now consider the discrete time process that starts at time $\tau = 0$ from the square containing the origin x_0, and successively visits new neighboring squares in the breadth first search manner [2]. Over time $\tau = 0, 1, 2, \ldots$, the status of each square is characterized by two random variables, which respectively take the values from the following two sets: (i) $\{visited, unvisited\}$, and (ii) $\{undetermined, open, closed\}$.

At the beginning $\tau = 0$ label all squares as undetermined and unvisited. We start from the origin x_0 conditioning on the event $Z_0 \geq z^*$. Denote by S_0 the square containing x_0 and label S_0 as open. Any of the four squares, adjacent to S_0, label as open if it contains at least k points in $\mathcal{X}$ adjacent to x_0 such that $Z \geq z^*$, otherwise label as closed. The probability that a square adjacent to S_0 is open, given that S_0 is open, is at least $(1 - \varepsilon(k))(1 - \alpha^k)$. Moreover those four probabilities are independent, since positions of the points in $\mathcal{X}$ and their risk factors are both i.i.d., as well as mutually independent. Now label S_0 as visited. Inductively, for any open and unvisited square S_i, consider its undetermined adjacent squares S_j. Probability that S_j is open, given that S_i is open, is at least $(1 - \varepsilon(k))(1 - \alpha^k)$, and these probabilities are independent by the same argument. Label S_i as visited and iterate the process.

This process is equivalent to site percolation in $\mathbb{Z}^2$, and creates a tree, rooted at x_0, which is a subgraph of the component in the original graph containing x_0. Thus the existence of an infinite component in site percolation in $\mathbb{Z}^2$ implies the existence of an infinite component in the weighted random connection model. This is satisfied if

$$(1 - \varepsilon(k))(1 - \alpha^k) \geq p_c. \tag{3.8}$$

We now analyze (3.8) in detail. First, it follows $\varepsilon(k) \leq 1 - p_c/(1 - \alpha^k)$. Second, given $k = k_0$ and chosen $\varepsilon_0 = \varepsilon(k_0)$ such that (3.8) holds, from (3.6) it suffices

$$\lambda \delta^2 H(k_0/\lambda\delta^2) \geq \log(1/\varepsilon_0). \tag{3.9}$$

Consider $J(y) := y^{-1}H(y) = \log y - 1 + y^{-1}$ for $(0, +\infty)$, where $J(0) := +\infty$ and $J(+\infty) := +\infty$. $J(y)$ is strictly decreasing on $[0, 1]$ and strictly increasing on $[1, \infty]$. So define J_L to be the restriction of $J(y)$ on $[0, 1]$, and denote by J_L^{-1} its inverse. Similarly, define J_R to be the restriction of $J(y)$ on $[1, +\infty]$, and denote by J_R^{-1} its inverse. Letting $y := k/\lambda\delta^2$, (3.9) becomes

$$J(y) \geq k_0^{-1} \log(1/\varepsilon_0). \tag{3.10}$$

The solution of (3.10) in $[0, 1]$ satisfies $0 < y \leq J_L\left(k_0^{-1} \log(1/\varepsilon_0)\right) \leq 1$ or $y \geq J_R\left(k_0^{-1} \log(1/\varepsilon_0)\right) \geq 1$. The lower bound on λ is given by the solution for $y \in (0, 1)$ and yields

$$\lambda \geq \frac{k_0}{\delta^2 J_L^{-1}(k_0^{-1} \log(1/\varepsilon_0))}. \tag{3.11}$$

All the previous conditions can be combined into an optimization problem. That is, we finally obtain that an infinite component exists with positive probability when the density λ of the Poisson point process $\mathcal{X}$ is at least $\lambda \geq \lambda_0$, where λ_0 is obtained by the following

$$\lambda_0 = \inf_{\delta \in \mathbb{R}^+, k \in \mathbb{N}} \frac{k}{\delta^2 J_L^{-1}\left(k^{-1} \log \frac{1 - \alpha(\delta)^k}{1 - \alpha(\delta)^k - p_c}\right)}$$

$$\text{subject to} \quad k \geq \frac{\log(1 - p_c)}{\log \alpha(\delta)} \tag{3.12}$$

$$\alpha(\delta) = 1 - g((2\delta, \delta)) \max_{z \geq 0} \mathbb{P}(Z \geq z)\sigma(z, z).$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The following example illustrates the previous method.

Example 3.2. Let $Z_i \sim \exp(1)$ be i.i.d. exponential random variables for $i = 0, 1, 2, \ldots$, and the random connection function $g(x) = e^{-|x|}$ for $x \in \mathbb{R}$. For any $z_i, z_j \in \mathbb{R}^+$ define the combined risk function to be $\sigma(z_i, z_j) = \exp(-1/z_i - 1/z_j)$, and $\sigma(z_i, 0) = \sigma(0, z_j) = \sigma(0, 0) = 0$. First, we maximize $\mathbb{P}(Z \geq z)\sigma(z, z)$,

$$\max_{z \geq 0} \mathbb{P}(Z \geq z)\sigma(z, z) = \max_{z \geq 0} \exp(-z - 2/z) = \exp(-2\sqrt{2}) \approx 0.059106. \tag{3.13}$$

Second, the function $\alpha(\delta)$ is given by

$$\alpha(\delta) = 1 - \exp(-\delta\sqrt{5} - 2\sqrt{2}). \tag{3.14}$$

From (3.12), λ_0 is given by

$$\lambda_0 = \inf_{\delta \in \mathbb{R}^+, k \geq k_0(\delta)} \frac{k}{\delta^2 J_L^{-1}\left(k^{-1} \log \frac{1 - \left(1 - \exp\left(-\delta\sqrt{5} - 2\sqrt{2}\right)\right)^k}{1 - p_c - \left(1 - \exp\left(-\delta\sqrt{5} - 2\sqrt{2}\right)\right)^k}\right)} \tag{3.15}$$

where

$$k_0(\delta) = \frac{\log(1 - p_c)}{\log\left(1 - \exp(-2\sqrt{2} - \delta\sqrt{5})\right)}. \tag{3.16}$$

The minimal value for k is given by k_0, which grows 'fast', see Figure 1. In order to find the global
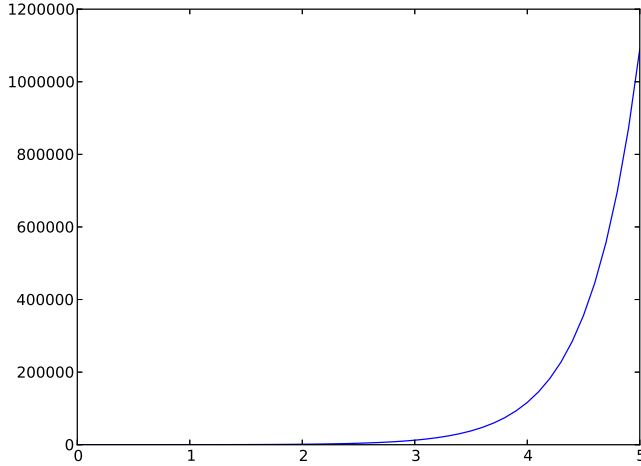

FIGURE 1. $k_0(\delta)$ as the function of δ.

minimum of (3.12), we need to analyze the function

$$V_\delta(k) = \frac{k}{\delta^2 J_L^{-1}\left(k^{-1} \log \frac{1 - \alpha(\delta)^k}{1 - \alpha(\delta)^k - p_c}\right)}. \tag{3.17}$$

Given δ and having k_0 from (3.16), the behavior of $V_\delta(k)$ for $k \geq k_0$, see (3.17), is neither monotonic nor analytically determinative. Hence we numerically estimate (3.17) for: (i) $\delta = 1.5, 1.6, \ldots, 3.5$ in Figure 2 (left), and (ii) $\delta = 1.5, 1.6, \ldots, 2.5$ in Figure 2 (right). Note the difference in the range of k in Figure 2 (left) and Figure 2 (right). Numerically, we obtain the global minimum of $V_\delta(k)$ to be ≈ 175. This is attained at $\lambda \approx 2$ and $k \approx 150$. The provided bound is not expected to be tight.
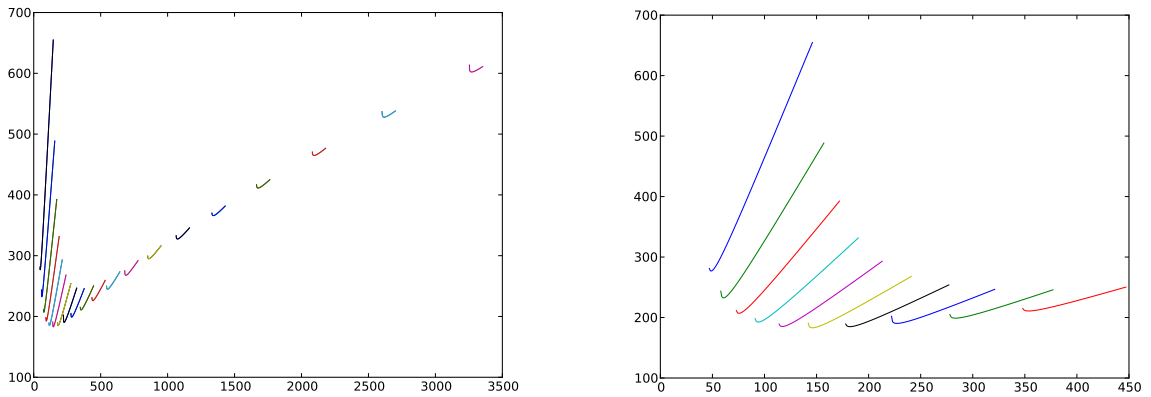
FIGURE 2. $V_\delta(k)$ as the function of k, for $k \geq k_0$ where $\delta = 1.5, 1.6, \ldots, 3.5$ (left) and $\delta = 1.5, 1.6, \ldots, 2.5$ (right).

References

[1] L.J.S. Allen, C.T. Bauch, C. Castillo-Chavez, D. Earn, Z. Feng, M.A. Lewis, J. Li, M. Martcheva, M. Nuño, J. Watmough, M.J. Wonham. *Mathematical Epidemiology (Lecture Notes in Mathematics / Mathematical Biosciences Subseries)*. F. Brauer, P. van den Driessche, J. Wu eds., Springer, 2008.

[2] N. Alon, J. Spencer. *The probabilistic method*. John Wiley & Sons Inc., 2000.

[3] H. Andersson, T. Britton. *Stochastic epidemic models and their statistical analysis*. Springer Lecture Notes in Statistics, 151. Springer-Verlag, New York, 2000.

[4] L. Arriola, M. Hyman. *Being sensitive to uncertainty*. Computing in Science and Engineering, 9 (2007), No. 2, 10–20.

[5] H.T. Banks, M. Davidian, J.R. Samuels Jr., K.L. Sutton. *An inverse problem statistical methodology summary*. Mathematical and statistical estimation approaches in epidemiology. G. Chowell, J.M. Hyman, L.M.A. Bettencourt, C. Castillo-Chavez eds., Springer, (2009), 249–302.

[6] M. Franceschetti, R. Meester. *Random Networks for Communication: From Statistical Physics to Information Systems*. Cambridge University Press, 2007.

[7] G. Grimmett. *Percolation*. Springer Verlag, 1999.

[8] R. Meester, R. Roy. *Continuum Percolation*. Cambridge University Press, 1996.

[9] M.D. Penrose. *Random Geometric Graphs*. Oxford. University Press, Oxford, 2003.