



Modelling epidemics on d -cliqued graphs

Laura P. Schaposnik^{a,b} and Anlin Zhang^c

^aUniversity of Illinois at Chicago, Chicago, IL, USA; ^bFreie Universität Berlin, Berlin, Germany; ^cCanyon Crest Academy, San Diego, CA, USA

ABSTRACT

Since social interactions have been shown to lead to symmetric clusters, we propose here that symmetries play a key role in epidemic modelling. Mathematical models on d -ary tree graphs were recently shown to be particularly effective for modelling epidemics in simple networks. To account for symmetric relations, we generalize this to a new type of networks modelled on d -cliqued tree graphs, which are obtained by adding edges to regular d -trees to form d -cliques. This setting gives a more realistic model for epidemic outbreaks originating within a family or classroom and which could reach a population by transmission via children in schools. Specifically, we quantify how an infection starting in a clique (e.g. family) can reach other cliques through the body of the graph (e.g. public places). Moreover, we propose and study the notion of a *safe zone*, a subset that has a negligible probability of infection.

ARTICLE HISTORY

Received 4 October 2017
Accepted 11 December 2017

KEYWORDS

Epidemic dynamics; cliques;
symmetric graphs

1. Introduction

The study of epidemic propagation in networks (social and biological) has been of much interest to biologists and mathematicians for a long time, but only recently have graph theory, number theory and computer science taken researchers to several breakthroughs. Moreover, as highlighted in [Pastor-Satorras and Vespignani \(2001\)](#), the importance of local clustering in networks has been widely recognised, and thus understanding the connectivity of the networks is of utmost importance for developing effective control measures (e.g. quarantine, vaccinations or specific treatments).

The mathematical foundation for epidemiology can be traced to the early 1900s with the work of Ronald Ross: he produced the first mathematical model of mosquito-borne pathogen transmission using mosquito spatial movement in order to reduce malaria from an area [Mandal, Sarkar, and Sinha \(2011\)](#). In the 1920s, Lowell Reed and Wade H. Frost developed the *Reed-Frost model* ([Brauer & Castillo-Chavez, 2001](#)), which improved on Ross's model by modelling how an epidemic behaves with respect to time. In 1927, [Kermack and McKendrick \(1927\)](#) created the *SIR model*, which categorized people into the three states *Susceptible*, *Infectious* and *Removed*. More recently, Matt J. Keeling and

CONTACT Laura P. Schaposnik  schapos@uic.edu, lauraschaposnik@gmail.com

Ken T. D. Eames used contact networks to better represent a community (Keeling & Eames, 2005). The use of contact networks which are adapted to reflect certain particular characteristics of society has been of much use when doing mathematical modelling of epidemics. This modelling is performed by seeing the network as a graph where vertices represent individuals, and edges encode the interactions amongst people: two people, seen as vertices, are connected by an edge in the graph whenever they are related (and thus an interaction could exist).

Particular shapes of contact networks have been studied in recent years, from grid contact network to represent fields and study fungal infections, to networks with regular d -ary tree structure, used to study SARS outbreaks in Hong Kong [e.g. see Riley et al., 2003]. Moreover, in Seibold and Callender (2016), the authors considered the implications of disease spread on perfect d -ary trees. Our work builds on their research, extending their results to more general networks and index cases.

We dedicate this paper to study local clustering in networks and its impact on epidemic modelling by incorporating an important method: the analysis of the *symmetries* which networks have through the appearance of cliques. This perspective is inspired by the appearance of symmetric relations amongst members of subgraphs within a network. An overall question we aim at understanding is *how an epidemic outbreak originating within a family or classroom could reach a population by transmission via children in schools*. This setting would also appear when modelling networks accounting for relationships between siblings in family trees, or when predicting the expected number of infected families. To answer these questions, we introduce a novel type of graphs, the *d -cliqued trees*, to study epidemics in regular networks which contain symmetric clusters both through next generation models as well as through general discrete time models, which allows one to assign probabilities of infection P_{inf} and recovery P_{rec} to track how an outbreak would affect the population modelled via our contact network.

Through d -cliqued trees \mathcal{A}_d^λ of degree d and height λ introduced in Definition 3.3, in what follows we investigate how the outbreak duration of an epidemic originated within a classroom propagates through society depending on the size of the contact network considered (the height λ), as well as on the size of the classroom (the degree d). Since we aim at highlighting the importance of symmetries to approach these realistic settings, we will elaborate on how our probabilities P_i of outbreaks lasting i time units (ticks) differ from those in Seibold and Callender (2016) – see, for example Figure 1.

2. A model on d -ary trees

Symmetries should appear in networks when considering clusters of vertices which share some common relation (for example, the members of a family and their symmetric relations, or cities connecting to a common airport). In 2016, researchers studied epidemic modelling on d -ary trees (Seibold & Callender, 2016) using the IONTW platform: in this paper, we present a new approach to generalize the study to *d -cliqued tree graphs*, given by regular trees with added edges creating cliques, or symmetric and completely connected subnetworks. We dedicate this section to give a brief overview of regular trees and d -ary trees in Section 2.1, and of the main results of Seibold and Callender (2016) in Section 2.2.

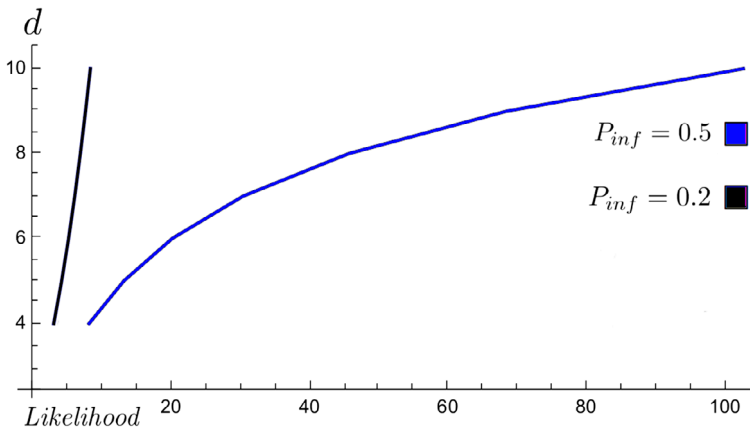


Figure 1. The quotient of the probability of an outbreak lasting two times in our model divided by the same probability for the model in Seibold and Callender (2016), more details in Figure 16.

2.1. Background: regular trees and d -ary trees

In modelling epidemics on simple graphs, vertices denote individuals, and edges the possible ways to transmit an infection from one individual to another. Following the notation of Seibold and Callender (2016), we refer to the *index node* I of the graph as the vertex where the infection originates.

Definition 2.1: A tree is an undirected graph in which any two vertices are connected by exactly one path. A d -regular tree is a tree graph for which all non-terminal vertices have degree d (i.e. have d adjacent vertices), and the terminal ones have degree 1. The degree of a d -regular tree is d .

A *rooted tree* is a tree in which one vertex x_0 has been designated the *root*. In such trees, the *parent* of a vertex v is the vertex connected to it on the path to the root x_0 , and v is called a *child* of the parent vertex. The *height* is the length, in number of edges, from a terminal vertex to the root. We assume that all terminal vertices of the d -regular tree have the same height $\lambda \in \mathbb{N}$ and call this the *height* of the d -regular tree. The *network diameter* L of a d -regular tree is the number of edges in the longest path between two vertices. Regular trees are closely related to d -ary trees (Figure 2).

Definition 2.2: A d -ary tree is a rooted tree in which each node has at most d children.¹ A perfect d -ary tree is one where non-terminal nodes have d children, and terminal nodes have none.

It is known that a d -regular tree on n vertices exists if and only if $d - 1$ divides $n - 2$. Finally, for any graph G , we denote by $d(v, w)$ the length of the shortest path between two vertices $v, w \in G$, which will become useful in later sections (e.g. for Lemma 3.6).

2.2. Outbreak modelling through the IONTW platform

The Infections On NeTWORKs platform (IONTW) (Just, Callender, & Lamar, 2015; Wilensky, 1999) is an agent-based modelling platform that shows theoretical predictions for disease spread and simulates disease transmission on graphs (see Figure 3). In Seibold and Callender (2016), the authors studied how outbreak duration of a given disease is affected

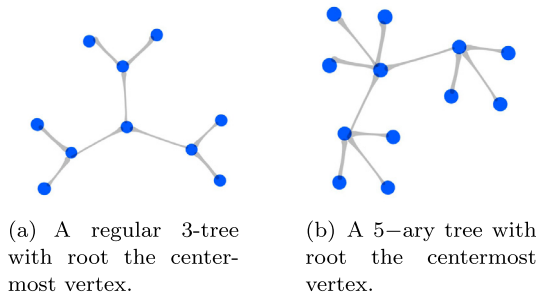


Figure 2. Comparison of two types of trees. Figure (a) gives a network studied in Seibold and Callender (2016).

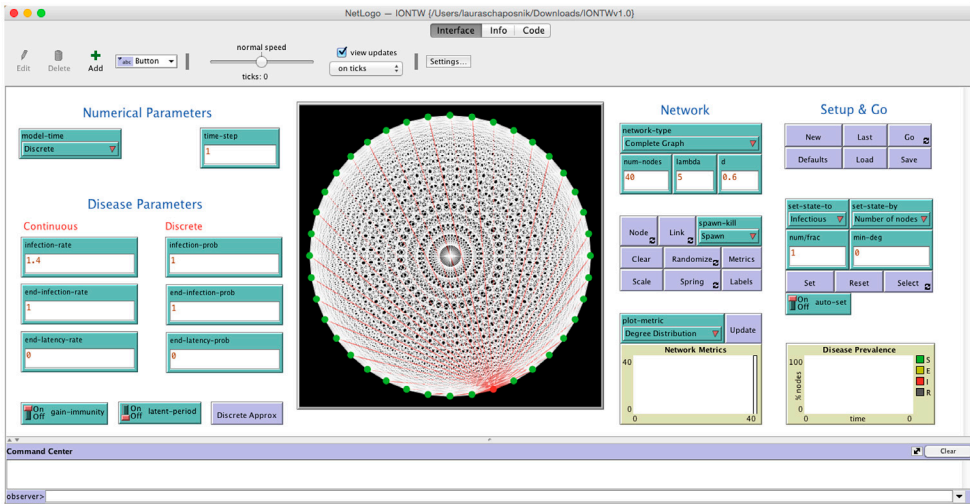


Figure 3. Screenshot of the IONTW platform, which can be used to model epidemic outbreaks for certain types of graphs. Note: In the figure, a complete graph in 40 vertices.

by the size of a given contact network whose structure resembles perfect d -ary trees via IONTW. It is important to note that whilst the authors referred to the networks as ‘regular tree graphs’, the notion of regular graph (where every vertex has the same incidence degree) was not considered. Hence, the correct terminology for their networks would seem to be indeed the one of ‘perfect d -ary trees’.

In Seibold and Callender (2016), the authors initially assumed homogeneity of hosts, i.e. that the probabilities of infection P_{inf} are identical for all hosts in the population, as are the probabilities of recovery P_{rec} . The *outbreak duration* δ is the number of ticks, or units of time, from the introduction of the infection in the index node until all hosts have recovered. We will denote by δ_λ the expected average outbreak duration, and by P_i the probability the outbreak lasts exactly i ticks.

In a next generation model, one assumes that $P_{rec} = 1$, i.e. infected hosts are moved to the recovered group at some time step, and $P_{inf} = 1$, i.e. that all nodes will eventually get infected. The lower bound for δ_λ on regular tree graphs with $d > 1$ is $\lambda + 1$ when the

origin I is placed in the root, since the maximum distance from there to any node is λ , and we need an additional time step for the last node to recover. Hence, in this case, when the index is placed in the root, by definition of expected value, one has

$$\delta_\lambda := \sum_{i=1}^{\lambda+1} i \cdot P_i.$$

Finally, it was shown in [Seibold and Callender \(2016\)](#) that the probabilities P_i of the duration spanning i time units, when the origin I is the root of a perfect d -ary tree, is given by

$$P_1 = (1 - P_{\text{inf}})^d \tag{1}$$

$$P_2 = \sum_{r=1}^d \binom{d}{r} P_{\text{inf}}^r (1 - P_{\text{inf}})^{d-r} (1 - P_{\text{inf}})^{d-r} \tag{2}$$

⋮

$$P_{\lambda+1} = 1 - (P_1 + P_2 + \dots + P_\lambda). \tag{3}$$

For each P_i term, the outer sum denotes nodes adjacent to the infection’s origin I which is placed in the tree’s root, whilst the inner sum denotes nodes that are infected $i - 1$ edges away from the infection’s origin. Moreover, from [Seibold and Callender \(2016\)](#), the outbreak duration δ is bounded as follows:

$$\lambda + 1 \leq \delta \leq 2\lambda + 1 \text{ when } d > 1, \tag{4}$$

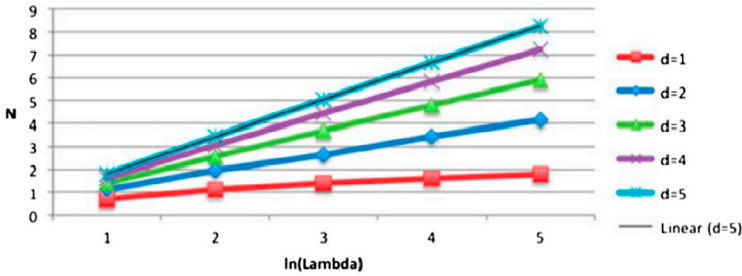
$$\frac{\lambda}{2} \leq \delta \leq \lambda + 1 \text{ when } d = 1. \tag{5}$$

Much of the work in [Seibold and Callender \(2016\)](#) was also dedicated to the study of the change a network went through when the degree d or height λ of the graph varied. In particular, through the IONTW platform, the authors obtained the following descriptions in [Figure 4](#), which should be compared with the ones in [Section 3](#) which we obtain without the use of IONTW.

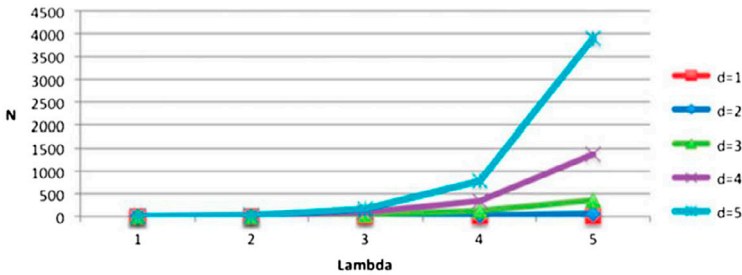
Through the understanding of the probabilities P_i the authors looked into the duration of outbreaks depending on the placement of the index node, as well as on the size of the network, although an explicit expression of P_i was only given when I was taken to be the root of the graph. Through a next-generation model where P_{inf} is set to 1, and when the index case is the root of the graph, the disease duration was described as a function of population size, determined by varying values of the height λ for degree $d = 2$ networks as in [Figure 5\(a\)](#), and for different values of d with the same height λ as in [Figure 5\(b\)](#).

In particular, it was observed that the duration will persist for $\lambda + 1$ ticks regardless of d when the index case is located in the root of the graph ([Seibold & Callender, 2016](#)), and formulas for the mean duration δ_λ of a disease outbreak were provided for such case:

$$\delta_\lambda = \begin{cases} \frac{1}{\sum_{i=0}^{\lambda} d^i} \sum_{j=0}^{\lambda} ((d^j)(\text{Max}(\lambda - j, j) + 1)) & d = 1; \\ \frac{1}{\sum_{i=0}^{\lambda} d^i} \sum_{j=0}^{\lambda} ((d^j)(\lambda + j + 1)) & d > 1. \end{cases} \tag{6}$$



(a) Number of nodes N in a d -ary graph generated in IONTW for various values of λ and d .



(b) Semi-log plot of number of nodes N in a d -ary tree graph generated in IONTW.

Figure 4. Figures from Seibold and Callender (2016) obtained through IONTW on the network’s size variation.

3. Geometry of d -cliequed networks

Our primary research question relates to how one can model epidemic outbreaks in networks that resemble family trees where the relations between siblings are taken into account. We shall begin by introducing these new types of networks, modelled on d -cliequed graphs, in Section 3.1. Since we are interested in understanding how the duration of a given disease is affected by the size of a d -cliequed network, we proceed in Sections 3.2 and 3.3 to study how the scaling of the height λ and the degree d affect the network.

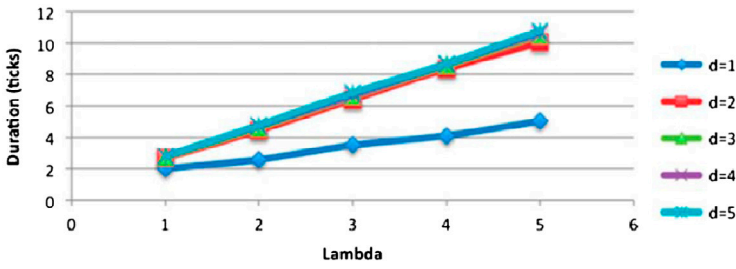
3.1. A new type of network modelled on d -cliequed graphs

Since symmetries in social networks can sometimes be modelled through cliques, in what follows we will extend the work of Seibold and Callender (2016) to a more general setting of regular trees with added cliques.

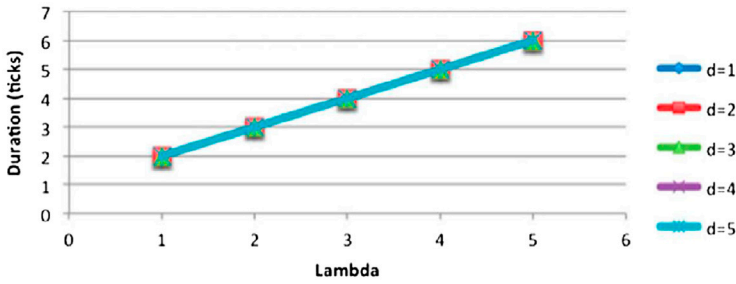
Definition 3.1: A clique is a subset of a graph such that all pairs of vertices in the subset are connected by an edge. A clique with d vertices is called a d -clieque.

Definition 3.2: The clique number $\omega(G)$ of G is the size of the largest clique in G , and a maximum clique is a clique with $\omega(G)$ vertices.

Cliques are especially useful in modelling epidemics, as they can represent completely related social groups (e.g. families or classrooms). We will look at regular tree graphs with



(a) Disease duration in terms of network size for randomly chosen index case using IONTW.



(b) Disease duration in terms of network height.

Figure 5. Figures from Seibold and Callender (2016) obtained through IONTW on disease durations.

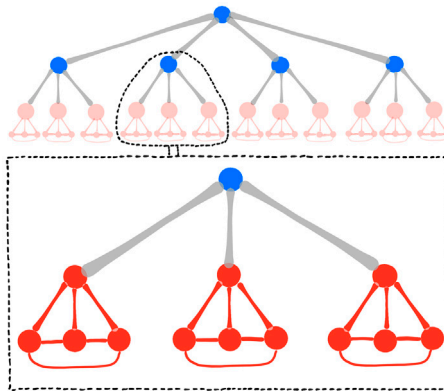


Figure 6. Graph \mathcal{A}_4^3 : a 4-regular tree completed to have terminal 4-cliques (in red), and body vertices (in blue).

added cliques and analyse how duration, P_{inf} and P_{rec} can be expressed in terms of different variables.

Definition 3.3: A d -cliqued tree graph \mathcal{A}_d^λ of height λ is a regular tree graph of degree d and height λ , with added edges to form d -cliques on the terminal vertices. The body of \mathcal{A}_d^λ is the set of vertices which are not in any clique.

By construction, the number of d -cliques is $d(d-1)^{\lambda-2}$, and $\omega(\mathcal{A}_d^\lambda) = d$. The graph \mathcal{A}_d^λ can be separated into levels, where the root makes up level 0, adjacent nodes are contained in level 1, and so on, until the leafs make up level λ , the level where cliques are formed.

3.2. Size scaling with λ and d

Since d -cliqued networks are modelled on modifications of d -graphs to which one has added edges, the size of the network will depend on the height λ and the degree d of the underlying regular tree. In particular, we can show the following.

Proposition 3.4: *The total number of edges of \mathcal{A}_d^λ is*

$$d^2(d-1)^{\lambda-2} \cdot \frac{d-2}{2} + \sum_{i=0}^{\lambda-1} d(d-1)^i. \tag{7}$$

Proof: Since the number of vertices of a d -regular tree of height λ is $1 + \sum_{i=0}^{\lambda-1} d(d-1)^i$, this is also the number of vertices of a d -cliqued tree \mathcal{A}_d^λ . Moreover, the number of edges in a d -regular tree of height λ is $\sum_{i=0}^{\lambda-1} d(d-1)^i$. Hence, since d -cliques have $d(d-1)/2$ edges, the number of edges that one needs to add to a d -regular tree of height λ to obtain the d -cliqued graph \mathcal{A}_d^λ is

$$\underbrace{\left(\frac{d(d-1)}{2} - (d-1) \right)}_{\text{edges missing to make a complete } d\text{-clique}} \cdot \overbrace{d(d-1)^{\lambda-2}}^{\text{number of cliques made}}$$

and thus the total number of edges of \mathcal{A}_d^λ is as in (7). □

3.3. Clique and diameter scaling with λ

In what follows, we shall study the dependence of the size of the network on the height λ and the degree d of the graphs \mathcal{A}_d^λ . Note that from its definition, the number of vertices in the body of \mathcal{A}_d^λ is $1 + \sum_{i=0}^{\lambda-3} d(d-1)^i$. In order to study the cliques of \mathcal{A}_d^λ , note that these have been formed through the outermost vertices, and one will have a clique for every vertex at the $\lambda - 1$ level of the graph. Since the first level has d vertices but subsequent ones increase by a factor of $(d-1)$, the number of cliques in the whole graph is $d(d-1)^{\lambda-2}$. Then, one can understand the growth of the number of cliques in \mathcal{A}_d^λ in terms of λ as shown in Figure 7(a). Moreover, the number of edges in \mathcal{A}_d^λ can also be seen in terms of the degree d and height λ via Equation (7), as illustrated in Figure 7(b).

3.4. Distances between vertices

When studying epidemics in d -cliqued trees, one needs to consider the different positions that the index vertex (origin of the infection) may take. These positions may be classified into two cases:

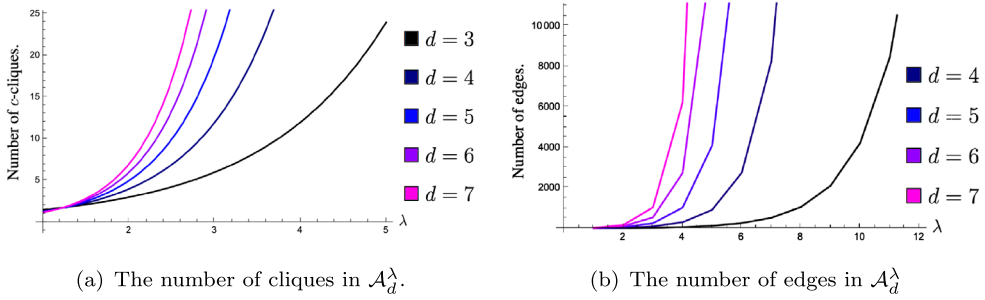


Figure 7. Clique number and edge growth of \mathcal{A}_d^λ in terms of the height λ .

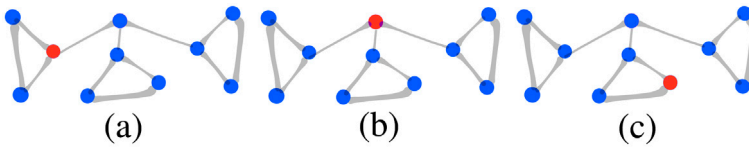


Figure 8. Placements of an infection’s origin (red) in \mathcal{A}_3^2 . Figures (a), (b) = Case (I); Figure (c) = Case (II).

- **Case (I):** when the infection’s origin I is adjacent to the body of \mathcal{A}_d^λ , as in Figure 8(a), or is in the body of \mathcal{A}_d^λ , as in Figure 8(b). In this case, the shortest path between I and any vertex is in the underlying regular tree and so the results of [Seibold and Callender \(2016\)](#) hold.
- **Case (II):** when the infection’s origin I is in a clique and adjacent to only vertices in a clique as in Figure 8(c). Note that from its definition, the body of \mathcal{A}_d^λ has

$$1 + \sum_{i=0}^{\lambda-3} d(d-1)^i$$

vertices, and as seen before the whole graph has $d(d-1)^{\lambda-2}$ cliques. The number of cliques in \mathcal{A}_d^λ depending λ is shown in Figure 7(b).

In order to study durations of disease outbreaks on d -cliqued networks, one first needs to understand the distances between nodes and the infection index, as well as how many nodes are at each distance. In what follows we shall answer these questions.

Definition 3.5: Denote by D_x the shortest path from a vertex x to the root of the underlying tree of \mathcal{A}_d^λ . Its size $|D_x|$ is the distance from x to the root.

From the above analysis, one can see that the most interesting case for us will be when the index case is in the outermost leaf of the network, and hence in what follows we shall fix such an I . That is, an index case I such that $|D_I| = \lambda$ as in Figure 9.

Lemma 3.6: For x a vertex in \mathcal{A}_d^λ , one has the following:

$$d(x, I) = \begin{cases} |D_x| + \lambda & \text{if } |D_x \cap D_I| = 0; \\ |D_x| + \lambda - 2\alpha & \text{if } |D_x \cap D_I| = \alpha > 0. \end{cases} \tag{8}$$

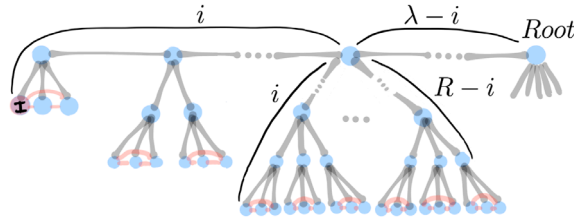


Figure 9. The main branch of I .

Proof: Consider first the case of $|D_x \cap D_I| = 0$, for which one knows that x is not in the main branch of the index vertex I , and hence to connect to the index node I one needs to go through the root of the tree. Therefore, the shortest path is given by $D_x \cup D_I$, and the first part of the lemma follows.

For x such that $|D_x \cap D_I| = \alpha > 0$, there exists $v_0 \in D_x \cap D_I$ such that

$$\alpha = |D_{v_0}| \geq |D_v|$$

for all $v \in D_x \cap D_I$. Thus,

$$d(x, I) = d(x, v_0) + d(v_0, I) \tag{9}$$

$$= (|D_x| - \alpha) + (\lambda - \alpha) \tag{10}$$

$$= |D_x| + \lambda - 2\alpha \tag{11}$$

and the lemma follows. □

In order to calculate the probabilities of infections lasting certain amount of time, and thus find the induced safe zones, one needs to understand the number of vertices at different distances from the infection index I .

Proposition 3.7: *The number N_R of vertices v of \mathcal{A}_d^λ for which $d(v, I) = R$ is*

$$N_R = \begin{cases} (d-1)^{R-\lambda} + \sum_{R/2 \leq i < \lambda} (d-2)(d-1)^{R-1-i} & R \geq \lambda; \\ 1 + \sum_{R/2 \leq i < R} (d-2)(d-1)^{R-1-i} & R < \lambda. \end{cases} \tag{12}$$

Proof: Consider first those nodes $v \in \mathcal{A}_d^\lambda$ for which $d(v, I) = R < \lambda$. By definition of \mathcal{A}_d^λ , and as seen in Figure 9, for each $R/2 \leq i < R$ there are

$$(d-2)(d-1)^{R-1-i} \tag{13}$$

vertices x for which

$$d(v, I) = R \quad \text{and} \quad |D_v \cap D_I| = \lambda - i. \tag{14}$$

Moreover, for $i = R$ there is only one vertex. Indeed, from the above lemma those vertices x satisfy $d(v, I) = |D_v| - \lambda + 2i = R$, and thus the proposition for $R < \lambda$ follows summing over all possible i , where $R/2 \leq i < R$.

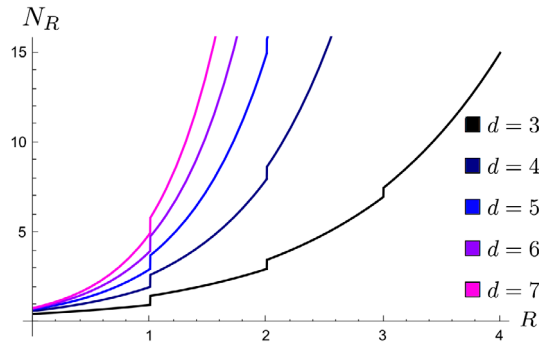


Figure 10. The number N_R of vertices at distance $R < \lambda$.

In order to understand the case of $R \geq \lambda$, note that those nodes $v \in \mathcal{A}_d^\lambda$ for which $d(v, I) = R \geq \lambda$ satisfy $|D_v \cap D_I| = \lambda - i$ for $0 \leq i \leq \lambda$. These cases can be classified as follows:

- (a) $|D_v \cap D_I| = 0$, or
- (b) $|D_v \cap D_I| = \lambda$, or
- (c) $|D_v \cap D_I| = \lambda - i$ for $0 < i < \lambda$.

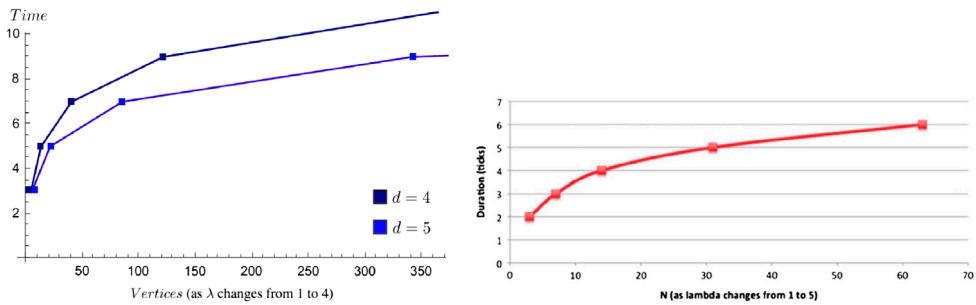
When considering case (a), note that by Lemma 3.6 one needs to consider points v in all branches not containing I of \mathcal{A}_d^λ , such that $|D_v| = R - \lambda$. By definition of a perfect $(d - 1)$ -ary tree, there are $(d - 1)^{R-\lambda}$ such points.

Consider next the case (b) above, and note that when $i = 0$, then there is only one vertex v satisfying the condition, which is the root, and thus corresponds to $R = \lambda$. Finally, consider case (c): from (13) to (14) and Figure 9, for each $R/2 \leq i < \lambda$, one has $(d - 2)(d - 1)^{R-1-i}$ vertices v for which $d(v, I) = R$ and $|D_x \cap D_I| = \lambda - i$. The proposition then follows for $R \geq \lambda$ summing over all cases (a), (b) and (c). □

The results from the above proposition can be visualised through Mathematica for a fixed value of λ as in Figure 10. In the following sections, we will use these results to study different models of epidemic outbreaks on d -cliqued tree networks. Finally, one should note that what has been shown in Lemma 3.6 and Proposition 8 serves to obtain the probabilities P_i of an outbreak lasting i ticks within d -ary trees when the index case is in the outermost vertex. These contact networks are those studied in Seibold and Callender (2016), but such probabilities were not given by the authors. Because of this, we calculated them here since we needed them to derive several useful comparisons between our model and that of Seibold and Callender (2016), and in particular to produce Figures 1 and 16.

4. Duration of an outbreak using a next-generation model

Recall that in a next-generation model, one sets the probability of infection to $P_{\text{inf}} = 1$, and the recovery probability to $P_{\text{rec}} = 1$. Then, in this case, the outbreak duration is $2\lambda + 1$, independently of the degree d . The examples in Figure 11 (a) for our model should be compared to (Seibold & Callender, 2016, Figure 6) appearing in Figure 11(b) for the model on perfect d -ary trees.



(a) Disease duration in a next-generation model with $P_{inf} = 1$ on \mathcal{A}_d^λ .

(b) Disease duration in 2-ary trees from Figure 6 of Seibold-Callender (2016).

Figure 11. Comparison of disease duration in a next-generation model.

As seen before, the degree d of \mathcal{A}_d^λ determines the number of nodes of the graph once the height λ is fixed. However, for SIR models on these graphs, just as in the case of [Seibold and Callender \(2016\)](#), the degree d has no effect on duration when the root is the index case, and only the height λ matters. On the other hand, since the height determines the network size, one can see how it influences the duration of the outbreak as in [Figure 11\(a\)](#) giving the duration scaling as $2\lambda + 1$ with population size stated as before.

4.1. Duration scaling with degree d and height λ

From the previous analysis, when the index case is located in the root, the outbreak will have the shortest duration regardless of the value of d , since the maximum distance from the root to any given vertex is the height λ . Taking into account the additional $+1$ step required to have the last infected nodes to become removed one has that the duration of the outbreak needs to be at most $\lambda + 1$. Note that, in particular, this case is equivalent to the one in [Seibold and Callender \(2016\)](#).

Equivalently, the work for d -ary trees will also hold when a random index node is chosen: since the diameter of the graph \mathcal{A}_d^λ is 2λ , the duration is bounded above by $2\lambda + 1$ when $d > 1$. Note that when $d = 1$, the graph \mathcal{A}_1^λ only exists for $\lambda = 1$, and it consists of two vertices connected by one edge. Hence, in this case, a random index will be equivalent to having the index in the root. As done before, we will denote by δ_λ the expected average outbreak duration, and P_i the probability the outbreak lasts exactly i ticks. Then, $\delta_\lambda = 2$ for $d = 1$ and the bounds of duration δ_λ otherwise are therefore given by

$$\lambda + 1 \leq \delta_\lambda \leq 2\lambda + 1 \text{ when } d > 1. \quad (15)$$

4.2. Duration scaling with degree d and the index case location

As mentioned before, the position of the origin of the infection (the index case I) within the graph will greatly determine the bounds on the outbreak durations. Moreover, both in Case (I), as well as in Case (II) of [Section 3.4](#), the outbreak durations for a next generation model will return results equivalent to those in [Seibold and Callender \(2016\)](#). To see this,

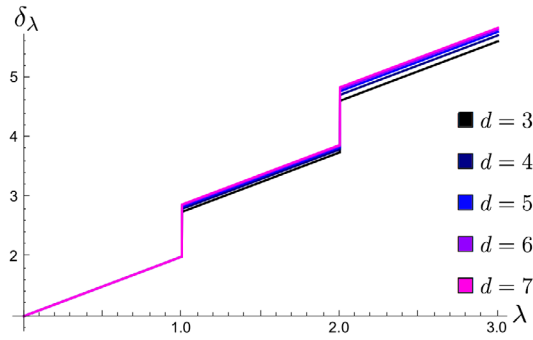


Figure 12. The mean duration outbreak δ_λ .

one needs to keep in mind that in a next generation model, the duration will always be $\max\{d(v, I) + 1 | v \in \mathcal{A}_d^\lambda\}$.

As in the case of d -ary trees, when a random index case is introduced away from the outermost vertex, the mean duration is given by the weighted average of the duration for each node multiplied by the number of nodes with that duration, this is, the weighted average of the vertices in each level multiplied by the duration they would produce. Hence, for each value of $\lambda > 1$, the work in [Seibold and Callender \(2016\)](#) still holds and the expected outbreak duration δ_λ is

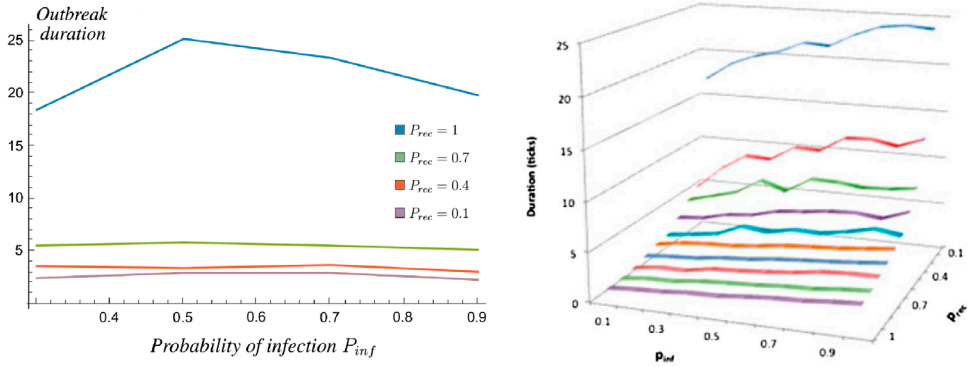
$$\delta_\lambda = \frac{1}{\sum_{i=0}^{\lambda} d^i} \sum_{j=0}^{\lambda} ((d^j)(\lambda + j + 1)), \quad (16)$$

and is depicted in [Figure 12](#).

When $\lambda = 1$ the results are quite different. Indeed, in this case, the graph \mathcal{A}_d^1 is the complete graph on $d + 1$ vertices, and thus all nodes are infected within 1 tick, whatever the choice of infection index is. Hence, the duration outbreak in this case is 2, different from [Equation \(6\) of Seibold and Callender \(2016\)](#) for d -ary trees. Moreover, these differences become even more apparent when allowing the probabilities of infection P_{inf} and recovery P_{rec} to vary, as in [Figure 13](#).

5. Duration of an outbreak using a general discrete time model

When considering discrete time models instead of next-generation models, outbreaks are specified in terms of varying probabilities of infection P_{inf} and of recovery P_{rec} , which can take values between 0 and 1. As an example, if $P_{\text{inf}} = 0.25$, then the adjacent node(s) to an infectious node will have a 25% probability of becoming infectious at the next time step. In particular, when P_{inf} decreases, some vertices may escape infection and thus the infection outbreak might last less than $2\lambda + 1$ ticks (recall that in [Figure 13](#) one could see the influence of the different probabilities on the duration outbreak). In what follows, we shall study the probabilities P_j of an outbreak lasting j tick (time units), and how these correlate to the scaling of different parameters.



(a) The average outbreak duration δ_λ depending on the recovery probability in \mathcal{A}_d^λ , averages taken over 20 simulations in IONTW. (b) Duration on a 4-regular tree graph with $\lambda = 1$ from Seibold-Callender (2016) through IONTW.

Figure 13. Durations for graphs with $\lambda = 1$, and $d = 4$.

5.1. Duration scaling with population size

Given an index node I in an outermost vertex, the values of the probabilities P_j for $j < \lambda$ will be much higher than in Seibold and Callender (2016). To calculate the average duration as a weighted average of all the possible durations, as well as to calculate the probabilities P_j for $j \leq \lambda$, one needs to consider the chance of an outbreak lasting j ticks and no one else getting infected after that time. For instance, by construction,

$$P_1 = (1 - P_{inf})^{d-1}. \tag{17}$$

As a comparison with Seibold and Callender (2016), consider an outbreak with the index case as one of the outermost vertices of \mathcal{A}_3^λ , this is, a vertex only adjacent to one of the 3-clique’s vertices. Setting the recovery probability $P_{rec} = 1$ but leaving $P_{inf} < 1$ unfixed, there are two nodes connected to the infectious index node. Then, there are three possible options for infection at the next time step: either no node becomes infected (with probability $(1 - P_{inf})^2$), both nodes become infected with probability P_{inf}^2 , or only one node becomes infected. Since it could be either adjacent node, this has probability of

$$2 \cdot (P_{inf}(1 - P_{inf})).$$

Through the same analysis as above, for generic degree d , all possible combinations of infectious nodes need to be taken into account. Hence, since the index node has $d - 1$ adjacent vertices, in \mathcal{A}_d^λ it can infect k vertices in $\binom{d-1}{k}$ ways, each with probability

$$P_{inf}^k(1 - P_{inf})^{d-1-k}.$$

To calculate P_2 , one would need the infection to last 2 ticks and no more, and thus all infected nodes at time 2 would not be allowed to infect anyone else. The calculation becomes more complicated in this setting than in d -ary trees since nodes which were not infected at time j within the clique might become infected at time $j + 1$ through a non-minimal path to the index node. Hence, when calculating the total probabilities, one

needs to differentiate between whether the vertices connected to the body of the clique are infected or not. In what follows, we shall illustrate how to calculate these probabilities.

Recall that the index case is adjacent to $d - 2$ vertices of the clique who are only adjacent to other vertices of the clique, and one vertex v_0 of the clique adjacent to the body of \mathcal{A}_d^λ , each which may be infected with probability P_{inf} , or not with probability $(1 - P_{\text{inf}})$. If v_0 is not infected, then to calculate P_2 , it should remain not infected at time 3, which will happen with probability

$$\underbrace{(1 - P_{\text{inf}})^2}_{v_0 \text{ uninfected at time 2 and 3}} \cdot \left(\underbrace{\sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{d-2-k}}_{\text{only } k \text{ infected at time 2}} \underbrace{(1 - P_{\text{inf}})^{d-2-k}}_{\text{no uninfected get infected at time 3}} \right). \quad (18)$$

When the vertex v_0 is infected, something similar happens to make the outbreak last 2 ticks, except that the neighbours of v_0 outside the clique need to be taken into account. Hence, this case contributes towards P_2 with the following

$$\underbrace{P_{\text{inf}}(1 - P_{\text{inf}})}_{v_0 \text{ does not infect outside clique}} \cdot \left(\underbrace{\sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{d-2-k}}_{\text{only } k \text{ infected at time 2}} \underbrace{(1 - P_{\text{inf}})^{d-2-k}}_{\text{no uninfected get infected at time 3}} \right) \quad (19)$$

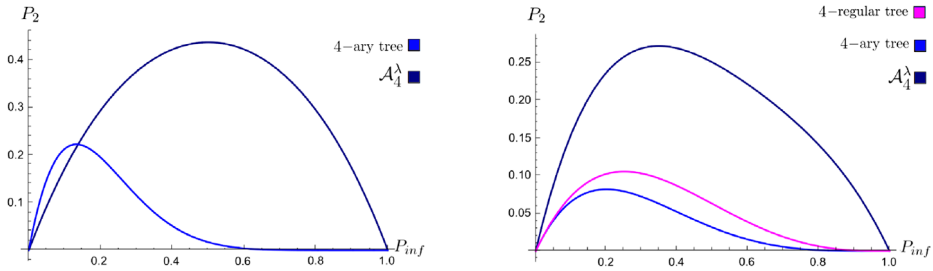
Then, to calculate P_2 one needs to sum Equations (18) and (19), leading to

$$\begin{aligned} P_2 &= ((1 - P_{\text{inf}})^2 + P_{\text{inf}}(1 - P_{\text{inf}})) \left(\sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{2d-4-2k} \right) \\ &= \sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{2d-3-2k}. \end{aligned} \quad (20)$$

Whilst one could compare this probability with the one found in [Seibold and Callender \(2016\)](#), the latter was obtained for an index case in the root and thus the comparison would not be too enlightening (see [Figure 14\(a\)](#)). However, we can deduce what the probabilities would have been for an index case in the same position as the one we are taking in \mathcal{A}_d^λ , in a vertex only adjacent to clique members.

As an example, in the setting of [Seibold and Callender \(2016\)](#), which is for d -ary trees, one has that when $d = 4$ the probability of the outbreak lasting 2 ticks is

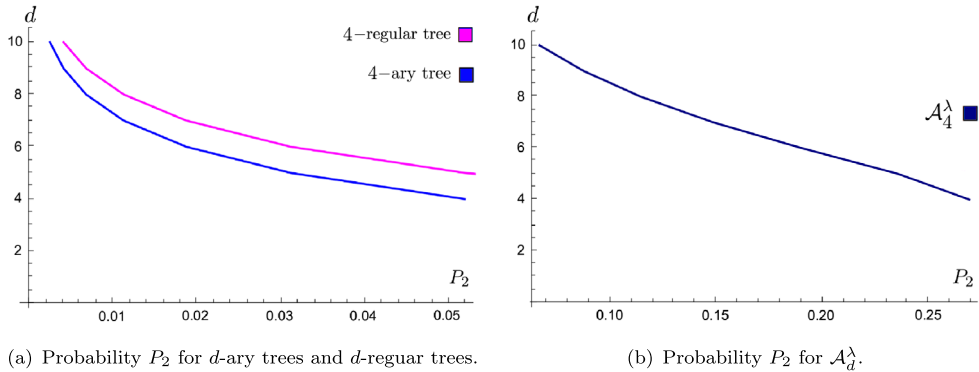
$$P_2 = P_{\text{inf}}(1 - P_{\text{inf}})^4. \quad (21)$$



(a) Root index of 4-ary tree as in Eq. (2) from Seibold-Callender (2016), compared with outermost index in \mathcal{A}_4^λ .

(b) Index case in outermost vertex of the 4-ary tree as in Eq. (21), and of the 4-regular tree as in Eq. (22), compared to \mathcal{A}_4^λ .

Figure 14. The values of P_2 in terms of the probability of infection P_{inf} for the graph \mathcal{A}_4^λ of Equation (20), and different index cases for the 4-ary tree of Seibold and Callender (2016), and d -regular trees.



(a) Probability P_2 for d -ary trees and d -regular trees.

(b) Probability P_2 for \mathcal{A}_d^λ .

Figure 15. The probability P_2 that an outbreak lasts 2 ticks for our model on d -cliqued networks and for the model in Seibold and Callender (2016), for index node in outermost vertex.

Moreover, one may want to also compare this to the setting of d -regular trees, in which case for $d = 4$ one has

$$P_2 = P_{inf}(1 - P_{inf})^3. \tag{22}$$

Comparing all of these settings, one can see how the chances of the outbreak lasting 2 ticks are much higher in the case of \mathcal{A}_d^λ , as seen in Figure 14(b) below.

5.2. Duration scaling with degree d

Whilst the height λ of the graph \mathcal{A}_d^λ will not influence the values of P_i for $i < \lambda$, the variation of the degree d will have a significant effect. In particular, it will enhance the differences between our model and the one studied in Seibold and Callender (2016). Consider for instance $P_{inf} = 0.4$. Then, for different values of the degree d , the probability P_2 for d -ary trees and for d -regular trees can be compared to the one for our d -cliqued networks \mathcal{A}_d^λ in Figure 15.

Although we have focused previously on the probability P_2 , similar analysis leads to equivalent results for P_i for higher values of i . Each probability P_i will be given by a nested sum of $i - 1$ terms. The outermost sum refers to the nodes directly adjacent to the index case inside the clique (which also appeared in P_2). The innermost sum refers to the nodes that experience infection and are at a distance of $i - 1$ from the root, and thus these probabilities can be deduced through Proposition 3.7, as we shall see below. To this end, recall that for \mathcal{A}_d^λ one has

$$P_1 = (1 - P_{\text{inf}})^{d-1};$$

$$P_2 = \sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{2d-3-2k}.$$

Then, for each $3 \leq i \leq d - 1$ one needs to separate the study of P_i into 2^i cases depending on whether the vertex v_0 in the index's clique adjacent to the body of \mathcal{A}_d^λ is infected or not at each step, as done for the case of P_2 .

To illustrate the technique, we shall show here how these calculations are done for P_3 . In this case, if the vertex v_0 is not infected at time 2 or 3, then the probability P_3 is given by

$$\underbrace{\left((1 - P_{\text{inf}})^3 \sum_{k=1}^{d-2} \binom{d-2}{k} P_{\text{inf}}^k (1 - P_{\text{inf}})^{d-2-k} \left[\sum_{j=1}^{d-2-k} \binom{d-2-k}{j} P_{\text{inf}}^j (1 - P_{\text{inf}})^{2(d-2-k-j)} \right] \right)}_{\substack{\text{no more infected} \\ \text{no further body vertex infected}}} \star$$

$\underbrace{\hspace{10em}}_{\text{no more infected}}$
 $\underbrace{\hspace{10em}}_{\text{only } k \text{ infected at time 2}}$
 $\underbrace{\hspace{10em}}_{j \text{ get infected at time 3}}$

When v_0 is infected at time 3, then the probability P_3 is similar to the one before, except that the only vertex in the body of \mathcal{A}_d^λ which is adjacent to v_0 need not get infected in the next step. Hence, in this case, through the term \star above one obtains a contribution of

$$\underbrace{P_{\text{inf}}(1 - P_{\text{inf}})^3}_{v_0 \text{ infected at time 3}} \star \tag{23}$$

If the vertex v_0 was infected at time 2, then it could infect its adjacent body vertex or not, and this will give two different cases equivalent to the cases in Equation (19) for P_2 . If no one in the body is infected, then one has that the contribution to P_3 is as in Equation (23). On the other hand, if v_0 infects someone at time 3, then that vertex needs not to infect anyone at time 4. Since this vertex has $d - 1$ adjacent nodes (e.g. see Figure 9), one has that in this case the contribution to P_3 is given by

$$\underbrace{P_{\text{inf}}^2(1 - P_{\text{inf}})}_{v_0 \text{ infects body vertex}} \cdot \underbrace{(1 - P_{\text{inf}})^{d-1}}_{\text{no further body vertex infected}} \star$$

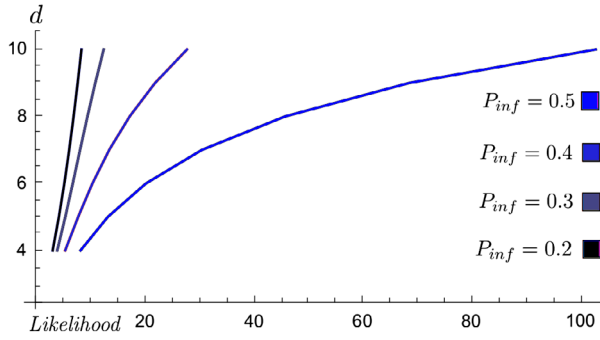


Figure 16. The likelihood of P_2 depending on the degree d for different values of P_{inf} . Notes: This is, the ratio of P_2 for \mathcal{A}_d^λ and P_2 for d -ary trees, both with index vertex in outermost layer.

Summing over these four cases, one has

$$\begin{aligned}
 P_3 &= \left[(1 - P_{inf})^3 + 2P_{inf}(1 - P_{inf})^3 \right. \\
 &\quad \left. + P_{inf}^2(1 - P_{inf})(1 - P_{inf})^{d-1} \right] \cdot \star \\
 &= \left[(1 + 2P_{inf} + P_{inf}^2)(1 - P_{inf})^3 \right. \\
 &\quad \left. + (1 - P_{inf})^{d-3} \right] \cdot \star
 \end{aligned}$$

Therefore, one has that

$$P_3 = \left[(1 - P_{inf}^2)^2 + (1 - P_{inf})^{d-2} \right] \cdot \star \tag{24}$$

In order to compare the probability of an outbreak lasting 2 times for a network modelled through [Seibold and Callender \(2016\)](#) and for a d -cliqued network, we consider the ratio between P_2 for the latter by the former as seen in Figure 16. In particular, one can see that the higher the probability of infection P_{inf} , the higher is the probability P_2 for \mathcal{A}_2^λ compared to the probability P_2 for d -ary trees with the same index root, and the same will apply for the other P_i .

6. Future work

The work presented here introduces a new perspective to study epidemics in networks with complete subnetworks, and thus we shall mention here a few of the many paths through which one may generalize our results.

6.1. Safe zones in d -cliqued networks

In order to contain an infection outbreak, one would like to understand how the removal of edges or vertices of a graph creates a safe zone of vertices which have a very low probability of an infection. For this, we introduce the following definition:

Definition 6.1: Given $1 \gg \varepsilon > 0$, a safe zone $\mathcal{S}_\varepsilon^t$ of a graph G is a connected subset of vertices of G that have probability of infection at most ε at time t . The size $|\mathcal{S}_\varepsilon^t|$ is the number of vertices in it.

The structure of the safe zone depends on the shortest paths from I to other vertices, as well as on the probability of infection P_{inf} , and of recovery P_{rec} . The paths between vertices can be understood via Lemma 3.6 and Proposition 3.7, and are closely related to the correspondence between the number of edges in \mathcal{A}_d^λ , the degree d and height λ , which from Equation (7) can be seen in Figure 7. We strive to find the number of edges that can be removed in the contact network, depending on the nodes which are initially infected.

We have seen that in many cases the outbreak duration in our networks will be quite different from the duration found in Seibold and Callender (2016). In particular, the case of \mathcal{A}_4^1 appearing in Figure 13 should be compared to the case of $\lambda = 1, d = 4$ of Seibold and Callender (2016, Figure 12) appearing in Figure 13. Moreover, when $1 \leq i \leq d - 1$ the probabilities P_i of an outbreak lasting i ticks increase quite drastically, since in \mathcal{A}_d^λ our index node I is within a clique, and thus all vertices not infected in one step may still be infected in the other. Hence, the safe zones for \mathcal{A}_d^λ will be smaller than those for perfect d -ary trees of Seibold and Callender (2016). For instance in \mathcal{A}_3^2 with $P_{\text{inf}} = 1$ there is a safe zone with $|\mathcal{S}_\varepsilon^1| = 7$, two with $|\mathcal{S}_\varepsilon^2| = 3$, and two with $|\mathcal{S}_\varepsilon^3| = 2$.

6.2. Epidemics in random networks

Understanding epidemics on random networks is an open problem which has attracted attention from researchers in many areas of mathematics. For this, it would be most important to understand connectivity of the random networks under consideration, as well as their clique density. One should note that whilst connectivity is important, the understanding of *safe zones* becomes much harder for connected graphs than for sparse graphs. In particular, we expect to understand whether having many cliques preclude the existence of safe zones in random networks.

7. Conclusion

The importance of precise and efficient mathematical modelling of epidemics in order to analyse biological and social networks is widely accepted. As these systems are often symmetric, our results mostly focus on contact networks with symmetric clusters (cliques), providing an enlightening setting from which to study society. We have introduced several new mathematical objects which appear to be of importance when studying epidemic dynamics in those social groups that have complete subnetworks, describing subgroups of people who are all related to each other.

In order to account for regular social groups where interactions are allowed amongst the youngest members of the network, we introduced a generalization of regular trees to *d-cliqued tree graphs* \mathcal{A}_d^λ (see Definition 3.3). These objects allow us to model, for example, social groups where children are connected to all their school class mates and not only their teachers.

In order to study infections in these new graphs, we extended the results of Seibold and Callender (2016) to our setting. We considered different ways in which an outbreak could originate and showed that when it originated in the root of the graph the results of Seibold and Callender (2016) still hold. On the other hand, we proved that when the index I of the graph (the origin of the infection) is one of the youngest nodes (the red node in Figure

8(c)), then its siblings will be infected immediately, whilst in the setting of [Seibold and Callender \(2016\)](#), they wouldn't.

Without the use of the IONTW platform, we were able to study the characteristics of d -cliqued tree graphs equivalent to those for d -ary trees of [Seibold and Callender \(2016\)](#), and showed that in many cases safe zones are smaller in our setting. Moreover, when the index vertex is within a network's clique and not adjacent to the body of the network, though the probabilities P_i of an outbreak lasting i ticks one could calculate the expected average duration δ_d^λ of the graph \mathcal{A}_d^λ as the weighted average of each of the possible durations $\delta_d^\lambda = \sum_{i=1}^{2\lambda+1} iP_i$.

Whilst in [Seibold and Callender \(2016\)](#) the model on perfect d -ary trees was shown to be greatly influenced by the size of the graph when determining the duration of an outbreak, even more so than the number d of branches which has minimal effect after the root is infected, we have shown that in our model the height is irrelevant when calculating P_i for small values of i . In fact, we have described how these values depend mainly on the degree d of the graph, which emphasises the difference between the two models.

When considering next generation models where $P_{\text{inf}} = P_{\text{rec}} = 1$, we saw that not much analysis was needed. However, interesting mathematics appeared when implementing more general discrete models for the index case being in the youngest node, through which the different values of P_{inf} influenced the duration of the outbreak. One should note that even for this type of index, the probabilities P_i for the d -ary trees were not given in [Seibold and Callender \(2016\)](#), and thus our methods help, in particular, to understand the simpler model of d -ary trees. Since the distance between the index case and other vertices are a very important factor for both scenarios, we dedicated a section to give formulas for the number of nodes at each distance from the index case, which are then used to calculate each probability P_i of an outbreak lasting exactly i ticks or time units, both for the setting of [Seibold and Callender \(2016\)](#) as well as for ours.

Finally, by considering the ratio between P_i for d -ary trees and for our \mathcal{A}_d^λ , we were able to quantise the different likelihoods of an outbreak lasting P_i in the two different models. In particular, as mentioned before, the higher the probability of infection P_{inf} is, the higher is the probability P_2 for \mathcal{A}_d^λ compared to the probability P_2 for d -ary trees with the same index root, and equivalent results apply for other P_i . In other words, the chance of an outbreak lasting two time units is much higher (sometimes 100 times higher) for our new graphs, showing the importance of accounting for relations between multiple individuals (see [Figure 16](#)).

Note

1. A vertex v is the child of a vertex w iff the unique path from v to the root contains w .

Acknowledgements

Much of the research in this paper was conducted by the second author under the supervision of the first one, as part of the MIT PRIMES-USA program. Both authors would like to thank MIT PRIMES-USA for the opportunity to conduct this research together, and in particular Tanya Khovanova for her continued support. The authors are also thankful to Kevin Ren, Fidel I. Schaposnik and James A. Unwin for their comments on the paper, and the referees of the journal for their useful corrections.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work of LPS is partially supported by NSF DMS [grant number DMS-1509693]; Alexander von Humboldt Fellowship.

References

- Brauer, F., & Castillo-Chavez, C. (2001). *Mathematical models in population biology and epidemiology*. New York (NY): Springer.
- Just, W., Callender, H. L., & Lamar, D. M. (2015). *Exploring transmission of infectious diseases on networks with NetLogo*. Retrieved 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(4), 295–307.
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115(772), 700–721.
- Mandal, S., Sarkar, R. R., & Sinha, S. (2011). Mathematical models of malaria – A review. *Malaria Journal*, 10, 202.
- Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63, 066117.
- 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.
- Seibold, C., & Callender, H. L. (2016). Modeling epidemics on a regular tree graph. *Letters in Biomathematics*, 3(1), 59–74.
- Wilensky, U. (1999). *Netlogo*. Center for connected learning and computer-based modeling. Evanston, IL: Northwestern University.