Modeling the transmission of COVID-19 in small populations: A case study of Southern Methodist University

Daniel Chavez, Eli Laird, Brandilyn Stigler

August 2021

Abstract

In this paper we show how the canonical SEIR model may not be suitable for modeling the spread of SARS-CoV-2 within a small community and at short time scales. We propose a novel model that encapsulates the granularity of small population dynamics and highlight its two most unique features: (1) an exponentially decaying and sinusoidal transmission coefficient, which is a function of time, population, and relative risk the population as a whole takes on, and (2) a recovery coefficient that decays exponentially and is in sinusoidal phase with the transmission coefficient. The two rate functions are studied, and their predictive usefulness is shown when substituted for the constant coefficients that control population flux through the categories of the model. A simple stability analysis is performed and a proxy for a dynamic reproductive number is discussed.

1 Introduction

The canonical Susceptible-Exposed-Infected-Recovered (SEIR) model is a powerful system of ordinary differential equations (ODEs) that give predictive insight into epidemic dynamics. From their inception, this system of ODEs has been linked together through constant rate coefficients and the timescale of observation is dependent upon how fast the infection of interest propagates through the population via the force of infection, whether or not the population can recover, become re-infected or, in some cases, not recover at all. However, SEIR models break down in their useful capacity when applied to small-scale populations and for short timescales. Using non-constant transmission rates that are functions of time and additional variables was used successfully in [3] to enhance the validity and accuracy of simulations when compared to real data.

In that spirit, we highlight an analysis providing evidence that the transmission and, presumably recovery rates, of SARS-CoV-2 may not be constant in Section 3. We propose functions to describe these two rates, which we then incorporate into a SIR model in Section 5. We conclude with a discussion of the model including predictions.

2 Background: Standard SEIR Model

The standard *Susceptible (S) Exposed (E) Infected (I) Recovered (R)* or *SEIR* model is widely used within the study of epidemiology to model the interactions between individuals of different subsections of a population (for an introduction, see [5, 1]). The *susceptible* group of the population represents the individuals who have not yet contracted the disease but could possibly contract it in the future. The *exposed* group represents those who have come into contact with an infected person but have not yet developed symptoms of the disease nor are infectious. The *infected* group contains those of the population who have contracted the disease and have developed symptoms are infectious. Lastly, the *recovered* group of the population represents individuals who have recovered from the disease and are no longer experiencing symptoms. A standard *SEIR* model is defined by the following differential equations (see Equations 1) and depicted below (see Figure 1):

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{I(t)S(t)}{N} \\ \frac{dE}{dt} &= \beta \frac{I(t)S(t)}{N} - \delta E(t) \\ \frac{dI}{dt} &= \delta E(t) - \gamma I(t) \\ \frac{dR}{dt} &= \gamma I(t) \end{aligned} \tag{1}$$

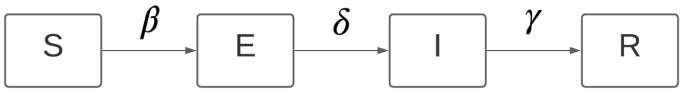


Figure 1: Standard SEIR diagram for Equations (1).

This compartmental model tracks the movement of individuals from one sub-group to another while conserving the total population N. Therefore at any given instance in time $S(t)+E(t)+I(t)+R(t) = N$. Individuals in a given sub-group transition to other sub-groups at rates defined by model parameters β,

δ, and γ. The parameter β represents the rate at which individuals transition from the *susceptible (S)* to the *exposed (E)* group and is commonly referred to as the *transmission rate*. The parameter δ represents the time period from when an individual is exposed to the disease to when that individual begins experiencing symptoms. This rate is often called the *incubation rate* and is generally defined as $\frac{1}{Latent\ Period}$, where *Latent Period* is in days. The parameter γ models recovery period or $\frac{1}{Days\ Until\ Recovered}$.

3 Applying an SEIR model to SARS-CoV-2 data

In this section we offer evidence that SARS-CoV2 cannot be described by an SEIR model by demonstrating that the transmission and recovery rates are not constant for data collected from the Dallas metropolitan area.

3.1 Parameters

To evaluate our model, we use data gathered from [4]. The parameters used to tune our SEIR model were estimated by referring to work done by [7]. In this study, we estimated the incubation period to be an average of 7 days, resulting in an incubation rate of 0.14. The recovery rate is estimated to be about 21 days or a rate of 0.048. All parameters can be found in Table 1 below.

Parameter	Value
Transmission Rate (β)	0.4
Incubation Rate (δ)	0.14
Incubation Period ($\frac{1}{\delta}$)	7 days
Recovery Rate (γ)	0.048
Recovery Period ($\frac{1}{\gamma}$)	21 days

Table 1: SEIR model parameters gathered from [7].

3.2 Data

To determine the appropriate parameter values to fit the standard model, we fit the model to confirmed COVID-19 cases on the Southern Methodist University (SMU) campus, cumulative COVID-19 cases in Dallas county, and cumulative cases in the United States as a whole. The SMU dataset [10] was updated daily and spans from August 16 to the end of the 2021 Spring semester. This dataset records the number of cumulative cases and active cases of students and faculty on SMU campus. The data also provides the number of individuals who are

quarantined or isolated in on-campus isolation dorms or elsewhere. From this dataset, we used the cumulative and active case counts to fit a standard SEIR model and to evaluate the model parameters. The cumulative case count logs the total number of COVID-19 cases that have occurred since August 8th 2020. The active case count keeps track of the total amount of people who currently are experiencing symptoms of COVID-19.

The data on the Dallas County COVID-19 cases [9] was provided by Johns Hopkins University Center for Systems Science and Engineering. This dataset is updated hourly by the Associated Press and offers the cumulative COVID-19 cases and death rates for Dallas County and the US. The table from [9] that contains the Dallas data is named `2_cases_and_deaths_by_county_timeseries` and contains values for

- `state`
- `date`
- `total_population`
- `cumulative_cases`

From this dataset, we utilized the `cumulative_cases` and `total_population` data values to fit our SEIR model. Since the SMU dataset starts on August 8, 2020, we filtered the Dallas County dataset to only include data from August 8 onwards. The source [9] includes data for all counties in the US. To filter only the Dallas County data, we used the search query below.

```
SELECT * FROM 2_cases_and_deaths_by_county_timeseries
WHERE state="Texas" AND location_name="Dallas"
```

The United States COVID-19 data, also provided by John Hopkins University, is found in the GitHub repository found in [2]. This repository aggregates data from various sources including the World Health Organization (WHO), the European Centre for Disease Prevention and Control (ECDC), and the United States Center for Diseases Control (CDC).

3.3 Analyses

Using the datasets described in the previous section, we approximated the parameters of our custom model by fitting each dataset to a standard SEIR model [1]. Using the popular Python library SciPy [11], we were able to determine the best fit parameter values using a non-linear least squares method.

To estimate β and γ for the SMU SEIR model, we made use of published values from [7] to establish a baseline for comparison, then we fit a standard

SEIR model to the SMU, Dallas County, and United States data for two distinct time periods; one being a 40-day period and the other a 171-day period. These time periods were chosen based on the amount of SMU data that was available. As a result of this process, we discovered surprising evidence suggesting that the model parameters, β and γ, in a standard SEIR model are dynamic rather than constant.

We made use of SEIR parameter values gathered in a study of the spread of COVID-19 in Wuhan, China [7] as a baseline to determine whether our fitted values for the SMU, Dallas, and US data were reasonable in magnitude.

When fitting the SEIR model to the SMU data with the non-linear least squares optimization, we obtained β and γ values of 0.99 and 0.25 respectively for the 40 day time period. There are two problems with these values. The first issue is that the fitted β and γ values vary significantly when compared to parameter values of 0.4 and 0.048 found in [7]. The second issue is with the β value of 0.99: in the least squares optimizer, we fixed the interval of β and γ parameters to be between [0,1]. We found, unsurprisingly, that too few data points in combination with an insufficiently long time-scale resulted in the β value reaching the boundary of the interval. Were no fixed interval to exist, this would mean convergence to reasonable value would never occur for the β coefficient.

There are a few possible explanations for this failed model fit. The significant difference in size between the SMU population of 10,000 and the Wuhan population of around 11 million suggests that our population size was much too small to effectively model COVID-19 with a SEIR model. A second explanation may be that the length of time for data collection is too small. A third explanation is that the parameters for this disease are not constant (for examples of models with varying parameters, see [3, 6, 8]).

To test the first hypothesis, we fit an SEIR model to the relatively larger population of Dallas. With a population of around 2.5 million, we expected the SEIR model to outperform the SEIR model fit on the SMU data and to more closely resemble the model fit to Wuhan data in [7], however; this fit resulted in β and γ values of 0.09 and 0.05 respectively for the 40-day time period; see Table 2 and Figure 2. Here the β values for Dallas in the 40-day period are roughly 4 times smaller that the Wuhan β values respectively. While the Dallas population is significantly larger than the SMU population, the difference between the Wuhan and the Dallas populations still remains large at around 8.5 million, suggesting that even the Dallas population is too small to fit a SEIR model.

Then we fit an SEIR model to United States COVID-19 cases. The model fit to the United States data produced β and γ values of 0.087 and 0.059 respectively for the 40-day time period. Here the β value is 5 times smaller than

Model	Cumulative Cases	Days	Beta	Gamma
SMU	534	40	0.99	0.25
SMU	990	171	0.58	0.35
Dallas	83306	40	0.090	0.05
Dallas	262738	171	0.0273	2.31E-10
US	7052359	40	0.087	0.059
US	26436155	171	0.029	2.00E-10

Table 2: Beta (β) and Gamma (γ) parameter values for SEIR models fit on SMU, Dallas, and US datasets.

the value for Wuhan.

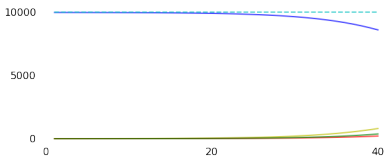
Next, we tested the second hypothesis by varying the time period of the data used to fit the model. For the previous test, which varied the population size, the time span of the data was only 40 days. For this test, the data spanned 171 days. For the SEIR model fit to the 171-day for SMU, the β and γ values were 0.58 and 0.35 respectively. Here the SMU data's β value is roughly 45 percent larger than the Wuhan value of 0.4 and the γ value is roughly 7 times larger than the Wuhan value. When fitting to the Dallas 171-day data, the β and γ values were 0.0273 and $2.31e-10$ respectively. Compared to the Wuhan values, the Dallas β value was roughly 15 times as small and the γ value was about 2e8 times smaller. Lastly, the US data for the 171-day period resulted in β and γ values of 0.029 and 2e-10. Here the β value is 13 times smaller than Wuhan's value while the γ value is 2.4e8 times smaller.

Considering that the application of small, medium, and large population sizes and the short and long time spans of 40 and 171 respectively failed to produce β values that closely matched the magnitude of the values for the city of Wuhan, we suggest that the parameters for the SEIR model are not constant.

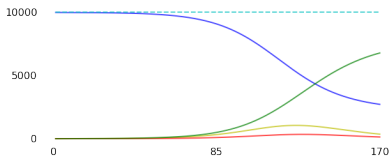
In summary, the canonical SEIR model is a poor fit for the COVID-19 virus when analyzing small populations at short timescales, as its limitations of constant parameters cannot capture the true transmission dynamics.

4 Proposed Novel Model for SARS-CoV-2

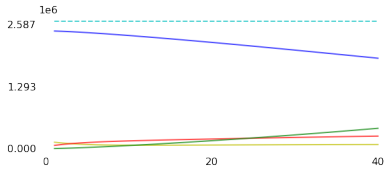
In an effort to more realistically model the spread of the virus through a micro scale scenario i.e. a university campus or other similarly sized population, the SIR model must be extensively modified both categorically and with respect to its group composition. To increase the granularity of analysis, the overall population has been split into two groups: commuters, which include off-campus students, employees, and non-residential faculty; and residents, which consist



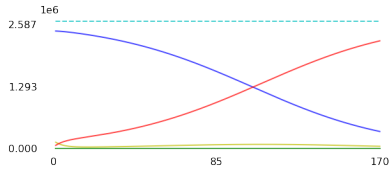
A. SMU 40 Days



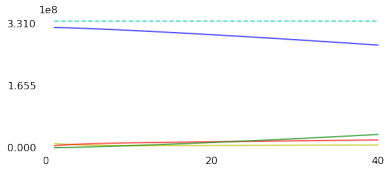
B. SMU 171 Days



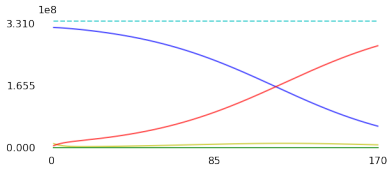
C. Dallas 40 Days



D. Dallas 171 Days



E. US 40 Days



F. US 171 Days

Figure 2: Each image A-F plots the standard SEIR curves for SMU, Dallas, and US populations. Plots A, C, and E plot each population's curve for a 40 day time period. Plots B, D, and F plot each population's curve for a 171 day period. The curves included are the susceptible (blue), infected (red) and recovered (green) curves. Note that the y-axis scale for plots C, D, E, and F is in millions of people, while the scale for A and B is the actual number of people.

solely of on-campus students. This bifurcation provides deeper insight into both groups and their behaviors and relevant parameters. Both groups are composed of five categorical states: susceptible (an individual yet to be infected), exposed (a susceptible individual having made contact with an infectious non-isolated, non-quarantined individual), infectious (an infected individual capable of propagating further infection to the larger population), quarantined and isolated (individuals in quarantine or in isolation, assuming they have been infected; however they do not interact with the larger population and are rendered null vectors), and lastly recovered (recovery is a suitable category in this case rather than removed, as no one from the SMU community died as a result of infection).

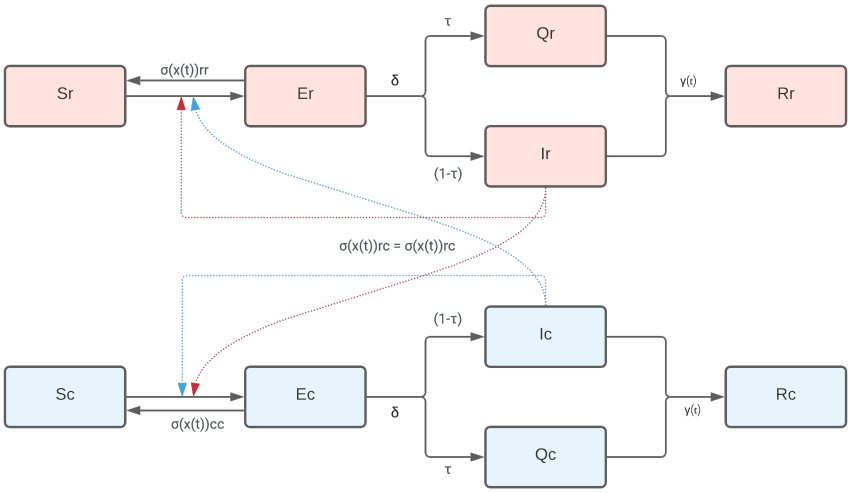


Figure 3: Model Wiring Diagram: Two populations are considered: residential students (red nodes in the top row) and commuter students (blue nodes in the bottom row). $S_{\_}$ refers to susceptible where Sr is the susceptible residential population and Sc is the susceptible commuter population. Similarly $E_{\_}$, $Q_{\_}$, $I_{\_}$, $R_{\_}$ refer to exposed, quarantined and isolated, infected, and removed respectively, and the relevant population by denotation of c or r. Black arrows show the flux of the populations through through their respective states, while the dashed arrows represent the possibility of intra- and inter-group transmission. Note $\sigma(x(t))$ is the exposure function, δ is the incubation rate, τ is the rate of isolation/quarantining, and $\gamma(t)$ is the recovery function.

This model places a heavy emphasis on what canonical models consider to be β, the transmission turnover rate as a result of susceptible portion of a population interacting with the infected fraction. Many efforts have been made to reliably measure β, but there exists no method for a population the size of a relatively small university campus. In an effort to more accurately model

transmission at the small-scale, the canonical β has been replaced with an transmission function, $\sigma(x(t))$, with the hopes of accounting for the granularity of such a small population. The transmission function is dependent on $x(t)$, where at the individual scale, $x(t)$ represents the normal amount of persons an individual might come into infectious proximity of as a function of time. The continuous transmission function $\sigma(x(t))$ is right-shifted by a value η, the number of persons an individual comes into contact with regularly, with a maximal transmission value of 1, and minimum value of zero; as the risk a person takes on increases, $\sigma(x(t))$ increases as, signifying high exposure risk sensitivity to incremental persons encountered. As relative risk lessens, $\sigma(x(t))$ decreases, showing less incremental risk taken on per person encountered. Risk in this model is represented by a computed parameter $\bar{r}$, and as this parameter tends to zero, so does $\sigma(x(t))$.

In the SEIQR model, persons encountered are assumed to be quasi-infectious and homogeneously mixed; that is persons encountered may or may not have the virus. What matters here is the risk. The assumption is that the rate at which the virus spreads is proportional to a function of the relative risk an individual takes on.

4.1 Defining the risk function

In this section, we define a function, called the *risk scalar function* $\bar{r}$, that describes the risk for an individual in terms of physical area, amount of time spent in an area, population density, proximity to other individuals, and proportion of individuals wearing masks in a given area. The computed risk value is then used as constant parameter in the $\sigma(x(t))$ function.

An individual's baseline risk is established by the average number of same persons they frequently interact with on a daily basis. If every individual within the population were treated as a node, then every edge within a reasonable distance (the high-risk infectious threshold being less than six feet according to the CDC) could be considered a feasible transmission pathway between individuals. From the simple quadratic relation,

$$f(\eta) = \frac{\eta^2 - \eta}{2} \tag{2}$$

where $f(\eta)$ represents the total amount of edges between η nodes, it is easy to calculate how many possible transmission routes a virus could take locally among persons in a susceptible populations given the introduction of infectious nodes. If on average a person comes into contact with the same η individuals on a daily basis, a $f(\eta)$ transmission pathway network exists, where a virus poses the risk of infection via a transmission pathway. Additional risk is taken on when a susceptible individual chooses to interact with an environment that

has a higher population density which correlates to an increased exposure to many more potential transmission pathways. If individuals choose to expose themselves to a new environment, with an area A and population density ρ, the number of persons they could encounter is $A \times \rho$ with the number of transmission pathways

$$\frac{\rho^2 A^2 - \rho A}{2}. \tag{3}$$

Risk, represented later by the risk scalar function $\bar{r}$, is proportional to the relative percentage increase of transmission pathways a person is exposed to, with relative percentage found by dividing Equation 3 by Equation 2, yielding

$$\frac{\rho^2 A^2 - \rho A}{\eta^2 - \eta}.$$

If

$$(\rho^2 A^2 - \rho A) = (\eta^2 - \eta),$$

then a person takes on no additional risk from his established baseline exposure.

There is a distinction between total transmission pathways and local network pathways an individual is exposed to. Within a present time frame, one cannot truly know if exposure to an additional pathway will or will not carry a pathogen; because of this, the total new pathways an individual is exposed to are accounted for, as this represents the additional risk the individual takes on. Total transmission pathways are not all feasible routes of transmission, nor are they all active; pragmatically, only a very small percentage are actually active. Total transmission pathways, or *Edges*, $[E](t)$, are given by:

$$[E] = \frac{(S+I)^2 - (S+I)}{2} - \frac{S^2 - S}{2} - \frac{I^2 - I}{2} = SI$$

The susceptible and infectious networks are subtracted out of the *Edges* because neither group demonstrates intra-group transmission. The resultant equation simplifies down to a simple two-group network. New cases that arise from the previous day's edges represent the apparent percent activation of those edges *or* the total active transmission pathways from the previous time-frame. The apparent activation of possible transmission pathways is a metric that can only be calculated with retrospect, which makes such calculations for future use impossible. To mitigate this, it is assumed that only local transmission pathways within six feet are considered active. Quantitatively, this is represented by the *proximity constant* p_c within the the risk scalar function $\bar{r}$.

Canonical models dictate that rates used for population flow from one category to the next be constant; however, the behavior of the virus at the micro-scale does not behave in such a fashion. Rather than a fixed rate constant, the transmission function $\sigma(x(t))$ is used. Additionally, because it is the magnitude

of the risk that dictates the rate at which persons flow from S to E categories, if an individual takes on $\sigma(x)$ risk in an environment (likelihood of exposure), then the likelihood they were not exposed is represented by $(1 - \sigma(x))$, which is shown by the ability of *Exposed* persons being able to flow back to the susceptible population.

The exposed pathway then feeds into two categories I and Q dictated by an actual incubation rate α, with directional control given by τ, where τ represents the percentage of exposed persons put into quarantine or isolation (Q category). The unique feature of this model is that it is assumed that the Q categories from either population, although infectious, do not come into contact with the susceptible populations. Both Q and I flow into the R category at a rate γ. Because this model is multi-group, (split into residential and commuter populations), each group has five respective ODEs, with interaction possible in the event any infectious person from either population comes into contact with any susceptible person from either population.

While this particular compartmental model will not be further analyzed, its two functions, $\sigma(x(t))$ and $\gamma(t)$ will be investigated further to show their usefulness and predicative capability.

4.1.1 The risk function at the micro- and macro-scales

At the individual scale, the $\sigma(x(t))$ is defined as

$$\sigma(x(t)) = \frac{e^{\bar{r}(x(t)-\eta)}}{1 + e^{\bar{r}(x(t)-\eta)}} \left(\frac{2e^{\bar{r}x(t)}}{1 + e^{\bar{r}x(t)}} - 1 \right),$$

and shows the risk a person takes on relative to their normalized η value. As described in the beginning of section 4.1, $\bar{r}$ is a function of time. An individual will frequent different areas, denoted A_i, that each have their own associated risk value, r_{A_i}.The greater the risk scalar function value $\bar{r}$,

$$\bar{r} = \sum_{i=1}^{n} (r_{A_i} \cdot t_{A_i})$$

the greater the incremental risk taken on by the individual. Every component within the $\bar{r}$ function is from the individual's perspective, and accordingly is applied to each area the individual frequents. The final $\bar{r}$ value generated is the weighted average (by percent of day spent in an area) of all the component r_{A_i} values,

$$r_{A_i} = \frac{(\rho_{A_i}^2 {A_i}^2 - \rho_{A_i} A_i)}{\eta^2 - \eta} \cdot p_{c_{A_i}} \cdot (1 - m_{A_i}),$$

where $p_{c_{A_i}}$ is the *proximity constant* which conveys the percentage of edges or transmission pathways of a given length in the area of interest, and m_{A_i} is the percentage of individuals in an area A_i that are wearing masks.

Area of Interest	F	ρ	A	p_c	m	$\%t$
Dorm room	1,468	0.0092	189.1	0.4053	0	35
Lab	50	0.0134	823.4	0.1189	0.99	2
Lecture	50	0.0222	882.7	0.1118	0.99	8
Auditorium	10	0.0176	2,914	0.03703	0.99	2
Gym	1	0.0118	9000	0.01254	0.5	4
Dining Hall	2	0.0513	29,816	0.00389	0.1	8
Library	1	0.0278	144,625	0.000816	0.75	17
Going Out	1	0.25	3000	0.1852[1]	0	21
Outside	1	-	∞	0	0	2

Table 3: Components of the risk scalar function: F refers to the area incidence, $\rho = \frac{people}{ft^2}$, A is area of the region given in ft^2, p_c the percent of transmission pathways under 6 ft, m = proportion of people wearing masks and $\%t$ refers to the amount of time people are in the given area.

However, modeling interactions at the individual scale, and then applying those interactions uniformly across the entire population assumes everybody within the population also behaves uniformly. In reality, we know every individual has a characteristic η value, and it follows that every individuals' characteristic $\sigma(x(t))$ is shifted to their respective η value.

Accounting for all behaviors is detrimentally complex, so rather than build up from the individual scale, a solution might be to view behavior from a campus perspective, where the only constraint on population behavior is the time spent in various areas, because η is fixed to the amount of the University population on campus. From this perspective, r_{A_i} represents the fraction of relevant transmission/contact pathways in a given area type relative to the entire campus network. To maintain congruity between perspectives, an additional component F representing the incidence of a given area type (*i.e.* 1500 dorm rooms, 2 dining halls, etc.,) must be added to the r_{A_i} function for the campus' perspective:

$$r_{A_i} = F \cdot \frac{(\rho_{A_i}^2 {A_i}^2 - \rho_{A_i} A_i)}{\eta^2 - \eta} \cdot p_{c_{A_i}} \cdot (1 - m_{A_i}).$$

Summing across time produces a time-weighted expression introduced at the beginning of the section, which we call the *risk scalar function.*

As $\bar{r}$ approaches zero, so too does $\sigma(x(t))$. This intuitively makes sense as reduction of risk-laden behaviors would lower the transmission rate. Lower social densities, higher mask-usage rates, and less time spent in high-risk areas all contribute to the lowering of the $\bar{r}$ value.

[1]When going off-campus, the proximity constant was assumed to be at least 5 $\times$ higher than usual (Table 3).

4.2 Derivation of the proximity constant

Lastly the generation of the *proximity constant*, p_c, for a given area can be found by integrating the probability density function from 0 to L, (the edge length of interest), for an arbitrary amount of randomly distributed nodes in a given square area, A. This process is of interest as every relevant area on campus was approximated to be a square for simplicity.

The distance between two nodes on a Cartesian plane may be given by:

$$\ell = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Let x_1, x_2, y_1, and y_2 be independent and pulled from identical normal distributions. The difference of two independent standard uniform random variables is well known to have a standard triangular distribution. Were the expectation value of the distance between randomly distributed nodes to be calculated for a square area of one, we would have the following:

$$\langle \ell \rangle = \int_0^1 \int_0^1 \sqrt{(\Delta x)^2 + (\Delta y)^2} 2(1 - \Delta x) 2(1 - \Delta y) d\Delta x d\Delta y$$

The solutions to "square-line picking" problems are well known [12] (see the Appendix for an alternate method of computing), and the exact probability density functions are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$; however, while the PDFs are exact, they are pieced together, and are ostensibly continuous to the unaware eye. For the purpose of simplicity, a function was created to approximate the true PDF for edge lengths within a square. The function contains two fitting constants C and k. Starting with the appropriate normalized triangular distribution, the derivation of the approximating function is found as follows:

$$\int_0^L N(kL - \Delta x) dx = 1$$

$$N \left[kL\Delta x - \frac{\Delta x^2}{2} \right]_0^L = 1 \longrightarrow N = \frac{1}{L^2(k - \frac{1}{2})}$$

To find the function that approximates the PDF up to L, the side-length of the square with area A:

$$\int \int N(kL - \Delta x) N(kL - \Delta y) d\Delta x d\Delta y$$

Let Δx and Δy be represented by x and y respectively,

$$C(x,y) = \int\int N(kL-x)N(kL-y)dxdy = N^2\left(k^2L^2xy - \frac{kLx^2y}{2} - \frac{kLy^2x}{2} + \frac{x^2y^2}{4}\right)$$

Assuming the triangular distributions are identical, then we can make the substitution $x = y$. Therefore

$$C(x) = N^2\left(k^2L^2x^2 - kLx^3 + \frac{x^4}{4}\right),$$

and the PDF may be found by taking the derivative with respect to x,

$$P(x) = N^2\left(2k^2L^2x - 3kLx^2 + x^3\right).$$

Finally the tuning constant, C, will be added to better the fit,

$$P(x) = C \cdot N^2\left(2k^2L^2x - 3kLx^2 + x^3\right),$$

where k and C were found to be 1.09 and .975 respectively through experimental simulations. The approximating function can be made a function of A, the square area of interest where $L^2 = A$, and for each area of interest within the model the generalized function is:

$$P_{A_i}(\ell) = \frac{C}{A_i(k-\frac{1}{2})^2}\left(2k^2A_i\ell - 3k\sqrt{A_i}\ell^2 + \ell^3\right),$$

where ℓ is the edge/transmission pathway length. For every area of interest there is a p_c value that essentially confers the percent of edges that are high-risk *in* that area. The motivation behind this computation can be appreciated when areas with higher levels of social density are taken into account. The number of edges is proportional to $(\rho_{A_i}^2 {A_i}^2 - \rho_{A_i}A_i)$, and were an arbitrary or poor estimate of p_c to be used, $\bar{r}$ would be very inaccurate. p_c is then:

$$p_c = \int_0^6 P_{A_i}(\ell)d\ell$$

5 Mathematical Analysis: Representing R_t

The reproductive number R_0 can be calculated when the model rates are constants. In our case as the transmission and recovery rates are functions, we

develop R_t as a dynamic (in time) representation of the reproductive number.

Using canonical equations, dividing $\frac{dI}{dt}$ by $\frac{dS}{dt}$ and then separating variables yields:

$$\int_0^t dI = -1 \int_0^t ds + \frac{N}{R_0} \int_0^t \frac{1}{S} ds,$$

and when integrated, we arrive at the following time independent equation:

$$I_t - I_0 = -(S_t - S_0) + \frac{N}{R_0}(\ln(S_t) - \ln(S_0))$$

$$I_t = I_0 + S_0 - S_t + \frac{N}{R_0} ln\left(\frac{S_t}{S_0}\right)$$

However, when applied to a university setting these equations are not practical for numerous reasons. These equations rely on canonical assumptions to retain their validity when used for predictive calculations. Additionally, normal SIR models account for only one wave of an infection, so to even attempt a predictive calculation for a whole academic year, two calculations would be needed for two semesters taking into account two separate initial conditions; even then, the results would still be poor. Similarly, if we were interested in hypothetical outbreak size (assuming that the infectious population started at 0 and will end at 0, that S_0 is approximately equal to N, and that the outbreak size can be given in terms of the fraction of the population):

$$I(t) \Longrightarrow I(\infty)$$

$$S(t) \Longrightarrow S(\infty)$$

$$I_\infty - I_0 = S_0 - S_\infty + \frac{N}{R_0}\left(\ln\left(\frac{S_\infty}{S_0}\right)\right),$$

$$1 - \frac{S_\infty}{N} = f,$$

(R_0 can be solved for as a function of the fraction of population succumbing to the infection overall)

$$R_0 = \frac{-\ln(1-f)}{f},$$

we would need to calculate two fractions, one per semester, or whatever unit of time of interest. The sum of the fractions would theoretically give the total outbreak size, but that is only under canonical conditions, so in reality, the resultant calculation would be null as well. A dynamic reproductive number R_t is more appropriate for a small population within a short time-scale. Such a function would take into account the "surge-like" nature of a disease, seeing as even at the university-scale, at least two actively-infectious peaks were observed. Such a function would assume the form:

$$R_t = \frac{\sigma(x(t))}{\gamma(t)}.$$

The function σ takes an input of people frequenting campus from the population at time t according to the function $x(t)$, outputting a transmission value. The value $x(t)$ would necessarily rise from and go to zero at the beginning and end of each day respectively, as generally all people rise and sleep with the same periodicity (locally); across the course of a semester the function as a whole would decay, because people are removed from the susceptible population by infection, leaving less individuals for the virus to spread to, and the force of infection would periodically decay as well, making it harder for the virus to spread within the campus community. Finally an element of periodicity would be added to account for the first and subsequent smaller waves of the virus through the community at large. A *trial* $x(t)$ function could be:

$$x(t) = N \cdot sin^2(\pi t)e^{-t/k} \cdot sin^2(\omega t)$$

The expression $sin(\pi t)$ ensures the function goes to zero at the end of each day, $e^{\frac{-t}{k}}$ ensures the function decays as time progresses with k being time in days, and the $sin(\omega t)$ allows for the simulation of the "waves" of the virus, with ω controlling the number of surges as seen in Figure 4. The $sin(t)$ functions are squared to ensure no negative values. A value of 0.0571 for ω will be derived in the *Discussion*. As ω increases, the crests in the $\sigma(x(t))$ function become closer together, giving the appearance of a constant coefficient, akin to the classical β.

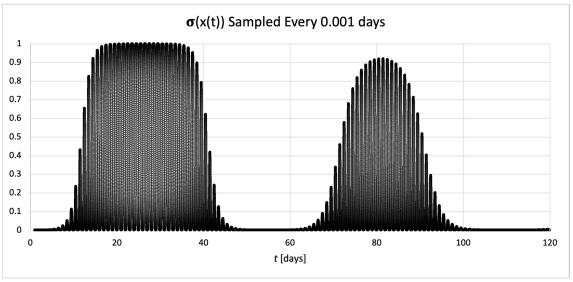
At the micro-scale $\gamma(t)$ is not constant either. As viral surges occur, the number of people of recovering will periodically rise and fall in phase with surge-timing. This periodicity, controlled by $sin^2(\omega t)$, must rise and fall to a maximum value, γ_m [to be interpreted as the fastest recovery period possible]. Those with the best health will of course on average recover much faster and gain effective-immunity to the virus sooner than others, which means that over the course of a semester, the recovery period will gradually become longer, because those with poorer health are still recovering. A *trial* $\gamma(t)$ function could be:

$$\gamma(t) = \gamma_m \cdot sin^2(\omega t)e^{\frac{-\gamma_m t}{k}}$$

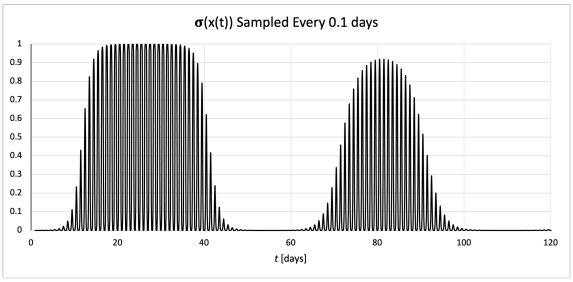
The dynamic reproductive number can then be tracked as described earlier. Classically, $R_0 = \frac{\beta}{\gamma}$, but this can also be interpreted geometrically as $R_0 = \frac{\int_0^t \beta dt}{\int_0^t \gamma dt}$, as canonically, β and γ are both constant coefficients. Congruently, if we take:

$$R_k = \frac{\int_0^k \sigma(x(t))dt}{\int_0^k \gamma(t)dt}$$

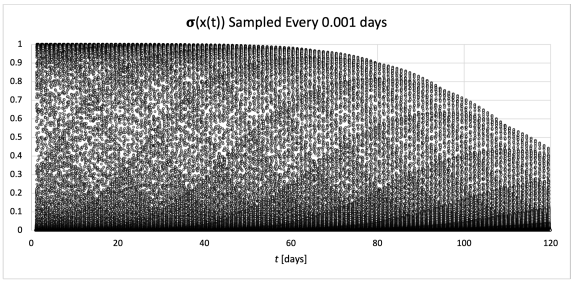
and $R_k \cdot S_0 > N$, then an epidemic will occur within the community. A more rigorous and intensive extension of this intuition can be seen by a stability analysis; for the simple SIR system substituted with $\sigma(x(t))$ and $\gamma(t)$, the equations and characteristic equation of its Jacobian matrix are [for a cleaner look, all $F(t) = F$]:



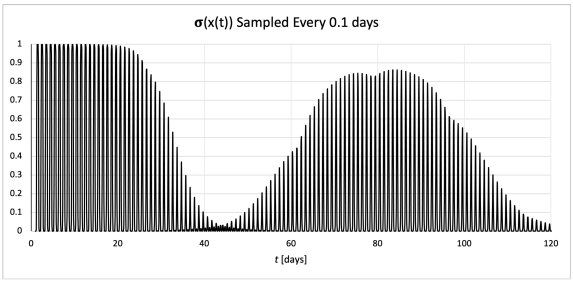
A: $\omega = 0.0571$



B: $\omega = 0.0571$



C: $\omega = 100,000$



D: $\omega = 100,000$

Figure 4: For all $\sigma(x(t))$ plots parameters are: population size $N = 10,000$, average persons present on campus $n = 3,500$, time period $k = 110$ [semester length in days], and risk scalar $\bar{r} = 0.001949$ [calculated from data in 3]. For **Plots A** and **B** $\omega = 0.0571$ and for **Plots C** and **D** $\omega = 100,000$. For very small values of ω, sampling interval *does not* make a noticeable difference in function appearance; however, for very large ω values, sampling interval creates a very noticeable difference in function appearance. At very large ω values, $\sigma(x(t))$ appears to become constant (only being forced to decay by the negative exponential term) as the time gap between each crest shrinks, approaching the behavior of the canonical β transmission coefficient which is constant. **Plots B** and **D** are meant to mimic two viral surges within a semester; ω and sampling-interval in the case of **Plot D**-were chosen towards that end. For very large ω values, sampling at discrete and equally spaced time intervals must occur, otherwise canonical results will be observed (only one peak and very inaccurate results for small populations and time-scales).

$$\frac{dS}{dt} = -\sigma S\frac{I}{N}$$
$$\frac{dI}{dt} = \sigma S\frac{I}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

$$(I_3\lambda - \mathbb{J}) = \begin{bmatrix} \lambda + \frac{\sigma I}{N} & \frac{\sigma S}{N} & 0 \\ \frac{-\sigma I}{N} & \lambda - \left(\frac{\sigma S}{N} - \gamma\right) & 0 \\ 0 & -\gamma & \lambda \end{bmatrix} = 0$$

with determinant

$$\lambda^3 - \lambda^2 \cdot \left(\frac{\sigma}{N} \cdot (I - S) + \gamma\right) + \lambda \cdot \left(\frac{\sigma\gamma I}{N}\right) = 0.$$

Assuming that at the beginning of the semester, an initial coordinate

$$E_0 = (S(t_0), I(t_0), R(t_0)) = (\approx N, \approx 0, 0)$$

exists, the determinant simplifies to

$$\lambda^3 + \lambda^2 \cdot (\gamma(t) - \sigma(x(t))) = 0$$

$$\lambda_1 = (\sigma(x(t)) - \gamma(t)), \rightarrow (\lambda_2, \lambda_3) = 0.$$

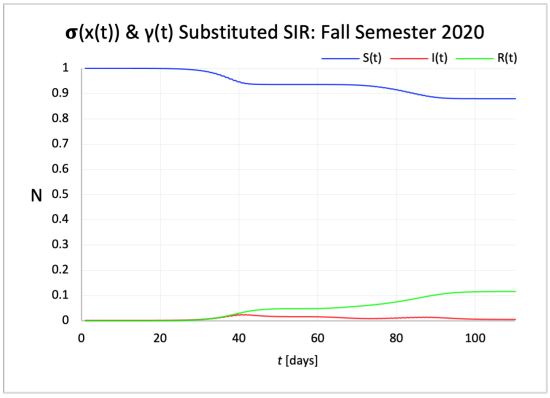
At t_0 however, both $\sigma(x(t))$ and $\gamma(t)$ will be zero due to the $sin^2(\pi t)$ term, hence it is intuitive to extend the meaning of the eigenvalue solution as follows:

$$\lambda_1 = \int_0^k \sigma(x(t))dt - \int_0^k \gamma(t)dt$$

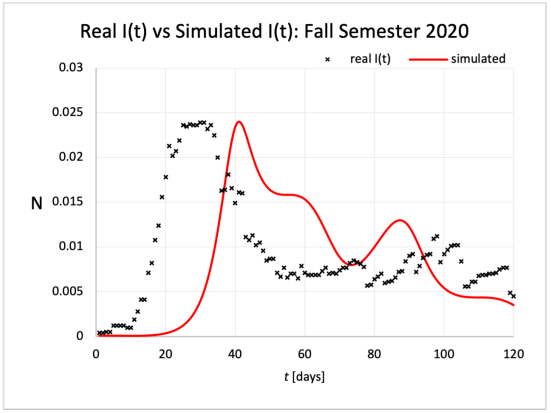
An *independent* and congruent conclusion was found in [8]. If we broaden our view to the total duration at point t_k and $\lambda_1 < 0$ then no significant outbreak will occur and the system will essentially be asymptotically stable at the starting point. Replacing β within the canonical SIR model with $\sigma(x(t))$ and γ for $\gamma(t)$ shows an immediate improvement in predictive capability not possible with constant parameters as shown in in Figure 5.

6 Discussion

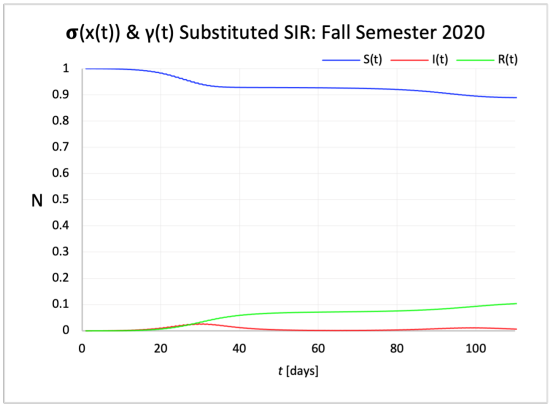
Here we proposed that canonical compartmental models with constant parameters were ill suited for modeing the spread of SARS-CoV-2 on a college campus



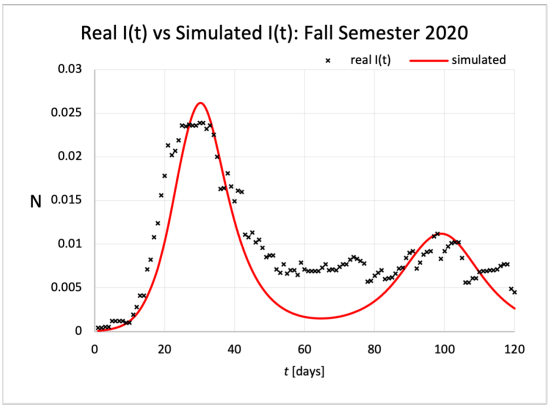
A: $\omega = 0.0571$; sampled every 0.1 days



B: $\omega = 0.0571$; sampled every 0.1 days



C: $\omega = 100,000$; sampled every 0.1 days



D: $\omega = 100,000$; sampled every 0.1 days

Figure 5: For all plots parameters are: population size $N = 10,000$, average persons present on campus $n = 3,500$, time period duration $k = 110$ [semester length], risk scalar $\bar{r} = 0.001949$, $\gamma_m = 0.2941$ for **Plots A** and **B** [8.4-day avg. recovery period], and $\gamma_m = 0.2817$ for **Plots C** and **D** [8.3-day avg. recovery period]. **Plots A** and **C**: SIR plots were created using Euler-integration with a time step of 0.1 and using the $\sigma(x(t))$ values being sampled every 0.1 days; S_0, I_0, and R_0, were set to 0.9999, 0.0001, and 0 respectively. **Plots B** and **D**: zoomed in y-axis.

on a short timescale. To address this we developed an ODE-based model for the spread of the disease for the Southern Methodist University campus, accounting for both residential and commuter student populations. The model includes *variable* transmission and recovery rates, from which we derived a dynamic expression R_t representing the reproductive number.

The practicality of R_t is that it gives insight as to when surges might occur. The relative peak numbers of active-Infectious cases occur soon after $\sigma(x(t))$ reaches a local maximum. The insight here is that $sin^2(\omega t)$ controls not only the viral surge periodicity, but can tell us when they well occur as well. For $\omega \leq 0.2$,when $\frac{d}{dt}\sigma(x(t)) = 0$, $\frac{d}{dt}sin^2(\omega t) = 0$ also [at the crest of each surge]; it is then relatively easy to find the time-windows around t_{surge} when surges in active-infectious cases will occur soon after. If $\omega \geq 0.2$ then the difference in time between the crest of each surge becomes too small to be of practical informative use and the number of surges becomes too numerous; i.e., the crests become so numerous and close, $\sigma(x(t))$ is practically constant from crest to crest.

$$\frac{d}{dt}sin^2(\omega t) = 0$$

$$2\omega sin(\omega t)cos(\omega t) = 0$$

$$\omega sin(2\omega t) = 0$$

Then in this instance, let n be constrained to the domain of non-negative integers so that we may find the times t of importance as follows:

$$2\omega t_n = \pi + 2n\pi$$

$$t_n = \frac{\pi(2n+1)}{2\omega}$$

where $n \geq 0$. In order to calculate a feasible value for omega, let us assume that there will be evenly spaced viral surges, the number of which will be denoted by the letter S, within the the time length k. It will take $\frac{k}{S2}$ days to reach the first transmission value crest at t_0. ω can then be rearranged for as follows:

$$\frac{k}{S2} = \frac{\pi(2n+1)}{2\omega}, n = 0$$

$$\omega = \frac{\pi S}{k} \longrightarrow \frac{\pi * 2}{110} = 0.0571$$

Using the the ω value of 0.0571 calculated, the relevant values for t calculated are days 28 and 83 for $n = 0, 1$ respectively; the t value of 138 for $n = 2$ is not of interest as $t_2 \geq k$. A reasonable assumption to make would be that the peak in active-infectious cases would occur 1-14 days after t_0 and t_1 due to the varying incubation rate of the virus.

Trying to describe the granular aspects of viral transmission within a small population and constrained time frame has proven to be very difficult; the lack

of data-points, the variance in how other institutions have handled and implemented COVID protocols is significant enough that comparisons are essentially useless. The largest improvement in the SIR model was seen when the the transmission parameter was allowed to assume variable values at essentially discrete points in time separated by a qauntized time interval. The insight here is that if N is sufficiently large, then the transmission coefficient may be treated as constant; if not, a more rigorous treatment is necessary. The interactions between all relevant individuals occur at discrete points in time, but the time gap between each interaction is so small due to very large values of N and how said population is distributed geographically. As seen in Figure 4 **Plot C**, when ω is very large, $\sigma(x(t))$ is essentially constant, only being forced to decay by the negative exponential. The fact that this essentially constant function has to be sampled at equivalently spaced intervals- discrete quantized intervals-and fits the raw data relatively well when used for SIR modeling, lends evidence to the theory that the transmission parameter is not constant at the fundamental level; it only appears so at the larger scale.

In fact, the $\sigma(x(t))$ trial function bears reminiscent characteristics of the radial wave-functions used to model the electronic states of the hydrogen atom. The similarity and successful usefulness is slightly eerie. The most appropriate analogy would be that at the micro level, classical models simply are not adept to describing what is fundamentally going on. Transmission is not truly constant; it is discrete because interactions between susceptible and infectious individuals are also discrete. Person A infects all of person B [persons are discrete] and so on and so forth at different discrete points in time. It is only when we zoom out to the macro scale that these discrete infections and overlapping interactions appear continuous and constant; *many* mutually exclusive interactions are happening simultaneously and very rapidly, giving the appearance of continuity, and hence a constant parameter. It is now clear why more robust and intricate models are necessary for an apt understanding at the small scale and this is the beginning of that effort.

There are limitations to our study. The two most prominent are (1), the inability to successfully replicate the results shown here when applied to another similarly sized institution because of lack of data, and (2) the assumptions made to simplify most college-students extremely volatile social behavior. Less importantly, the impact of age distribution on SMU's population was ignored as an overwhelming majority of the population is between the ages of 18-23 which essentially lumps the entire population into the same level of risk according to the CDC.

References

[1] Fred Brauer, Pauline van den Driessche, and Jianhong Wu, editors. *Compartmental Models in Epidemiology*, pages 19–79. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[2] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. JHU CSSE COVID-19 Dataset, 2020.

[3] Jin Wang Chayu Yang. Modeling the transmission of covid-19 in the us – a case study. *Infectious Disease Modelling*, 6:195–211, 2021.

[4] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534, February 2020.

[5] Leah Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, first edition, 2005.

[6] D Greenhalgh and R Das. Modelling epidemics with variable contact rates. *Theor Popul Biol*, 47(2):129–79, 1995.

[7] Can Hou, Jiaxin Chen, Yaqing Zhou, Lei Hua, Jinxia Yuan, Shu He, Yi Guo, Sheng Zhang, Qiaowei Jia, Chenhui Zhao, Jing Zhang, Guangxu Xu, and Enzhi Jia. The effectiveness of quarantine of Wuhan city against the coronavirus disease 2019 (COVID-19): A well-mixed SEIR model analysis. *Journal of Medical Virology*, 92(7):841–848, 2020.

[8] Xinzhi Liu and Peter Stechlinski. Infectious disease models with time-varying parameters and general nonlinear incidence rate. *Applied Mathematical Modelling*, 36(5):1974–1994, 2012.

[9] The Associated Press. Johns Hopkins COVID-19 case tracker. 2020.

[10] Southern Methodist University. Confirmed COVID-19 cases on campus. Available at https://www.smu.edu/coronavirus/cases, 2020.

[11] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[12] Eric W. Weisstein. Square line picking. Available at https://mathworld.wolfram.com/SquareLinePicking.html.

7 Appendix

7.1 Equations

$$
\begin{aligned}
\frac{dS_r}{dt} &= E_r(1-\sigma_{rr}(x(t))) - S_r\left(\sigma_{rr}(x(t))\frac{I_r}{N_r} + \sigma_{rc}(x(t))\frac{I_c}{N_c}\right)\\
\frac{dS_c}{dt} &= E_c(1-\sigma_{cc}(x(t))) - S_c\left(\sigma_{cc}(x(t))\frac{I_c}{N_c} + \sigma_{cr}(x(t))\frac{I_r}{N_r}\right)\\
\frac{dE_r}{dt} &= S_r\left(\sigma_{rr}(x(t))\frac{I_r}{N_r} + \sigma_{rc}(x(t))\frac{I_c}{N_c}\right) - E_r(1+\delta-\sigma_{rr}(x(t))\\
\frac{dE_c}{dt} &= S_c\left(\sigma_{cc}(x(t))\frac{I_c}{N_c} + \sigma_{cr}(x(t))\frac{I_r}{N_r}\right) - E_c(1+\delta-\sigma_{cc}(x(t))\\
\frac{dI_r}{dt} &= \delta E_r(1-\tau) - \gamma(t)I_r\\
\frac{dI_c}{dt} &= \delta E_c(1-\tau) - \gamma(t)I_c\\
\frac{dQ_r}{dt} &= \delta E_r\tau - \gamma(t)Q_r\\
\frac{dQ_c}{dt} &= \delta E_c\tau - \gamma(t)Q_c\\
\frac{dR_r}{dt} &= \gamma(t)(Q_r+I_r)\\
\frac{dR_c}{dt} &= \gamma(t)(Q_c+I_c)
\end{aligned}
$$

7.2 Transforming the risk function to space-dependent contacts

To transform the $\sigma(x(t))$ function into a function dependent on risk-laden transmission pathways or contacts (in this context, contacts are defined to be interactions within 6 ft) rather than persons encountered, we must scale the the x-axis accordingly:

$$\sigma(x) = \frac{e^{\bar{r}(x-\eta)}}{1+e^{\bar{r}(x-\eta)}}\left(\frac{2e^{\bar{r}x}}{1+e^{\bar{r}x}} - 1\right) \longrightarrow \sigma(x) = f(x)g(x)$$

where

$$
\begin{aligned}
f(x) &= \frac{e^{\bar{r}(x-\eta)}}{1+e^{\bar{r}(x-\eta)}}\\
g(x) &= \left(\frac{2e^{\bar{r}x}}{1+e^{\bar{r}x}} - 1\right).
\end{aligned}
$$

Solving $f(x)$ for x, we get

$$x = \frac{ln\left(\frac{-f(x)}{f(x)-1}\right)}{\bar{r}} + \eta$$

Recall that

$$E(x) = \bar{r}\left(\frac{x^2 - x}{2}\right),$$

which gives the number of transmission pathways in the total network that are high risk. Substituting the expression for x into E, we have

$$\frac{E_f(x)}{\bar{r}} = \frac{\left(\frac{ln\left(\frac{-f(x)}{f(x)-1}\right)}{\bar{r}} + \eta\right)^2 - \left(\frac{ln\left(\frac{-f(x)}{f(x)-1}\right)}{\bar{r}} + \eta\right)}{2}$$

Similarly we get the following expression when we solve g for x:

$$\frac{E_g(x)}{\bar{r}} = \frac{\left(\frac{ln\left(\frac{g(x)+1}{1-g(x)}\right)}{\bar{r}}\right)^2 - \left(\frac{ln\left(\frac{g(x)+1}{1-g(x)}\right)}{\bar{r}}\right)}{2}.$$

Rearranging for $\sigma(E(x))$, we get

$$\sigma(E(x)) = f(E_f(x))g(E_g(x))$$

$$\sigma(E(x)) = \frac{e^{\frac{\bar{r}(1-2\eta+\sqrt{1+8\frac{E_f(x)}{\bar{r}}})}{2}}}{1+e^{\frac{\bar{r}(1-2\eta+\sqrt{1+8\frac{E_f(x)}{\bar{r}}})}{2}}}\left(\frac{e^{\frac{\bar{r}(1+\sqrt{1+8\frac{E_g(x)}{\bar{r}}})}{2}}-1}{1+e^{\frac{\bar{r}(1+\sqrt{1+8\frac{E_g(x)}{\bar{r}}})}{2}}}\right),$$

where $\sigma(E(x))$ gives the transmission coefficient as a function of the risk-laden transmission pathways/contacts.

7.3 Alternate method for computing the distribution of edge lengths

A transformed Gaussian distribution will be sufficient to calculate shorter edge-lengths; it must be transformed and scaled such that the $x - axis$ is only continuous from 0 to ∞ because no edge-length may be negative:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

A change of variables is required to make the function begin at the origin; we will put x in terms of y such that the function will only be real from 0 to ∞:

$$y = \frac{x^2}{\sqrt{kA}} \longrightarrow x = y^{\frac{1}{2}}(kA)^{\frac{1}{4}}$$

$$P(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y^{\frac{1}{2}}(kA)^{\frac{1}{4}}-\mu)^2}{2\sigma^2}\right)}$$

Assuming we can transform the standard deviation and mean into functions dependent on A, in its form as the median edge length $\frac{\sqrt{2A}}{2}$,

$$\mu = \frac{\sqrt{2A}}{2} \times \frac{1}{c_1}$$

$$\sigma = \frac{\sqrt{2A}}{2} \times \frac{1}{c_2}$$

we now have a probability density function that is dependent on the given area of interest; as the area increases the curve will elongate and flatten, and as the area decreases, the curve will contract and spike. This is intuitive as the total area dictates the space nodes may occupy, and therefore the distribution of relevant edge lengths. The constants c_1 and c_2 are for fitting purposes and maybe be found numerically via empirical simulations. The unknown k may be found by integrating $P(y)$ from from 0 to ∞, and once found, the function will be pseudo-normalized (dependent on the choices of c_1 and c_2):

$$\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\left(\frac{(y^{\frac{1}{2}}(kA)^{\frac{1}{4}}-\mu)}{\sigma}\right)^2} dy = 1$$

Using u-substitution,

$$\alpha = \frac{(y^{\frac{1}{2}}(kA)^{\frac{1}{4}} - \mu)}{\sigma} \longrightarrow dy = \frac{2\sigma(\alpha\sigma+\mu)}{(kA)^{\frac{1}{2}}})d\alpha,$$

applying the substitution, extracting the constants, and fully expanding we have:

$$\int_{\frac{-\mu}{\sigma}}^\infty \frac{\frac{2\sigma(\alpha\sigma+\mu)}{(kA)^{\frac{1}{2}}}}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}(\ \alpha)^2} d\alpha = 1$$

$$\frac{2\sigma}{(kA)^{\frac{1}{2}}\sqrt{2\pi\sigma^2}} \int_{\frac{-\mu}{\sigma}}^\infty (\alpha\sigma+\mu) e^{\frac{-1}{2}(\ \alpha)^2} d\alpha = 1$$

$$\frac{2\sigma}{(kA)^{\frac{1}{2}}\sqrt{2\pi\sigma^2}} \left(\sigma \int_{\frac{-\mu}{\sigma}}^\infty \alpha e^{\frac{-1}{2}(\ \alpha)^2} d\alpha + \mu \int_{\frac{-\mu}{\sigma}}^\infty e^{\frac{-1}{2}(\ \alpha)^2} d\alpha\right) = 1,$$

Subsequent integration yields:

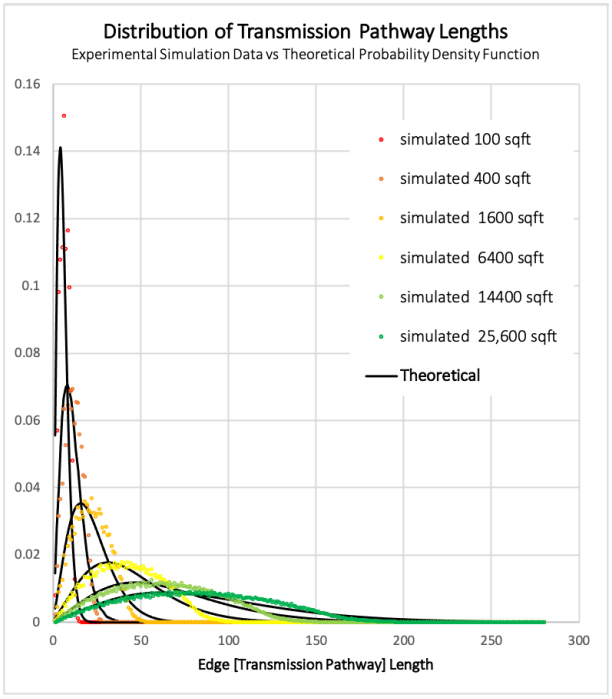
$$\frac{2\sigma}{(kA)^{\frac{1}{2}}\sqrt{2\pi\sigma^2}} \left(\sigma e^{\frac{-1}{2}\left(\frac{-\mu}{\sigma}\right)^2} + \mu\sqrt{2\pi}\right) = 1$$

which can be rearranged and simplified to solve for k:

$$k = \frac{2}{A\pi} \left(\sigma e^{\frac{-1}{2}\left(\frac{-\mu}{\sigma}\right)^2} + \mu\sqrt{2\pi}\right)^2$$

$$k = \frac{1}{\pi}\left(\frac{1}{c_2}e^{\frac{-1}{2}\left(\frac{-c_2}{c_1}\right)^2} + \frac{1}{c_1}\sqrt{2\pi}\right)^2$$

The constants c_1 and c_2 were empirically found to be 0.905 and 2.505 respectively, yielding a value of approximately 2.4572212 for the constant k.



For each of the areas listed in Table 2, p_c could be generated by integrating $\int_0^6 P(l(x))dl(x)$ numerically.