Descr 4
Form. 4
Qual 4
Depth 4
Mech 4

20/20 fantastic!

Sex, Death, and Scale-Free Graphs: Modeling the AIDS Epidemic

Simon Maccracken Stump
Scientific Computing

April 28, 2006

1 Background

Earlier this year, I was chosen to represent Harvey Mudd College in the Interdisciplinary Contest in Modeling (ICM). Our project involved creating a model to simulate the effect the virus HIV would have on a handful of countries around the world, and test the effect of various proposed solutions. While brainstorming possible models, we felt that a scale-free sexually transmitted infection (STI) model would be the most realistic, and the most interesting. However, due to time constraints, we decided to program a simple suseptible-infected-resistant (SIR) model instead. For my final project, I decided to design the scale-free STI model, and see how successful it would be.

SIR models assume a randomly mixed population, in which every person is equally likely to give a disease to every other person in the population. The exact makeup of the population is unimportant. This works well for air borne diseases, such as a cold, the flu, or SARS. However, there are many diseases, referred to as social diseases, in which the exact social networks can make a difference. STIs often fall into this category [5].

In a scale-free model, each individual person is represented as a node on a graph. If two persons are able to transmit a disease to one another, this is represented with an edge between the two nodes. Under a scale-free model, the probability density function for the number of edges a node has is a power series. The result of this is that most persons in the society have fewer

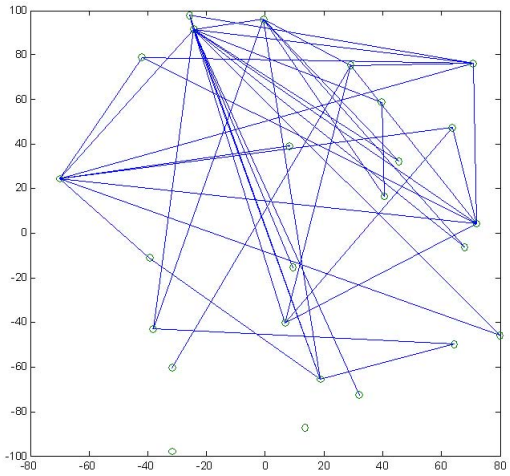

Figure 1: A scale free model is one in which the probability density function for the number of edges per node is a power series, resulting in most nodes having a few edges, and a few having many edges.

than average vertices, and a small number of people have a large number of vertices [5]. An example graph can be seen in Figure 1.

2 The Model

My simulation works as follows. First, the program builds a society of people. The program starts with a simple graph. Then, it adds a node one at a time, until it has a full starting population. For every new node, the probability that node will be connected with node i is

$$p(connect) = \frac{V_i + 1}{N} \tag{1}$$

where V_i is the number of edges on node i, and N is the total number of nodes, counting the new one. This processed was taken from [6]. Once this population is created, a preset number of people are randomly selected to become infected with HIV. Those who are infected have a random level of

infection (see below). This process takes $\Omega(N^2)$ steps, where N is the number of individuals in the population. However, it was often the case for multiple trials that a single population would be created, and used every time.

Once we have built the population, we run it for a preset number of years. Each year, the following events happen. First, health deteriorates. Second, HIV transmissions occur. Third, people die. And last, new people are born.

2.1 Health

All persons have $N+2$ possible levels of health. A health of 0 means healthy. A health of 1 means in the first year of HIV infection. Similarly, a health of 2 means in the second year, and so on. A health of N (attained after being infected with HIV for N years) indicates that a person's disease has progressed to the level of AIDS. A person's infection cannot be worse than this, although it can kill them. Finally, a health of -1 is assigned to a dead individual. The process of adjusting people's levels of infection takes $\Omega(N)$ steps.

2.2 Transmissions

Normally, people who study infectious diseases talk about a disease's R_0, in order to determine whether the diseases will persist. Basically, if I became infected with a disease, then on average I will infect R_0 others before the diseases has run its course. The R_0 value for AIDS under this model is

$$R_0 = (chance\ of\ transmission) * (life\ span) * (\#\ of\ contacts) \quad (2)$$

Thus, R_0 should vary from individual to individual.

To determine whether HIV is transmitted, the population is scanned. Whenever there is an edge between someone who is infected and someone who isn't, there is a set random probability of transmission. This probability is increased in someone with AIDS, because of an increased viral load. In some cases, it may be realistic for this probability to be reduced if either individual has a high number of edges (this would represent the government pushing for condom use in brothels, needle exchange programs, and other

methods of specifically targeting those most at risk). A person who has just become infected cannot infect others until the next year. This entire process takes $\Omega(N^2)$ steps.

2.3 Death

Every year, a person has a set chance of dying. If the person has a fully progressed case of AIDS, the person suffers an increased likelihood of dying (often, this is actually an automatic death). If a person dies, their health is set to -1, and all of their vertices are removed. This process is $\Omega(N^2)$.

2.4 Births

The number of births is equal to the birth rate multiplied by the population. The overall population gain is multiplied by a random number, to represent small fluctuations in birth rates. Each person then gains a random number of contacts in the same fashion as when the population is created. The birth process takes $\Omega(M^2+NM)$ steps, where N is the number of people currently in the population, and M is the number of new births.

3 HIV at Mudd

As a test of the model, we considered the question, "Could a strain of HIV become endemic at Mudd?" This question was sparked in part by the 2005 National College Health Assessment (NCHA), where 1 individual (of the 1/3 of campus who responded) indicated that he or she was HIV+ [1]. Because we are a school of only 700 (a medium size for the model), I felt that this would be a perfect opportunity to test this.

Under this scenario, connectivity was set to 1.12 (this gave us contact levels that were similar to those indicated in [1]). Due to problems setting up an initial graph, all trials began with the same initial population. I assumed that no one would die of AIDS, because the incubation time was much higher than the 4 years most students spend in college. I assumed a 25% annual

turnover in the student body, and a 15% transmission rate of HIV annually (this would be expected if the person was not on anti-retroviral medication, and if condoms were not used [2]), giving us that $R_0 \approx 0.45$(# of partners). These both were worst case scenario situations, and were chosen to determine if HIV had any chance of becoming endemic at Mudd. The simulation was run for 25 years.

I ran several trials in which 3 members of the initial student body had HIV. Under no circumstances (short of an unrealistically high transmission rate) were any students HIV+ at the end of the simulation. In order to determine if this was an effect of initial conditions, I ran 5 trials, in which 100 student (14% of the initial population, similar to the current state in Zimbabwe) had HIV. After 25 years, there was an average of around 6-7 students in the population with HIV (see Figure 2). With numbers like these,

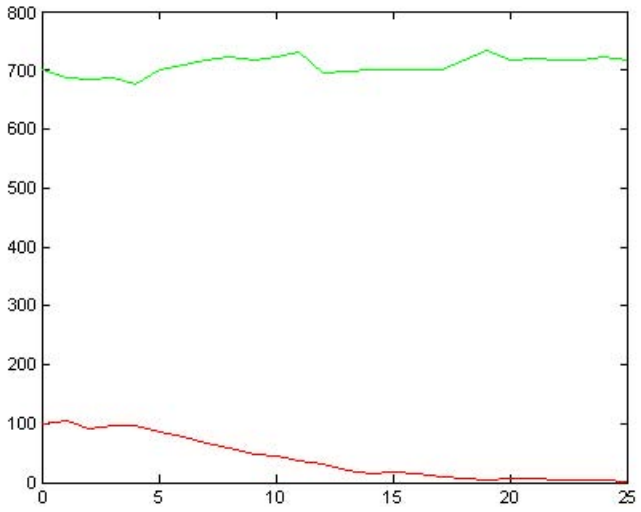

Figure 2: Even with an extremely high initial population of infected students, HIV does not appear to be capable of becoming endemic at Mudd without additional outside. The green bar represents the total population size. The red bar represents the number of people infected with HIV.

it seems that Mudd need not worry about an endemic HIV problem, unless

either there is an influx from an outside population, or there is a sudden and prolonged increase in heroin use.

4 HIV in Africa

4.1 Latin Hypercube Sampling

In order to run sensitivity tests on my AIDS model, we chose to use Latin Hypercube Sampling (LHS), as described in [3]. The advantage of LHS is that you can analyze data with far fewer calculations than would be needed for a factoral experiment (consider, for 4 parameters, trying 5 values each, would require $4^5 > 1000$ trials, while the LHS analysis we did took only 16). The first step in LHS is to choose a probability distribution function for each variable, and then divide each pdf into N equiprobable groups (where $N > 4K/3$). Then, a point from each group is drawn without replacement, and N trials are arranged. Once each of these N trials has been run, statistical tests are run to determine the correlation between each variable and the outcome.

4.2 Setup

For this project, we ran 16 simulations, and varied 4 parameters. The first parameter was how long a person could live with HIV before it advanced to AIDS. The second parameter varied was the average number of contacts a person would have. The third parameter was how easily HIV was spread. The last parameter was how easily HIV was spread by those within the top 10 percentile of contacts (this ended up being anyone with 9 or more contacts). There were two variables we kept track of, to determine the "health" of a population. The first was the final population size. Because every population should grow to be approximately similar population levels in the absence of AIDS, this was an indicator of the size of a population crash. The second health statistic was the end number of people infected with HIV. We hoped that these two statistics would give similar results. Each simulation was run 10 times, and the average end values were used. All parameters except promiscuity were taken from [2]. I assumed that the low expected life span

was likely in part due to the high prevalence of AIDS, and as such I chose rather to use the average expected life span in Africa [4].

Because the initial setup of the population was so variable, and had such a major effect on the outcome of the population, this was regulated. Also, due to technical trouble, we the starting population was only 999 people. All trials were run with the same initial population, and with all members having the same level of infection.

4.3 Statistics

To determine the effect each parameter had on overall AIDS levels, we ran least-squares correlative tests. Because our health measurements appeared very non-normal, we ran tests on their rank (i.e., the largest population received a 16, the next largest a 15, and so on) [3]. Our degree of correlation, r, was determined by

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} \tag{3}$$

where $SS_{ij} = \Sigma(ij - \overline{ij})$ [7].

4.4 Results

We found that our two ratios of health were related. In fact, as can be seen in Figure 3, larger populations tended to have a lower number of HIV cases.

We ran statistical tests on our data, the results of which can be seen in Table 4.4. Simulations with the same parameters gave similar results. For example,

Parameter	Corr. with Population	Corr. with Health
Time to Progression	0.3147	-0.0118
Trans. in High Risk Groups	-0.302	-.7223
Connectivity Level	0.7250	0.3471
Transmission Rate	-0.3843	0.1122

The results indicate that reducing the transmission rates of those most at risk and reducing promiscuity levels would have the biggest effect on the population. Increasing the time it takes for HIV to progress to AIDS only

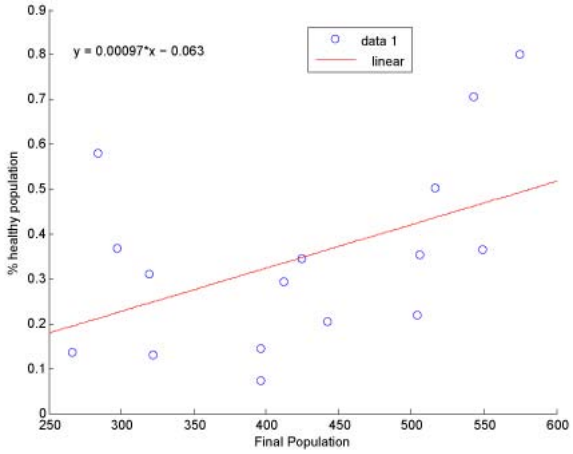

Figure 3: There appeared to be a positive correlation between end population levels end healthy population.

affected the end population. Lowering the transmission rates increased the end population also, but had the paradoxical effect of increasing the level of HIV in the final population. This may have been caused by statistical noise.

A later simulation gave somewhat different results, in which increasing the time to HIV progression severely hurt the population, and the transmission rate in high risk groups made less of a difference. Due to time constraints, we were unable to run more tests. This should be done in the future, however, to determine what the optimal number of simulations would be. In the later simulation, we found that after simulations with the same set of parameters, population levels had an average standard deviation of 5%, and population health had an average standard deviation of 11%.

4.5 Conclusion

It appears that according to the model, the most successful plans for lessening the impact of AIDS on Zimbabwe would

?

This seems the most important part!!
But I remember from your talk.

5 Problems

There are several problems with the current model, which could be improved upon. The first is that we assume that if two people come into contact with one another, they will stay in contact for the rest of their lives, no matter how many other people they come into contact with. The second problem is that, in our model, a person cannot transmit HIV to others the same year they come into contact with it. This would tend to underestimate the spread. Another problem is that everyone in the model has the same chance of transmitting HIV per contact. This is not realistic, because various factors, such as the use of condoms, or whether the person is infected with other STIs can affect how likely HIV will be transmitted. A further model could take into account such things as "uses condoms," and allow these people to have lowered transmission rates. The fourth problem is that chance of transmission does not scale with number of contacts. To explain this, I will give an example. If a person has only one other partner, and the partner has HIV, then we will say that the person's chance of contracting HIV is $(1 - X)$. On the other hand, if the person has 300 contacts, each of whom is infected with HIV, then the probability that person will contract HIV is $(1 - X^{300})$. This would only be the case if the latter person were having 300 times as much sex as the first person, which seems unrealistic. This could perhaps be fixed with some type of asymptotic curve. The final problem with the model is that it assumes static behavior. This may not be unrealistic, however it means that we are unable to determine what increased HIV testing would accomplish.

6 Acknowledgments

I would like to thank Benj and Marguerite for inspiration, for helping design this model, and for all of the research they conducted during the ICM. I truly could not have done this without them. I would also like to thank Professor DePillis for help with scale free models. Finally I would like to thank Dean Guy for giving me data from the NCHA.

7 Bibliography

References

[1] American College Health Association. National college health assessment spring 2005.

[2] B. Azose, M. Leeds, and S. Stump. Talking about my generation's big crisis: Hiv/aids recommendations to the un and the who. Technical report, Harvey Mudd College, 2006.

[3] S. M. Blower and H. Dowlatabdi. Sensitivity and uncertainty analysis of complex models of disease transmission: an hiv model as an example. *International Statistical Review*, 2:229–243, 1994.

[4] The consortium for Mathematics and its Applications. 2006 icm problems. http://www.comap.com/undergraduate/contests/mcm/contests/2006/problems/.

[5] F. Lilijeros, C. R. Edling, and L. A. N. Amaral. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection*, 5:189–196, 2003.

[6] Alex Popkin. Math 164 final report: Virus-spreading in scale-free networks. Technical report, Harvey Mudd College, 2004.

[7] Wolfram Research. Correlation coefficient. http://mathworld.wolfram.com/CorrelationCoefficient.html.