# Analysis of Importance of Brief Encounters for Epidemic Spread

**P. Dawson** [a]

[a]*Land Division, Defence Science and Technology Organisation, 506 Lorimer St, Fishermans Bend, VIC 3207, Australia*
*Email: peter.dawson@dsto.defence.gov.au*

**Abstract:** Social network research has discovered that in many human networks both the frequency and duration of social interactions approximately follow a power-law with cumulatively short duration interactions being vastly more common than long duration interactions. Conversely it is known that longer duration interactions lead to a greater chance of disease transmission. This article examines mathematically if these briefer, more numerous, encounters are more important to epidemic spread than longer encounters. The focus is on respiratory diseases for which there is evidence that transmission is dominated by close range aerosol transmission.

The method involves mathematically deriving how many people a single infectious person will infect over the course of a day, in a population whose contact durations are governed by a power-law profile. Experimental evidence for use of this profile is cited. The effect of progressively eliminating the shortest and longest links for the network is then studied.

The mathematical analysis is validated by simulating disease outbreaks over a freely available, experimentally obtained high resolution school temporal-contact network dataset.

The results show that while the vast majority of contacts are very short, they should not play a great role in respiratory disease transmission, and that it is the few number of long interactions that dominate the disease spread. An estimate for the factor that the basic reproduction rate for a disease can be reduced by, by culling long interactions, is derived.

This work has practical applications. Firstly it indicates that partial quarantining that consists of cutting long interactions (and not replacing these by many short interactions) should be effective for many diseases. Secondly it shows that when modelling disease spread in a population, a model could be greatly simplified by ignoring quick interactions while not losing a large amount of accuracy.

*Keywords: Epidemics, Networks, Probability, Snowball sample*

# 1 INTRODUCTION

One of the most common ways to model an epidemic is to model person-to-person transmission via a contact network. This is very relevant for modelling the spread of respiratory disease, such as influenza, where, based on the work of Atkinson and Wein (2008), it can be assumed that disease transmission is dominated by close range airborne transmission. Therefore as an approximation surface contact transmission can be ignored, meaning the epidemic model can be restricted to the spread of disease along links in the contact network.

In general, the probability of a contagious individual passing the disease onto a susceptible person is dependent on the length of time the individuals spend together. Hence as well as capturing which people interact in a contact network, it is important to capture when and for how long people interact.

The Snowball Method, Koskinen et al. (2010), can be used to reconstruct contact network statistics through interviewing only a subset of the population. This is done by taking an initial sample of nodes, tracking which other nodes the initial sample is linked to, and repeating this process for another generation of nodes. These statistics can be used to simulate the entire network.

Due to practical reasons the sample interviewed for the Snowball Method should be kept to the minimum that yields good results. However for a disease contact network, if every short interaction a person experiences is counted and tracked, the node degree is very high, meaning that after two generations, the number of people to be tracked is of the order of $O(d^2)$ where $d$ is the average degree. This process could quickly become infeasible. A solution is to cut-down on which links are tracked, however any cut must be systematic and justified. It is expected that by far the most contacts would be those that last only a short time, Kostakos et al. (2007, 2010); Barrat et al. (2010), hence eliminating all contacts below a certain time period from consideration is the logical approach. However it was not understood if it is these very contacts that play the most important role in transmitting disease.

Previous work has studied the effect of eliminating the shortest and longest contacts from an epidemic simulation, Kostakos et al. (2007, 2010), however this work does not answer the key question because it assumes that the probability of transmission is 1 (hence all contacts, no matter how brief, succeeded in transmitting the virus), a reasonable assumption for the software virus they were chiefly simulating, but an unrealistic assumption for a biological disease.

This paper tackles the question of whether the large number of short contacts people have are more or less important for epidemic spread than their smaller number of longer duration contacts. As well as being of benefit for gathering interaction network data and epidemic modelling, a large benefit to studying this problem is that it answers an important question in public health: during an epidemic is the spread best controlled by quarantine measures that reduce brief encounters or persistent ones?[1]

The mathematical framework in this paper work draws on experimental evidence, based upon sensor data, that for many populations the number of contacts an individual has approximately follows a power-law distribution, and that the duration of these contacts also approximately follows a power-law distribution. This results in a probability distribution function (PDF) for the duration of contact between any two individuals in a population to be described by a power-law function Kostakos et al. (2007, 2010); Barrat et al. (2010).

The analysis, based upon the power-law approximation, is then compared to simulating epidemics over a very high-resolution temporal school contact dataset, consisting of sensor based data measured over the course of a school day, which measured how long any two individuals were within 3 m of each other (see supplementary material of Salathe et al. (2010)).

## 2 ASSUMPTIONS

This work makes some key assumptions.

- It is approximated that all contacts between individuals have the same probability distribution function (PDF) for duration of contacts (this ignores that some people are inherently more social than others and when simplifying to the level of whether a contact exists or not, results in a random social network with link duration described by a power-law, rather than any other type such as a scale free network, such as in Barabasi and Albert (1999)). As will be seen in section 5 this approximation does not reduce the accuracy of the results for calculating initial $R_0$.

---

[1] An example of reducing brief encounters would be quarantining people in their homes. An example of reducing persistent interactions would be asking people to stay in separate rooms when home.

- The disease being modelled requires the simultaneous presence of a contagious and a susceptible person for transmission (contagious people do not contaminate the environment for later infection of others).
- The immune system works in such a way that susceptible individuals have a constant chance per time unit of contracting the disease when interacting with an infectious person, Salathe et al. (2010).
- The approximation is made that all interactions have the same probability to infect per unit time (assuming an averaged type of interaction and an averaged immune response). This approximation should result in accurate results for large populations.
- That the power-law index for close contact durations measured by Kostakos et al. (2007) is relevant to this network.
- That all brief interactions have a cutoff in that none occur in less than a certain time window, i.e. it takes a non-zero time to move in and out of the distance of interest that defines a contact. All interactions of less than the cutoff are ignored, resulting in intermittency in the PDF.
- The interaction network shall have a daily cycle, leading to maximum interaction times of one day.
- That the index case remains infectious for the full simulated day when looking at infections in the first 24 hours.

## 3 PROBABILITY DISTRIBUTION FUNCTION OF CONTACT DURATION

From the assumptions the probability distribution function of contact persistence between individuals is taken to be a power-law distribution. At first we shall assume the power to be $b = 1.3$. This value was found to describe the distribution of cumulative contact times for proximity interactions based upon coincident bluetooth signals received from individuals phones at sentinel sensors, Kostakos et al. (2007). In section 5 this parameter shall be calibrated to be $1.17$ when applied to a higher resolution high-school dataset.

The distribution is assumed to be applied each day; i.e. each day for each possible contact, the PDF determines if the contact exists, and if it exists, how long it persists for. Beyond serving to limit the requirement for temporal resolution of the network, the lower cutoff in contact times avoids the singularity in the power-law spectrum at time $t = 0$.

To generate a PDF based on the power-law distribution, take a population of $N$ people with an average number of contacts per day of $c$. This information is used to scale the PDF $\Phi(t)$.

$$\Phi(t) = r\delta(t) + qt^{-b}\Theta(t - t_{\min})\Theta(t_{\max} - t) \tag{1}$$

where $r = 1 - c/N$. Scaling gives a $q$ of:

$$\int_0^{t_{\max}} \Phi(t)\, dt = 1 \quad \rightarrow \quad q = \frac{(1-r)(1-b)}{t_{\max}^{1-b} - t_{\min}^{1-b}} \tag{2}$$

For illustrative purposes we take the example of a population of $N = 3000$, with $c = 200$. We set $t_{\min} = 1$ min and $t_{\max} = 24 \times 60$ min. This gives $r = 14/15$ and $q = .023$.

This can be used to calculate how many contacts a person will have on average within a range of durations, the cumulative time of these contacts and the fraction of the total cumulative time compared to those of all contacts. The cumulative time for all contacts within a range of durations $[t_1, t_2]$ is:

$$\int_{t_1}^{t_2} dt\, \Phi(t)t = q\left[\frac{t^{2-b}}{2-b}\right]_{t_1}^{t_2} \tag{3}$$

The results are displayed in Table 1.

Contacts lasting less than 15 minutes and 30 minutes contribute respectively 3.5% and 6.1% of the total average contact time of an individual. From the perspective of modelling disease transmission, not including these contacts should not greatly affect the epidemic (especially considering that the model's infection probability will generally be calibrated to achieve a reasonable $R_0$) while eliminating 63% and 72% of the contacts in the network respectively, simplifying the task of capturing network data.

Similarly from the perspective of quarantining, cutting the few longest contacts can have a dramatic effect on reducing disease transmission. This shall be explored analytically next.

Table 1: Prevalence of contacts, binned by duration, by average frequency, total time taken for contacts, and fraction of time for these contacts compared to all contacts.

| Contact durations | Average frequency | Cumulative time (min) | Fraction of contacts by time |
|---|---|---|---|
| 1-5 min | 86 | 200 | 0.013 |
| 5-15 min | 39 | 350 | 0.022 |
| 15-30 min | 19 | 400 | 0.026 |
| 30-60 min | 15 | 650 | 0.042 |
| 1-2 hr | 12 | 1100 | 0.068 |
| 2-4 hr | 10 | 1700 | 0.11 |
| 4-8 hr | 8.2 | 2800 | 0.18 |
| 8-24 hr | 9.9 | 8400 | 0.54 |

## 4 PROBABILITY OF TRANSMISSION

The assumptions lead to a description of transmission equivalent to radioactive decay. The longer a sick person interacts with a susceptible one, the less likely the susceptible person is to escape infection. Thus the probability to infect $P$ for a contact of cumulative time $t$ is:

$$P(t) = 1 - (1-p)^t = (1 - e^{-t/T}) \qquad (4)$$

where $p$ is the probability to infect per time unit (Salathe et al. (2010)), $T = \tau/\log(2)$, and $\tau$ is the 'half-life.' In this case we have arbitrarily set $T$ to 1440 min (one day), equivalent to a transmission probability of 0.632 over a 24 hour contact. In practice $T$ is not known and depends on the details of contacts between individuals. However $R_0$ is generally approximately known and when using this methodology it is recommended that $T$ is chosen to calibrate the $R_0$ this methodology predicts.

Now assume one person is infectious in a sea of susceptibles, what is the effect on how many people will be infected when including all the contacts versus only those over 5 min, 15 min, 30 min and 1 hr in length? This is done by integrating the product of the PDF, the probability to infect and the susceptible population size.

$$I(t_1, t_2) = \int_{t_1}^{t_2} (N-1)\Phi(t)P(t)\,dt \quad = \quad (N-1)q\left(\frac{t_2^{1-b} - t_1^{1-b}}{1-b} - T^{1-b}\Gamma(1-b, t_1/T, t_2/T)\right) \quad (5)$$

$$\Gamma(1-b, t_1/T, t_2/T) \quad = \quad \int_{t_1/T}^{t_2/T} z^{-b}e^{-z}\,dz \qquad (6)$$

where $\Gamma$ is the generalised incomplete gamma function. For our simulated population, the effect of eliminating quick and long links are listed in Table 2. As can be seen eliminating all links under 30 min in duration (corresponding to 72% of links) only reduces the number of cases by 7.8%, indicating these links are not likely to be critical for modelling an epidemic (although to fully conclude this a complete knowledge of the time evolving network is needed, for example cutting these links could create persistent isolated subnetworks.)

It is also clear from the table that the few most persistent links correspond to most disease transmission. Eliminating all contacts over 4 hours reduces transmission by 68% while affecting only 9.1% of contacts. A reduction of 68% would reduce the $R_0$ for many diseases below 1.

It is interesting to consider the case of $T >> t_{\max}$ in this system. As $t_1 < t_2 < t_{\max}$, equation (5) approaches in this large $T$ limit:

$$I(t_1, t_2) \approx \frac{(N-1)q}{T}(t_2^{2-b} - t_1^{2-b}) \qquad (7)$$

In this case the effect of cutting all links between duration $t_2$ and $t_{\max}$ is to reduce $R_0$ by a factor of:

$$I(t_{\min}, t_2)/I(t_{\min}, t_{\max}) \approx \frac{t_2^{2-b} - t_{\min}^{2-b}}{t_{\max}^{2-b} - t_{\min}^{2-b}} \qquad (8)$$

Care should be taken when using this large $T$ approximation of $I(t_{\min}, t_2)/I(t_{\min}, t_{\max})$ that $b \neq 2$.

Table 2: Average number of infections over one day for a single index case when culling quick contacts (left two columns) and culling long contacts (right two columns).

| Contact durations | Average number of infections | Contact durations | Average number of infections |
|---|---|---|---|
| 1 min-24 hr | 9.0 | 1 min-5 min | 0.14 |
| 5 min-24 hr | 8.8 | 1 min-15 min | 0.38 |
| 15 min-24 hr | 8.6 | 1 min-30 min | 0.66 |
| 30 min-24 hr | 8.3 | 1 min-1 hr | 1.1 |
| 1 hr-24 hr | 7.9 | 1 min-2 hr | 1.8 |
| 2 hr-24 hr | 7.2 | 1 min-4 hr | 2.9 |
| 4 hr-24 hr | 6.0 | 1 min-8 hr | 4.7 |
| 8 hr-24 hr | 4.3 | 1 min-24 hr | 9.0 |

## 5  COMPARISON TO REAL DATA

In this section the previous epidemic behaviour is corroborated by modelling epidemic spread on a high resolution temporal contact network for a school. The network data is measured in Salathe et al. (2010). This data was gathered by placing wireless sensors on 95% of a school population and recording duration of contacts in units of 20 s (where a contact was defined by a period of signal strength corresponding to a separation of less than 3 m).

The results for fitting a power-law probability distribution to the duration of links in the school dataset are displayed in Figure 1. The result is:

$$\Phi(n) = \begin{cases} .761 & n = 0 \\ .0587\, n^{-1.17} & 4 \leq n \leq 200 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

for a population of 789. The power-law is in terms of discrete time units. The time unit is that of the data, equivalent to 20 s. Beyond n=200 the number of contacts drops off rapidly, hence is approximated in the distribution as having no links of duration greater than 200 units. It should be noted using the same duration PDF for each link in the network will result in a random network with link durations described by a power-law, where-as the authors of Salathe et al. (2010) identified the school network to be a small-world network. In this work where only the those infected by an initial seed case are measured, this difference will not matter.[2]

To corroborate the results of the previous section an identical epidemic with a corresponding $T = 4320$ units = 1 day was run over the network, by selecting an individual as the index case of an epidemic, calculating how many people they will on average infect over a day, and repeating this for the whole population to find the overall average. This is repeated when systematically culling links of duration $< n$ for all $n$ and $\geq n$ for all $n$. These calculations were then repeated based upon the probability distribution equation (9).

The average number of infected during the first day becomes the standard $R_0$ used in the literature, if the infectious period of an individual is set to the school day. The results for this analysis appear in Figure 2.

As can be seen from Figure 2, the power-law PDF model accurately reproduces the behaviour of the system for contacts of less than 200 units, when scaling the results by the maximum number infected. The figure also shows that links of greater than 200 time units do not play a large role due to their scarcity. Figure 2 (b) shows that links of short duration contribute little to the spread of disease, in that the numbers infected are only significantly reduced when most links (when starting with the briefest) are culled.[3] Subfigures (c) and (d) show that eliminating the longest duration links quickly reduces the relative expected number of infected, and hence could potentially reduce $R_0$ below 1 for many diseases.

---

[2]The different type of network would have a measurable effect if following the full course of an epidemic, however even in that case the network on a single day can only approximate the network on other days and hence a loss of fidelity would always occur when modelling the full time-course of the epidemic based upon the measured network.

[3]Related to this behaviour, due to the high average node degree of the network of 299.85, when culling the briefest links, isolated sub-networks are slow to form. Eliminating 90% of all links (corresponding to all links of 20 min and below), only separates 4.6% of nodes from the giant component of the network, supporting the finding that quick links can be ignored in epidemic modelling without great impact on the model.
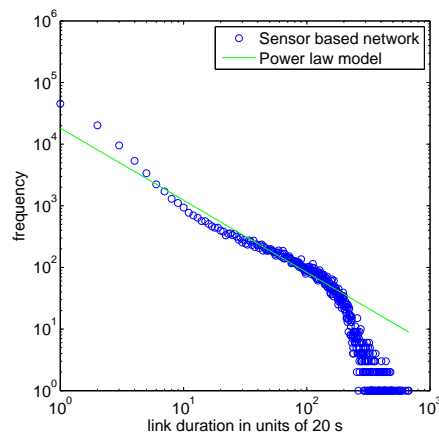
Figure 1: Fitting the power-law model for contact duration to the school dataset. The data closely follow a power-law for time units 4 to 200.

## 6   CONCLUSIONS

In this paper a framework for analysing the importance of quick and long interactions for respiratory disease transmission is presented, based on modelling the duration of contact lengths between individuals by a power-law. Experimental evidence for this profile is presented. The result is an argument that the many quick interactions people have do not play a large role in respiratory disease transmission, and hence are not critical to model in agent based interactions.

Conversely the evidence of the paper also indicates the $R_0$ of diseases can be reduced by a significant factor (of the order of 2 to 3) by eliminating the most enduring interactions (assuming that these interactions are not replaced by many shorter interactions).

In the limit that the time scale for disease transmission over an average interaction, $T$, for a specific disease (of equation (4)) is much greater than the duration of the longest interactions, a simple equation for estimating the reduction in $R_0$ by cutting the longest interactions is presented in equation (8).

The methodology is of most use if details of the contact network are known, such as the longest common interactions during a day, and exponent $b$ of the power law, however even without this level of knowledge the general principles derived are still useful.

The findings here assume that transmission of respiratory disease such as influenza is dominated by close range aerosol transmission, based on the analysis of Atkinson and Wein (2008). In truth while the analysis of Atkinson and Wein (2008) represents the best knowledge of the importance of different transmission routes, its results are not conclusive, hence surface contact transmission may play more a more significant role for some respiratory diseases. If so the effect of environment contamination would need to be added to this analysis.

This paper does not examine the difference between disease spread in the home and elsewhere. It is likely that interactions in the home will be on average some of the longest, per day, individuals would have, and also have a greater chance to infect per unit time. If so this would strengthen the argument for the longest interactions resulting in most respiratory disease spread.

### REFERENCES

Atkinson, M.-P. and L.-M. Wein (2008). Quantifying the routes of transmission for pandemic influenza. *Bull Math Biol 70*(3), 820–867.
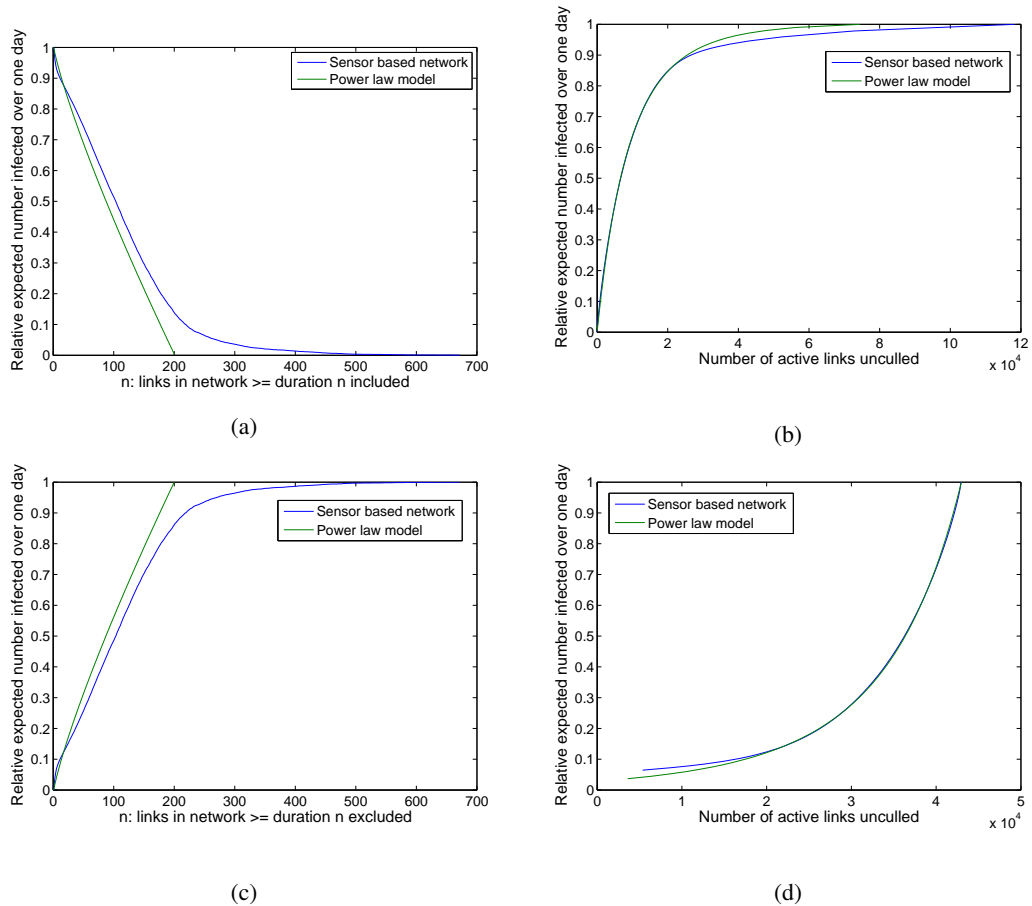
(a)

(b)

(c)

(d)

Figure 2: Numbers infected in the first day by an average index case (relative to peak value) versus: (a) the culling of all links below duration $n$; (b) the number of links still active when culling links from briefest to longest; (c) the culling of all links of duration $\geq n$; and (d) the number of links still active when culling from longest to briefest. The blue lines are for disease spread over links in the high-school data set and the green lines are for links based on the power-law model. When deriving these results for (c) and (d), links $\leq 1$ min long were ignored as the power-law model fit is not accurate for these very quick interactions. However these ignored links contribute only 4% to total infections, hence ignoring them does not alter the main message.

Barabasi, A.-L. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*(5439), 509–512.

Barrat, A., C. Cattuto, V. Colizza, J. Pinton, W. Van den Broeck, and A. Vespignani (2010). High resolution dynamical mapping of social interactions with active rfid. *PLoS ONE 5(7)*(e11596).

Koskinen, J., G. Robins, and P. Pattison (2010). Analysing exponential random graph (p*) models with missing data using bayesian data augmentation. *Statistical Methodology 7*, 366–384.

Kostakos, V., E. O'Neill, A. Penn, G. Roussos, and D. Papadongonas (2010). Brief encounters: Sensing, modeling and visualizing urban mobility and copresence networks. *ACM Trans. Comput.-Hum. Interact. 17*(1), 1–38.

Kostakos, V., E. O'Niell, and A. Penn (2007). Brief encounter networks. *CoRR abs/0709.0223*.

Salathe, M., M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences* (10.1073/pnas.1009094108).