

2016

# Modeling epidemics on a regular tree graph

Claire Seibold

Hannah L. Callender

*University of Portland*, [highland@up.edu](mailto:highland@up.edu)

Follow this and additional works at: [https://pilot scholars.up.edu/mth\\_facpubs](https://pilot scholars.up.edu/mth_facpubs)



Part of the [Mathematics Commons](#)

---

## Citation: Pilot Scholars Version (Modified MLA Style)

Seibold, Claire and Callender, Hannah L., "Modeling epidemics on a regular tree graph" (2016). *Mathematics Faculty Publications and Presentations*. 12.

[https://pilot scholars.up.edu/mth\\_facpubs/12](https://pilot scholars.up.edu/mth_facpubs/12)

This Journal Article is brought to you for free and open access by the Mathematics at Pilot Scholars. It has been accepted for inclusion in Mathematics Faculty Publications and Presentations by an authorized administrator of Pilot Scholars. For more information, please contact [library@up.edu](mailto:library@up.edu).



# Modeling epidemics on a regular tree graph

Claire Seibold and Hannah L. Callender

Department of Mathematics, University of Portland, Portland, OR, USA

## ABSTRACT

We will first provide a brief introduction to models of disease transmission on so-called contact networks, which can be represented by various structures from the mathematical field of graph theory. These models allow for exploration of stochastic effects and incorporation of more biological detail than the classical compartment-based ordinary differential equation models, which usually assume both homogeneity in the population and uniform mixing. In particular, we use an agent-based modelling platform to compare theoretical predictions from mathematical epidemiology to results obtained from simulations of disease transmission on a regular tree graph. We also demonstrate how this graph reveals connections between network structure and the spread of infectious diseases. Specifically, we discuss results for how certain properties of the tree graph, such as network diameter and density, alter the duration of an outbreak.

## ARTICLE HISTORY

Received 30 November 2015  
Accepted 16 April 2016

## KEYWORDS

Infectious diseases; contact networks; agent-based modelling; regular tree graphs; stochastic simulations

## 1. Introduction

Infectious diseases in populations of hosts pose a significant risk to humans, animals and plants. In turn, the spread of such diseases threatens the economic and ecological health of society. Growing populations increase the likelihood – as well as the consequences – of a major outbreak. The shape of a community, including factors such as distance between individuals and number of connections, greatly influences the outbreaks that occur. The ability to model various scenarios is valuable in helping prepare societies and prevent further harm, as we can develop a better understanding of the disease dynamics and, in turn, explore the most effective control measures, such as quarantine or vaccinations (Just et al., 2015a).

There are several approaches to such models, the most common of which is through compartment-based ordinary differential equation models. A typical compartment-based model includes three possible states: Susceptible, Infectious and Removed. Fittingly, this is called an SIR model and was originally proposed by Kermack and McKendrick (1927). In diseases for which an SIR model is appropriate, we begin with a population of susceptible hosts and introduce an *index case* (the first infectious host). As a pathogen is spread from that initial case to other susceptible hosts, the newly infected hosts move to an infectious

**CONTACT** Hannah L. Callender  [callende@up.edu](mailto:callende@up.edu)

state (and hence the infectious compartment) and lastly to the removed state, at which they remain permanently. For more information about such models, refer to [Allen et al. \(2008\)](#), [Hethcote \(2000\)](#).

*Contact networks* can be used to offer even more realistic results by adding complexity to the possible interactions modelled in a compartment-based model. Often these models are idealized networks representing different facets of a community such as physical proximity or other forms of contact ([Keeling & Eames, 2005](#)). These contact networks, which can be described as mathematical graphs, provide a network of interaction for diseases that satisfy the assumptions of the contact network being imitated. Here, two individuals in the population (two nodes on the graph) are connected by an edge in the graph whenever there is a potential for these two individuals to interact. For example, a grid-shaped graph could represent a crop field, and using this grid contact network, one could model the fungal infections or insect swarms plaguing the crop. Indeed, various types of contact networks have been used throughout the infectious disease modelling literature ([Eubank et al., 2004](#); [Grijalva et al., 2015](#); [Miller & Volz, 2013](#); [Riley, 2007](#)).

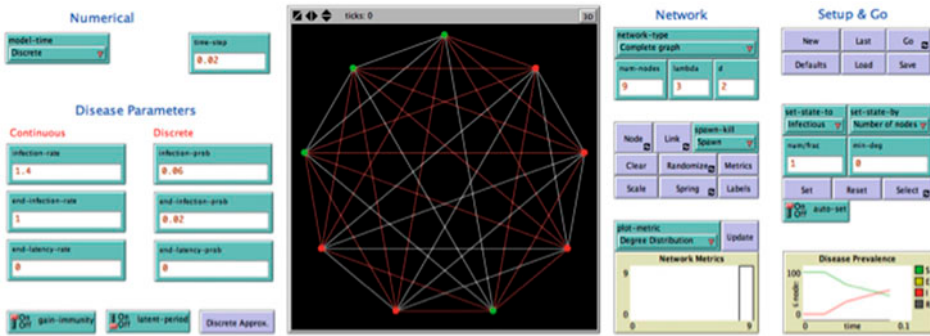
Here, we wish to focus on contact networks that take the form of regular tree graphs. Tree graphs have been used for specific outbreaks such as the SARS cases in Hong Kong, in part because they contain no loops [Riley et al. \(2003\)](#). Although other types of tree graphs, such as random tree graphs, may better represent a wider variety of real world contact networks, our aim is to first develop an understanding of the implications of disease spread on regular tree graphs, which are simpler to analyse [Shapiro & Delgado-Eckert \(2012\)](#), in hopes that these results can inform future study of more realistic network structures.

To further improve upon the traditional SIR compartment-based model, which is often modelled using deterministic differential equations, we investigate the effects of stochasticity in disease dynamics on contact networks through the use of an agent-based modelling approach. Agent-based modelling provides a framework within which we can assign probabilities of infection and recovery to individuals and monitor the effects of such probabilities over time and across multiple simulation runs.

Our primary research question relates to how outbreak duration of a given disease is affected by the size of a given contact network whose structure resembles a regular tree graph. We will elaborate on how connections between network structure and outbreak duration are not straightforward for such a contact network, and we will approach this problem from several different angles for a more complete analysis of relationships between network structure and disease duration. Additionally, we provide various means by which one can calculate the probability of a specific duration, given certain disease parameters and a known size of regular tree graph contact network.

## 2. Model and simulation details

To perform our simulations, we use a software tool called IONTW [Just et al. \(2015b\)](#) which was built from NetLogo [Wilensky \(1999\)](#), an agent-based programming language and modelling environment. IONTW, which stands for Infections On NeTWORKs, allows users to run simulations on various network models of disease transmission. A sample interface is shown in [Figure 1](#). IONTW offers a variety of adjustable parameters such as the type of contact network, initial number of hosts in the susceptible, infectious, or removed states, type of model (SIR, SEIR, SIS, etc.), probabilities of infection and recovery,



**Figure 1.** IONTW interface for modelling a disease whose contact network is such that each node is connected to every other node, known as a complete graph.

and much more. See Just et al. (2015a, 2015b), for further details. The software also has the capability to simulate both discrete and continuous time models. Our work in this paper uses variations of the discrete-time simulation option since we are dealing with probabilities of infection or recovery over the course of a specified time unit instead of rates of infection or recovery.

As mentioned above, IONTW allows the user to simulate the spread of a variety of diseases by allowing for control of values of the infection probability,  $p_{inf}$ , and end infection (recovery) probability,  $p_{rec}$ . In IONTW, these parameters are given the names *infection-prob* and *end-infection-prob*. These are the probabilities that a node will move from the susceptible to infectious state, or infectious to removed state, respectively. Here, we assume homogeneity of hosts, that is, the probabilities of infection are identical for all hosts in the population, as are the probabilities of recovery.

Currently, we also assume a fixed contact network; that is, the probability of effective contact between two nodes is fixed and positive – or 0 if there is no edge – over a given time interval, and the network itself does not change over the desired time course. We refer to units of time as *ticks*, so that the outbreak *duration* can be measured as the number of ticks from the introduction of the index case until the last infectious host has transitioned into the removed state. We elect to keep time units general so that results are applicable to diseases modelled on any time scale. Currently, the IONTW software is unable to simulate diseases where populations interact on multi-time scales.

As mentioned above, we seek to use IONTW to explore disease duration on the rooted regular tree graph, which we will refer to as simply the tree graph. The tree graph is centred on a root node, and contains no closed loops (or cycles). The *regular* tree graph is a tree graph with the same number of branches connected to every non-terminal node. Examples of trees are found in river deltas, subway systems, the cardiovascular system and the namesake plants. Although few cases of regular tree graphs exist in nature, analysis of disease on this type of graph is a first step in providing greater understanding towards other types of contact networks of similar structure, such as other types of tree graphs.

The IONTW application constructs a tree graph using two parameters,  $\lambda$  and  $d$ . The first,  $\lambda$ , is equivalent to the height of the tree, i.e. the largest distance between any node and the *root*, the node from which all other nodes emerge. The parameter  $d$  is the *degree* of the root, the number of nodes connected to the root by an edge. Once  $\lambda$  and  $d$  are selected, a

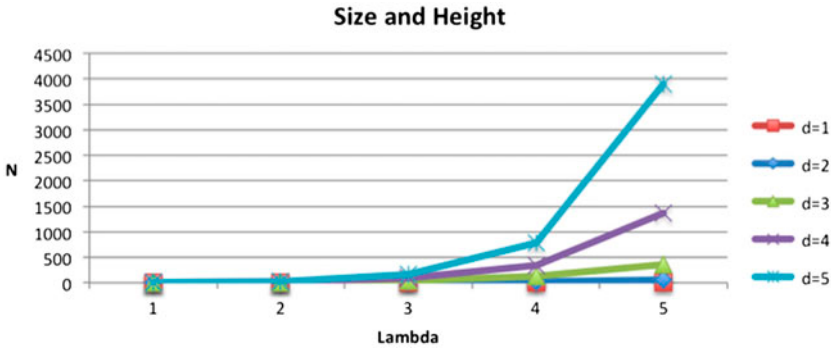


Figure 2. Number of nodes in a regular tree graph generated in IONTW for various values of  $\lambda$  and  $d$ .

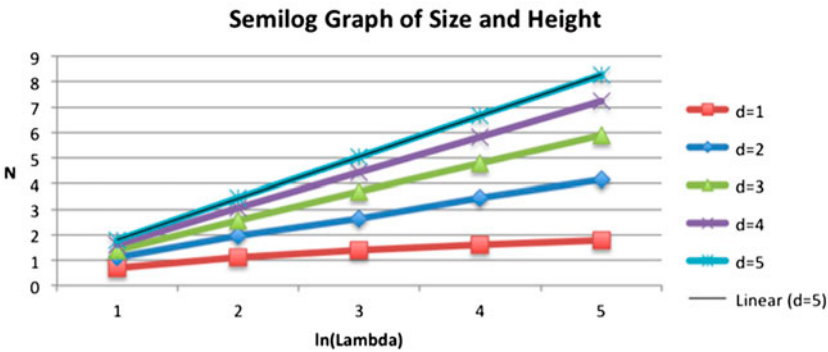


Figure 3. Semi-log plot of number of nodes in a regular tree graph generated in IONTW for various values of  $\lambda$  and  $d$ .

degree of  $d + 1$  is automatically assigned to all other nodes except the *leaves*, which have a degree of 1. A unique property of this type of tree graph is that there are  $n - 1$  edges for the  $n$  nodes; any additional nodes added would create a loop and would break the required properties of tree graphs.

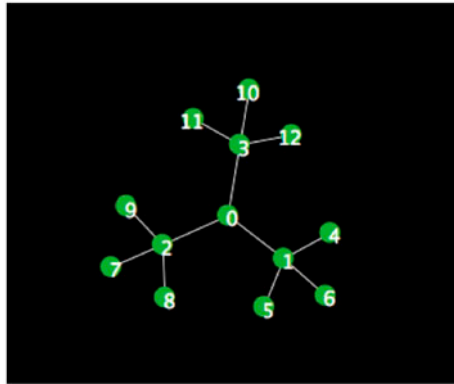
### 3. Size scaling with $\lambda$ and $d$

Since we are interested in analysing the relationship between network size and disease duration, we must note the dependency of tree graph size on the network parameters  $\lambda$  and  $d$ . For fixed values of  $d$ , as  $\lambda$  grows larger, observe how the number of nodes appears to grow exponentially in Figure 2. To see this more clearly, Figure 3 shows a semi-log plot of the same data. We see that the total number of nodes,  $N$ , does indeed exhibit exponential growth as a function of  $\lambda$ , yet different values of  $d$  produce different rates of growth.

We write the number of nodes,  $N$ , as a function of  $\lambda$  and  $d$  as follows:

**Theorem 1:** *Let  $N$  be the number of nodes in a regular tree graph. Then*

$$N = \sum_{i=0}^{\lambda} d^i.$$



**Figure 4.** Regular tree with  $d = 3$  and  $\lambda = 5$ . Node 0 is the root.

**Proof:** Using induction on  $\lambda$ , the base case is  $\lambda = 0$ , where there exists only the root, therefore  $N = 1 = d^0$ .

Induction Step: Assume  $N_k = \sum_{i=0}^k d^i$  for some  $k \geq 1$ . The total includes all nodes a distance of  $k$  away from the root. The number of leaves is  $d^k$  by our assumption. Therefore, the  $k+1$  step will have  $d$  nodes attached to each previous leaf. There are  $d^k * d$  new leaves, or  $d^{k+1}$  nodes added to the previous structure. Hence,  $N_{k+1} = \left(\sum_{i=0}^k d^i\right) + d^{k+1} = \sum_{i=0}^{k+1} d^i$ .  $\square$

#### 4. Diameter scaling with $\lambda$

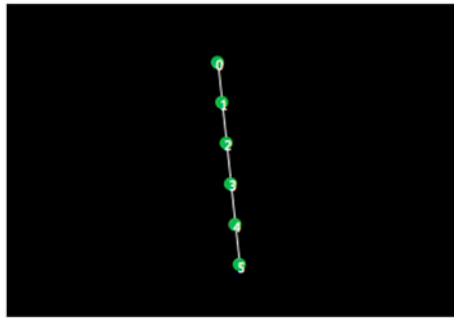
The network *diameter* is defined as the size of the largest of all the shortest paths between every two nodes in a graph. In IONTW, the regular tree graph is constructed such that when  $d > 1$ , the root – labelled node 0 in IONTW – acts as the centre and is exactly half the diameter from all leaves (see Figure 4). In the case when  $d = 1$ , the root is an endpoint on a contiguous line of nodes (see Figure 5). Hence, we can calculate the diameter,  $L$ , of the regular tree from  $\lambda$  as follows:

$$L = \begin{cases} \lambda & : d = 1 \\ 2\lambda & : d > 1 \end{cases} \quad (1)$$

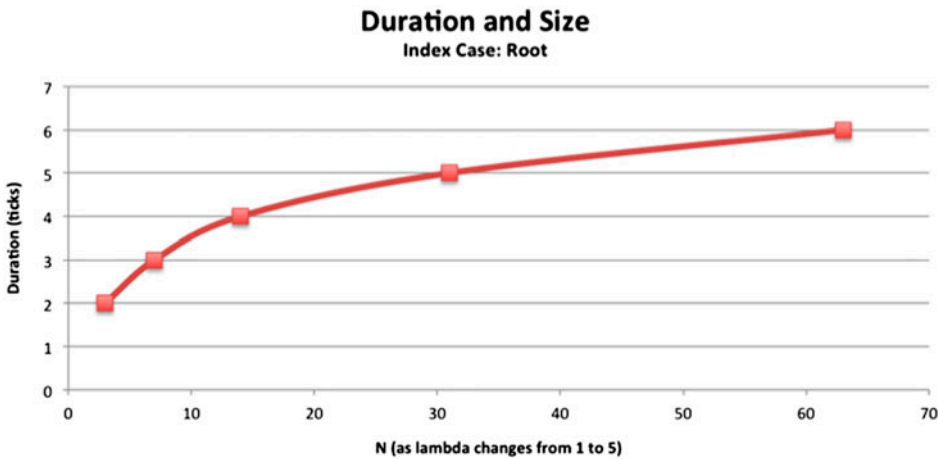
### 5. Determining duration of an outbreak

#### 5.1. Using a next-generation model

The next-generation model in IONTW is created from a simplified discrete-time model measured with time  $\Delta t$ . In this type of model, after exactly one time step, it is assumed that any infectious hosts cease to be infectious and move into the removed category. Hence, since there is a 100% chance of infectious hosts becoming removed at the next time step, the recovery probability,  $p_{rec}$ , is set to 1. Thus the simplest case for determining disease duration in a next generation model would be when the infection probability,  $p_{inf}$ , is also



**Figure 5.** Regular tree with  $d = 1$  and  $\lambda = 5$ . Node 0 is the root.



**Figure 6.** Disease duration in a next-generation model with  $p_{inf} = 1$  on a regular tree graph as a function of population size, as determined by varying values of  $\lambda$  (with corresponding duration) and  $d = 2$ . Note that here we are setting the root to be the index case.

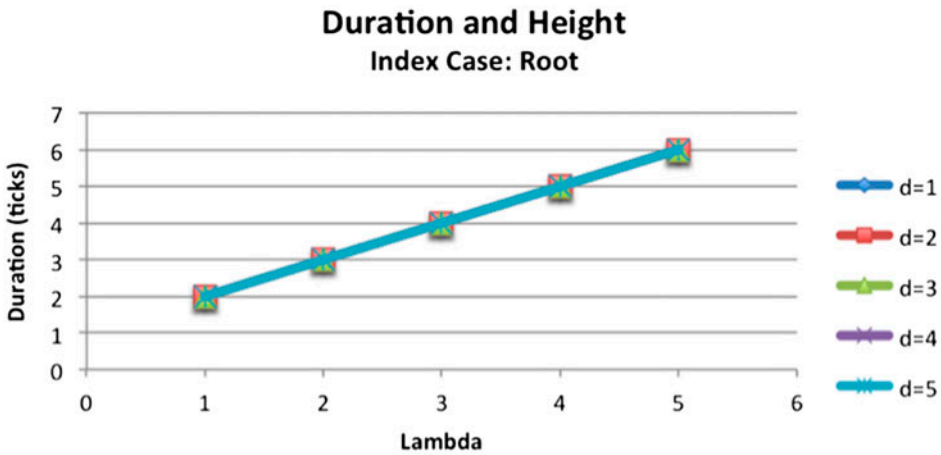
set to 1. This case will be our initial focus, and we discuss additional cases in subsequent sections.

### 5.1.1. Duration scaling with population size

We know  $d$  changes the number of nodes, yet for SIR models on a regular tree graph, it has no effect on duration when the root is the index case. It is interesting, however, to observe the relationship between disease duration and network size (as defined by number of nodes and determined by  $\lambda$  and  $d$ ). Figure 6 displays the duration as a function of size (for  $1 \leq \lambda \leq 5$ ) in the next generation model with  $p_{inf} = 1$ .

### 5.1.2. Duration scaling with $d$

As we just noted, when the root is the index case, the outbreak will have the shortest duration regardless of  $d$ , since the root is on average the shortest distance from all other nodes. In other words, the maximum distance from the root to any given node is  $\lambda$ . This creates a lower bound for an outbreak’s duration:  $\lambda + 1$ . The extra 1 tick accounts for the additional time step required for the last infectious nodes to become removed. Figure



**Figure 7.** Disease duration in a next-generation model with  $p_{inf} = 1$  on a regular tree graph as a function of tree height,  $\lambda$ , for various values of  $d$ . Note that here we are setting the root to be the index case.

7 provides a reference for observing these results with varying tree heights,  $\lambda$ , providing another view of the results depicted in Figure 6. We see that the duration will persist for  $\lambda + 1$  ticks regardless of  $d$ .

If a random node is introduced as the index case, and if  $d > 1$ , the diameter of the graph is  $2\lambda$ . Therefore, in a next-generation model with  $p_{inf} = 1$ , duration is bounded above by  $2\lambda + 1$ . The case when  $d = 1$  is more unique; the graph is created such that the root will be the furthest node from the single leaf. A random index case will on average be close to halfway between the root and the leaf. The bounds of duration,  $\delta$ , are therefore given by

$$\frac{\lambda}{2} \leq \delta \leq \lambda + 1 \quad : d = 1 \quad (2)$$

$$\lambda + 1 \leq \delta \leq 2\lambda + 1 \quad : d > 1 \quad (3)$$

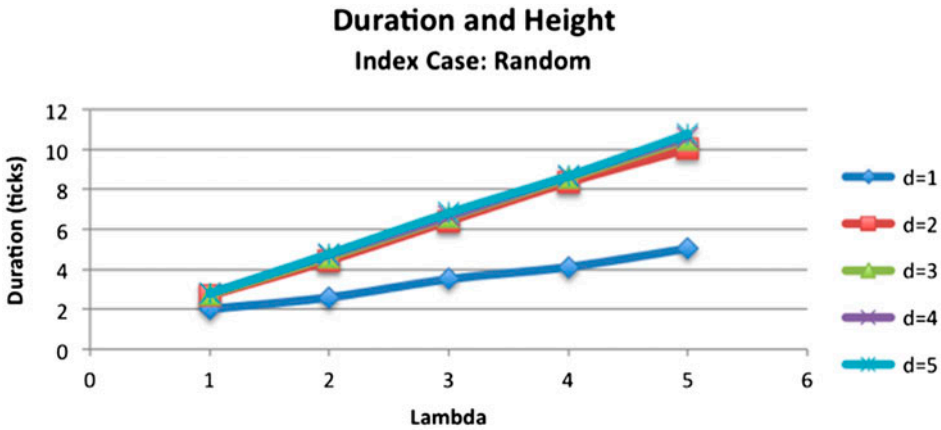
Figure 8 provides a picture of how duration scales with  $\lambda$ , for various values of  $\lambda$  and  $d$ , when the index case is chosen at random.

### 5.1.3. Duration scaling by index case location

As mentioned in subsection 5.1.2, there is a notable difference in duration between instances where the index case is the root and when the index case is a randomly selected node in the body of the tree graph. The duration difference between when an index case is a root or a leaf is depicted in Equation 3. In a next generation model with  $p_{inf} = 1$ , the duration will always equal 1 plus the maximum distance between the index case node and any other node in the network. For example, a leaf, whose maximum distance to any other node is  $2\lambda$ , will produce an outbreak with duration =  $2\lambda + 1$ .

In a model where  $p_{inf} = p_{rec} = 1$  and with a random index case, the mean duration equals the weighted average of the duration for each node multiplied by the number of nodes with that duration. There are various ways to categorize these nodes, one of which is by location. A tree graph can be visualized as a series of levels, the root makes up level 0, adjacent nodes are contained in level 1, and so on, until the leafs make up level  $\lambda$ . Nodes in each level have in common a maximum distance to all other nodes and hence, when





**Figure 8.** Disease duration in a next-generation model with  $p_{inf} = 1$  on a regular tree graph as a function of tree height,  $\lambda$ , for various values of  $d$ . Here the index case is chosen randomly by IONTW. Data points represent averages from 100 simulations.

these nodes are chosen as an index case, they produce the same duration: Maximum distance + 1. The maximum distance is the distance between the current level to the root, plus the height of the tree. Therefore, the mean duration equals a weighted average of the nodes in each level multiplied by the duration they would produce.

Demonstrated below, there are three parts to the sum that make up the mean duration,  $\delta_\lambda$ , for random index cases. The first term seen in the equations below is the total number of nodes (refer to Theorem 1). The second sum multiplies the number of nodes at each level ( $d^j$ ) by the duration for outbreaks occurring at that level. In the case where  $d = 1$ ,  $\lambda$  is equal to the diameter of the graph, and the duration is one more than the maximum distance possible, either between the leaf and  $j$  ( $\lambda - j$ ) or the root and  $j$  ( $j$ ). All other cases of  $d > 1$  have a more predictable distance of  $\lambda + j$ .

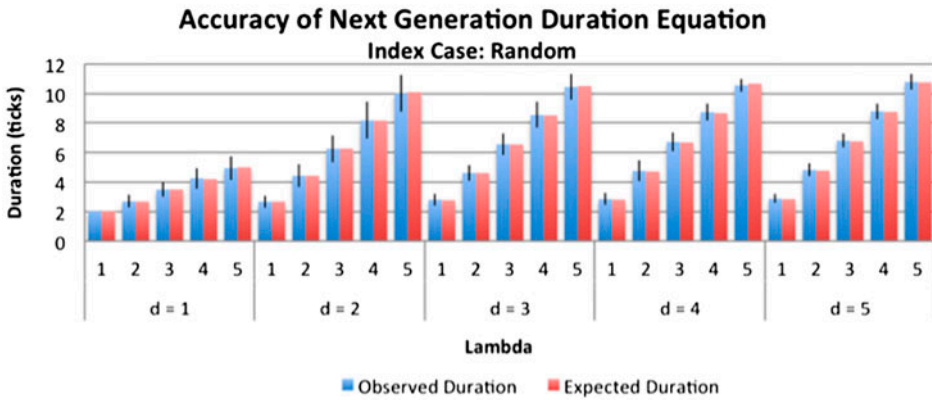
$$\delta_\lambda = \begin{cases} \frac{1}{\sum_{i=0}^{\lambda} d^i} \sum_{j=0}^{\lambda} ((d^j)(\text{Max}(\lambda - j, j) + 1)) & : d = 1 \\ \frac{1}{\sum_{i=0}^{\lambda} d^i} \sum_{j=0}^{\lambda} ((d^j)(\lambda + j + 1)) & : d > 1 \end{cases} \quad (4)$$

Results of expected and observed durations are shown in Figure 9. Each data point for the observed results is the mean duration of 1000 trials. For all data, the deviation between expected and observed is less than 2%, and 84% of results have a deviation smaller than 1%.

**5.2. Using a general discrete-time model with a random index case**

In general discrete-time models, outbreaks are specified in terms of varying transition probabilities ( $p_{inf}$  and  $p_{rec}$ ) and size of time step  $\Delta t$ . Here, we use a unitless time step equal to 1 tick. These models are similar to next-generation models except that  $p_{inf}$  and  $p_{rec}$  can take values between 0 and 1. For instance,  $p_{inf} = 0.5$  means that adjacent node(s) to an

Downloaded by [University of Portland] at 10:49 13 November 2017



**Figure 9.** Expected and observed mean duration in a next-generation model with  $p_{inf} = 1$  on a regular tree graph as a function of tree height,  $\lambda$ , for various values of  $d$ . Note that we are setting the index case to be random and calculating the observed duration upon the possible index case locations. Error bars show one standard deviation above and below the mean duration computed from 1000 runs.

infectious node will have a 50% probability of becoming infectious at the next time step. When  $p_{inf}$  decreases, some hosts may escape infection; likewise, as  $p_{rec}$  decreases, hosts will, on average, stay infectious for a longer time.

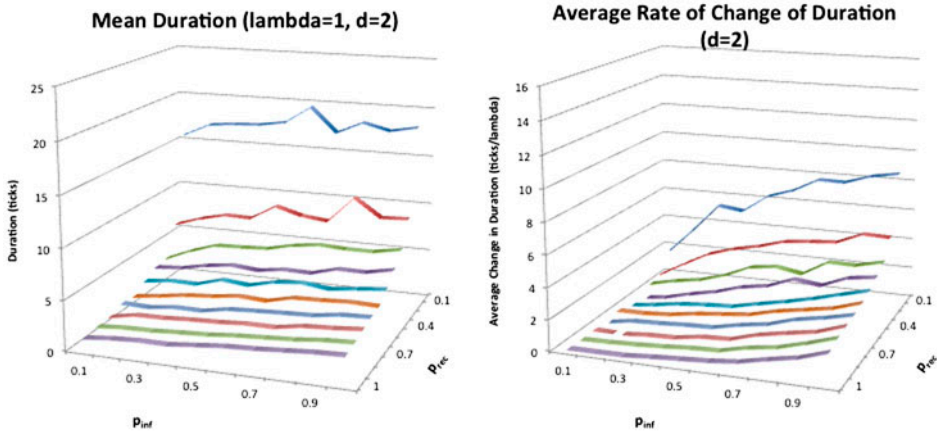
When  $p_{inf}$  is closer to 0 than 1, the likelihood of adjacent nodes acquiring infection is lower, leading to a smaller *final size*, defined to be the total number of hosts that experience infection throughout the time course of the disease; in turn, this leads to fewer opportunities for the infection to stay active. Hence the average duration will approach 0 as  $p_{inf}$  approaches 0. A higher  $p_{rec}$  value will also reduce the average duration by removing infectious nodes sooner.

When  $p_{inf}$  is closer to 1 than 0, nodes are more likely to become infected from contact. With a greater proportion of nodes infectious, the duration is largely affected by  $p_{rec}$ . For  $p_{rec}$  close to 1, the duration approaches  $2\lambda + 1$ , the value found in the next generation model. At low  $p_{rec}$ , nodes will stay infectious for a very long time, therefore duration values increase much faster. These results are shown in Figures 10, 11, and 12 using a randomly assigned index node. Notice the leftmost side of each figure. When  $p_{inf}$  is close to 0 (in these trials the lowest value was 0.1), the mean duration is closer to 0. Larger values of  $p_{inf}$ , combined with a high  $p_{rec}$  value result in a duration close to  $\lambda + 1$ . For all values of  $p_{inf}$ , duration grows rapidly as  $p_{rec}$  decreases.

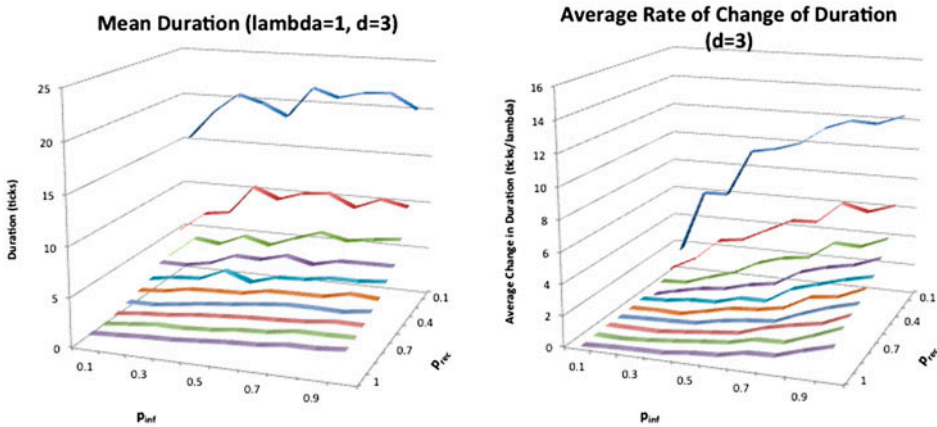
## 6. Calculating probabilities of duration when $p_{inf} \neq 1$

The infection probability,  $p_{inf}$ , and the end-infection (or recovery) probability,  $p_{rec}$ , greatly influence not only average duration but also its variability. However, if we know these probabilities for a given disease, there is a systematic approach to calculating the chances that the duration will last any given number of ticks. The average duration will be a weighted average of all the possible durations.

Let us consider an outbreak with the root as an index case. Initially, we shall hold  $p_{rec} = 1$  and  $d = 2$ . The probability of one node getting infected is equivalent to the infection probability,  $p_{inf}$ . If there are two nodes connected to the infectious root node,



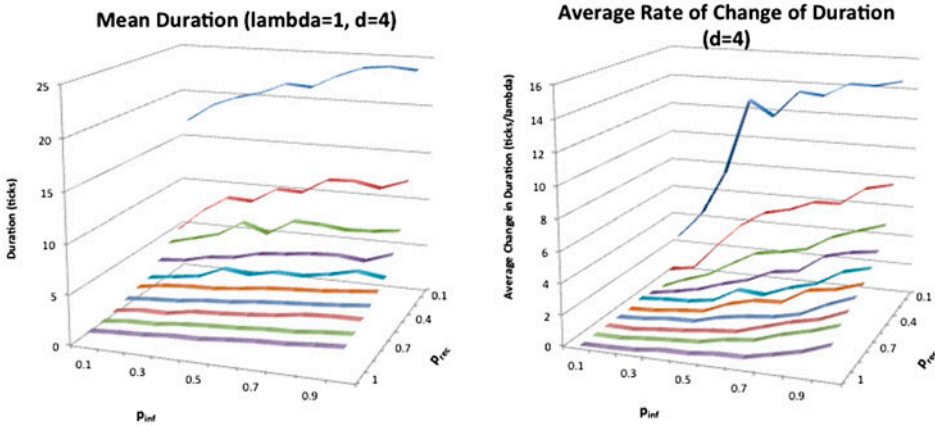
**Figure 10.** Results for duration as a function of  $p_{inf}$  and  $p_{rec}$  on a regular tree graph with  $d = 2$ . An initial case with  $\lambda = 1$  is shown, as well as the average change as a network increases in height up until  $\lambda = 5$ . Trials were conducted with a random index case. Each data point represents the average over 100 simulation runs.



**Figure 11.** Results for duration as a function of  $p_{inf}$  and  $p_{rec}$  on a regular tree graph with  $d = 3$ . An initial case with  $\lambda = 1$  is shown, as well as the average change as a network increases in height up until  $\lambda = 5$ . Trials were conducted with a random index case. Each data point represents the average over 100 simulation runs.

there are three possible options for infection at the next time step, one of which can happen in two different ways: 1) neither node becomes infected with probability  $(1 - p_{inf})^2$ ; 2) one of the two nodes becomes infected with probability  $2(p_{inf} * (1 - p_{inf}))$ ; 3) both nodes become infected with probability  $p_{inf}^2$ .

Each node will either contribute to the spreading infection with probability  $p_{inf}$  or not contribute with probability  $(1 - p_{inf})$ . Since we are adding up all the possible events, we will account for all possible combinations of infectious nodes. For instance, the root has  $d$  adjacent nodes so it can infect  $\binom{d}{0}$ ,  $\binom{d}{1}$ ,  $\binom{d}{2}$ , ..., or  $\binom{d}{d}$  nodes. For any  $0 \leq r \leq d$ , there is probability  $p_{inf}^r$  that  $r$  nodes are infected at the next time point. Additionally, there is



**Figure 12.** Results for duration as a function of  $p_{inf}$  and  $p_{rec}$  on a regular tree graph with  $d = 4$ . An initial case with  $\lambda = 1$  is shown, as well as the average change as a network increases in height up until  $\lambda = 5$ . Trials were conducted with a random index case. Each data point represents the average over 100 simulation runs.

a probability  $(1 - p_{inf})^{d-r}$  that all other adjacent nodes are not infected at the next time point.

For nodes at different distances,  $i$ , from the root, we look at how many nodes there are within that corresponding distance (i.e.  $\sum_{j=0}^i d^j$ ) and then count the ways they could be infected (using the above method of combinations). The number of infectious nodes at distance  $i$  will influence the number of nodes that could be infected at distance  $i + 1$ . For instance, if two nodes are infected by the index case, there are  $2d$  nodes which could possibly become infectious at the next time point. The probability that the duration is equal to 3 is equivalent to the probability of the outbreak reaching nodes two edges away from the index and not continuing any further. That is calculated as the probability that the outbreak ends at the nodes at a distance of 3, multiplied by the double sum of the probability that nodes at a distance of 2 are infected, respectively, by nodes adjacent to the index case.

When  $p_{rec} = 1$ , outbreak duration will stop one time step after the moment when no additional nodes are infected. The probability of a duration of  $\delta$  is equal to the sum of the probability that some node—or nodes—a distance of  $\delta - 1$  from the index case are infected and the probability that none of the nodes at a distance  $\delta$  from the index case are infected. Therefore, consider  $k$  infected nodes and  $d * k$  adjacent susceptible nodes.  $(1 - p_{inf})^{d*k}$  is the probability that none of the adjacent nodes will become infected.

With the previous work, we can continue to multiply our sums for each probability: Let  $P_i$  denote the probability that the duration is  $i$  ticks. We can use the fact that all probabilities must add to 1 to find the  $P_{\lambda+1}$  term. Remember  $\lambda$  represents the maximum distance a node could exist from the root.

From above, we know that when  $p_{rec} = 1$  and the index case is the root, the maximum duration is  $\lambda + 1$ . The expected average duration,  $\delta_\lambda$ , of a graph with height  $\lambda$  and a root

index case is a weighted average of each of the possible durations:

$$\delta_\lambda = 1 * P_1 + 2 * P_2 + \dots + (\lambda + 1) * (P_{\lambda+1}) = \sum_{i=1}^{\lambda+1} i * P_i \quad (5)$$

Using the reasoning provided above, each of the respective probabilities is calculated as follows:

$$P_1 = (1 - p_{inf})^d \quad (6)$$

$$P_2 = \sum_{r=1}^d \binom{d}{r} p_{inf}^r (1 - p_{inf})^{d-r} (1 - p_{inf})^{d*r} \quad (7)$$

$$P_3 = \sum_{r=1}^d \left[ \binom{d}{r} p_{inf}^r (1 - p_{inf})^{d-r} * \left[ \sum_{s=1}^{d*r} \binom{d*r}{s} p_{inf}^s (1 - p_{inf})^{d*r-s} (1 - p_{inf})^{d*s} \right] \right] \quad (8)$$

$$P_4 = \sum_{r=1}^d \left[ \binom{d}{r} p_{inf}^r (1 - p_{inf})^{d-r} * \left[ \sum_{s=1}^{d*r} \binom{d*r}{s} p_{inf}^s (1 - p_{inf})^{d*r-s} * \left[ \sum_{u=1}^{d*r*s} \binom{d*r*s}{u} p_{inf}^u (1 - p_{inf})^{d-u} (1 - p_{inf})^{d*u} \right] \right] \right] \quad (9)$$

⋮

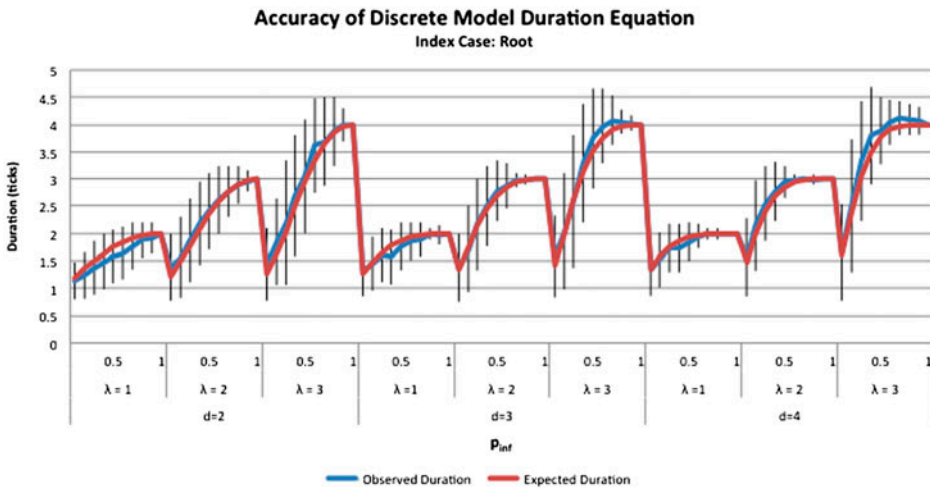
$$P_{\lambda+1} = 1 - (P_1 + P_2 + \dots + P_\lambda) \quad (10)$$

A  $P_i$  term will contain an  $i - 1$  amount of nested sums. The outermost sum refers to the nodes directly adjacent to the root, the innermost sum refers to the nodes that experience infection and are at a distance of  $i - 1$  from the root.

IONTW allows us to easily test these predictions against the simulations. As shown in Figure 13, for various tree degrees,  $d$ , tree heights,  $\lambda$ , and for a wide range of  $p_{inf}$  values, our expected durations are in strong agreement with the observed mean durations.

## 7. Future work

The final size offers a different measure of the severity of an outbreak. The trials conducted in Section 5 can be replicated for random or set index cases to collect data for final size. A final size of 0.3, meaning 30% of the population experienced infection, is the threshold at which point outbreaks are commonly classified as major outbreaks. For each  $\lambda$  and  $d$  there will exist a value of  $p_{inf}$  and  $p_{rec}$  that allow the final size to reach this threshold. Consider how an increasing infection probability leads to larger outbreaks by spreading infection quicker, and a decreasing recovery probability leads to larger outbreaks by keeping infection active for longer. Let the point  $(x_0, y_0)$  be the coordinate pair for the final size to reach this threshold for major outbreak on a given regular tree graph, with  $p_{rec}$  on the  $x$ -axis and  $p_{inf}$  on the  $y$ -axis. Then  $x \leq x_0$  and  $y \geq y_0$  will on average cause a major outbreak. Consequently,  $x > x_0$  and  $y < y_0$  will on average cause a minor outbreak with final size less than 0.3. Future work will include a detailed investigation of how contact



**Figure 13.** Expected and observed mean duration in a discrete model with  $p_{rec} = 1$  and values of  $p_{inf}$  ranging from 0.1 to 1 in increments of 0.1, for several values of  $\lambda$  and  $d$ . Note that we are setting the root to be the index case and calculating an expected duration based upon Equation 5.

network size and structure,  $p_{inf}$ , and  $p_{rec}$  all work together to affect the final size of an outbreak.

In this paper, all studies were conducted with unchanging parameters within each individual trial. In reality, control measures attempt to delay or reduce an outbreak by altering parameters before a major outbreak occurs. For instance,  $p_{inf}$  can be reduced through behaviour modification such as washing hands, and  $p_{rec}$  can be increased by caring for infected individuals. The threshold of values that cause a major outbreak provides the first step towards directing efforts of control measures. Consider a disease whose initial infection probability,  $p_{inf}$ , is very close to  $y_0$ , while the respective recovery probability,  $p_{rec}$ , is much smaller than  $x_0$ . Control measures that are focused on shifting the infection probability below  $y_0$  will prevent major outbreaks much more efficiently than similar attempts to change the probability of recovery.

We have already begun to conduct trials to observe the  $x_0, y_0$  threshold for regular tree graphs. For graphs of various  $\lambda$  and  $d$ , there are patterns emerging as to the influence of the size on the threshold. Initially, we observe that  $d$  has minor effects while  $\lambda$  increases the infection probability necessary for major outbreaks, at a decreasing rate of change. This opens more opportunities not only to predict the threshold for regular tree graphs, but also to predict the threshold on irregular tree graphs.

We would also like to further our investigations on irregular tree graphs, since these graphs are more realistic due to the fact they are able to represent contact networks that are missing the symmetry of a regular tree graph. Irregular tree graphs will contain fewer nodes than a regular tree graph with the same  $\lambda$  and  $d$ , where here  $d$  would denote maximum degree of any node in the tree. With fewer nodes and edges, the spreading infection on an irregular tree graph will likely follow tendencies of an infection spreading on a regular tree graph with smaller  $\lambda$  and  $d$ . The observations of a regular tree graph could be a bound on the maximum duration and final size of an outbreak on an irregular graph. Our initial



efforts will involve a comparison of the differences in number of nodes and resulting outbreak duration and final size between the two types of graphs.

## 8. Conclusion

A regular tree graph provides one type of simplified structure to observe patterns in the movement of outbreaks on a contact network. The outbreaks that fit this network are not necessarily those of pathogens and infectious hosts. One example is the information flow in the military or a large business, where the secretary of defense or a CEO is the root and has connections to several departments who are not directly in contact with each other. Each department has a director who shares the message with the next person in the chain of command. In general, military structure can be represented by a tree graph because personnel only pass orders to those directly below them in their respective unit. An 'outbreak' of an order in the military would respond similarly to an outbreak beginning at the root and following a very high infection probability (because orders are very likely to be followed).

Other examples are seen in networks such as subway systems, water channels, hereditary trees or plant networks. Infectious diseases can be traced through public transportation on land through buses and subway systems, or on water through ships that transport infection in the form of humans, rats or even contaminated cargo. Hereditary (genetic) disorders such as colour blindness, cystic fibrosis or ALS also represent outbreaks of a different variety on tree networks. Plants pass xylem and other molecules through tissue from the roots to all parts above ground, in a pattern similar to a tree network.

Although these scenarios may result in loops, as individuals can be infected multiple times, the tree network creates a simplified situation for observations. The height of the graph is greatly influential in determining the duration of an outbreak, even more so than the number of branches which has minimal effect after the root is infected. Duration can be calculated for next generation models where  $p_{inf} = p_{rec} = 1$ . In more general discrete models, the duration can be calculated when the index case is the root, but becomes an increasingly complicated sum of sums. The distance between the index case and the furthest nodes becomes the most important factor with both scenarios.

## Funding

This work was partially supported by the University of Portland Butine faculty development [grant number (13561-397)] and the University of Portland student development fund.

## References

- Allen, L. J., Brauer, F., Van den Driessche, P., & Wu, J. (2008). *Mathematical epidemiology*. Berlin: Springer.
- 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, 180–184.
- Grijalva, C. G., Goeyvaerts, N., Verastegui, H., Edwards, K. M., Gil, A. I., Lanata, C. F., Hens, N. (2015). A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. *PLoS One*, 10, 9–10.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42, 599–653.

- Just, W., Callender, H. L., & Lamar, D. M. (2015a). Disease transmission dynamics on networks: Network structure vs. disease dynamics. In Robeva, R. (Eds.), *Algebraic and discrete mathematical methods for modern biology*. San Diego, CA: Academic Press. (pp. 217–235)
- Just, W., Callender, H. L., & Lamar, D. M. (Eds.). (2015b). *Exploring Transmission of Infectious Diseases on Networks with NetLogo*. Retrieved [3/3/2016] from <http://www.ohio.edu/people/just/IONTW/>
- Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2, 295–307.
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. In *proceedings of the royal society of london a: Mathematical, physical, and engineering sciences*; 115, p. 700-721. The Royal Society
- Miller, J. C., & Volz, E. M. (2013). Incorporating disease and population structure into models of SIR disease in contact networks. *PloS One*, 8, 1–4.
- Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316, 1298–1301.
- Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Anderson, R. M. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, 300, 1961–1966. Retrieved 3/3/2016, from <http://www.sciencemag.org/content/300/5627/1961.abstract> doi:10.1126/science.1086478
- Shapiro, M., & Delgado-Eckert, E. (2012). Finding the probability of infection in an SIR network is NP-hard. *Mathematical biosciences*, 240, 77–84.
- Wilensky, U. (1999). *Netlogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University: Evanston, IL. Retrieved 3/3/2016, from <http://ccl.northwestern.edu/netlogo/>

## Appendix 1. Next generation model duration equation

Code used to calculate Equation 4 in Python and results shown in Figure 9.

```
#calculate the average duration of a graph with a random index node.
#This only holds for a next gen graph with infection-prob=end-infection-prob=1.
#Enter lambda (height) and d.

height = #enter lambda value
d = #enter d value

avgduration = 0
calc = 0
temp = 0

for i in range(0, height+1):
    calc += d ** i
N = calc

for j in range(0, height+1):
    temp += (height + 1 + j) * (d**j)
avgduration = temp/N
print ('duration =', avgduration)
```



## Appendix 2. Discrete model duration equation

Code used to calculate Equation 5 in Python and results shown in Figure 13.

```
import math

def nCr(n,r): #function to calculate combinations.
    f = math.factorial
    return f(n) / f(r) / f(n-r)

def Prob(d,time,ip): #function to calculate probability that a duration will last 'time'
    if time == 1: #case when duration is 1
        return float(math.pow(1-ip,d))
    elif time == 2: #case when duration is 2
        q = 0
        for r in range(1,d+1):
            b = float(nCr(d, r))*(math.pow(ip, r))*(math.pow((1-ip),(d-r)))
            c = float(math.pow((1-ip),d*r))
            q += float(b*c)
        return q
    elif time == 3: #case when duration is 3
        f = 0
        for s in range(1,d+1):
            ad = float(nCr(d,s))*(math.pow(ip, s))*math.pow((1-ip),(d-s))
            e = 0
            for u in range(1,d*s):
                b = float(nCr(d*s, u))*(math.pow(ip,u))*(math.pow((1-ip), (d*s)-u))
                c = float(math.pow(1-ip, d*u))
                e += float(b*c)
            f += float(ad*e)
        return f

def delta(d,lambd,ip): #function to add up weighted averages of possible durations
    tots = 0
    d = d
    totalprobs = 0
    for i in range(1,lambd+1): #all possible durations when root is index case
        if i == lambd + 1: #the largest duration,
            #P_lambd+1 = 1 - (P_1 + P_2 + ... + P_(i-1))
            tots += i*(1-totalprobs)
        else:
            temp = Prob(d,i,ip) #run Prob function on each possible duration
            totalprobs += temp #add up for calculating largest duration, P_lambd
            tots += i*temp #multiply probability by duration, i, to get weighted average
    tots += (i+1)*(1-totalprobs)
    print (tots)
    return tots

delta(x,y,z) #calculate average duration for d = x, lambda = y, infection-probability = z.
```