THE UNIVERSITY OF HULL

Using Small World Models to Study Infection Communication and Control

being a Thesis submitted for the Degree of PhD

in the University of Hull

by

Paul Sefton Ganney B.Sc. (Hons), M.Sc., Cert.Ed.

April 2011

Abstract

The modeling of infection transmission has taken many forms: The simple Susceptible-Infected-Removed (SIR) model yields good epidemiological results, but is not well suited to the modeling of the application of interventions. Attention has focused in recent years on graph (network) models and especially on those exhibiting the small-world properties described by Watts and Strogatz in "Nature" in 1998. This thesis examines such graph models, discovering several attributes which may yield improved results. In order to quantify the effects of these proposals, a classification system was developed together with a Goodness-of-Fit (GoF) measure. Additionally, a questionnaire was developed to reveal the operational organisational structure of the NHS Trust being examined. The resultant theoretical model was implemented in software and seeded with a graph derived from this questionnaire. This model was then examined to determine the effectiveness of these proposals, as measured via the GoF. The additional features proving beneficial were shown to be: full directionality in the graphs; modeling unknown paths via a new concept termed an "external path"; the division of the probability of infection transmission into three components; the seeding of the model with one derived from an organizational questionnaire. The resulting model was shown to yield very good results and be applicable to modeling both infection propagation and control.

Contents

List of Figures

List of Tables

List of Public Output

Publications

A classification system for hospital-based infection outbreaks

Paul S. Ganney, Maurice Madeo, Roger Phillips

Computational and Mathematical Methods in Medicine, Vol. 11, No. 4, December 2010, 297-311.

This paper is a condensed version of chapter 5 of this thesis.

Conferences – Accepted and Presented, with Published Abstracts

A Framework for Investigating the Spread of Infection in a Hospital Environment Utilising a Stochastic and Small-World Model

Ganney P

IPEM Annual Scientific Meeting, Bath, 17 September 2003.

Proceedings of the Annual Scientific Meeting of the Institute of Physics & Engineering in Medicine, 2003 p 105

Modelling Infection Propagation using Small World Theory

Paul S. Ganney

Clinical Biosciences Graduate Research Conference, 8 December 2006.

Proceedings of Clinical Biosciences Institute Graduate Research Conference, 2006 pp 17-18

A Graph Model of a Hospital Environment for Modelling Infection Propagation

Ganney P S

IPEM Annual Scientific Conference, Cardiff, 11 September 2007.

Proceedings of the 13th Annual Scientific Meeting of the Institute of Physics & Engineering in Medicine, 2007 p 24

Acknowledgements

I would like to thank the Hull And East Yorkshire Hospitals NHS Trust for their support of the initial years of this work, particularly Chris Langton and Glyn Davies. Special thanks go to Roger Phillips for his supervision, Maurice Madeo for very helpful discussions, Steve Strogatz for an inspiring meeting at short notice and Rachel Ganney for her patience, proofreading and overall support and encouragement.

1 Introduction

This research is concerned with the transmission of an infection within a hospital environment. It presents contemporary work in the general fields of graph (or network) theory and infection propagation/control alongside work linking the two together. Additionally, new work is presented defining new concepts and investigations. This chapter of the thesis sets the outline context for the ones that follow (a fuller examination is in Chapter 2) and outlines the hypothesis that the thesis investigates.

1.1 Context

An infection is an invasion of the human body by a parasitic organism[1] which attaches itself to the body, the inside of the body or to another organism present within the body and in so doing contaminates the host body and causes disease[2] (Smith et al. 1997; Martin 2002). This disease is usually the result of local cellular injury and may become systemic if such microorganisms gain access to the lymphatic or vascular systems. The idea of infections being caused by such parasites (otherwise known as germs theory (Martin 2002)) is relatively recent in medical history, having reached acceptance in the late 19th Century[3] through the work of scientists such as Louis Pasteur[4], Joseph Lister[5] and Robert Koch[6]. Although controversial when first proposed, it is now generally accepted and forms a cornerstone of modern medical practice.

---

[1] Whilst this definition does include such parasites as tapeworms and flukes, this thesis is concerned more with microorganisms. Such invasion by matzoons (e.g. intestinal worms) are more commonly termed an infestation (Macpherson 2002), although Last (2008) reserves this term for a surface invasion (e.g. scabies).

[2] The normal growth of the usual bacterial flora in the intestinal tract meets most of this definition but is not usually considered an infection. The same applies to the bacteria that normally inhabit the mouth.

[3] There is a reference in *On Agriculture* by Varro, published in 36BC, to minute organisms that cannot be seen, that enter the body and cause disease, but this was a minority view.

[4] Pasteur showed that organisms in the air spoil food.

[5] Lister used antiseptics to prevent germs in the air from causing infection.

[6] Koch was the first to link a specific organism to a specific disease, in this case Anthrax.

Almost all infections that humans can acquire are transmitted from another human or from an animal[7]. This transmission may use an intermediate object but the beginning and end of the transmission is human/animal. It therefore follows that if the routes of transmission can be made safe then an infection will not propagate but instead will die out. During the bubonic plague of 1665-1666 villages attempted to restrict the spread of the plague by sealing themselves off. In one noted case (Eyam in Derbyshire) the village sealed itself off once plague struck, thus containing the spread, but at huge cost to the local population as approximately 50% of the village contracted the disease and died (www.eyamplaguevillage.co.uk, amongst others).

The most successful implementations of tackling the route of transmission have been in the development and worldwide availability of vaccines, which render the vaccinated individual immune from the infection, thus destroying the transmission route. There are varying degrees of success of vaccination programmes, with the near-eradication of smallpox[8] being probably the most successful. Conversely, reduction in take-up has seen an increase in infections such as Measles (Jansen et al. 2003).

However, not all infections currently have vaccines. Some infections are proving very difficult to develop vaccines for (Graham et al. 2009) whilst others (such as the common cold) mutate so quickly that a search for a vaccine may forever prove fruitless. In the absence of a suitable vaccine, or during an outbreak (if the vaccine requires significant time for an individual to develop immunity), an alternative and speedier approach is required.

[7] The other transmission routes: indirect (soil or surface contamination) and airborne are considered, for this research, to have originated with another human and are discussed in Chapter 6.

[8] Certified by the WHO in 1977 as being no longer present in the general population, with samples only existing in laboratories.

As has already been noted, the human-to-human transmission route is the one followed by most infections[9] and is certainly the one that is easiest to understand and therefore control. It follows that an alternative method of infection control to the vaccine method is to remove or restrict the transmission route (Hawker et al. 2005). In cases where the infection is not susceptible to vaccine or the effects are not especially life-threatening, understanding the transmission routes leads to better prediction and therefore improved coping strategies.

Experiments on infection propagation within a population are not generally ethical to carry out. Mathematical models (e.g. Ancel Meyers et al. 2003; Carrat et al. 2006; Dezsö & Barabási 2002; Eames 2008; Moslonka-Lefebvre et al. 2009; Pastor-Satorras & Vespignani 2001; Saramaki & Kaski 2004; Small & Tse 2005; Toroczkaia & Guclu 2007; Vanderpas et al. 2009; Verdasca et al. 2004; Witten & Poulter 2007) do not contain such restrictions and may be run and re-run multiple times in order to investigate the effect of randomness as well as differing control regimes and methods.

In order to understand the transmission routes, the contact network of all humans (but especially those who are infected) must be known and mapped (see Auerbach et al. 1984 for a noted attempt at this). This contact network will not be static, but dynamic, reflecting the constant shifting in the human population, the movements of individuals and the social networking that takes place.

1.2 Scope

The total contact tracing approach described above may be a theoretical method for understanding and controlling infection, but is certainly not a practical one at present. Therefore, in order to be implementable and thereafter usable, the idea must be reduced in one or more ways. By so reducing the scope of the theoretical model a practical model may be produced.

[9] Malaria is one of the most prolific diseases in the world today, killing between 1 and 3 million people annually from 350-500 million cases. It is spread by mosquitoes. Such diseases, no matter how deadly, are the minority.

Firstly, the population is limited to a defined geographical area or structure (e.g. Ancel Meyers et al. 2006; Griffith 2003). This eliminates a large part of the population who thus do not require modelling. Secondly, a slowly-changing population is required. This reduces the dynamic elements of the model and means that sufficient structure will always remain. Thirdly, a population where most of the contact network is prescribed is envisaged. This enables the contact network to be more easily collected, as it is prescribed by the organisational structure and procedures.

A hospital environment fulfils these criteria. It is limited in geographical area to a few sites, staff turnover is low (and staff and patient movement controllable), and the main contact points are function-related. Additionally, hospital-acquired infection is an area of considerable interest at present (e.g. Chadwick et al., 2000; Ganney 2003; Gleizes et al., 2006; Gould, 2006; Lau et al. 2004; Lynn et al., 2004; Pittet et al. 2000) and therefore much data exists for testing and validating a model with.

The infections that are to be studied are therefore limited to those that, whilst occurring outside of the hospital environment, cause major problems when contracted and are propagating within it. Infections such as MRSA and Norovirus are particularly prevalent in institutions such as hospitals, care homes and cruise ships (Barker et al. 2004; CDC 2006; Lynn et al. 2004), mostly due to the contained environment but also due to the reduced/suppressed immune systems of those that are there (patients and residents respectively – this does not apply to cruise ships).

1.3 Justification of Research

There has been much interest in recent years in the two main strands of this work: in models of hospital-acquired infection (e.g. Noakes et al. 2006; Vanderpas et al. 2009) and in models of human interaction (most notably work following Milgram 1967). However, there has not been any published work linking the two together. Whilst the work of Ancel Meyers et al. (2003; 2006) has looked at using network models to investigate contact networks that may be used to transmit infection, it

has concentrated on large municipal models and not investigated specifically hospital ones. Therefore such work has lacked the detail that this work brings.

Mathematical models of infection propagation throughout populations (particularly when used for epidemic prediction) have normally been of a compartmental-deterministic statistical variety (e.g. Vanderpas et al. 2009). These give good results in terms of numbers contracting the disease and may show up some propagation routes. They are very good when used to model reservoir or general proximity-based infections but, by their nature, cannot accurately model more contact or close proximity-based ones. A simplistic view of a deterministic verses network approach would be to say that both will tell you how many will become infected, but the network model is more likely to tell you where (or possibly even who) they are likely to be. A network model therefore better lends itself to an investigation into control regimes that are liable to be short-term, such as confinement, pharmaceuticals or targeted immunisation.

1.4 Hypothesis

The hypothesis of this thesis is:

- That it is possible to produce a network model of a hospital environment that incorporates individuals.
- That such a model can be used to demonstrate the effects of an infection within the hospital.
- That such a model can be used to investigate the efficacy of differing infection control regimes, especially those influencing the transmission routes.
- That such a model will be able to inform infection control decisions prior to and during an outbreak, thus reducing its duration and overall effect.
- That a method of categorising and thereby comparing outbreaks can be devised.
- That such a categorisation may be adapted to determine how realistic an outbreak on the proposed model is.
- That such a model may lead to the development of one that is superior to existing models.

1.5 Aim and Objectives of Research

The aim of this research is to produce a network model of the social contacts that exist within a hospital environment. Once created, this model will then form a tool that can be used to investigate the likely effects of an infection outbreak and give planners opportunity to investigate differing methods of controlling such an outbreak.

In order to achieve this, existing approaches will be examined and any shortcomings will be identified. Theoretical concepts will be proposed in order to overcome these. These theoretical concepts will then be modelled in software. Data will be sourced to test, derive parameter values and validate the model in order to produce an effective, reliable tool for studying infection propagation and control within a hospital environment.

A classification system will be determined and used to compare real and modelled outbreaks in order to determine how realistic the models are.

1.6 An Overview of the Thesis Content

This thesis is structured into 7 Chapters and 5 Appendices. A brief description for each Chapter or Appendix is given below.

Chapter 2 examines the existing literature in the fields of Infection Propagation and Control, Graph Theory and a particular branch of Graph Theory, Small World Theory. The study of Infection Propagation and Control is limited to those parts that are relevant for this thesis. The section on Graph Theory is similarly limited to the study of how properties (such as infections, packages or gossip) move within social structures. The Small World Theory section describes and examines the two key publications in this field and then goes on to describe other relevant work within it.

Chapter 3 describes the problem that is being addressed, examining the hospital environment (with particular reference to infection propagation) and describes four specific infections that are particular problems within such an environment. Finally, other infection propagation and control modelling methods are examined.

Chapter 4 presents the mathematical model that has been developed from and in response to the information presented in Chapters 2 and 3. It then describes the high-level solution that this thesis proposes and the extensions to graph theory that this proposal necessitates. It then describes how these elements are used in the model, showing the implementation of each and contextualising them.

Chapter 5 describes a categorisation methodology developed for this research. It examines the usefulness of this method in describing an outbreak and then investigates how this categorisation may be used in validating computer-generated models.

Chapter 6 describes the case study undertaken in this thesis. It describes the structure of the chosen organisation, the collection of organisational data and the construction of a software model of it, based upon the model presented in Chapter 4. The model (basic and enhanced) is investigated using the categorisation method described in Chapter 5. The model is then used to analyse and investigate differing outbreaks and infection control methods and the results and discussion are presented. Finally, three scenarios of using the model are presented.

Chapter 7 presents some conclusions from the work undertaken so far and describes future work and plans.

Appendix A describes the scripting language developed to repeatedly generate the models used within this thesis.

Appendix B describes the structure of saved computer files from generated models.

Appendix C describes the random models the software produces – their generation, visualisation and outputs/metrics.

Appendix D gives a very brief and high-level description of the software created to undertake the research presented in this thesis.

Appendix E contains large sets of data that have been summarised within the main chapters, but are referred to therein.

Finally, the references used within this thesis are given.

2 Background and Literature Review

2.1 Introduction

The study of the spread of infection by simulation has traditionally been studied utilising a compartmental-deterministic model. Whilst giving admirable results in numerical terms, they have not modelled the propagation throughout a social network and therefore do not lend themselves well to studying that propagation or methods of preventing such spread.

In recent years there has been a great deal of interest in the use of graphs to model networks, both physical and logical. One such avenue of research has been in the spread of information, rumour or infection throughout a population (e.g. Antal & Balogh, 2009 on belief systems). It has been reflected that such studies will aid the understanding of propagation, as the structure of a network affects the operations performed upon it.

Although graphs have been studied since the 1950s, it was the publication of a brief paper in 1998 in Nature by Watts & Strogatz (see description in 2.4.2.2) that caught the imagination of many researchers and their small-world model formed the basis of a further flurry of publications.

The "small-world" effect was itself not a new concept, though, having been the subject of a paper in 1967 by Milgram (see description in 2.4.2.1). This paper, whilst clearly novel, was regarded as little more than a curious result and was largely overshadowed by his work on authority and obedience (especially his famous experiment, published in 1963 and repeated several times by others, where subjects were instructed to give "electric shocks" to an actor).

Regardless of how old the ideas may be, they have now spawned many avenues of investigation, of which the research described in this thesis forms a part. Of particular interest here is the spread of infection, and methods of understanding, preventing or constraining this.

The fight against the spread of infection is of paramount importance. To fight it, one has to have an understanding of how it works and how it spreads. Realistic models are therefore important tools in gaining this understanding and in allowing researchers to investigate different scenarios of spread and different techniques for combating it. Additionally, in epidemic research, there is great value in models that can predict the signs of infections that will become epidemics, and those which will not, from early data. The threat of pandemics of new diseases (or variants of existing ones)[1] give rise to demands for mathematical and computer models of infection propagation partly to understand the mechanisms involved, but mostly to inform the countermeasure strategies.

Many infectious diseases spread through direct person-to-person contact. Respiratory-borne diseases like influenza, tuberculosis, meningococcal meningitis and SARS, spread through the exchange of respiratory droplets between people in close proximity to each other. Sexually transmitted diseases like HIV, genital herpes and syphilis spread through intimate sexual contact, yet some (such as HIV) are more easily caught by women than by men during heterosexual encounters. (Ancel Meyers et al., 2006, p 401)

As already mentioned, the study of social, technological and biological networks of various kinds has been the subject of a large number of recent publications (see Strogatz, 2001; Newman, 2003; Witten & Poulter, 2007 for comprehensive reviews). One of the principal practical applications of such work is in modelling the spread of disease, especially that which may lead to an epidemic. Of particular interest are diseases that spread across networks of people by utilising the physical contacts between them ("physical" is here taken to include "in close proximity (spatial and temporal)" in order to include air-borne infections such as those listed above). Such networks may easily be mapped (and hence modelled) utilising a simple representation of people as dots and contacts as lines between them.

[1] SARS, "Bird Flu" and "Swine Flu" were recent examples..

Mathew has contact with Mark and Luke

Mark has contact with Matthew and John

Luke has contact with Matthew and Mark

John has contact with Mark

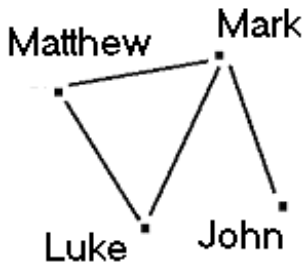


Figure 2.1: A simple social network

Such networks are simple to draw (assuming that a robust definition of "social contact" can be established[2]) and, once so modelled, it is appropriate to apply graph theory techniques to their analysis. In practice, however, such a definition appears to have proved to be difficult to establish for large populations and therefore "realistic" networks are normally developed and their properties investigated[3].

The class of graphs generally assumed to be "most realistic" are those termed "scale-free", after the observation by Barabási & Albert (1999) that "*a common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution*". However, as Li et al. (2005) observe, there is little agreement over what really constitutes a scale-free network, a matter their paper seeks to resolve.

An alternative starting point to such investigations was proposed by Watts & Strogatz (1998). This began from a simple, uncomplicated model and added elements of randomness to it by re-wiring the links between vertices. Their work is discussed in 2.4.2.2., below.

---

[2] The largest database of people subjected to network analysis is probably the Internet Movie Database of screen actors, where a "contact" is defined as "appearing in the same film as". This is clearly robust and therefore makes the database amenable to such analysis. Possibly the most famous such analysis is the "Kevin Bacon Game", the object of which is to link a chosen actor to the actor Kevin Bacon in as few contacts as possible (the "Bacon number"). (Smith, 1996)

[3] One notable exception is the work of Ancel Meyers, who has mapped the social networks of Vancouver (Griffith, 2003) – see figure 2.14.

There have been many recent publications on the subject of network or graph analysis, often characterising and analysing the structure of the graph as much as any operations that takes place upon it. Such characterisation of network anatomy is important as structure always affects function: "*the topology of social networks affects the spread of information and disease*" (Strogatz 2001, p.268), an assertion that appears not to always be appreciated.

The mechanics of infection propagation and control are described in 2.2, moving on to a description of the relevant aspects of Graph Theory (2.3), including terminology and metrics, before describing small world theory (2.4) paying particular attention to the papers by Milgram (2.4.2.1) and Watts & Strogatz (2.4.2.2).

Some methods of modelling disease propagation are briefly described in 2.5, with a fuller examination of the applications of graphs within an infection propagation context (2.6) including some elements of control (2.6.3).

2.2 Infection Propagation and Control

Infections propagate through several methods, but may be loosely categorised as:

- Person-to-person (via no intermediaries – fluid exchange, blood exchange etc. Examples include AIDS/HIV and hepatitis B).
- Person-to-air-to-person (airborne droplets or small particle aerosols. Examples include chickenpox and tuberculosis).
- Person-to-fomite-to-person. (Non-living third party. Examples include legionellosis and tetanus).
- Person-to-lifeform-to-person. (Plant or animal. Examples include malaria and encephalitis, which are spread by mosquitoes, and rabies).
- Person-to-ingestible-to-person (where "ingestible" includes food, fluids such as water and pellets such as faecal or aerosol. Examples include norovirus and clostridium difficile).

(Examples from Hawker et al., 2005 and South Australia Department of Health, 2005).

Whilst each has differing characteristics, the overall pattern is of an infection passed from one person to another, possibly via a third party. The ability of an infection to survive such third-party transmission[4] will affect the overall progress of an outbreak and thereby influence the control methods that may be deployed.

The environment may have an effect on transmission. Influenza transmission, for example, is facilitated by overcrowding and enclosed spaces (Hawker et al., 2005, p 136) as "anyone within a metre of an infected person who coughs or sneezes is at risk of inhaling infected droplets."[5]

Ignoring the issue of treating the infected (by which point they may well have passed the infection on), control methods can be viewed as falling into five main groups:

- Containment – restricting the movement of infected cases. This may be individual (e.g. "stay at home") or group (e.g. the cancellation of mass population movements and gatherings, such as football matches).
- Isolation – remove infected cases and place in a special environment. This may also include the erection of barriers, for example in the case of insect-borne infections.
- Separation/Segregation – identify and break lines of transmission (e.g. if schools are seen as primary reservoirs, closing them will reduce transmission rates). This may also include gathering infected cases together, thereby separating the infected from the susceptible (see 2.5.1 below). Other approaches include excluding highly susceptible individuals from high-risk environments (in a hospital setting this would apply to visitors, not patients).

[4] For example, Influenza A and B survive for 24-48 hours on hard nonporous surfaces such as stainless steel but for only five minutes on hands. (Bean et al, 1982, p 47) – although the BMA web site (www.bma.org.uk/health_promotion_ethics/diseases/viralrespiratorydisorders.jsp?page=3) stated this as being ten minutes when accessed in 2009.

[5] "Administration of Holy Communion during a Flu Pandemic", Church of England website, 2009.

- Information (e.g. influenza transmission is reduced by increased personal hygiene (Hawker et al., 2005, p 138). The use of anti-bacterial handwash has reduced on-ward infection rates (Pittet et al., 2000, p 1307). Both of these measures were heavily supported by public information programmes).
- Immunisation. This requires not just the development of a vaccine (in itself not a simple task) but also the identification of the most effective deployment of the vaccine. Mass immunisation programmes such as mumps and measles, carried out during childhood, are not appropriate to the kind of outbreak considered here which requires an immediate response – in this case it may be because the vaccine (if it exists) only confers short-term immunity (e.g. the influenza vaccine is re-administered annually in the UK) or the disease mutates regularly enough for previous vaccines to be ineffective (e.g. the delay in the production of a vaccine for Swine flu in 2009). The most common approach is to identify the most susceptible (e.g. the elderly), although it may include likely carriers (e.g. schoolchildren) and those in both groups (e.g. healthcare workers). This latter group may also be identified via a "minimisation of disruption" identification scheme.

2.3 Graph Theory

Graph theory is a large subject area and only those concepts pertinent to this thesis are rehearsed here. Likewise, only relational graphs (as opposed to spatial[6]) are considered.

2.3.1 Terminology

Graph theory is a branch of mathematics devoted to the analysis of networks. A *graph* in this context is a collection of *vertices* and *edges*, which interconnect them. It has been used to describe and investigate physical concepts such as road systems, electrical circuits, atomic bonds and computer networks (especially the

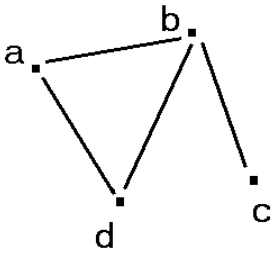
[6] Relational graphs are one where the construction depends upon relationships alone; Spatial graphs are ones where the construction depends on distances between vertices. Only relational graphs exhibit the small-world effect.

Internet); relational concepts such as matches played between football teams; and social concepts such as “friend” and “acquaintance”. The graph is therefore an abstract representation of the circuit, history or social relationship that is under study.

A *graph* comprises a finite non-empty set V of *vertices* (alternatively “nodes” or “sites” from percolation theory) and a finite set E of *edges* joining them to one another (note that this can be empty). E is mathematically defined as a bag. This is Wilson (1985)’s approach and is simpler (and therefore easier to analyse mathematically) than the more rigorous requirement that a graph has an end-point function ∂ such that, for each $e \in E$, $\partial(e)$ is the set of vertices which e joins. Thus, for each $e \in E$, the set $\partial(e)$ contains one or two vertices. (If $\partial(e)=\{v,w\}$ then e joins v and w. If $\partial(e)=\{v\}$ then e is a loop).

A *path* is a route traversed between two vertices along edges where each intermediate vertex is distinct (and therefore, by implication, each edge is also distinct). A path’s *length* is the number of edges so traversed. This is a special case of a *walk* (where loops and repeated edges are possible) and a *trail* (where repeated vertices are possible but all edges are distinct). A *path* is therefore the shortest *walk* or *trail*.

An *adjacency matrix* is a mathematical representation of a graph, where the elements are 1 or 0 indicating whether or not two vertices are connected by an edge. For non-simple graphs (*simple* is defined on the next page), higher values may be used to indicate the number of edges between a vertex pair.



$V = \{a,b,c,d\}$
$E = \{ab, ad, bc, bd\}$

Note that order in the sets is unimportant.

The path (c,d) has one intermediate vertex (b) and its length is 2 (edges bc and bd). This is the shortest path as another (via b and a) exists with length 3.

The trail (c,b) passes through all vertices (c, b, a, d, b or c, b, d, a, b) but uses each edge only once. There also exists a shorter trail with length 1.

The walk (c,b) may complete the loop involving a and d multiple times. It's length is therefore one of 1, 4, 7, 10, 13, …

The adjacency matrix is

$$\begin{array}{c} \begin{matrix} a & b & c & d \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \end{array} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$

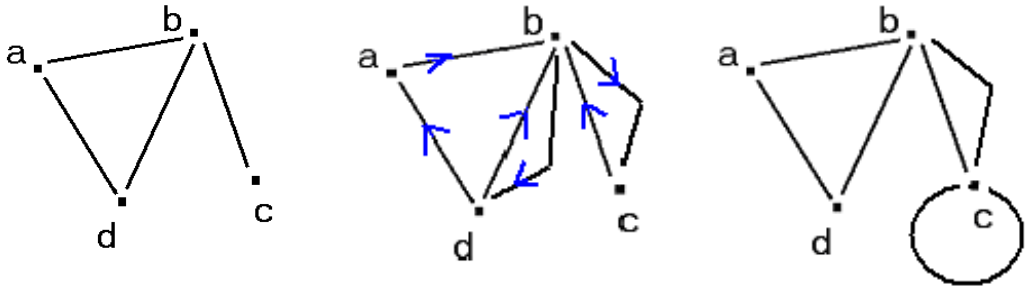
Figure 2.2: A simple graph

A *simple graph* is one in which multiple edges between the same pair of vertices or edges connecting a vertex to itself are forbidden. That is, if an edge exists between two vertices, then it is unique; loops do not exist. A *non-simple* graph is therefore one in which there may be multiple edges between two vertices and loops (edges that start and end at the same vertex) are permitted. Thus Wilson (1985)'s approach espoused above is sufficient if only simple graphs are considered.

A *connected* graph is one in which a path exists from every vertex to every other vertex in the graph. A *fully connected* graph is one in which an edge exists between every vertex and every other vertex within the graph. Hence the graph in Figure 2.2 is connected, but not fully connected. (*Fully connected* is sometimes referred to as *complete* and the complete graph with n vertices denoted K_n).

A *directed graph* (or *digraph*) is one in which the edges exhibit an inherent direction, so that an edge from a to b does not imply the existence of an edge from b to a. An *undirected* graph is therefore one in which the edges are directionless, implying that any relationship between the vertices at the terminus of the edge is symmetric. A *strongly connected* directed graph ("strongly connected" is meaningless for undirected graphs) is one where paths exist from *a* to *b* and from *b* to *a* for all *a,b*. (Durr, Mhalla & Lei, 2003, p 1).

The Giant Strongly Connected Component (GSCC) is the largest set of vertices for which you can move between any two in the set by following edges in the correct direction. (A giant component is one containing more than 50% of the vertices in a graph). A random graph constructed by placing *n* nodes on a plane, then randomly connecting pairs until *m* links have been constructed gives an expected single giant component when $m>n/2$ (Strogatz, 2001, p 271).



Simple, undirected. If the edge *bc* is removed the remaining graph has a GSCC comprising {*a, b, d*}

Simple, directed. Also strongly connected, although the removal of any one edge (apart from *db*) would remove this property.

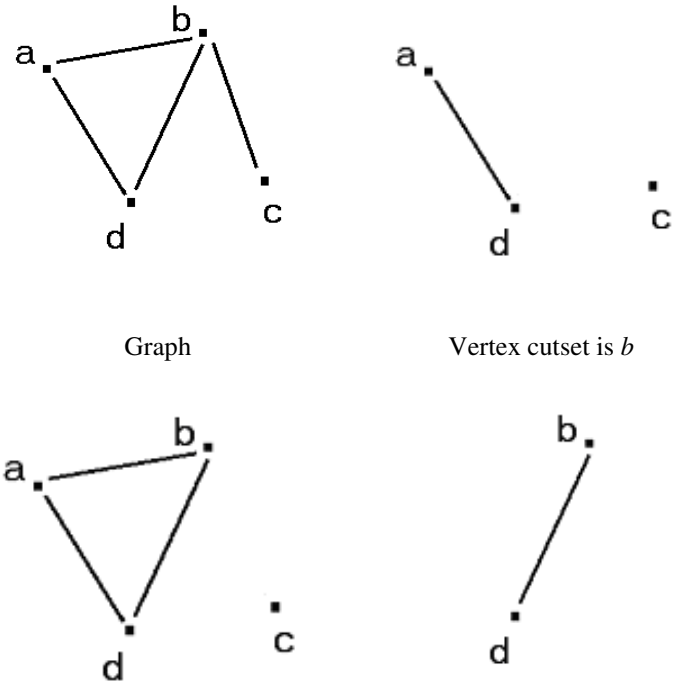
Non-simple (or *multigraph*), undirected

Figure 2.3: Classes of graph

A *subgraph* is a graph comprised of a subset of vertices and edges from the main graph. i.e. a graph H is a *subgraph* of a graph G if $V(H) \subseteq V(G)$ and every edge of H is also an edge of G. This is written as $H \leq G$.

A *cutset* of a graph is a set of vertices or edges (depending on the context) that, if removed from the graph, disconnects the graph. A cutset is therefore a subgraph.

One other term was introduced via SWT (Small World Theory – see section 2.4), but is described here for completeness. This term, *neighbourhood* (Γ_v), is the subgraph that consists of the vertices adjacent to *v* but does not include *v* itself.



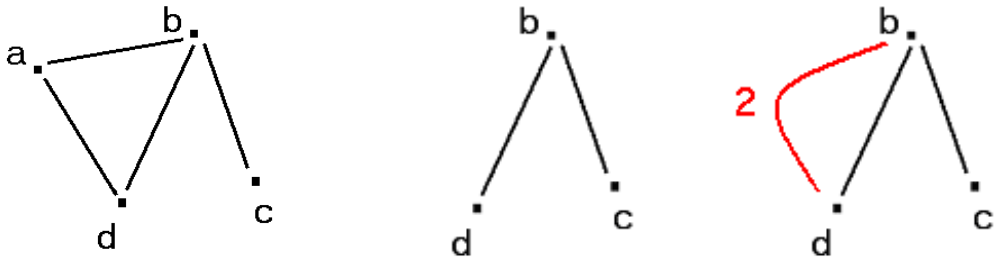
Graph Vertex cutset is *b*

Edge cutset is *bc* Neighbourhood of *a*

Figure 2.4: Subgroups of a graph

This research introduces two new concepts: *external path* and *system*, which are briefly described here but explored more in chapter 4.

An *external path* is a path between two vertices in a sub-graph that utilises edges that are part of the graph, but not of the sub-graph. Therefore, as the vertices in the graph that are not in the sub-graph are unknown to the sub-graph, only the path's length is known. (See 4.2.2.2 for a fuller discussion).



Graph Sub-graph External path (*bd*), length 2

Figure 2.5: Illustration of an external Path

A *system* is comprised of a sub-graph together with the set of external paths. The adjacency matrix for the system is thus equivalent to that for the original graph, as they are equivalent save for the knowledge of the vertices on the external paths.

2.3.2 Metrics

The *degree* of a vertex v, k_v, is the number of edges incident with a given vertex v. For directed graphs this term is meaningless and is replaced by *indegree* ($\bar{\rho}(v)$) and *outdegree* ($\bar{\rho}(v)$), being the number of edges terminating and originating at v, respectively. Whilst it is generally agreed that the degree of a graph is the mean of the degrees of the graph's vertices, there is a lack of consistency over the indegree and outdegree of a graph. Some literature (such as Sengupta, 1998) refers to the *maximal in/outdegree*, being the largest in/outdegree of the vertex set. Others (such as Durr et al., 2003, p 1) take the view that a graph has outdegree k iff each vertex has outdegree k, and likewise for indegree. This latter approach would lead to a view that the in/outdegree of a graph should be defined as the average in/outdegree of the vertices. However, these values are always equal (as it is the same set being considered, only in reverse). Therefore three metrics are considered here and defined as follows: the *in/outdegree* is taken to be the average of the in (or out) degrees of the vertices. *Indegree* is taken to be the maximal indegree of the vertices and, likewise, *outdegree* is taken to be the maximal outdegree of the vertices.

A graph for which every vertex has the same degree is called a *regular* graph. The graph is *k-regular* when every vertex is of degree k. In a graph where the vertices

have an average of degree z, there are Nz edges in the graph ($\frac{1}{2}Nz$ if the graph is undirected) and z is called the *coordination number* of the graph (i.e. is the same as the degree).

The *vertex-connectivity* of a graph ($\kappa(G)$) is the size of the smallest vertex cutset of G and the *edge-connectivity* of a graph ($\lambda(G)$) is the size of the smallest edge cutset of G. These may be thought of as measures of the fragility of the graph.

Two other important metrics were introduced through SWT, but are described here for completeness. These are *characteristic path length* and *clustering coefficient*.

The *characteristic path length* of a graph ($L(G)$) is the median of the means of the shortest path lengths connecting each vertex to all other vertices. It is the "*average degree of separation*" (Comparing Watts and Strogatz, 1998 with Newman, 2000)

The *clustering coefficient* of a vertex v, γ_v, is the average fraction of pairs of neighbours of a vertex that are also neighbours of each other. It is calculated as

$$\gamma_v = \frac{|E(\Gamma_v)|}{\binom{k_v}{2}} = \frac{\text{total edges in } \Gamma(v)}{\text{total possible edges in } \Gamma(v)}$$

$\gamma_v = 1$ implies that the corresponding graph consists of $\frac{n}{k+1}$ disconnected, but individually complete, sub-graphs (cliques) and $\gamma_v = 0$ implies that no neighbour of any vertex v is adjacent with any other neighbour of v. Equivalently, γ_v is the probability that two vertices in Γ_v will be connected. The clustering coefficient of a graph, γ, is the mean of all γ_v. $\gamma=1$ for a fully connected graph.

A *community* is a group of vertices within which connections are dense. Connections between communities are sparser. The problem of dividing a graph into communities is often referred to as *graph partitioning*.

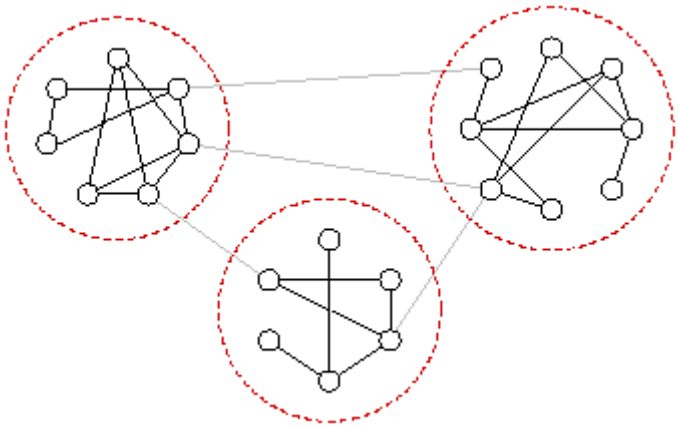


Figure 2.6: Communities within a graph

The *modularity, Q,* of a graph is defined as the difference between the fraction of edges that fall within communities and the expected number of edges for a truly random graph. As such, nonzero values for Q represent deviations from randomness and in practice a value of about 0.3 is a good indicator of significant community structure in a network. (Clauset et al., 2004, p 2; Kernighan & Lin, 1970 quoted in Newman 2004b, p 4).

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(c_v, c_w)$$

where

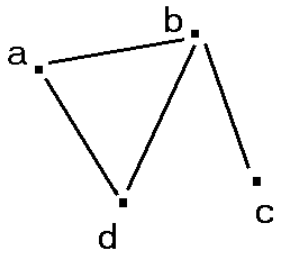
i.e. $A_{vw} = \begin{cases} 1 \text{ if vertices } v \text{ and } w \text{ are connected} \\ 0 \qquad\qquad \text{otherwise} \end{cases}$

m = number of edges in the graph

$\delta(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

k_v = the degree of vertex v

where the graph is divided into communities such that vertex v belongs to community c_v.



$$k = \frac{2+3+1+2}{4} = 2$$

$$\kappa = 1(b)$$

$$\lambda = 1(bc)$$

$$L = median\left\{\frac{4}{3}, \frac{3}{3}, \frac{5}{3}, \frac{4}{3}\right\} = \frac{4}{3}$$

$$\gamma = \frac{1+\frac{1}{2}+0+1}{4} = \frac{5}{8}$$

$$Q = \frac{1}{8}\sum\left(\frac{1}{4} - \frac{1}{4} + \frac{1}{2} + \frac{5}{8} + \frac{1}{4} - \frac{1}{4}\right) = \frac{9}{64} = 0.14$$

$$Q_c = \frac{1}{8}\sum\left(\frac{1}{4} + \frac{0}{4} + \frac{1}{2} + \frac{0}{8} + \frac{1}{4} + \frac{0}{4}\right) = \frac{8}{64} = 0.125$$

Figure 2.7: Metrics for a simple, undirected graph. Q_c is the modularity calculated assuming c is a separate community.

2.4 Small World Theory

2.4.1 Introduction

Small World Theory (SWT) is essentially a part of Graph Theory, although it was not originally perceived as such. SWT was originally observed within social science (see Milgram's work, below) and the networks that were examined were therefore social ones. The two most-cited (and, by implication, most influential) papers on SWT are by Milgram (1967) and Watts & Strogatz (1998) and are discussed below.

2.4.2 Literature Review

2.4.2.1 Milgram, 1967

This paper is the most cited method of investigating whether the small-world phenomenon actually exists and is believed to have coined the term to characterise social networks. It is the only real experiment ever conducted (although it has been repeated, with varying degrees of success). It originates from social science and as such was not originally described mathematically.

Milgram introduces the term "small world" as "*Almost all of us have had the experience of encountering someone far from home, who, to our surprise, turns*

out to share a mutual acquaintance with us. This kind of experience occurs with sufficient frequency so that our language even provides a cliché to be uttered at the appropriate moment of recognising mutual acquaintances. We say, "My it's a small world."" (Milgram, 1967 p.61)

Milgram constructed and undertook an experiment to investigate social connectivity. This involved passing letters from initial recipients to one of two named yet personally unknown targets, via strong social contacts (known on "first name terms"). The heuristic (see 4.4.11.1) is thus "a person known to you who is more likely to know the final target than you are". These *chains* (and specifically, their lengths in terms of numbers of intermediaries) are the focus of the paper.

Milgram noted the "small world effect" whereby two seemingly unconnected persons at disparate locations have a mutual acquaintance, thus shortening the chain between them. Milgram was interested in whether any two persons, anywhere in the world, could be so linked, or whether there existed social cleavages that could not be bridged.

Milgram abstracted this experiment to represent people and networks by points and lines, restating the small world problem as "*Given any two of these points chosen at random from this universe of 200 million points, through how many intermediate points would we pass before the chosen points could be connected by the shortest possible path?*" (Milgram, 1967, p.63)

Of the 160 chains started in Nebraska, only 44 completed. Of an unspecified number started in Boston, 20 completed. Most of the chains had large initial steps (in terms of physical distance) and gradually decreased as they reached their targets. Of 145 participants in the study, 114 passed the message to a person of the same gender (FF: 56, MM: 58) and 31 cross-gender (FM: 18, MF: 13). Messages did not reach the final target through a broad range of acquaintances, but 48% came through only three. This led Milgram to surmise that certain channels are better transmitters than others and to note that physical distance and social distance are not the same thing.

Milgram, being a social scientist, never defined his model in mathematical terms. Applying these shows the model as being a connected graph where some vertices ("superhubs") are more connected than others.

i.e. for a subset of vertices U, $\exists v \in U \subseteq V(G)\, s.t. k_v >> k\ where (|U| << |V(G)|)$ $and\ \forall w \in \{V(G) - U\}, k_w << k_v$

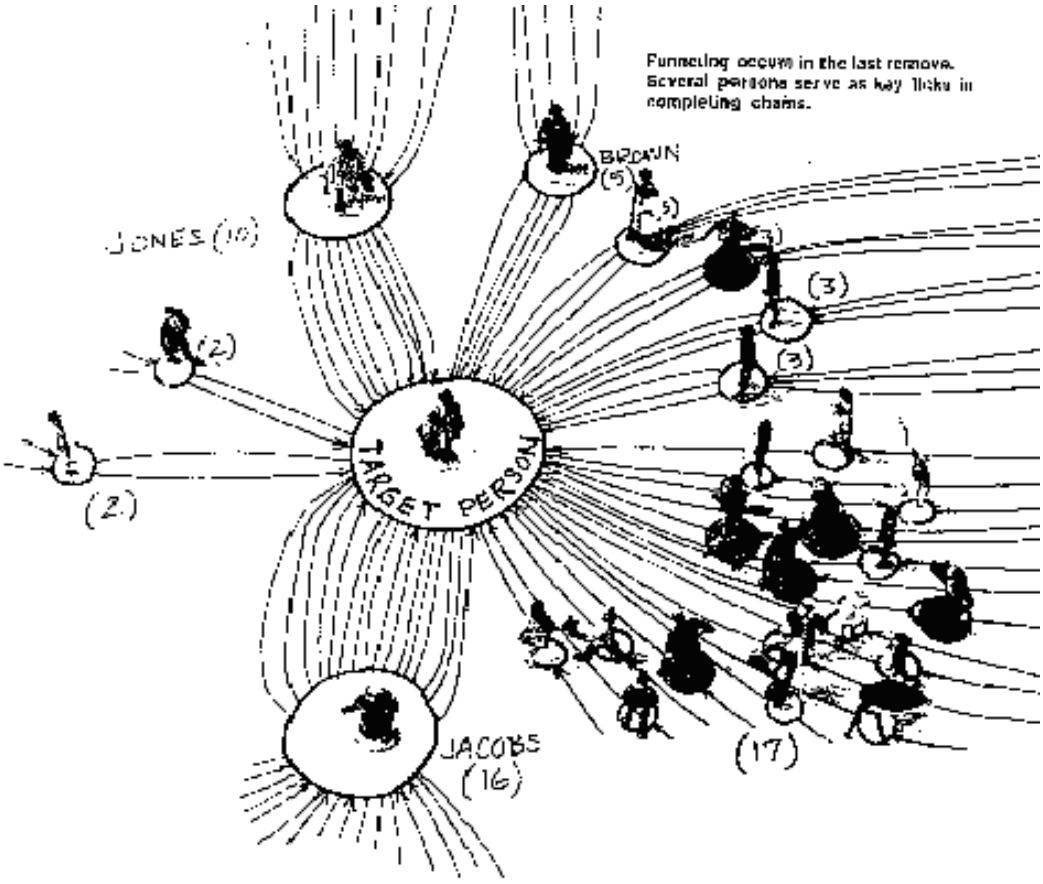


Figure 2.8: An illustration of a superhub and the funnelling effect. From Milgram, 1967, p. 67.

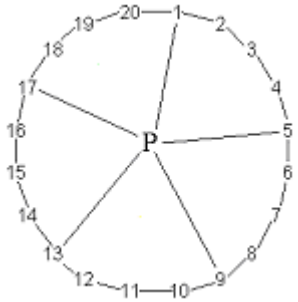


Figure 2.9: A simple superhub as a graph. The "P" in the centre is a hospital porter. From Ganney, 2003

Although Milgram noted that some chains did not complete, the paper did not investigate the reasons for this. Despite an interest in social cleavage and Milgram's own theorem that two populations were disparate if no chain could be found between any two persons (one from each population), it is therefore peculiar to note that his model does assume that the paths exist and that they were not traversed for some other, unknown, reason. If he had included them and not assumed the paths exist he would have produced a disconnected graph.

There has been some debate over the validity of Milgram's original experiment (Kleinfeld, 2000), its poor control and sources of error (Newman, 1999, p 1) and in attempts to repeat it. In particular, most of the letters in the original experiment simply never arrived. A repudiation of such doubts appeared in the Sunday Telegraph (28/11/04) from Prof. Thomas Blass of the University of Baltimore[7] who pointed to a recent Internet replication involving 24,000 people from 166 countries (Dodds, Muhamad & Watts 2003). However an e-contact is much less robust than a social or physical one (although some doubt must exist over how robust Milgram's heuristic was) and certainly has no relevance to a study of disease propagation. Certainly the "six degrees of separation" were postulated by Milgram to cover the USA but have been extrapolated to cover the entire world, without any research justification (Kleinfeld, 2000).

It has been noted that the two most surprising discoveries in Milgram's work are that short chains do exist and that people should be able to find them knowing so little about the target individual, leading to speculation that "cues" must exist within the social framework (Kleinberg 1999). Certainly it would have helped if Milgram specified the occupations of the targets (the paper does not state whether this was done and omissions such as this are not unusual) but this is one example of such a "cue".

[7] Milgram's Biographer – although his web site www.stanleymilgram.com, appears to be more interested in the obedience experiments and only mentions the small world one in passing

Regardless of the failure to "prove" the small world phenomenon in a sociological context, the general result that a short chain of acquaintances can connect two randomly selected people has been subsequently verified and is widely accepted (Korte and Milgram, 1970 - quoted in Newman, 1999, p 1). The concepts, however dubiously discovered, do explain a feature that is real: the ability of an infection to reach a point far distant to its origin in a far shorter time than might be expected. A further point, made by Watts (1999, p 19) that the chain length may have been overstated due to a non-optimal contact being selected makes this result even more remarkable.

2.4.2.2 Watts and Strogatz, 1998

This paper introduces many of the ideas now associated with SWT, in particular *characteristic path length, L.* Watts and Strogatz denote the clustering coefficient as *C* in their paper - this is replaced by γ here (following Watts, 1999) for consistency.

This investigation begins with a substrate of a 1-lattice with k connections per vertex (i.e. a ring in which every point is connected to one or more immediate neighbours on its left and the same number on its right). Individual edges are then re-wired at random (probability φ, sometimes referred to as the "short-cut probability" (Newman & Watts 1999b, p 3)), allowing some local edges to become long-distance edges (or *shortcuts*), but keeps the same number of edges overall. The only restrictions are that the graph should remain simple (i.e. edges are unique and no loops exist).

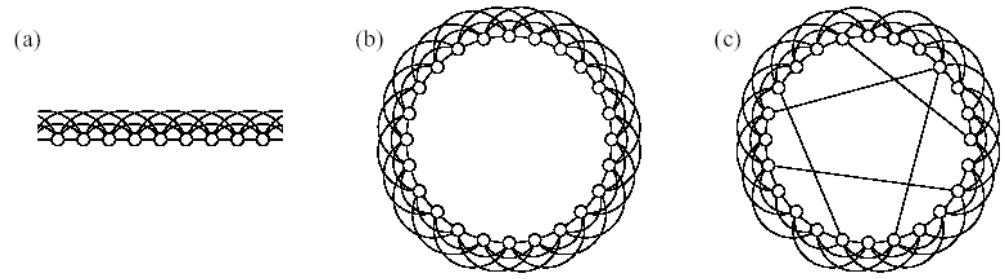


Figure 2.10: (a) A one-dimensional lattice with each site connected to its z nearest neighbours, where in this case z=6. (b) The same lattice with periodic boundary conditions, so that the system becomes a ring. (c) The Watts-Strogatz model is created by rewiring a small fraction of the links (in this case five of them) to new sites chosen at random. From Newman, 1999, p 3.

Watts and Strogatz discovered that, as a completely ordered graph starts to become randomised, there is little change in the characteristic path length. However, as the graph becomes more randomised a threshold is quickly reached, at which point the average path length plummets and the graph becomes a small world. For example, starting from 1,000 nodes in a ring world, each of which had ten adjacent edges, the characteristic path length is approximately 38 (a big world). Only one percent of random links is required to flip this graph from big world to small world (L =3).

This immediate drop in L caused by the introduction of a few long-range edges is characteristic of the small world effect in which the lengths of paths are small compared with the number of vertices in the graph. "Any network in which the lengths of such chains [of acquaintances] are small compared with the number of people in the network is said to display the small-world effect" (Newman, 1999, p 1).

This paper then explores models that can be tuned between the extremes of completely regular (p=0) and completely random (p=1), where p is the rewiring probability.

The models examined have $n >> k >> \ln(n) >> 1$ ($k >> \ln(n)$ guarantees that a random graph will be connected (Bollobás, 2001, p 447)). The authors discovered that

$$L \sim \frac{n}{2k} >> 1 \; and \; \gamma \sim \frac{3}{4} \; as \; p \longrightarrow 0$$

$$L \approx L_{random} \sim \frac{\ln(n)}{\ln(k)} \; and \; \gamma \approx \gamma_{random} \sim \frac{k}{n} << 1 \; as \; p \longrightarrow 1$$

Watts and Strogatz' paper describes the result that as the regular lattice is a highly clustered, large world where L grows linearly with n and the random world is a poorly clustered, small world where L grows linearly with n, one may suspect that large γ is always associated with large L and small γ with small L. However, there is a broad interval of p over which $L(p)$ is almost as small as L_{random} yet
$\gamma(p) >> \gamma_{random}$.

There is an immediate drop in *L(p)* caused by the introduction of a few long-range edges ("short cuts") and at local level (reflected by $\gamma\ (p)$) the transition to small world is almost undetectable.

The "small world effect" is thus quantified as one in which the lengths of paths are small compared with the number of vertices within the graph. Large-world graphs have the average distance between two nodes increasing linearly with system size; small-world graphs increase logarithmically.

The authors present some data on three real-world examples, not hand-picked for their results, but chosen because complete wiring diagrams were available.

Network	N	L_{actual}	L_{random}	γ_{actual}	γ_{random}
Movie actors	225226	3.65	2.99	0.79	0.00027
Neural network	282	2.65	12.4	0.28	0.05
Power grid	4941	18.7	2.25	0.08	0.0005

Table 2.1: The number of nodes *N*, characteristic path length *L*, and clustering coefficient γ, for three real-world networks. The last column is the value which γ would take in a random graph with the same size and coordination number. (*N* from Newman 2000)

Table 2.1 shows γ to be very much larger than γ_{random}, thus the networks are clustered. Also, L is similar to L_{random} thus demonstrating the small-world phenomenon: $L \geq L_{random}$ *but* $\gamma >> \gamma_{random}$

The "rewired" models developed by Watts and Strogatz mirror well the real-world networks that the paper investigated, having both a small-world effect and are clustered. However, hubs are not considered to be important (as Milgram did), unless one is formed by the application of the rewiring algorithm. These models therefore mimic only some aspects of the structure of networks of social interactions (Newman & Watts 1999b, p 1).

Watts and Strogatz found that their models displayed many of the characteristics of true random graphs even for $\varphi<<1$, and it seems to be in this regime that the model's properties are most like those of real-world social networks (Newman & Watts 1999b, p 3). Watts and Strogatz argue that their random re-wiring model

captures two crucial parameters of social networks: there is a simple underlying structure that explains the presence of most edges, but a few edges are produced by a random process that does not respect this structure. This reflects the assertions by Gravovetter (1973 and 1983) that weak ties are more important than strong ties as they are the ones that tend to link social groups (within which ties are strong, i.e. if A and B are linked and B and C are linked, then A and C are very probably also linked). Thus it may be that Milgram's heuristic, weak as it was, was wholly appropriate to the task.

2.4.2.3 Other Work

2.4.2.3.1 Random Graphs

The most difficult part of describing a social network by using graphs is in the creation of the edges. It is unclear how well a person needs to be aware of another for them to be deemed "connected". A simple "have met" algorithm is as poor as the "handshakes" one of folklore. Even Milgram's original criterion of "someone you know on a first-name basis" (Milgram, 1967, p 64) assumes a social stratum of niceties that would not be breached. This may explain why there has been a certain reluctance to repeat the experiment, along with more interest being expressed in examining the Internet Movie Database (the "Kevin Bacon Game[8]" (Smith, 1996)) and Internet topologies where a connection is more easily and robustly defined.

Due to these difficulties in acquiring models of the real world, attention has migrated to the construction and analysis of "realistic" models, constructed using an element of randomness. There is also the issue (as noted by Witten & Poulter (2007, p 197)) that a fully connected graph is actually equivalent to the traditional stochastic model, therefore randomness is to be desired in constructing the

[8] This gives rise to Bacon numbers (see footnote 2) and likewise the Erdös number (linking those who have authored papers with Erdös, then those who have authored papers with those who have authored papers with the Hungarian mathematician Paul Erdös and so on) and then Bacon- Erdös numbers branching out from the few individuals who appear in both lists, of which the MIT mathematician Daniel Kleitman has possibly the lowest (3), having authored with Erdös and appeared in the film "Good Will Hunting" with Minnie Driver, who has a Bacon Number of 1.

network. As the network model is more difficult to analyse, randomness might be better viewed as being essential rather than desirable.

Bollobás (2001) describes random graphs where a simple model of a social network is constructed using a pre-determined number of vertices and a pre-determined number of edges are randomly assigned. Such graphs have been studied extensively by the mathematics community, especially by Erdös and Rényi (1959). Newman (2000) shows that it is easy to see that a random graph shows the small-world effect (see following section), wherein there is a logarithmic increase in the number of degrees of separation (the maximum path length in order to reach all vertices in the graph) with the size of the network (graph).

Random graphs are simple to construct, yet do not represent the real world and real social networks. It is intuitively obvious (yet often ignored) that a pair of "friends" of a person are very likely to also be friends, leading to the clustering described by Watts & Strogatz (1998). A random graph does not show clustering (as Table 2.1, above, illustrates).

If p_k denotes the fraction of nodes with degree k, then random graphs predict a bell-shaped Poisson distribution for p_k. However, for real networks, p_k is highly skewed and decays much more slowly than a Poisson. (Strogatz, 2001, p 274)

2.4.2.3.2 Variants of Watts-Strogatz

Most subsequent work on small world models has been performed on a variant of the Watts-Strogatz model suggested by Newman and Watts (1999a). This preserves the underlying original structure by adding random links and not removing any. This prevents sections of the lattice from becoming disconnected (with the attendant infinite path length problems that this introduces that render the graph difficult to analyse). Although this is not a problem for numerical simulations, it is for mathematical analysis, as is the maintenance of a simple graph: therefore Newman and Watts (1999b) allow loops and non-unique edges in order to have uniform distributions of rewiring probability. Newman and Watts

contend that small-world graphs have a high effective dimension even for quite moderate values of φ, and thus are in some sense close to being random graphs.

In this revised model, *Renormalisation Groups* are used to group together sets of nodes. However, there appears to be no logical grouping: in one case, nodes are simply paired up: in another, triplets are used. It would appear that different results might be found by simply using a different starting node to form the groups.

In random graph work the problem of infinite path lengths (and therefore infinite average path lengths) has been overcome by averaging the reciprocal of the vertex-vertex distance, but this approach does not seem to have been tried for the Watts-Strogatz model (Newman 2000 p 4).

Newman (1999) demonstrated that the Watts-Strogatz model is based on random graphs and does not include superhubs. These have been investigated by Kasturirangan (1999) and Dorogovtsev & Mendes (1999) who start from the same ring lattice as Watts and Strogatz but, instead of rewiring or adding in new edges, add in new vertices which are "superconnected". These new vertices (or, to be specific, the paths through them) provide the shortcuts required to demonstrate the small-world effect.

Another method of generating superhubs was proposed by Albert et al. (1999, quoted in Newman, 2000 p 825) during their studies of the structure of the World Wide Web. The starting point was a number of vertices and a power law spread of degrees. Selecting a vertex at random, an edge was created between it and another randomly selected vertex if the result would bring the overall distribution of degrees closer to the required power law. Despite matching the measured properties of the World Wide Web quite closely, this model does not show clustering, a property that Adamic (1999, quoted in Newman, 2000 p 825) demonstrated exists in the Web. Although this makes it an unrealistic model of the structure it was seeking to mimic, it remains an alternative method of creating superhub-based graphs which may prove to be of use when the target structure is known to include superhubs.

Kasturirangan (1999 p 12) has noted that the connection of vertices arbitrarily far apart with uniform probability (in order to create shortcuts) is a poor representation of at least some real-world situations. Kleinberg (1999) notes that, in the real world, people are surprisingly good at finding short paths between individuals (as Milgram's experiment shows) given only local knowledge about the structure of the network. Granovetter (2003) noted that there is an important question (not so far resolved) of how much people know about their own social networks and why this matters, whilst speculating that shorter paths may bring rewards (Granovetter, 1995, quoted in Granovetter 2003). Granovetter has demonstrated that no algorithm exists which is capable of finding such paths on Watts-Strogatz-type graphs, given only local information. Kleinberg (1999) therefore proposed another variant on the Watts-Strogatz model, in which the distance (in this case measured across the underlying lattice) between the vertices is tuned by utilising an inverse power law of distance as the probability of an edge forming. On such graphs, there exists a simple algorithm for finding a short path between two given vertices, making use only of local information. It should be noted, however, that this model therefore mimics well a graph where the "normal" social distance is similar to the geographic distance (i.e. is well modelled by an underlying lattice) but it is not clear how well this would model a graph where "normal" social distance is not similar to the geographic distance (i.e. a small world one).

Kasturirangan (1999) also investigated the changes required within a social network to make it change from being a "large world" into a "small world". The paper investigates multiple-scale graphs and asserts that multiple-scale is a stable property of a graph and that the distribution of length scales within a graph was a more useful study than the effect of introducing disorder.

2.5 Disease Modelling

The modelling of a disease and of the propagation of it has clear benefits: it contains no risk to life and such models are repeatably generatable allowing multiple scenarios to be investigated from exactly the same (rather than similar) starting conditions.

The classical mathematical approach to disease modelling (in particular, disease spreading) either ignores the structure of the social network altogether or treats populations as spatially distributed in a continuous medium. Typically, the first case uses an SIR model (see 2.5.1 below and 3.3.1) and subdivides the population into three sub-populations whose number, size and interaction determine the transmission of disease. This approach has been utilised effectively in the modelling of infection in well-mixed populations (May & Nowak, 1994; Murray, 1993 – both quoted in Watts, 1999 p 167) with an emphasis on the detailed dynamics of disease transmission rather than the relationships between subpopulations.

The second classical approach introduces a spatial dependency to the subpopulations involved and is typified by reaction-diffusion equations (Murray, 1993 – quoted in Watts, 1999 p 167). Here questions of the stability of equilibria and the analytic tractability of solutions tend to dominate.

A third approach began to appear in the late 1980s that took greater account of the fact that populations are often inherently discrete and exhibit high levels of structure (see Sattenspiel & Simon, 1988 for one such approach).

None of these approaches, however, treats the spread of an infection within a population as a function of the structure of that population, which would seem to be a mistake.

2.5.1 Models

Mathematical modelling of virus spreading and epidemics has generally utilised one of two models, both of which can be utilised on graphs:

1. a susceptible-infected-susceptible model (SIS), in which vertices are either "healthy" or "infected". At each time step a healthy vertex becomes infected with probability v if it is connected to at least one infected vertex. An infected vertex is cured with probability δ, defining an effective spreading rate of $\lambda \equiv \frac{v}{\delta}$. It is assumed that an infected vertex is capable of

passing the infection to a susceptible (healthy) one. The behaviour of the SIS model is well understood for vertices in a regular lattice or random network (Anderson & May, 1991; Nowak & May, 2000, both quoted in Dezsö & Barabási, 2002 p 1).

2. a three-compartment model, the compartments being those who are susceptible to the disease, those who already infected and those who have recovered/died (SIR: Susceptible, Infected, Removed). The probabilities of movement between the three compartments are fixed, and are the same for each individual within a compartment, as is the probability of contact with an individual from one of the other compartments. The case where movement is possible from Removed to Susceptible (where short-term immunity is conveyed, for example) is often termed SIRS. Another variant adds an "Exposed" stage, where a vertex is infected but is not (yet) capable of passing the infection to a susceptible (SEIR model).

Ng, Turinici & Danchin (2003) demonstrate that the standard SIR model does not describe a SARS outbreak well and instead utilise a double SIR model (SEIRP[9]). This is unusual, but does show the limitations of the standard SIR model.

2.6 Application

2.6.1 Application to Transmissions Between People

Whilst the concepts of mapping and rewiring social networks are interesting, it is the investigation of the communications along them (using the broadest definition of "communication") that makes such studies useful. Most human communication takes place directly between individuals and, specifically, the spread of disease occurs primarily by person-to-person contact (therefore attempts to replicate Milgram's experiment by e-mails, although strong, is not as useful in this context as Milgram's weak heuristic). The structure of a social network has a huge impact on the nature of epidemics (Newman 2000 p 1) – it is intuitively obvious that a network with long paths between individuals will be more resistant to epidemic

[9] The additional parameters are Exposed (infected but not contagious) and Protected (a period of immunity).

than one where the paths are short. Toroczkaia & Guclu (2007) investigated this network via independent agents and collision dynamics: the infection having a probability of being passed on at each collision. Most models, though (and certainly all the ones under consideration here) use links to describe this infection probability rather than random interactions, although this work did show that such an approach can account for the observed qualitative differences between the degree distributions of contact graphs of diseases with short infectivity period (such as air-transmitted diseases) or long infectivity periods (such as HIV).

2.6.2 Application to Infection Propagation

As all infections under consideration pass from person to person by contact (which may also include person-fomite-person spread), the propagation of infection through a network is easily modelled by setting a number of vertices in the graph to be infected, then allowing the infection to move from vertex to vertex along the edges, much as Milgram (1967) modelled the flow of information between subjects. Modifications to this simple idea are introduced via the SIR or SIS models described above and other techniques described later.

2.6.2.1 Small World Graphs

Watts & Strogatz (1998) describe a very simplified model for examining the spread of infectious disease, wherein ring graphs represent the social contacts, with random re-wiring. At time t=0 a single infective individual is introduced. Each infection lasts one unit of time, during which healthy neighbours are infected with probability r. The disease therefore either spreads along the graph, or it dies out (having infected part of the graph). The critical infectiousness, r_{half} (half the population infected), decreases rapidly for small p (the rewiring probability). The time for total infection, $T(p)$, resembles the $L(p)$ curve, illustrating that infectious diseases spread much more quickly in a small world. This model illuminates the dynamics of the infection as an explicit function of structure. Other models indicate that network structure influences the speed and extent of disease transmission.

Newman and Watts (1999b) took this idea further and introduced a fraction, q, of the population that is susceptible to the disease, indicated by a two-state variable

associated with each vertex. This work investigated the point at which an infection became an epidemic and discovered this to be the percolation point for site percolation (see Figure 2.11) with probability q on the small-world graph, the position of this point being strongly influenced by the small-world nature of the graph. Although the paper does not make this clear, as the disease can only spread between susceptible individuals, this paper therefore investigates the subgraph of these vertices and, as q increases, the probability of the subgraph being connected increases, and thus the probability of epidemic. When q is small, the subgraph is less likely to be connected and thus an epidemic is less likely.

Moslonka-Lefebvre, Pautasso and Jeger (2009) showed that the incidence of epidemic is not dependant upon the starting point within a network, but is negatively related to the correlation coefficient between the in- and out-degree for the structures, unless the networks are sparsely connected. If this is so, then clustering plays a significant role. For small-size scale-free directed networks to have a lower epidemic threshold than other network structures, there needs to be a positive correlation between the number of links to and from nodes. When this correlation is negative (one-way scale-free networks), the epidemic threshold for small-size networks can be higher than in non-scale-free networks. The paper shows that clustering does not necessarily have an influence on the epidemic threshold if connectance is kept constant. Additionally, Eames (2008) demonstrated the importance of random contacts in a clustered system: without them, the spread of infection is greatly reduced. With random contacts, parts of the network that are otherwise inaccessible may be reached. In a result comparative to Watts-Strogatz, Eames also showed that very few random contacts are required to increase the spread of infection.

2.6.2.2 Percolation

Whilst most work on disease propagation focuses on what is known as “site percolation”, there have been studies into bond percolation also, as reported by Newman (2003).

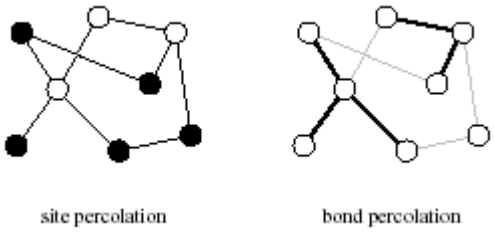


Figure 2.11: Site and bond percolation on a network. In site percolation, vertices ("sites" in the physics parlance) are either occupied (solid circles) or unoccupied (open circles) and studies focus on the shape and size of the contiguous clusters of occupied sites, of which there are three in this small example. In bond percolation, it is the edges ("bonds" in physics) that are occupied or not (black or gray lines) and the vertices that are connected together by occupied edges that form the clusters of interest. (From Newman, 2003 p 38)

In the case of infection propagation, site percolation investigates infected individuals, whereas bond percolation investigates infection transmission routes. Contact rate and infective time translate directly into a bond occupation probability (Witten & Poulter 2007 p 198), showing that the two models have great similarity.

2.6.2.3 Competing Pathogens

Competing pathogens, such as strains of influenza that show cross-immunity, or pathogens which kill the host, may also be modelled (Newman, 2005). The particular case studied here is the one where the first pathogen has passed through a population, causing an epidemic that leaves some fraction immune or dead, followed by a second pathogen at a later time. The author notes that the model could similarly represent two outbreaks of the same disease. The model uses a generalised SIR model, together with a probability of transmission (termed *transmissibility, T*) for the edges. Newman found that co-existence of the pathogens is only possible for intermediate values of *T*. Two phase transitions therefore exist: the standard epidemic transition below which the first pathogen is unable to spread, and the point at which the first disease removes so large a fraction from the population that not enough remain to support the spread of the second. There are other experiments that could be performed on such a model by varying the cross-effects of the pathogen, although this does not appear to have been done.

2.6.2.4 Semi-Directed Graphs

Ancel Meyers et al. (2006) described the use of semi-directed graphs, in which some edges are directed and others are undirected, in order to model the fact that some diseases transmit better in one direction than another and that infected individuals will seek out certain people (especially healthcare workers (HCWs)) whereas the converse is not true and is also not true of the same individual when uninfected. This paper shows that in semi-directed networks the probability of an epidemic and the expected fraction of the population infected during such an epidemic may be different, in contrast to the many conventional models that assume the equality of these two epidemiological values (although this is true for undirected graphs). The model was applied to assess the role of HCWs in disease transmission and containment. One useful measure introduced in this work is the risk to individuals of infection, expressed as a function of their degree. However, this requires the distribution (or generating) function of degrees to be known.

This paper also presents a case study in hospital-based transmission of respiratory disease (possibly the only one so published), which is built upon previous work to simulate urban contact networks[10]. In this, the undirected-degree distribution is roughly exponential and the in- and out-degree distributions solely determined by the flow of infected people into health care facilities. Assuming each non-HCW individual to have three directed edges pointing to randomly selected HCWs (i.e. an out-degree of three and an in-degree of zero) gives HCWs out-degrees of zero and in-degrees of 409-530. The paper reports that, for diseases close to the epidemic threshold, the probability of epidemic in the semi-directed graph is more than double that of the simpler undirected graph. The effect of intervention (in this case travel restrictions and isolation) was then examined, by removing appropriate edges from the graph. Vaccination prior to an outbreak was modeled by removing a vertex and all of its edges from a graph. Other interventions, such as the use of facemasks, surgical gowns and hand washing were modeled by lowering the

[10] Based on demographic data from Vancouver, British Columbia.

probability of transmission. The levels of infection were shown to increase drastically if/once an HCW is infected. Although not noted by the authors, this shows these vertices to be the superhubs of Milgram's work. Having introduced some directionality into the graph, it is not clear why full directionality was not investigated.

2.6.2.5 Weighted Graphs

Newman (2004a) investigated weighted graphs. These have normally been avoided or ignored, as they are perceived as being harder to analyze than unweighted ones. Whilst it is reasonable to study simpler cases (unweighted) before moving onto more complex or complicated ones (weighted), Newman shows that weighted graphs can in many cases be analysed using a simple mapping from a weighted graph to an unweighted multigraph, allowing standard techniques for unweighted graphs to be applied to weighted ones. Weighted graphs are often used in sociological studies, with negative weights indicating animosity. A multigraph is a graph where non-unique edges are allowed, that is, one in which there may be multiple edges between a vertex pair. Thus the translation from weighted edge to multigraph is intuitive as they yield the same adjacency matrix. Newman shows that the method of Girvan and Newman (2002, quoted in Newman 2004b p 4) finds communities within weighted networks. Indeed, the weights often expose community structure that a simple unweighted one (i.e. created by removing the weights from each edge) does not. It is presumed (but not make clear) that the subsequent method of Clauset et al. (2004) also does this.

2.6.2.6 Social Context

Ganney (2003) considered the context in which the exposure to infection takes place. Using three sizes of locations (individual, small and large) the context modified the susceptibility of the individuals. When the wards (originally classified as "small") were re-classified as "large", the clear peaks of the original lifespan graph became less pronounced (see Figure 2.12).

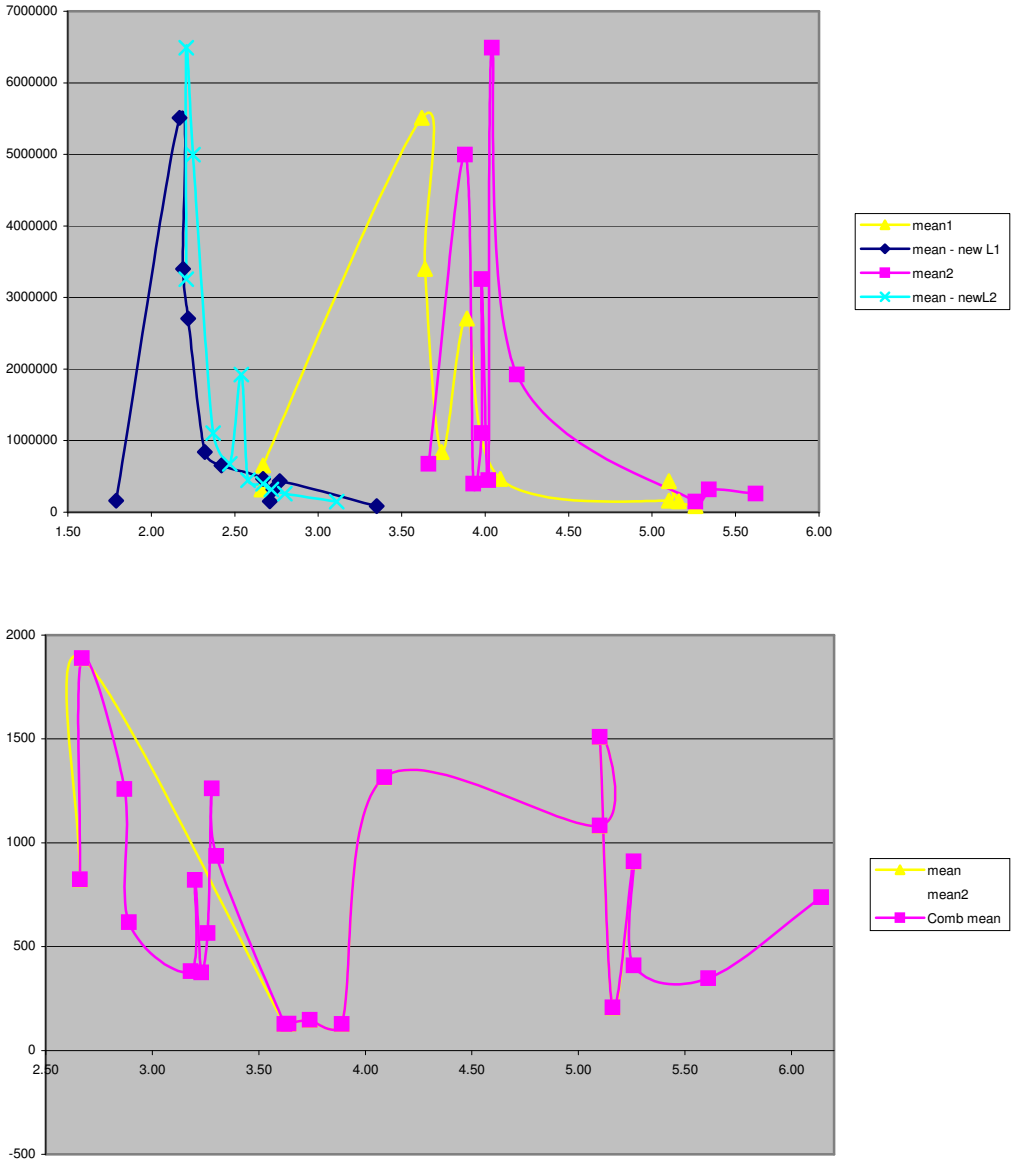


Figure 2.12: The upper graph shows the mean lifespan against mean path length for four relationship models with wards classified as "small" spaces. The lower graph shows similar data where wards have been re-classified as "large".

2.6.3 Infection Control

Once a disease process has been mapped onto a graph and its spread analysed, it is intuitive to experiment with infection control procedures, a topic undertaken by Dezsö & Barabási (2002) and Cohen, Havlin & ben-Avraham (2003). Dezsö & Barabási quote the result of Pastor-Satorras & Vespignani that for free-scale graphs with $\gamma \leq 3$ the epidemic threshold (λ_c) vanishes; that is all diseases, regardless of infectiousness, will spread and prevail. Dezsö & Barabási propose

that this effect is due to vertices with a large number of edges (superhubs, although not described as such in the paper) as, once infected, they pass on the disease to a significant fraction of vertices in the system. Using a simple SIS model, this work found that randomly distributing cures throughout such a graph (to infected vertices only) had no effect, whereas even a weak biasing of the distribution towards infected superhubs re-established the epidemic threshold, thus allowing the disease to die out naturally. This is because $\lambda_c = \frac{< k >}{< k^2 >} = \frac{k_0 - m}{k_0 m}\left(\ln \frac{k_0}{m}\right)^{-1}$ where vertices with degree $k>k_0$ are healthy (Lloyd & May, quoted in Dezsö & Barabási, 2002, p 2). Therefore, the more superhubs cured (the lower k_0 is), the larger the value of λ_c. The practical problem, of course, is to identify these superhubs (equivalent to finding the vertex cutset). Whilst the approach of identifying and treating/inoculating the superhubs is intuitive, and has been implemented (e.g. Esu-Williams, 1995), there appears to have been no interaction between modelling and implementation.

Cohen, Havlin & ben-Avraham (2003) studied instead the SIR model and a novel immunisation program. In this, a random number of vertices were selected. For each vertex, one edge was randomly selected and the vertex at the terminus of this edge was immunised. This yields better-connected vertices (Feld, 1991; Newman, 2003 – both quoted in Cohen, Havlin & ben-Avraham, 2003), although some vertices may be selected more than once. This research discovered that this immunisation technique was more efficient than a purely random one, dramatically reducing the immunisation threshold for all studied cases. A further advantage of this technique is that it does not require global knowledge of the graph, only local knowledge of the selected vertices. Despite the effectivenes of this approach, the social issues surrounding an implementation make it unlikely to succeed.

Infection control methods have generally focussed upon reducing the spreading rate of the infection, hoping to reduce it to a point at which it will die out naturally. However, as scale-free graphs have an epidemic threshold of zero (Pastor-Satorras & Vespignani 2001 p 3200), this method will never eradicate a

disease within such a model. Techniques such as those described above therefore attempt to re-introduce the epidemic threshold by targeting the structure of the graph.

An alternative method of infection control might be to discover the "fire breaks" (or "fault lines") within a network – i.e. identify the edge cutset rather than the vertex cutset. Although this has not yet been investigated, the detection of communities has received some attention. The work of Zachary, for example, (1977 - quoted in Newman, 2004a) describes a social network that fragmented into two communities that can be seen in the original analysis. Many methods have been proposed for discovering such communities, although all seem to require foreknowledge of the number of communities to discover. Once this is known, methods such as hierarchical clustering (Scott, 2000 quoted in Newman, 2004a) are very good at detecting the boundaries and membership of the communities. The use of modularity in the algorithm described by Clauset et al. (2004) would appear to be a reliable method for detecting and determining the number of communities where this is unknown a priori. This method implements a greedy optimization that seeks for the maximal increase (or, if there is none, the minimal decrease) in modularity, Q. The resulting dendogram is cut at the point where Q is maximal in order to reveal the communities.

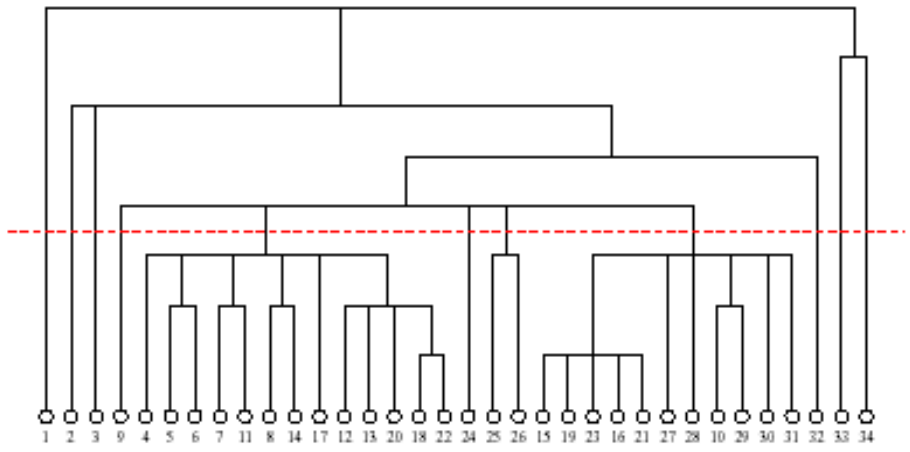


Figure 2.13: A dendogram of the community described by Zachary (1977). A cross-section of the tree at any level will give the communities at that level. The cross-section indicated by the dotted line corresponds to the community division discovered by the hierarchical clustering method. Taken from Newman, 2004a p 5.

2.7 Discussion

The problem with the compartment models (SIS, SIR) is that all susceptible people do not face the same risk of contracting a disease. Along with natural susceptibility, the level of social interaction needs to be taken into account. Ancel Meyers (Griffith 2003) has modelled the pattern of interactions in a city[11] utilising municipal contact networks.

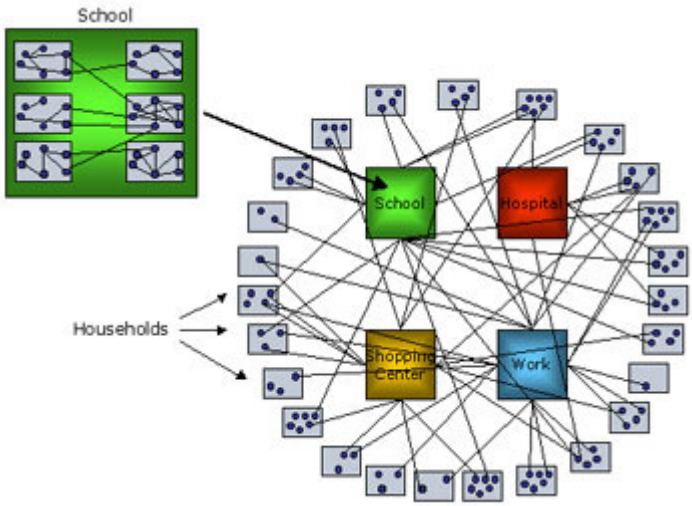


Figure 2.14: A municipal contact network. Diagram from Griffith (2003).

Ancel Meyers' subsequent work involved modelling a psychiatric institution in Indiana in order to understand the spread of walking pneumonia. Ancel Meyers discovered that "*while the focus is generally on preventing the spread of walking pneumonia from patient to patient, caregivers play a much more important role in the large-scale spread of respiratory infections across such a facility*." (Griffith 2003)

Clearly this work is based upon graph theory with the municipal contact network centres and caregivers providing superhubs and the long-range links required of small world theory, thus giving credence to the models so constructed.

[11] In this case Vancouver.

Another problem with the two main models studied (SIS and SIR) is that neither allow for a vertex to be "temporarily removed" from the set of susceptible vertices. Whilst some diseases (e.g. measles) confer a life-long immunity and some (such as the common cold) confer none, others (e.g. Respiratory Syncytial Virus (RSV)) do confer a period of immunity, as do all vaccines (e.g. Chickenpox). Neither model takes account of natural passive immunity, probably because it is fairly rare except between mother and child. There is also the question of the degree of infectiousness and the level of susceptibility of an individual. Some pathogens, such as Chickenpox/Shingles, are highly contagious, so will have high transmission probabilities (87% attack rate in susceptible exposed children. Hawker et al., 2001 p 72). Others, such as SARS, will have low ones (1.2% attack rate in hospital workers overall, with the highest of 2.3% being for nonmedical support staff. Lau et al., 2004 p 1399).

Most work on SWT and on the use of graphs to model disease propagation uses simple, undirected graphs. The assumption, therefore, is that a disease is as likely to pass in one direction between people as the other. This, however, is not true. As mentioned above, HIV passes easier from men to women than vice versa and an uninfected Healthcare Worker is more likely to be sought out by an infected member of the public than an infected Healthcare Worker is to seek out that same uninfected member of the public. (Ancel Meyers et al., 2006 p 403 - quotes HIV data from Italian Study Group on HIV Heterosexual Transmission). Watts himself notes that friendships are not symmetric (Watts, 1999 p 5).

Whilst the mixture of directed and undirected edges in the Ancel Meyers et al. (2006) model adds a useful level of complexity, edges cannot exist in both directions with different transmission probabilities: edges are either fully two-way, or fully one-way. Matters were simplified by assigning the same transmission probability to all undirected edges (T_u) and the same transmission probability to all directed edges (T_d), as well as studying the special case where T_u=T_d.

As mentioned above, it has been identified that the treating of superhubs greatly increases the probability of eradicating a disease yet that it remains difficult to

identify these superhubs. As also described above, parallel work has taken place in attempting to identify structure within a society. There is clearly scope for these two approaches to be married together and a new method of infection control to be investigated. This may also tie in with recent work on edge percolation.

Despite Milgram's interest in social cleavages, little, if any, work has been done in examining disease propagation on networks so cleaved. Likewise, little, if any, work has been done on examining the effects of a graph mutating during the lifetime of the disease, except in the case where the disease removes vertices (by causing either immunity or death).

As noted above, one surprising discovery of Milgram's work showed that paths often exist which are not visible to first analysis. However, if these paths are there, no matter how hidden, then infection can propagate along them. The second surprising result, that these paths can be found using only local knowledge, increases the likelihood of a disease so propagating as overall structure does not need to be known.

The reason most investigation has concentrated on simple undirected graphs is that these graphs lend themselves to mathematical analysis. As this research aims to develop a more "realistic" model, only computer simulations will be possible.

2.8 Summary

In summary, the following shortcomings of current published models are identified:

- SWT has generally been applied to large, even global, populations. Whilst this emphasises the effect of a shortcut upon chain length, the theory should be equally applicable in the modelling of a semi-closed environment such as a hospital. Smaller populations have been examined, but have done so in isolation, ignoring any possible effects due to the wider system.
- SWT utilises random connections. In experiments, the reason for a link's revelation has been in finding someone. Other reasons for links' existence are not explored.

- The connections are generally two-way. Real-world experience suggests that links have different weightings and different transmission properties/probabilities in each direction.
- The strength of the links is not considered: all are viewed as being equal.
- SWT assumes static connections, i.e. once a model has been formed, its properties are investigated but the network does not alter. Real-world experience suggests that links form and reform (and the strengths alter) with time. However, for swiftly spreading infections, this may not be pertinent. Random changes in a graph results in all networks converging onto random graphs (Witten & Poulter, 2007 p 204) so the underlying initial model may not be relevant. Verdasca et al. (2004) did present a dynamic model where the vertices and edges are created afresh at each iteration with good results, but that level of "dynamic" is probably excessive – certainly the computations were "intensive" (Witten & Poulter, 2007 p 204).
- The infections investigated merely propagate: they do not die out and infected individuals do not recover. Whilst this is true (and an appropriate statistic) for SARS, HIV/Aids or Foot & Mouth (e.g. Saramaki & Kaski, 2004; Chen, 2001; Small & Tse, 2005), it is not for Norovirus or MRSA (both of which are prevalent within the NHS).
- The simulation methods run on the connection models are rather basic. It assumes an overall probability of infection, rather than implementing different factors (for example, the probability of the link being in place at a particular time is one factor generally ignored).
- Some infections, such as Influenza, are cyclic according to season. No models so far appear to allow for this "dormant" and "rampant" phases of an infection, although Saramaki & Kaski (2004) have partly modelled this by using different strains of influenza.
- All models consider the effect of only a single infection taking place at once. For example, Influenza A and Influenza B often co-infect and interfere with one another. Immunity from one does not grant immunity from the other. Likewise, one infection may suppress the immune system leading to an enhanced effect of a secondary one.

- Few models consider the impact of a treatment or vaccine being introduced during the outbreak, although post-infection immunity is considered.

3 Infection in a Hospital Environment

3.1 Introduction

It is estimated that "1 in 11 of all in-patients has a hospital acquired infection at any one time" (Daily Telegraph, 25/2/2005) with Teaching Hospitals (such as Hull Royal Infirmary) having a higher rate (11.2%) to non-teaching (8.4%) (ibid). Whilst there is some progress in limiting the propagation through hygiene programmes, a suitable, realistic model would allow investigators to conduct "what-if" experiments in order to determine additional defences and strategies as well as determining the likely effects of uncontrolled outbreaks.

The fight against the spread of infection is of paramount importance. To fight an infection within a population, one has to have an understanding of how it is transmitted between members of the population and how it then spreads within that population. Realistic models are important tools in gaining this understanding and in allowing researchers to investigate different scenarios of spread and different techniques for combating it. Additionally, in epidemic research, there is great value in models that can predict the signs of infections that will become epidemics, and those which will not, from early data. The threat of pandemics of new diseases[1] give rise to demands for mathematical and computer models of infection propagation partly to understand the mechanisms involved, but mostly to inform the countermeasure strategies.

The study of the spread of infection by simulation has traditionally been studied using a compartmental-deterministic model. Whilst giving admirable results in numerical terms, such studies have not modelled the propagation throughout a social network and therefore do not lend themselves well to studying methods of preventing such spread.

The problem with the compartment models when studying a host-to-host infection is that they presume that all susceptible people face the same risk of contracting a

[1] SARS was a recent case – "Swine Flu" is the current.

disease. This assumption holds provided that all susceptible people are given the same level of exposure to infected people. Whilst this may be argued as being reasonable for a common source infection[2], this is not a realistic proposition for host-to-host[3], as an infected person will encounter only a subset of the population under consideration. If an initial infection is evenly distributed, then it is reasonable to infer that the intersection of these subsets evenly covers the whole population. However, infections are rarely (if ever) so distributed. Therefore, along with natural susceptibility, the level of social interaction needs to be taken into account.

There has been much interest in recent years in using graph or network models to study connections between individuals and between groups within a population (e.g. Milgram 1967, Watts & Strogatz 1998, Ancel Meyers et al. 2006). Connections link individuals but can also be a conduit for transmission of properties such as information, letters or infection. The study of how a property propagates within a population can be modelled by creating a graph for the real-world scenario and imposing upon it the transmission properties of the property under consideration.

In order to investigate the properties of infection propagation within a population, previous research has utilised regular graphs or ones enhanced by random alterations or additions. This has had the advantage that the graphs are controllable and limitable whilst providing sufficient complexity to make the examination of the process non-trivial. However, there remains the question as to how realistic these randomly generated structures are and therefore how relevant the results may prove to be.

Firstly four infections of particular interest to this type of research are described, especially Norovirus (3.2.1.4) which is modelled in later chapters. The chapter concludes with a brief examination of some alternative modelling methods.

[2] A common source infection is one that arises from a contaminated source, such as food or water.

[3] A host-to-host infection is one that is transmitted from an infected individual to a susceptible one.

3.2 The Hospital Environment

A hospital environment is a highly complex network of interactions between hospital departments, individual staff, patients and visitors. It can be viewed as several interconnecting compartments, where the connections are formed by architecture, patient movement and staff function. As each compartment has a pre-defined role, the compartments that it connects to are liable to be limited but certainly will be known and unlikely to alter significantly. It can therefore be viewed as a network of communities.

3.2.1 Infection Propagation and Control

Many infectious diseases spread through direct person-to-person contact. Respiratory-borne diseases like influenza, tuberculosis, meningococcal meningitis and SARS, spread through the exchange of respiratory droplets between people in close proximity to each other. Sexually transmitted diseases like HIV, genital herpes and syphilis spread through intimate sexual contact, yet some (such as HIV) are more easily caught by women than by men during heterosexual encounters. (Ancel Meyers et al., 2006)

Four infections are of particular interest in this thesis: Clostridium Difficile, Influenza, MRSA and Norovirus. They are all transmitted by close contact or proximity, are highly contagious and have been shown to thrive in institutional settings, especially hospitals. These are now described[4].

3.2.1.1 Clostridium Difficile (C.Diff)

Clostridium Difficile (CD) causes 25% of cases of antibiotic-associated diarrhoea and a greater proportion of more severe disease. Elderly, hospitalised patients are at the greatest risk. There is a background rate of CD in most hospitals with occasional outbreaks. Typically diarrhoea will start within a few days of commencing antibiotics.

CD is asymptomatically carried by 2-3% of adults and (possibly symptomatically) 50% of neonates (less than 1 year old).

[4] Except where noted, the descriptions are drawn from Hawker et al, 2005.

CD is transmitted from symptomatic infected individuals via contact (including the hands of uninfected healthcare workers) or through the build-up of spores in the environment or on contaminated fomites such as commodes. Spread does not occur from an asymptomatic (absence of diarrhoea) infected individual. Transmission from infected patients to medical and nursing staff has been recorded, but is usually mild and short-lived.

CD-associated disease (CDAD) occurs when the gastrointestinal tract of a susceptible individual is colonised by a pathogenic strain of CD. Factors such as age, antibiotic treatment, cytotoxic agents, intensive care, naso-gastric intubation, concurrent illnesses and alteration in gut motility increase the risk of acquiring CDAD.

Infection control advice is that symptomatic patients should be isolated and gloves and aprons should be worn by staff, along with adherence to handwashing protocols.

3.2.1.2 Influenza

Influenza is a virus that is life-threatening in the elderly and chronically unwell. During epidemics it is a major cause of morbidity. It causes annual wider epidemics of varying sizes and pandemics at other times. It affects all ages with the highest incidence in children, although most hospitalisations and deaths are among the elderly.

Community outbreaks are common between November and March, lasting 6-10 weeks, peaking at around 4 weeks and are responsible for between 3,000 and 30,000 deaths each winter.

The A and B viruses alter gradually resulting in a significant epidemic every few years with rapid spread and a 10-20% attack rate. Influenza A may change abruptly leading to a subtype for which there is little or no population immunity and causing major pandemics. Recent occurrences were in 1918 (20-40 million deaths worldwide), 1957 and 1968. Despite the huge numbers affected during a

pandemic, more deaths occur due to the steady accumulation of normal influenza activity.

Influenza in humans is transmitted via the respiratory secretions of infected individuals, via air-borne droplets or small particle aerosols. Transmission is enhanced by enclosed and overcrowded spaces. Spread in such cases is rapid and attack rates high.

The reservoir for influenza A is primarily aquatic fowl. Influenza B only affects humans.

Influenza incubates in 1-3 days (occasionally up to 5), with the infectious period lasting for 1 day before the onset of symptoms to 3-5 after this point in adults. Children have been observed at 3 days before to 9 days after. The infectious dose is low.

Immunity develops and protects for many years against the same strain. Cross immunity to other strains does occur.

Immunisation programmes exist and are very effective, reducing the risk of hospital admissions and death.

3.2.1.3 Methicilin-Resistant Staphyloccus Aureus (MRSA)

Levels of MRSA in UK hospitals have risen since 1995 and have become a major public health concern. This increase has been attributed to the appearance of strains with epidemic potential, increasingly susceptible patients, failure to maintain good hospital hygiene, as well as greater bed usage, throughput of patients and inter-ward transfers.

S.aureus is a common cause of infection, ranging from mild skin sepsis to life-threatening septicaemia. MRSA as a proportion of S.aureus has increased in England and Wales from 2% in 1990 to 42% in 2000, but appears to have now stabilised.

In 2003, MRSA rates per 1000 bed days ranged from 0.04 to 0.33, although it is not clear whether these were community- or hospital-acquired.

The reservoir for MRSA is colonised or infected humans (and, rarely, animals). Colonisation sites are mainly skin, whilst discharges from wounds and other lesions are the main sources in infected individuals. Infection is via contact and invasion usually via broken skin.

The incubation period is 4-10 days and an infected individual is contagious until the infection/colonisation is eradicated. Risk factors include prolonged hospital stay, intensive care and surgical procedures.

Control measures include a reduction in patient movement, isolation, clearance of MRSA using topical or systemic antibiotics and adherence to hospital hygiene. Contact with infants and other susceptible groups should be avoided and school-age children should not attend whilst infectious.

3.2.1.4 Norovirus

Norwalk-like viruses (NLV) are members of the calicivirus family, first discovered in 1972 following an outbreak of gastroenteritis in Norwalk, Ohio (Kapikian et al., 1972 p 1075), and more recently renamed Norovirus (Fauquet et al., 2005 – approved as official genus for NLVs, CDC, 2006). They may also be referred to as small round structured viruses (SRSV). Noroviruses are the most common cause of gastroenteritis in Europe. Nearly 50% of all gastroenteritis outbreaks reported for England and Wales were due to Noroviruses, a figure that is similar to data reported for other European countries including Finland, Sweden, the Netherlands and Germany (Lopman, Brown & Koopmans, 2002). Spread, especially in institutions, may be rapid, although the illness may be mild.

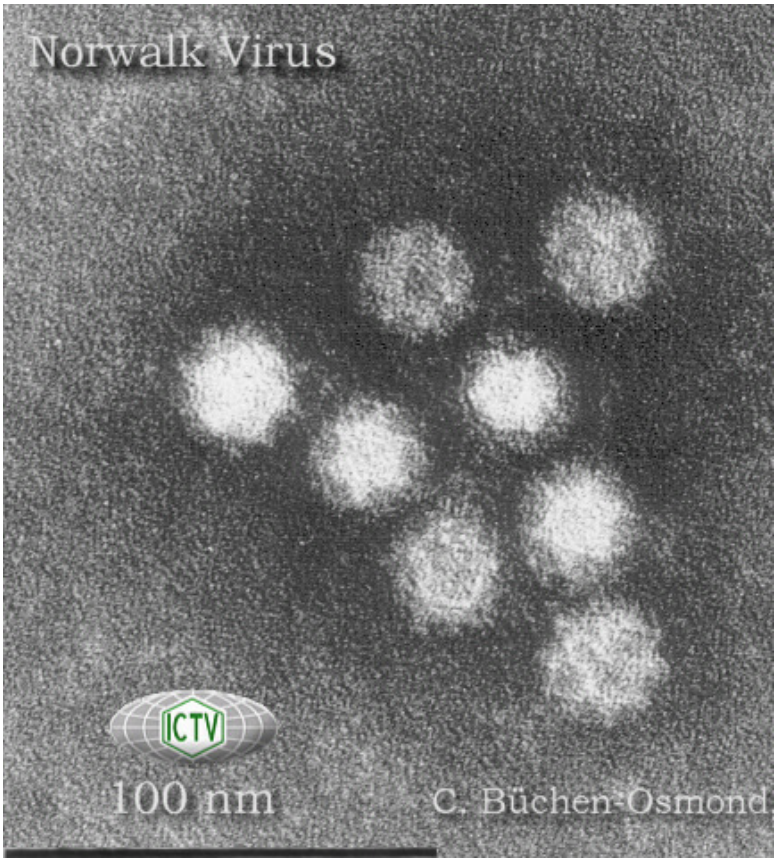


Figure 3.1: Electron micrograph of Norwalk virus. (Fauquet et al. , 2005)

The incidence of norovirus infection is likely to be at least 1% of the population per year and all age groups are affected (Caul, 1996 p 959). Incidence is highest in young children but the severe infection is more prevalent in the elderly, especially the institutionalised. Infection occurs all year round, with a peak in the UK during the cooler months. However, in 2002 a summer peak was also recorded (Lopman et al., 2002). Immunity appears to be short-lived and only to the specific strain, meaning that individuals are likely to be repeatedly infected during their lifetimes. There is some evidence that a genetic susceptibility may exist, with people of blood group O being at greatest risk of severe infection (CDC, 2006).

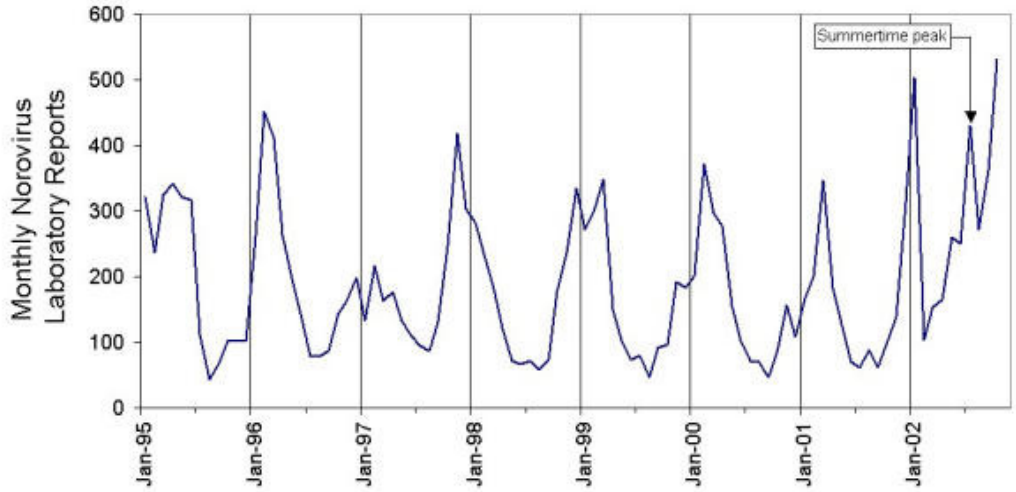


Figure 3.2: Laboratory reports of confirmed Norovirus infections in England and Wales, 1995 to 2002. From Lopman et al., 2002.

Norovirus infection lasts 12-60 hours, causing stomach cramps followed by vomiting (especially forceful) and/or diarrhoea. The shortness of the symptoms means that affected individuals rarely seek medical attention but the effect of an outbreak in a hospital environment can have a significant effect on its activities.

The reservoir for Norovirus is humans and transmission is via the faecal-oral route, including food contamination. Person-to-person spread is either direct (including aerosol transmission) or indirect (via contaminated surfaces). The indirect route leads to its high secondary attack rate.

The incubation period is 15-50 hours (Hawker et al., 2005 p 169 – although CDC, 2006 quotes 24-48 with a median of 33-36 hours) with the infectious period commencing prior to symptoms (excretion of the virus in faeces appearing a few hours beforehand (Chadwick et al., 2000 p 3)) and lasting until 48 hours after the cessation of symptoms, but is highest in the first 48 hours. The attack rate is high (around 50%) and post-infection immunity brief (a few months). Norovirus is extremely infectious, requiring as little as 10-100 virons to cause infection with 10^{11} virons per gram being excreted in the stool and 30 million in a vomiting incident. (Boone & Gerba, 2007 p 1691). The high attack rate is so high means that by the time an outbreak on a ward has been detected, most susceptible

individuals will have been exposed to it, especially if vomiting is prominent (Chadwick et al., 2000 p 1). There is no evidence to suggest that an infected person can become a long-term carrier of norovirus (Bresee et al., 2002 quoted in Vanderpas et al., 2009 p 220). Asymptomatic infection (and especially the transmission via it) is not well understood, but may occur in as many as 30% of infections (CDC, 2006).

There are three points at which SRSVs may be controlled: introduction to the hospital; containment at ward level; measures to prevent spread to other wards. Infected individuals are usually isolated until 48 hours after cessation of primary symptoms. Hygiene controls are also utilised, as are barrier methods (gloves and aprons). Infected individuals should restrict (preferably completely removing) their movement, especially to uninfected areas, whilst the infection "burns out" in the infected area. Staff should avoid moving from infected to non-infected wards unless 48 hours elapse beforehand. Where movement between wards is necessary, non-infected wards should be visited first.

3.2.1.5 Summary of Infections

Whilst each of the four infections described above have their individual characteristics, the properties that are used in constructing the model described in Chapter 4 may be summarised in Table 3.1

Property	C.Diff	Influenza	MRSA	Norovirus
At Risk	Elderly, hospitalised patients.	All ages, esp. children and elderly. Elderly most at risk of fatality.	All.	All ages but especially young children and the elderly.
Increased Risk	Age, antibiotic treatment, cyotoxic agents, intensive care, naso-gastric intubation, concurrent illnesses, alteration in gut motility.	November to March.	Prolonged hospital stay, intensive care, surgical procedures.	Cooler months.
Transmission	From symptomatic individuals via contact (including via hands of uninfected staff) OR build-up of spores in the environment OR contaminated fomites.	Respiratory secretions of infected individuals. Enhanced by enclosed and overcrowded spaces.	Contact and invasion usually via broken skin.	Faecal-oral route, including food contamination.
Property	**C.Diff**	**Influenza**	**MRSA**	**Norovirus**
Attack Rate	0.49%-2.25%	10-20%	0.35%-10% (54% recorded for one burns unit study)	50%
Incubation	From 1-2 months to a few days	1-3 days	4-10 days	15-50 hours
Infectious Period	Symptoms (esp. diarrhoea) present.	1 day before to onset of symptoms to 3-5 days after this point.	Until the colonisation is eradicated.	48 hours after cessation of symptoms.
Immunity	No.	Mainly to same strain, for many years.	No.	Brief – a few months. May only be to the specific strain.
Immunisation Programme	No.	Yes, very effective.	No.	No.
Control Advice	Isolation. Gloves and aprons. Handwashing.		Reduce patient movement. Isolation. Antibiotics. Hygiene.	Isolation. Hygiene. Barrier methods. Restricted patient and staff movement.

Table 3.1: A summary of the properties of four infections, used to develop the model described in Chapter 4.

3.2.1.6 Types of Epidemic

The two main types of epidemic are common source and host-to-host. These are compared in Figure 3.3.

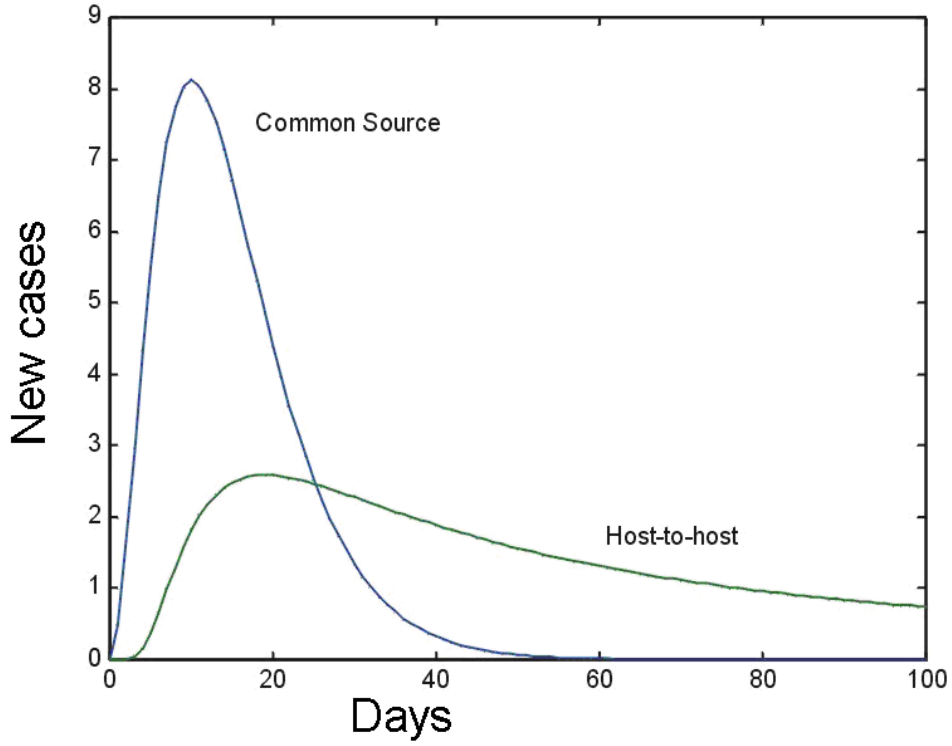


Figure 3.3: Common source epidemics usually produce more new cases earlier and faster than host-to-host epidemics. Once the infected source is closed, sealed, or removed, the common source epidemic usually abates rapidly. Host-to-host epidemics are slower to grow and slower to diminish. From: uhavax.hartford.edu/bugl/histepi.htm

3.3 Other Modelling Approaches

Graph or Network modelling is not the most common method of modelling infection propagation and control. The main method is statistical/deterministic whilst geographical methods are growing in popularity, especially with increased computing power becoming available. Examples of these are now described.

3.3.1 The Standard SIR Model

The SIR (susceptible-infected-removed) model divides the population into three groups: those that may become infected (S), those that are infected (I) and those that have been infected and are either immune or deceased (R). At time t the number in each group is $S(t)$, $I(t)$ and $R(t)$. It is worth noting that in practice, only

R(t) can truly be known. These three functions are governed by the nonlinear differential equations

$$\frac{ds}{dt} = -rS(t)I(t)$$
$$\frac{dI}{dt} = rS(t)I(t) - aI(t)$$
$$\frac{dR}{dt} = aI(t)$$

Where *r* is the infection rate and *a* the removal rate of infectives.

The SIR model is useful in predicting two figures: the limiting value R_∞, the total number of infected people at the end of the epidemic, and the basic reproduction number R_o, the average number of infections caused when an infected individual is introduced into a susceptible population. Thus the SIR model is very good at describing the overall life and total effect of an infection, but it cannot describe effectively how the infection will spread. As it is a statistical model, it is poor when the numbers of cases are low (15-25). It is also, as Ng et al. (2003) point out, poor when a double epidemic is taking place. Their SEIRP (Susceptible-Exposed-Infected-Removed-Protected) model overcomes these limitations, but the geographic spread of the infection is still not described. The most common extension is to SEIR (Susceptible-Exposed/Latent-Infected/Infectious-Removed/Recovered) which introduces an extra differential equation to the model.

Deterministic models such as these can provide good indicative data but are poor at the limits where a few individuals are involved. SEIR can lead to a steady state whereas a stochastic model is more likely to see the infection end. (Vanderpas et al., 2009 p 220).

3.3.2 Geographic Information Systems (GIS)

GIS was initially a visualisation tool rather than a predictive model. As public health bodies began to generate larger data sets and store them electronically, so the desire to be able to visualise this data grew.

GIS in its simplest form maps the infection occurrences onto a geographic map (see Figure 3.4) and animates using a time-based measurement, thus showing how the infection has spread and how and when it either grew into an epidemic or died out. It is therefore predominantly a reflective tool, rather than a predictive one.

The power of a GIS comes from the ability to aggregate and visualise large data sets and thereby discover patterns in the data – in this way, GIS can be used to spot an epidemic earlier than might be realised using conventional reporting tools. For example, a threshold might be set for a certain number of cases in a certain sized area within a certain time frame. Traditionally this has been monitored by dividing a larger area into set smaller ones and counting the occurrences in each smaller area. The GIS can aggregate in many different ways, thus determining whether an area that is of the correct size but crosses one of these divisions contains sufficient cases to trigger an alert.

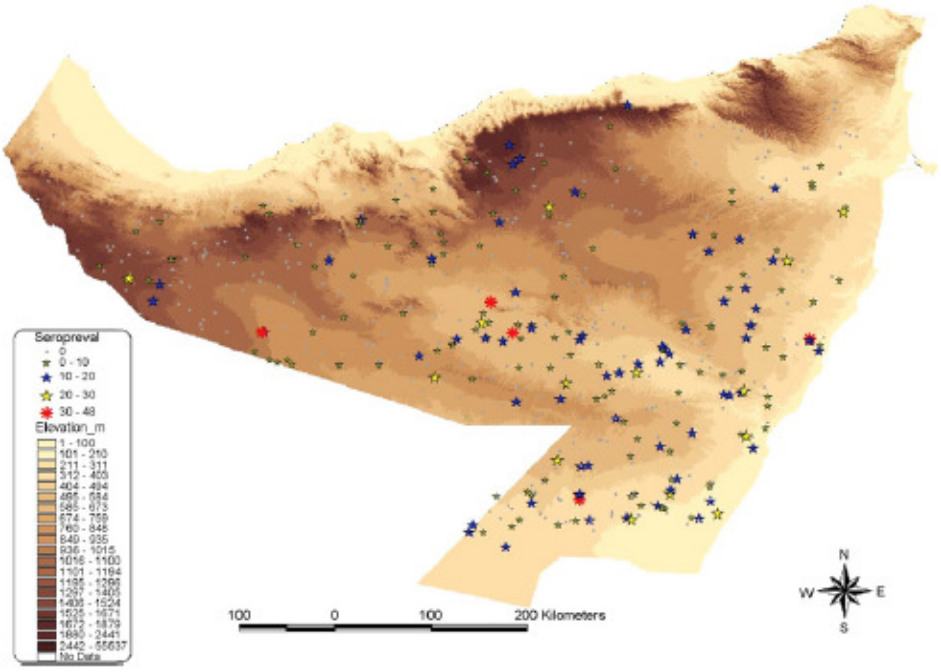


Figure 3.4: Rift Valley Fever spatial distribution in relation to ground elevation, from Soumare et al., 2007 p 253

3.3.3 Summary

The three forms of model all have their uses. For hospital-based research, a GIS is not sufficiently detailed to provide useful information – the plotting of infected

areas can also be derived from an SIR and from a graph model, especially as the data set is not likely to be that large. An SIR (or variant) model gives good overall information on the spread of an infection and the possibility of reaching epidemic levels but it does not investigate the routes of transmission as directly as a graph model. It is also poor when the number of cases is low (Trust data for Norovirus, for example, indicates average cases of less than four per ward – see chapter 5 for a fuller description) and therefore may not be as suitable as a graph model for investigating interventions.

4 Modelling Infection Propagation on a Graph

4.1 Introduction

The classical mathematical approach to disease modelling (in particular, disease spreading) either ignores the structure of the social network altogether or treats populations as spatially distributed in a continuous medium. Typically, the first case uses an SIR model (see 3.3.1) and subdivides the population into three sub-populations whose number, size and interaction determine the transmission of disease. This approach has been utilised effectively in the modelling of infection in well-mixed populations (May & Nowak, 1994; Murray, 1993 – both quoted in Watts, 1999 p 167) with an emphasis on the detailed dynamics of disease transmission rather than the relationships between subpopulations.

The second classical approach introduces a spatial dependency to the subpopulations involved and is typified by reaction-diffusion equations (Murray, 1993 – quoted in Watts, 1999 p 167). Here questions of the stability of equilibria and the analytic tractability of solutions tend to dominate.

A third approach began to appear in the late 1980s that took greater account of the fact that populations are often inherently discrete and exhibit high levels of structure (see Sattenspiel & Simon, 1988 for one such approach).

None of these approaches, however, treats the spread of an infection within a population as a function of the structure of that population.

Mathematical modelling of virus spreading and epidemics has generally utilised one of two models, both of which can be utilised on graphs:

- a Susceptible-Infected-Susceptible model (SIS), in which vertices are either “healthy” or “infected”. At each time step a healthy vertex becomes infected with probability ν if it is connected to at least one infected vertex. An infected vertex is cured with probability δ, defining an effective spreading rate of $\lambda \equiv \frac{\nu}{\delta}$. The behaviour of the SIS model is well

understood for vertices in a regular lattice or random network (Anderson & May, 1991; Nowak & May, 2000, both quoted in Dezsö & Barabási, 2002).

- a three-compartment model, the compartments being those who are susceptible to the disease, those who already infected and those who have been removed either through immunity or death (SIR: Susceptible-Infected-Removed). The probabilities of movement between the three compartments are fixed, and are the same for each individual within a compartment, as is the probability of contact with an individual from one of the other compartments.

An enhanced theoretical model (utilising both new and improved concepts) that addresses the issues as set out in Chapter 3 is now proposed.

Firstly the high-level model is outlined, in terms of the shortcomings of currently published models and then some proposed extensions to graph theory that address these.

Secondly the issue of the whether a directed or undirected graph should be used is addressed in 4.3.1, with the additional properties the elements require being described in 4.3.2. The model is then described using mathematical (set theory) terminology in 4.4.

Finally, the application of this model to infection propagation and infection control is described in 4.5 and 4.6 respectively.

4.2 The High-Level Solution

4.2.1 Introduction

The following shortcomings of current published models were identified in Chapter 2:

1. SWT has generally been applied to large, even global, populations. Whilst this emphasises the effect of a shortcut upon chain length, the theory should be equally applicable in the modelling of a semi-closed

environment such as a hospital. Smaller populations have been examined, but have done so in isolation, ignoring any possible effects due to the wider system.

2. SWT utilises random connections. In experiments, the reason for a link's revelation has been in finding someone. Other reasons for links' existence are not explored.
3. The connections are generally two-way. Real-world experience suggests that links have different weightings and different transmission properties/probabilities in each direction.
4. The strength of the links is not considered: all are viewed as being equal.
5. SWT assumes static connections, i.e. once a model has been formed, its properties are investigated but the network does not alter. Real-world experience suggests that links form and reform (and the strengths alter) with time. However, for swiftly spreading infections, this may not be pertinent. Random changes in a graph results in all networks converging onto random graphs (Witten & Poulter, 2007 p 204) so the underlying initial model may not be relevant. Verdasca et al. (2004) did present a dynamic model where the vertices and edges are created afresh at each iteration with good results, but that level of "dynamic" is probably excessive – certainly the computations were "intensive" (Witten & Poulter, 2007 p 204).
6. The infections investigated merely propagate: they do not die out and infected individuals do not recover. Whilst this is true (and an appropriate statistic) for SARS, HIV/Aids or Foot & Mouth (e.g. Saramaki & Kaski, 2004; Chen, 2001; Small & Tse, 2005), it is not for Norovirus or MRSA (both of which are prevalent within the NHS).
7. The simulation methods run on the connection models are rather basic. They assume an overall probability of infection, rather than implementing different factors (for example, the probability of the link being in place at a particular time is one factor generally ignored).
8. Some infections, such as Influenza, are cyclic according to season (and some are particularly affected by school holidays) (see Figure 4.1, below). No models so far appear to allow for these "dormant" and "rampant"

phases of an infection, although Saramaki & Kaski (2004) have partly modelled this by using different strains of influenza.

9. All models consider the effect of only a single infection taking place at once. For example, Influenza A and Influenza B often co-infect and interfere with one another. Immunity from one does not grant immunity from the other. Likewise, one infection may suppress the immune system leading to an enhanced effect of a secondary one.
10. Few models consider the impact of a treatment or vaccine being introduced during the outbreak, although post-infection immunity is considered.

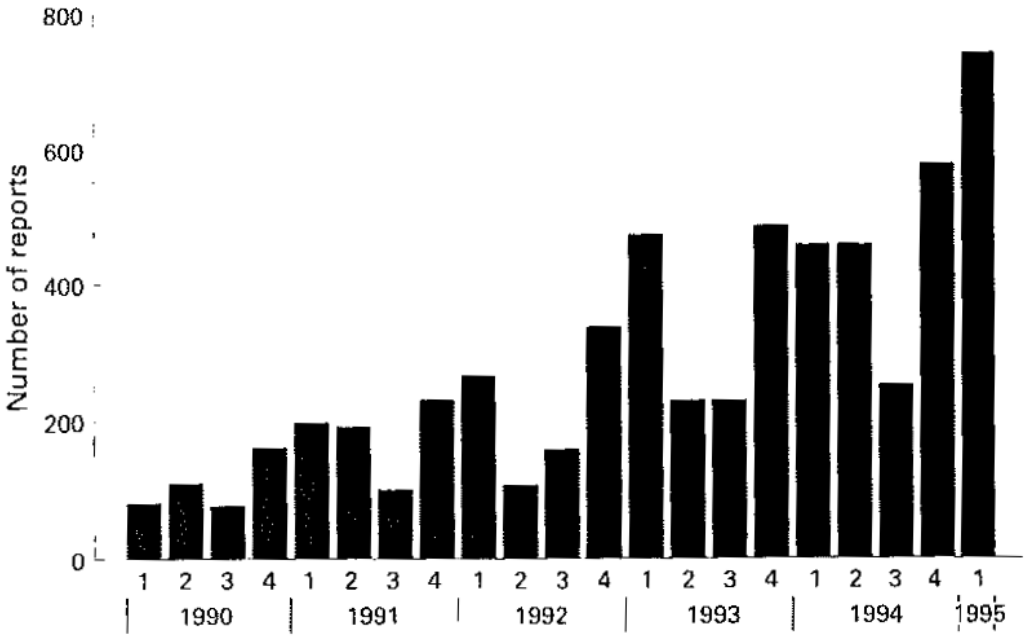


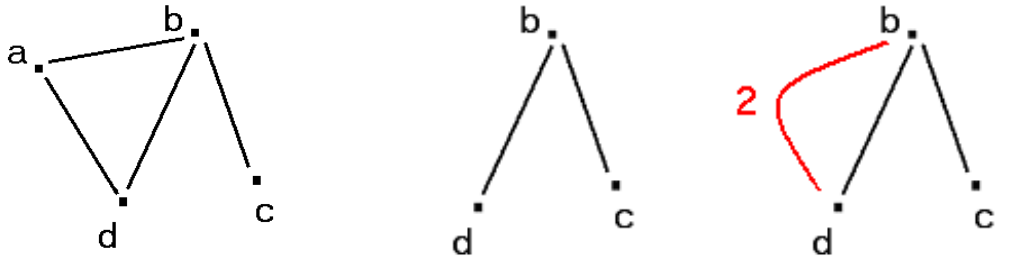
Figure 4.1: Seasonal distribution of SRSVs in England and Wales 1990-1995. Data for 1993-1995 are provisional. Taken from Caul, 1996 p 960.

4.2.2 Extensions to Graph Theory

This research introduces two new concepts: *external path* and *system* (and specifically the *semi-closed system*). These were originally described in 2.3.1, but are expounded upon here.

An *external path* is a path between two vertices in a sub-graph that utilises edges that are part of the graph, but not of the sub-graph. Therefore, as the vertices in the

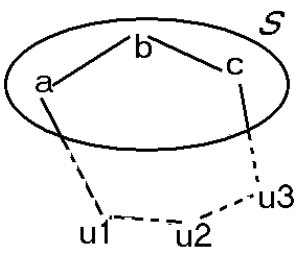
graph that are not in the sub-graph are unknown to the sub-graph, only the path's length is known.



Graph Sub-graph External path (*bd*), length 2

Figure 4.2: Illustration of an external Path, reproduced from Figure 2.5

A *system* is comprised of a sub-graph together with the set of external paths. This leads also to the concept of the *semi-closed system*. That is, there exist paths between the vertices under consideration that lie outside of the environment which utilise non-modelled vertices. Whilst essentially the same concept, the difference in terminology comes from the viewpoint. For example, when looking at a hospital as a system, it is semi-closed. When looking at a city as a system, the hospital within it is a sub-graph.



The semi-closed system, *S*, comprises the vertices {*a*, *b*, *c*}. The full system comprises the vertices {*a, b, c, u1, u2, u3*}.

The path (*a,c*) utilising the intermediate vertex *b* is an interior path.

The path (*a,c*) utilising the non-modelled vertices {*u1, u2, u3*} is an exterior path.

Figure 4.3: A semi-closed system

In order to address the shortcomings of current research outlined in Chapter 2, new concepts are required and are now described.

4.2.2.1 Information

For the problem of modelling information flow, and especially that of modelling how a vertex finds a short path to another vertex, a new concept, *information,* is

required. This is that knowledge of the context of the social networks (subgraphs) to which a connected vertex is connected yet the vertex itself is not. In the terms of the Milgram experiment (described in 2.4.2.1), this is akin to being asked to pass the letter to a lawyer in Boston. If you know someone in Boston - or someone who is a lawyer - it would seem reasonable to pass the letter to them, expecting that their social network is more likely to encompass the target.

However, for disease modelling this kind of problem does not apply, for diseases do not "seek out" certain vertices. This concept, although interesting, will therefore not be considered for this research (see Granovetter 2003 for a fuller discussion). There is the case where an infected individual will seek out a medical practitioner (i.e. the individual seeks another – however the infection has not sought that individual), but this is implemented through the strength of an edge and the dynamic elements of the model (see 4.2.2.3 and 4.2.2.4).

4.2.2.2 External Paths

An external path (see 4.2.2) is a new concept that represents those social connections that exist, yet are outside the scope of the analysis. An example might be two members of staff who have children in the same class: they may be unaware of the path between them, as it is outside of their local knowledge, yet nonetheless it exists. An external path is one where the terminating (initial and final) vertices are within the model, yet it includes vertices that are not.

This concept assists in understanding how seemingly disparate people can pass infections, through an unknown chain of contacts. As these contacts are of necessity unknown, the model will have to create them randomly, based upon known incidences of remote infection. These known incidences will determine the number, as well as the length, of the external path.

The concept of the external path provides two mechanisms. Firstly, it provides the ability of an infection to move more rapidly through the population than expected, exhibiting an ability to "jump" or take "short cuts". This is the underlying principle of the small-world group of models. Secondly, this concept provides a mechanism whereby an infection that has been eradicated from a population re-

appears without warning, as an instance is actually progressing through an external path. The infection has therefore been eradicated from the local population, but not from the global (see Appendix C for a visualisation of this). With infections such as Norovirus, which appears "spontaneously", this is the case. This concept has been termed a *long-cut* in keeping with the short-cut of a small world model.

In practice, external paths are added to an existing model. This is commensurate with the practice of Newman & Watts (1999a) in adding rather than rewiring edges, thereby avoiding cleavages and the subsequent infinite path lengths this would bring.

4.2.2.3 Remodelling

Social contacts come and go. It is therefore reasonable that a model should reflect this by adding and removing vertices, edges and external paths during the simulation. To date only limited implementations of this have taken place. This may be long-term (for example, due to moving employment) or short-term (for example, due to sickness caused by the infection under consideration or to non-working days or changing shift patterns). This feature of the model reflects the changing nature of a semi-closed society and, in particular:

- A person leaving employment or another person joining.
- A person being temporarily absent through sickness, annual leave or non-working days (e.g. weekends).
- A new social contact being established within the organisation, or one ceasing (e.g. through specialist committee work).
- A person changing social group through shift working or internal transfer.
- A person seeking out another for treatment or advice.

The extent to which a graph will be remodelled will depend upon the speed of infection (a fast-moving one will not allow time for the structure to remodel significantly, if at all) and the virulence of the infection (an infection with strong effects will cause more temporary absence than one with weak or less obvious expression).

A vertex that is infected yet remains within the model represents one of the following cases:

- An individual who attends despite illness.
- An individual who is a carrier (i.e. is infected and therefore infectious yet displays no symptoms).
- An infection which has an infectious stage prior to the outbreak of symptoms.

Remodelling due to the progress of the disease may be appropriate, to represent an intervention where all staff are advised not to attend upon first instance of symptoms.

4.2.2.4 Strength of Edge

In a graph that will remodel (change) during the lifetime of the simulation, the possibility of an edge being removed should not be purely random. For example, a link between mother and daughter will be stronger than one between work colleagues, even if the chance of infection propagation is stronger between the work colleagues, due to frequency of contact.

The strength of an edge is a new property that represents this concept, in that the stronger the edge is the less likely it is to be removed. In the mathematical model described in 4.4, the strength is a probability value. This can therefore be thought of as the probability that an edge will survive the remodelling process.

4.2.2.5 Probability of Transmission

Although Newman (2004a) has investigated weighting edges in order to represent differences in the probability of transmission, this was only for integer values. In this research, this concept is uniquely expressed in two parts: firstly, a resistance to change for a vertex (one per infection in the model) and secondly a continuous probability value (the *modification chance*) for the edge or external path in question (one per infection in the model). Note that, unlike previous models, the probability is of an infection being resisted rather than of being accepted. This is felt to be more generalisable, but as both are probability values, they are easily linked (i.e. resistance = 1-acceptance).

It could be argued that these two values (one for the vertex and one for the edge) could be combined into one (the edge only). For a static model this would give the same effect and simplify the implementation. However, for the dynamic model proposed here, this simplification would make the implementation more complex as any new edge that formed would first require the resistance to change for the vertex to be determined from the pre-existing edges' transmission probabilities and then to be combined into the transmission probability for the new edge. Additionally, the dynamic model allows for vertices to become separated (i.e. the graph to become disconnected). If the transmission probabilities were only recorded in edges, then such isolated vertices would lose one of their properties, a matter that would become important if (or when) the vertex should become re-connected.

Having determined that the probability of transmission per infection is best represented by two values, one for the vertex and one for the edge, it can then be seen how these two values represent the reality that they are modelling.

The modification chance of an edge represents the probability of an infection being transmitted between two individuals. It is composed of several components:

1. The type of social contact. This reflects the fact that infections have different transmission media: some (such as influenza) require the sharing of an air space; others (such as HIV) require bodily contact.
2. The predisposition of an individual to seek out another. This encompasses the effect observed by Ancel Meyers et al. (2006 p 401) that an infected member of the public will seek out a healthcare worker, whereas the converse is not true.
3. Gender bias. This encompasses the observation by Milgram (1967 p 65) that his messages were three times more likely to be passed to a person of the same gender than to someone of the opposite gender, with no significant differences between the genders otherwise.

The resistance to change of a vertex represents the fact that not all individuals are as susceptible to an infection – one specific instance of this is the different transmission rates of HIV to male and female, mentioned above.

4.2.2.6 Context

As has been noted above, the environment in which social contact takes place may have a bearing upon the modification chance. Whilst a rumour may pass more easily via e-mail, HIV will not. The context in which the edge exists is therefore a modifier upon the modification chance, thereby breaking down the possibility of the transmission of the modifier into three elements: context, modification chance and resistance to change.

4.2.2.7 New Metric – Path Length Matrix

The Path Length Matrix (PLM) is an extension of the Adjacency Matrix and records information about the shortest paths between vertices.

$P=[p_{ij}]$ where p_{ij} is the shortest path between vertices i and j.[1]

It is calculated by first forming the Adjacency matrix, A, from the set of edges E. Then each row is examined in turn in order to build up paths of length 2, then 3 etc. The algorithm is described in both mathematical notation and pseudo-code (see 4.4 for notation):

$$\forall v, w \in V, v \neq w$$
$$\quad if\ \exists p \in P\ s.t.\ p.e1.v1 = v\ and\ p.el.v2 = w$$
$$\quad then\ \forall u \in V, u \neq v, u \neq w$$
$$\quad\quad if\ \exists q \in P\ s.t.\ q.e1.v1 = u$$
$$\quad\quad\quad and\ q.el.v2 = v$$
$$\quad\quad\quad and \neg\exists r \in P\ s.t.\ r.e1.v1 = u$$
$$\quad\quad\quad and\ r.elv2 = w\ and\ |\ r\ | < (|\ p\ | + |\ q\ |)$$
$$\quad\quad then\ p_{vu} = |\ p\ | + |\ q\ |$$

[1] This is not the same thing as a transitive closure (TC). A TC is a graph where each edge represents a path in the original graph. A PLM is a matrix that describes the lengths of these shortest paths – information that is missing from a TC.

```
while(changes made to matrix)
{
 for(i=0;i<no. of vertices;i++)
 {
  for(j=0;j<no. of vertices;j++)
  {
   if(path exists from i to j)
   {
    for(k=0;k<no. of vertices;k++)
    {
     if(path from k to i exists
              and a shorter path doesn't exist from k to j) set (k,j)=(k,i)+(i,j)
    }
   }
  }
 }
}
```

The Path Length Matrix is used in the calculation of some statistics, especially the Characteristic Path Length. This calculation then involves taking the mean of each row (assuming the notation is row:column = from:to, otherwise columns are used) and taking the median of these means.

Algorithms already exist for the calculation of shortest paths, in particular those of Floyd and Dijkstra (see http://www-unix.mcs.anl.gov/dbpp/text/node35.html for a description). However, these algorithms are general-purpose ones for edges of varying length. In the model presented here, the edges are considered to be of equal length (taken to be unity for simplicity). The methods of Floyd and Dijkstra complete in N^3 and FN^3 comparisons, respectively (where F is a constant empirically shown to be approximately 1.6). They also require additional storage for the intermediate matrices and sets, respectively.

The PLM algorithm described above completes in a maximum of N^3 comparisons and requires no additional storage (beside the usual counters and index markers). For this particular case, the PLM algorithm is therefore more efficient than either the Floyd or Dijkstra ones.

Note that the PLM does not compute all walks or trails between all vertices, but only the shortest paths between all vertices (where a path exists). Likewise, it does not compute a walk or trail from a vertex to itself. (See 2.3.1 for terminology).

This PLM is not the same as a transitive closure, as has been suggested. Although both are useful in solving reachability questions (i.e. "Does a path exist between a selected pair of vertices?"), they provide different information. A transitive closure is a graph with the same vertices as the original graph and edges (i,j) iff a path exists between vertices i and j in the original graph. The adjacency matrix of this graph is therefore used to answer the reachability questions. A transitive closure (or its adjacency matrix) therefore records where paths exist (which, for a connected graph, is redundant) whereas a PLM records the lengths of these paths.

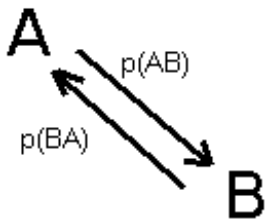
4.2.3 Directed vs. Undirected Graphs

Most of the published work on the use of graphs to model disease propagation uses simple, undirected graphs. The assumption, therefore, is that a disease is as likely to pass in one direction between two people as the other. This, however, is not true. HIV, for example, passes easier from men to women than vice versa and an uninfected Healthcare Worker is more likely to be sought out by an infected member of the public than an infected Healthcare Worker is to seek out that same uninfected member of the public. (Ancel Meyers et al., 2006 p 401 - quotes HIV data from Italian Study Group on HIV Heterosexual Transmission).

The reason most investigations have concentrated on simple undirected graphs is that these graphs lend themselves to mathematical analysis. As this research aims to develop a more "realistic" model, directed graphs are used meaning that only computer simulations are currently possible.

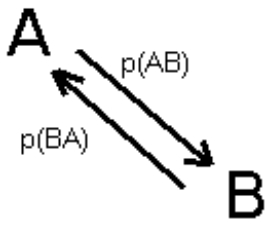
The use of directed graphs allows all four cases of host-to-host infection propagation to be modelled, as shown in Figure 4.4.

1. An infection has equal probability of passing between A and B



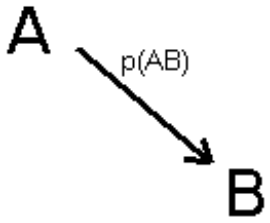
p(AB)=p(BA)

2. An infection has differing probabilities of passing from A to B and from B to A



p(AB)≠p(BA)

3. An infection has a probability of passing from A to B, but will not pass from B to A



4. An infection cannot pass from A to B nor from B to A

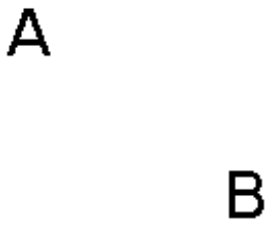


Figure 4.4: The four cases of host-to-host propagation.

Undirected graphs can be used to model cases 1 and 4, but cannot model cases 2 and 3. Case 3 may be seen as a special case of case 2, with p(BA)=0 (a null edge) and case 4 as a case where p(AB)=p(BA)=0. If this approach is utilised, then a fully connected graph may be used, whereupon there is no computational overhead for the addition or removal of edges (see Dynamic Model Operations, below) but there is a large storage overhead. In the case examined in chapter 6 with 376 vertices and 3347 edges, 137278 additional null edges would have been required if this approach had been taken. The computational overhead of edge addition and removal was therefore deemed acceptable.

4.2.4 Additional Properties

In order to use a graph to model the progress of an infection through a population, the elements of the graph require additional new properties in order to better describe the entities that they are modelling. These additional properties are described in sections 4.2.4.1 to 4.2.4.3, and in mathematical notation in section 4.4.

4.2.4.1 Vertex Properties

A vertex, representing a person or location, has a current status and a future status (together with a time until that status change takes place). A vertex also has a probability of a change of status. These properties are linked together (using a Susceptible-Infected-Immune-Susceptible model) and shown in Table 4.1.

Current Status	**Probability of a Change of Status**	**Future Status**	**Time to Status Change**
Susceptible	Probability of becoming infected if exposed to an infected host	Infected	Immediate for exposure to infected host, delayed for exposure via an external path
Infected	Probability of infection being overcome by the host	Immune	Average life of the infection – modified by the probability of a status change
Immune	Probability of the immunity ending	Susceptible	Period of Immunity – modified by the probability of a status change

Table 4.1: The progression between vertex statuses.

4.2.4.2 Edge Properties

An edge, representing a possible contact between two hosts, has a probability of a change of status being transmitted via this edge and a probability that this route of contact will continue to exist.

4.2.4.3 Modifier Properties

A modifier is an entity that modifies the status of a vertex and is passed throughout the graph via edges and external paths. As such it has a status that it will transfer a vertex to, and a statistical lifespan (normally distributed with specified mean and standard deviation, plus a tail length).

A vertex in a non-normal and non-null status contains a modifier, which may therefore be passed on without the vertex relinquishing the modifier.

4.3 The Theoretical Model

As all infections under consideration pass from person to person by contact, the propagation of infection through a network is easily modelled by introducing an infection into the structure (i.e. by setting a number of vertices in the graph to be infected), then allowing the infection to move from vertex to vertex along the edges, much as Milgram modelled the flow of information between his subjects. Modifications to this simple idea are introduced via the SIR or SIS models described above and other techniques described later.

A problem with the two main models studied (SIS and SIR) is that neither allow for a vertex to be "temporarily removed" from the set of susceptible vertices (leading to a SIRS model). Whilst some diseases (e.g. measles) confer a life-long immunity and some (such as the common cold) confer none, others (e.g. Respiratory Syncytial Virus (RSV)) do confer a period of immunity, as do all vaccines (e.g. Chickenpox). Neither model takes account of natural passive immunity, probably because it is fairly rare except between mother and child. There is also the question of the degree of infectiousness and the level of susceptibility of an individual. Some pathogens, such as Smallpox, are highly contagious, so will have high transmission probabilities. Others, such as SARS, will have low ones.

The enhanced model, as described here, addresses these limitations.

4.4 Description of Mathematical Model

The graph model discussed in Chapter 2, together with the extensions described above may be described in a mathematical manner, particularly via set theory notation. This description now follows.

Firstly some nomenclature in use is described and some basic sets are defined in 4.4.1.

4.4.2 – 4.4.6 defines the major entities of the model: Vertices, Edges, Paths, Graphs and Systems.

Three new concepts are defined in 4.4.7 – 4.4.9.

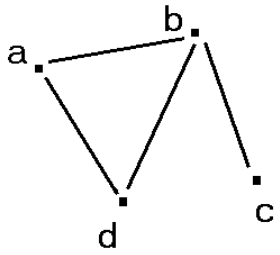
4.4.10 describes the movement of a process across the model in terms of status changes to Vertices.

The creation of an initial structure, a synthetic graph, is defined in 4.4.11.

4.4.12 defines the modification of the model and stopping conditions for the model are defined in 4.4.13.

Finally, some limitations of this model are described in 4.4.14.

As this is an abstraction of the model, the Figures 4.5 below and 4.2 above may prove useful in contextualising the concepts. The model as defined is still very abstract, so, in order to enhance the readability of this section, an implementation-specific description including contextualisation is added to each concept.



V = {a,b,c,d}
E = {ab, ad, bc, bd}

Note that order in the sets is unimportant.

The path (c,d) has one intermediate vertex (b) and its length is 2 (edges bc and bd). This is the shortest path as another (via b and a) exists with length 3.

The trail (c,b) passes through all vertices (c, b, a, d, b or c, b, d, a, b) but uses each edge only once. There also exists a shorter trail with length 1.

The walk (c,b) may complete the loop involving a and d multiple times. It's length is therefore one of 1, 4, 7, …

The adjacency matrix is

$$\begin{array}{c} \begin{matrix} a & b & c & d \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix} \end{array}$$

The path length matrix is

$$\begin{array}{c} \begin{matrix} a & b & c & d \end{matrix} \\ \begin{pmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 1 & 1 & 2 & 0 \end{pmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix} \end{array}$$

Figure 4.5: A simple graph

4.4.1 Preliminary Definitions and Nomenclature

N is the set of natural numbers (including zero[2])

Z is the set of integers

R is the set of real numbers

Φ is the empty set

[2] There exists debate over whether zero should be included in the set of natural numbers, dependent on which branch of mathematics is being considered. In the context of this work, zero is included.

Let **Q** be the set of probability values

$Q = \{x \in R \mid 0 \leq x \leq 1\}$

Let **T** be the set of statuses that a vertex may take

T={Null, Normal, Status 1, Status 2, …, Status n}

Let **X** be the set of contexts that an edge may be in

X={Context 1, Context 2, …}

N(μ,σ) is a random value from a normal distribution with mean μ and standard deviation σ.

Except where specifically noted, all sets described herein are unordered.
Except where specifically noted, all sets described herein are unique, in that all elements in a set occur exactly once.

Properties of entities are described using dot notation, in that a.b represents the property or instance b of instance a. Similarly, a.b.c represents the property or instance c of the instance or property b which is a property or sub-member of instance a. Multiplications are therefore represented by *.

4.4.1.1 Application

A status is the infection state: Normal (uninfected), Status 1 (infected by infection 1), etc. This permits multiple different infections to be released into the model simultaneously.

A context is an environment in which the social contact takes place: it is the setting that the edge may be in. This represents the difference in infection rate in, say, a large open space and a small enclosed one.

4.4.2 Vertices

Let *v* be a 7-tuple of the form

(ordinal *o*, status *s*, resistance to change $\{c_{null}, c_{normal}, c_1 \ldots c_n\}$, reversion $\{r_{null}, r_{normal}, r_1 \ldots r_n\}$, immunity $\{i_{null}, i_{normal}, i_1 \ldots i_n\}$, status to change to *s2*, time until change *t*)

where

$o \in Z, o$ unique

$s \in T$

$c_j \in Q, 1 \leq j \leq n$

$r_j \in Q, 1 \leq j \leq n$

$i_j \in Z, 1 \leq j \leq n$

$s2 \in T$

$t \in Z$

$n = |T| - 2$

Then *v* is a vertex and **V** is the set of such vertices.

4.4.2.1 Application

A vertex represents a person. This person therefore has a current health status, a resistance to each infection, a probability of reverting from the current infection, a period of immunity to each infection following infection and a delayed time to infection, representing an infection to be transmitted via an external path. c_{null}, c_{normal}, r_{null}, r_{normal}, i_{null} and i_{normal} are meaningless in this implementation as resistance to, reversion from and immunity to *null* and *normal* is meaningless.

4.4.3 Edges

Let *e* be a 5-tuple of the form

(vertex *v1*, vertex *v2*, modification chance $\{m_{null}, m_{normal}, m_1 \ldots m_n\}$, strength *s*, context *x*)

where

$v1 \in V$

$v2 \in V$

$m_j \in Q \qquad 1 \leq j \leq n$

$s \in Q$

$n = |T| - 2$

$x \in X$

Then *e* is an edge between vertices *v1* and *v2* and **E** is the set of all edges between all $v \in V$.

e is denoted $e_{a,b}$ where *e.v1.o=a* and *e.v2.o=b*.

Note that in the general case $e_{a,b} \neq e_{b,a}$ as these are distinct edges.

4.4.3.1 Application

An edge represents a direct social contact between two people. There is a probability that an infection may be transmitted via this contact, which may be different for each infection within the model. The strength of the edge represents the strength of the social contact, i.e. how likely it is to withstand changes in circumstance. The context represents the current environment, i.e. a modifier upon the modification chance.

The model uses a modified SIRS (Susceptible-Infected-Recovered/Immune-Susceptible) process. The modification is in the transmission probability calculation: in previous models this has been represented as one figure – the probability of an infection passing from an infected vertex to an uninfected neighbour and thus infecting it. In the model here presented, the probability is divided into two components: the probability of the infection being passed via the edge, and the probability of the vertex becoming susceptible. These are referred to as the *modification chance* and *resistance to change* respectively. Whilst these figures may be combined into the more usual model, doing so does not allow the effect of an increased susceptibility in a location (nor a reduced one via, for example, a new infection control protocol being introduced) to be easily modelled. This feature will also allow for dynamic structures to be easily modelled.

4.4.4 Paths

Let p be a 1-tuple of the form

(edges $\mathbf{F}=\{e1,e2,\ldots,el\}$)

where

$e1, e2, \ldots, el \in E$

F is an ordered set as it describes the path in order of the edges traversed.

Then p is a path between vertices $e1.v1$ and $el.v2$ and **P** is the set of such paths between all $v \in V$.

The length of the path, $p.l = |p|$

The modification chance of the path, $p.m_j = \prod_{e \in p.F} e.m_j$

When $|p|=1$, p is denoted $p_{\{a,b\}}$ where $p.e1.v1.o=a$ and $p.e2.v2.o=b$

When $|p|=2$, p is denoted $p_{\{a,b,c\}}$ where $p.e1.v1.o=a$ and $p.e3.v2.o=c$ and $\exists v3 \in V \; s.t. \; v3.o = b$ and $\exists e_{a,b}, e_{b,c} \in E$

NB when $|p|=1$, $p_{\{a,b\}}=e_{a,b}$

4.4.4.1 Application

A path represents an indirect social contact between two people where all the intermediate people are known. Paths of length 1 are of no interest in this model (consisting, as they do, of a single edge and are therefore a direct social contact) and are omitted from the implementation.

4.4.5 Graph

A graph, **G**, is the collection of the set of vertices and the edges that connect them, i.e. **G**={**V**,**E**}, **V**≠Φ. **G** is viewed as the universe (sometimes termed "world") under consideration.

There exist three special cases:

- A connected graph,
  $G_c = \{V, E \mid V \neq \Phi, \forall v1, v2 \in V \; \exists \; p \in P \; s.t. \; p.e1.v1 = v1 \; and \; p.el.v2 = v2, l = \mid p \mid\}$
- A directed graph, $G_d = \{V, E \mid V \neq \Phi, \exists e_{a,b}, e_{b,a} \in E \; s.t. \; e_{a,b}.m \neq e_{b,a}.m\}$
- Therefore an undirected graph,
  $G_u = \{V, E \mid V \neq \Phi, e_{a,b}.m = e_{b,a}.m \; \forall e_{a,b}, e_{b,a} \in E\}$

4.4.5.1 Application

The graph is the limit of the model – the total population under consideration (e.g. Hospital, City, Conference).

4.4.6 System

Let **S** be 5-tuple of the form (**V_s**, **E_s**, **P_e**, addition *a*, removal *r)*

where

$V_s \neq \Phi$

$a \in Q$

$r \in Q$

and **P_e** is the set of external paths as defined in 4.4.7 below.

Then **S** is a system and is a sub-graph of **G**, such that

$S \subset G$

$V_s = \{v \mid v \in V, v \in S\}$

$V_s \neq \Phi$

$E_s = \{e \mid e \in E, e \in S\}$

$\forall (v, w) \in V_s, e_{v.v1.o,w.v1.o} \in E_s$

If **S≠G** then $\exists v \in G, v \notin S$. If the system is a connected graph then $\exists e \in G, e \notin S$ and $\exists p \in G, p \notin S$

4.4.6.1 Application

The vertices in a system are the part of the graph (population) that is known. The edges in it are also known. The paths are known to exist, but without the detail of the edges.

4.4.7 Internal and External Paths

A path *p* is internal when $p.e_i \in E_s \; \forall 1 \leq i \leq \mid p \mid$

A path *p* is external when $\exists i, 1 \leq i \leq \mid p \mid s.t.\; p.e_i \notin E_s, p.e_i \in E$

P_i is the set of internal paths in the graph

P_e is the set of external paths in the graph

It follows that |*p*|>1 for all external paths as *p* for which |*p*|=1 is an edge.

As not all *e* are defined for an external path, an external path is defined as follows:

Let *p* be a 5-tuple of the form (ordinal *o*, vertex *v1*, vertex *v2*, length *l*, modification chance $\{m_1 \ldots m_n\}$) where

$o \in Z, o$ unique
$v1 \in V$
$v2 \in V$
$l \in N, l > 1$
$m_j \in Q, 1 \leq j \leq n$
$n = |T| - 2$

4.4.7.1 Application

An external path represents an indirect social contact between two people where at least one of the intermediate people is unknown. In practice, this becomes the condition that none of the intermediates are known as an external path where only one intermediary is unknown can be decomposed into one or two internal paths and one external path.

4.4.8 Connectedness

The connectedness of a vertex, $v = |E_v|$ where $E_v = \{e \in E, e.v1 = v \text{ or } e.v2 = v\}$

For an undirected graph, connectedness is the same as degree. For a directed graph, they are different, with degree≤connectedness and connectedness=indegree+outdegree.

4.4.8.1 Application

The connectedness of a vertex represents the number of direct social contacts that that person has.

4.4.9 Modifier

Let m be a 5–tuple of the form (status s, lifespan mean $l\mu$, lifespan deviation $l\sigma$, lifespan tail lt, seasonal variant $\{sv_1, sv_2, sv_3, sv_4\}$) where

$s \in T, s \neq \text{Null}, s \neq \text{Normal}, \text{s unique}$

$l\mu \in R, l\mu \geq 0$

$l\sigma \in R, l\sigma \geq 0$

$lt \in N$

$sv_i \in R, sv_i \geq 0, 0 < i < 5$

Then m is a modifier and **M** is the set of such modifiers.

The distribution ($l\mu$,$l\sigma$) is only defined for positive values.

$\{sv_i\}$ is a tuple.

$|\mathbf{M}|=|\mathbf{T}|-2$

4.4.9.1 Application

A modifier represents an infection. It has a normally-distributed lifespan, together with a lifespan tail representing a period of infectiousness following the cessation of symptoms (see 6.2.2.3.9.1 for a fuller description around a specific example). It may have different effects according to the season. A modifier's status is the status into which it seeks to place a vertex, i.e. it is the infected state into which a person becoming infected will become.

4.4.10 Vertex Behaviour

At each t>0,

Action	Description
Form $V_n = \{v \mid v \in V_s, v.s \neq Normal\}$	Find all non-normal vertices
Form $E_n = \{e \mid e \in E_s, e.v1 \in V_n\}$	Find all vertices connected to non-normal vertices via edges
Form $P_{en} = \{p \mid p \in P_e, p.v1 \in V_n\}$	Find all vertices connected to non-normal vertices via external paths
$\forall e \in E_n$, determine whether *e.v2* will change status, using the season, $e.m_j$, *e.x*, $v2.c_j$ and $v2.i_j$, where *e.v1.s*=Status*j*.	Find all normal vertices connected via edges that will change status
If so, set *e.v2.s2*=*e.v1.s*, *e.v2.t*=1.	Mark these to change status on this iteration
$\forall p \in P_{en}$, determine whether *p.v2* will change status, using $p.m_j$, $v2.c_j$ and $v2.i_j$, where *p.v1.s*=Status*j*.	Find all normal vertices connected via external paths that will change status
If so, set *p.v2.s2*=*p.v1.s*, *p.v2.t*=*p.l*.	Mark these to change status on the iteration (next+|path|)
$\forall v \in V_n$, determine whether *v* will revert, using $v.r_j$ where *v.s*=Status*j*.	Find all non-normal vertices that will revert before the pre-allocated time
If so, set *v.s2*=Normal, *v.t*=1	Set these to revert on the next iteration
Form $V_c = \{v \mid v \in V_s, v.s2 \neq Null, v.t = 1\}$	Find all vertices about to change status
$\forall v \in V_c$, set *v.s*=v.s2, *v.s2*=Null, *v.t*=0 if *v.s*=Normal, randomly determine $v.i_j$, where *v.s* previously=Status*j* else set *v.s2*=Normal, *v.t* using N($m.l\mu$,$m.l\sigma$), $m.sv_k$, *m.lt* where *v.s*=Status*j*, season=*k* and *m.s*=*v.s*.	Change the status for these and set the appropriate time parameter and immunity
Form $V_{c2} = \{v \mid v \in V_s, v.s2 \neq Null, v.t > 1\}$	Find all vertices due to change status on a later iteration
$\forall v \in V_{c2}$, set *v.t*=*v.t*-1	Reduce the time to change by 1
Form $V_{i,j} = \{v \mid v \in V_s, v.i_j > 1, 1 \leq j \leq n\}$	Find all immune vertices
$\forall v \in V_{i,j}$, set $v.i_j$=$v.i_j$-1	Reduce the immunity interval by 1
If the season is to change, do so.	
Set *t*=*t*+1	Increment the time and proceed to the next iteration

Table 4.2: The process for altering the state of vertices in the system on each iteration.

4.4.10.1 Application

Vertex behaviour represents the transmission of an infection (or several infections) through the social network under investigation. The description above is written in a form that is easy to implement via a computer simulation.

4.4.11 Creation of Initial Structure

4.4.11.1 Synthetic Graph

Definition: A *heuristic* is "a way of directing your attention fruitfully" (Wikipedia). In the Artificial Intelligence problem of maze solving, for example, a heuristic is the rule that is applied in order to determine which branch at a junction is likely to be the most fruitful. In the context of this research, a heuristic is a set of rules that forms the synthetic graph, forming vertices, edges and external paths with the intention of making the resultant system "more realistic".

4.4.11.1.1 Method

Determine the number of statuses to be present in the system and form **T** (|**T**|=number of statuses+2).

Form $\mathbf{V_s}$ where $\forall v \in V_s, v.s = \text{Normal}, v.i = 0, v.x = \text{Context}\,1$, *v.s2*=Null

Apply a heuristic to form $\mathbf{E_s}$. Do this repeatedly until $\{\mathbf{V_s},\mathbf{E_s}\}$ is connected.

Some possible heuristics:

- **Random**: $\forall v \in V_s$ randomly select $w \in V_s, w \neq v$ and form e such that *e.v1*=*v*, *e.v2*=*w*. Randomly assign $e.m_j$ and *e.s*. Add *e* to $\mathbf{E_s}$, unless e_{v1v2} already exists.
- **Superhub**: $\forall v \in V_s$ randomly select $w_1, w_2 ... w_n \in V_s, w_i \neq v\ 1 \leq i \leq (n.|V|), n \in Q$. Evaluate the connectedness of each w_i and select the best connected (if several equally connected, all such) and for each form e_i such that $e_i.v1=v$, $e_i.v2=w_i$. Randomly assign $e_i.m_j$ and $e_i.s$. Add e_i to $\mathbf{E_s}$. This measure may also be viewed as "popularity", i.e. the more popular vertices become better connected, i.e. increase in popularity. Other heuristics for "seeking out" a vertex to which the examined vertex wishes to connect may be developed as extensions of this basic idea. The

opposite heuristic (seeking the least connected vertex to connect to) might be termed *resilience* and has applications in computer networks, for example.

- **Grouped**: Determine the number of groups, $g \in N, g > 2$ and $\forall v \in V_s$ randomly assign each *v* to one of the *g* groups. (Note that some groups may be empty – this is definitely true where $g > |V|$ and increasingly likely as $g \to |V|$) Connect all *v* in a group to all other *v* in the same group by creating edges between them (for directed graphs this will be two edges: one in each direction). Randomly determine the total number of edges between groups and then form them by randomly selecting two *v* in differing groups and form an edge between them. This will form communities, as described in 2.3.2.
- **Movement**: Form a clustered graph (similar to the ring lattice of Watts & Strogatz), for $n \in N, n > 2$ neighbours[3]. Then randomly select a number of vertices and move each to another position on the graph, thus forming new edges in the new neighbourhood whilst preserving the (now) long-range edges to the old neighbourhood. Note that the clustered graph that forms the starting point is undirected by definition (as is the final graph). It is also connected, so the heuristic will only ever be applied once.

Apply a heuristic to form $\mathbf{P_e}$

A possible heuristic:

- **Random**: $\forall v \in V_s$ randomly determine whether *v* is a terminus (start) for an external path. If so, randomly select $w \in V_s, w \neq v$ and form *p* such that *p.v1=v, p.v2=w*. Randomly assign *p.l* (with a maximum=no. of vertices) and $p.m_j$.

Form $\mathbf{S=\{V_s,E_s,P_e\}}$.

Randomly select $v \in V_s$ *and* $t \in T$ and assign *v.s=t, v.s2*=Normal and randomly assign *v.t.* (This seeds the initial structure with one non-normal vertex).

[3] In the software implementation, six neighbours (three on each side) are used – corresponding to the diagram in Figure 2.10.

Randomly assign $\mathbf{S}.a$ and $\mathbf{S}.r$ – for a static model these=0

Set time, t=0

Set season randomly, together with length of each season.

4.4.11.2 Application

In order to test these concepts and to examine the effect upon the model of different conditions, synthetic graphs may be constructed, using the methods here described. Varying the elements used to construct the initial structure constructs different models – the effects of these variations may then be studied.

4.4.12 Dynamic Model Operations

At each t>0,

Action	**Description**
$\forall e \in E_s$, randomly determine whether $e.m_i, 1 \le i \le n$ will alter.	Find all edges that will alter modification chance
If so, randomly alter $e.m_i$	
$\forall e \in E_s$, randomly determine whether $e.x$ will alter.	Find all edges that will alter context
If so, randomly alter $e.x$	
$\forall e \in E_s$, randomly determine whether $e.s$ will alter.	Find all edges that will alter strength
If so, randomly alter $e.s$	
$\forall p \in P_{es}$, randomly determine whether $p.m_i, 1 \le i \le n$ will alter.	Find all external paths that will alter modification chance
If so, randomly alter $p.m_i$	
$\forall p \in P_{es}$, randomly determine whether $p.l$ will alter.	Find all external paths that will alter length
If so, randomly alter $p.l$	
$\forall v \in V_s$, randomly determine whether v will be removed from the system	Find all vertices about to be removed
If so, remove v and E_r where $E_r = \{e \mid e.v1 = v \text{ or } e.v2 = v\}$	
$\forall e \in E_s$, randomly determine whether e will be removed from the system, using $e.s$	Find all edges about to be removed
If so, remove e	
$\forall p \in P_{es}$, randomly determine whether p will be removed from the system	Find all external paths about to be removed
If so, remove p. If $p.v1.t$>1 and $p.v1.s2$≠Normal, set $p.v1.t$=0, $p.v1.s2$=Null. If $p.v2.t$>1 and $p.v2.s2$≠Normal, set $p.v2.t$=0, $p.v2.s2$=Null.	
Randomly determine the number of new vertices to add to the system, $vn \in Z$	Determine whether any new vertices are to be added

For each new v, use the original heuristic to add it to the system via new edges	
Randomly determine the number of new edges to be added to the system, $en \in Z$	Determine whether any new edges are to be added
For each new e, use the original heuristic to add it to the system	
Randomly determine the number of non-modelled vertices (currently in external paths) that are to become part of the system, $vp \in Z$	Determine whether any non-modelled vertices are to become modelled and join the system
For each new v, randomly select an external path $p \in P_s$	
Randomly determine which of the non-modelled vertices on the path is to become modelled, v_j	
Set $p.l$=$p.l$-j and buffer $p.v1$ in v_b	
If $p.l$=0, delete p and form an edge from v to $p.v2$ using the original heuristic	
Otherwise, set $p.v1$=v	
If j>1, form an external path from v_b to v, using the original heuristic	
Otherwise, form an edge from v_b to v using the original heuristic	

Table 4.3: The algorithm (in separate steps) for the implementation of dynamic re-modelling.

4.4.12.1 Application

The dynamic model operations represent the changes that take place within a social network, with people joining and leaving and social contacts changing (including new ones forming and old ones breaking). It also represents and implements the alteration in the probabilities of an infection being transmitted between two individuals as the social contact changes. Without this modification, an infection may become quarantined by virtue of reaching a person from which it is impossible to progress. With it, an infection may get limited as the network cleaves.

4.4.13Stopping Conditions

Four possible stopping conditions exist:

1. The simulation is run for a fixed time interval, which is reached.
2. A preset proportion of vertices have a non-normal status (for example, 100%, 50% or 0%)
3. The model has reached a steady state where the number of vertices having non-normal status varies only slightly about a certain level.

4. The model is locked into a cyclic pattern where the same vertices are altering status in a repeated pattern.

4.4.13.1 Application

The stopping conditions represent an investigation into whether an infection will become an epidemic (option 2, 50% or greater), become steady and therefore difficult to eradicate (options 3 and 4), dies out (option 2 again, 0%) or will still be existent within the population at a future time (option 1).

4.4.14 Limitations of This Model

This model allows for only one non-Normal status per vertex (which is appropriate for the investigation of competing infections, such as different strains of influenza) and if a vertex is in a non-Normal status then it will not allow another to be implemented, queued or paralleled. The level of complexity caused by the interaction of non-competing disease states is therefore unmodelled.

Only one system is studied – multiple inter-related systems are therefore not modelled. (If the systems are joined by common vertices (1 or more), then the second system may be regarded as forming many of the external paths. If the systems are joined by common vertices (1 or more) on an external path, then this system, by nature of being unable to retain a status, is completely unmodelled). This is only a significant problem if the second system is, or forms part of, the reservoir for the infection under consideration.

Removal of an external path removes the terminating vertex's impending status change (if there is one). This assumes that no other external path caused it. This may be solved by adding a flag to the status change in order to note which external path the status change is transmitted by, but is not a feature of this model.

Vertices may not change from being modelled to non-modelled, as this reduces the specificity of the model.

At present the concept of *information* has not been modelled. Whilst this is an interesting problem in itself, it is not really relevant to the current research: whilst

a person may be interested in finding a short path to another person, a pathogen has no such desire. Likewise, susceptibles are unlikely to seek out the infected (although Granovetter, 2003 p 774 does note that there are exceptions to this – children are often socialised with the infected in order to gain immunity for later life, for example.). Therefore the ability to find a short path utilising only local knowledge is outside the current scope.

4.5 Application of This Model to Infection Control

Various infection control methodologies exist. The five most common of these are now described, together with their implementation in the model.

4.5.1 Inoculation/Vaccination/Immunisation

4.5.1.1 Description

Although originally three different processes, inoculation, vaccination and immunisation are, for the purposes of this research, the same.

The process involves the administering of a live, weakened or dead pathogen to a subject with the objective that the immunity that would follow a full infection is conferred to the host without undergoing the symptoms that would normally accompany it.

Although such a process may take place after a host has become infected (as in the case of experimental AIDS, cancer and Alzheimer's disease vaccines), the definition adopted here is of a pre-infection administered immunogen which stimulates the immune system. Post-infection administration is included in treatment (4.5.2).

Some models (Ancel Meters et al., 2006, for example) have approached this by removing the vertex and all connected edges from the model. This works for long-term immunity, but does not do so for any immunity that is less than the lifespan of the outbreak.

4.5.1.2 Implementation

Action	Description
Randomly form $V_v = \{v \mid v \in V_s, v.s = Normal\}$	Find all vertices to be vaccinated against status j
$\forall v \in V_v$, set $v.i_j$=N(μ,σ)	Set the immune period from the efficacy of the vaccination

Table 4.4: The implementation in the model of a method of simulating a vaccination programme. The random nature of the selection recognises that immunisation cannot be forced, but is elective. What this does not model, though, is peer pressure either for or against the inoculation (e.g. the MMR vaccine's perceived and reported yet unproven links to Autism have reduced the uptake of it, Jansen et al., 2003).

4.5.2 Treatment

4.5.2.1 Description

A treatment is a course of action that is followed in order to hasten the end of the infection. Typically this will be pharmaceutical, but may involve factors such as enforced rest.

4.5.2.2 Implementation

Action	Description
Randomly form $V_t = \{v \mid v \in V_s, v.s \notin \{Null, Normal\}\}$	Find all vertices to be treated for current non-normal status
$\forall v \in V_v$, determine if the treatment is effective (randomly determined using parameters of the treatment and status). If so, set $v.s2$=Normal, $v.t$=int(v.t/te)	Set the future status to Normal. te is the efficacy of the treatment, a reduction in the length of infection. A rapid-acting treatment will therefore have a large te.

Table 4.5: The implementation in the model of simulating a treatment programme. The random nature of the selection recognises the differing willingness of patients to seek treatment. This will be affected by the severity of the symptoms and the level of publicity surrounding an outbreak (e.g. the Flu Pandemic hotline and associated publicity in 2009).

4.5.3 Firebreaks/Isolation/Containment

4.5.3.1 Description

A firebreak is a term originally applied to forestry, where a gap is created in the forest which fire is unable to pass across. Thus an uncontrolled fire is contained within a region surrounded by firebreaks and thereby isolated. Frequently these firebreaks will double as roads or may utilise natural features such as streams. The creation of a firebreak in crop farming will sometimes involve the setting of controlled fires in order to create gaps across which fire cannot pass, due to the combustible material having already been consumed.

In animal infection control terms, a firebreak often refers to a cull of healthy livestock surrounding an infected area. In the UK the authority for such a cull lies with Defra and is contained in the Animal Health Act of 1981. Although only designed for a small range of infections, the amendments of 2002 and 2003 (to include Foot & Mouth, Avian Influenza and Newcastle Disease) show that its use can be extended swiftly if required.

In human infection terms, a firebreak was employed by many areas during plagues (particularly the Black Death circa 1350) where communities would shut themselves off from outsiders in order to prevent the disease entering (and, in some notable cases, leaving). This approach had varying degrees of efficacy, as the plague was not exclusively human-borne. Another approach is to immunise all potential hosts in a geographical area surrounding the outbreak.

For the purposes of this research, a firebreak is similar to isolation or containment approaches, in that an attempt is made to contain the infection within a controlled area in which it may be left free to progress unhindered. All new cases are quickly placed within it and entry to and from it is subject to stringent regimes. Contact to and from those within the infected area (often including staff working there) is reduced so that possible routes for the infection to progress to other areas are minimised. The firebreak is thus created by breaking contacts along which the infection may pass.

4.5.3.2 Implementation

Action	**Description**
Form $V_{fb} = \{v \mid v \in V_s\}, v.s \notin \{Null, Normal\}$ $E_{fb} = \{e \mid e \in E_s\}, e.v1.s \in V_{fb} \cap e.v2.s \in V_{fb}$ $P_{fb} = \{p \mid p \in P_s\}, p.v1.s \in V_{fb} \cap p.v2.s \in V_{fb}$	Find all edges and paths attached to a vertex in a non-normal state.
$\forall e \in E_{fb}, set\ e \notin E_s$ $\forall p \in P_{fb}, set\ p \notin P_s$	Remove them from all contact with other vertices
Form $E_i = \{e \mid e.v1 \in V_{fb}, e.v2 = v_i\} \cap \{e \mid e.v1 = v_i, e.v2 \in V_{fb}\}$	Set contacts with the selected isolation location (v_i) only

Table 4.6: The implementation in the model of an isolation procedure.

4.5.4 Hand Washing/Handrubbing Regimes

4.5.4.1 Description

The demonstration of the efficacy of handwashing is usually attributed to Ignaz Semmelweis, who is known as "the father of infection control". During an appointment in obstetrics, he observed a discrepancy in the post-delivery mortality rates for physicians and medical students (13-18%) compared to that for midwife trainees and midwives (2%). He reasoned that the difference was that physicians and medical students handled corpses during autopsies, whereas midwives and their trainees did not. By instituting a programme of handwashing with a chloride of lime solution the mortality rates fell to about 2%. (Best & Neuhauser, 2004).

According to the PHLS (www.phls.org), hand-washing by health care staff before and after close contact procedures is the single most important measure for controlling and preventing the spread of hospital infection. However, compliance is very low (8.6% as reported in Tibballs, 1996, Table 2). Alcohol gel dispensers have improved compliance considerably, possibly due to the removal of the time-constraint of handwashing (estimated at 30-60 minutes per hour for 100% compliance by Hugonet et al., 2002) and their introduction being paired with a publicity campaign, both amongst medical staff and patients. (Hugonnet et al, 2002 p 1037 reported an increase in compliance in intensive care units from 38.4% to 54.5%). Ancel Meyers et al. (2006) implemented this by lowering the probability of transmissibility of the infection. As this model uses resistance rather than conductance, the parameter is increased.

4.5.4.2 Implementation

Action	**Description**
Randomly form $V_{hw} = \{v \mid v \in V_s\}, \lvert V_{hw}\rvert = \lvert V_s\rvert * k_{hw}$	Find all vertices to be complying with the hand-washing regime. k_{hw} is the proportion of vertices complying.
$\forall v \in V_{hw}$, set $v.c_i = (v.c_i) * R_i$ where $m_i \in \{M \cap \{\text{contact} - \text{borne modifiers}\}\}$	Increase their resistance to change for all infections that are contact-borne. R_i is the increased resistance (i.e. >1) for modifier i.

Table 4.7: The implementation in the model of a handwashing regime.

4.5.5 Personal Protection (Barrier Methods)

4.5.5.1 Description

As all infections have to enter the host, a barrier to entry is the most effective method of preventing infection. The most appropriate barrier method will depend upon the infection: air-borne infections may be stopped by breathing masks, contact-borne infections by the use of gloves or overalls.

There are many documented examples of success in control of infection spread using full contact isolation methods such as gowns (e.g. Harbarth et al., 2000 reported in Fleming Forum, 2004). Additionally, Boyce et al., 1994 (reported in Fleming Forum, 2004) showed the difference between two VRE outbreaks – the one where gloves were used was controlled, the other not. Ancel Meyers et al. (2006) model this by lowering the probability of transmissibility of the infection. As this model uses resistance rather than conductance, the parameter is increased.

4.5.5.2 Implementation

Action	**Description**
Randomly form $V_{bm} = \{v \mid v \in V_s\}, \lvert V_{bm}\rvert = \lvert V_s\rvert * k_{bm}$	Find all vertices to be using a barrier method. k_{bm} is the proportion of vertices complying.
$\forall v \in V_{bm}$, set $v.c_i = (v.c_i) * R_i$ where $m_i \in \{M \cap \{\text{contact/proximity} - \text{borne modifiers}\}\}$	Increase their resistance to change for all infections that are contact- or proximity-borne (depending upon the barrier method used). R_i is the increased resistance (i.e. >1) for modifier *i*.

Table 4.8: The implementation in the model of a barrier regime.

4.5.6 Inspiring New Approaches

One advantage given by simulations and models is the ability to try differing scenarios from the same initial conditions and with different (and experimental) containment/eradication techniques. This ability is not present in the "real world" as the most risk-adverse policies must be followed. In a model it does not matter if the entire world becomes infected as the simulation can be rewound and re-run. Thus differing methodologies can be investigated (provided a method of implementing them within the model can be determined) and evaluated.

4.5.6.1 Segregation of Multi-Site Services

One such possible approach is the segregation of multi-site services, removing some links by limiting the movement of staff and samples. By using a simulation, each link can be removed/reduced in turn in order to identify the most effective ones to modify. Combinations of links may then be investigated, and so on. Equally, due to the ease of implementation in a model, services may be duplicated on multiple sites in order to reduce movement and thereby connections between them. If such duplication were to be shown to provide a major enhancement to the reduction of infection propagation then such data could form a part of a business case for this service development. This analysis is similar to that of identifying communities and superhubs, which (as has already been noted) aids in the reduction of infection propagation.

4.6 Conclusion

This chapter has set out the proposed model, describing it in both mathematical and descriptive terms. The new and enhanced concepts have been described in terms of application to an infection transmission model.

The new concepts, and therefore the contribution of this part of the thesis to knowledge, are:

- The use of *information* to find short paths.
- The addition of *external paths* to model the contribution of the full population to the semi-closed environment under consideration.
- The use of *remodelling* to model the dynamic nature of social networks.
- The *strength of an edge* to assist in a realistic determination of how the model should remodel.
- The separation of *transmission* into three parts: *resistance to change, modification chance* and *context* to produce more realistic models.
- The new metric: *path length matrix* used in the calculation of some statistics.
- The implementation of *fully directed graphs* to produce more realistic models.

- The concept of a *system* to describe the known and unknown parts of the semi-closed environment.

A key part of this contribution is in producing more realistic models. It is the determination of "more realistic" that the next chapter sets out to address.

Appendix A – Scripting Language

Purpose

A scripting language was added to the research software in order to automate simple and repetitive tasks. This improved the research times and allowed such things as overnight untended runs, when required.

Three areas of the research were deemed to benefit from this approach:

- random model construction (due to the dual needs of multiple experiments and length of time taken for construction of large models)
- database model construction (due to the need for multiple experiments)
- classification (due to the need to repeat the classification during development of the systems described in chapter 5).

It is also possible to create a model by writing a text file to describe the structure. This is described in Appendix B.

Syntax

The script file is a text file (default extension: .txt) with a series of statements (in lower case) on separate lines. They are executed sequentially as no flow control (aside from the repeat mechanism described below) was deemed necessary.

The first line of the file begins “Small World Script File” and may have a suffix of “-Database”, “-Classify” or “-Random” in order to denote the type of script it is. No suffix indicates a random model construction script.

A line in the script file that commences with “*” is treated as a comment and is ignored.

The following syntax applies to this appendix:

{value 1 | value 2 | etc.}	One of these values must be selected and used.
<value>	An explicit value must be used in this position in the statement.
[n-m]	An explicit numeric value between n and m (inclusive) must be used in this position in the statement.

Model Construction

This version of the script file constructs random models in the form of one per line in the file. The format of the line is:

<Heuristic>;<Number of vertices>;<Directed>;<Add Paths>;<Probability>;<Use Paths>;<Number of models to Build>;<Calculate Statistics>;<Number of Vertices to Select (Heuristics 1 and 3) OR Number of Groups to Make (Heuristic 2)>;<Edges to Make>;<File Name for results>

and the possible values are:

Heuristic	0 – random until connected 1 – superhub 2 - grouped
Number of vertices	Integer number of vertices to place in the model
Directed	Y – model is directed N – model is not directed
Add Paths	Y – external paths are created for the model N – external paths are not created
Probability	Integer [0-100] of probability that a path exists from a vertex.

Use Paths	Y – external paths are used in assessing connectivity N – external paths are not used
Number of models to Build	Integer number of models to build
Calculate Statistics	Y – statistics are calculated for the model (and saved in the log file) N – statistics are not calculated
Number of Vertices to Select (Heuristics 1 and 3) OR Number of Groups to Make (Heuristic 2)	Integer parameter for the heuristic. Number of Vertices to Select is a percentage of the total, not an absolute number.
Edges to Make	Number of edges to make on each pass for heuristic 2.
File Name for results	Name of text file (*.txt) to be created to receive the results. If the file already exists, it is overwritten. This may be blank.

Database

This version of the script file creates one model as per the commands within it (but see "times", below).

It should be noted that the default is for the messages to be "full", there is no maximum run time, the stopping criteria is No non-normal vertices, excelCR is "on" and pathsinstats is "off".

classify	Runs a full classification list.
dfr3file <name>	The name of the file to receive the DFR3 output. If this is the first run, the file is opened, truncated and the current date and time is written to it.
edges <%age probability>	Adds edges to the model, at the specified %age. These paths are only between named individuals, not wards/departments/services.
excelCR {on | off}	If "on", carriage returns in the summary log file are replaced with tabs.
file <name>	The name of the file (run log) to receive the output from this run. The file is opened, truncated and the current date and time is written to it.

handwash <infection number (digit)> <%age increase> <%age probability>	After each time step, each non-ward vertex with the status indicated by the (single-digit) infection number has this probability of being part of a handwashing protocol, thereby increasing the resistance to change by the %age given.
heading <text>	The text is added to the summary log file on a separate line, if open.
infect <infection number (digit)> <name of vertex>	Sets the specified vertex to a status indicated by the (single-digit) infection number. If either does not exist, the operation fails and reports an error. Additionally, the status to change to is set to Normal and a random time (as determined by the infection's parameters) is set.
isolate <infection number (digit)> <%age probability>	After each time step, each non-ward vertex with the status indicated by the (single-digit) infection number has this probability of being isolated, removing all edges from and to it from the model.
max time <value>	Sets the maximum time that the model can run for.
messages {off | full | infected | ward}	Specifies the type of messages to place into the run log: "off" – none. "full" – all (whether vertex will change, together with parameters showing why), plus a list of current non-normal vertices at each time step. "infected" - only non-normal vertices are listed at each time step. "ward" – only non-normal wards are listed at each time step.
paths <%age probability>	Adds paths to the model, at the specified %age. Sets the "use paths in statistics" flag to true.
pathsinstats {on | off}	If "on", paths are used in the calculation of statistics.
reset	Resets the model, clearing all vertex flags and setting all Vertices to Normal status with no pending status changes. The run log file is closed if open and the time is set to zero.
run model	Runs the model once.

set <infection number (digit)> <parameter (digit)> <value>	Sets the specified parameter of the specified infection to the specified value. Parameter values: 0 - StatusEdgeModification (int) 1 - StatusLifespanMean (int) 2 - StatusLifespanSD (int) 3 - StatusLifespanTail (int) 4 - StatusVertexImmunityMean (int) 5 - StatusVertexImmunitySD (int) 6 - StatusVertexResistanceMean (float) 7 - StatusVertexResistanceSD (float) 8 - StatusVertexReversion (float) Values calculated from these for edges and vertices are re-calculated.
statistics	Calculates the model's statistics and outputs them to the file specified in the sumfile command, if open.
stopping <value>	Sets the stopping criteria as: 0 - No non-normal vertices 1 - No future changes 2 - No non-normal vertices or future changes 3 – Specified % of non-normal vertices 4 - Pre-defined time elapsed
sumfile <name>	The name of the file to receive the summary output (summary log). If this is the first run, the file is opened, truncated and the current date and time is written to it.
timenotclass {on | off}	If "on", the length of the outbreak (time) is recorded instead of the classification.
times <value>	The script is run the specified number of times
treat <infection number (digit)> <%age probability>	After each time step, each non-ward vertex with the status indicated by the (single-digit) infection number has this probability of being treated, setting the status to Normal.
vaccinate <infection number (digit)> <%age probability>	After each time step, each non-ward vertex has this probability of being vaccinated, giving infinite immunity to the status indicated by the (single-digit) infection number.

Classification

This version of the script classifies all outbreaks in the database, according to the parameters set in the script.

It should be noted that the default values are for GoF to be used for Goodness of Fit, with p=3.

addfile {on | off}	If "on" then an existing log file is added to. If "off" then this file is truncated first.
classify {staff | patients}	Starts the classification of recorded outbreaks for either staff or patients.
dfr {on | off}	If "on", DFR3 is used to calculate Goodness of Fit.
e1r {on | off}	If "off" then E1 and R are excluded from the Goodness of Fit measure (GoF and MRSSD only).
gof <text>	If the log file is open, the Goodness of Fit is calculated and the line "Classification <text> Goodness of fit= " is added to the log file, followed by the Goodness of Fit value.
heading <text>	The text is added to the log file on a separate line, if open.
mrssd {on | off}	If "on", MRSSD is used to calculate Goodness of Fit.
openfile	If the log file (run.log) is not open, it is opened.
p [1-6]	The p value (for GoF) is set to the required integer.
recurrence {on | off}	Determines whether recurrence is checked for or not.
startpoints {on | off}	If "on", the start points of the outbreaks (name of initial infected vertex) are calculated and output to the log file.

Random

This version of the script is a hybrid of the model construction and database forms. The format is one line for a model to be constructed (always 1 – the parameter for number of models to build is ignored), followed by commands from

the database set. Once the random model has been built, the infections are loaded from the database.

There is one additional command, as below:

wards <value>	This number of vertices are randomly assigned to be wards. These are also assigned random severity values, in the ratio 34:18:6, as observed in the database.

Appendix B – Saved Model Structure

Models constructed by the software can be saved for re-loading later. As these are saved as text files, these can be manipulated in order to effect minor modifications, or even written completely by hand in order to reflect a particular structure. The first header is read by the import routine, the other headers are ignored (as are the blank lines) so could contain any text that is helpful to the author – the only requirement is that a line is present. The text listed here for the headers is that which is generated by the software.

The format of these files is as follows:

Header	"Small World Model Version ", followed by “1” or “2” (Version 2 includes names for vertices).
Overall parameters 1 header	"Statuses, Vertices, Paths, Edges"
Statuses	<number of statuses>;<first status name (usually “Normal”)>;<second name>;etc…
Number of vertices in the model	<integer>
Number of external paths in the model	<integer>
Number of edges in the model	<integer>
Overall parameters 2 header	“Normal Status, Heuristic, Connected, Directed, Calc Stats, VerticesToCheck%/GroupsToMake, EdgesToMake”
Normal Status	<integer>, the index of the Normal status (usually 0).
Heuristic	<integer>, the heuristic used to construct the model (see Appendix A).
Connected	“Yes” – the model is connected “No” – the model is not connected
Directed	“Yes” – the model is directed “No” – the model is not directed
Calc Stats	“Yes” – the statistics are to be calculated on load “No” – the statistics are not to be calculated
VerticesToCheck%/GroupsToMake	<integer> - additional parameter for heuristic (see Appendix A).
EdgesToMake	The maximum number of edges to create after loading the model until it is connected or this number is reached.

Blank line	
Header for vertices	"Vertices"

For each vertex:

Header	"Immunity, Context, Ordinal, Status, StatusTo, TimeTo, Resistance, Reversion, Name"
Immunity	<immunity for first status – [0-100]>; <immunity for second status [0-100]>;etc…
Context	<integer>, the context for this vertex
Ordinal	<long integer>, the unique ordinal for this vertex.
Status	<integer>, the current status of this vertex.
StatusTo	<integer>, the status this vertex will change to.
TimeTo	<integer>, the number of time units in the future at which this status change will happen. (-1 indicates no future change – the case when the vertex is in normal status).
Resistance	<resistance to change for first status – [0-100]>; <resistance to change for second status [0-100]>;etc…
Reversion	<reversion chance for first status – [0-100]>; <reversion chance for second status [0-100]>;etc…
Name – *version 2 only, otherwise not present*	<textual name of this vertex>

Blank line	
Header for edges	"Edges"

For each edge:

Header	"Modification, Ordinal, Strength, Vertex1, Vertex2"
Modification chance	<modification chance for first status – [0-100]>; <modification chance for second status [0-100]>;etc…
Ordinal	<long integer>, the unique ordinal for this edge.
Strength	<integer [0-100]>, the strength of this edge.
Vertex1	<long integer>, the ordinal for the starting vertex for this edge.
Vertex2	<long integer>, the ordinal for the terminating vertex for this edge.

Blank line	
Header for external paths	"Paths"

For each external path:

Header	"Modification, Ordinal, Vertex1, Vertex2, Length"
Modification chance	<modification chance for first status – [0-100]>; <modification chance for second status [0-100]>;etc…
Ordinal	<long integer>, the unique ordinal for this external path.
Vertex1	<long integer>, the ordinal for the starting vertex for this external path.
Vertex2	<long integer>, the ordinal for the terminating vertex for this external path.
Length	<integer>, the length of this external path.

Blank line	

Appendix C – Random Models

Creation of Models

The theoretical models as described in chapter 4 are created in software (Microsoft Visual C++ version 6 was used) with classes representing the vertices, edges and paths. Additionally there is a database which contains the details and parameters for the infections under investigation. For models based upon the Trust, these parameters are also contained within the database.

A purely random model is created by specifying:

- the number of vertices to be used
- whether the graph is to be directed or undirected
- the heuristic to use to connect them (together with any additional parameters for the selected heuristic)
- whether the model is to include external paths (and, if so, the percentage probability of one being added)
- whether to use the paths in determining connectivity
- the number of models to build
- a file name for the results of the build

Once built, a model may be run by providing some initial conditions, either randomly altering a set number of vertices or one specific vertex to a specified status.

The model can then be run, after specifying stopping criteria (one of):

- Run until all vertices are normal
- Run until no vertices will change
- Run until all vertices are normal & none will change
- Run until % of non-normal vertices <= a specified value
- Run until Time= a specified value

Additionally, a maximum run time and a file name for results can be specified.

A model based upon the Trust is created using the database, which specifies the vertices and edges derived from the questionnaire described in chapter 6. It can then be run in the same way as a random model.

Visualisation

The most obvious visualisation is to represent vertices as points and edges and external paths as lines joining them. Due to the number of vertices involved in a realistic model, such visualisation is of limited use, as Figure C.1 demonstrates.

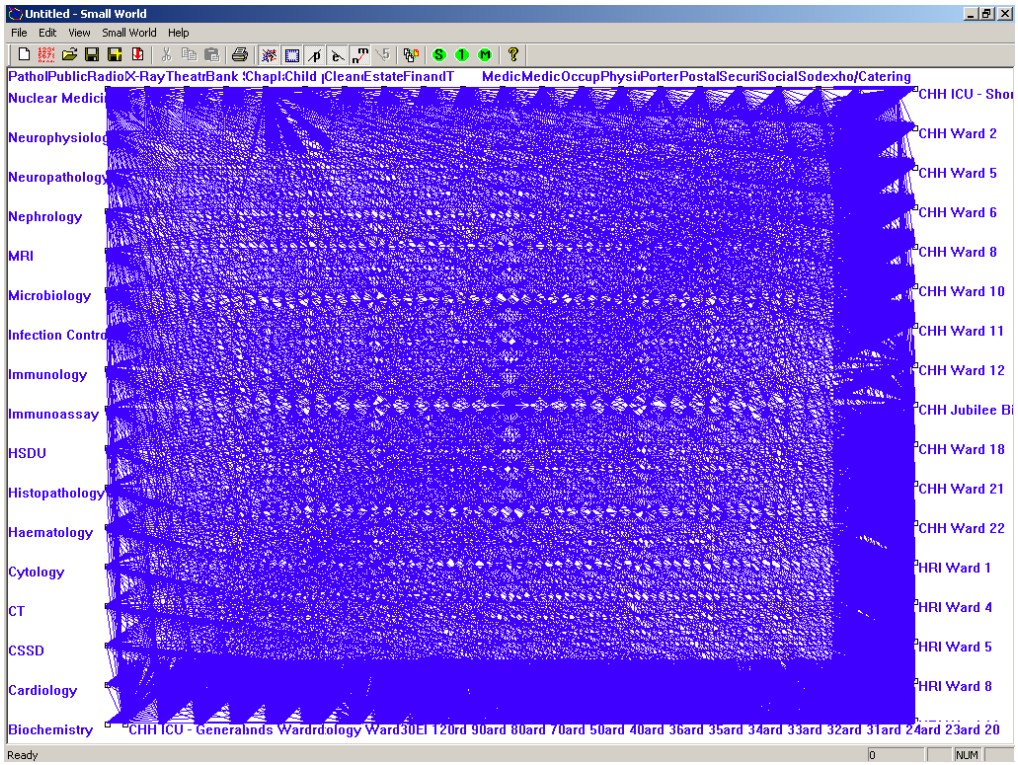


Figure C.1: Visualisation of all three hospital sites, showing linkages between wards and services – 76 vertices and 2663 edges.

However, on a smaller scale, this can be useful in demonstrating the build-up of the model (Figures C.2 and C.3) and in following the progress of an outbreak (Figure C.4).

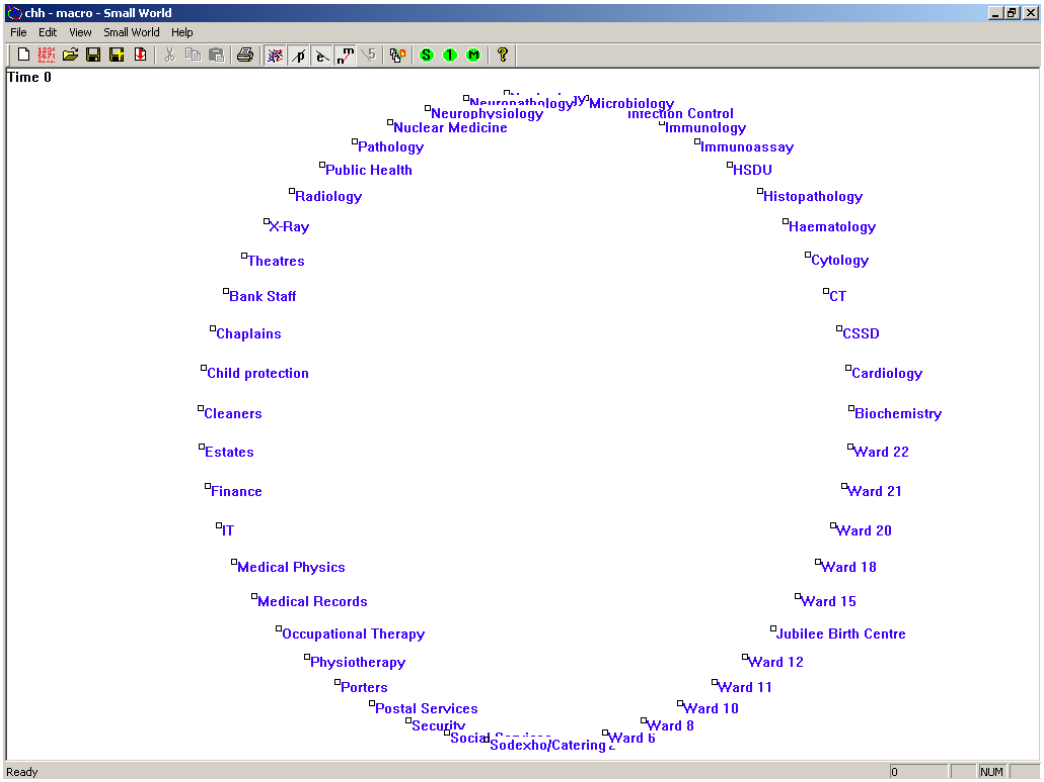


Figure C.2a: Castle Hill Hospital and services, with the vertices arranged in a circular pattern.

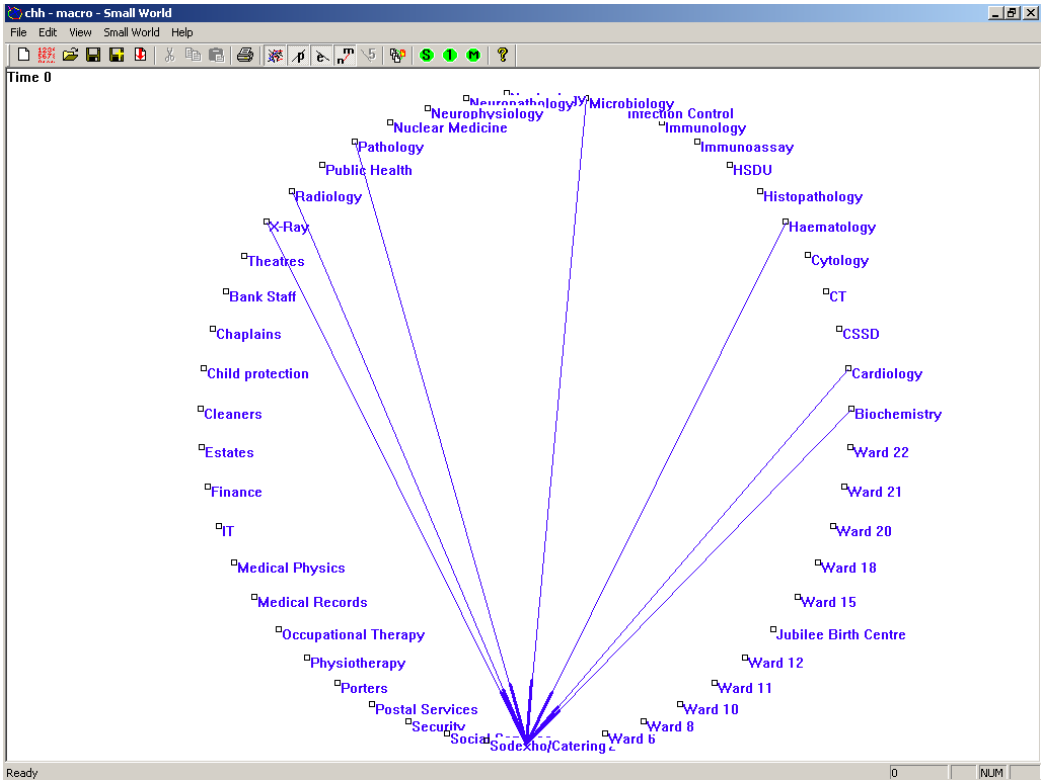


Figure C.2b: The most significant connections (i.e. daily contact) for one ward, with these edges overlaid on the circle as lines. The directionality of the edge is indicated by a thicker line indicating the origin. For simplicity, shared facilities and other shared staff are omitted.

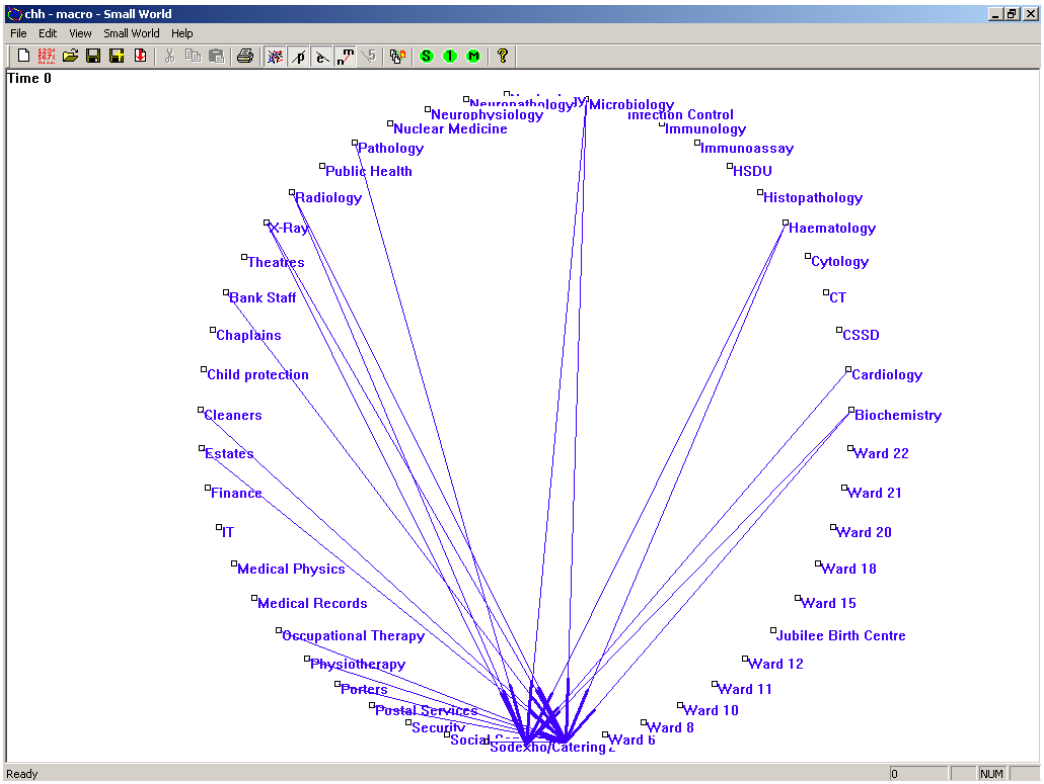


Figure C.2c: Two wards, mapped as above

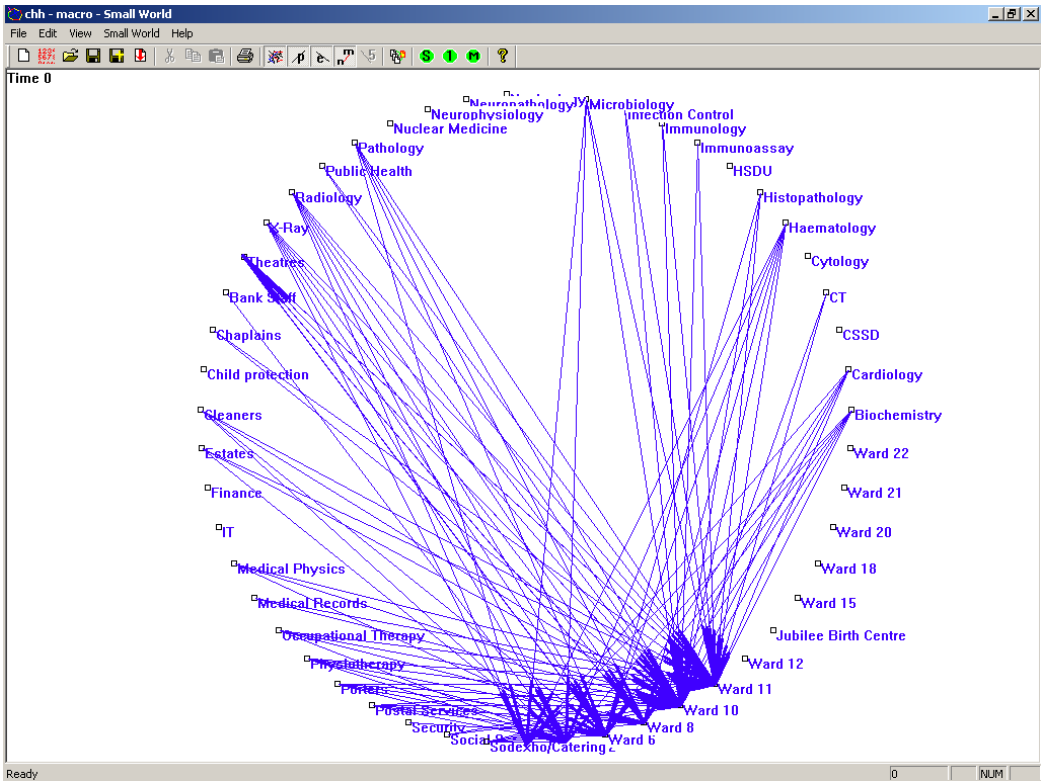


Figure C.2d: Six wards, as above

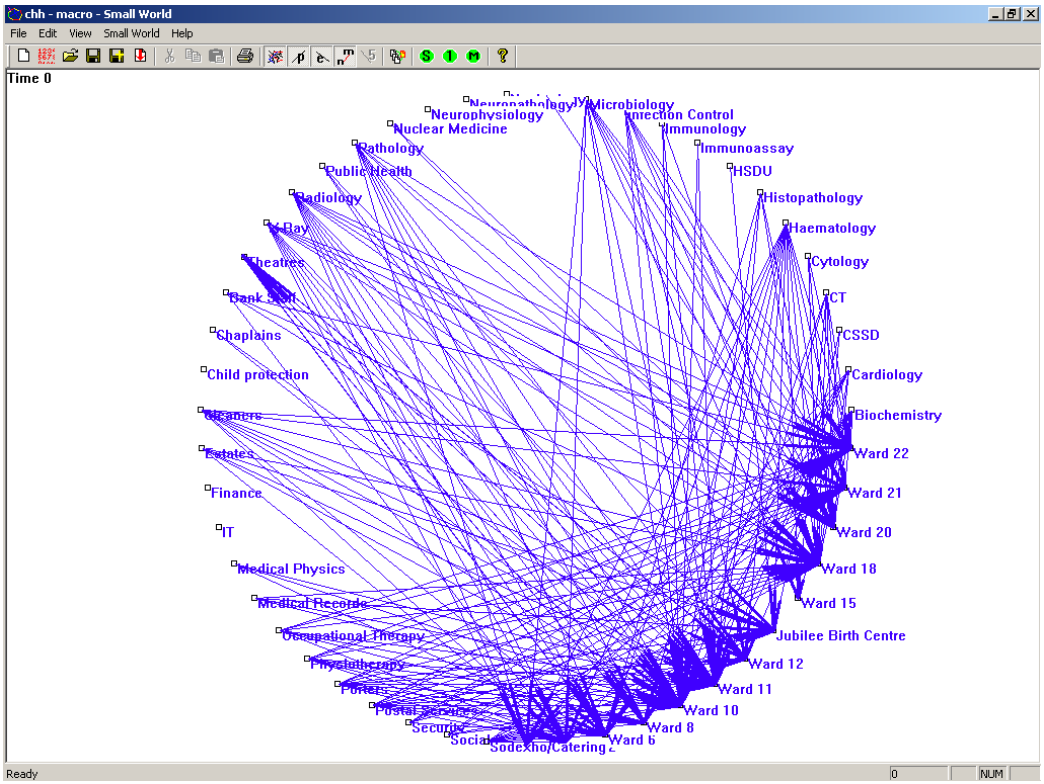


Figure C.2e: All 13 wards.

Some information may be drawn from Figure C.2e, in that it demonstrates the low level of connectivity of services such as Finance and IT, yet the high level of connectivity of Cleaners and Cardiology.

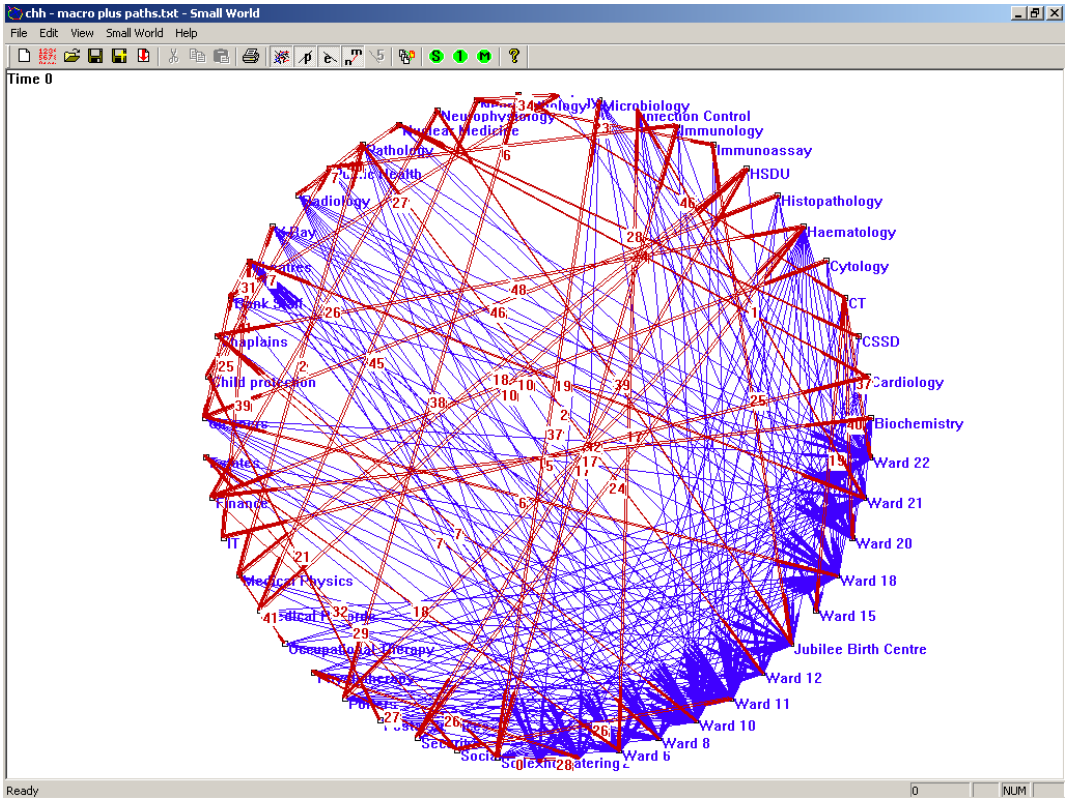


Figure C.3: The addition of randomly generated external paths onto the model shown in Figure C.2e.

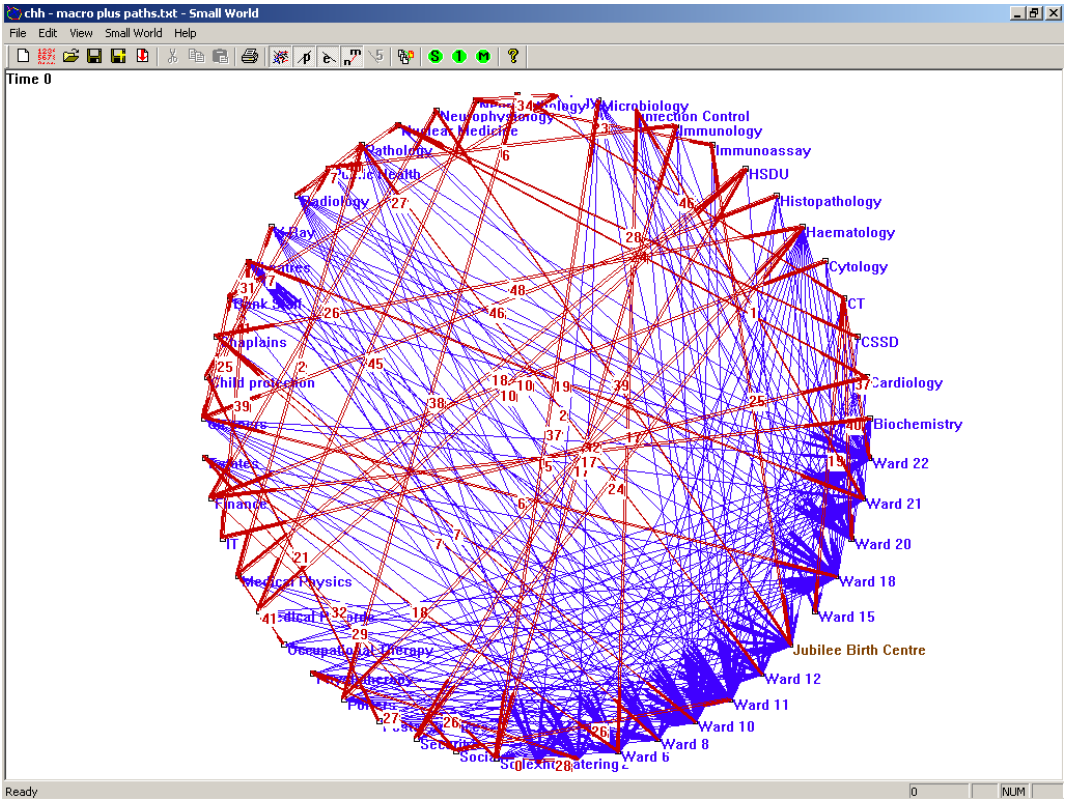


Figure C.4a: An infection is introduced into the model shown in Figure C.3, at the Jubilee Birth Centre (marked in brown).

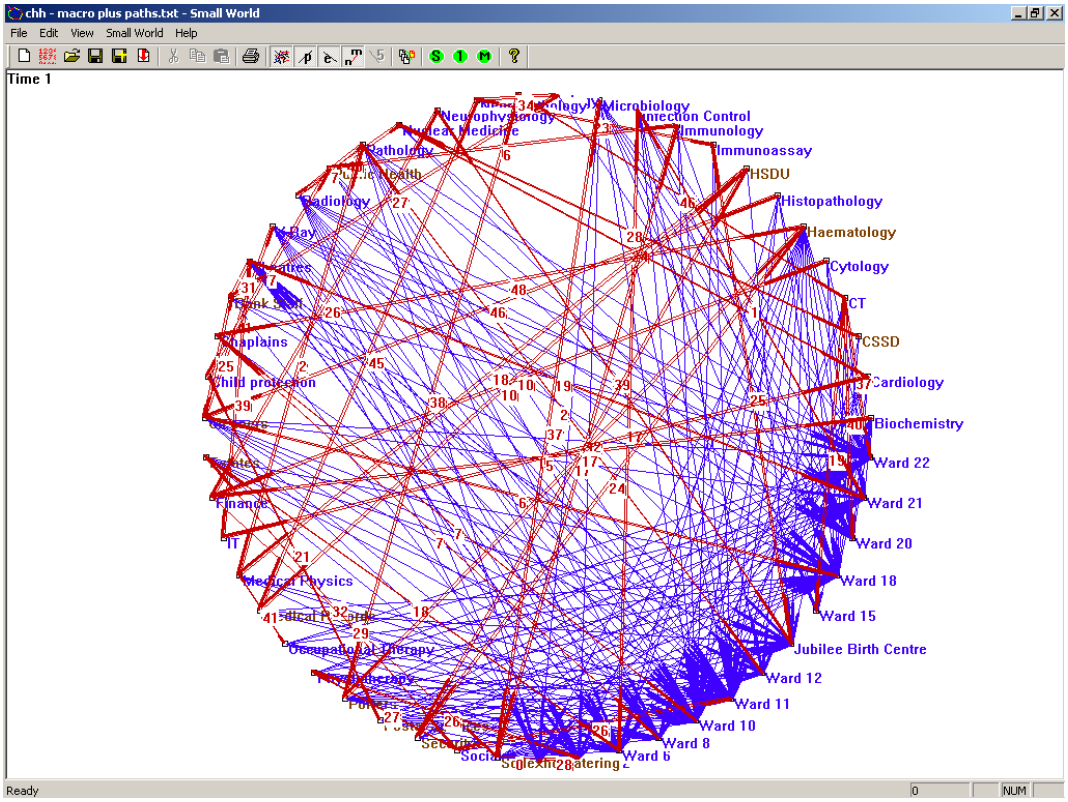


Figure C.4b: The infection progresses to HSDU, Haematology and CSSD.

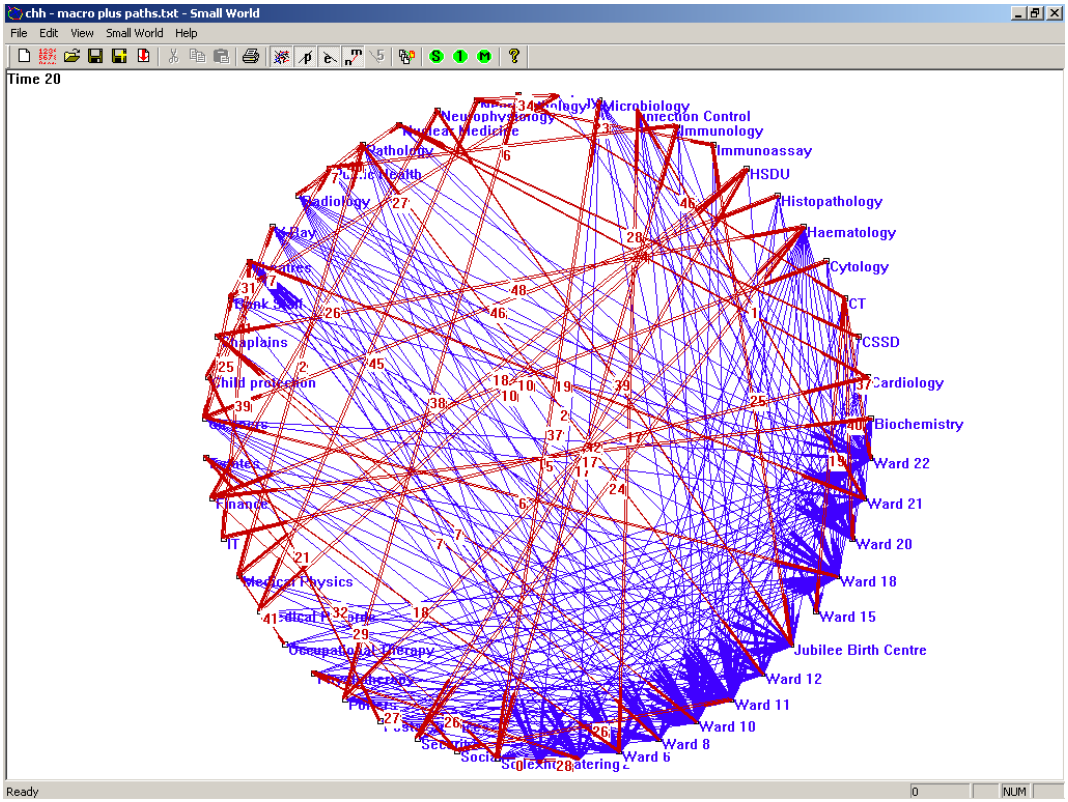


Figure C.4c: The infection dies out

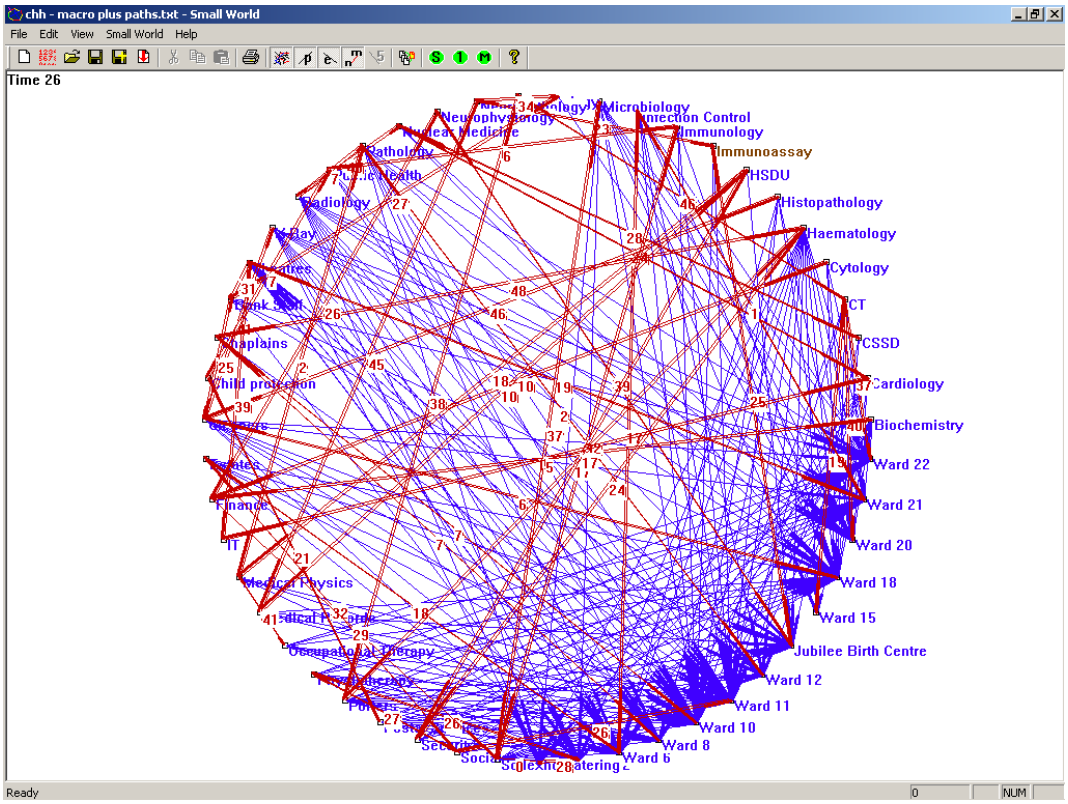


Figure C.4d: The infection re-appears due to the action of an external path

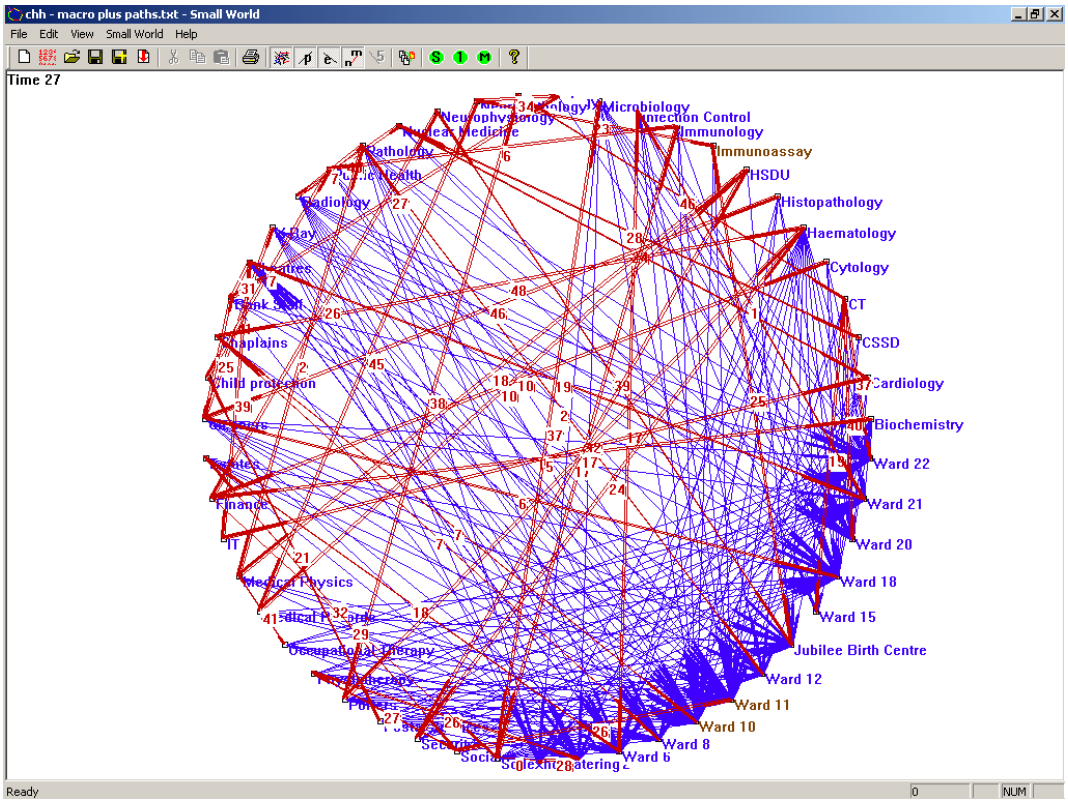


Figure C.4e: The infection outbreak continues to progress.

Outputs and Metrics

Statistics and metrics produced by the implementation of the model are (see Chapter 4 for descriptions):

- Heuristic
- Number of times heuristic applied
- Number of vertices
- Number of edges
- Number of paths
- % chance of paths
- Is connected (Y/N)
- In/OutDegree
- InDegree
- OutDegree
- Characteristic Path Length
- Clustering Co-efficient (edges only)
- Vertex Connectivity (<list of vertices removed>)
- Min InDegree
- Edge Connectivity (<list of edges removed>)

The files that are produced by the creation/running of models contain (not all in the same file):

- Time run started
- Time run finished
- Heading defined by script
- The eight DFR3 values, plus a sum
- Classification
- Goodness of Fit (GoF plus p, MRSSD and DRF3)
- Run time
- Future vertex changes (including parameters)
- Vertices that will not change (including parameters)
- Vertices with non-normal status
- Total number of vertices with non-normal status

- Peak number of wards infected
- Number of wards in the system
- Stopping condition
- Date outbreak from
- Date outbreak to
- Classification system
- Classification parameter values, with minimum, maximum and mode
- Outbreak patient days, ward days and average
- Start points for outbreaks
- Vertex resistance values
- Vertex reversion values
- Edge modification values

Appendix D – The Software

Introduction

Elements of the software have been described and outputs from it utilised throughout the main thesis. This appendix describes elements not covered elsewhere.

Overview

The model was constructed using C++ under Microsoft Visual Studio version 6 (together with the Microsoft Foundation Classes and thus implementing a Model-View-Controller paradigm). Although designed to create and manipulate directed graphs, the software can also handle undirected graphs by the simple method of placing two edges (one in each direction) for this case.

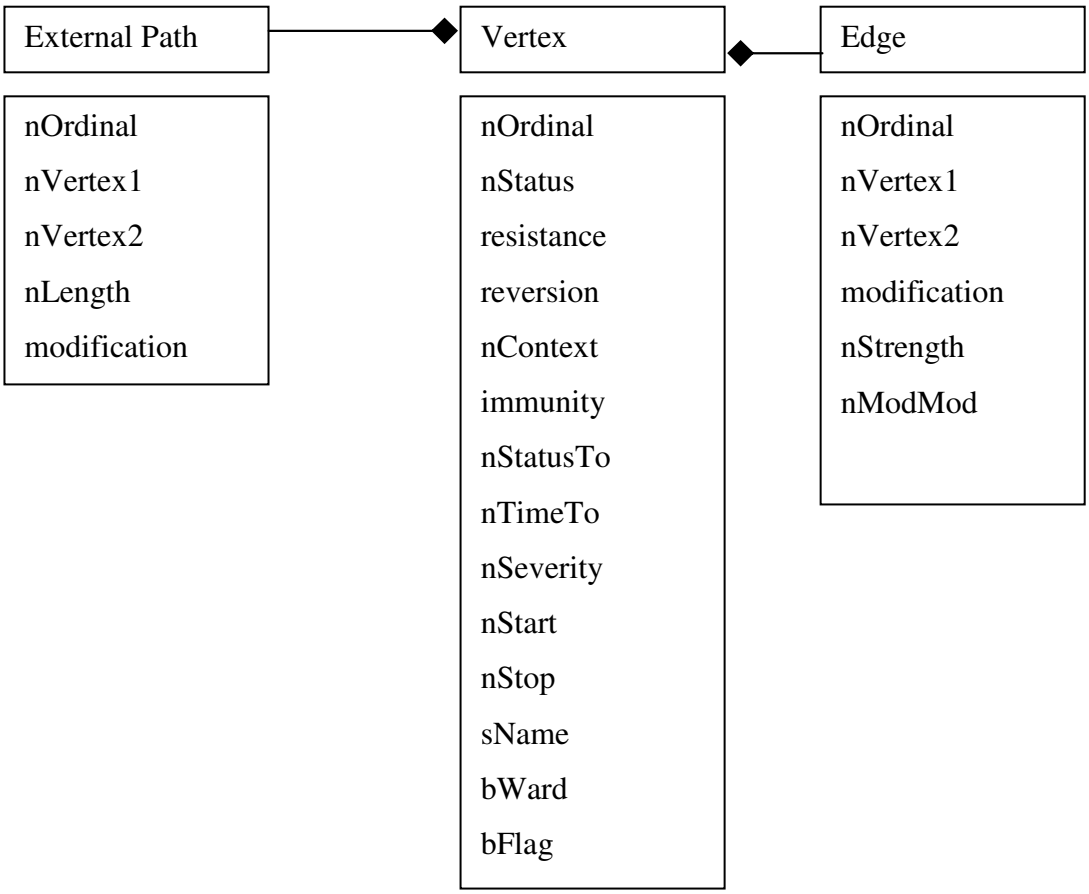


Figure D.1: High level software design showing classes, members and compositions.

The software is capable of creating random models and calculating the following statistics:

- Number of vertices
- Number of edges
- Number of paths
- Is connected
- In/Outdegree
- InDegree
- OutDegree
- Characteristic Path Length
- Clustering Co-efficient (edges only)
- Vertex Connectivity

Some of the algorithms for calculating these statistics have come from the literature (especially where a true mathematical method exists, e.g. InDegree). Others are my own, derived from textual description in the literature (e.g. characteristic path length).

Outputs

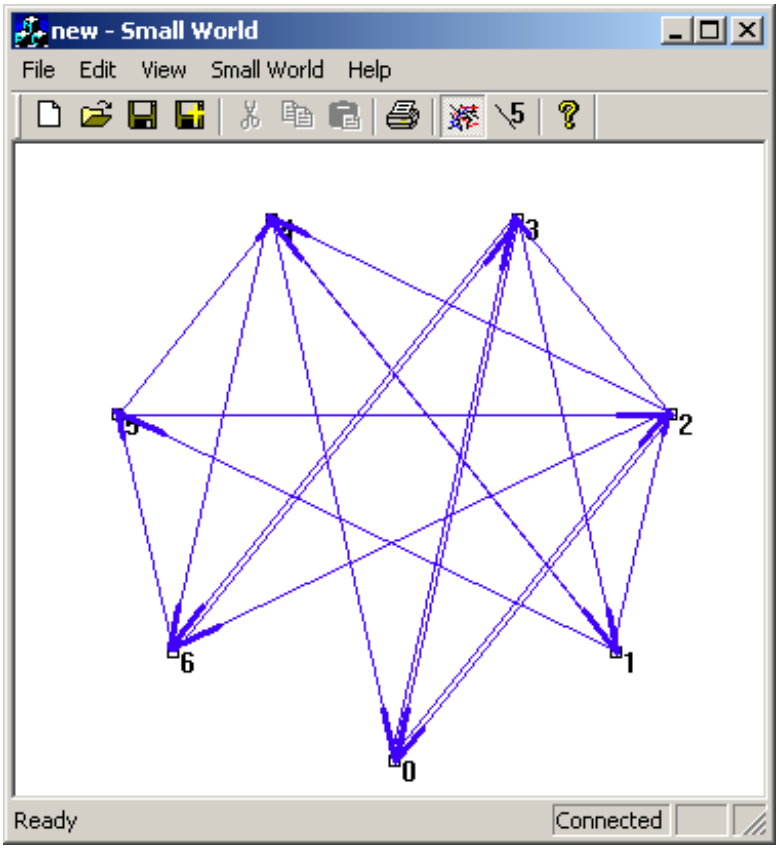


Figure D.2: A graphic representation of a 7-vertex directed graph. Thickened ends of lines show the originating vertex for the edges, that is the direction of the edge. For example, the edge between vertices 1 and 3 has a thickened end at vertex 3, showing that the direction of the edge is from vertex 3 to vertex 1.

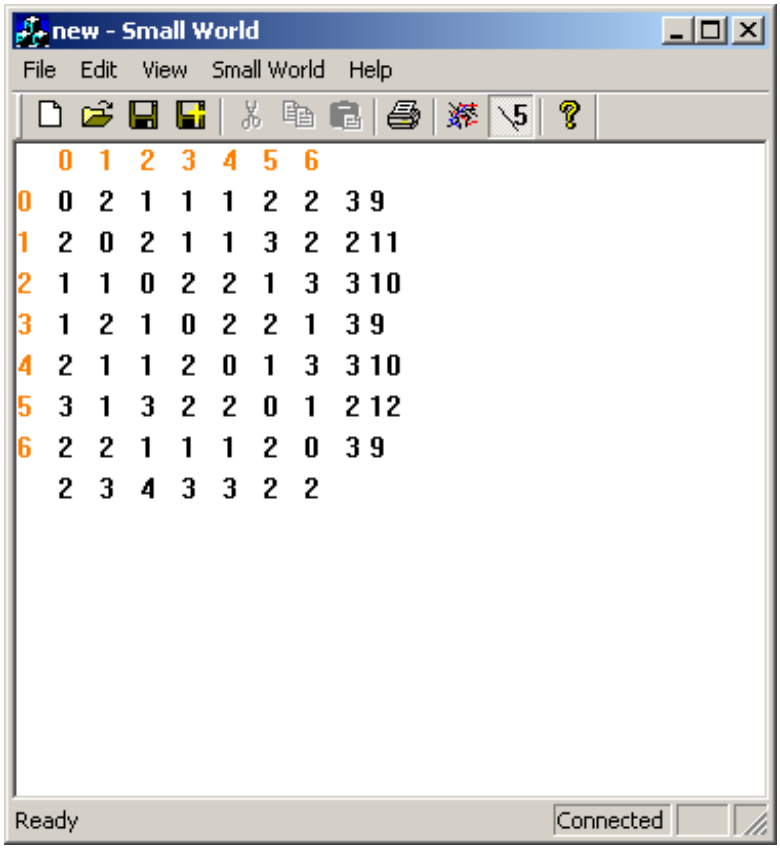


Figure D.3: The Path Length Matrix for the Graph in Figure D.2. This is read "from <row> to <column>", e.g. The path from vertex 1 to vertex 3 has length 2. The first additional column gives the vertex's outdegree and the additional row gives the indegree. The second additional column gives the sum of the lengths of paths originating from the vertex, which is used in calculating the average path length, itself used in calculating the characteristic path length.

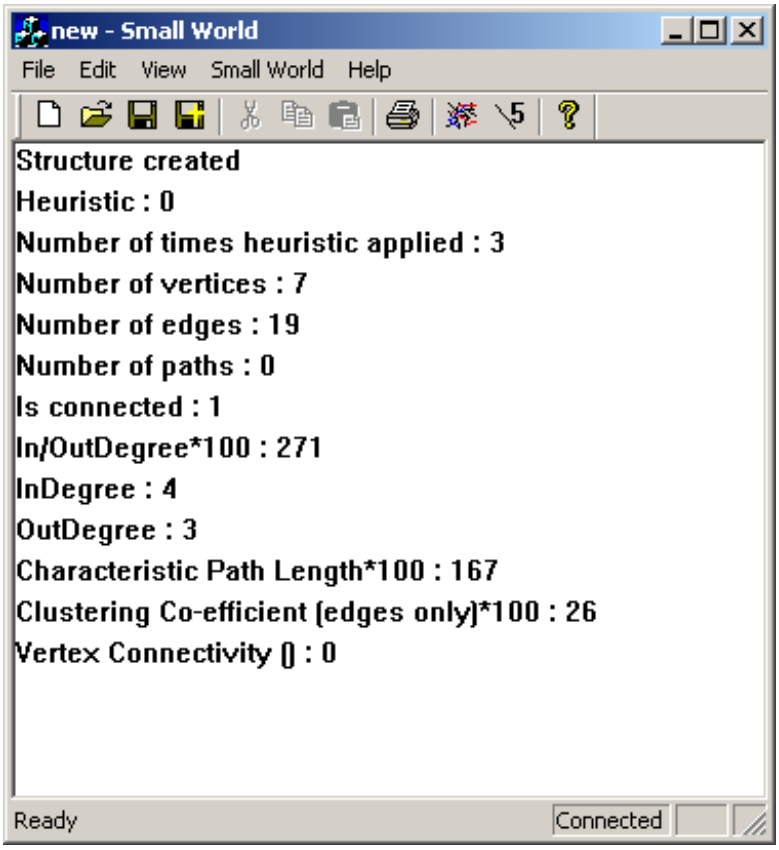


Figure D.4: The Statistics produced for the Graph in Figure D.2

Software Performance

The software as currently developed is not optimised. This is because it is a research tool rather than an operational one: therefore there exist within the software several options which would not be required in the production version. For example, the various types of classification comparison (See Chapter 5) are all implemented. Each time a classification is compared, the required classification type must be checked for, along with any parameters that this requires (e.g. p in GoF (5.7.1)).

As the software is a research tool, the algorithms are represented in code in their most explicit form. This is so that the code may be easily modified should the algorithm be enhanced in some way. Thus each step in the algorithm is coded separately and optimisations that may be gained from step combination are eschewed.

Furthermore, speed of coding was more desirable during development of the thesis than speed of execution therefore fully optimised algorithms, for example for sorting, were not sought. This was because it was the ideas that were being tested, not the code.

Simplified Complexity Analysis for Production Software

There are many strands to the software – this analysis examines one only, that of the production of a model based upon the questionnaire results, the running of an outbreak simulation upon it and the comparison with real results via the DFR3 method (5.7.4) to determine how realistic the model is. This has been selected as the most likely use of a production version of the software.

Notation:
Ni – Number of infections
Nw – Number of wards
Ns – Number of staff
Nv – Number of services
Ne – Number of edges
Nr – Number of real models used for DRF3

The basic model from the questionnaire completes in
2O(Ni)+O(Nw)+O(Ns)+2O(Nv).O(Ns)+ 6O(Nw).
The connectedness check is omitted, as once the questionnaire-based model has been shown to be connected, the computation time in establishing it again is unnecessary.

The outbreak simulation completes in O(Nw)(3+O(Ne))(O(Nw)+O(Nv)+O(Ns)).

The classification completes in 3O(Nw).

The DFR3 comparison completes in O(Nr).

The major factor in this analysis can therefore be seen to be N=Nw+Ns+Nv, in other words the number of vertices in the model (as Ne can be viewed as a

function of N). The simplified complexity analysis therefore yields the algorithm to complete in $O(N^3)$, with the dominant term being from the outbreak simulation.

Appendix E – Data Summarised in Thesis

Data produced for this thesis is usually presented in summarised form. There are some elements that are onerous to read, but are referred to in some detail. This data is summarised in the main thesis, but presented in full here.

Section 5.4.1 Staff data - Recurrence Not Checked For

Outbreak from 10/9/2004 to 4/10/2004. Classification 26665380

Outbreak from 6/10/2004 to 11/10/2004. Classification 11124900

Outbreak from 22/11/2004 to 7/12/2004. Classification 24456500

Outbreak from 16/12/2004 to 28/12/2004. Classification 23345500

Outbreak from 23/1/2005 to 1/2/2005. Classification 23344800

Outbreak from 11/3/2005 to 20/3/2005. Classification 24443620

Outbreak from 27/3/2005 to 14/4/2005. Classification 11158900

Outbreak from 3/5/2005 to 7/5/2005. Classification 01120900

Outbreak from 1/6/2005 to 12/6/2005. Classification 11144900

Outbreak from 9/7/2005 to 13/7/2005. Classification 11121900

Outbreak from 27/9/2005 to 1/10/2005. Classification 01120900

Outbreak from 3/10/2005 to 13/10/2005. Classification 24442500

Outbreak from 14/10/2005 to 20/10/2005. Classification 11133900

Outbreak from 26/11/2005 to 5/12/2005. Classification 11143900

Outbreak from 23/12/2005 to 27/12/2005. Classification 01120900

Outbreak from 30/12/2005 to 15/1/2006. Classification 23355400

Outbreak from 24/2/2006 to 21/3/2006. Classification 25466240

Outbreak from 28/3/2006 to 9/4/2006. Classification 24444700

Outbreak from 10/4/2006 to 17/4/2006. Classification 11132900

Outbreak from 19/4/2006 to 5/5/2006. Classification 23356600

Outbreak from 13/5/2006 to 9/6/2006. Classification 24467320

Outbreak from 14/6/2006 to 18/6/2006. Classification 01120900

Outbreak from 20/6/2006 to 24/6/2006. Classification 01120900

Outbreak from 11/7/2006 to 22/7/2006. Classification 11144900

Outbreak from 19/9/2006 to 5/10/2006. Classification 23353420

Outbreak from 12/10/2006 to 17/10/2006. Classification 23321720

Outbreak from 18/10/2006 to 22/10/2006. Classification 01120920

Outbreak from 11/11/2006 to 25/12/2006. Classification 25575160

Outbreak from 27/12/2006 to 1/1/2007. Classification 11121900

Outbreak from 3/1/2007 to 22/1/2007. Classification 24464460

Outbreak from 25/1/2007 to 29/1/2007. Classification 01120900

Outbreak from 30/1/2007 to 7/3/2007. Classification 25474120

Outbreak from 17/3/2007 to 29/3/2007. Classification 23342500

Outbreak from 19/4/2007 to 25/4/2007. Classification 11132900

Outbreak from 26/7/2007 to 2/8/2007. Classification 11132900

Outbreak from 12/10/2007 to 21/10/2007. Classification 11143920

Outbreak from 23/10/2007 to 30/10/2007. Classification 23332560

Outbreak from 1/11/2007 to 5/11/2007. Classification 01120900

Section 5.4.2 Staff data - Recurrence Checked For

Outbreak from 10/9/2004 to 11/10/2004. Classification 26675282

Outbreak from 22/11/2004 to 28/12/2004. Classification 25476102

Outbreak from 23/1/2005 to 1/2/2005. Classification 23344801

Outbreak from 11/3/2005 to 14/4/2005. Classification 24476222

Outbreak from 3/5/2005 to 7/5/2005. Classification 01120901

Outbreak from 1/6/2005 to 12/6/2005. Classification 11144901

Outbreak from 9/7/2005 to 13/7/2005. Classification 11121901

Outbreak from 27/9/2005 to 20/10/2005. Classification 24463302

Outbreak from 26/11/2005 to 5/12/2005. Classification 11143901

Outbreak from 23/12/2005 to 15/1/2006. Classification 24363202

Outbreak from 24/2/2006 to 24/6/2006. Classification 26497042

Outbreak from 11/7/2006 to 22/7/2006. Classification 11144901

Outbreak from 19/9/2006 to 22/10/2006. Classification 24372122

Outbreak from 11/11/2006 to 29/3/2007. Classification 26597062

Outbreak from 19/4/2007 to 25/4/2007. Classification 11132901

Outbreak from 26/7/2007 to 2/8/2007. Classification 11132901

Outbreak from 12/10/2007 to 5/11/2007. Classification 24363262

Section 5.4.3 Patient data - Recurrence Not Checked For

Outbreak from 28/8/2004 to 1/9/2004. Classification 01120900

Outbreak from 8/9/2004 to 2/10/2004. Classification 26667390

Outbreak from 4/10/2004 to 13/10/2004. Classification 11146900

Outbreak from 18/11/2004 to 30/12/2004. Classification 24476200

Outbreak from 9/1/2005 to 17/1/2005. Classification 11138900

Outbreak from 20/1/2005 to 28/1/2005. Classification 23331600

Outbreak from 7/2/2005 to 16/2/2005. Classification 01140900

Outbreak from 7/3/2005 to 17/5/2005. Classification 24487220

Outbreak from 3/6/2005 to 8/6/2005. Classification 11123900

Outbreak from 15/6/2005 to 21/6/2005. Classification 11137900

Outbreak from 9/7/2005 to 11/7/2005. Classification 11112900

Outbreak from 24/9/2005 to 17/10/2005. Classification 24467400

Outbreak from 29/11/2005 to 3/12/2005. Classification 11122900

Outbreak from 31/12/2005 to 12/1/2006. Classification 23345500

Outbreak from 14/2/2006 to 18/2/2006. Classification 01120920

Outbreak from 20/2/2006 to 11/4/2006. Classification 25587150

Outbreak from 13/4/2006 to 17/4/2006. Classification 01120900

Outbreak from 19/4/2006 to 29/4/2006. Classification 23348800

Outbreak from 9/5/2006 to 7/6/2006. Classification 25478220

Outbreak from 19/6/2006 to 3/7/2006. Classification 11158900

Outbreak from 13/7/2006 to 21/7/2006. Classification 11134900

Outbreak from 23/9/2006 to 3/10/2006. Classification 11148900

Outbreak from 10/11/2006 to 25/12/2006. Classification 25478160

Outbreak from 3/1/2007 to 24/1/2007. Classification 25567360

Outbreak from 27/1/2007 to 5/2/2007. Classification 11148900

Outbreak from 9/2/2007 to 14/3/2007. Classification 24478320

Outbreak from 16/3/2007 to 4/4/2007. Classification 24468400

Outbreak from 7/4/2007 to 21/4/2007. Classification 11158900

Outbreak from 27/7/2007 to 6/8/2007. Classification 11147900

Outbreak from 12/10/2007 to 17/10/2007. Classification 11123920

Outbreak from 25/10/2007 to 1/11/2007. Classification 11136900

Section 5.4.4 Patient data - Recurrence Checked For

Outbreak from 28/8/2004 to 13/10/2004. Classification 26677192

Outbreak from 18/11/2004 to 16/2/2005. Classification 25485002

Outbreak from 7/3/2005 to 17/5/2005. Classification 24487221

Outbreak from 3/6/2005 to 21/6/2005. Classification 23156202

Outbreak from 9/7/2005 to 11/7/2005. Classification 11112901

Outbreak from 24/9/2005 to 17/10/2005. Classification 24467401

Outbreak from 29/11/2005 to 3/12/2005. Classification 11122901

Outbreak from 31/12/2005 to 12/1/2006. Classification 23345501

Outbreak from 14/2/2006 to 21/7/2006. Classification 26598052

Outbreak from 23/9/2006 to 3/10/2006. Classification 11148901

Outbreak from 10/11/2006 to 21/4/2007. Classification 26598062

Outbreak from 27/7/2007 to 6/8/2007. Classification 11147901

Outbreak from 12/10/2007 to 1/11/2007. Classification 23165222

Section 5.7.4 Measure 4: DFR3

Staff data, recurrence not checked for:

Revised:

Outbreak from 24/2/2006 to 21/3/2006. Classification 25467340 (was 25466240)

Outbreak from 27/12/2006 to 29/1/2007. Classification 25573160 (was 3 outbreaks)

New:

Outbreak from 20/11/2007 to 23/12/2007. Classification 25476260

Outbreak from 1/1/2008 to 24/2/2008. Classification 25586120

Outbreak from 4/3/2008 to 31/3/2008. Classification 25465220

Outbreak from 19/4/2008 to 4/5/2008. Classification 24456460

Outbreak from 8/5/2008 to 20/5/2008. Classification 23344500

Outbreak from 21/5/2008 to 27/5/2008. Classification 23331700

Outbreak from 28/5/2008 to 2/6/2008. Classification 11122900

Outbreak from 23/8/2008 to 2/9/2008. Classification 11143900

Outbreak from 21/10/2008 to 7/11/2008. Classification 24452400

Staff data, recurrence checked for:

Revised:

Outbreak from 11/11/2006 to 29/3/2007. Classification 26596062 (was 26597062)

New:

Outbreak from 20/11/2007 to 31/3/2008. Classification 26597062

Outbreak from 19/4/2008 to 2/6/2008. Classification 25476162

Outbreak from 23/8/2008 to 2/9/2008. Classification 11143901

Outbreak from 21/10/2008 to 7/11/2008. Classification 24452401

Patient data, recurrence not checked for:

Revised:

Outbreak from 31/12/2005 to 10/1/2006. Classification 11148900 (was 23345500)

Outbreak from 9/5/2006 to 7/6/2006. Classification 25477220 (was 25478220)

Outbreak from 3/1/2007 to 24/1/2007. Classification 25568360 (was 25567360)

New:

Outbreak from 19/11/2007 to 18/12/2007. Classification 25477260

Outbreak from 27/12/2007 to 25/2/2008. Classification 25588160

Outbreak from 29/2/2008 to 29/3/2008. Classification 25578330

Outbreak from 1/4/2008 to 5/4/2008. Classification 11124900

Outbreak from 17/4/2008 to 3/5/2008. Classification 24457560

Outbreak from 7/5/2008 to 11/6/2008. Classification 24377220

Outbreak from 24/8/2008 to 27/8/2008. Classification 11112900

Outbreak from 19/9/2008 to 23/9/2008. Classification 01120900

Outbreak from 18/10/2008 to 8/11/2008. Classification 24467410

Patient data, recurrence checked for:

Revised:

Outbreak from 31/12/2005 to 10/1/2006. Classification 11148901 (was 23345501)

- same as recurrence not checked for

New:

Outbreak from 19/11/2007 to 11/6/2008. Classification 26598062

Outbreak from 24/8/2008 to 27/8/2008. Classification 11112901

Outbreak from 19/9/2008 to 23/9/2008. Classification 01120901

Outbreak from 18/10/2008 to 8/11/2008. Classification 24467411

Table 6.17

Combination	**Realistic**	**FI**	**Unrealistic**	**Mean Severity**
H24	76	24	0	1
H20	74	26	0	1
H08	74	25	1	1
H80	80	19	1	1
C05	69	29	2	1
H01	68	31	1	2
C18	71	29	0	2
H50	37	60	3	3

H24,H20	58	42	0	1
H24,H08	65	35	0	1
H24,H80	58	41	1	1
H24,C05	53	46	1	1
H24,H01	52	43	5	1.5
H24,C18	55	43	2	1.5
H24,H50	52	45	3	2
H20,H08	61	38	1	1
H20,H80	57	43	0	1
H20,C05	53	47	0	1
H20,H01	61	38	1	1.5
H20,C18	53	46	1	1.5
H20,H50	47	49	4	2
H08,H80	60	40	0	1
H08,C05	52	48	0	1
H08,H01	62	35	3	1.5
H08,C18	52	45	3	1.5
H08,H50	52	45	3	2
H80,C05	62	37	1	1
H80,H01	62	37	1	1.5
H80,C18	63	35	2	1.5
H80,H50	34	65	1	2
C05,H01	60	38	2	1.5
C05,C18	58	41	1	1.5
C05,H50	51	43	6	2
H01,C18	48	51	1	2
H01,H50	39	56	5	2.5
C18,H50	58	41	1	2.5
H24,H20,H08	48	52	0	1
H24,H20,H80	57	42	1	1
H24,H20,C05	48	52	0	1
H24,H20,H01	60	40	0	1.333333
H24,H20,C18	47	51	2	1.333333
H24,H20,H50	50	46	4	1.666667
H24,H08,H80	51	49	0	1
H24,H08,C05	58	42	0	1
H24,H08,H01	55	45	0	1.333333
H24,H08,C18	60	39	1	1.333333
H24,H08,H50	50	47	3	1.666667
H24,H80,C05	53	47	0	1
H24,H80,H01	64	36	0	1.333333
H24,H80,C18	50	47	3	1.333333
H24,H80,H50	48	48	4	1.666667
H24,C05,H01	46	54	0	1.333333
H24,C05,C18	51	49	0	1.333333
H24,C05,H50	54	44	2	1.666667
H24,H01,C18	53	46	1	1.666667
H24,H01,H50	52	46	2	2
H24,C18,H50	55	44	1	2
H20,H08,H80	50	50	0	1
H20,H08,C05	53	47	0	1
H20,H08,H01	61	39	0	1.333333
H20,H08,C18	56	43	1	1.333333

H20,H08,H50	50	47	3	1.666667
H20,H80,C05	45	55	0	1
H20,H80,H01	66	32	2	1.333333
H20,H80,C18	49	49	2	1.333333
H20,H80,H50	45	55	0	1.666667
H20,C05,H01	51	47	2	1.333333
H20,C05,C18	51	49	0	1.333333
H20,C05,H50	52	46	2	1.666667
H20,H01,C18	45	53	2	1.666667
H20,H01,H50	46	52	2	2
H20,C18,H50	50	45	5	2
H08,H80,C05	56	44	0	1
H08,H80,H01	64	35	1	1.333333
H08,H80,C18	51	47	2	1.333333
H08,H80,H50	51	46	3	1.666667
H08,C05,H01	48	50	2	1.333333
H08,C05,C18	43	54	3	1.333333
H08,C05,H50	51	47	2	1.666667
H08,H01,C18	53	44	3	1.666667
H08,H01,H50	51	49	0	2
H08,C18,H50	57	42	1	2
H80,C05,H01	56	42	2	1.333333
H80,C05,C18	53	46	1	1.333333
H80,C05,H50	56	41	3	1.666667
H80,H01,C18	54	45	1	1.666667
H80,H01,H50	48	50	2	2
H80,C18,H50	44	56	0	2
C05,H01,C18	59	40	1	1.666667
C05,H01,H50	41	54	5	2
C05,C18,H50	58	40	2	2
H01,C18,H50	46	53	1	2.333333
H24,H20,H08,H80	50	48	2	1
H24,H20,H08,C05	50	50	0	1
H24,H20,H08,H01	47	52	1	1.25
H24,H20,H08,C18	59	41	0	1.25
H24,H20,H08,H50	46	53	1	1.5
H24,H20,H80,C05	51	47	2	1
H24,H20,H80,H01	44	55	1	1.25
H24,H20,H80,C18	56	40	4	1.25
H24,H20,H80,H50	45	52	3	1.5
H24,H20,C05,H01	57	43	0	1.25
H24,H20,C05,C18	49	51	0	1.25
H24,H20,C05,H50	48	50	2	1.5
H24,H20,H01,C18	51	49	0	1.5
H24,H20,H01,H50	52	48	0	1.75
H24,H20,C18,H50	49	50	1	1.75
H24,H08,H80,C05	52	48	0	1
H24,H08,H80,H01	50	49	1	1.25
H24,H08,H80,C18	41	59	0	1.25
H24,H08,H80,H50	35	63	2	1.5
H24,H08,C05,H01	56	44	0	1.25
H24,H08,C05,C18	44	55	1	1.25
H24,H08,C05,H50	47	53	0	1.5

H24,H08,H01,C18	51	49	0	1.5
H24,H08,H01,H50	49	50	1	1.75
H24,H08,C18,H50	44	53	3	1.75
H24,H80,C05,H01	39	61	0	1.25
H24,H80,C05,C18	42	55	3	1.25
H24,H80,C05,H50	46	53	1	1.5
H24,H80,H01,C18	45	55	0	1.5
H24,H80,H01,H50	45	55	0	1.75
H24,H80,C18,H50	41	57	2	1.75
H24,C05,H01,C18	58	42	0	1.5
H24,C05,H01,H50	55	45	0	1.75
H24,C05,C18,H50	48	52	0	1.75
H24,H01,C18,H50	48	52	0	2
H20,H08,H80,C05	41	58	1	1
H20,H08,H80,H01	50	49	1	1.25
H20,H08,H80,C18	55	44	1	1.25
H20,H08,H80,H50	47	53	0	1.5
H20,H08,C05,H01	52	47	1	1.25
H20,H08,C05,C18	48	51	1	1.25
H20,H08,C05,H50	48	51	1	1.5
H20,H08,H01,C18	46	52	2	1.5
H20,H08,H01,H50	58	41	1	1.75
H20,H08,C18,H50	51	48	1	1.75
H20,H80,C05,H01	52	46	2	1.25
H20,H80,C05,C18	52	48	0	1.25
H20,H80,C05,H50	47	53	0	1.5
H20,H80,H01,C18	51	48	1	1.5
H20,H80,H01,H50	58	39	3	1.75
H20,H80,C18,H50	46	53	1	1.75
H20,C05,H01,C18	49	50	1	1.5
H20,C05,H01,H50	47	52	1	1.75
H20,C05,C18,H50	44	55	1	2
H20,H01,C18,H50	40	58	2	2
H08,H80,C05,H01	53	47	0	1.25
H08,H80,C05,C18	56	43	1	1.25
H08,H80,C05,H50	41	55	4	1.5
H08,H80,H01,C18	45	55	0	1.5
H08,H80,H01,H50	56	44	0	1.75
H08,H80,C18,H50	57	42	1	1.75
H08,C05,H01,C18	52	47	1	1.5
H08,C05,H01,H50	45	55	0	1.75
H08,C05,C18,H50	55	44	1	1.75
H08,H01,C18,H50	55	45	0	2
H80,C05,H01,C18	55	42	3	1.5
H80,C05,H01,H50	41	59	0	1.75
H80,C05,C18,H50	49	49	2	1.75
H80,H01,C18,H50	49	48	3	2
C05,H01,C18,H50	51	49	0	2
H24,H20,H08,H80,C05	51	49	0	1
H24,H20,H08,H80,H01	43	56	1	1.2
H24,H20,H08,H80,C18	48	51	1	1.2
H24,H20,H08,H80,H50	37	61	2	1.4
H24,H20,H08,C05,H01	51	47	2	1.2

H24,H20,H08,C05,C18	54	45	1	1.2
H24,H20,H08,C05,H50	48	52	0	1.4
H24,H20,H08,H01,C18	50	49	1	1.4
H24,H20,H08,H01,H50	43	55	2	1.6
H24,H20,H08,C18,H50	51	48	1	1.6
H24,H20,H80,C05,H01	41	58	1	1.2
H24,H20,H80,C05,C18	43	56	1	1.2
H24,H20,H80,C05,H50	41	58	1	1.4
H24,H20,H80,H01,C18	48	49	3	1.4
H24,H20,H80,H01,H50	48	51	1	1.6
H24,H20,H80,C18,H50	38	60	2	1.6
H24,H20,C05,H01,C18	57	42	1	1.4
H24,H20,C05,H01,H50	39	58	3	1.6
H24,H20,C05,C18,H50	42	57	1	1.6
H24,H20,H01,C18,H50	50	49	1	1.8
H24,H08,H80,C05,H01	54	46	0	1.2
H24,H08,H80,C05,C18	52	47	1	1.2
H24,H08,H80,C05,H50	52	46	2	1.4
H24,H08,H80,H01,C18	61	39	0	1.4
H24,H08,H80,H01,H50	45	54	1	1.6
H24,H08,H80,C18,H50	51	46	3	1.6
H24,H08,C05,H01,C18	56	43	1	1.4
H24,H08,C05,H01,H50	48	50	2	1.6
H24,H08,C05,C18,H50	48	51	1	1.6
H24,H08,H01,C18,H50	40	60	0	1.8
H24,H80,C05,H01,C18	48	52	0	1.4
H24,H80,C05,H01,H50	54	46	0	1.6
H24,H80,C05,C18,H50	55	44	1	1.8
H24,H80,H01,C18,H50	45	51	4	1.8
H24,C05,H01,C18,H50	43	56	1	1.8
H20,H08,H80,C05,H01	43	55	2	1.2
H20,H08,H80,C05,C18	45	55	0	1.2
H20,H08,H80,C05,H50	38	60	2	1.4
H20,H08,H80,H01,C18	43	57	0	1.4
H20,H08,H80,H01,H50	42	58	0	1.6
H20,H08,H80,C18,H50	47	51	2	1.6
H20,H08,C05,H01,C18	57	43	0	1.4
H20,H08,C05,H01,H50	44	54	2	1.6
H20,H08,C05,C18,H50	40	59	1	1.6
H20,H08,H01,C18,H50	50	50	0	1.8
H20,H80,C05,H01,C18	44	56	0	1.4
H20,H80,C05,H01,H50	51	47	2	1.6
H20,H80,C05,C18,H50	48	52	0	1.6
H20,H80,H01,C18,H50	50	50	0	1.8
H20,C05,H01,C18,H50	43	56	1	1.8
H08,H80,C05,H01,C18	54	46	0	1.4
H08,H80,C05,H01,H50	44	54	2	1.6
H08,H80,C05,C18,H50	46	52	2	1.6
H08,H80,H01,C18,H50	46	53	1	1.8
H08,C05,H01,C18,H50	54	46	0	1.8
H80,C05,H01,C18,H50	40	59	1	1.8
H24,H20,H08,H80,C05,H01	45	55	0	1.166667
H24,H20,H08,H80,C05,C18	51	49	0	1.166667

H24,H20,H08,H80,C05,H50	37	60	3	1.333333
H24,H20,H08,H80,H01,C18	47	52	1	1.333333
H24,H20,H08,H80,H01,H50	40	59	1	1.5
H24,H20,H08,H80,C18,H50	37	62	1	1.5
H24,H20,H08,C05,H01,C18	48	51	1	1.333333
H24,H20,H08,C05,H01,H50	49	49	2	1.5
H24,H20,H08,C05,C18,H50	38	62	0	1.5
H24,H20,H08,H01,C18,H50	46	53	1	1.666667
H24,H20,H80,C05,H01,C18	41	57	2	1.333333
H24,H20,H80,C05,H01,H50	42	58	0	1.5
H24,H20,H80,C05,C18,H50	46	52	2	1.5
H24,H20,H80,H01,C18,H50	45	55	0	1.666667
H24,H20,C05,H01,C18,H50	49	51	0	1.666667
H24,H08,H80,C05,H01,C18	43	56	1	1.333333
H24,H08,H80,C05,H01,H50	41	54	5	1.5
H24,H08,H80,C05,C18,H50	42	57	1	1.5
H24,H08,H80,H01,C18,H50	52	47	1	1.666667
H24,H08,C05,H01,C18,H50	44	52	4	1.666667
H24,H80,C05,H01,C18,H50	45	53	2	1.666667
H20,H08,H80,C05,H01,C18	39	60	1	1.333333
H20,H08,H80,C05,H01,H50	42	57	1	1.5
H20,H08,H80,C05,C18,H50	44	54	2	1.5
H20,H08,H80,H01,C18,H50	44	54	2	1.666667
H20,H08,C05,H01,C18,H50	49	51	0	1.666667
H20,H80,C05,H01,C18,H50	45	53	2	1.666667
H08,H80,C05,H01,C18,H50	47	49	4	1.666667
H24,H20,H08,H80,C05,H01,C18	45	55	0	1.285714
H24,H20,H08,H80,C05,H01,H50	40	60	0	1.428571
H24,H20,H08,H80,C05,C18,H50	50	49	1	1.428571
H24,H20,H08,H80,H01,C18,H50	46	52	2	1.571429
H24,H20,H08,C05,H01,C18,H50	43	55	2	1.571429
H24,H20,H80,C05,H01,C18,H50	45	54	1	1.571429
H24,H08,H80,C05,H01,C18,H50	37	60	3	1.571429
H20,H08,H80,C05,H01,C18,H50	48	50	2	1.571429
H24,H20,H08,H80,C05,H01,C18,H50	40	58	2	1.5

Table 6.17: The results for modelled outbreaks with multiple start points.

Table 6.18

%age Prob'y of Non-Ward Vertex Being Vacc'd at Each Time Step	%age Prob'y of Non-Ward, Non-Norm Vertex Being Treated at Each Time Step	%age Prob'y of Non-Normal Vertex Being Remodelled at Each Time Step	%age Prob'y of Non-Ward Vertex Increasing Resistance to Change	Min Time for Outb'k	Max Time for Outb'k	Mean Time for Outb'k
0	0	0	0	3	101	27
5	0	0	0	3	101	20.94
10	0	0	0	3	32	16.87
20	0	0	0	3	28	14.79
30	0	0	0	3	25	12.6
40	0	0	0	3	24	12.72
50	0	0	0	3	26	11.24
70	0	0	0	3	24	8.92
90	0	0	0	3	29	10.01
100	0	0	0	3	24	8.97
0	5	0	0	2	47	22.47
0	10	0	0	2	101	21.64
0	20	0	0	2	101	19.19
0	30	0	0	2	51	17.62
0	40	0	0	2	57	18.85
0	50	0	0	2	50	17.35
0	70	0	0	2	40	15.13
0	90	0	0	2	32	12.14
0	100	0	0	2	33	13.18
0	0	5	0	3	101	23.41
0	0	10	0	3	101	23.34
0	0	20	0	3	51	16.8
0	0	30	0	3	44	16.13
0	0	40	0	3	36	12.36
0	0	50	0	3	41	12.9
0	0	70	0	3	28	11.55
0	0	90	0	3	32	9.4
0	0	100	0	3	27	9.81
0	0	0	5	3	101	23.1
0	0	0	10	3	40	19.87
0	0	0	20	3	101	18
0	0	0	30	3	27	15.19
0	0	0	40	3	23	13.08
0	0	0	50	3	26	14.28
0	0	0	70	3	34	12.37
0	0	0	90	3	23	10.32
0	0	0	100	3	27	10.86

0	0	0	0	3	101	27
0	0	0	10	3	40	19.87
0	0	0	30	3	27	15.19
0	0	10	0	3	101	23.34
0	0	10	10	3	40	16.33
0	0	10	30	3	101	14.15
0	0	30	0	3	44	16.13
0	0	30	10	3	33	14.05
0	0	30	30	3	22	10.81
0	10	0	0	2	101	21.64
0	10	0	10	2	42	19.5
0	10	0	30	2	30	13.76
0	10	10	0	2	47	18.27
0	10	10	10	2	38	16.01
0	10	10	30	2	26	14.18
0	10	30	0	2	101	15.99
0	10	30	10	2	33	14.25
0	10	30	30	3	23	11.76
0	30	0	0	2	51	17.62
0	30	0	10	2	42	16.69
0	30	0	30	2	23	13.01
0	30	10	0	2	57	19.8
0	30	10	10	2	38	15.54
0	30	10	30	2	101	14.23
0	30	30	0	2	47	15.72
0	30	30	10	2	28	12.89
0	30	30	30	2	28	12.26
10	0	0	0	3	32	16.87
10	0	0	10	2	27	15.01
10	0	0	30	3	25	12.91
10	0	10	0	3	32	14.74
10	0	10	10	3	30	14.11
10	0	10	30	3	26	11.59
10	0	30	0	3	101	12.67
10	0	30	10	3	29	11.67
10	0	30	30	3	24	10.03
10	10	0	0	2	34	15.28
10	10	0	10	2	101	15.09
10	10	0	30	2	26	13.38
10	10	10	0	2	28	14.82
10	10	10	10	3	30	13.97
10	10	10	30	2	24	12.83
10	10	30	0	2	30	13.26
10	10	30	10	2	28	12.82
10	10	30	30	2	25	11.97
10	30	0	0	2	33	16.89
10	30	0	10	2	32	14.2
10	30	0	30	2	26	12.08
10	30	10	0	2	36	15.01
10	30	10	10	2	27	13.42
10	30	10	30	2	24	12.39
10	30	30	0	2	30	14.09
10	30	30	10	2	30	13

10	30	30	30	2	27	11.45
30	0	0	0	3	25	12.6
30	0	0	10	3	33	11.3
30	0	0	30	3	26	11.17
30	0	10	0	3	24	11.43
30	0	10	10	3	27	11.58
30	0	10	30	3	29	11.91
30	0	30	0	3	101	10.88
30	0	30	10	3	25	11
30	0	30	30	3	27	9.76
30	10	0	0	2	25	12.97
30	10	0	10	2	25	12.19
30	10	0	30	2	101	12.06
30	10	10	0	2	27	11.94
30	10	10	10	2	23	11.54
30	10	10	30	2	26	10.57
30	10	30	0	2	29	11.07
30	10	30	10	3	25	9.79
30	10	30	30	2	28	10.39
30	30	0	0	2	21	10.81
30	30	0	10	2	29	12.42
30	30	0	30	2	28	10.98
30	30	10	0	2	25	11.7
30	30	10	10	2	25	10.22
30	30	10	30	2	24	10.79
30	30	30	0	2	24	11.2
30	30	30	10	2	24	9.53
30	30	30	30	2	30	10.47

Table 6.18: The length of outbreaks (minimum, maximum and mean) for varying levels of intervention. The upper 37 rows are for single interventions (i.e. down to the blank line). The lower 81 are for multiple interventions. As some of these are also for single interventions, these rows are repeated for clarity. Note that the probability of a vertex being vaccinated, etc, equates to that percentage of vertices being vaccinated. Aside from treatment, the minimum can never be less than 3, as this equates to the one day of the infection plus the 48 hours post-asymptomatic (see 6.2.2.3.9.1).

References

Ancel Meyers L, Newman MEJ, Martin M & Schrag S, 2003, *Applying network theory to epidemics: control measures for Mycoplasma pneumoniae outbreaks*. Emerging Infectious Diseases [online] 9, 204-210. Available: http://www.cdc.gov/ncidod/EID/vol9no2/02-0188.htm [Accessed 28/12/05]

Ancel Meyers L, Newman M E J & Pourbohloul B, 2006, *Predicting Epidemics on Directed Contact Networks*. J. Th. Biol. [online] 240:3 pp 400-418. Available: ScienceDirect [Accessed 23/02/10]

Antal M & Balogh L, 2009, *Modeling belief systems with scale-free networks,* Neural Networks [online] 22(10), pp1359-1371. Available: ScienceDirect. [Accessed 22/02/10]

Auerbach DM, Darrow WW, Jaffe HW & Curran JW, 1984, *Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact*. Am J Med. [online] 1984 Mar;76(3):487-92. Available: PubMed. [Accessed 20/05/10]

Barabási, A L & Albert, R, 1999, *Emergence of Scaling in Random Networks,* Science [online] 286 p.509 Available: arXiv [Accessed 21/11/05].

Barker J, Vipond I B & Bloomfield S F, 2004, *Effects of cleaning and disinfection in reducing the spread of Norovirus contamination via environmental surfaces*, Journal of Hospital Infection [online] Volume 58, Issue 1, September 2004, Pages 42-49 [online] Available: ScienceDirect [Accessed 06/06/08]

Bean B, Moore BM, Sterner B, Peterson LR, Gerding DN & Balfour HH, 1982, *Survival of Influenza Viruses on Environmental Surfaces,* The Journal of Infectious Diseases, [online] Vol. 146, No. 1 (Jul., 1982), pp. 47-51 Available: Jstor [Accessed 10/12/09]

References

Best M & Neuhauser D, 2004, *Ignaz Semmelweis and the birth of infection control*, Quality and Safety in Health Care [online] 2004;13:233-234 doi:10.1136/qshc.2004.010918 Available: BMJ.com [Accessed 18/05/10]

Bollobás B, 2001, *Random Graphs*, 2nd edition, Cambridge: University Press

Boone, SA & Gerba, CP, 2007, *Significance of Fomites in the Spread of Respiratory and Enteric Viral Disease,* Applied And Environmental Microbiology [online], Mar. 2007, p. 1687–1696 Vol. 73, No. 6. Available: American Society for Microbiology [Accessed 06/06/08]

Carrat F, Luong J, Lao H, Sallé A, Lajaunie C & Wackernagel H, 2006, *A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics*. BMC Medicine [online] 2006, 4:26 doi:10.1186/1741-7015-4-26 Available: Biomedcentral [Accessed 28/12/06]

Caul E O. 1996, *Viral gastroenteritis: small round structured viruses, caliciviruses and astroviruses. Part II. The epidemiological perspective*. H Clin Pathol [online] 49: 959-964 Available: PubMedCentral [Accessed 23/08/06]

Centers for Disease Control and Prevention (CDC), 2006, *Norovirus: Technical Fact Sheet* [online], CDC. Available: http://www.cdc.gov/ncidod/dvrd/revb/gastro/downloads/noro-factsheet.pdf [Accessed 29/10/08]

Chadwick P R, Beards G, Brown D, Caul E O, Cheesbrough J, Clarke I, Curry A, O'Brien S, Quigley K, Sellwood J & Westmoreland D, 2000, *Management of hospital outbreaks of gastro-enteritis due to small round structured viruses,* Journal of Hospital Infection [online] (2000) 45: 1–10. Available: ScienceDirect [Accessed 09/08/06]

References

Chen S, 2001, *Foot-and-Mouth Disease Spread in This Small World*. [online], MSc thesis, University of Sussex. Available: http://www.informatics.susx.ac.uk/easy/Publications/Online/MSc2001/ThesisI.pdf [Accessed 10/12/04]

Clauset A, Newman M E J & Moore C, 2004, *Finding community structure in very large networks,* Phys. Rev. E [online] 70, 066111 Available: arXiv [Accessed 22/12/05]

Cohen R, Havlin S & ben-Avraham D, 2003, *Efficient Immunization Strategies for Computer Networks and Populations,* Phys Rev Lett [online] 91, 247091. Available: arXiv [Accessed 6/1/06]

Curtis V & Cairncross S, 2003, *Effect of washing hands with soap on diarrhoea risk in the community: a systematic review*, The Lancet Infectious Diseases [online] Vol 3 May 2003. Available: ScienceDirect [Accessed 04/03/10].

Dezsö Z & Barabási A-L, 2002. *Halting viruses in scale-free networks*. Phys Rev E [online] 65, 055103. Available: arXiv [Accessed 6/1/06]

Dodds P S, Muhamad R, Watts D J, 2003, An experimental study of search in global social networks. *Science* 2003 August 8;301(5634):827-829.

Dorogovtsev S N & Mendes J F F, 1999, *Exactly solvable analogy of small-world networks* Statistical Mechanics [online]. Available: arXiv [Accessed 29/12/04].

Durr C Mhalla M & Lei Y, 2003, *Quantum query complexity of graph connectivity,* Quantum Physics [online]. Available: arXiv [Accessed 25/11/05]

Eames K T D, 2008, *Modelling disease spread through random and regular contacts in clustered populations*, Theoretical Population Biology [online] 73:1, 104-111. Available: ScienceDirect [Accessed 18/08/09]

References

Erdös P & Rényi A, 1959, On random graphs, *Publicationes Mathematicae* 6, 290-297.

Esu-Williams E, 1995, *Sexually transmitted diseases and condom interventions among prostitutes and their clients in Cross River State* [online], Health Transition Review, Supplement to Volume 5, 223-228 Available: http://htc.anu.edu.au/pdfs/Esuwill1.pdf [Accessed 13/01/06]

Fauquet CM, Mayo MA, Maniloff J, Desselberger U & Ball LA (eds), 2005, *8th Reports of the International Committee on Taxonomy of Viruses,* Elsevier. [online] Available: http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fr-fst-a.htm [Accessed 30/10/08]

Fleming Forum, 2004. *MRSA – The facts. Now please wash your hands –just how important is hand hygiene in stopping the spread?* Fleming Forum[online]. Available: http://www.flemingforum.org.uk/slides/handwashing.pdf [Accessed 14/05/07]

Ganney P, 2003, A Framework for Investigating the Spread of Infection in a Hospital Environment Utilising a Stochastic and Small-World Model, *Proceedings of the Annual Scientific Meeting of the Institute of Physics & Engineering in Medicine, 2003,* p 105

Ganney P S, Madeo M & Phillips R, 2010, A classification system for hospital-based infection outbreaks, *Computational and Mathematical Methods in Medicine.* Vol 11, No. 4, December 2010, pp 297-311.

Gleizes O, Desselberger U, Tatochenko V, Rodrigo C, Salman N, Mezner Z, Giaquinto C & Grimprel E, 2006, *Nosocomial Rotavirus Infection in European Countries: A Review of the Epidemiology, Severity and Economic Burden of Hospital-Acquired Rotavirus Disease,* The Pediatric Infectious Disease Journal [online] January 2006 - Volume 25 - Issue 1 - pp S12-S21. Available: lww.com [Accessed 19/05/10]

References

Gould D, 2006, *Nurses' hands as vectors of hospital-acquired infection: a review*, Journal of Advanced Nursing [online] Volume 16 Issue 10, Pages 1216 – 1225. Available: Interscience [Accessed 19/05/10]

Graham B S,Ledgerwood J E & Nabel G J, 2009, *Vaccine Development in the Twenty-First Century: Changing Paradigms for Elusive Viruses,* Clinical Pharmacology & Therapeutics [online] (2009); 86 3, 234–236. doi:10.1038/clpt.2009.128. Available: Nature.com [Accessed 20/05/10]

Granovetter M S, 1973, The Strength of Weak Ties, *American Journal of Sociology,* 78(6), 1360-80.

Granovetter M S, 1983, The Strength of Weak Ties: A Network theory revisited, *Sociological Theory,* 1:203-33.

Granovetter M, 2003, Ignorance, Knowledge, and Outcomes in a Small World, *Science,* 2003 301, pp 773-774.

Griffith V, 2003, *Predicting the path of infectious diseases* [online], University of Texas at Austin. Available: http://www.utexas.edu/features/archive/2003/meyers.html [Accessed: 21/12/05]

Hamming R W, 1950. Error detecting and error correcting codes. *Bell System Technical Journal* 29(2) 147-160.

Hawker J, Begg N, Blair I, Reintjes R & Weinberg J, 2005, *Communicable Disease Control Handbook,* Oxford: Blackwell Publishing.

Hugonnet S, Perneger TV, Pittet D , 2002. *Alcohol-Based Handrub Improves Compliance With Hand Hygiene in Intensive Care Units,* Arch Intern Med. [online] 162: 1037-1043 Available: archint.ama-assn.org [Accessed 15/5/2007]

References

Jansen V A A, Stollenwerk N, Jensen H J, Ramsay M E, Edmunds W J, & Rhodes C J, 2003, Measles Outbreaks in a Population with Declining Vaccine Uptake, *Science* Vol. 301. no. 5634, p. 804

Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM., 1972, Visualisation by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J Virol 1972*; 10: 1075–81.

Kasturirangan, R, 1999, *Multiple scales in small-world graphs* Disordered Systems and Neural Networks [online]. Available: arXiv [Accessed 29/12/04].

Kleinberg J, 1999, *The Small-World Phenomenon: An Algorithmic Perspective,* [online], Cornell University Computer Science Department technical Report. Available: http://www.cs.cornell.edu/home/kleinber/swn.d/swn.html [Accessed 23/12/05]

Kleinfeld J, 2000, *Could It Be a Big World After All? - What the Milgram Papers in the Yale Archives Reveal About the Original Small World Study* [online], University of Columbia. Available: http://www.columbia.edu/itc/sociology/watts/w3233/client_edit/big_world.html [Accessed 3/11/04]

Last J M, 2008, *A dictionary of epidemiology*, 4th edition. Oxford: University Press.

Lau J T F, Yang X, Leung P, Chan L, Wong E, Fong C & Tsui H, 2004, *SARS in Three Categories of Hospital Workers, Hong Kong. Emerging Infectious Diseases* [online] 10:8, August 2004 pp 1399-1404. Available: http://www.cdc.gov/ncidod/EID/vol10no8/pdfs/04-0041.pdf [Accessed 01/04/09]

Levenshtein V I, 1966, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10 (1966):707–710.

References

Li J S, Sexton D J, Mick N, Nettles R, Fowler V G Jr, Ryan T, Bashore T, Corey G R, 2000, *Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis*, Clin Infect Dis. [online] 2000; 30 :633 –638. Available: PubMed [Accesed 30/12/09]

Li L, Alderson D, Doyle J, Tanaka R & Willinger W, 2005, *Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications (Extended Version),* Primary version in Internet Mathematics [online] Available: arXiv [Accessed 13/01/06]

Lindesmith L, Moe C, Lependu J, Frelinger JA, Treanor J, Baric RS, 2005, *Cellular and humoral immunity following Snow Mountain virus challenge*. J Virol. 2005 Mar;79(5):2900-9. [online] Available: PubMed [Accessed 30/05/08]

Lopman B.A., Brown D.W., Koopmans M, 2002, Human caliciviruses in Europe. *Journal of Clinical Virology*, 24, 137–160.

Lynn S, Toop J, Hanger C & Millar N, 2004, *Norovirus outbreaks in a hospital setting: the role of infection control*. Journal of the New Zealand Medical Association [online], Vol 117 No 1189. Available: New Zealand Medical Association [Accessed 15/6/2007]

Macpherson G (ed), 2002, *Black's Medical Dictionary*, 40th edition, London: A&C Black.

Mandelbrot B B, 1985, Self-Affine Fractals and Fractal Dimension, Phys. Scr. 32 257-260.

Martin E A (ed), 2002, *Oxford Concise Medical Dictionary*, 6th edition, Oxford: Oxford University Press.

References

Matsui SM & Greenberg HB, 2000*, Immunity to Calicivirus Infection,* The Journal of Infectious Diseases [online], volume 181, pages S331–S335. Available: Chicago Journals [Accessed 26/6/2007]

Mayhall C G, 2004, *Hospital Epidemiology and Infection Control,* 3rd Edition, New York: Lippincott Williams & Wilkins

Milgram S, 1967. The small-world problem, *Psychology Today* 2, pp 61-67.

Moslonka-Lefebvre M, Pautasso M & Jeger M J, 2009, *Disease spread in small-size directed networks: Epidemic threshold, correlation between links to and from nodes, and clustering,* Journal of Theoretical Biology [online] 260:3 pp 402-411. Available: ScienceDirect [Accessed 18/08/09].

Newman M E J, 1999, *Small Worlds – the Structure of Social Networks* [online], Santa Fe Institute. Available: http://www.santafe.edu/research/publications/workingpapers/99-12-080.pdf [Accessed 16/11/04].

Newman M E J, 2000, *Models of the Small World a review,* J. Stat. Phys. [online] 101, 819-841 Available: ArXiv [Accessed 20/12/05]

Newman M E J, 2003, *The structure and function of complex networks*. SIAM Review [online] 45, 167-256. Available: arXiv [Accessed 6/1/06]

Newman M E J, 2004a. *Detecting community structure in networks*. Eur. Phys. J. B 38, 321-330. [online] Available: http://www-personal.umich.edu/~mejn/papers/epjb.pdf [Accessed 28/12/05]

Newman M E J, 2004b, *Analysis of weighted networks*. Phys.Rev.E. [online] 70:5 Available: arXiv [Accessed 28/12/05]

References

Newman M E J, 2005, *Threshold effects for two pathogens spreading on a network,* Phys. Rev. Lett. [online] 95, 108701. Available: arXiv [Accessed 22/12/05]

Newman, M E J & Watts, D J, 1999a, *Renormalization group analysis of the small-world network model*, Physics Letters A [online] 263 pp341-346. Available: arXiv [Accessed 29/12/04].

Newman M E J & Watts, D J, 1999b, *Scaling and percolation in the small-world model*, Physical Review E [online] 60 pp7332-7342. Available: arXiv [Accessed 29/12/04].

Ng T W, Turinici G, Danchin A, 2003. *A double epidemic model for the SARS propagation,* BMC Infectious Diseases [online] 3:19 Available: Biomedcentral [Accessed 29/5/07].

Noakes C J, Beggs C B, Sleigh P A & Kerr K G, 2006, *Modelling the transmission of airborne infections in enclosed spaces*, Epidemiology and Infection [online] 134:05, October 2006, pp 1082-1091. Available: Cambridge Journals [Accessed: 30/12/09].

Pastor-Satorras R & Vespignani A, 2001, *Epidemic Spreading in Scale-Free Networks,* Physical Review Letters [online] 86:14 3200-3203. Available: http://www.asu.edu/clas/csdc/events/pdf/vespignani.pdf [Accessed 24/2/10]

Pittet D, Hugonnet S, Harbarth S, Mourouga P, Sauvan V, Touveneau S, Perneger T, 2000, *Effectiveness of a hospital-wide programme to improve compliance with hand hygiene,* The Lancet [online], Volume 356, Issue 9238, pp 1307-1312 Available: Elsevier [Accessed 11/12/09]

Richter C F, 1935, An instrumental earthquake magnitude scale*, Bulletin of the Seismological Society of America* v. 25 no. 1(1935), pp. 1-32

References

Saramaki J & Kaski K, 2004, *Predicting Development of Epidemics with A Dynamic Small-World Model* [online], (Preprint). Available: http://www.lce.hut.fi/~jsaramak/epidemic.pdf [Accessed: 29/10/04]

Sattenspiel L & Simon, C P, 1988, The spread and persistence of infectious diseases in structured populations, *Mathematical Biosciences*, 90, pp 367-383

Sengupta A, 1998, *C343/H343/A594 Data Structures* [online], Indiana University. Available: http://www.cs.indiana.edu/classes/c343-seng/readings/Jan2998.sgm [Accessed 03/03/06]

Simpson R H, 1974, The hurricane disaster potential scale, *Weatherwise* 27 (1974) (4), pp. 169–186.

Small M & Tse C K, 2005, *Small World and Scale Free Model of Transmission of SARS,* International Journal Of Bifurcation And Chaos In Applied Sciences And Engineering [online] 5:5, pp 1745-1756, (Preprint) Available: http://www.eie.polyu.edu.hk/~ensmall/pdf/IJBCXX.pdf [Accessed 29/10/04]

Smith A D, Datta S P, Howard Smith G, Campbell P N, Bentley R & Mckenzie H A (eds), 1997, *Oxford Dictionary of Biochemistry and Molecular Biology*, Oxford: Oxford University Press.

Smith D D, 1996, *The Kevin Bacon Game: Properties of the Kevin Bacon Absorbing Set* [online], Colorado State University. Available: http://www.eskimo.com/~bleach/docs/kb.html [Accessed 5/11/04].

Soumare B, Tempia S, Cagnolati V, Mohamoud A, Van Huylenbroeck G & Berkvens D, 2007, *Screening for Rift Valley fever infection in northern Somalia: A GIS based survey method to overcome the lack of sampling frame*, Veterinary Microbiology [online] 121 (2007) 249–256. Available: ScienceDirect [Accessed 26/2/10].

References

South Australia Department of Health, 2005, *The ways infectious diseases spread* [online], Government of South Australia Department of Health. Available: http://www.health.sa.gov.au/PEHS/Youve-got-what/ways-disease-spread-06.pdf [Accessed 11/12/09]

Strogatz S, 2001, Exploring Complex Networks, *Nature*, 410 pp 268-276

Tibballs J, 1996. *Teaching hospital medical staff to handwash*. Med J Austral [online] 164: 395-398. Available: Medical Journal of Australia Online [Accessed 01/03/10]

Toroczkaia Z & Guclu H, 2007, Proximity networks and epidemics . *Physica A: Statistical Mechanics and its Applications* 378:1, 68-75

Vanderpas J, Louis J, Reynders M, Mascart G & Vandenberg O, 2009, Mathematical model for the control of nosocomial norovirus. *J. Hosp. Infect*., 71 pp 214-222.

Verdasca J, Telo da Gama M M, Nunes A, Bernadino N R, Pacheco J M & Gomes M C, 2004, *Recurrent epidemics in small world networks*, Journal of Theoretical Biology [online], 233 (2005) pp 553–561. Available: ScienceDirect [Accessed 24/2/10]

Ward R I, Bernstein D I, Knowlton D R, Sherwood J R, Young E C, Cusack T M, Rubino J R & Schiff G M, 1991, *Prevention of Surface-to-Human Transmission of Rotaviruses by Treatment with Disinfectant Spray*, J. Clin. Microbiol. [online], 29 pp 1991-1996, No. 9 Available: PubMed [Accessed 12/06/08]

Watts D J, 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press

Watts D J & Strogatz S H, 1998, Collective dynamics of 'small-world' networks. *Nature*, 393 pp 440-442.

References

Wikipedia, *the free encyclopaedia* [online]. Available: http://en.wikipedia.org.

Wilson R J, 1985, *Introduction to Graph Theory*, 3rd edition. Harlow, Essex: Longman.

Witten G & Poulter G, 2007, *Simulations of infectious diseases on networks,* Computers in Biology and Medicine [online] 37:2, 195-205. Available: ScienceDirect [Accessed 18/08/09]