

Utilising Adjacency Matrices in Biological Systems: A Focus on Disease Modelling

Alberto-John THORNTON

Introduction

Adjacency matrices are powerful tools for representing complex biological interactions within network theory. They capture the presence or absence of connections between nodes in a network, which can represent entities such as proteins, genes, species, or individuals. This report explores the utility of adjacency matrices in modelling disease spread within biological systems.

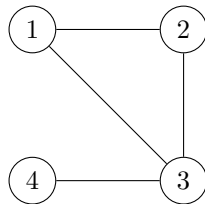
Adjacency Matrix Overview:

An adjacency matrix is a square matrix used to represent a finite graph. The elements of the matrix indicate whether pairs of vertices (nodes) are adjacent (connected) or not in the graph.

Example:

Consider a simple undirected graph with 4 vertices:

- Vertex 1 is connected to Vertex 2 and Vertex 3.
- Vertex 2 is connected to Vertex 1 and Vertex 3.
- Vertex 3 is connected to Vertex 1, Vertex 2, and Vertex 4.
- Vertex 4 is connected to Vertex 3.



The adjacency matrix A for this graph is:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

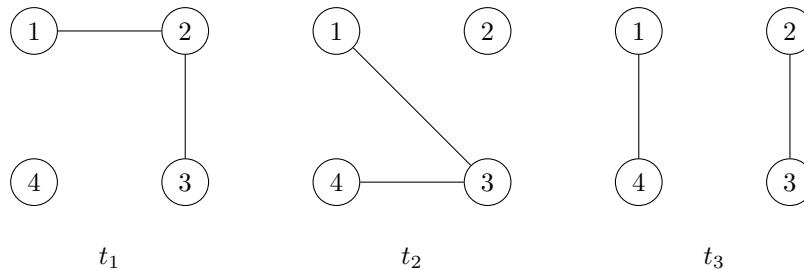
Adjacency matrices are particularly useful for representing graphs because they provide a simple and efficient way to capture the relationships between nodes. They allow for easy visualization and quick computation of graph properties, making them ideal for analyzing complex networks in various fields, including biology.

Temporal Networks:

Unlike static networks, temporal networks consider the timing and sequence of interactions, which are crucial for accurately modelling processes like the spread of infectious diseases. This dynamic aspect requires adaptations to traditional network models.

Example:

Consider a temporal network where interactions between nodes change over time. For example, nodes 1, 2, 3, and 4 have different connections at different time steps.



The SIR Model and Adjacency Matrices

The Susceptible-Infected-Recovered (SIR) model is a widely used framework in epidemiology for understanding the spread of infectious diseases within a population. This compartmental model divides the population into three categories:

- **Susceptible (S):** Individuals who are not yet infected but are at risk of becoming infected.
- **Infected (I):** Individuals who have been infected and can transmit the disease to susceptible individuals.

- **Recovered (R):** Individuals who have recovered from the infection and are assumed to have immunity.

The SIR model is governed by a set of differential equations that describe the rates at which individuals move from one compartment to another:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

where: - β is the transmission rate, representing the probability of disease transmission per contact between a susceptible and an infected individual. - γ is the recovery rate, representing the proportion of infected individuals who recover per unit time [1, 2].

Incorporating Adjacency Matrices

Adjacency matrices can be integrated into the SIR model to represent the contact network of individuals within a population. In this context, the adjacency matrix A is a square matrix where each element A_{ij} indicates whether there is a direct interaction between individuals i and j .

By incorporating an adjacency matrix, the SIR model can account for the specific patterns of interactions within a population, leading to more accurate simulations of disease spread. The modified equations for the SIR model using an adjacency matrix are:

$$\begin{aligned}\frac{dS_i}{dt} &= -\beta S_i \sum_j A_{ij} I_j \\ \frac{dI_i}{dt} &= \beta S_i \sum_j A_{ij} I_j - \gamma I_i \\ \frac{dR_i}{dt} &= \gamma I_i\end{aligned}$$

where S_i , I_i , and R_i are the susceptible, infected, and recovered states of individual i , respectively. The term $\sum_j A_{ij} I_j$ represents the sum of interactions that individual i has with infected individuals j [3].

Applications in Disease Modelling

Temporal networks play a crucial role in accurately modelling the spread of infectious diseases, as they account for the dynamic nature of interactions between individuals over time. This section discusses the application of temporal adjacency matrices in disease modelling, highlighting key findings from the paper "Infections on Temporal Networks—A Matrix-Based Approach" [3].

Study Overview

The paper by Holme (2015) introduces a method to incorporate temporal dynamics into the SIR model using adjacency matrices. The approach enhances the traditional SIR model by accounting for the time-varying nature of interactions between individuals, leading to more realistic simulations of disease spread.

Modelling Disease Spread with Temporal Networks

In the study, the authors used temporal adjacency matrices to model how diseases propagate through different types of networks. Temporal adjacency matrices allow the model to capture the sequence and timing of interactions, which are critical for understanding the spread of infections. We will delve into three examples shown in the paper.

Example 1: Social Contact Networks

One application discussed in the paper is the modelling of disease spread in social contact networks. The authors applied their model to data from a conference, where the interactions between participants were recorded over time. By using temporal adjacency matrices, the model was able to simulate the rapid spread of infections in a highly dynamic setting, providing insights into how outbreaks can occur and spread in such environments.

Construction of the Social Contact Network:

- During the conference, interactions between participants are recorded using proximity sensors.
- Each node in the network represents one of the 113 participants.
- An edge between two nodes is established whenever a face-to-face interaction is detected over a period of 20s.
- These interactions are recorded over different time steps, creating a temporal network where the adjacency matrices vary with time.

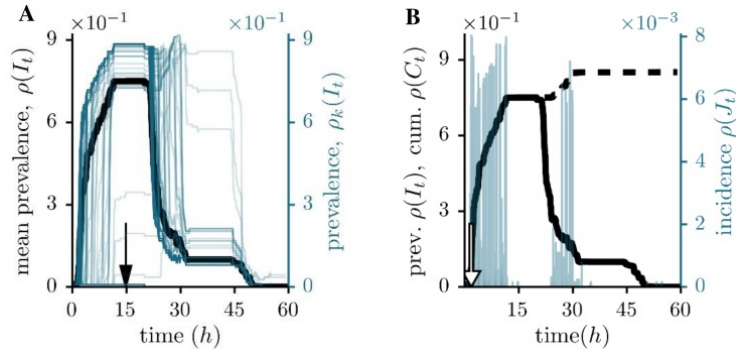


Figure 1: Mean prevalence and cumulative incidence over time in a social contact network [3]. **(A)** The comparison shows individual single-source outbreaks (blue, right axes) and the averaged prevalence (black, left axes). The infectious period is fixed at $\tau = 20$ hours. The arrow at $t = 14.8$ hours indicates the peak of the averaged prevalence. **(B)** Mean prevalence $\rho(I_t)$ (solid curves), incidence $\rho(J_t)$ (blue bars, right axes), and cumulative incidence $\rho(C_t)$ (dashed curves). The arrow points to the maximum averaged incidence at $t = 1.8$ hours.

Explanation of the SIR Process:

- The SIR (Susceptible-Infected-Recovered) model is applied to the social contact network to simulate the spread of an infectious disease shown in Figure 1.
- The process starts with one infected individual (node).
- The model assumes a fixed recovery time of 20 hours, chosen due to the dataset's characteristics (a night break of several hours).
- The infection spreads through the network based on the contact interactions recorded.

Key Observations:

- **Mean Prevalence ($\rho(I_t)$):** This measures the proportion of infected individuals over time. Figure 1 shows a rapid increase in infections within the first few hours due to the high interaction rate typical at conferences.
- **Cumulative Incidence ($\rho(C_t)$):** This measures the total number of infections over time. Figure 1 shows two stable phases (plateaus), indicating periods with no new cases after around 21 hours.
- **Plateaus in the Cumulative Incidence:** These phases highlight periods where the infection spread is contained temporarily, possibly due to breaks or reduced interaction periods during the conference.

- The infection process concludes before the third day, affecting about 85% of all nodes on average, as indicated by the cumulative incidence.

Comparison with Time-Aggregated Network:

| | $\rho(\mathbf{C}_T^{\text{temporal}})$ | $\rho(\mathbf{C}_T^{\text{static}})$ |
|---------------------|--|--------------------------------------|
| social interactions | 0.85 | 1.0 |
| sexual contacts | 0.027 | 0.88 |
| livestock-trade | 0.0048 | 0.15 |

We list the average fraction of individuals, which have been infected up to the observation time for the temporal networks ($\rho(\mathbf{C}_T^{\text{temporal}})$) and the corresponding time aggregated graph ($\rho(\mathbf{C}_T^{\text{static}})$). We use the same recovery times as in Figs 2–4.

doi:10.1371/journal.pone.0151209.t001

Table 1: Comparison Between Temporal Networks and Time Aggregated Networks

An equivalent experiment was conducted on a time-aggregated network so a comparison could be made. The same disease parameter was used, and it was assumed that the infection could only spread to the nearest neighbor within one time step, resulting in every node being infected independent of the seed of infection, as shown in Table 1 with a cumulative incidence of 1.0. It was found that the difference with the temporal network was only 17 nodes, therefore, regarding the total impact of an outbreak, a temporal representation of the data might not be necessary.

Overall, while temporal networks offer deeper insights into the dynamics of disease spread, for the purpose of assessing the total impact of an outbreak, a time-aggregated network representation might often be sufficient.

Example 2: Sexual Contact Networks

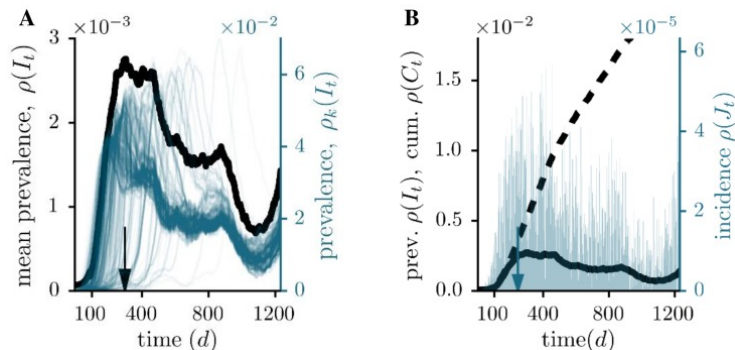


Figure 2: Mean prevalence and cumulative incidence over time in a sexual contact network [3]. **(A)** This panel compares individual outbreaks from a single source (blue, right axes) with the average prevalence (black, left axes). The infectious period is set to $\tau = 91$ days. The arrow marks the peak of the averaged prevalence. **(B)** This panel shows the mean prevalence $\rho(I_t)$ (solid lines), incidence $\rho(J_t)$ (blue bars, right scale), and cumulative incidence $\rho(C_t)$ (dashed lines). The arrow indicates the time of the highest average incidence.

In this dataset, 16,730 participants from a Brazilian escort website volunteered to record their sexual interactions online, as shown in Figure 2. Spanning a period from September 2002 to October 2008, this dataset allows for the study of the dynamics of sexually transmitted diseases. Using a 91-day infectious period (τ) based on HIV-1 transmission characteristics and ignoring the first 1000 days to avoid early-stage effects, the mean prevalence peaks after 305 days with a slow decline and considerable incidence throughout, as shown in Figure 2. The highest new infections occur after 245 days, reflecting the network’s steady growth and increasing interactions.

Comparison with Time-Aggregated Network:

In contrast to Example 1, the temporal characteristics of the sexual contacts graph result in a prolonged disease outbreak. This is because the network is growing, and an increasing number of interactions are recorded. Comparing the temporal network’s cumulative incidence (2.7%) to the static network (15%), we can see that the total impact would be greatly overestimated if we use the aggregated data, as shown in Table 1.

Key Findings:

- **Peak After 305 Days:** The mean prevalence of infections in the sexual contact network peaks after 305 days. This peak signifies the point at

which the infection has spread most widely through the network, with the highest proportion of individuals being infected at the same time. This delayed peak, compared to more immediate outbreaks in other networks, reflects the slower, more sustained transmission typical in networks characterized by ongoing and varied interactions over time.

- **Significance of Prolonged Outbreak:** The extended duration of the outbreak in the sexual contacts network highlights the importance of considering temporal dynamics in disease modeling. Unlike static analyses, temporal networks capture the ongoing nature of interactions and the resultant sustained transmission, leading to more accurate predictions and better-targeted interventions.
- **Higher Accuracy in Temporal Networks:** The temporal network's mean cumulative incidence is 2.7%, significantly lower than the 15% observed in the time-aggregated network. This stark difference demonstrates that using static, aggregated data can greatly overestimate the total impact of the outbreak.

Overall, this example emphasizes the critical importance of considering temporal dynamics in network analysis to avoid overestimation of disease impact.

Example 3: Livestock Trade Networks

This dataset includes over 900,000 trade transactions between 70,286 agricultural premises and traders from January 1 to July 20, 2011, at a one-day resolution. Livestock trade is a major transmission route for animal-related diseases. In the event of a potential outbreak, the detected node (the animal) would be isolated from the trade network. Therefore, the detection period is primarily determined by the detection time; in this example, a 14-day detection period was used. The epidemic evolves with a mean prevalence of 64 days and an incidence of 60 days, both much longer than the infectious period, indicating slow network mixing.

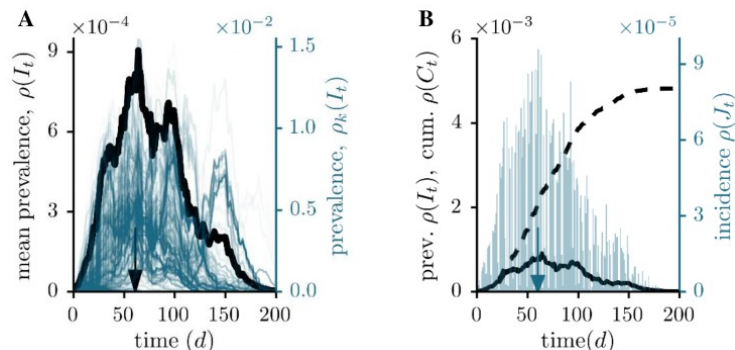


Figure 3: Mean prevalence and cumulative incidence over time in a livestock trade network [3]. **(A)** This panel shows a comparison between individual outbreaks from a single source (blue, right axes) and the average prevalence (black, left axes) with an infectious period set at $\tau = 14$ days. The arrow indicates the point of highest average prevalence. **(B)** This panel illustrates the mean prevalence $\rho(I_t)$ (solid lines), incidence $\rho(J_t)$ (blue bars, right axes), and cumulative incidence $\rho(C_t)$ (dashed lines). The arrow points to the time of peak average incidence.

Comparison with Time-Aggregated Network

Similar to the sexual contact network example, the temporal characteristics of the livestock trade network result in a prolonged disease outbreak. This is due to the slow mixing within the network, as indicated by the longer mean prevalence and incidence periods (Figure 3). Comparing the temporal network’s cumulative incidence (0.5%) to the static network (15%), we can see that the total impact would be greatly overestimated if we use the aggregated data, as shown in Table 1. These findings confirm previous observations that time-respecting paths can lead to a considerable improvement compared to static network analysis.

Key Findings

- **Prolonged Disease Outbreak and its Implications:** The temporal characteristics of the livestock trade network lead to a prolonged outbreak. This extended duration is due to the slow mixing within the network, as indicated by the longer mean prevalence (64 days) and incidence (60 days) periods. The slow mixing implies that an epidemic can persist for a longer duration within the network, affecting different nodes over extended periods. This results in a prolonged outbreak, making it more challenging to control and eradicate the disease.
- **Higher Accuracy in Temporal Networks:** The temporal network’s cumulative incidence is 0.5% (around 340 nodes), significantly lower than the overestimated 15% observed in the time-aggregated network. This

stark difference demonstrates that using static, aggregated data can greatly overestimate the total impact of the outbreak.

The livestock trade network analysis highlights the critical role of temporal dynamics in modeling disease spread. Slow network mixing results in prolonged outbreaks, making control more challenging. Temporal network analysis offers more accurate estimates of outbreak impact compared to static models, underscoring its importance for effective disease management.

Conclusion

Temporal networks are shown to offer a nuanced and realistic representation of disease spread, capturing the intricacies of interactions over time. This report's analysis indicates that while static networks provide a simpler approach, they often lead to overestimations and fail to accurately predict outbreak dynamics.

By examining the spread of infections in social contact, sexual contact, and livestock trade networks, it becomes clear that temporal networks are indispensable for understanding the true nature of disease propagation. The temporal approach allows for:

- **Accurate Disease Modelling:** By capturing the sequence and timing of interactions, temporal networks provide a more precise prediction of disease spread.
- **Improved Public Health Interventions:** Understanding the detailed dynamics of disease spread helps in designing targeted and effective intervention strategies, thereby enhancing public health responses.
- **Avoiding Overestimation:** Static networks tend to overestimate the impact of outbreaks, potentially leading to resource misallocation. Temporal networks offer a more balanced and realistic assessment.

However, it is important to note that temporal networks may not always be necessary or useful in every context. For instance, in the case of social contact networks, the difference in total outbreak impact between temporal and time-aggregated networks was minimal. This suggests that for certain types of analysis, where the overall impact is the primary concern, a static approach may suffice, offering simplicity and reduced computational complexity.

Future research should continue to refine these models, exploring new ways to enhance the accuracy and applicability of temporal networks in various biological contexts. With the further addition of data, new models can be refined and developed to be more accurate. Nonetheless, researchers should carefully consider the context and goals of their analysis when deciding whether to employ temporal or static network representations.

References

- [1] Anderson, R. M., & May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- [2] Kermack, W. O., & McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772), 700-721.
- [3] Holme, P. (2015). "Infections on Temporal Networks—A Matrix-Based Approach." *Journal of Complex Networks*.