Modelling disease spread through random and regular contacts in clustered populations

K.T.D. Eames

Department of Biological Sciences and Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

Received 15 March 2007
Available online 9 October 2007

Abstract

An epidemic spreading through a network of regular, repeated, contacts behaves differently from one that is spread by random interactions: regular contacts serve to reduce the speed and eventual size of an epidemic. This paper uses a mathematical model to explore the difference between regular and random contacts, considering particularly the effect of clustering within the contact network. In a clustered population random contacts have a much greater impact, allowing infection to reach parts of the network that would otherwise be inaccessible. When all contacts are regular, clustering greatly reduces the spread of infection; this effect is negated by a small number of random contacts.

Keywords: Epidemic; Network; Clustering; Pair-wise approximation

1. Introduction

When trying to understand the spread of an infectious disease through a population, many different factors are important. Pathogen genetics, host immune response, host behaviour, and public health interventions all have a role. Here, we focus on the part played by the social behaviour of the host population, specifically considering a human population.

Infections are often transmitted through close contact. Sexually transmitted infections are spread by intimate physical contact whereas common airborne infections such as influenza require less close proximity (Hethcote and Yorke, 1984; Anderson and May, 1992; Wallinga et al., 1999). Vector or water-borne diseases do not generally rely on an interaction between infector and infected but instead tend to be spatially localised (Anderson and May, 1992). In this paper we will consider infections that are transmitted through direct interaction between individuals, and observe that the relevant definition of “interaction” is pathogen-specific.

Many contacts are long lasting. Those who share houses, workspaces, or those who are long-term sexual partners interact repeatedly and over a long period of time (Edmunds et al., 1997; Martin and Yeung, 2006). It is no surprise that much transmission of infectious diseases takes place within the close household, since these individuals interact with high frequency (Eichner, 2003; Ferguson et al., 2006).

Some contacts can be considered as random events: interactions that take place once and only once. Those who sit next to each other on a train, or bump into each other in the street, for example, would fall into this category (Eichner, 2003; Ferguson et al., 2006).

Other contacts are harder to categorise: if an infection lasts for a fortnight then a barber visited each month or an aunt seen every Christmas is a contact that will not be repeated during the infectious period, and is, by this measure, indistinguishable from a random interaction; however, such contacts may have a place in the wider social network—linked through mutual friends, for instance—that gives them more significance than a random event.

For the purposes of this paper, we will categorise contacts as regular or random: regular contacts are those that are seen repeatedly during the course of an infection,
random contacts are those that will not be encountered multiple times by an infectious individual. This categorisation of contacts as regular (repeated) or random (one-off) interactions simplifies greatly the reality of human social behaviour, but it allows the tractable exploration of the effect of different types of interaction without the need for large scale data collection to parameterise detailed simulations.

When modelling the spread of an infection, the two distinct types of interaction require different treatment. Regular interactions can be viewed as a contact network (Watts and Strogatz, 1998; Bearman et al., 2004; De et al., 2004; Eames and Keeling, 2004; Eubank et al., 2004; Andre et al., 2006), with individuals appearing as nodes and those individuals who have regular contact with each other being joined by lines. This network representation illustrates very clearly that each individual has, in general, regular contact with only a small subset of the population.

Random, one-off, contacts can be well captured with more familiar mean-field models (Anderson and May, 1992). These models, which have been applied to a huge range of infection scenarios, assume that all individuals in a population are capable of interacting, in contrast to the network paradigm that limits the number of interactions. Mean-field models adhere nicely to the idea of brief, unpeated, contacts, resulting in the force of infection experienced by a susceptible individual being proportional to the level of infection present in the whole population.

If mean-field and network models with the same contact intensity are compared we would expect infection to spread more slowly within a network. This is because of a local “burning out” of susceptible individuals and is manifested by correlations in infection status emerging between connected individuals—infected individuals tend to be linked to other infected individuals, by whom they were infected or to whom they have transmitted infection; therefore the number of contacts between infected and susceptible individuals is reduced and the spread of infection slowed (Dickmann et al., 1998; Keeling, 1999).

It has been shown that the structure of a contact network has important implications for the spread of infection. Individuals with particularly high numbers of contacts can dominate the dynamics and are key targets for interventions (Kretzschmar et al., 1996; Rothenberg et al., 1998; De et al., 2004; Andre et al., 2006), an effect also shown by high-mixing subgroups in mean-field models (Hethcote and Yorke, 1984; Anderson and May, 1992). It has also been shown that highly clustered networks, those in which two linked individuals are particularly likely to have other contacts in common, result in reduced epidemic spread; in clustered networks infection tends to be confined to highly interconnected cliques and is less likely to emerge into the wider population (Rothenberg et al., 1996; Keeling, 1999; Eubank et al., 2004) although it may be more likely to infect all of a well-connected cluster (Newman, 2003). Clustering makes contact tracing a more effective control measure since it allows multiple routes for intervention efforts to follow to reach infected individuals (Eames and Keeling, 2003; Tsimring and Huerta, 2003; Kiss et al., 2005). Further, by increasing the chance that two infected individuals will compete to infect the same susceptible individual, clustering alters the evolutionary pressures that a pathogen experiences (Boots and Sasaki, 1999; Read and Keeling, 2003; Boots et al., 2004).

In this paper we will examine the impact of regular and random contacts on disease spread, looking particularly at the differences between clustered and unclustered social networks. We will see that disease spreads further and more rapidly when contacts are random, and that the difference between regular and random contacts is most apparent in highly clustered populations.

2. Methods

Mixing patterns within a population display a combination of order and randomness. This observation is particularly apparent when looking at the spatial arrangement of contacts; many contacts are with nearby individuals while some are between individuals a large distance apart (Read and Keeling, 2003; Ferguson et al., 2006). Modelling approaches have dealt with this fact in a number of ways: notably, models of foot-and-mouth infection using an interaction strength based on separation (Ferguson et al., 2001; Keeling et al., 2001) with, in these cases, nearby farms being much more likely to interact than those widely separated; an alternative approach, sparked by the work of Watts and Strogatz (1998), is to use small-world networks, i.e. networks that look in the main like lattices, with individuals connected to their nearest neighbours, but with some links rewired to allow long-distance connections (Watts and Strogatz, 1998; Boots and Sasaki, 1999; Tsimring and Huerta, 2003). Each of these models idealises spatial regularity and randomness as local and long-distance links. However, none contains the distinction between social regularity and randomness that we seek to investigate here: all links are of the same type once they have formed, rather than some being long-lasting and others short lived. The models can be seen as spatially extended versions of either the mean-field or network approach but not as an amalgam of the two.

Some previous models have been developed to combine network and mean field interactions (Ball et al., 1997; Ferguson et al., 2001; Eames and Keeling, 2003; Kiss et al., 2006). The approach taken here follows the path laid by these. In this paper we will investigate disease dynamics in a population that has both random and regular contacts, which we will model as mean-field and network interactions, respectively.

2.1. Models

We will look at two modelling approaches: stochastic and deterministic. The stochastic approach is more natural but it requires that every link within the network is known; this is not a problem when using simulated networks but it will be a concern when applying the model to a real-world population. Stochastic models allow individual variation to be included with relative ease providing there are appropriately detailed data (Eubank et al., 2004; Ferguson et al., 2006). The deterministic approach only provides an approximation to the true mixing behaviour but requires far less information to parameterise. By allowing the calculation of analytical results, it allows a more straightforward evaluation of the impact of a pathogen over a wide range of population-level behavioural parameters such as the mean number of contacts and the level of clustering.

We will investigate the case of a pathogen that confers lasting immunity following infection and which allows individuals to be classified as either susceptible (S), infected and infectious (I) or recovered (R). This commonly used SIR classification is a useful approximation to the many complexities of the infection process within an individual (Hethcote and
Yorke, 1984; Anderson and May, 1992). Regular contacts are treated as a static network while random interactions are modelled as mean-field mixing.

2.1.1. Stochastic model

Combining regular and random interactions in a stochastic model is straightforward. Regular contacts are represented as a network, through which infection can spread along links between connected individuals. Random contacts are thought of as short-lived interactions with members of the population chosen at random and are modelled as mean-field interactions.

The rate at which a susceptible individual becomes infected is therefore:

\[
\tau_N \times \# \text{ infected regular contacts} \\
\tau_M \times \# \text{ random contacts} \\
\times \text{ proportion of population infected,}
\]

where \(\tau_N\) and \(\tau_M\) are the rates of transmission through regular and random contacts, respectively. Throughout, we will use the subscripts \(N\) and \(M\) to represent parameters related to the regular (network) and random (mean-field) interactions, respectively. In this vein, we define \(k_N\) and \(k_M\) to be the number of regular and random contacts, respectively, and we define the total number of contacts, \(K\), by \(K = k_N + k_M\). In the mean-field case, \(k_M\) can be thought of as the number of contact events per unit time, whereas the \(k_N\) network contacts are assumed to be fixed.

The appropriate network to use depends on the population being studied and the relevant definition of a contact—for instance, for a given population, sexual contact networks will differ from social interaction networks. Rather than attempting an accurate representation of a particular population, here we will look at simplified network configurations. Throughout, for simplicity, we will assume that the network is homogeneous, with each individual having the same number, \(k_N\), of contacts. Following Keeling et al. (1997), we define the clustering coefficient, \(\phi\), to be the proportion of triples that form triangles, i.e. the proportion of contacts of any given individual that are connected to each other.

We use an iterative process to form clustered networks: first, a proportion of the desired links are made by connecting randomly chosen nodes; then, with some probability, triangles are formed by adding links between individuals with a mutual contact (with the restriction that no individual may have more than \(k_N\) contacts and that no two individuals may be connected to each other twice). These two steps are repeated until the desired number of contacts has been reached. Clustering can be increased by either increasing the probability of forming triangles or by decreasing the proportion of the desired number of contacts made each iteration. In contrast to other models (for example Watts and Strogatz, 1998; Ferguson et al., 2001; Read and Keeling, 2003), the probability of any two individuals becoming connected does not depend on the distance between them in any abstract or realistic space. Aside from the clustering introduced the networks formed will resemble random graphs. The network formation method used, although not enabling a network with a specified value of \(\phi\) to be formed, permits a range of networks with different values of \(\phi\) to be simply generated. The mechanism chosen allows clustering to reach high levels without the network breaking up into a collection of disconnected components (Newman, 2003).

A fuller discussion of the formulation of pair-wise models can be found elsewhere (Rand, 1999); briefly, we let \([A]\) be the number of individuals of type \(A\) (either \(S\), \(I\), or \(R\) to represent susceptible, infected, or recovered, respectively); \([AB]\) the number of contacts in the network between a type \(A\) individual and a type \(B\) individual; and \([ABC]\) the number of \(A - B - C\) triples within the network. For ease of book-keeping, pairs are counted once in each direction, thus \([AB] = [BA]\) and \([AA]\) is even. With this notation, we can form a system of differential equations describing the spread of infection through the population passing along both regular network contacts and through random mean-field interactions:

\[
[S] = -\tau_N[S][I] - \tau_M k_M [S][I]/P,
\]

\[
[I] = \tau_N[S][I] + \tau_M k_M [S][I]/P - \phi[I],
\]

\[
[R] = \phi[I],
\]

\[
[SS] = -2\tau_N[SS][I] - 2\tau_M k_M [SS][I]/P,
\]

\[
[SI] = \tau_N[SSI] - [SI] + \tau_M k_M [SS][I]/P - [SS][I]/P - \phi[SI],
\]

\[
[SR] = -\tau_N[ISR] - \tau_M k_M [SR][I]/P + \phi[SI],
\]

\[
[I^2] = 2\tau_N[ISI] + [SI] + 2\tau_M k_M [SI][I]/P - 2\phi[I^2],
\]

\[
[IR] = \tau_N[ISR] + \tau_M k_M [SR][I]/P + \phi[I^2] - [IR],
\]

\[
[RR] = 2\phi[IR],
\]

where \(\phi\) is the recovery rate and \(P = [S] + [I] + [R]\) is the population size. To close the system we need to be able to approximate the triples that appear in terms of pairs and singletons. In an unclustered network we can use the approximation (Keeling et al., 1997)

\[
[ABC] \approx \frac{(k_N - 1)[AB][BC]}{k_N} [B],
\]

which assumes that the triple is made up of two pairs that have a central individual in common but are otherwise independent. In the case of clustered networks, this assumption is no longer appropriate and instead we use

\[
[ABC] \approx \frac{(k_N - 1)[AB][BC]}{k_N} [B] \left( 1 - \phi + \phi \frac{ACP}{[A][C]k_N} \right),
\]

which is made up of two parts: the approximation, as above, for unclustered triples (i.e. those that do not form a triangle) along with the correlation between the outside members of the triple if the triple happens to form a triangle (Keeling et al., 1997). With the application of this approximation, the system of equations can be closed and numerically iterated to approximate the behaviour of an epidemic on a network with a given number of regular contacts, \(k_N\), and a given clustering coefficient, \(\phi\).

One advantage of the deterministic model over its stochastic counterpart is the scope it gives for obtaining easily interpreted analytic results. In particular here we will derive expressions for the initial growth rate of the epidemic as functions of the clustering coefficient, \(\phi\), and the distribution of contacts between regular \((k_N)\) and random \((k_M)\).

3. Results

3.1. Unclustered populations

We begin by investigating the difference between regular and random contacts within an unclustered population. The network of regular contacts in such a population is approximately tree-like, with very few triangles and therefore very little social clustering.
Fig. 1 shows typical time series of epidemics as the proportion of random contacts is varied. We see that the stochastic and deterministic models are in close agreement and both predict that there is a difference in the initial epidemic growth rate, with growth rates being largest when all contacts are random. This difference is translated into a difference in the final size of the epidemic, with more individuals being infected when there are few regular contacts (Diekmann et al., 1998; Keeling, 1999).

Using the pair-wise approximation, we will explore the exponential growth rate of the epidemic during its early stages in more detail. We can derive an expression for the early stage growth rate of an epidemic: we have that

$$ R = \frac{(\tau N + \tau M k_M)}{g}. $$

The early exponential growth of the epidemic is characterised by a phase during which \([SI]/[I] = \lambda\), say, giving a growth rate of \(\tau N + \tau M k_M - g\). From this growth rate we can define a quantity, \(R\), which has analogous properties to the basic reproductive ratio, \(R_0\) (the number of secondary cases produced by an index case in an otherwise susceptible population, Anderson and May, 1992):

$$ R = \frac{(\tau N + \tau M k_M)}{g}. $$ (15)

\(R\) and \(R_0\) are not the same thing in this network model, since the initial growth rate does not translate into number of secondary cases in the way that it does for simple mean-field models, but it shares some important properties: in particular, \(R > 1\) implies that the epidemic initially grows whereas \(R < 1\) implies that there is no spread.

To evaluate \(R\) we must calculate the value taken by \([SI]/[I]\) in the early stages of the epidemic. Following the methods of Keeling (1999) we calculate the early quasi-equilibrium value of \(\lambda\) as follows:

$$ \lambda = 0 \Rightarrow [SI] = [I] \Rightarrow \lambda \Rightarrow $$ (16)

$$ (\tau_N[S][I] - [SI] - [SI]) + \tau_M k_M [I]/P[S][S] $$

$$ - [SI] = g[SI]/[I] $$

$$ \Rightarrow (\tau_N[S] + \tau_M k_M [S]/[I] - g)[SI]/[I]. $$ (17)

Using the triples approximation (Eq. (11)) and making the substitution \([SI] = \lambda [I]\) and linearising at the early stage of the epidemic, when \([I] \ll 1\), \([S] \approx P\), and \([SS] \approx k_N P\) we obtain

$$ \lambda = \frac{R g k_N}{2 \tau_N + R g}, $$ (18)

which gives the result:

$$ R = \frac{\tau_M k_M + \tau_N (k_N - 2) + \sqrt{(\tau_M k_M + \tau_N (k_N - 2))^2 + 8 \tau_M k_M \tau_N}}{2g}. $$ (19)

We observe a number of things: first, when there are no regular contacts \((k_N = 0)\) the result simplifies to the expected result: \(R = \tau_M k_M / g\). When there are no random contacts \((k_M = 0)\) we have \(R = \tau_N (k_N - 2) / g\) (Keeling, 1999). We see, as previously observed, that an epidemic spreads more slowly when restricted to pass through a network (Diekmann et al., 1998; Keeling, 1999). If we define \(R_N = \tau_N (k_N - 2) / g\) and \(R_M = \tau_M k_M / g\) the expression simplifies to

$$ R = \frac{R_M + R_N + \sqrt{(R_M + R_N)^2 + 8 R_M R_N (k_N - 2)}}{2}. $$ (20)

Further straightforward manipulation shows that \(R_M + R_N < R < R_M + \tau_N k_N / g\). Thus the growth rate of an epidemic that spreads through both regular and random contact types exceeds the sum of the growth rates of an epidemic when restricted to each contact type in turn; the random spread speeds up the spread of an epidemic through a network of regular contacts. As expected, though, the growth rate is lower than it would be if all contacts were random.

Fig. 2 shows the effect on an epidemic of transferring interactions from regular to random. \(K = k_N + k_M\) is held constant but the fraction of random contacts is varied. We see that populations with more random contacts experience greater transmission, and that the effect is stronger when there are fewer contacts in total: when there are few contacts the network and the mean-field models are more different.

### 3.2. Clustered populations

We now look at epidemics on clustered networks. The models for these networks closely resemble those discussed
above; in the stochastically generated networks, triples are joined to form triangles (as described above), thus increasing the amount of clustering, and in the deterministic case the triples approximation (Eq. (12)) is used to include the effect of clustering in the network of regular contacts.

In the deterministic case similar calculations can be made to derive an expression for the initial growth rate of the epidemic. However, the added complexity of the triples approximation means that a simple closed form for the growth rate cannot be obtained.

3.2.1. Regular contacts only

We begin by looking at the case of a clustered network of regular contacts with no random interactions, i.e. \( k_M = 0 \). This section considers solely the influence of clustering within a social network on the spread of infection.

We see in Fig. 3 that the stochastic model predicts that clustering within a network can have a huge effect on the spread of an epidemic: particularly when the number of contacts is small, clustering prevents infection from emerging from densely connected cliques in to the wider population, and therefore the progress of infection is slowed (Rand, 1999). We can obtain the same result by looking at the initial epidemic growth rate using the deterministic pair-wise model. As before, we have \( R = \tau_N \lambda / g \), where at the early stage equilibrium value of \( \lambda \) we have

\[
(\tau_N (\text{SSI} - [\text{SI}] - [\text{SI}]) - g[\text{SI}])[\text{I}] = (\tau_N [\text{SI}] - g[\text{I}])[\text{SI}].
\]

(21)

In the case of a clustered population, the term \([\text{SI}]\) is no longer negligible if such a triple forms a triangle (since infected individuals are highly correlated). In this case, the expression reduces to

\[
\lambda = (k_N - 1) \left( 1 - \phi + \frac{\phi \lambda}{K_N} \right) - 1 - \frac{(k_N - 1) \phi [\text{II}]}{K_N [\text{I}]}.
\]

(22)

Hence, when the clustering coefficient, \( \phi \), is non-zero we must also derive the early stage quasi-equilibrium value of \([\text{II}]/[\text{I}]\). We let \( \alpha = [\text{II}]/[\text{I}] \) and the same methodology gives

\[
\alpha = \frac{2 \tau_N \lambda}{\tau_N \lambda + g - 2 \tau_N \lambda^2 \phi (k_N - 1)/k_N^2}.
\]

(23)

Substituting this back in we obtain that \( \lambda \) is the solution of the equation:

\[
\lambda = (k_N - 1) \left( 1 - \phi + \frac{\phi \lambda}{K_N} \right) - 1 - \frac{(k_N - 1) \phi}{k_N^2} - \frac{2 \tau_N \lambda}{(\tau_N \lambda + g - 2 \tau_N \lambda^2 \phi (k_N - 1)/k_N^2)}.
\]

(24)

This can be solved numerically with ease. We see from Fig. 3b that it leads to the same conclusions as the stochastic model: that clustering in a network noticeably impedes the spread of infection. Increasing \( \phi \) leads to a decrease in \( \lambda \) and therefore a decrease in epidemic growth rate. Further, we see that the lower the number of contacts the greater the impact of clustering.

Eq. (24) also allows us to calculate analytically the critical value of the clustering coefficient \( \phi \) for which an epidemic cannot take off (see Rand, 1999 for a similar calculation in the case of an epidemic in which individuals are once again susceptible on recovery): at this point, \( R = 1 \) and \( \lambda = g/\tau_N \) giving the critical value of \( \phi \) to be the solution of the quadratic

\[
\phi^2 \frac{g}{\tau_N} \frac{(k_N - 1)}{k_N} \left( \frac{k_N - g}{\tau_N} \right) + \phi \left( \frac{g^2}{\tau_N^2} + \frac{g}{\tau_N} - k_N \right) + \frac{k_N^2}{(k_N - 1)} \left( k_N - \frac{g}{\tau_N} - 2 \right) = 0.
\]

(25)
Fig. 4 shows how the critical value of $\phi$ depends on both the transmission parameter and the neighbourhood size. No matter how large the transmission rate or the neighbourhood size, the model predicts that there is always a level of clustering that will suffice to prevent the epidemic from taking off. As $\tau_N/g \rightarrow \infty$ or $k_N \rightarrow \infty$ the critical value of $\phi$ approaches 1, i.e. a fully clustered population with all individuals with a mutual contact being connected. Such high clustering coefficients would require the network to be split into multiple separate cliques with everyone within each clique being connected; this is unlikely to be seen in any real population, hence clustering is anticipated to be capable of preventing an epidemic only for smaller values of the transmission rate or neighbourhood size.

3.2.2. Regular and random contacts

To the clustered network model described in the previous section we can add random connections leading to mean-field spread of infection. We can thereby investigate the difference between regular and random contacts in a clustered society.

Again, the adjustment to the models is slight, with mean-field spread being included in the models. Once again, although no simple result is forthcoming, the same approach as above can be used to explore the initial growth rate of the deterministic epidemic. Here we have

$$R = \left(\tau_N^* + \tau_M k_M^*\right)/g,$$

where here $\lambda$ is the solution of

$$\lambda = (k_N - 1)\left(1 - \phi + \frac{\phi \lambda}{k_N}\right) - 1$$

and $\phi > 1$.

As before $z = [II]/[I]$ evaluated at the early quasi-equilibrium level given by

$$z = \frac{2\tau_N \lambda}{\tau_M k_M^* + \tau_N^* g - 2\tau_N^* \lambda^* (k_N - 1)/k_N^*}.$$  (27)

As before, numerical solution is straightforward. Fig. 5 shows the influence of clustering on an epidemic in a population with both regular and random contacts. Both for the stochastic final size and deterministic growth rate, we reach the same conclusions. When contacts are random rather than regular in an unclustered network the speed and final size of an epidemic are increased by a noticeable but not enormous amount (the effect being greatest when the total number of contacts is small); in a clustered network, the impact is far larger; when clustering is particularly high a few random contacts can be the difference between an epidemic that fails to take off and one that infects half the population.

Again, we can look for the critical level of clustering that reduces $R$ to below one (Fig. 6). We see in this case that the more random contacts there are the smaller the region of parameter space in which clustering of regular contacts can prevent epidemic spread. The presence of random contacts slightly enlarges the region in parameter space where an epidemic in an unclustered population can take off and also dramatically reduces the possible impact of clustering. The greater the proportion of random contacts, the lower the capacity of clustering to influence an epidemic.

4. Conclusions

A wide range of interactions is implicated in disease transmission: from contacts whom we meet every day to those whom we will never see again. From the point of view of the infected individual, the distinction is between those contacts who will be seen multiple times during the infectious period of the pathogen and those who will not.
It makes more than merely a descriptive difference whether contacts are regular or random. Regular, repeated, contacts constrain the spread of infection since effectively an infected individual “wastes” some contact events by meeting again people whom he has already infected. Therefore, when contacts are random, one-off, events infection spreads faster and further.

The simple models of social interactions described here suggest that, in the case of an unclustered network, the impact of an epidemic is increased by a small but potentially significant amount when contacts are random rather than regular. However, when regular contacts are clustered then random contacts make a huge amount of difference to the epidemic. In the absence of random mixing, when interactions are limited to a network of regular contacts, high levels of clustering can prevent an epidemic from taking place; a few random contacts allow wider transmission, even in highly clustered networks. In a clustered network infection is restricted to tightly knit cliques with little interaction with the rest of the population; the presence of random contacts allows infection to emerge and spread from clique to clique.

In the case of unclustered networks, random contacts serve to reduce the correlation between interacting individuals. In clustered networks, however, random contacts act more like the long-range links introduced in small-world models (Watts and Strogatz, 1998), allowing infection to pass between different regions of the network. However, unlike in small-world models, in the model investigated here randomness does not relate to the spatial arrangement of links but to the nature of the interactions.

Both the stochastic and deterministic approaches taken here support the conclusions that clustering is an important aspect of social mixing patterns whose impact is reduced by random, irregular, interactions. However, neither model is capable of capturing the complex processes that lead to real-world social network formation. In the real world, clustering may arise from a number of different routes, such as through sharing homes, schools, or workplaces, or through spatial proximity, or through common social activities (Ball et al., 1997; Rothenberg et al., 1998; Watts and Strogatz, 1998; Ferguson et al., 2001, 2006; Read and Keeling, 2003; De et al., 2004; Eubank et al., 2004), and there are many ways to generate models of clustered networks, some of which are based on approximations of these social processes. The models presented here aim to make as few assumptions as possible about how social structure arises and so provide an abstract arena in which to investigate particular aspects of behaviour. While the broad conclusions are expected to apply more generally, different modelling approaches will give different quantitative predictions of the effects of clustering and of interaction patterns.
To know the relative importance of regular and random contacts, it is vital to be able to assess the level of clustering within social networks. This is no easy task, and requires extensive surveying of social mixing behaviour (Edmunds et al., 1997). Information about behaviour within households or workplaces, for example, though interesting and useful, is obtained from a naturally clustered subset of the population, so does not provide a sensible estimate of social clustering. Accurate surveys are needed to determine both how many regular and how many random contacts we have and also the amount of clustering displayed by the contacts.

Acknowledgments

The author thanks EPSRC and Emmanuel College, Cambridge, for financial support, Matt Keeling for helpful discussions, and two anonymous referees for their constructive suggestions.

References


