

IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

BSC PROJECT

Pandemic Modelling:

Assessing the Validity of Average Shortest Path Length
as an Intermediate Explanation for the Phase Transition
Phenomenon in a Small World Network

Author:

02020640

Supervisor:

Dr. Dave CLEMENTS

Assessor:

Dr. Tim EVANS

Team Members:

02020580

02074553

Word count: 5,300

May 4, 2024

IMPERIAL

CONTENTS

I	Declaration of Work Undertaken	2
II	Introduction	2
III	Theory: Percolation Theory	3
IV	Methodology	5
V	Results	9
VI	Discussion	19
VII	Conclusion	23
VIII	Acknowledgement	23
IX	References	23

Abstract

This computational study simulated the spread of a disease in a finite population using small-world networks, adopting the Susceptible-Infected-Recovered model of epidemiology. A quantitative study of the non-linear phase transition phenomenon was conducted by considering the average shortest path length of the network. The point of criticality was found to be at $(2 \pm 0.2) \cdot 10^3$. For systems with an average shortest path lengths higher than the critical value, the disease was found to be in the contained state, otherwise, the disease infected the entire population. The effects of adopting mitigation measures were also investigated to reveal that mitigation measures increase the average shortest path length of a network.

I. DECLARATION OF WORK UNDERTAKEN

Amongst the team of three (Alex Coleman, Charlie Stamp and myself), the division of work throughout the project varied. During the development of the main simulation, the work was carried out collaboratively. When stand-alone subtopics were investigated, the work was carried out individually to maximise the possible avenues we could explore as a team.

II. INTRODUCTION

THIS study utilised principles from network and complexity science to conduct a computational epidemiology investigation.

Network and Complexity Science

Network and complexity science was first recognised as a discipline in the 1980s [1] and now labelled as ‘the science of the 21st century’ [2]. The emergence of complexity science can be simultaneously accredited to Paul Erdős and Alfréd Rényi for their works on random networks in graph theory [3] and Mark Granovetter on social networks [4]. This field of discipline is not clearly defined, in fact, contemporary books spend a lot of efforts on defining the discipline itself. Fundamentally, the discipline is a study of complex graph systems and the macroscopic phenomena that arise from these systems, underpinned by graph theory. The essence of complexity science is highlighted in *The Strength of Weak Ties* by Granovetter, commenting on the inherent relationship between micro-level interactions to macro-level patterns, as well as ‘the cohesive power of weak ties’ [4]. Complexity science is an interdisciplinary field: its insights are relevant in statistical physics, economics and sociology [5] [6].

Epidemiology

Epidemiology is the study of the spread of a disease. The contemporary approach to modelling in epidemiology originated from the works of Ronald Ross, who named the discipline ‘*a priori pathometry*’ (*the mathematical study of epidemics*) [7]. In his works, Ross outlined the compartmental mathematical framework which is the current standard in epidemiology. Epidemics of infectious diseases have been a recurring phenomenon in human history. With four pandemics since the turn of the century: SARS (2003), H1N1 (2009), Ebola (2013), COVID-19 (2019) [8], there is evidence to suggest pandemics are happening at a higher frequency [9]. The value of epidemiology arises from its abilities to build an accurate, predictive tool to guide public health efforts and mitigation strategies. In 2009, the course and time evolution of the H1N1 pandemic was accurately predicted months before the peak of the pandemic using mathematical models [10].

The aim of this investigation was to investigate the macroscopic phenomenon of percolation with a disease spreading in a finite small-world network and modelling the effects of mitigation adoption.

III. THEORY: PERCOLATION THEORY

The study of complexity is concerned with the emergent macroscopic behaviours of a network when a set of simple rules are repeatedly applied to a system, where the behaviours observed are not encoded in the rules [11]. One example of a set of systems which demonstrates emergent behaviours are the Bernoulli percolation models.

The Bernoulli site percolation model consists of a finite 2D lattice of size $n \cdot m$, where n and m are the number of sites along the x and y direction respectively. There exists a parameter called the occupation probability p_o : this is the probability that a site is occupied. The emergent behaviour of interest here is the formation of percolating clusters. A cluster is a group of nearest neighbouring occupied sites. A percolating cluster is defined as a cluster which percolates between opposite boundaries, touching the boundaries on both sides of the lattice. For a finite lattice, it is intuitive that at low p_o , the probability of forming a percolating cluster is low. For a $p_o \approx 1$, we expect a high probability of forming a percolating cluster. As demonstrated in Fig. 1, we observe a percolating cluster for the first time for when $p \approx 0.59$. This is the percolation threshold p_c , also known as the point of criticality, where the system undergoes phase transition. The phases in different systems are different depending on the subject of interest, the relevant phases in the site percolation model are I) the absence of a percolating cluster and II) the existence of a percolating cluster. Percolation theory provides insights into the phase transitions and critical phenomena of more complicated systems [12].

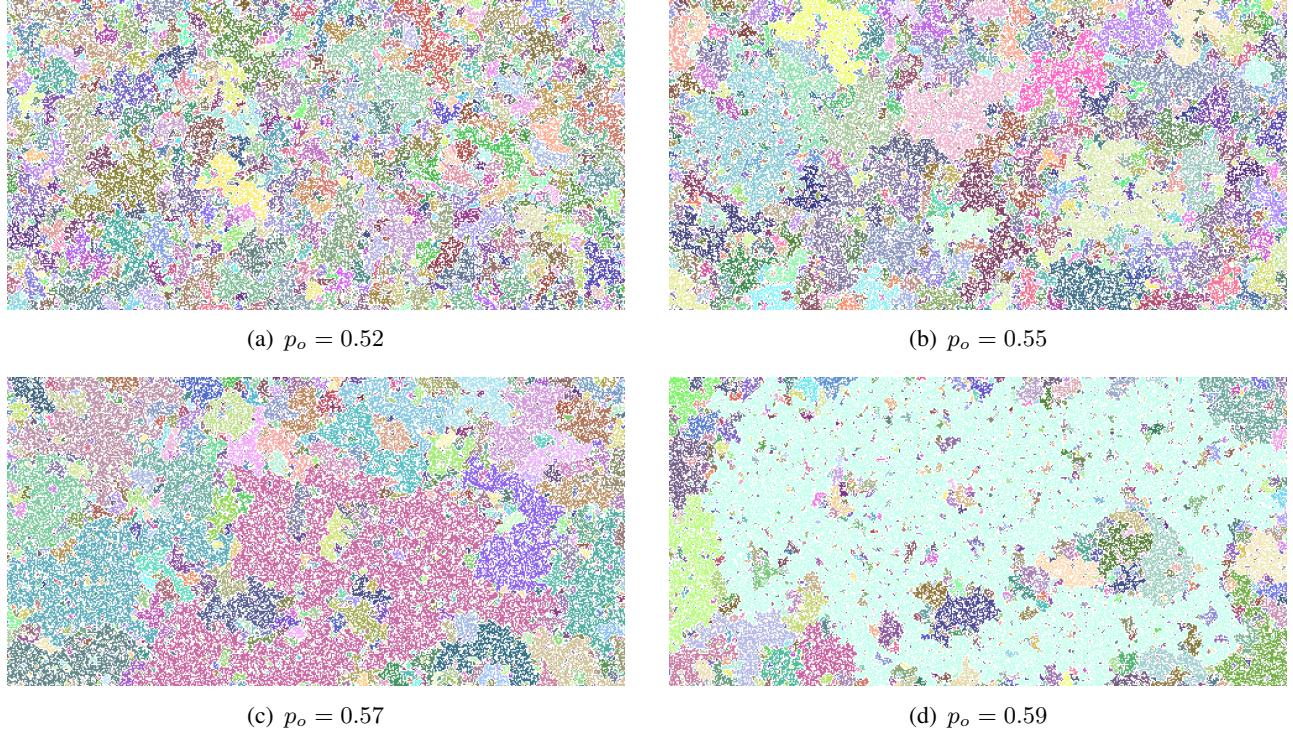


Fig. 1: Site percolation demonstration with a 5000×2500 lattice and varying the probability of occupation p_o . A homogeneous area of colour indicates the existence of a cluster. For $p_o = 0.52$ (1(a)) and $p_o = 0.55$ (1(b)), the cluster sizes are small and no percolating cluster exists. Between $p_o = 0.57$ (1(c)) and $p_o = 0.59$ (1(d)), the cluster size increases non-linearly. An infinite, percolating cluster is observed for the first time for $p_o = 0.59$.

Percolation theory originated from the study of the flow of a fluid through a medium with randomly blocked channels [4]. By considering a network of connected individuals capable of spreading an infection through the network, the study of the spread of a disease becomes a percolation problem investigating

whether the disease can permeate through the network [13]. The relevant phases in the pandemic model are I) a contained disease and II) a disease outbreak percolating through the population.

Percolation theory is a study of connectivity and clustering. There exists parameters to quantify the connectivity of a network, including the cliqueness (the existence of maximally connected clusters) and the average degree (average number of connections an agent has). The parameter utilised in this investigation is the Average Shortest Path Length (ASPL), the mean shortest distance between two nodes, calculated by Eq. 1. For a non-weighted calculation, each edge traversed contributes 1 to the distance $d(s, t)$. For a weighted calculation, each edge contributes w to the distance, where w is the weighting of the edge, as demonstrated in Fig. 2.

$$a = \sum_{s,t \in V, s \neq t} \frac{d(s, t)}{n(n - 1)} \quad (1)$$

where V is the set of nodes in the network G , $d(s, t)$ is the shortest path from the source node s to the target node t , and n is the number of nodes in G [14].

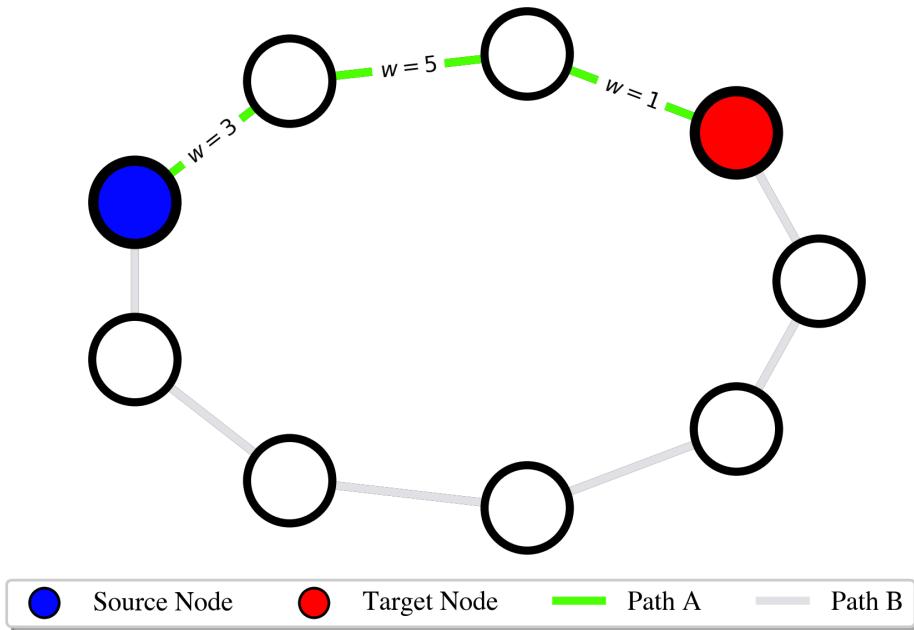


Fig. 2: This figure demonstrates the principle behind shortest path length calculations in a ring network. The shortest path length is, intuitively, the shortest number edges that can connect the source node s and the target node t . Path A is the shortest path between s and t , path B is not. The weighted length of path A is $3 + 5 + 1 = 9$.

In a pandemic model, there are layers of complexity and many interacting factors. The hypothesis of this investigation is that the ASPL encapsulates the effects of variations in all topology parameters, and that the point of criticality of the system can be defined in terms of the ASPL.

IV. METHODOLOGY

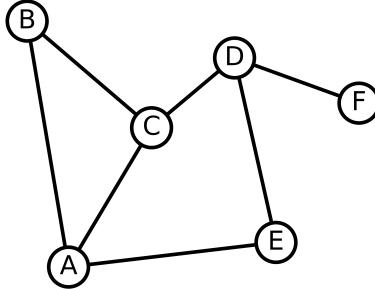


Fig. 3: Visualisation of a simple random network generated with NetworkX. The circles labelled $\{A, B, C, D, E, F\}$ are nodes, the straight lines between nodes are edges.

Networks are composed of nodes and edges, where the edges connect the nodes, as shown in Fig. 3. In this simulation, a node represented an individual in the population and an edge represented interactions between individuals. The geometry of the network was representative of social locations rather than physical locations, as the edges represented interactions, rather than physical distance between agents. The infection traveled through the network along edges: i.e. it could only pass from node to another if there existed an edge between the two nodes. The simulation adopted an agent-based approach by defining individual-level interactions ('interaction-based modelling').

This investigation considered small-world networks generated using the NetworkX module according to the Watts-Strogatz model [15]. The Watts-Strogatz model generates small-world networks by initializing a ring network where each node is connected to its k nearest neighbors, where k is the average number of connections. Each edge is then iterated through with a probability of rewiring an edge, this probability is the randomness of the network.

Simulation Workflow

The simulation adopted the basic compartmental model in epidemiology – the Susceptible-Infected-Recovered (SIR) Model [16]. At any time, a node belonged to one of the mutually exclusive compartments. In this investigation, the compartments were: susceptible, infected and recovered. Nodes were characterised by these compartments depending on their infection status. By considering the number of new infections at each time step, the cumulative case count was calculated.

Temporal data was the information of interest, so the network was evaluated at each time-step. The time scale utilised here is adaptive, dependent on the purpose of the investigation, the time-steps can be representative of an hour, a minute or a month. At each time-step, the number of nodes in each compartment and other relevant information such as the average shortest path length were recorded. Each node had a finite infectious period, this was characterised by the parameter 'Time to Recover' (TTR). TTR indicated the number of time-steps for which the node remains infectious for. The microscopic interactions of the simulation is demonstrated in Fig. 4.

The probability of infection p and reduced probability of infection p_r were defined as the probability of transmitting the infection from an infected node to a connected susceptible node at a time-step. For a node adopting mitigation the probability of infection was p_r , otherwise, it was p .

There existed a hierarchy of weighting attached to edges, as demonstrated in Fig. 5. If at least one of the nodes attached to an edge was infected and was not adopting mitigation, the probability attached to the edge was p . The probability of infection was p_r only when both edges attached to the node were adopting mitigation. This was based on the evidence that mitigation measures, e.g. masking, are only effective when both the infected and susceptible individuals are adopting the measures [17].

To verify the validity of the simulation, small scale simulations were ran. The behaviour of the system and status of the nodes at each time-step was verified. Inherently, it is difficult to verify a large scale

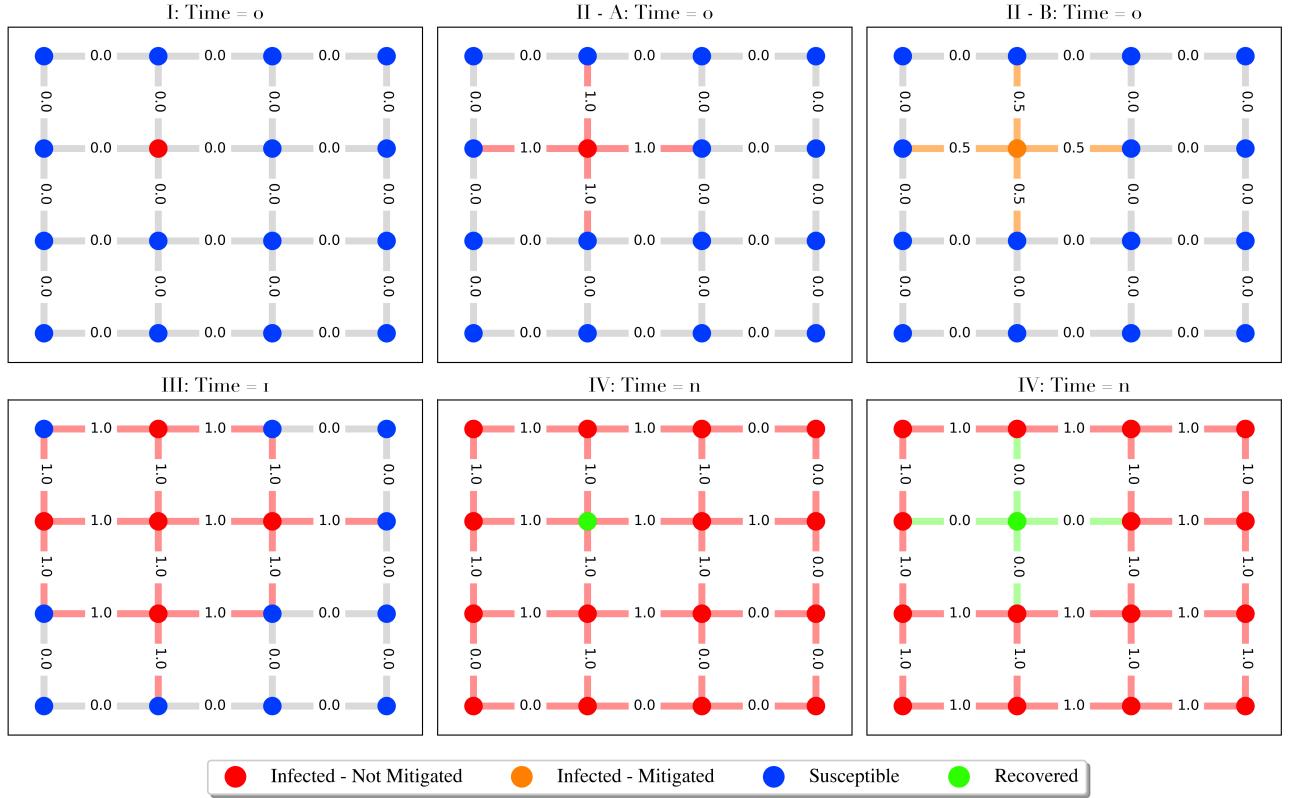


Fig. 4: Visualisation of the workflow for an infected node, created using NetworkX. I: The simulation is initiated with a single sick node. II: All the edges attached to the sick node are characterised with a weighting on that edge. The weighting is p , the probability of infection for when no mitigation was adopted (A) and p_r , the reduced probability of infection, if mitigation was adopted (B). III: At each time-step, all edges connecting one infected node and one susceptible node are queried, and new nodes are infected according to the probability of infection on the weighting. IV: At the end of the infectious period, indicated by TTR, the node is recovered and the weighting on all the edges attached to the node are set to zero. Here, TTR = n time-steps.

network, due to the emergence of unexpected, non-linear phenomena. However, the simulation could be verified with known results. For example, for $p = 0$ we expect the disease to not propagate and for $p = 1$ we expect the disease to outbreak in the population quickly, infecting 100% of the population.

To account for the probabilistic and stochastic nature of infection propagation, all simulations were run for a large number of trials ($\sim 10^2$). This helped with the identification of general trends and elimination of anomalous behaviour.

ASPL Calculation

The average shortest path length was calculated by an algorithm in NetworkX. The formula used for the calculation is given by Eq. 1. The weighting used to calculate ASPL is $1/p$ where p is the probability weighting on that edge. For example, for an edge with probability weighting = 0.5, the edge contributed $1/0.5 = 2$ to the path length.

Assumptions and Simplifications

In order to conduct an effective investigation, simplifications were made, leading to the following assumptions:

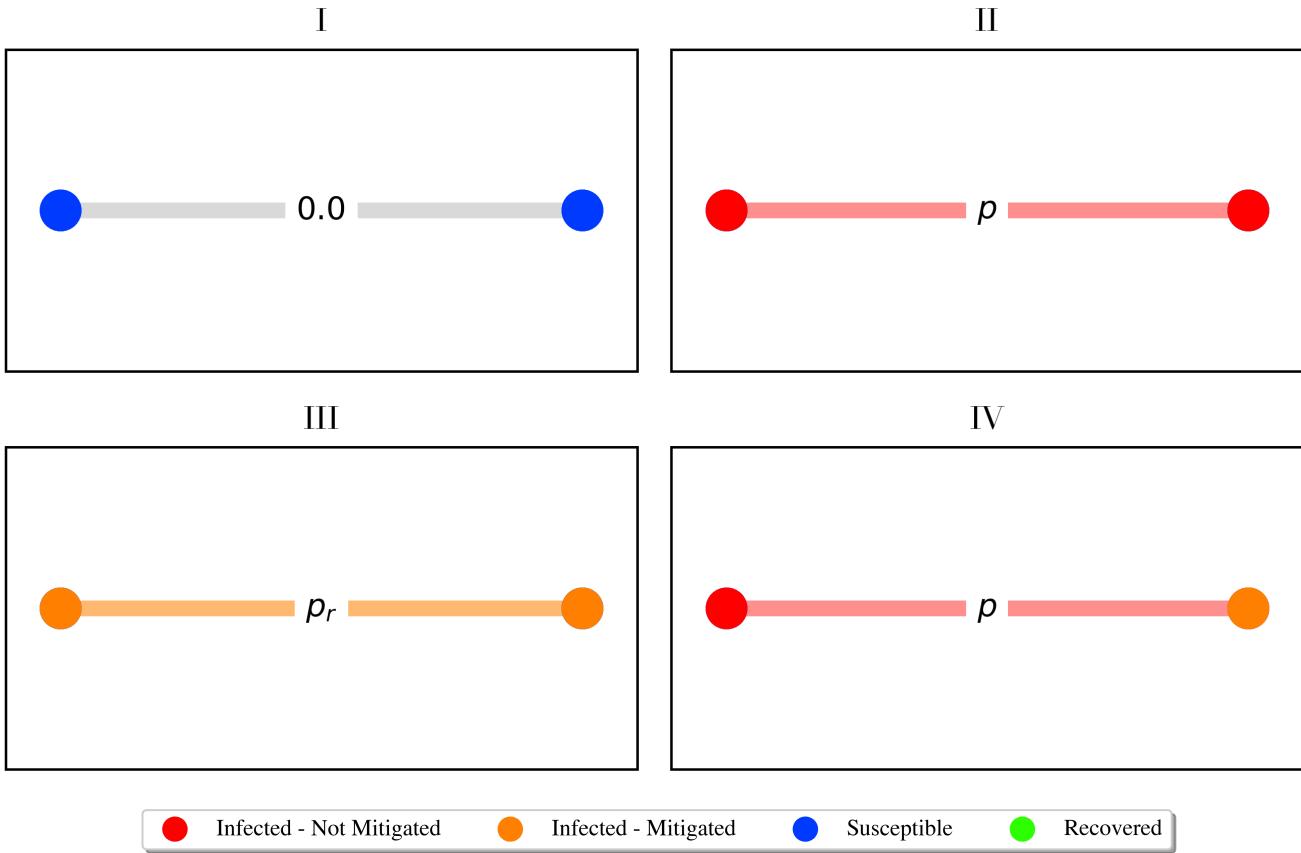


Fig. 5: This visualisation presents the dependency of edge weighting on the status of the connected infectious nodes. I. when no infected node is present, the edge weighting is 0. II. when both nodes are infected and not adopting mitigation, the weighting is p . III. when both nodes are infected and adopting mitigation, the weighting is p_r . IV. when both nodes are infected and only one node is adopting mitigation, the weighing is p .

- 1) The probabilities of infection p, p_r are constant.
- 2) TTR is constant for all individuals.
- 3) The number of nodes and edges remain constant.
- 4) No reinfection: once a node has been infected, it cannot be infected again.
- 5) Mitigation adoption is not directional.

The implications of these assumptions will be explored in the Discussion section.

Investigation Procedure

Firstly, the dependency of ASPL on topology parameters needed to be determined. The topology parameters are variables that affect the structure of the network. In this investigation, the following topology parameters were considered: the number of nodes, the average number of connections (degree), the randomness of the network and the probability of infection.

Secondly, the relationship between ASPL and the phase of the network needed to be determined. This required quantifying the percolating status of the network. This was achieved through three different measurements:

- 1) Percentage of population infected: this is the proportion of the total network infected at the end of the simulation¹.
- 2) Propagation time: the number of time-steps it takes for the infection to percolate through the network. If at the end of the simulation, there are still susceptible nodes, the propagation time was assumed to be $1 \cdot 10^9$ as an approximation to infinity. Since with the absence of infectious nodes and no possibility of spontaneous infection, theoretically, it would take an infinite amount of time for the infection to percolate through the network.
- 3) Probability of outbreak: out of a large number of trials ($\sim 10^2$), the percentage of trials where the network was in a percolating state. When the proportion of the total network infected was 100%, the network was considered to be in a percolating state.

In the context of this investigation, a percolating state is when the disease is in an outbreak state, i.e. when the disease can permeate through the entire population.

¹The end of the simulation is marked by when there are no more infectious nodes. Given the assumption that there are no spontaneous infections or reinfection, this marks the end of the disease propagation in the network.

V. RESULTS

Preliminary Results

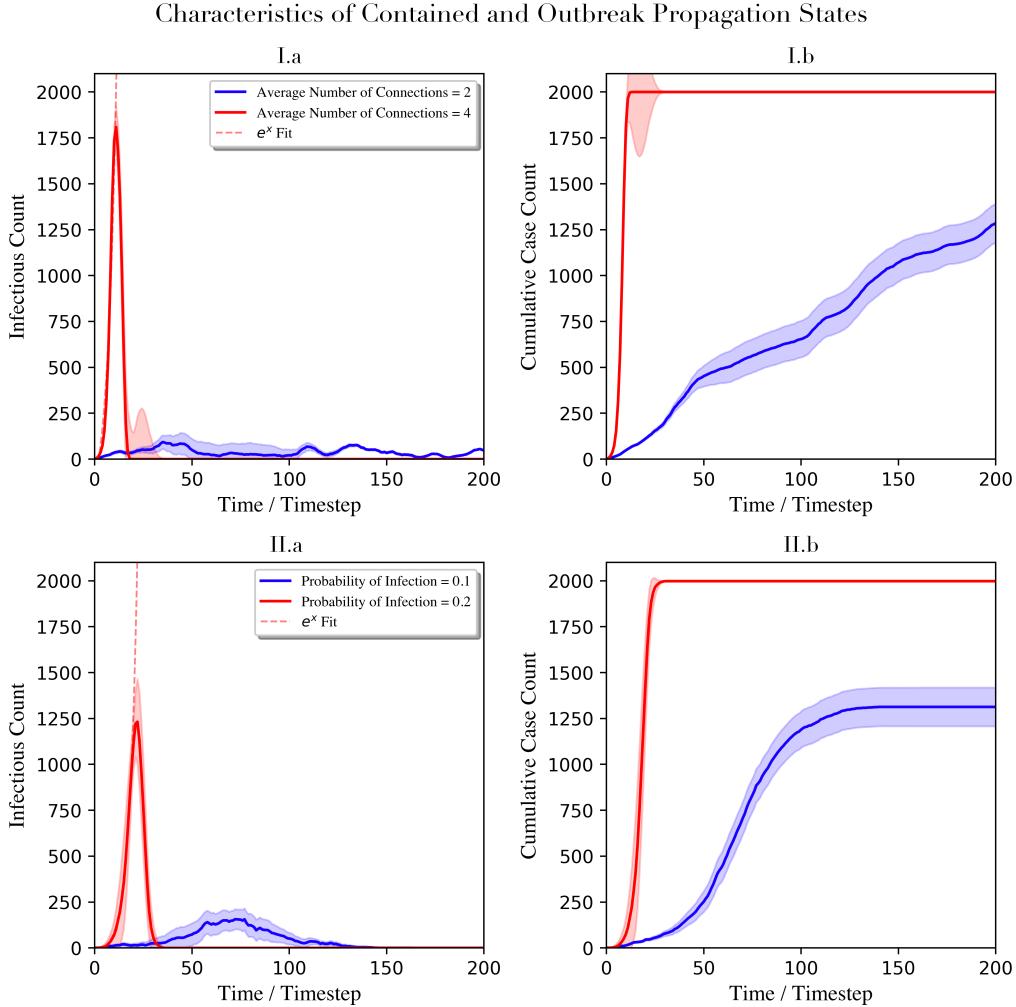


Fig. 6: Observation of the two distinct propagation phases: outbreak state (red) and contained state (blue), achieved by varying the average number of connections (I) and the probability of infection (II). I.a and II.a demonstrated that for the outbreak states, the initial rise of infected counts to the peak could be well-modelled by an exponential with a coefficient of determination of $R^2 = 0.997 + / - 0.001$. I.b and II.b show the cumulative case counts (the total number of cases up to a certain time): for the outbreak states, the infection spread to the entire population nearly instantly. The shaded area indicates 1σ of variation in recorded value across 20 simulation trials.

The preliminary results considered the temporal data of the compartments when varying the average number of connections and the probability of infection. As demonstrated in Fig. 6, there exists two distinct phases in the network: the contained state and the outbreak state. The percentage of population infected in a network in the outbreak state was consistently 100%, whereas for a contained state, only a proportion of the population was infected. Moreover, the results pointed towards the existence of a point of criticality in terms of the average number of connections and probability of infection. Distinctly different propagation characteristics could be observed: for the outbreak states, the initial rise in infectious cases could be well modelled by an exponential, whereas for contained states, the R^2 for an exponential fit was consistently negative, indicating a very poor fit [18]. The quality of the fit was quantified using the coefficient of

determination calculated by Eq. 2, where y is the value used for curve-fitting from the dataset, \hat{y} is the value predicted by the fit and \bar{y} is the mean y value of the dataset [19].

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2} \quad (2)$$

ASPL Dependencies on Topology Parameters

The four topology parameters (number of nodes, probability of infection, average number of connections and the randomness of a small world network) considered all had a well-mapped relationship to ASPL.

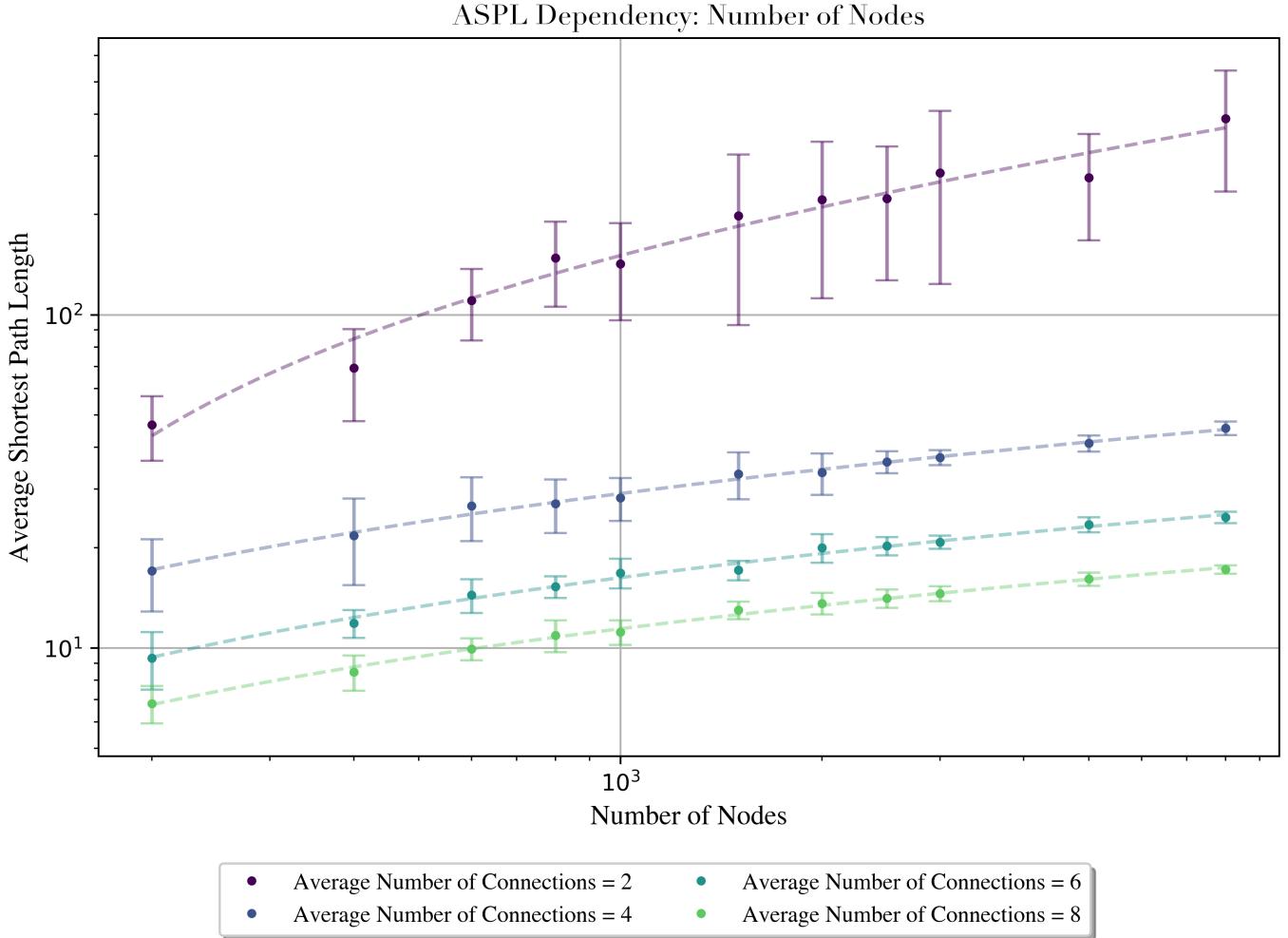


Fig. 7: The relationship between number of nodes and ASPL followed a power function relation (Eq. 3) with $R^2 = 0.984 \pm 0.009$ shown by the dashed line. The error bars indicate a 1σ variation in ASPL value. For a higher k : i) a increased ii) b decreased and iii) c decreased.

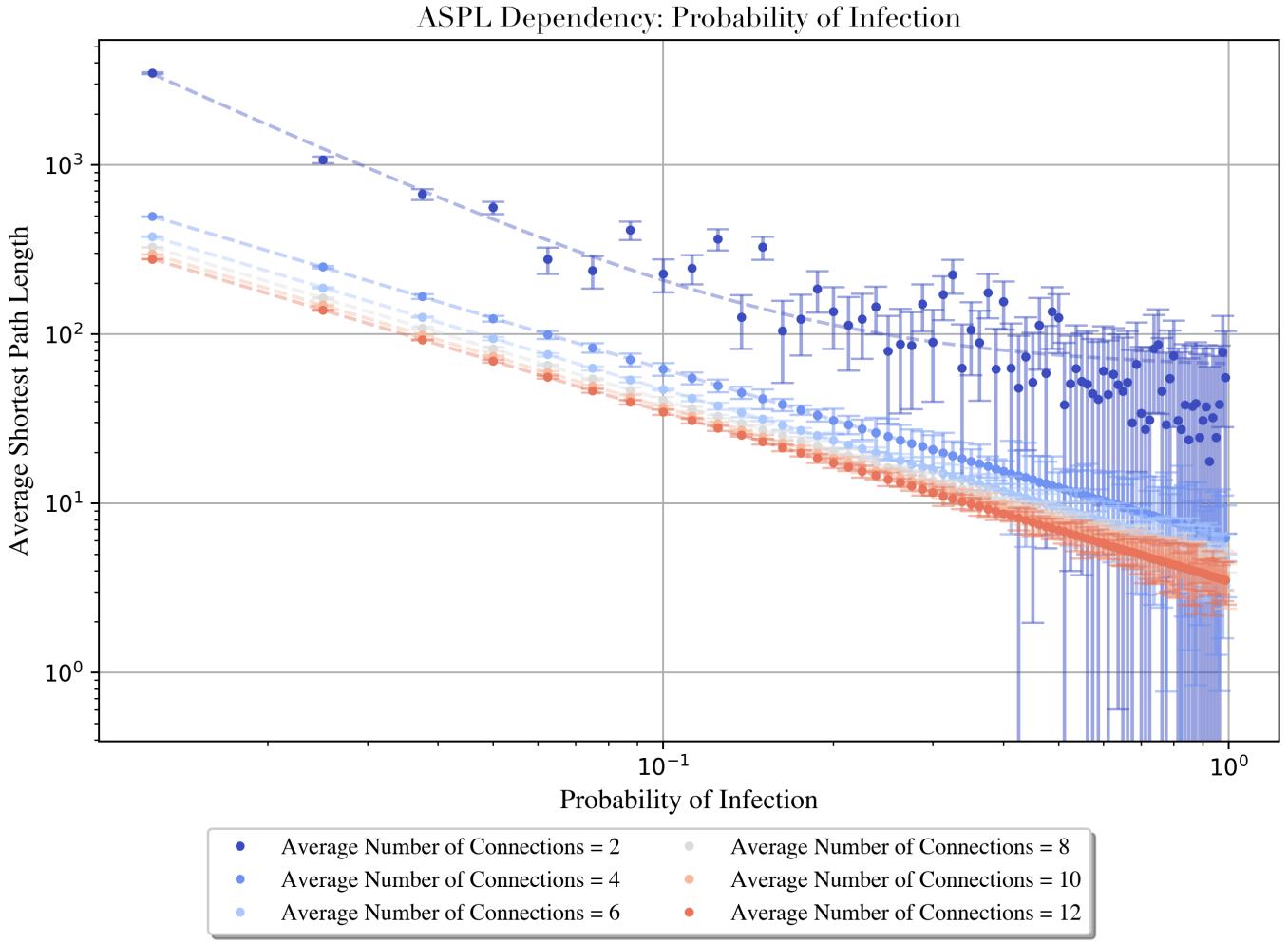


Fig. 8: The relationship between probability of infection and ASPL followed a power function, with $R^2 = 0.997 \pm 0.003$. For a higher k : i) a decreased and ii) c decreased. The error bars indicate 1σ variation.

The relationship between the ASPL and the number of nodes followed a power function as shown in Eq. 3 where a, b, c are fit parameters.

$$f(x) = a \cdot x^b + c \quad (3)$$

The fit follow the power function with a positive exponent parameter b in the range $0 < b < 1$, as demonstrated in Fig. 7. As the average number of connections k varied, the fit parameters changed accordingly, leading to the following observations for higher k : i) ASPL increased steeply initially, ii) the ASPL value plateaued quickly with the increase in number of nodes and iii) the ASPL was lower generally.

The relationship between the ASPL and the probability of infection also followed a power function, with a negative exponent parameter b , as demonstrated in Fig. 8. The goodness of fit was exceptional, with $R^2 = 0.996 \pm 0.003$. The exponent parameter b was independent of k , with a value of $= -1.09 \pm 0.09$. For a higher k : i) the initial increase in p led to a steeper decrease in ASPL and ii) the ASPL was lower generally.

The relationship between ASPL and the average number of connections is an exponential decay modelled by Eq. 4 with $R^2 = 0.998 \pm 0.001$, as seen in Fig. 9. In terms of ASPL, as the number of nodes increases:

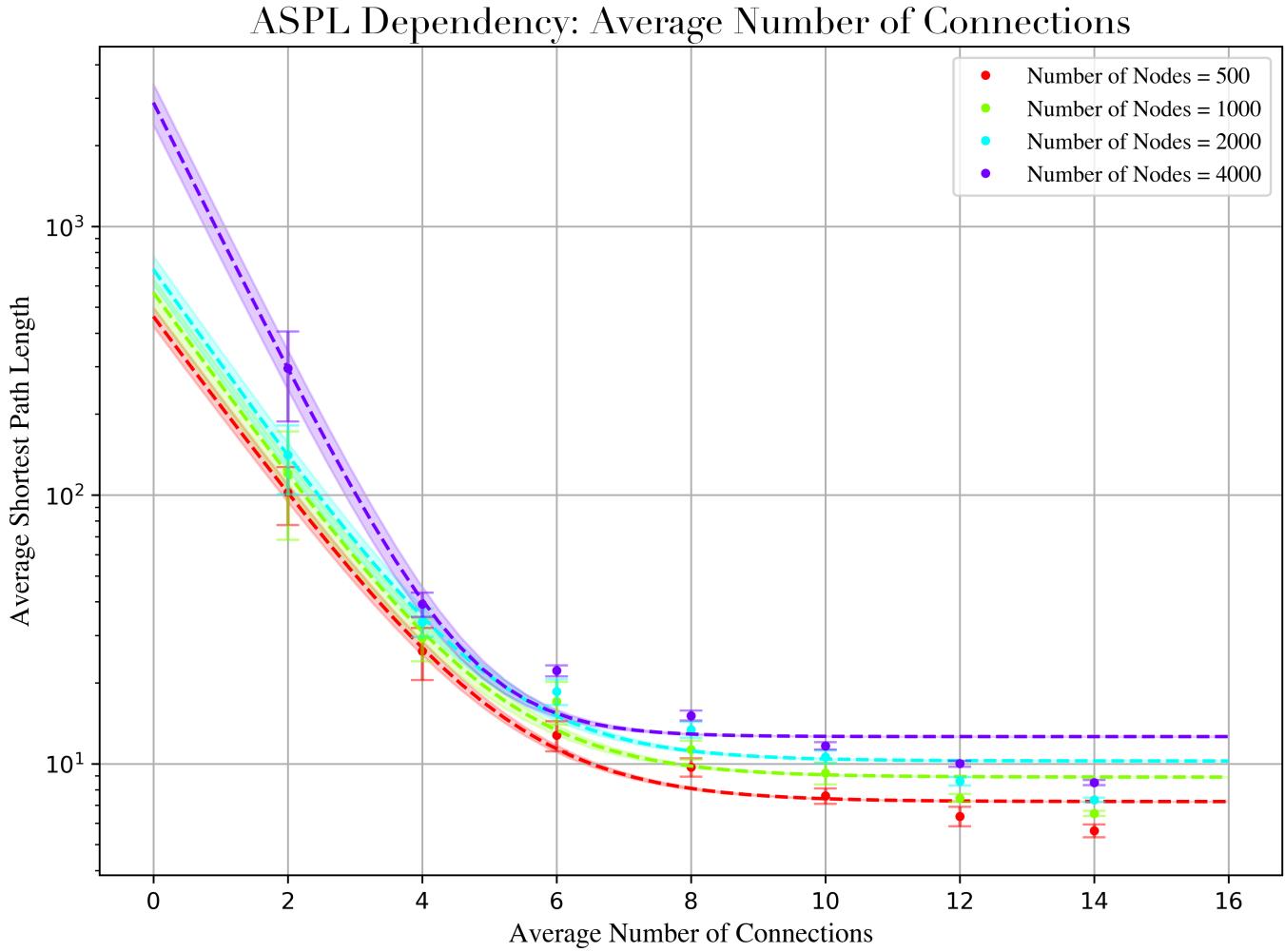


Fig. 9: The relationship between the average number of connections and ASPL followed an exponential decay relationship (Eq. 4), with a $R^2 = 0.9982 \pm 0.0004$. The error bars indicate a 1σ variation. The shaded region indicate the range of the possible fit functions. As the number of nodes n increased: i) a increased, ii) b increased and iii) c increased.

i) the steep ASPL drop happens at a higher average number of connections , ii) ASPL decreases more sharply and iii) ASPL is higher.

$$f(x) = a \cdot e^{-bx} + c \quad (4)$$

Lastly, as the randomness of a network increased, the ASPL decreased, as demonstrated in Fig. 10. The power function relationship between the number of nodes and ASPL was preserved. As randomness increased, the entire curve was shifted down, decreasing the ASPL.

Percolating State Dependence on ASPL

Two distinct phases and the location of phase transition could be identified by considering ASPL and the percentage of population infected, as seen in Fig. 11. The results demonstrated that the network was in at outbreak state at low ASPL, for $\text{ASPL} < 1 \cdot 10^3$, where the disease had percolated through the population. The network was in a contained state at high ASPL, for $\text{ASPL} > 4 \cdot 10^3$, where the disease was only capable of infected a small percentage of the population. The uncertainties introduced a finite

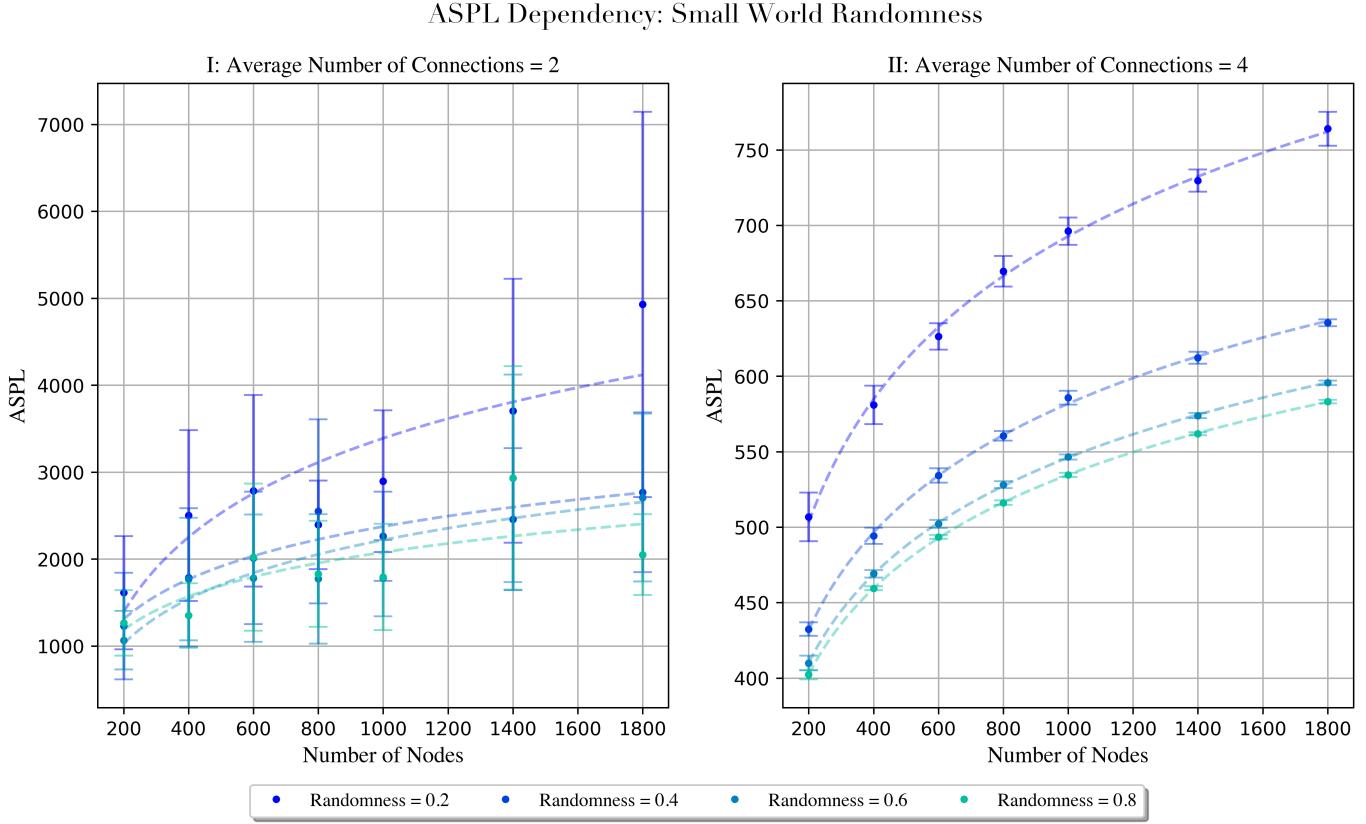


Fig. 10: Randomness of a small-world network shifted the position of the curve relating the number of nodes and ASPL. For $k = 2$: the power function fit has $R^2 = 0.75927 + / - 0.0695$. For $k = 4$, $R^2 = 0.9993 \pm 0.0005$.

range over which the critical value was located. Identical patterns could be observed by consideration of the propagation time and probability of outbreak as seen in Fig. 12 and Fig. 13 respectively.

The three metrics considered above agree on the following relationship between ASPL and the percolating state: the network is in an outbreak state for low ASPL and in a contained state for high ASPL. The critical threshold is the value which ‘marks the arrival’ of an infinite connected component [20]. In the context of pandemic modelling, this is the value for which the first outbreak state happens, i.e. the highest ASPL value for which the probability of outbreak is greater than 0. By considering the propagation time and the probability of outbreak, the critical threshold $\text{ASPL}_{\text{critical}}$ could be located at $(2 \pm 0.2) \cdot 10^3$.

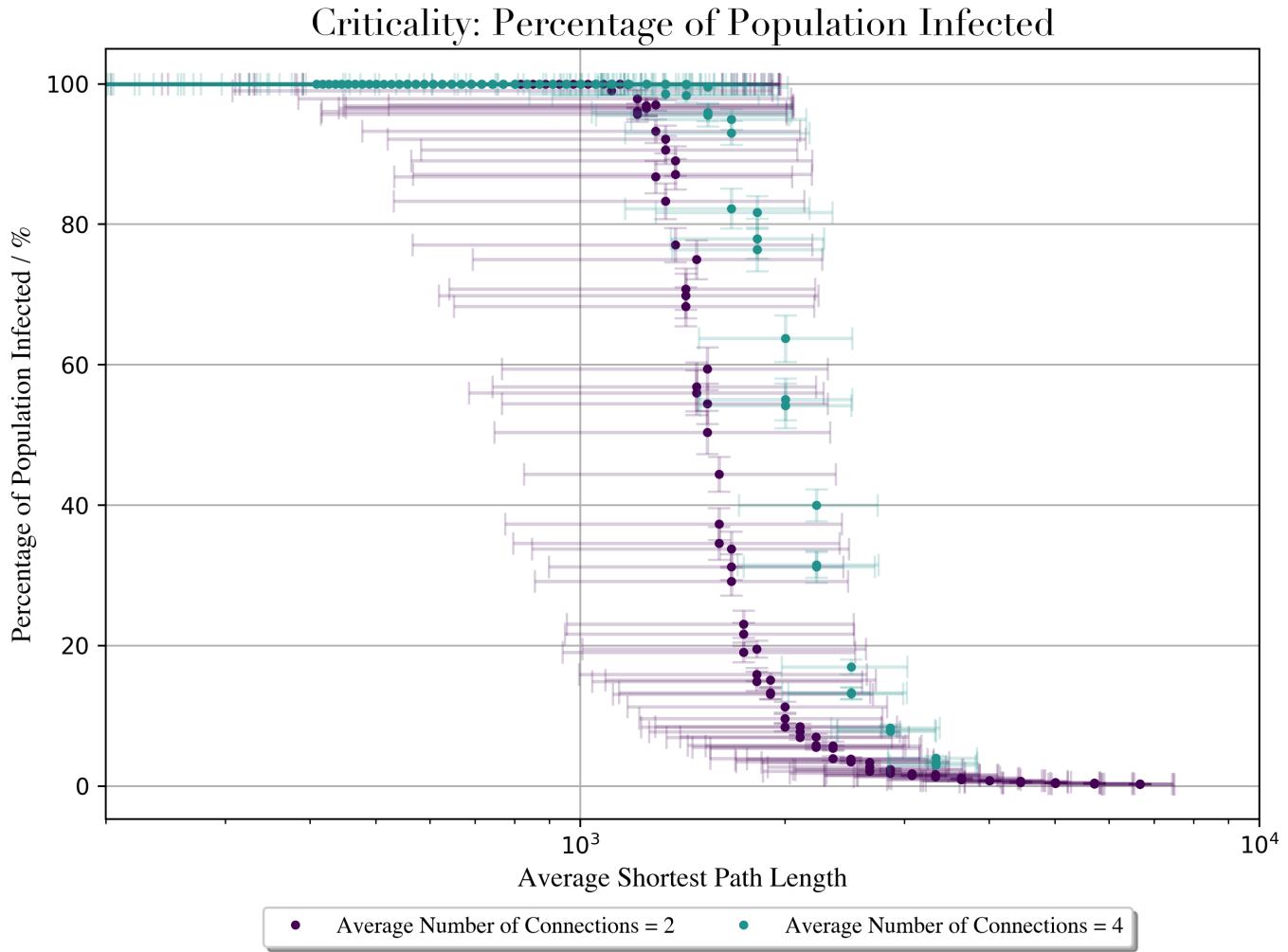


Fig. 11: The relationship between ASPL and percentage of population infected followed a sigmoidal curve. For $\text{ASPL} < 1 \cdot 10^3$, 100% of the population was infected on average. For $\text{ASPL} > 3 \cdot 10^3$, 0% of the population was infected on average. The ASPL errors indicate 1σ variation and the errors on the percentage of population infected error was the standard error of the mean from $\cdot 10^2$ simulations.

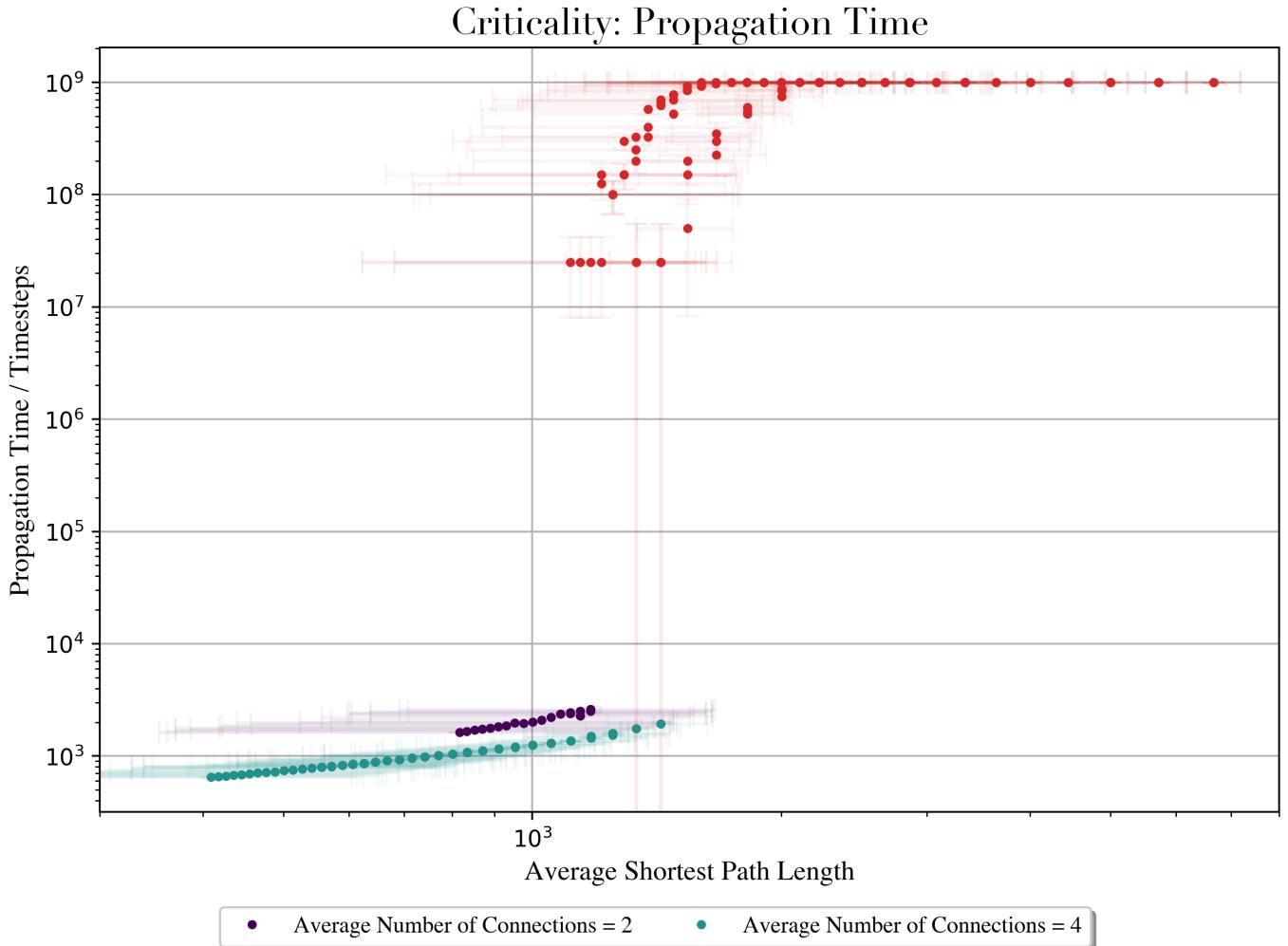


Fig. 12: The results here are the mean values over 20 simulations with error bars indicating the standard error on the mean. At low ASPL ($\text{ASPL} < \sim 1.5 \cdot 10^3$), the network was in an outbreak state with a finite propagation time and the relationship between ASPL and propagation time was a power function. The red markers indicate when at least one of the trials was in a contained state. At high ASPL ($\text{ASPL} > 2 \cdot 10^3$), the network was in a contained state.

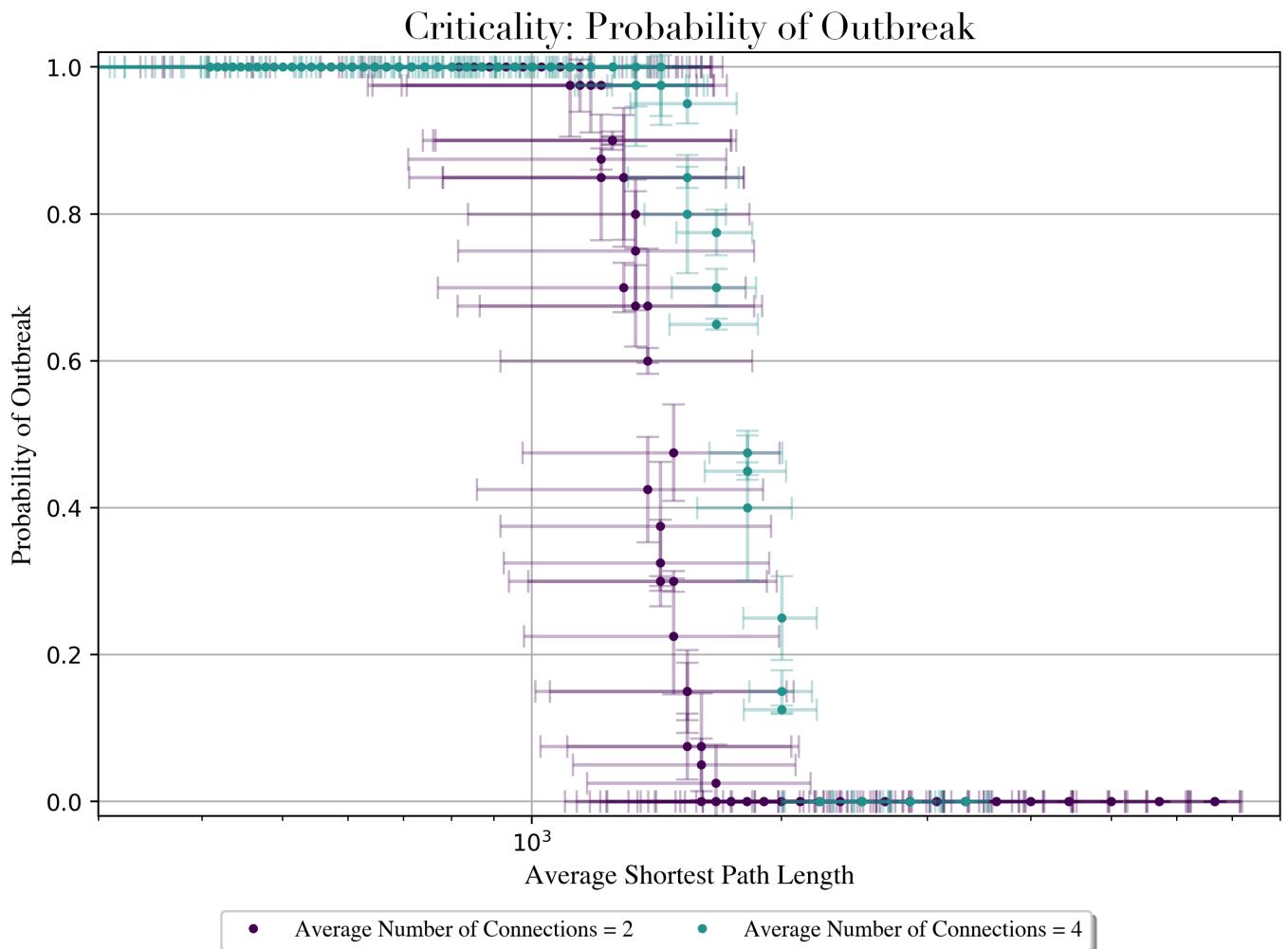


Fig. 13: The probability of outbreak considered the percentage of trials that are in an outbreak state (i.e. 100% of the population was infected). Again, the outbreak and contained state could be observed at low and high ASPL respectively.

Mitigation Measures

Investigation of networks adopting mitigation demonstrated that mitigation adoption and the reduced probability of infection (i.e. mitigation effectiveness) are parameters that ASPL depends on, as shown in Fig. 14. It can be noted that mitigation effectiveness was only significant for when adoption was $> 40\%$. Mitigation adoption works by effectively increasing the ASPL until it is beyond the critical value so that the network is in a contained state. Simulations with networks adopting mitigation demonstrated similar results as before, with the network in a contained state at high ASPL and an outbreak state at low ASPL, as seen in Fig. 15. For a more effective mitigation measure, i.e. a lower p_r , the ASPL was higher at higher rates of adoption.

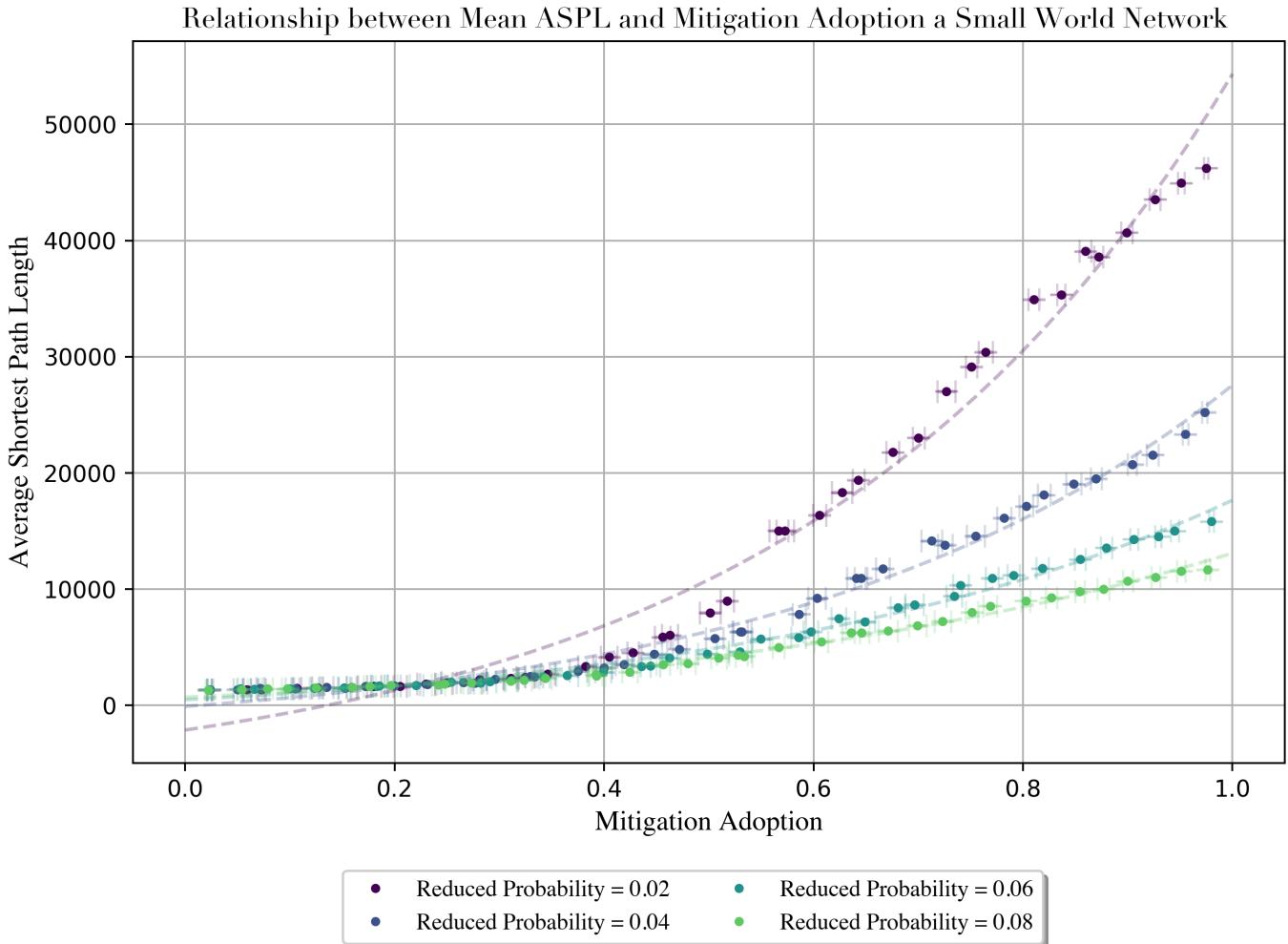


Fig. 14: The relationship between mitigation adoption and ASPL followed an exponential function (dotted line) with $R^2 = 0.986 \pm 0.002$. The rate of exponential growth was dependent on the reduced probability when adopting mitigation: the growth is steeper for a lower reduced probability. The errors in mitigation adoption represent the range in actual adoption over $\cdot 10^2$ simulations.

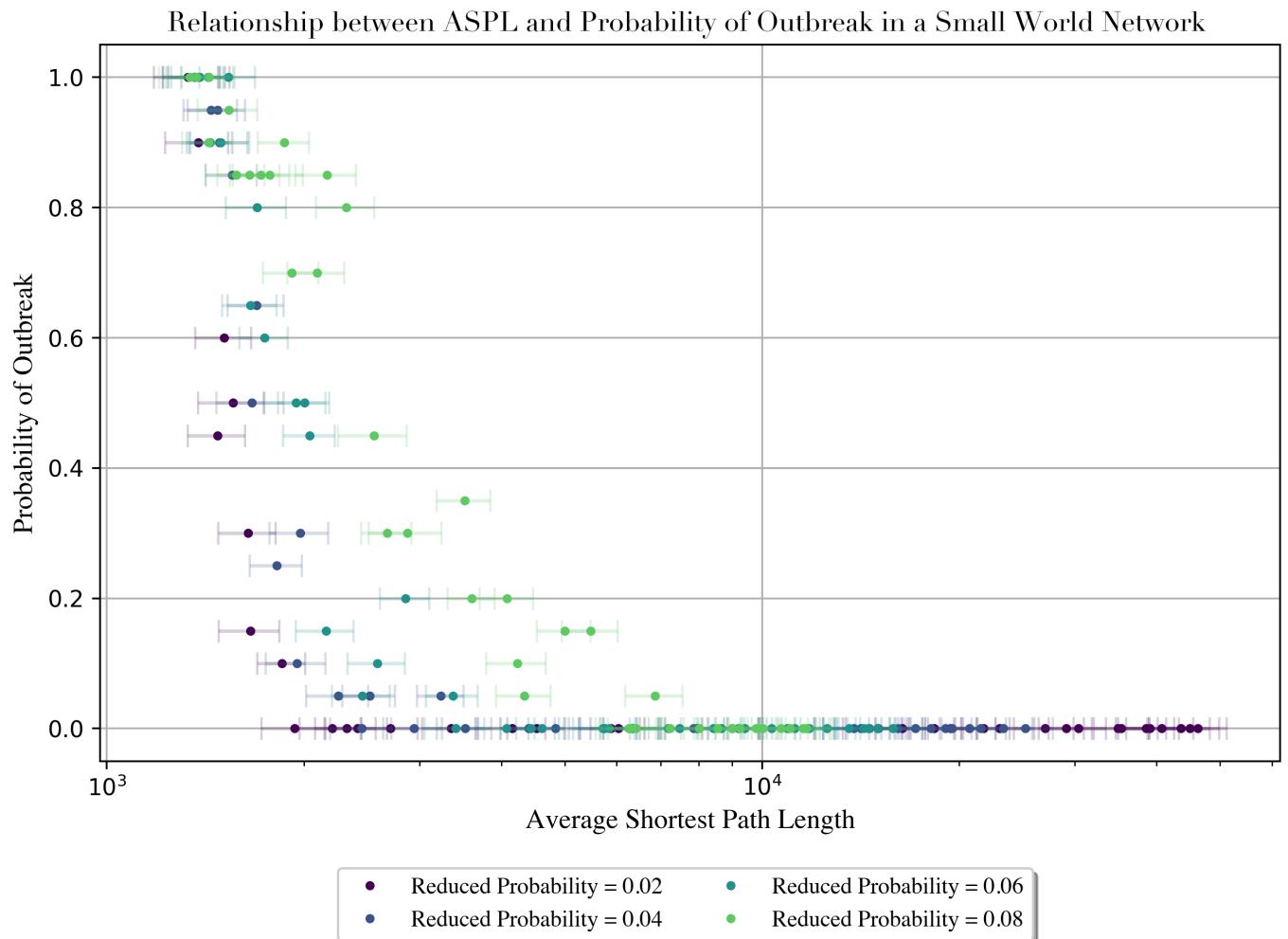


Fig. 15: For high ASPL ($\text{ASPL} > 7 \cdot 10^3$) the disease was contained with a 0% probability of outbreak. At low ASPL ($\text{ASPL} \approx 1 \cdot 10^8$) the disease was in an outbreak state with 100% probability of outbreak. The errors in ASPL represent 1σ variation from $\cdot 10^2$ simulations.

VI. DISCUSSION

During the investigation, an outbreak state was defined to be when 100% of the population was infected by the disease at the end of the simulation. This definition was used to determine the propagation time and the probability of outbreak. In the context of a pandemic, this definition is inaccurate. From a public health standpoint, a disease is considered to be significant when any unusual spikes in infection cases happens, without infecting the entire population [21]. This implies that $\text{ASPL}_{\text{critical}} = (2 \pm 0.2) \cdot 10^3$ is an overestimation.

It should also be recognised that a percolating state is not a binary status. In percolation theory, there exists a continuous variable called the percolation strength [22]. In site percolation, this is the percentage of sites which belong to the infinite percolating cluster. The non-linear transition zone of the results have a finite width and is ‘smeared out’, this is limited by the size of the simulations networks [13]. For a bigger simulation network, the width of the transition zone is narrower, as demonstrated in Fig. 16. Given a limited, small network simulated here, this introduced a finite uncertainty range to the location of the critical value.

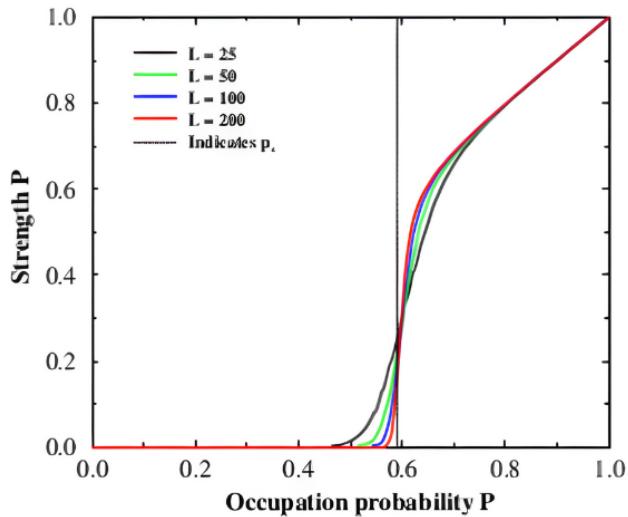


Fig. 16: Percolation strength as a function of the occupation probability in a site percolation lattice of dimension $L \cdot L$. The non-linear feature around the point of criticality is steeper for larger L , i.e. a larger networks. [13]

Interpretation of ASPL

ASPL is also known as the characteristic path length or the average geodesic distance, it is an intuitive characterisation of the size of the world represented by the network and its connectivity [23]. In this investigation, the weighting used to calculate ASPL was $1/p$: this can be interpreted as the expected number of time steps it took for the infection to pass through an edge. Hence, the weighted ASPL can be interpreted as the average expected shortest time it took for the infection to pass from one node to another. The validity of this interpretation is confirmed by considering the frequency distribution (Fig. 17) of the number of time steps it took to pass an edge. The distribution can be well-approximated by a log-normal distribution with the peak coinciding with $1/p$, as demonstrated in Fig. 18. The distribution was significantly wider for lower probabilities, indicating this interpretation of ASPL is more inaccurate for lower probability values.

With this interpretation, ASPL is a distance. It then becomes intuitive that whether the infection could percolate through the network will depend on how quickly the infection traveled and how long the infection

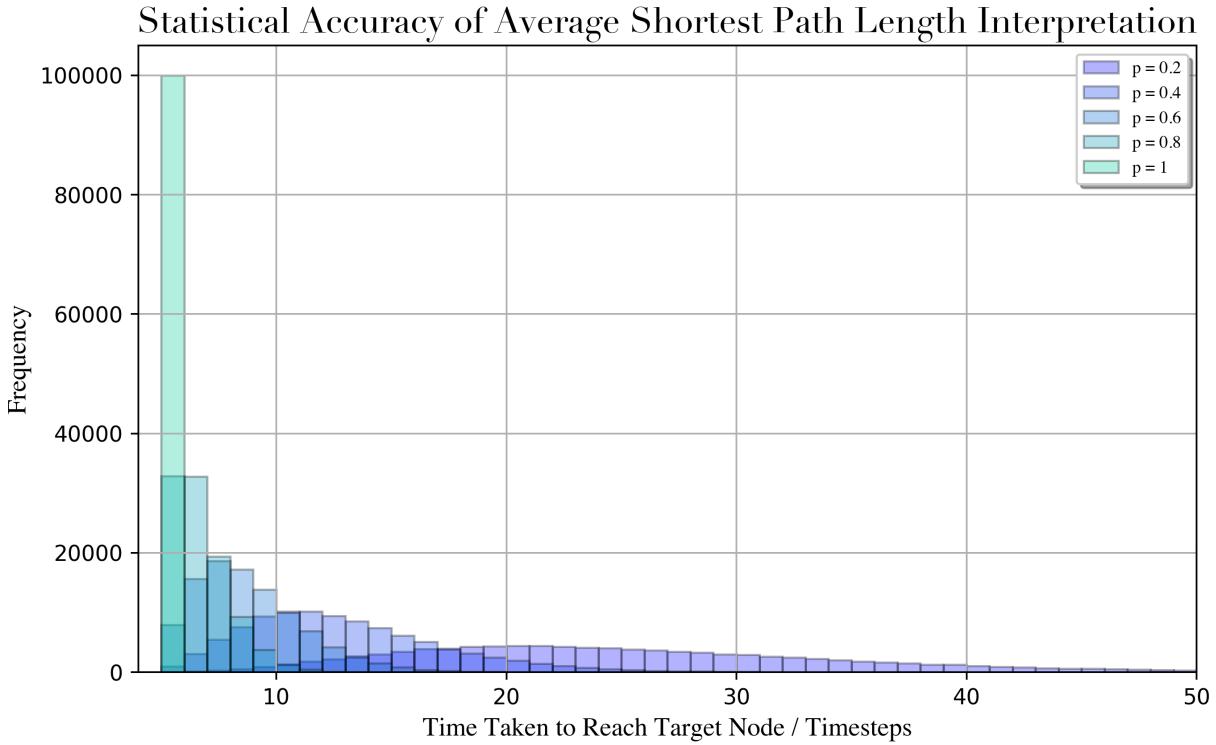


Fig. 17: The frequency distribution of the number of time-steps taken to cross an edge for different probability weightings for $1 \cdot 10^5$ trials. As the probability of infection p decreased, the curve was shifted to the right and became wider.

was travelling for. Fig. 19 verified the validity of this interpretation by considering the relationship between TTR and ASPL_{critical}. For a higher TTR, ASPL_{critical} was higher: for an infection that remains with the individual for a longer period of time, it was able to propagate and percolate through a larger distance and hence a network with a higher ASPL.

Wave Interpretation

With the interpretation of a disease spreading through a network as a travelling wave [4]. The waveform of infection can be thought of as travelling at a constant speed v_i through the network with the edges as its paths of propagation. The magnitude of the distance travelled is the path length, adjusted by the weighting. There also exists a recovery wave. The recovery wave travels at a different speed to the infection wave v_r , its speed dependent on TTR. Whilst the infectious wave speed v_i is affected by topology parameters and faces resistance from the network, the recovery wave speed v_r is unaffected by topology parameters and is only dependent on the number of time-steps a disease is infectious for. For example, an edge with mitigation adoption takes a significantly longer time to traverse, hence slowing down the travel of the infection waveform. If $v_r < v_i$, the network is in an outbreak state, as the infection wave can reach the entire population. For $v_r > v_i$, the recovery wave is able catch up and overtake the infectious wave before the entire population is infected. In the wave interpretation, ASPL_{critical} is the limiting value where the recovery waveform can catch up with the infection waveform before the entire population is infected. For a low TTR, v_r is high, as individuals recover quickly. As a result, an infection wave with a higher v_i can be caught up with.

Utilising the wave interpretation, the introduction of mitigation measures introduces obstacles in the path of the infection wave, slowing down the travelling waveform and allowing the recovery wave to catch up. This interpretation is a possible explanation for the insignificance of mitigation effectiveness at

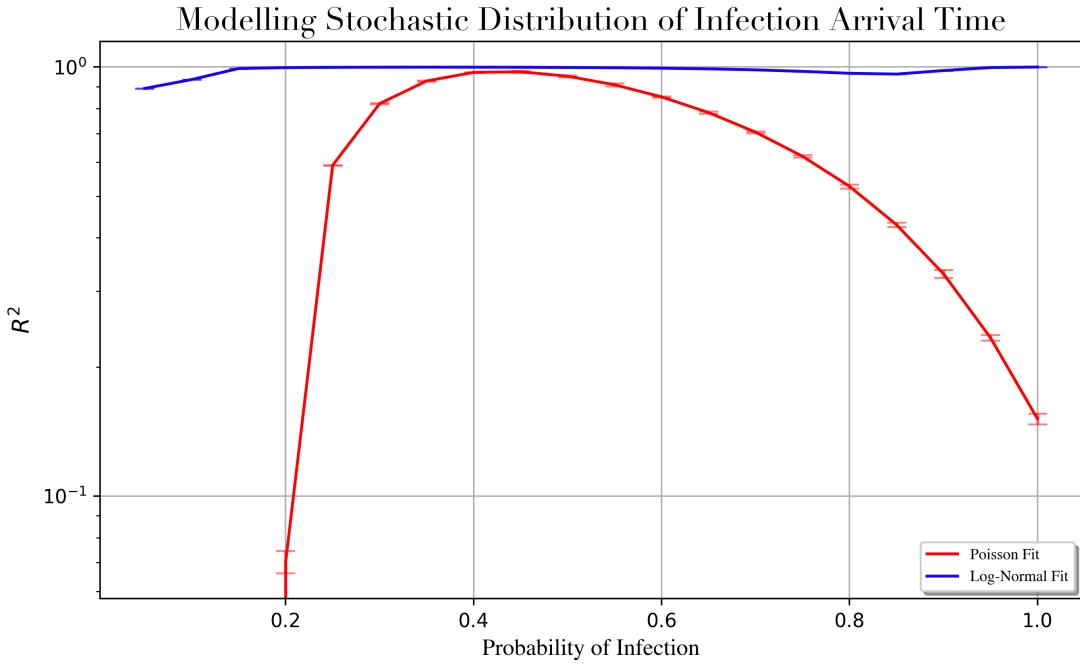


Fig. 18: When fitting the frequency distributions of time-steps taken to cross an edge for different probability weighting, the coefficient of determination varied for Poisson and log-normal function fits. The log-normal fits performed consistently well, for $R^2 \approx 0.99$ consistently. The performance of Poisson fits varied significantly. Its performance peaked for $p \approx 0.4$ and was comparable to a log-normal fit. For $p < 0.2$, R^2 was negative for Poisson fits, indicating a very poor fit.

low adoption rates: at low adoption rates, the probability of a wavefront coinciding with a mitigated edge is low, hence the overall effect of the mitigation effectiveness is negligible.

Interpretation of ASPL Dependencies

With the interpretation of ASPL as the distance through with a wavefront travels, the ASPL dependencies on topology parameters become clear. The results showed that as the probability of infection approaches 0, ASPL approaches infinity: for a disease where $p \sim 0$, we expect it to take an infinite amount of time to infect a target. This is equivalent to having to travel an infinite distance or a very slow travelling wavefront. Similarly, as the average number of connections approached 0, ASPL approached infinity: for a disease that is not easily transmittable, we expect a very slow travelling wavefront.

How Realistic is this Simulation?

The assumptions mentioned previously all have an impact on how realistic the simulations are. Assumption 1 assumed homogeneous transmission: research has shown that that susceptibility and transmission is age- and gender-dependent [24]. Assumption 2 assumed homogeneous recovery time, this is clearly wrong as the health of the infected individual can drastically increase the recovery time [25]. Assumption 3 assumed a constant population size: whilst this is not realistic, the net changes in population size over a short period of time is negligible. Assumption 4 neglected the possibility of reinfection: as witnessed during the COVID-19 pandemic, most diseases develop variants that can reinfect individuals who have previously been infected [26]. Lastly, Assumption 5 assumed the effectiveness of mitigation was not directional, i.e. the effectiveness is identical independent of whether the infected or the susceptible individual is adopting mitigation. In the case of airborne diseases, whilst any adoption of mitigation

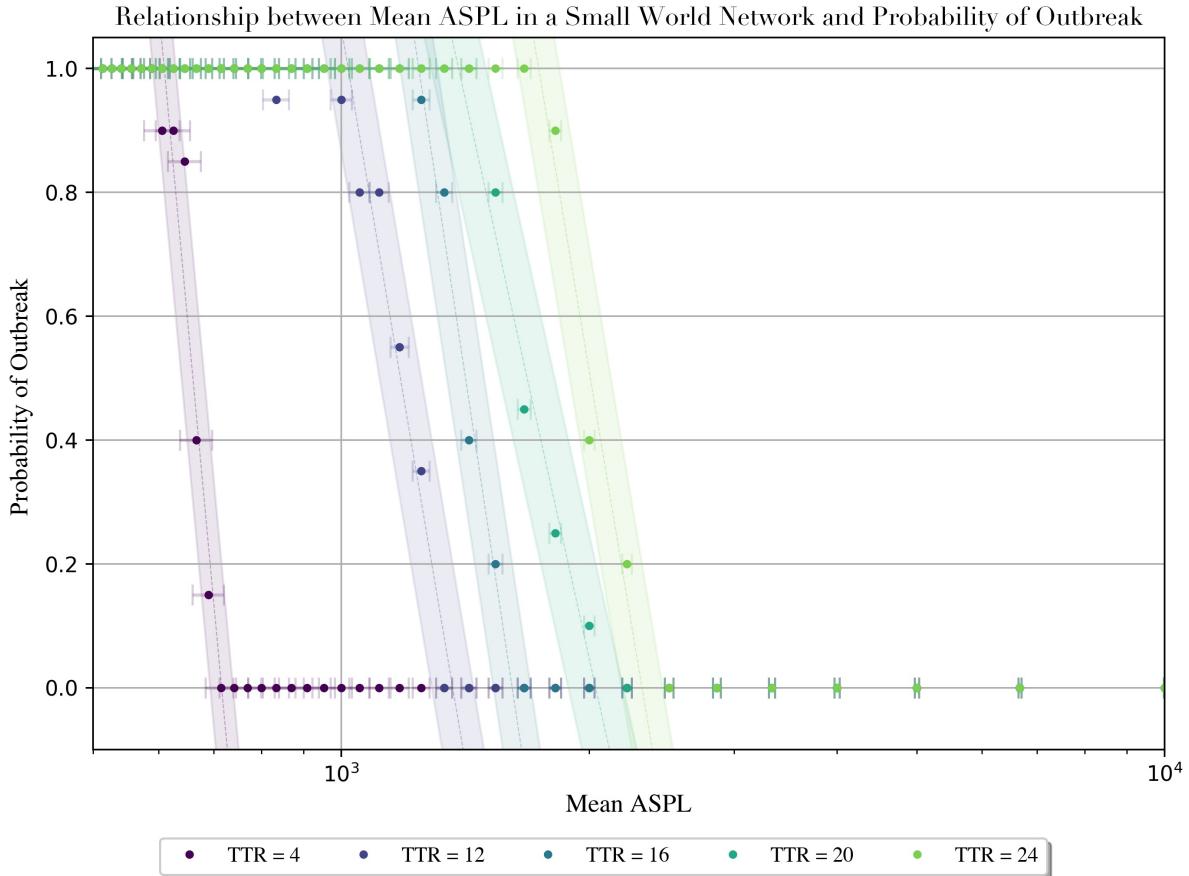


Fig. 19: The phase transition zone shifted with a variation in 'Time To Recover' (TTR). The dotted lines indicate a linear fit of the data points within the phase transition zone with the shaded region indicated the range of possible fits. The gradient of the linear fits did not change significantly with variations in TTR. The general shape of the relationship is preserved, but the phase transition zone is shifted to a high ASPL for higher TTR.

measures is productive, its effectiveness is directional and significantly more effective when adopted by the infectious agent [27].

A major weakness of the investigation was the limitation on network sizes. Majority of the simulations were of networks of the dimensions $\sim 10^5$. This is significantly smaller than the population of most cities around the world. Moreover, a larger network might demonstrate macroscopic behaviours not observable for smaller networks.

This investigation looked exclusively at small-world networks. A realistic model for social networks needs to capture nature of social interactions with an appropriate clustering coefficient and an accurate degree distribution. The clustering coefficient quantifies the existence of clusters, where a cluster is a group of well-interconnected nodes. A high clustering coefficient indicates the existence of a large number of clusters. The degree distribution describes the proportion of nodes that have different degrees (number of connections). Whilst small-world networks exhibit a high clustering coefficient, representative of families in social networks, it has an exponential degree distribution, which is not characteristic of social networks [28]. Some have argued that scale-free networks present a more realistic social network due to its power-law degree distribution, but its low clustering coefficient is its downfall [29].

VII. CONCLUSION

To conclude, this investigation was successful in exploring the role of ASPL as an intermediate explanation for the phase transitions of a disease spread in a finite network. The simulation built can be utilised as a predictive tool for disease spread, whilst the results derived from considering mitigation measures can be used to inform efficient mitigation policies. It must be recognised that this study was limited by time and resources, and would significantly benefit from future work and developments. Efforts should be dedicated to simulating with larger networks, implementing features to enhance the accuracy of the simulations and constructing a more realistic social network by considerations of cliques and degree distributions.

VIII. ACKNOWLEDGEMENT

I would like to express my appreciation to Dr. Dave Clements for his stimulating suggestions and support throughout the duration of this project. I would also like to acknowledge the precious resources that Professor Kim Christensen shared with us to contribute towards our understanding of the theoretical foundations.

IX. REFERENCES

- [1] Downey, A.B. (2012). *Think Complexity*. ‘O’Reilly Media, Inc.’
- [2] Albert-László, B.(2016). *Network science*. [online] Cambridge: Cambridge University Press. Available at: <https://www.cambridge.org/fr/academic/subjects/physics/statistical-physics/network-science?format=HB&isbn=9781107076266>.
- [3] Erdős, P and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6: 290, 1959.
- [4] Ziff, R.M. (2021). Percolation and the pandemic. *Physica A: Statistical Mechanics and its Applications*, 568, p.125723. doi:<https://doi.org/10.1016/j.physa.2020.125723>.
- [5] Zeppini, P. and Frenken, K. (2018). Networks, Percolation, and Consumer Demand. *Journal of Artificial Societies and Social Simulation*, 21(3). doi:<https://doi.org/10.18564/jasss.3658>.
- [6] Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry.—Part I. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 92(638), pp.204–230. doi:<https://doi.org/10.1098/rspa.1916.0007>.
- [7] Sun, J. and Sathasivam, S. (2023). Nonlinear dynamic model for COVID-19 epidemics using the Gaussian distributed wiring small-world network technique. *Alexandria Engineering Journal*, 81, pp.243–255. doi:<https://doi.org/10.1016/j.aej.2023.09.015>.
- [8] Poorolajal, J. (2021). The global pandemics are getting more frequent and severe. *Journal of Research in Health Sciences*, 21(1), pp.e00502–e00502. doi:<https://doi.org/10.34172/jrhs.2021.40>.
- [9] Balcan, D. et al. (2009). Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine*, 7: 45.
- [10] Namatame, A. and Chen, S.H. (2016). *Agent-Based Modeling and Network Dynamics*. Oxford University Press.
- [11] Wierman, J.C. (1978). On critical probabilities in percolation theory. *Journal of Mathematical Physics*, 19(9), pp.1979–1982. doi:<https://doi.org/10.1063/1.523894>.
- [12] Bela Bollobás and Riordan, O. (2006). *Percolation*. Cambridge University Press eBooks, pp.1–35.
- [13] Christensen, K. and Moloney, N.R. (2005). *Complexity and Criticality*. London: Imperial College Press.
- [14] networkx.org. (n.d.). *average_shortest_path_length — NetworkX 3.2.1 documentation*. [online] Available at: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.shortest_paths.generic.average_shortest_path_length.html.
- [15] networkx.org. (n.d.). *networkx.generators.random_graphs.watts_strogatz_graph — NetworkX 2.6.2 documentation*. [online] Available at: https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.watts_strogatz_graph.html.

- [16] Smith, D. and Moore, L. (2004). The SIR Model for Spread of Disease - The Differential Equation Model, *Mathematical Association of America Periodical Convergence*
- [17] Wang, Y., Deng, Z. and Shi, D. (2020). How effective is a mask in preventing COVID-19 infection? *Medical Devices & Sensors*, [online] 4(1). doi:<https://doi.org/10.1002/mds3.10163>.
- [18] Chicco, D., Warrens, M.J. and Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, [online] 7(5), p.e623. doi:<https://doi.org/10.7717/peerj-cs.623>.
- [19] Menard, S. (2000). Coefficients of Determination for Multiple Logistic Regression Analysis. *The American Statistician*, 54(1), pp.17–24. doi:<https://doi.org/10.1080/00031305.2000.10474502>.
- [20] Grimmett, G. (1999). *Percolation*, 2nd ed. Berlin: Springer-Verlag.
- [21] Fischer, R.S.B. (2020). *What's the Difference between Pandemic, Epidemic, and Outbreak?* [online] JSTOR Daily. Available at: <https://daily.jstor.org/whats-the-difference-between-pandemic-epidemic-and-outbreak/>.
- [22] Stauffer, D. (1979). Scaling theory of percolation clusters. *Physics Reports*, 54(1):1–74. URL: [http://dx.doi.org/10.1016/0370-1573\(79\)90060-7](http://dx.doi.org/10.1016/0370-1573(79)90060-7), doi:10.1016/0370-1573(79)90060-7.[SA94]
- [23] Browne, C.A., Amchin, D.B., Schneider, J. and Datta, S.S. (2021). Infection Percolation: A Dynamic Network Model of Disease Spreading. *Frontiers in Physics*, 9. doi:<https://doi.org/10.3389/fphy.2021.645954>.
- [24] Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M. and Eggo, R.M. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, [online] 26(8), pp.1–7. doi:<https://doi.org/10.1038/s41591-020-0962-9>.
- [25] Voinsky, I., Baristaite, G. and Gurwitz, D. (2020). Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients. *Journal of Infection*, 81(2), pp.e102–e103. doi:<https://doi.org/10.1016/j.jinf.2020.05.026>.
- [26] Pinto, L.M., Nanda, V., Sunavala, A. and Rodrigues, C. (2021). Reinfection in COVID-19: A scoping review. *Medical Journal Armed Forces India*, 77, pp.S257–S263. doi:<https://doi.org/10.1016/j.mjafi.2021.02.010>.
- [27] Du, M. (2021). Mitigating COVID-19 on a small-world network. *Scientific Reports*, 11(1). doi:<https://doi.org/10.1038/s41598-021-99607-z>.
- [28] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), pp.440–442. doi:<https://doi.org/10.1038/30918>.
- [29] networkx.org. (n.d.). *clustering — NetworkX 3.2.1 documentation*. [online] Available at: <https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.cluster.clustering.html> [Accessed 2 Apr. 2024].