

# Mathematical model of biofilm-mediated pathogen persistence in a water distribution network with time-constant flows

SADIQAH AL MARZOOQ<sup>1,2</sup>, ALVARO ORTIZ-LUGO<sup>1</sup> and  
BENJAMIN L. VAUGHAN, Jr.<sup>1</sup>

<sup>1</sup>*Department of Mathematical Sciences, McMicken College of Arts and Sciences,  
University of Cincinnati, Cincinnati, Ohio 45221-0025, USA*

*emails: s\_almarzooq@yu.edu.sa, ortizlaa@mail.uc.edu, vaughabn@ucmail.uc.edu*

<sup>2</sup>*Al Yamamah University, P.O. Box 45180, Riyadh 11512, Saudi Arabia*

(Received 23 May 2017; revised 3 May 2018; accepted 9 May 2018; first published online 6 June 2018)

In industrialized nations, potable water is often provided through sophisticated water distribution systems. If pathogenic bacteria are introduced into the water distribution network, the presence of a biofilm can lead to biofilm-assisted retention of the pathogens, affecting the potability of the water. To study the dynamics of planktonic and biofilm-bound pathogens within the large network of pipes in a water distribution system, we develop a network model governing the concentration of introduced pathogens within the bulk fluid and the biofilms lining the pipes. Under time-constant flow regimes within the network, we prove that the long-time behaviour of the entire network is dependent on the Lyapunov exponents for each connection in the network when viewed in isolation and the network connectivity. An efficient algorithm is developed for predicting the long-time behaviour of the pathogen population within large networks using the network's topological ordering. The algorithm's predictions are validated using numerical simulations of the full non-linear system on a range of water distribution network sizes.

**Key words:** 05C20 directed graphs, 34D20 stability, 92B05 general biology and biomathematics, 92D99 biology and other natural sciences.

## 1 Introduction

The availability of potable (drinking) water is essential for human health and wellness. In industrialized nations, many but not all consumers receive potable water through drinking water distribution systems. These systems consist of a network of pipes along with storage facilities (reservoirs, water towers, etc.) and pressurized components (pumping stations) to deliver the water to consumers.

Water quality is the measure of the condition of water within the network relative to some standard for the safety of the water for human contact and consumption. Contaminants such as microorganisms, salts, metals, pesticides, herbicides, pharmaceuticals, and radioactive materials can negatively affect the water quality. In particular, biological contamination from accidental exposure to fecal matter due to cracks, back-flow incidents,

unsecured reservoirs, or leaching pipes can create a significant health risk (LeChevallier *et al.*, 2003; Momba *et al.*, 2000; Murray *et al.*, 2006; Uber *et al.*, 2004). The presence of coliform bacteria such as *Escherichia coli* are used as a marker for fecal contamination. Waterborne diseases such as botulism (*Clostridium botulinum*), cholera (*Vibrio cholerae*), dysentery (*Shigella dysenteriae*), and typhoid fever (*Salmonella typhi*) are thought to be responsible for 1.8 million deaths worldwide every year with an estimated 88% due to unsafe water supplies (World Health Organization, 2002).

There has been extensive research on the design of water distribution networks to optimize flow rates and water quality while reducing costs as well as the optimal maintenance of existing systems. For an introduction to water engineering, see Mays (2010) and Davis (2011). Of particular interest is the dynamics of non-native bacteria, possibly pathogenic, released into the water distribution system. Deterioration of water quality as it passes through drinking water distribution systems, especially as it relates to microbial contamination, has been well documented (Besner *et al.*, 2002; Camper *et al.*, 1996; Carabeş *et al.*, 2011; Carter *et al.*, 2000; LeChevallier *et al.*, 1996; Power and Nagy, 1999; Skraber *et al.*, 2005; van der Kooij, 1992; Zhang and DiGiano, 2002).

In water pipes, indigenous bacteria colonize the solid surfaces in structures called biofilms. Biofilms are aggregates of various microorganisms that attach themselves to solid surfaces using secreted extracellular polymeric substances. Biofilms are ubiquitous and are a crucial element in a wide range of harmful as well as beneficial phenomena including tooth decay, remediation of wastewater, maintenance of naval ships, bacterial growth on surgical implants, and complications in cystic fibrosis patients. While the existence of heterotrophic biofilms in a drinking water distribution system does not directly correlate with human health risk (Bartram, 2003), the presence of a biofilm may encourage the retention and possibly even growth of pathogens released within the system.

Bacteria within a biofilm have a distinct survival advantage compared with planktonic (free-swimming) bacteria in a water distribution system. It has been observed that biofilms assist in the survival of microorganisms under inhospitable conditions found within distribution systems through improved nutrient absorption and improved resistance to disinfectants (Camper, 1998; LeChevallier *et al.*, 1996; Szabo, 2006). In particular, the improved resistance to disinfectants makes the removal of potentially pathogenic bacteria difficult.

The attachment of planktonic pathogens to an established biofilm can lead to a reservoir of contamination (Camper, 1998) where pathogens survive disinfection and possibly grow, reappearing long after their initial introduction into the system (Camper and McFeters, 2000; Szabo, 2006) when detachment occurs and mobilizes them once again within the distribution system. Changing flows related to changes in the instantaneous demand within water distribution systems have been shown to contribute to detachment of biofilm and increases in water turbidity (Lehtola *et al.*, 2006).

A significant body of research on the growth and structure of bacterial biofilms has appeared over the last four decades. A classic resource on the subject is the Costerton *et al.* (1995) review in *Annual Review of Microbiology*, and, more recently, Flemming and Wingender (2010) provide a comprehensive discussion of the biofilm matrix in *Nature Reviews: Microbiology*. There has also been extensive research in the mathematical

modelling of biofilms. A comprehensive resource is a review by the IWA task force on biofilm modelling (Eberl *et al.*, 2006).

A key mechanism in the dynamics of non-native bacteria is the attachment/detachment processes, which involves the exchange of bacteria from the biofilm to the fluid (detachment) and vice versa (attachment). In the case of detachment, large-scale pieces of the biofilm can separate from the mass in a process called sloughing (Trulear and Characklis, 1982; van Loosdrecht *et al.*, 2002). When the detachment involves individual or small numbers of bacteria, the process is called erosion (van Loosdrecht *et al.*, 2002).

Of particular interest is the macroscale dynamics of non-native bacteria introduced into a water distribution network. The microscale and mesoscale dynamics of biofilms have been extensively studied theoretically (Eberl *et al.*, 2006) but there is a need to incorporate this research with the larger scale network dynamics. This requires network-level models of fluid dynamics within the network (Eiger *et al.*, 1994; Rossman, 2000; Rossman and Boulos, 1996) coupled with models of various complexity describing the interaction between the network topology, fluid mechanics of the flow, the planktonic bacteria, and the biofilms. Previous work on modelling the dynamics of biological contaminants have focused on quantitative behaviour such as numerical measures of water quality (Murray *et al.*, 2006; Uber *et al.*, 2004). We are interested in qualitative predictions of the system's behaviour such as possible colonization of indigenous biofilms within the network by pathogens, leading to persistent biofouling of the drinking water, and the predictions of the critical pipes that dominate the contamination.

There exists a significant body of research on dynamical systems on networks. For example, coupled oscillators models on networks in neuroscience where the oscillators can exhibit phase locking and synchronization (Ashwin *et al.*, 2016). The classic model is the Kuramoto model (Kuramoto, 1984; Strogatz, 2000), which typically has an all-to-all or nearest-neighbour network topology. Other examples include epidemiological models on networks (Brauer *et al.*, 2012), where subpopulations are modelled as nodes on a graph whose edges describe the interconnection between the different subpopulations, and dynamics of influence on social networks (Castellano *et al.*, 2009). For more thorough reviews, see Barrat *et al.* (2008), Vespignani (2012), and Porter and Gleeson (2016).

In these different examples of dynamical systems on networks, master stability conditions and functions could be used to understand the stability of their steady states. By relating the eigenvalues of adjacency matrix to the dynamics, one can determine the stability of the steady states for the entire network. In this paper, our approach differs in that we do not examine master stability conditions or functions but we use the structure of the adjacency matrix to identify specific locations within the network where the trivial steady state would be stable or unstable.

Additionally, water distribution networks often have network topologies that can both make parts of the analysis easier or more difficult. For example, distribution networks typically have sources (towers, tanks, reservoirs, etc.) and sinks (usages points) that create a directed flow within the network. This property is what we use in this paper to aid in our analysis. Distribution networks can also contain directed cycles or be a dynamic network where the connectivity changes within the directed graph, both of which complicate the analysis. While we do not study these cases here, the results of this paper suggest those

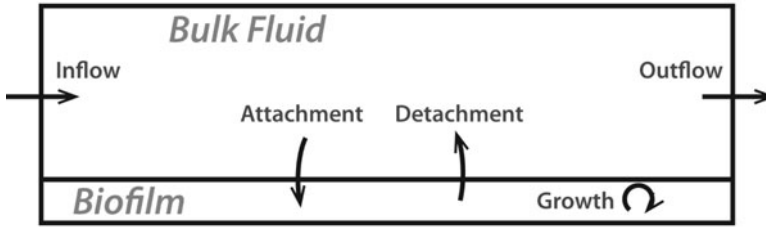


FIGURE 1. The compartmental model of pathogen dynamics within a single connection.

cases will be tractable by our analytical framework and are currently the subject of future study.

This paper is organized as follows. We introduce a model for the dynamics of non-native bacteria in a single network connection with a time-constant flow and discuss its dynamics in Section 2. We introduce a model for the dynamics of non-native bacteria in a distribution network with a time-constant flow in Section 3. Proofs regarding the linear stability of the trivial steady for the network model is given in Section 4. An efficient algorithm for determining the long-time behaviour of the pathogen concentration within the network is presented in Section 5. Results are given for example networks in Section 6 and conclusions and future directions are discussed in Section 7.

### 2 Single connection model

A water distribution system is constructed from a network of collection points, purification facilities, pipes, pumping stations, storage components such as reservoirs, water towers, water tanks, and usage points. For our model, we will assume that the system consists of a collection of one or more pipes that we call a ‘connection’ and ignore the other system components. For a single connection, we develop a compartment model governing the concentration of pathogenic bacteria. We assume that the interior surface of a fluid-filled connection is lined with a drinking water biofilm that is native to the distribution system. Pathogenic bacteria are introduced into the connection with the total concentration in the connection partitioned into two compartments: planktonic bacteria in the bulk fluid and sessile (biofilm-bound) bacteria. We assume that the pathogens can attach to the biofilm, detach from the biofilm into the bulk fluid, grow within the biofilm, and be transported out of the connection with the effluent as shown in Figure 1. Since our goal is to introduce a model for a distribution network, we assume that growth of the pathogens within the bulk fluid is negligible for mathematical simplicity.

Assuming well-mixed compartments, the concentrations of planktonic pathogens in the bulk fluid,  $C_f$ , and the biofilm-bound pathogens,  $C_b$ , are governed by the system

$$\begin{cases} C'_f &= -\frac{Q}{V}C_f + \beta\frac{S_A}{V}C_b - \alpha C_f \\ C'_b &= \alpha\frac{V}{S_A}C_f - \beta C_b + rC_b\left(1 - \frac{C_b}{K}\right), \end{cases} \tag{2.1}$$

where  $Q$  is the fluid flow rate in  $\text{hr}^{-1}$ ,  $V$  is the fluid volume in the connection in mL,  $S_A$  is the surface area of the connection in  $\text{cm}^2$ ,  $\alpha$  is the attachment rate in  $\text{hr}^{-1}$ ,  $\beta$  is the detachment rate in  $\text{hr}^{-1}$ ,  $r$  is the logistic growth rate in  $\text{hr}^{-1}$ , and  $K$  is the pathogen carrying capacity of the biofilm in colony forming units per square centimeter ( $\text{CFU cm}^{-2}$ ). The units of  $C_f$  are  $\text{CFU cm}^{-3}$  and  $C_b$  are  $\text{CFU cm}^{-2}$ . All parameters are assumed to be positive constants.

Since we are interested in the dynamics of pathogens introduced into the system, we assume that the pathogens are initially only in the bulk fluid and there were no pathogens initially within the biofilm. Hence, the initial conditions are  $C_f(0) = C_0$  and  $C_b(0) = 0$ .

Using the scales  $C_f = K (S_A/V) u$ ,  $C_b = Kv$ , and  $t = r^{-1}\tau$ , the dimensionless equations are

$$\begin{cases} u' &= \hat{\beta}v - (\hat{\alpha} + \hat{Q})u \\ v' &= \hat{\alpha}u - \hat{\beta}v + v(1 - v), \end{cases} \tag{2.2}$$

with the initial conditions  $u(0) = u_0$  and  $v(0) = 0$ , where  $\hat{\alpha} = \alpha/r$ ,  $\hat{\beta} = \beta/r$ ,  $\hat{Q} = Q/(Vr)$ , and  $u_0 = C_0V/(S_AK)$ .

The dynamics of the dimensionless system (2.2) are straightforward. There are two possible steady states: the trivial steady state  $(u^*, v^*) = (0, 0)$ , which we denote as the ‘washout’ steady state where there are no pathogens in the system, and

$$(u^*, v^*) = \left( \frac{\hat{\beta}}{\hat{\alpha} + \hat{Q}}v^*, 1 - \frac{\hat{\beta}\hat{Q}}{\hat{\alpha} + \hat{Q}} \right), \tag{2.3}$$

which we denote as the ‘persistence’ steady state. The persistence steady state, (2.3), is positive when either

$$\hat{\beta} \leq 1 \tag{2.4}$$

or

$$\hat{\beta} > 1 \quad \text{and} \quad \hat{Q} < \frac{\hat{\alpha}}{\hat{\beta} - 1} = Q_c. \tag{2.5}$$

It is trivial to show that the eigenvalues are real so there are no oscillations in the system and the persistence steady state is linearly stable when  $\hat{\beta} \leq 1$  or  $\hat{\beta} > 1$  and  $\hat{Q} < Q_c$ . In the phase plane, shown in Figure 2 for the three possible cases, it is straightforward to prove that all trajectories for  $u, v \geq 0$  approach the linearly stable steady state.

The first condition ( $\hat{\beta} \leq 1$ ) indicates that the pathogens reproduce faster or at the same rate as they detach from the biofilm so they persist regardless of the fluid flow. The second condition ( $\hat{\beta} > 1$  and  $\hat{Q} < Q_c$ ) indicates that the pathogens detach from the biofilm faster than they grow in the biofilm but they are not removed from the system by the fluid rapidly enough, giving them enough time to reattach and persist in the system.

The dynamics of the washout steady state mirror the persist steady state. For  $\hat{\beta} > 1$  and  $\hat{Q} > Q_c$ , the washout steady state is linearly stable. Here, the fluid flow rate is large enough to remove the bacteria from the system. For  $\hat{Q} < Q_c$  or  $\hat{\beta} \leq 1$ , the trivial steady state is linearly unstable.

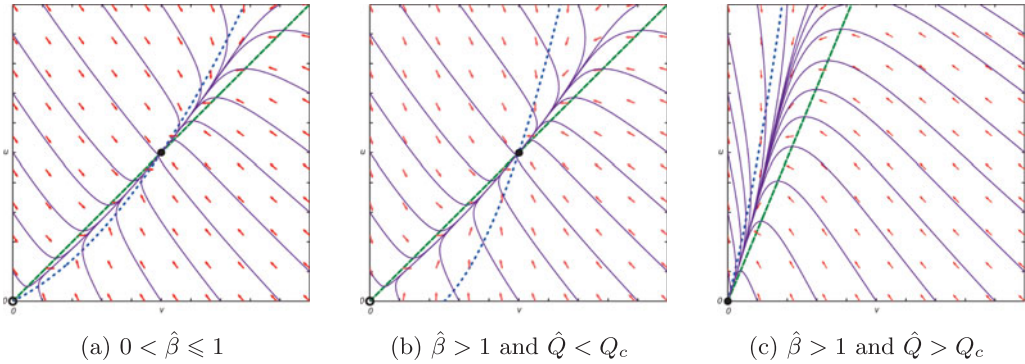


FIGURE 2. Phase portraits of the non-linear system (2.2) for the three cases. The dotted blue line indicates the  $v$ -nullcline, the green dashed line the  $u$ -nullcline, the red arrows that direction field, and the purple lines are individual trajectories. An open circle denotes a linearly unstable steady state and a filled circle a linearly stable steady state. The case where  $\hat{\beta} < 1$  is shown in (a). In this case, there are two steady states and all trajectories approach the non-trivial steady state. For  $\hat{\beta} > 1$ , there are two steady states when  $\hat{Q} < Q_c$  with all trajectories approaching the non-trivial steady state as shown in (b). When  $\hat{\beta} > 1$  and  $\hat{Q} > Q_c$ , there is only the trivial steady state and all trajectories approach it as shown in (c).

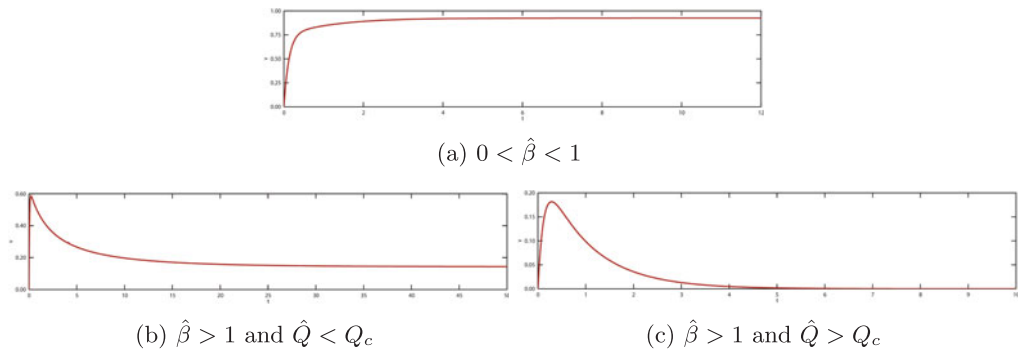


FIGURE 3. Numerical results for the three different cases. The pathogen concentration within the biofilm,  $\hat{v}$ , is shown in (a) for the case  $\hat{\beta} < 1$ . Here, the pathogen increases until it reaches the non-zero steady state, indicating the pathogens will persist for all  $t > 0$ . For  $\hat{\beta} > 1$  and  $\hat{Q} < Q_c$ , shown in (b), the pathogen concentration approaches a non-negative steady state, again indicating the pathogens persist within the pipe for all  $t > 0$ . When  $\hat{\beta} > 1$  and  $\hat{Q} > Q_c$ , the trivial steady state is stable and the pathogen washes out of the system, as shown in (c).

Numerical results using MATLAB's ode45 function (Shampine and Reichelt, 1997) are shown in Figure 3 and confirm the analysis. For  $\hat{\beta} < 1$ , the pathogens always persist in the connection and the pathogen concentration in the biofilm increases monotonically. For  $\hat{\beta} > 1$ , the system undergoes a bifurcation where the pathogens transition from persisting to washing out of the connection as  $\hat{Q}$  increases through the critical flow rate  $Q_c$ . There is an initial increase in the pathogen concentration within the biofilm before decreasing to the persist steady state or zero depending on  $\hat{Q}$ .

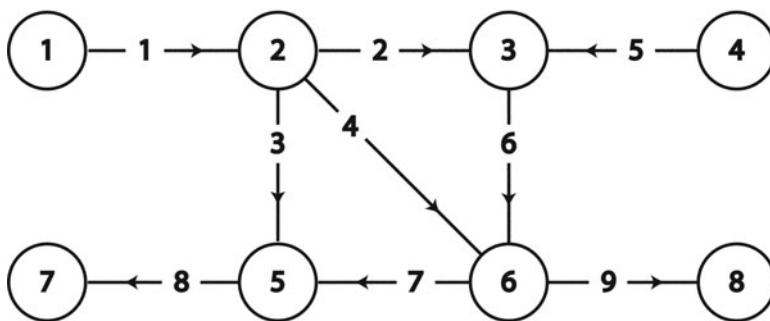


FIGURE 4. A model water distribution network. The circles are junctions and the lines are connections with the arrow indicating the direction of the flow. Junctions 1 and 4 are water sources, 7 and 8 are usage points, and 2, 3, 5, and 6 are interior junctions.

### 3 Network model

To model the entire distribution network, we assume that it can be described as a collection of connections joined at junctions, as shown in Figure 4, with the direction of the flow through the connections indicated by the arrows. In this example distribution network, junctions 1 and 4 are water sources, junctions 7 and 8 are usage points, and the remaining junctions are internal junctions. When modelling real-world networks, a connection in the network could be a collection of individual pipes within the actual distribution network, where the connection would have some averaged or aggregated properties.

Mathematically, we model the network as a directed graph (digraph)

$$D = (V(D), A(D)) = (\{v_1, \dots, v_N\}, \{a_1, \dots, a_M\}), \tag{3.1}$$

with  $N$  vertices (junctions)  $v_1, \dots, v_N$  and  $M$  arcs (connections)  $a_1, \dots, a_M$ . We will use vertex and junction as well as connection and arc interchangeably.

For each arc  $a$ , we have an incident map  $I_D(a) = (u, v)$  where the vertex  $u$  is the tail of  $a$  and  $v$  is the head of  $a$ . An arc  $a$  is called incident out of  $u$  and incident into  $v$ . For our model, we assume the digraph is simple (no multiple arcs or graph loops) and is strongly connected since an unreachable vertex or arc would not be part of the distribution network and can be ignored.

The incidence matrix of digraph,  $\mathbf{A}(D) \in \mathbb{Z}^{N \times M}$ , is defined by

$$\mathbf{A}(D)_{i,j} = \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ arc is incident out of the } i^{\text{th}} \text{ vertex} \\ -1 & \text{if the } j^{\text{th}} \text{ arc is incident into the } i^{\text{th}} \text{ vertex} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

For each vertex  $v$ , we define the outdegree of  $v$ ,  $\text{deg}^+(v)$ , as the number of arcs that are incident out of  $v$ , the indegree of  $v$ ,  $\text{deg}^-(v)$ , as the number of arcs that are incident into of  $v$ , and the degree of  $v$  as  $\text{deg}(v) = \text{deg}^+(v) + \text{deg}^-(v)$ . We define a leaf vertex  $v$  as a vertex where either  $\text{deg}^-(v) = 0$  or  $\text{deg}^+(v) = 0$  and we denote a non-leaf vertex as an interior vertex.

To derive our model for the pathogen concentration within the network, we need to construct an internal incidence matrix,  $\mathbf{A}_I(D) \in \mathbb{Z}^{N_I \times M}$ , where  $N_I$  is the number of interior

vertices, by deleting the rows that correspond to each leaf vertex.  $A_I(D)$  is a submatrix of  $A(D)$  where each row that corresponds to a leaf vertex is removed. For the digraph shown in Figure 4, the incidence and the internal incidence matrices are

$$A(D) = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

and

$$A_I(D) = \begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 & 0 & -1 \end{bmatrix}. \tag{3.4}$$

### 3.1 Determining the flow rate

To compute the dynamics of pathogens within a water distribution network, we need a method for computing the fluid flow profile within the network. For each connection in the network, we define a fluid flow rate,  $Q$ , and for each junction in the network, we define the fluid pressure and the system demand. The system demand at each junction models the water removed from the network by various usage points. Since our focus is on the dynamics of the pathogens and not on the actual flow, we will assume that the flow is constant in time throughout the network and each connection can be modelled as a linear resistance pipe with pressures defined at each junction. The flow rate through a linear resistance pipe is

$$Q = \frac{P_a - P_b}{R}, \tag{3.5}$$

where  $P_a$  and  $P_b$  are the pressures at the ends of the pipe and  $R$  is the hydraulic resistance of the pipe. For a network, the flow rates are computed using the incidence matrix,  $A$ , in the relation

$$r_f \circ q = A^T p, \tag{3.6}$$

where  $q \in \mathbb{R}^M$  is the vector of flow rates for each connection,  $r_f \in \mathbb{R}^M$  is the vector of hydraulic resistances for each connection, the  $\circ$  operator denotes the Hadamard (element-wise) product, and  $p \in \mathbb{R}^N$  is the vector of pressures at the junctions.

Note that we use the incidence matrix for a directed graph before we know the direction of flow in each connection. This is compensated by the sign of the flow rate, where  $Q > 0$  indicates that the fluid flows in the direction of the arc from the tail to the head and  $Q < 0$  indicates flow from the head to the tail. For convenience, we will assume  $Q > 0$  for all connections and compensate by flipping the head and tail vertices of the corresponding arcs.



For each interior junction, we prescribe a demand,  $D_i$ , and for each leaf junction, we prescribe a pressure,  $P_i$ . From (3.6) and the pressure conditions at the leaf junctions, we have  $M + (N - N_J)$  equations for  $M + N$  unknowns. The remaining  $N_J$  equations come from conservation of mass at each interior junction (Kirchhoff's law). At each interior junction, we have the relationship

$$\underbrace{\sum_{k=1}^{M_{\text{inflow}}} q_j}_{\text{Flow Into Junction from Connections}} - \underbrace{\sum_{k=1}^{M_{\text{outflow}}} q_k}_{\text{Flow Out of Junction to Connections}} = \underbrace{D}_{\text{Demand at Junction}}. \tag{3.7}$$

The equations for the entire network, using the incidence matrix, is

$$\Delta \mathbf{A} \mathbf{q} = \mathbf{d}, \tag{3.8}$$

where  $\Delta \in \mathbb{R}^{N \times N}$  is a diagonal matrix with  $\Delta_{i,i} = 1$  if the  $i^{\text{th}}$  junction is an interior junction and  $\Delta_{i,i} = 0$  otherwise. The vector  $\mathbf{d} \in \mathbb{R}^N$  is the vector of the demand at each junction. For leaf junctions, the values of  $\mathbf{d}$  are set to zero. Combining (3.6), (3.8), and the conditions on the pressure at non-interior junctions, the linear system governing the flow rates through the connections and the pressures at the junctions is

$$\begin{bmatrix} \mathbf{I} - \Delta & \Delta \mathbf{A} \\ \mathbf{A}^T & \mathbf{R}_f \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \tag{3.9}$$

where  $\mathbf{p}$  is the vector of pressures at each junction,  $\mathbf{I}$  is the identity matrix,  $\mathbf{R}_f = \text{diag}(r_f)$  is the diagonal matrix of hydraulic resistances for each connection, and  $\mathbf{b} = \Delta \mathbf{d} + (\mathbf{I} - \Delta) \mathbf{p}_0$ . Here,  $\mathbf{p}_0 = P_i$  for each leaf junction. Solving this system provides the flow rate  $\mathbf{q}$ , which we use to define the flow rate matrix  $\mathbf{Q} = \text{diag}(\mathbf{q})$ .

Since we are only considering distribution networks without pumps, it is not possible to have a directed cycle, which is a path of arcs and vertices where a vertex is reachable starting from itself, in the directed graph. Since the fluid flows from higher pressure to lower pressure, a directed cycle in the network would require the sequence of pressures at each vertex in the path satisfy

$$P_0 > P_1 > \dots > P_j > \dots > P_N. \tag{3.10}$$

Since we assume we have a directed cycle,  $P_N = P_0$ , which implies that  $P_0 > P_0$ . Hence, we cannot have a directed cycle.

### 3.2 Pathogen concentration model

Similar to the single connection model, we will partition the pathogen concentration within each connection into two compartments: bulk fluid and biofilm bound. For each junction, we assume that the biofilm compartment is negligible and only assume a bulk fluid compartment. For each compartment, we assume the concentration of pathogens

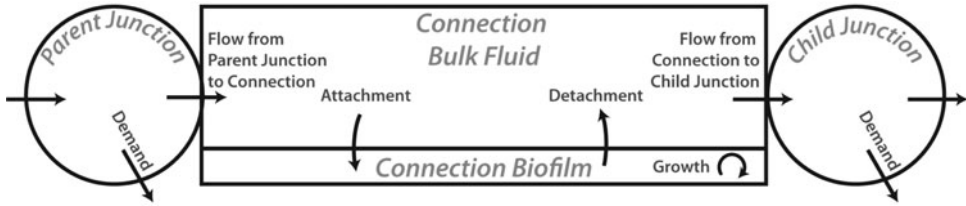


FIGURE 5. Compartmental network model of pathogen dynamics for a single connection and two junctions.

change due to flow in from the parent junction, flow out to the child junction, and attachment/detachment from the biofilm. We assume no growth in the bulk fluid (see Figure 5). For the biofilm compartment in the connection, we assume attachment/detachment from the bulk fluid and logistic growth. For the junctions, we assume the pathogen concentration only changes due to flow in from parent connections, flow out to child connections, and flows out of the system due to demands at the junctions. The governing equations are

$$\begin{aligned}
 \mathbf{u}' &= \mathbf{Z}\mathbf{c} - \mathbf{Q}\mathbf{u} + \mathbf{f}(\mathbf{u}, \mathbf{v}), \\
 \mathbf{c}' &= \mathbf{W}\mathbf{u} - (\mathbf{T} + \mathbf{D})\mathbf{c}, \\
 \mathbf{v}' &= \mathbf{g}(\mathbf{u}, \mathbf{v}),
 \end{aligned}
 \tag{3.11}$$

where  $\mathbf{u}$  is the vector of the pathogen concentration within the bulk fluid for each connection,  $\mathbf{v}$  is the vector of the pathogen concentration within the biofilm for each connection, and  $\mathbf{c}$  is the vector of the pathogen concentration within the fluid in the interior junctions. The vector functions  $\mathbf{f}(\mathbf{u}, \mathbf{v})$  and  $\mathbf{g}(\mathbf{u}, \mathbf{v})$  describe the attachment/detachment/growth dynamics of the pathogens. They are diagonal operators and are defined as

$$\mathbf{f}(\mathbf{u}, \mathbf{v}) = \mathbf{\Gamma}^{-1}\mathbf{B}\mathbf{v} - \mathbf{A}\mathbf{u}
 \tag{3.12}$$

and

$$\mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{\Gamma}\mathbf{A}\mathbf{u} + (\mathbf{R} - \mathbf{B})\mathbf{v} + \mathbf{Y}^{-1}\mathbf{R}(\mathbf{v} \circ \mathbf{v}),
 \tag{3.13}$$

where the  $\mathbf{v} \circ \mathbf{v}$  indicates the Hadamard product of the vectors. The matrices  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{\Gamma}$ ,  $\mathbf{R}$ , and  $\mathbf{Y} \in \mathbb{R}^{M \times M}$  are all diagonal matrices. For the  $i^{\text{th}}$  connection,  $A_{i,i} = \alpha_i$  is the pathogen attachment rate from the bulk fluid to the biofilm,  $B_{i,i} = \beta_i$  is the pathogen detachment rate from the biofilm to the bulk fluid,  $\Gamma_{i,i} = \gamma_i = V_i/S_{A_i}$  is the bulk fluid volume to connection surface area ratio,  $Y_{i,i} = K_i$  is the carrying capacity of the pathogens within the biofilm, and  $R_{i,i} = r_i$  is the pathogen growth rate in the biofilm. Each are considered to be constant and positive so each diagonal matrix is invertible.

The matrices  $\mathbf{Z}$  and  $\mathbf{W}$  indicate the flow into the connections and junctions, respectively. The matrix  $\mathbf{Q}$  is a diagonal matrix that denotes the flow rate through the connections into the child junctions with  $Q_{i,i} > 0$ . The matrices  $\mathbf{T}$  and  $\mathbf{D}$  denote the flows out of the junction into the child connections and demands at each junction, respectively. The inflow

matrices are defined by

$$\mathbf{Z} = \mathbf{Q} (\mathbf{A}_I^T < 0) \in \mathbb{R}^{M \times N_I} \quad \text{and} \quad \mathbf{W} = (\mathbf{A}_I > 0) \mathbf{Q} \in \mathbb{R}^{N_I \times M}, \tag{3.14}$$

and the outflow matrices are  $\mathbf{T} = (\mathbf{A}_I < 0) \mathbf{Q} (\mathbf{A}_I^T < 0) \in \mathbb{R}^{N_I \times N_I}$  and  $\mathbf{D} \in \mathbb{R}^{N_I \times N_I}$ , which is a diagonal matrix where  $\mathbf{D}_{i,j} = D_i$  is the demand at the  $i^{\text{th}}$  junction. Note that the notation  $\mathbf{M} = (\mathbf{A} < 0)$  indicates

$$\mathbf{M}_{i,j} = \begin{cases} 1 & \text{if } \mathbf{A}_{i,j} < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

Following EPANET’s model of mixing at junctions (Rossman, 2000), we assume that under realistic flow conditions the pathogen concentration within the junctions respond quickly to changes in concentration so we use a quasi-steady-state assumption,  $c' \approx 0$ , to obtain

$$c \approx (\mathbf{T} + \mathbf{D})^{-1} \mathbf{W}u. \tag{3.16}$$

Substituting (3.16) into (3.11), we obtain the quasi-steady-state network model for the pathogen concentrations within the connections:

$$\begin{aligned} u' &= (\mathbf{K} - \mathbf{Q})u + f(u, v), \\ v' &= g(u, v), \end{aligned} \tag{3.17}$$

where the matrix  $\mathbf{K} = \mathbf{Z}(\mathbf{T} + \mathbf{D})^{-1} \mathbf{W} \in \mathbb{R}^{M \times M}$  denotes the connection-to-connection inflows. It is a singular matrix with a zero diagonal since we have assumed that there are no flow loops in the digraph.

#### 4 Linear stability of the trivial steady state for the network model

To understand the long-time behaviour of the pathogen concentration in the entire network, we examine the linear stability of the trivial (washout) steady state where the pathogen concentration goes to zero in all the connections. We linearize the non-linear system (3.17) about  $u = v = 0$  to obtain the linear system  $x' = \mathbf{J}x$  where  $x = [u, v]^T$  and

$$\mathbf{J} = \begin{bmatrix} \mathbf{K} - \mathbf{Q} - \mathbf{A} & \Gamma^{-1} \mathbf{B} \\ \Gamma \mathbf{A} & \mathbf{R} - \mathbf{B} \end{bmatrix} \tag{4.1}$$

is the Jacobian matrix for the non-linear system linearized around the trivial steady state. The stability of the trivial steady state depends on the real part of the eigenvalues of the Jacobian matrix  $\mathbf{J}$ .

To predict the linear stability of the trivial steady state for a network with  $M$  connections, we need to compute all  $2M$  eigenvalues of  $\mathbf{J}$ . Using QR decomposition to approximate the eigenvalues would require  $\mathcal{O}(8M^3)$  operations, which could be computationally inefficient and the computation could suffer from roundoff errors for large distribution systems. To reduce the computational cost, we use the structure of the

network and the Jacobian matrix to develop an efficient approach that we prove predicts the linear behaviour of the system near the trivial steady state.

We begin by noting that all the matrices in  $\mathbf{J}$  are diagonal except for the matrix  $\mathbf{K}$ , which describes the inflows into each connection. The matrix  $\mathbf{K}$  will have a zero diagonal since we assume that the network has no self-loops. Since there are no directed cycles as shown in Section 3.1, the digraph will have a topological ordering, which is an ordering of the vertices such that for every arc from vertex  $u$  to vertex  $v$ , the vertex  $u$  will appear in the ordering before  $v$ . We can subsequently order the arcs so that every arc incident into vertex  $v$  appears in the sequence before all arcs incident out of  $v$ . For an arbitrary digraph with no directed cycles, we can compute a topological ordering using Kahn (1962) algorithm. From the topological sorting, we prove the following:

**Theorem 4.1** *There exists a permutation matrix  $\mathbf{P} \in \mathbb{R}^{M \times M}$  that transforms the inflow matrix  $\mathbf{K}$  into a lower triangular matrix.*

**Proof** For an arbitrary ordering of arcs in the digraph, we apply a modified version of Kahn (1962) algorithm that sorts arcs instead of vertices. Let  $\mathcal{S}$  be the set of all arcs whose tail vertex is a source (a vertex that is not the head of any arc in the digraph) and let  $v_j$  represent the head of  $a_j$ . Remove an arc  $a_j$  from  $\mathcal{S}$  and add it to the end of the ordered set  $\mathcal{T}$ . Now remove the  $a_j$  from the digraph and if  $v_j$  is now a source vertex, then add all of the arcs that incident out of  $v_j$  and add them to  $\mathcal{S}$ . Remove the next arc from  $\mathcal{S}$  and repeat until  $\mathcal{S}$  is empty. Since the digraph has no directed cycles, every arc in the digraph will be placed in a topological ordering within the set  $\mathcal{T}$ . The permutation matrix, defined by

$$\mathbf{P}_{i,j} = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ element of } \mathcal{T} \text{ is the arc } a_j \\ 0 & \text{otherwise,} \end{cases} \quad (4.2)$$

transforms a matrix such that elements are in a topological ordering. Let  $\mathbf{L} = \mathbf{PK}$ . Since  $\mathbf{K}_{i,j}$  denotes the flow from arc  $a_j$  into  $a_i$ , the head vertex of  $a_j$  is the tail vertex of  $a_i$ . It follows that if  $\mathbf{L}_{m,n} \neq 0$ , when  $n > m$  then the head vertex of arc  $a_n$  is the tail vertex of arc  $a_m$ . This contradicts the topological ordering. Hence,  $\mathbf{L}$  is a lower triangular matrix.  $\square$

We will now assume that  $\mathbf{K}$  is a lower triangular matrix with zero diagonal elements without loss of generality.

Next, we prove that the eigenvalues of  $\mathbf{J}$  are real and are independent of the matrix  $\mathbf{K}$ .

**Theorem 4.2** *The eigenvalues of  $\mathbf{J}$  given in (4.1) are real and independent of  $\mathbf{K}$ .*

**Proof** The eigenvalues  $\lambda$  of  $\mathbf{J}$  satisfy the matrix equation:

$$(\mathbf{J} - \lambda\mathbf{I})\mathbf{x} = 0, \quad (4.3)$$

where  $\mathbf{x} = [\mathbf{u}, \mathbf{v}]^T$ . This is equivalent to

$$(\mathbf{K} - \mathbf{Q} - \mathbf{A} - \lambda\mathbf{I})\mathbf{u} + \Gamma^{-1}\mathbf{B}\mathbf{v} = 0, \tag{4.4}$$

$$\Gamma\mathbf{A}\mathbf{u} + (\mathbf{R} - \mathbf{B} - \lambda\mathbf{I})\mathbf{v} = 0, \tag{4.5}$$

so solving for  $\mathbf{u}$ , (4.5) yields

$$\mathbf{u} = \mathbf{A}^{-1}\Gamma^{-1}(\mathbf{R} - \mathbf{B} - \lambda\mathbf{I})\mathbf{v}, \tag{4.6}$$

since  $\mathbf{A}$  and  $\Gamma$  are invertible.

We now substitute the value of  $\mathbf{u}$  from (4.6) into (4.4) to get

$$(\mathbf{K} - \mathbf{Q} - \mathbf{A} - \lambda\mathbf{I})\mathbf{A}^{-1}\Gamma^{-1}(\mathbf{R} - \mathbf{B} - \lambda\mathbf{I})\mathbf{v} - \Gamma^{-1}\mathbf{B}\mathbf{v} = 0. \tag{4.7}$$

Since all the matrices except for  $\mathbf{K}$  are diagonal matrices, we can simplify (4.7) to obtain

$$[(\mathbf{K} - \mathbf{Q} - \mathbf{A} - \lambda\mathbf{I})(\mathbf{R} - \mathbf{B} - \lambda\mathbf{I}) - \mathbf{A}\mathbf{B}]\mathbf{v} = 0. \tag{4.8}$$

The eigenvalue problem is now of the form  $\mathbf{L}\mathbf{v} = 0$ , where

$$\mathbf{L} = (\mathbf{K} - \mathbf{Q} - \mathbf{A} - \lambda\mathbf{I})(\mathbf{R} - \mathbf{B} - \lambda\mathbf{I}) - \mathbf{A}\mathbf{B} \tag{4.9}$$

is a lower triangular matrix since  $\mathbf{K}$  is lower triangular and the other matrices are diagonal. Thus, for the null space not to be empty,  $\mathbf{L}$  must be singular so  $\det(\mathbf{L}) = 0$ . This implies that

$$\prod_{i=1}^N L_{i,i} = 0, \tag{4.10}$$

so

$$\prod_{i=1}^N (q_i + \alpha_i + \lambda)(r_i - \beta_i - \lambda) + \alpha_i\beta_i = 0, \tag{4.11}$$

since  $\mathbf{K}_{i,i} = 0$ ,  $\mathbf{R}_{i,i} = r_i$ ,  $\mathbf{B}_{i,i} = \beta_i$ ,  $\mathbf{Q}_{i,i} = q_i$ , and  $\mathbf{A}_{i,i} = \alpha_i$ . This yields the characteristic equations

$$(q_i + \alpha_i + \lambda)(r_i - \beta_i - \lambda) + \alpha_i\beta_i = 0, \tag{4.12}$$

for  $i = 1, 2, \dots, M$  and if  $\lambda$  satisfies (4.12) for any  $i$ , then  $\lambda \in \sigma(\mathbf{J})$ . The characteristic equations do not depend on any elements of  $\mathbf{K}$  so the eigenvalues are independent of  $\mathbf{K}$ . The determinant of the quadratic equation (4.12) is

$$(r_i - \beta_i - Q_i - \alpha_i)^2 - 4[(Q_i + \alpha_i)(r_i - \beta_i) + \alpha_i\beta_i] = (r_i - \beta_i + Q_i + \alpha_i)^2 + 4\alpha_i\beta_i > 0, \tag{4.13}$$

so all  $\lambda \in \sigma(\mathbf{J})$  are real. □

Since the communication between connections in the network occurs through the matrix  $\mathbf{K}$ , Theorem 4.2 proves that the linear stability of the trivial steady state of the network is the same as the linear stability of the trivial steady state when each connection is viewed in isolation outside of the network. This allows us to examine the connections independently

and we are not required to examine the linear stability for the entire network. If one connection in isolation has a linearly unstable trivial steady state, then the network will have a linearly unstable trivial steady state and if all connections have a linearly stable trivial steady state in isolation, then the trivial steady state for the entire network will be linearly stable.

To proceed, we will need the following definitions. The set of children of the  $i^{\text{th}}$  arc,  $a_i$ , in the digraph is defined as

$$\mathcal{C}(a_i) = \{a_j \mid \text{the tail of } a_j = \text{the head of } a_i\}, \tag{4.14}$$

the set of parents of the  $j^{\text{th}}$  arc as

$$\mathcal{P}(a_j) = \{a_i \mid \text{the head of } a_i = \text{the tail of } a_j\}. \tag{4.15}$$

The set of all descendants of the  $i^{\text{th}}$  arc is

$$\mathcal{D}(a_i) = \{a_j \mid a_j \in \mathcal{C}(a_i) \vee a_j \in \mathcal{D}(a_k) \text{ for any } a_k \in \mathcal{C}(a_i)\}, \tag{4.16}$$

and the set of ancestors of the  $j^{\text{th}}$  arc is defined as

$$\mathcal{A}(a_j) = \{a_i \mid a_j \in \mathcal{D}(a_i)\}. \tag{4.17}$$

These sets satisfy the following relations:

$$\mathcal{C}(a_i) \subseteq \mathcal{D}(a_i), \quad \mathcal{P}(a_i) \subseteq \mathcal{A}(a_i), \quad \mathcal{D}(a_i) \cap \mathcal{A}(a_i) = \emptyset. \tag{4.18}$$

The last relation denotes that there are no directed cycles in the digraph as demonstrated in Section 3.1. Since the elements of  $\mathbf{K}$  indicate the inflows into each connection, we note that the element  $K_{i,j} \neq 0$  implies that the arc  $a_j$  is a parent of the arc  $a_i$  ( $a_j \in \mathcal{P}(a_i)$ ).

We now prove the eigenvectors of the Jacobian matrix  $\mathbf{J}$  depend on the matrix  $\mathbf{K}$  and the non-zero elements of the eigenvectors indicate the descendants of individual connections.

**Theorem 4.3** *The eigenvectors of  $\mathbf{J}$  given by (4.1) depend on the matrix  $\mathbf{K}$  and the non-zero elements of the  $i^{\text{th}}$   $\mathbf{v}$ -eigenvector correspond to all the descendants of the  $i^{\text{th}}$  connection.*

**Proof** To find the eigenvectors of  $\mathbf{J}$ , we solve  $\mathbf{L}\mathbf{v} = 0$ . Notice that  $\mathbf{L}$  can be written as  $\mathbf{L} = \mathbf{C}\mathbf{K} + \mathbf{D}$ , where  $\mathbf{K}$  is a lower triangular with zero diagonal matrix and  $\mathbf{C}$  and  $\mathbf{D}$  are a diagonal matrices with entries

$$C_i = r_i - \beta_i - \lambda \tag{4.19}$$

and

$$D_i = -[(q_i + \alpha_i + \lambda)(r_i - \beta_i - \lambda) + \alpha_i\beta_i], \tag{4.20}$$

so that

$$\mathbf{L} = \begin{bmatrix} D_1 & 0 & \cdots & \cdots & \cdots & 0 \\ C_2K_{2,1} & D_2 & 0 & \cdots & \cdots & 0 \\ C_3K_{3,1} & C_3K_{3,2} & D_3 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & D_{N-1} & 0 \\ C_MK_{M,1} & \ddots & \ddots & \ddots & C_MK_{M,M-1} & D_M \end{bmatrix}. \tag{4.21}$$

There will be three different cases based on the algebraic and geometric multiplicity of the eigenvalues  $\lambda_i$ . We will assume that all the eigenvalues have multiplicity of one, so  $\lambda_i \neq \lambda_j$ , if  $i \neq j$  for all  $i, j = 1, 2, \dots, 2M$ . Also, we will assume that  $C_i \neq 0$  for  $i = 1, 2, \dots, M$ . The proof for higher multiplicities or when  $C_i = 0$  for some  $i$  follows the same arguments given below.

We start by choosing  $\lambda_1$  such that  $D_1 = 0$ . This implies that  $v_1 = z$  is an arbitrary non-zero constant. Looking that the second row, we have

$$C_2K_{2,1}v_1 + D_2v_2 = 0, \tag{4.22}$$

so

$$v_2 = -\frac{C_2K_{2,1}}{D_2}v_1 = -\frac{C_2K_{2,1}}{D_2}z, \tag{4.23}$$

since  $D_2 \neq 0$  under the assumption  $\lambda_2 \neq \lambda_1$ . Note that  $v_2 = 0$  if  $K_{2,1} = 0$ , which implies that  $a_1 \notin \mathcal{P}(a_2)$ .

For the third row, we have

$$C_3(K_{3,1}v_1 + K_{3,2}v_2) + D_3v_3 = 0, \tag{4.24}$$

so that

$$v_3 = -\frac{C_3}{D_3}(K_{3,1}v_1 + K_{3,2}v_2) = -\frac{C_3}{D_3}\left(K_{3,1} - \frac{C_2K_{3,2}K_{2,1}}{D_2}\right)z. \tag{4.25}$$

To determine if  $v_3 = 0$ , we examine the values of  $K_{3,1}$ ,  $K_{3,2}$ , and  $K_{2,1}$ . If  $a_1 \notin \mathcal{P}(a_3)$  and either  $a_2 \notin \mathcal{P}(a_3)$  or  $a_1 \notin \mathcal{P}(a_2)$ , we have  $v_3 = 0$  since  $K_{3,1} = K_{3,2}K_{2,1} = 0$ . This implies that  $a_1 \notin \mathcal{A}(a_3)$  and there is no directed path from  $a_1$  to  $a_3$ . If either  $a_1 \in \mathcal{P}(a_3)$  or  $a_2 \in \mathcal{P}(a_3)$  and  $a_1 \in \mathcal{P}(a_2)$ , we have  $v_3 \neq 0$  and  $a_1 \in \mathcal{A}(a_3)$ . Note that it is not possible for both  $K_{3,1} \neq 0$  and  $K_{3,2}K_{2,1} \neq 0$  since that implies that the arc  $a_1$  is a parent to arc  $a_2$  and  $a_3$  as well as implying  $a_2$  is a parent to  $a_3$ , making  $a_1$  a grandparent of  $a_3$ . This is a contradiction since an arc cannot be a parent and grandparent to a single arc without the digraph having a self-loop.

For any row  $i > 3$ , we have

$$C_i \sum_{j=1}^{i-1} K_{i,j}v_j + D_iv_i = 0. \tag{4.26}$$

Solving (4.26) for  $v_i$  yields the recursive relation

$$v_i = -\frac{C_i}{D_i} \sum_{j=1}^{i-1} K_{i,j} v_j = ac, \tag{4.27}$$

where  $a$  is a constant that depends on  $\mathbf{K}$ ,  $\mathbf{C}$ , and  $\mathbf{D}$ . If the  $j^{\text{th}}$  connection is a parent of the  $i^{\text{th}}$  connection,  $j \in \mathcal{P}(i)$ , then  $K_{i,j} \neq 0$ . Likewise, if  $j \notin \mathcal{P}(i)$ , then  $K_{i,j} = 0$ . Hence, (4.27) reduces to

$$v_i = -\frac{C_i}{D_i} \sum_{\substack{j=1 \\ j \in \mathcal{P}(i)}} K_{i,j} v_j, \tag{4.28}$$

and, by induction,  $v_i = 0$  if  $a_1 \notin \mathcal{A}(a_i)$ . As a consequence, the non-zero entries of  $\mathbf{v}$  correspond to the descendants of the first connection.

For the  $m^{\text{th}}$  eigenvalue,  $\lambda_m$ , with  $m > 1$ , we assume the ordering of the eigenvalues are such that  $D_m = 0$  when  $\lambda = \lambda_m$  for  $m = 1, 2, \dots, N$ . Note that we still are assuming the algebraic multiplicity of the eigenvalues are one so  $D_i \neq 0$  for  $i \neq m$ . From the first row, we have  $D_1 v_1 = 0$  so  $v_1 = 0$  since  $D_1 \neq 0$ . Likewise, we have

$$C_2 K_{2,1} v_1 + D_2 v_2 = 0, \tag{4.29}$$

from the second row, so  $v_2 = 0$ , since  $v_1 = 0$  and  $D_2 \neq 0$ . We now assume  $v_j = 0$  for  $j = 1, \dots, i - 1$ . We have

$$C_i \sum_{j=1}^{i-1} K_{i,j} v_j + D_i v_i = 0 \tag{4.30}$$

for the  $i^{\text{th}}$  row with  $i < m$ . Since  $v_j = 0$  for  $j = 1, \dots, i - 1$  and  $D_i \neq 0$ , we have, by induction,  $v_i = 0$  for  $i < m$ .

We now examine the  $m^{\text{th}}$  row of  $\mathbf{L}$ :

$$C_m \sum_{j=1}^{m-1} K_{m,j} v_j + D_m v_m = 0. \tag{4.31}$$

Since  $v_j = 0$  for  $j = 1, \dots, m - 1$  and  $D_m = 0$ , we have  $v_m = c$ , where  $c$  is an arbitrary non-zero constant. We proceed as above when  $\lambda = \lambda_1$  to obtain that the non-zero entries of  $m^{\text{th}}$   $\mathbf{v}$  correspond to the ancestors of the  $m^{\text{th}}$  connection. □

### 5 Algorithm for predicting long-time behaviour in the network

To determine the stability of the trivial steady state for the entire network, one could linearize the system and compute the  $2M$  eigenvalues of the Jacobian matrix. If any eigenvalue has a positive real part, the pathogens could persist within the network, based upon the initial pathogen distribution within the network. For large networks, the eigenvalues can be computed using a method such as QR factorization, which is  $\mathcal{O}(8N^3)$  for this problem. For efficiency, we can use the results of Theorems 4.2 and 4.3 to reduce the effort in determining the eigenvalues and, hence, the stability of the trivial steady state to  $\mathcal{O}(N)$ .



We start by noting that the eigenvalues for the entire linearized network model and the eigenvalues of the linearized model of the individual connections in isolation are the same from Theorem 4.2. Hence, the trivial steady state for the entire network model is unstable if the trivial steady state for any individual connection is unstable—if the pathogens would persist for an individual connection, it will persist within the network.

We can additionally determine the distribution of the pathogens within the entire network using the eigenvectors computed for each individual connection. Let  $\mathbf{A}_p$  denotes the parent incidence matrix defined by the relation

$$\mathbf{A}_{p_{i,j}} = \begin{cases} 1 & \text{if the arc } a_i \in \mathcal{P}(a_j) \\ 0 & \text{otherwise.} \end{cases} \quad (5.1)$$

From Theorem 4.3, if the pathogens within a connection persist, the eigenvector will contain a non-negative entry for each child of that connection. This implies that the persistence of the pathogens within the parent causes the pathogens to persist within each child. When the pathogens would washout of a child connection in isolation but the child has an ancestor where the pathogens persist, the pathogen concentration within the child approach a non-zero steady state. We say that the child is ‘induced’ in this situation.

To summarize, each connection in isolation can be labelled as persistent or washout based on the bifurcation conditions for the individual connection (see Section 2). If any connection is classified as persistent, then the pathogens can persist within the network and if all connections are classified as washout, then the pathogens washout of the network. When the pathogens within a connection are persistent, all its descendants are either persistent (if they ‘persist’ in isolation) or they are induced (if they are ‘washout’ in isolation).

We can use the parent incidence matrix,  $\mathbf{A}_p$ , iteratively to determine which connections within the network are induced. The algorithm is as follows:

- (1) Compute the stability of the trivial steady state for each connection in isolation. Classify each connection as ‘persist’ or ‘washout’ using the bifurcation conditions.
- (2) Define the elements of the vector  $s^{(0)}$  as  $s_i^{(0)} = 1$  if the  $i^{\text{th}}$  connection is ‘persistent’ and  $s_i^{(0)} = 0$  if the  $i^{\text{th}}$  connection is ‘washout’.
- (3) Define the iteration matrix  $\mathbf{Z} = \mathbf{I} + \mathbf{A}_p^T$ .
- (4) Set  $s = s^{(0)}$ .
- (5) Repeat
  - (a)  $s = \mathbf{Z}s$ .
  - (b) If  $s_i^{(0)} = 1$ , then set  $s_i = 1$  for all  $i = 1, \dots, N$ .
  - (c) If  $s_i^{(0)} = 0$  and  $s_i > 0$ , then set  $s_i = 2$  for all  $i = 1, \dots, N$ .

Until  $\|s - s^{(0)}\| = 0$ .

After this algorithm is completed, the vector  $s$  indicates the connection is ‘persist’ when  $s_i = 1$ , ‘induced’ when  $s_i = 2$ , and ‘washout’ when  $s_i = 0$ .

Note that this approach indicates which connections are the biggest concerns since a connection that is labelled ‘persist’ can cause the pathogens to appear in all its descendants and if the pathogens were removed from all the ‘persist’ connections by some external means, then all the descendants would have the pathogens washout. This could aid in the efficient removal of pathogens from the network.

Using this algorithm, we effectively compute the result predicted by computing both the eigenvalues and eigenvectors of the linearized system. Since this algorithm using iterative matrix-vector multiplication, it is an  $\mathcal{O}(mN^2)$  algorithm, where  $m < N$  is the number of iterations and  $N$  is the number of connections in the network. Hence, computing both the linear stability and the predicted locations of contamination is  $\mathcal{O}(mN^2)$  complexity while using a method such as QR factorization is  $\mathcal{O}(8N^3)$  for both eigenvalues and eigenvectors. For a more detailed discussion of efficiency and timings of a serial implementation of the algorithms, see Appendix A.

## 6 Solutions for example networks

To validate the theorems and the algorithm given in Section 5, we examine some simple networks. Note that these networks are caricatures of real-world networks and were generated for testing and illustrative purposes.

### 6.1 Medium-sized network example

The first is a moderately sized network that consists of  $M = 11$  junctions and  $N = 11$  connections. The topology of the network and the flow through the network is shown in Figure 6 and the system parameters are given in Appendix B.

We determine the linear stability of each connection independently of the network and compare it to the steady state numerical solution. Since the linear stability assumes a general perturbation from the trivial steady state, we introduce pathogens into the bulk fluid in all connections and no pathogens within the biofilm in any connections.

Figure 7 shows the linear stability of the trivial steady state for each connection using the algorithm discussed in Section 5. The dashed green on black lines indicate connections where the linear stability analysis predicts the pathogens to washout. The red connections indicate the connections where the pathogens are predicted to persist and the yellow dotted lines on black denote connections that are predicted to be induced by upstream connections to have a non-zero pathogen concentration as  $t \rightarrow \infty$ . We can see that of the 11 connections, there are eight (4–11) that would have what non-zero pathogen concentrations for long times. Of these eight, four (4, 5, 6, and 11) would persist outside of the network and the other four connections (7, 8, 9, and 10) would have the pathogens washout if they were not in the network. This suggests that a possible approach to disinfection would be to focus on the four persistent connections (4, 5, 6, and 11). If the pathogens were removed from these connections, the other four connections (7, 8, 9, and 10) would have the pathogens washout due to the fluid flow. This observation suggests that connections 4, 5, 6, and 11 are possible vulnerable points within the network.

To confirm the long-time behaviour predicted by the linear stability analysis, we numerically simulated the entire network until a steady state was reached. Figure 8 shows

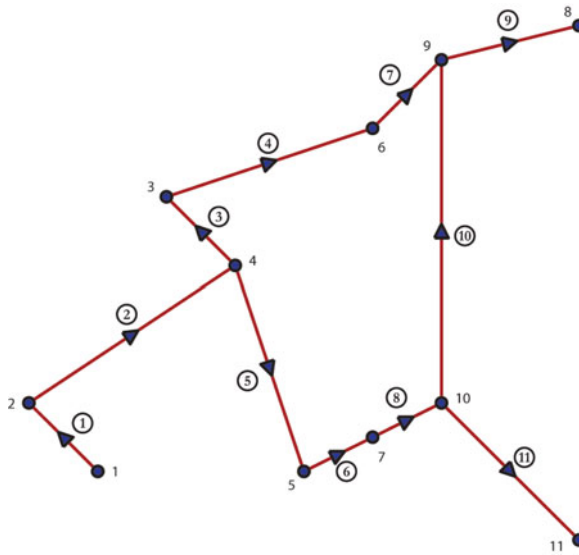


FIGURE 6. Topology and flow for the medium-sized network example. The lines indicate connections (pipes) and the filled circles indicate the junctions. The triangles indicate the direction of the flow through the connections. The circled numbers indicate the index of each connection and the uncircled numbers indicate the index assigned to each junction.

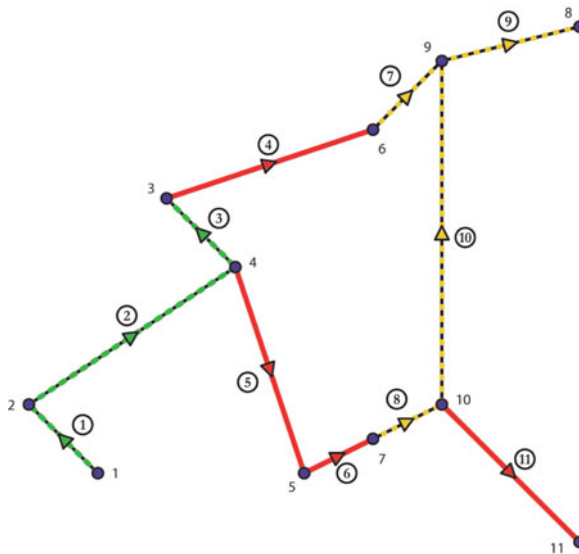


FIGURE 7. Linear stability of the trivial steady state for the medium-sized network example. The lines indicate the connections (pipes) and the circles indicate the junctions, which are numbered. The triangles indicate the direction of flow. The green dashed lines on black denote connections where the pathogens are predicted by the linear stability analysis to washout. The solid red lines indicate the connection where linear stability analysis predicts the pathogens persist and dotted yellow lines on black indicate the connections where the pathogens are predicted to be induced by upstream connections.

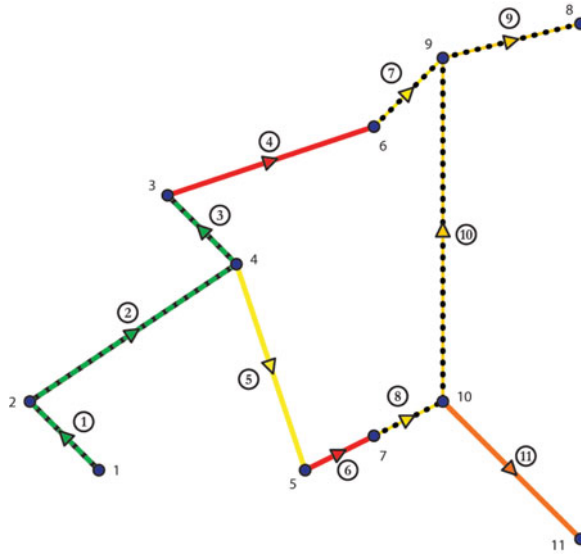


FIGURE 8. Steady-state numerical solution for the medium-sized network example. The lines indicate the connections (pipes) and the circles indicate the junctions, which are numbered. The triangles indicate the direction of flow. The dashed black-on-green lines denote connections where the pathogens will washout. Solid yellow, orange, or red lines indicate that the pathogens persist within that connection both in the network and in isolation. The dotted black-on-yellow/orange lines indicate the connections were induced by an upstream connection and have a non-zero pathogen concentration. The colour indicates the relative pathogen concentration as with the solid lines. The yellow, orange, and red colours indicate relatively low, intermediate, and high pathogen concentrations, respectively.

the steady-state numerical solution for this network. The dashed black-on-green lines indicate the pathogens are washed out of that connection, which is defined by their concentration going below a certain threshold. Here, the threshold is  $10^{-8}$ . The solid lines indicate that the pathogens persist within that connection (both in the network and in isolation) with the colour indicating a continuous mapping between low (yellow), intermediate (orange), and high (red) concentrations relatively to other connections in the network. The dotted black-on-colour lines indicate the connections were induced by an upstream connection and have a non-zero pathogen concentration with yellow indicating low relative concentration and orange intermediate. Both Figures 7 and 8 give equivalent results, even though connection 5 appears in different colours. To clarify this point, we emphasize that none of the ancestors of this connection is contaminated by pathogens which implies that the contamination is caused by persisting in isolation, and hence it appears with red colour in Figure 7. On the other hand, the pathogens that introduced into the fluid grow and are transported within the network and, as the network achieves the equilibrium, each connection has received a different amount of pathogens. The pathogens concentration within connection 5, which is labelled persistent, is lower than the concentration is connection 9, which is labelled induced.

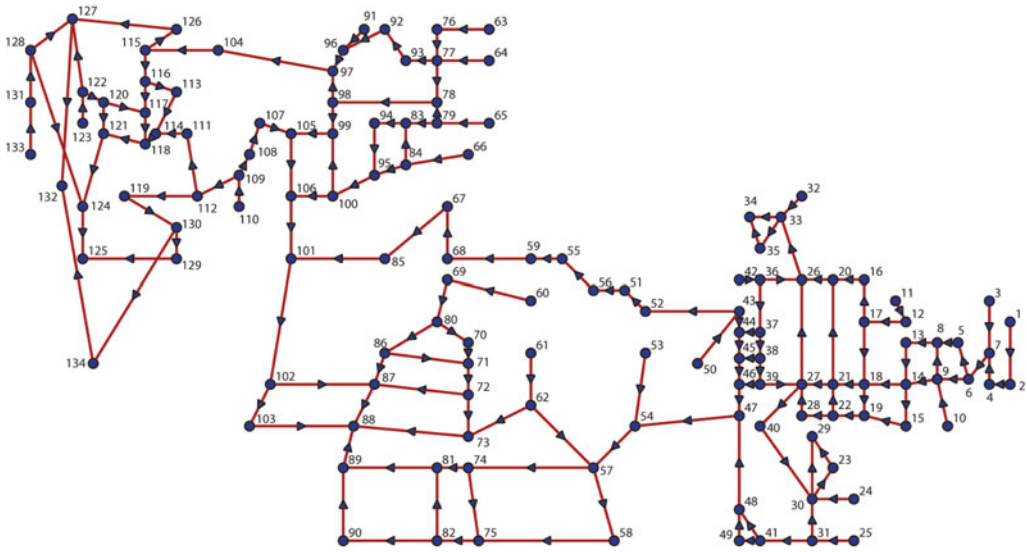


FIGURE 9. Topology and flow for the large network example. The lines indicate connections (pipes) and the circles indicate the junctions. The triangles indicate the direction of the flow through the connections. The numbers indicate the index assigned to each junction with each connection indexed sequentially starting for lower junctions to larger junctions.

## 6.2 Large network example

We present here an example to illustrate how our methodology scales to a larger water distribution network. Consider a water distribution network with  $M = 134$  junctions and  $N = 167$  connections. The network parameters given in Appendix C. The topology of the network and the flow through the network is shown in Figure 9.

We determine the linear stability of each connection independently of the network and compare it to the steady-state numerical solution. Since the linear stability assumes a general perturbation from the trivial steady state, we introduce pathogens into the bulk fluid in all connections and no pathogens within the biofilm in any connections.

Figure 10 shows the linear stability of the trivial steady state for each connection using the algorithm discussed in Section 5. The line style and colour scheme is the same as the one used in the figures in Section 6.1.

We can see that there are 30 connections where pathogens persist that induce another 52 connections to have non-zero pathogen concentrations. If one focuses the cleanup efforts on the 30 persistent connections, the other 52 connections could then be washed clean as long as there are not pathogens coming in from upstream. This observation could greatly reduce the cleanup costs from the pathogen introduction into a water distribution network and illustrates the possible vulnerable points within the network.

To confirm the long-time behaviour predicted by the linear stability analysis, we numerically simulated the entire network until a steady state was reached. Figure 11 shows the steady-state pathogen concentration within the network. Again, the line styles and colour schemes are the same as the ones used in Section 6.1.



FIGURE 10. Linear stability of the trivial steady state for the large network example. The lines indicate the connections (pipes) and the circles indicate the junctions, which are numbered. The triangles indicate the direction of flow. The green dashed lines on black denote connections where the pathogens are predicted by the linear stability analysis to washout. The solid red lines indicate the connection where linear stability analysis predicts the pathogens persist and dotted yellow lines on black indicate the connections where the pathogens are predicted to be induced by upstream connections.

We can see that Figures 10 and 11 are equivalent, indicating the algorithm described in Section 5 is able to predict the steady-state behaviour of the network, except for the connection between junctions 117 and 120. This is due to the different meanings of the colours in the figure. In the linear stability results, pathogens would persist in this connection when it is viewed in isolation and is labelled red since red indicates persist but in the numerical solution, this connection is labelled yellow. This does not indicate that this connection is induced by an upstream connection but it indicates that the pathogen concentration relative to the other connections within the network is lower.

## 7 Conclusions and future directions

The prevention and mediation of biological contamination of potable water in water distribution systems is an important aspect of maintaining public health. The interaction between non-native pathogens introduced into the system and the native drinking water biofilms can lead to increased biological retention times or possibly persistence leading to a 'reservoir of contamination' (Camper, 1998). In an attempt to study and predict when and where within the system pathogens could colonize the native biofilms, we have developed and analysed a mathematical model of the dynamics between non-native bacteria and native biofilms within large-scale water distribution networks under time-constant flow regimes. This model is able to predict when pathogens might colonize the network and critical components of the system that leads to persistence.

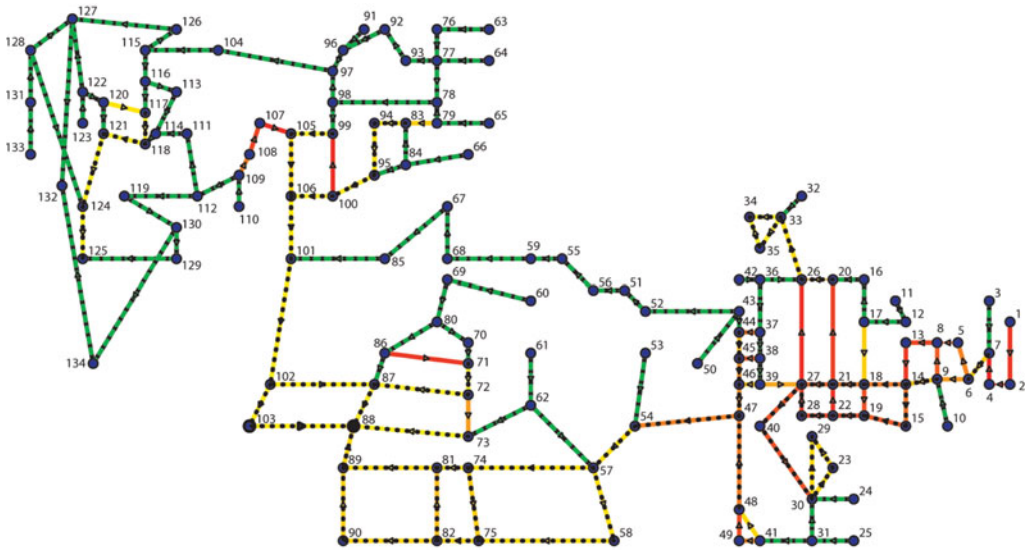


FIGURE 11. Steady-state numerical solution for the large network example. The lines indicate the connections (pipes) and the circles indicate the junctions, which are numbered. The triangles indicate the direction of flow. The dashed black-on-green lines denote connections where the pathogens will washout. Solid yellow, orange, or red lines indicate that the pathogens persist within that connection both in the network and in isolation. The dotted black-on-yellow/orange lines indicate the connections were induced by an upstream connection and have a non-zero pathogen concentration. The colour indicates the relative pathogen concentration as with the solid lines. The yellow, orange, and red colours indicate relatively low, intermediate, and high pathogen concentrations, respectively.

For large-scale systems of many connections, the approach of computing the Lyapunov exponents directly from the Jacobian matrix can be computationally expensive and subject to accumulation of round-off errors. To address this, we developed an efficient iterative approach to predicting the linear stability of the trivial steady state and we prove that the long-time behaviour of the linearized system is the same as predicted by the iterative algorithm. Results for example networks of different sizes are given and suggest that this approach can provide insight into vulnerable components of the distribution network. Since certain connections where pathogens might persist can cause pathogens to appear in downstream connections, the identification of critical connections by the algorithm could lead efficient mediation strategies where resources are directed towards disinfecting these critical connections. This would allow the pathogens to washout of downstream connections naturally.

This model and analysis are for time-constant flow networks, which do not model realistic networks in general. To model real-world networks, more accurate submodels need to be introduced such as time-varying flows within the network, models for water storage units such as water towers, flow reversal within connections from changing system demands, and pump-driven flow cycles within the network. This paper introduces a beginning framework for the modelling and analysis of these more complicated systems.

In the future, we will study these effects on the pathogen dynamics in the network as well as more sophisticated models for the fluid dynamics and the attachment and detachment processes.

### Acknowledgements

Research reported in this publication was partially supported by the Charles Phelps Taft Research Center.

### References

- ASHWIN, P., COOMBES, S. & NICKS, R. (2016) Mathematical frameworks for oscillatory network dynamics in neuroscience. *J. Math. Neurosci.* **6**(1), 2.
- BARRAT, A., BARTHELEMY, M. & VESPIGNANI, A. (2008) *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, U.K.
- BARTRAM, J. (2003) *Heterotrophic Plate Counts and Drinking-Water Safety: The Significance of HPCs for Water Quality and Human Health*, IWA Publishing, London, U.K.
- BESNER, M. C., GAUTHIER, V., SERVAIS, P. & CAMPER, A. (2002) Explaining the occurrence of coliforms in distribution systems. *J. Am. Water Works Ass.* **94**(8), 95–109.
- BRAUER, F., CASTILLO-CHAVEZ, C. & CASTILLO-CHAVEZ, C. (2012) *Mathematical Models in Population Biology and Epidemiology*, vol. 1, Springer, New York, USA.
- CAMPER, A. K. (1998) *Pathogens in Model Distribution System Biofilms*, American Water Works Association, USA.
- CAMPER, A. K. & MCFETERS, G. A. (2000) *Industrial Biofouling: Detection, Prevention and Control*, J. Walker, S. Surman & J. Jass (editors), Wiley, New York, USA.
- CAMPER, A. K., JONES, W. L. & HAYES, J. T. (1996) Effect of growth conditions and substratum composition on the persistence of coliforms in mixed-population biofilms. *Appl. Environ. Microbiol.* **62**(11), 4014.
- CARABEȚ, A., MIREL, I., FLORESCU, C., STĂNILOIU, C., GÎRBACIU, A. & OLARU, I. (2011) Drinking water quality in water-supply networks. *Environ. Eng. Manag. J.* **10**(11), 1659–1665.
- CARTER, J. T., RICE, E. W., BUCHBERGER, S. G. & LEE, Y. (2000) Relationships between levels of heterotrophic bacteria and water quality parameters in a drinking water distribution system. *Water Res.* **34**(5), 1495–1502.
- CASTELLANO, C., FORTUNATO, S. & LORETO, V. (2009) Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591.
- COSTERTON, J. W., LEWANDOWSKI, Z., CALDWELL, D. E., KORBER, D. R. & LAPPIN-SCOTT, H. M. (1995) Microbial biofilms. *Ann. Rev. Microbiol.* **49**(1), 711–745.
- DAVIS, M. L. (2011) *Water and Wastewater Engineering*, McGraw-Hill Higher Education, New York.
- EBERL, H., MORGENROTH, E., NOGUERA, D., PICIOREANU, C., RITTMANN, B., VAN LOOSDRECHT, M. & WANNER, O. (2006) *Mathematical Modeling of Biofilms*, IWA Publishing, London, U.K.
- EIGER, G., SHAMIR, U. & BEN-TAL, A. (1994) Optimal design of water distribution networks. *Water Resour. Res.* **30**(9), 2637–2646.
- FLEMMING, H. AND WINGENDER, J. (2010) The biofilm matrix. *Nat. Rev. Microbiol.* **8**(9), 623–633.
- KAHN, A. B. (1962) Topological sorting of large networks. *Commun. ACM* **5**(11), 558–562.
- KURAMOTO, Y. (1984) *Chemical Oscillations, Waves and Turbulence*, Springer, Berlin, Germany.
- LECHEVALLIER, M. W., WELCH, N. J. & SMITH, D. (1996) Full-scale studies of factors related to coliform regrowth in drinking water. *Appl. Environ. Microbiol.* **62**(7), 2201.
- LECHEVALLIER, M. W., GULLICK, R., KARIM, M., FRIEDMAN, M. & FUNK, J. (2003) The potential for health risks from intrusion of contaminants into the distribution system from pressure transients. *J. Water. Health* **1**(1), 3–14.



- LEHTOLA, M. J., LAXANDER, M., MIETTINEN, I. T., HIRVONEN, A., VARTIAINEN, T. & MARTIKAINEN, P. J. (2006) The effects of changing water flow velocity on the formation of biofilms and water quality in pilot distribution system consisting of copper or polyethylene pipes. *Water Res.* **40**(11), 2151–2160.
- MAYS, L. W. (2010) *Water Resources Engineering*, John Wiley & Sons, New York, USA.
- MO MBA, M. N. B., KFIR, R., VENTER, S. N. & CLOETE, T. E. (2000) An overview of biofilm formation in distribution systems and its impact on the deterioration of water quality. *Water S.A.* **26**(1), 59–66.
- MURRAY, R., UBER, J. & JANKE, R. (2006) Model for estimating acute health impacts from consumption of contaminated drinking water. *J. Water. Res. Pl.-ASCE* **132**(4), 293–299.
- PORTER, M. & GLEESON, J. (2016) *Dynamical Systems on Networks: A Tutorial*, vol. 4, Springer, Cham, Switzerland.
- POWER, K. N. & NAGY, L. A. (1999) Relationship between bacterial regrowth and some physical and chemical parameters within Sydney's drinking water distribution system. *Water Res.* **33**(3), 741–750.
- ROSSMAN, L. A. (2000) *EPANET 2: Users manual*, US Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory.
- ROSSMAN, L. A. & BOULOS, P. F. (1996) Numerical methods for modeling water quality in distribution systems: A comparison. *J. Water Res. Pl.-ASCE* **122**(2), 137–146.
- SHAMPINE, L. F. & REICHEL, M. W. (1997) The MATLAB ODE suite. *SIAM J. Sci. Comput.* **18**(1), 1–22.
- SKRABER, S., SCHIJVEN, J., GANTZER, C. & DE RODA HUSMAN, A. M. (2005) Pathogenic viruses in drinking-water biofilms: A public health risk? *Biofilms* **2**(02), 105–117.
- STROGATZ, S. H. (2000) From kuramoto to crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena* **143**(1–4), 1–20.
- SZABO, J. G. (2006) *Persistence of Microbiological Agents on Corroding Biofilm in a Model Drinking Water system Following Intentional Contamination*, PhD thesis, University of Cincinnati.
- TRULEAR, M. G. & CHARACKLIS, W. G. (1982) Dynamics of biofilm processes. *J. Water Pollut. Con. F.* **54**(9), 1288–1301.
- UBER, J., MURRAY, R. & JANKE, R. (2004) Use of systems analysis to assess and minimize water security risks. *J. Contemp. Water Res. Edu.* **129**(1), 34–40.
- VAN DER KOOIJ, D. (1992) Assimilable organic carbon as an indicator of bacterial regrowth. *J. Am. Water Works Ass.* **84**(2), 57–65.
- VAN LOOSDRECHT, M., HEIJNEN, J., EBERL, H., KREFT, J. & PICIOREANU, C. (2002) Mathematical modelling of biofilm structures. *A. Van Leeuw. J. Microb.* **81**(1), 245–256.
- VESPIGNANI, A. (2012) Modelling dynamical processes in complex socio-technical systems. *Nature Phys.* **8**(1), 32.
- WORLD HEALTH ORGANIZATION. (2002) *The world health report 2002: Reducing the risks, promoting healthy life*. Geneva, Switzerland.
- ZHANG, W. & DIGIANO, F. A. (2002) Comparison of bacterial regrowth in distribution systems using free chlorine and chloramine: A statistical study of causative factors. *Water Res.* **36**(6), 1469–1482.

## Appendix A Algorithm complexity and efficiency study

To demonstrate the efficiency of the proposed algorithms for computing the linear stability within the network as well as predicting locations of possible contamination, we ran a number of trial simulations on a rectangular water distribution network with randomly generated coefficients. Figure A 1 shows the topology and flow for a rectangular network

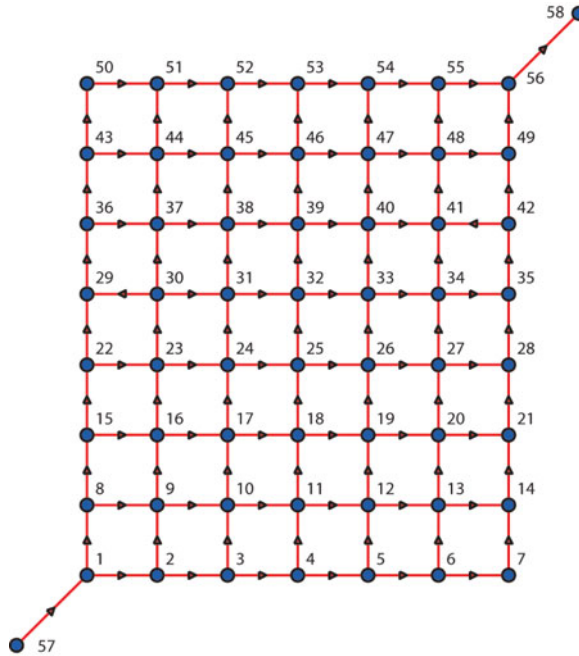


FIGURE A 1. Topology and flow for the medium-sized network example. The lines indicate connections (pipes) and the filled circles indicate the junctions. The triangles indicate the direction of the flow through the connections. The circled numbers indicate the index of each connection and the uncircled numbers indicate the index assigned to each junction.

with  $M = 7$  rows and  $N = 8$  columns of nodes plus a source and a sink. There are  $MN + 2 = 58$  nodes and  $2MN - (M + N) + 2 = 99$  connections. The parameters within each connection are chosen randomly from uniformly from the following ranges:  $\alpha \in (0, 1)$ ,  $\beta \in (0, 2)$ , and  $r \in (0, 1)$ . The volumes,  $V$ , and surface areas,  $S_A$ , of each connection is fixed at one as well as the pathogen carrying capacity,  $K$ , for simplicity. The system is driven by the fixed pressures  $P_{57} = 75$  at node 57 and  $P_{58} = 0$ . The demands at each junction are chosen randomly from a uniform distribution with the range  $D \in (0.5, 2.5)$  to allow the direction of flow in the network to change randomly.

To determine the efficiency of each algorithm, the initial network connectivity and flow is held constant for each trial of each algorithm. The linear stability and the predicted location of contamination are then computed 10 times and then averaged. Each simulation was performed on a single core of an Intel Core i7-970 processor with 24 GB of memory. The algorithms presented in this manuscript were implemented in MATLAB v8 (2012b) on Windows 7 SP1, while the direct computing of eigenvalues and eigenvectors was done using the built-in MATLAB eig function. Figure A 2 shows the results for networks with 200 to 1,500 connections. We can see that our algorithm for computing the eigenvalues of the system is significantly more efficient than direct computation. It is 1,700x faster than direct computation for a 200 connection network and shows a sub  $\mathcal{O}(N)$  complexity compared to  $\mathcal{O}(N^{2.49})$  for the MATLAB eig function. Note that differences

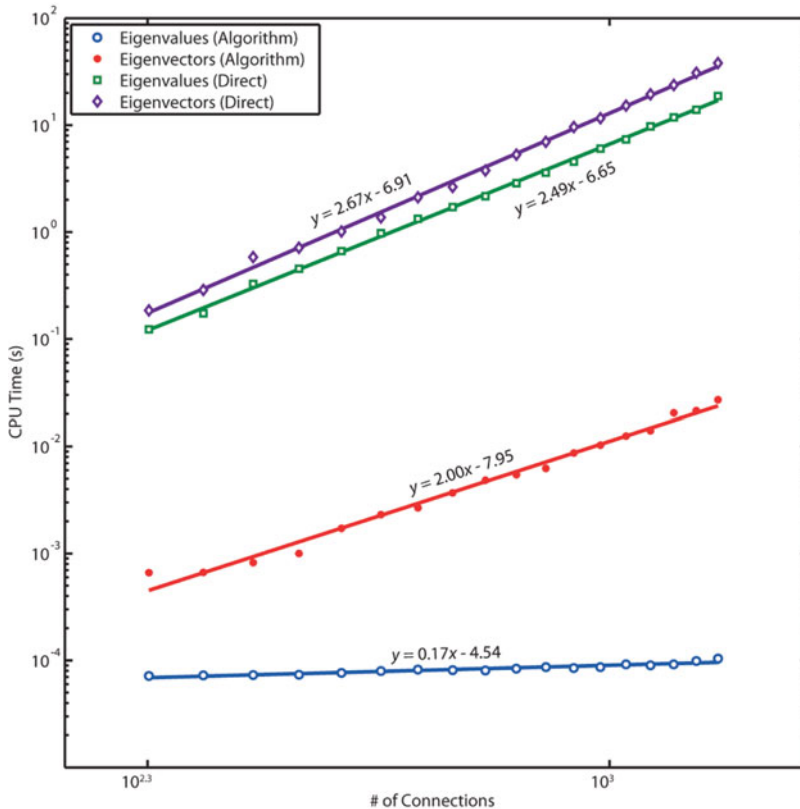


FIGURE A 2. Log-log plot of the clock time average for each algorithm. The markers indicate an average of 10 trials for different network sizes and the trend line is the linear least squares fit of the log of the data. The slope of each trend line is the exponent of the algorithmic complexity. The proposed algorithms are orders of magnitude faster than direct computation and also have a lower algorithmic complexity, which suggests that they will scale better as network size is increased.

between the theoretical and practical complexities is due to various optimizations within MATLAB. To compute the eigenvalues and the eigenvectors, our algorithm is 280x faster than direct computation for a 200 connection network and is complexity  $\mathcal{O}(N^2)$  compared to  $\mathcal{O}(N^{2.67})$  for MATLAB  `eig`  function. Note that our iterative algorithm only required 5–8 iterations to converge.

### Appendix B Parameters for moderate-sized network example

The medium network examples consists of  $M = 11$  junctions and  $N = 11$  connections as shown in Figure B 1. At the interior junctions, we assign a demand, which is zero. At each non-interior junction, we prescribe a pressure. For this example, we let  $P_1 = 29$ ,  $P_8 = 2$ , and  $P_{11} = 9$ . All the hydraulic resistances are set to one. We assume the parameters governing the pathogen dynamics are the same for each connection with  $r = 1$ , and  $K = 1$  with  $V = 1$  and  $S_A = 1$  except for  $\alpha$  and  $\beta$ , which are given in Table B 1.

Table B1. Attachment and detachment rates for each connection for the time-constant medium-sized network example

$i$	$\alpha_i$	$\beta_i$	$i$	$\alpha_i$	$\beta$	$i$	$\alpha_i$	$\beta_i$
1	1.00	6.0	5	0.01	0.8	9	5.00	5.0
2	0.01	11.0	6	1.00	0.3	10	1.00	2.0
3	1.40	9.0	7	1.00	2.0	11	1.00	2.0
4	0.01	0.2	8	1.00	2.0			

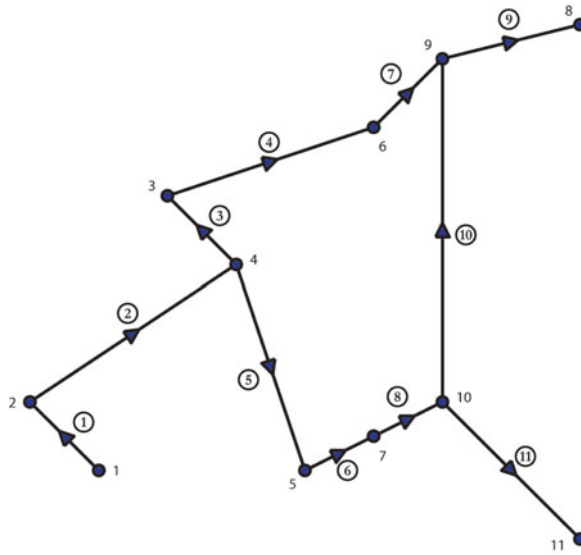


FIGURE B1. Connectivity for the time-constant flow medium network example.

### Appendix C Parameters for large network example

The large network examples consists of  $M = 134$  junctions and  $N = 167$  connections as shown in Figure C 1.

At the interior junctions, we assign a demand, which is zero except for the ones given in Table C 1. At each non-interior junction, we prescribe a pressure, which is given in Table C 2. All the hydraulic resistances are set to one. We assume the parameters governing the pathogen dynamics are the same for each connection and are  $\alpha = 1$ ,  $\beta = 2$ ,  $r = 1$ , and  $K = 1$  with  $V = 1$  and  $S_A = 1$ . This gives a critical flow rate of  $Q_c = 1$ .

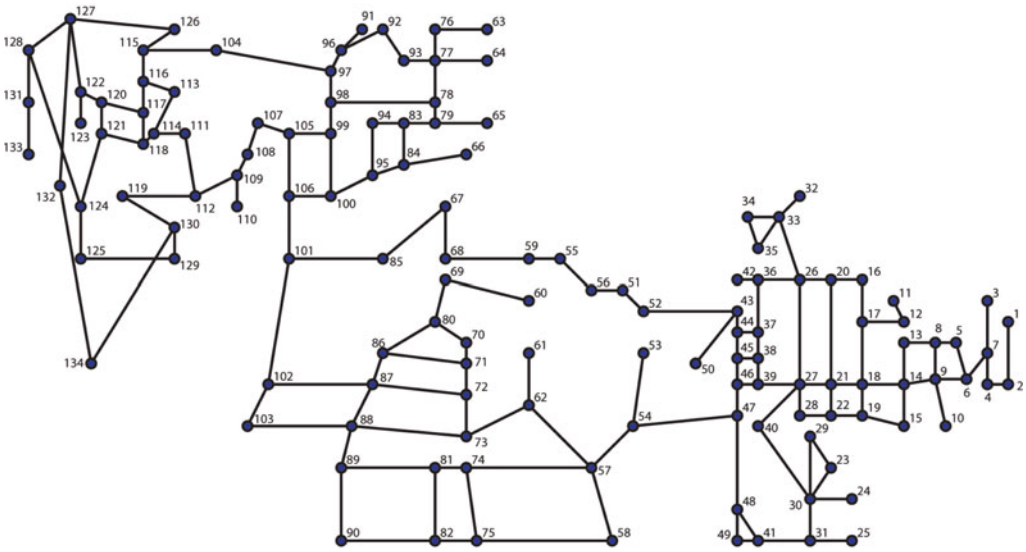


FIGURE C 1. Connectivity of the large network example.

Table C 1. Demands,  $D_i$ , at the interior junctions in the large network example

$i$	$D_i$	$i$	$D_i$	$i$	$D_i$
29	20	88	99	122	18
30	66	90	78	124	67
34	29	107	44	125	56
43	66	117	90	132	44
86	56				

Interior junctions without an entry have a zero demand and non-interior junctions are prescribed a pressure, which is given in Table C 2.

Table C 2. Pressures at the non-interior junctions in the large network example

$i$	$P_i$	$i$	$P_i$	$i$	$P_i$	$i$	$P_i$
1	1	25	0	60	1	66	1
3	2	32	0	61	1	91	0
10	0	42	1	63	0	110	0
11	0	50	0	64	0	123	1
24	0	53	0	65	0	133	1

Interior junctions do not have a prescribed pressure but have demands, which are given in Table C 1.