Transmission Trees, Superspreaders and Contact Tracing in Agent Based Models of Covid-19

Ernest Chang1, Ken Moselle2, Ash Richardson3

Victoria, BC

ABSTRACT

The agent-based model CovidSIMVL (github.com/ecsendmail/MultiverseContagion) is used to examine the fine structure of transmission paths in simulated Covid epidemics where the initial parameters are set to approximate epidemics with a single transmission source and R0 between 1.5 and 2.5.

The resulting Transmission Trees are characterized by breadth, depth, and generations needed to reach a target of 50% of the starting population (of 100) or self-extinction. Metrics derived for efficiency of an epidemic relate closely to the topology of the trees and their efficiency of transmission in space and time.

It can be shown that the notion of super-spreading individuals may be a statistical artefact of transmission tree growth, while superspreader events can be readily simulated with appropriate parameter settings. The potential of using contact tracing aggregate data to identify chain length and shared paths is explored as a measure of epidemic progression.

October 5, 2020

1 Ernest Chang, MD PhD(Computer Science – U Toronto)

­2 Ken Moselle, PhD (neuropsychology – UC Berkeley)

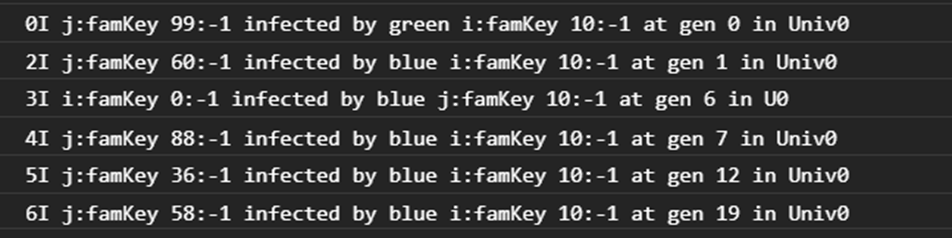
3 Ash Richardson, MSc (Mathematics – UVIC)

INTRODUCTION

As CovidSIMVL is a tool for specifying epidemics of varying intensities and progressions, the software can track all chains of transmission. This report is a study of the structure of the chains of transmission within specific synthetic epidemics, and what insights they yield with respect to the nature of the epidemic.

THE MODEL

We consider the traces for a synthetic epidemics, using console.log as the window into the detailed dynamics of the contagion. Please refer to the other Technical Reports and Handbook in github.com/ecsendmail/MultiverseContagion for background information.



This is a portion of console.log starting at the initiation of a trial, with a single Universe in action. The important point to note is that each infection is displayed with the identity of the infected node, and the node that was the infecting agent, and the generation at which this infection occurred.

Our model for study involves, as did Technical Report #004 on the parameters of Hazard Radius and MingleFactor, a population of 100 nodes (agents, persons) with one infective agent as the starting condition.

Clearly, the more active the contagion (high R0, large Hazard Radius, high MingleFactor, large population density), the more rapid the epidemic, and the more quickly agents become infected. We will characterize the dynamics of the progress through the enumeration of Chains of Transmission, starting from the initial agent (#10).

From the console.log below, we can generate a Transmission Tree, represented in Excel.



The reader is invited to verify the correctness of this table.

Parameters p=100 Hazard Radius=4 MingleFactor=0.90

At generation 1248, trial self-extinguished with 22 infections.

From the Transmission Tree (the table on the right), we can make the following observations:

1. The depth of the tree (maximum path length) is, at its maximum, 9.
2. The span of the tree (the number of distinct leaf nodes that have no descendants is 11. This is also the number of distinct chains (or paths to leafs).
3. The unique nodes in the tree (here it is 22).
4. The number of generations it takes to self-extinction or to reach 50 new infections (our target for the trials as in Tech Report #004). Here it is 1248
5. We have estimates of R0 for agents that have become inert at the time of termination.

SOME ASSERTIONS ON EFFICIENCY OF AN EPIDEMIC

We rate the efficiency of an epidemic by the largest number of infections in the shortest period of time. In the model with one initial infective agent, it is intuitively clear that the more parallel infective agents that are active, the less time it will take to infect a target population (whether it is 50% or all).

Thus, balanced trees with less depth (which are more infections depending on previous infections) and more breadth (more total leaves) will represent more efficient epidemic dynamics.

Trees such as balanced 1:5 (each parent node gives rise to 5 descendants)….are going to be more efficient than a highly unbalanced tree with one long pathway and many short branches coming off a level. At the extreme, one long pathway of k nodes with one terminal node is the least efficient, as it is a strict serial sequence of infections.

SOME LIMITED TRIALS

In the attempt to clarify and quantify the relationships between the structure of Transmission Trees for parameterized synthetic epidemics, we have run a limited number of trials, created potential metrics, and made some observations below.



The input variables are the first 2 columns. Notice that they are not in strict order of MingleFactor. The raw outputs are:

EndInf the number of transmissions at the end of the trial

Gen the number of generations to the end of the trial (reach 50 or extinction)

The derived measures from the Transmission Tree for each trial are:

Maximum Depth the longest path from the first node to a leaf (no descendants)

Average Depth the sum of the paths / number of distinct paths (leafs)

Leafs the count of nodes that have no descendants

Nodes the total number of New Infections (the initiator not included)

The Transmission Metrics considered for characterizing a trial are:

Leafs/MaxD this gives the ratio of the breadth vs depth of the Transmission Tree

Leafs/AvD breadth/average depth smooths out highly asymmetric trees

Nodes/MaxD total nodes incorporates the density of the shape of the tree

Nodes/AvD reflects the density of nodes over the average shape of the tree

Observations to this point

Recall the intuition that more efficient epidemics occur faster (Generations measure this), and therefore have more parallelism as reflected by broader, shallower trees. Such epidemics would have higher MingleFactors for a given Hazard Radius.

Following the four Transmission Metrics down the page in the Transmission Table above, we see a trend to higher values. This follows the expectation that breadth/depth should be higher for efficient epidemics and lower for slow or self-extinguishing epidemics.

The metric Leafs/AvD seem to be more internally consistent and aligned with MingleFactor (recall that this is a proxy for activity level and roaming in terms of agent behaviour) than the others, and without exclusion, we will adopt this metric as the go-to for now.

This table is sorted by MingleFactors, which increase down the page. We see that none of the metrics are uniformly increasing, though they all follow a trend to higher numbers with increasing MingleFactors. Other factors may need to be included.

SINGLE UNIFYING METRIC

The breadth and depth of a transmission tree may not cover all the important factors in a synthetic epidemic. Among other important factors are:

The Generation needed to run the trial to termination – the longer the less efficient

The number of transmissions before termination – the smaller the less efficient

In the row for MingleFactor 0.97, only 815 generations were needed compared to >1600 on each side of it, which makes the depth of the tree only 6 instead of 11 for its neighbours.

This epidemic stands out as being more efficient than those on either side. Not seeing any intrinsic reason for 0.97 to be different, we would attribute this result to the statistical stochastic nature of the simulation, which is why we need to repeat each trial several times.

Similarly, examine the trial for MingleF 0.94. This level of activity may be expected, based on its neighbours, to run to 50 new infections but in this trial self-extinguished at 25 infections, using 2588 generations, the largest in all trials.

To incorporate these factors, we put up two candidate Unifying Metrics

1. Unifying Metric T/E, which is defined as:

T/E = ((Leafs/AvD)/Gen)\*1000 the 1,000 scales to values near 1 to 10

we put the Gen into the denominator so that smaller Generations yield larger T/E efficiencies

1. Q, defined as

Q = T/E \* New Infections \* 100 again to scale to values near 1 to 10

The result of using these two Unifying Metrics can be seen in the following condensed table:



This table is sorted by Q.

In the Green Band, the epidemics are self-extinguishing (EndInf = 25,22,20,47,49) with relatively larger generations except for Gen=1248 and Gen=774 where the EndInf are only 22 and 20 respectively. The Q values are between 1 and 10.

In the next band, Q values are between 10 and 20. Even though MingF=1.01 did not create 50 new transmissions, the extinction came in only 1056 generations, with a smaller depth of Transmission Tree, so together these contributes to the efficiency Q of the epidemic. The stochastic nature produced a rapid epidemic which at 47, ran out of susceptible neighbours for the remaining infective agents. This is the nature of statistical randomness, and does not take away from the overall character of this epidemic.

The next band from Q values of 20 to 30, are characterized by lower Generations and smaller Average Depths than the previous two bands. This means that Q is driven by the actual epidemic progress than the parameters, which only set the framework, but not the unfolding in a stochastic universe.

The final band with Q values above 40, clearly have smaller Generations, and shallower depths to the Transmission Trees, reflecting the overall efficiency of the transmissions.

REPEATING THESE TRIALS OVER A LARGER RANGE OF PARAMETERS, AND OVER THE SAME PARAMETERS, WOULD BE NEEDED BEFORE THESE TENTATIVE OBSERVATIONS CAN GIVE MORE CONFIDENCE AND RELATE MORE DIRECTLY TO PARAMETERS AND R0.

CONSIDERATIONS ABOUT R0

The approach we have taken to R0 is to record the number of susceptibles that each agent in its lifetime have transmitted successfully. However, when we terminate at newInf=50, there may be a significant number of agents that have not completed their life-cycle …. we only know that for those which have become inert, we have their counts, and we can average over those to provide a value of R0.

These are of course the early infectives, because susceptibles never become Inert. However, the random nature of the simulation does not guarantee that all infectives to any point have the same number of infectives.

By examining the Transmission Trees, we see many that do not have a balanced structure. This means that the paths which are short had few transmissions. If the distribution of infections are random, any particular R0 at newInf=50 or at extinction is not necessarily reflective of the average over a large number of trials with a given parameter setting of Population, Hazard Radius and MingleFactor.

OVERLAPPING PATHS

Epidemics with more serial than parallel executions would share more paths that have common ancestors, and this characteristic may be of use in describing the epidemic dynamics, if we can capture the relationships.

With lower degrees of parallelism, there would be more common paths, and longer ones, such as in the Transmission Tree for HzR=4 and mF=0.94, with a Q of 1.72.



On the other hand, for HzR=5 and mF=2.0 HzR=4 we have a high Q value of 61 (very efficient epidemic), with many shorter paths.



Longer common paths are signs of a slower epidemic

DEGREE OF INFECTIVITY OF AGENTS

There are two ways to look at the role that individual agents (or persons) can have in the progress of an epidemic. One is to count the numbers of persons who become infected as a consequence of the individual being infective. This corresponds to tallying the size of the sub-tree of descendants of the individual in the Transmission Tree. This count might be termed the SPREAD COUNT of an infective agent.

The other is to count the number of DIRECT descendants of an agent, rather than all of the descendants that result from first level, second level, third level etc descendants. This is simply the Edge Count of a node in the Transmission Tree, and in the context of an epidemic, we can call it the CASE COUNT.

To illustrate these two concepts, we will compare two Transmission Trees, taken from the Appendix, for HzR=4, MingleFactor=0.91 and MingleFactor=1.04. The first is a trial that self-extinguishes after 22 new infections and 1,248 generations, with an R0 of 2.00, and the second completes 50 new infections in 855 generations with a terminating R0 of 3.41.

Their summary differences from the Transmission Table looks like this:



The first case (call it mF=0.90) has a Q value of 3.33, at the extreme end of inefficiency, while the second case (mF=1.04) has a trial which is very efficient, with a Q of 41,36, ending in 852 generations instead of 1,248.

Clearly the mF=1.04 trial has a larger leaf to depth ratio at 7.07, compared to 1.89, which shows a higher degree of parallel activity.

The two Transmission Trees taken from the Appendix look like this:

mF=0.90 mF=1.04



It is obvious that the Transmission Trees, which have been redrawn in their graph representations, are very different, with mF=0.90 being tall and skinny, with only node 33 having more than two Case Counts, and a maximum path length of 9, while the mF=1.04 is broad and shallower, with a maximum path length of 6.

We can compare their SPREAD COUNTs, and their CASE COUNTs:

mF=0.90 mF=1.04 mF=0.90 mF=1.04



Consider first the SPREAD COUNT. Efficient epidemics have broader, shallower shapes, so we expect that nodes higher up will have larger sub-trees. Therefore, it should take fewer nodes to account for 50% and 80% of the infected population.

If the spread was uniform, we might expect the SPREAD COUNT for each person to be the same (each person spreads to the same number of descendants), but with the stochastic tree of primary and secondary and tertiary spread, it is inevitable that some agents will infect a larger pool of descendants than others. When we sort the agents by their descendant counts, we get some agents with large numbers of descendants, and some with small numbers of descendants.

In broader trees, the difference in sizes will be greater than in skinny trees. This has to do with the k-distance between centroids represented by the roots of sub-trees, and will be left to another time. For now, we simply observe that to get from 50 to 25 to 10 (stepping past 50% and 80%) takes 1 step and 6 steps in mF=1.04, or 2% and 12% of the infected number.

For mF=0.91, to get from 22 to 11 to 4 (getting past 50% and 80%) takes 4 steps and 8 steps, or 18% and 36% of the infected number. This adds a quantitative and descriptive dimension to what it intuitively obvious: that the SPREAD COUNT is not uniformly distributed, and in fact, is skewed according to the distance in sizes between the clusters representing sub-trees.

In terms of CASE COUNTs, the tables show up to 2 case counts, and everything below is 1 or 0. Thus,mF=1.04 has 13 agents with PRIMARY infections of 2 or more agents, while MF=0.90 has 6 agents with infections of 2 or more agents. What is remarkable is that in neither case is the largest significantly related to the efficiency of the epidemic, considering how different they are.

COMMENT ON SUPERSPREADERS

If by “superspreader” we mean agents that have a high degree of primary or DIRECT infections, the stochastic nature of the agent-based model does not appear to generate large variations.

On the other hand, if by “superspreader” we are counting all descendants infected by primary and secondary (and so on) infections along all subsequent chains of infections, then the stochastic nature of the agent based model guarantees that there will be significant asymmetry depending on the temporal position of an agent in the transmission chains. These may not be physiological super-spreaders but statistical artifacts of the stochastic and probabilistic model of transmissions based on contact.

COMMENT ON SUPERSPREADER EVENTS

There are three ways that the term “superspreader” has found common use. They are:

1. A person who is highly infective with many PRIMARY infections of susceptibles
2. A person to whom large numbers of cases can be traced, and “held responsible”.
3. An event in which an “unexpected” number of new cases arise.

We have discussed (1) and (2) in the preceding paragraph. Here, we establish the conditions for a super-spreader event as modelled in CovidSIMVL, by using the same 100 persons, with 5 infective persons, a Hazard Radius of 20 and a MingleFactor of 10, to represent high density of persons (as in the Rose Garden White House Sept 26 event, and a MingleFactor of 10.

The following infections take place in the first Generation (hour)….

0I j:famKey 78:-1 infected by green i:famKey 10:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 89:-1 infected by green i:famKey 10:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 43:-1 infected by green i:famKey 11:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 62:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 71:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 34:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:2006 0I i:famKey 0:-1 infected by green j:famKey 13:-1 at gen 0 in U0

simulation.js:1990 0I j:famKey 59:-1 infected by green i:famKey 10:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 66:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 77:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 88:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:2006 0I i:famKey 7:-1 infected by green j:famKey 13:-1 at gen 0 in U0

simulation.js:1990 0I j:famKey 60:-1 infected by green i:famKey 10:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 96:-1 infected by green i:famKey 10:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 18:-1 infected by green i:famKey 11:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 22:-1 infected by green i:famKey 11:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 83:-1 infected by green i:famKey 11:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 56:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 24:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 23:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 53:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 65:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:2006 0I i:famKey 1:-1 infected by green j:famKey 12:-1 at gen 0 in U0

simulation.js:2006 0I i:famKey 4:-1 infected by green j:famKey 12:-1 at gen 0 in U0

simulation.js:1990 0I j:famKey 17:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 26:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 38:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 61:-1 infected by green i:famKey 12:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 16:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 20:-1 infected by green i:famKey 13:-1 at gen 0 in Univ0

simulation.js:1990 0I j:famKey 27:-1 infected by green i:famKey 14:-1 at gen 0 in Univ0

simulation.js:1990 36I j:famKey 40:-1 infected by blue i:famKey 12:-1 at gen 1 in Univ0

That would be 31 persons infected by the five active cases in the first hour, distributed among agents 10 to 14 in the counts 5-4-8-8-6. This is a simple demonstration of the power of density and mingling in the contagion spread of Covid.

POTENTIAL APPLICATIONS OF THE TRANSMISSION TREE CHARACTERIZATION

It is intuitively clear that rapid epidemics would have many infective agents, and that the chains of transmission would be short (if one could clearly trace them), and that the leafs (not having infected anyone else yet) at any one time would be plentiful compared to the length of the chains.

If contact tracing were seen as a sample survey of chains of transmission, and they were examined in the aggregate, and the transmission tree or trees to be reconstructed, then it may be possible to estimate the Generation, Depth (path length), Leafs and New Infections, and thus calculate Q.

Often it is somewhat difficult to know with the variations in daily counts and positivity which way the epidemic is going….up, down, sideways.

Contact tracing may be seen as a forward activity – given a positive, who have they contacted recently who need to be self-observed and tested, as well as a backward activity – who were they in contact with, and were any of these known cases, or should be tested if they were in common with other new cases.

The creation of Transmission Trees can be at a Provincial, or at a Regional, or City, or Minicipal, or Regional, or local epicenter level……any infective agent can be considered the root of a Transmission Tree, and the subsequent epidemic characterized from that point onward.

APPENDIX. TRANSMISSION TREES



Generations = 1248



Generations = 774



Generations = 1481



 Generations = 1748

Generations = 2588



Generations = 1056



















