

Pandemic: A Graph Evolution Story

Michael Dubé
Dept. of Computer Science
Brock University
St. Catharines, ON, Canada
Email: md12ol@brocku.ca

Sheridan Houghten
Dept. of Computer Science
Brock University
St. Catharines, ON, Canada
Email: shoughten@brocku.ca

Daniel Ashlock
Dept. of Mathematics and Statistics
University of Guelph
Guelph, ON, Canada
Email: dashlock@uoguelph.ca

Abstract—The Graph Evolution Tool (GET) was built to generate personal contact networks representing who can infect whom within a community. The tool is expanded in order to permit an infection scheme which divides the community into different districts, thus permitting within-district and between-district infections. The evolutionary algorithm comprising GET is expanded upon to simulate communities which include 512 individuals in up to eight districts, initially infecting one person in one district and spreading through a community. The overall goal is to generate communities that will maximize the length of an epidemic. The problem associated with adequately exploring the numerous parameters accompanying evolutionary algorithms is addressed using a point packing and insight from previous work. The Susceptible-Infected-Removed (SIR) model of infection was chosen as it provides a sufficient balance of simplicity and complexity for the problem.

I. INTRODUCTION

This paper considers the problem of evolving a personal contact network, an epidemic model in which an epidemic spreads only along the links of the network. It is well known that the structure of the contact network has a significant effect on the spread of epidemics [9].

The study uses the Local THADS-N representation, a generative representation for network induction. An evolutionary algorithm utilizing this representation is used to generate good personal contact networks for specific test problems that include maximizing *epidemic duration* [4] and *epidemic profile matching* [6]; the current study considers the problem of maximizing epidemic duration. This approach of using a generative representation to locate satisfactory networks for a test problem is termed *graph induction*.

The Local THADS-N representation has evolved over the course of several past publications which have created different versions of the representation, in each case using evolutionary algorithms to generate graphs with desired characteristics. In [3] the first operation of the representation, edge swap, was introduced as a potential universal operator for graph induction; this is shown within Fig. 3 as the Swap operation. Edge swap was used because it was able to take an initial personal contact network and change which members of the community came into contact with each other. This representation proved to be universal but exhibited poor performance, requiring a rather ornate evolutionary algorithm to permit it to function, the restarting-recentering algorithm [8]. To improve performance,

additional graph editing operations were added, detailed in Section II, which allowed for the addition and deletion of connections [1]; more recent work has focused on adding local variants of the operations [7], [11]–[13].

The personal contact networks used in this work are implemented as combinatorial graphs. Individuals are represented by the *vertices* of the graph and the connections between individuals, along which infection can spread, are represented by the *edges* of the graph. The terms network and graph are used interchangeably within this paper. A graph G is defined as a set of edges E and vertices V and is denoted $G(V, E)$. An edge is represented as an unordered pair $\{p, q\}$ in which p and q are vertices from V . Only undirected graphs are used in this study, meaning that infection can pass in either direction. A graph is said to be *sparse* if it has a relatively small fraction of its possible edges, given the number of vertices it possesses.

A *path* from vertex p to vertex q on graph G is a sequence of edges from E which connect p and q . The *distance* from p to q is the length of the shortest path between p and q . A graph is *connected* if there is a path from any vertex to any other vertex.

In [14], the researchers use personal contact networks to represent a human population and compare simulated epidemics against data from an outbreak of H1N1 in 2009 on a university campus. Their personal contact network is generated in a series of layers with each layer adding more edges to the network. The first layer has edges between members living in the same residence room. The second layer connects members who are enrolled in the same class. Successive layers represent less intense and less frequent interactions between individuals (e.g. attending the same university). The network begins as vertices without any edges and the layers are applied such that the graph is connected only after the final layer is applied. Thus, the final layer adds edges between several sub-networks which share no edges between one-another.

The current study combines a similar network structure as [14] with the Local THADS-N generative representation. The problem of maximizing epidemic duration is used to test personal contact networks that are arranged into a set of sub-networks. This arrangement is analogous to a set of districts or neighbourhoods, each with separate personal contact networks. Within a district individuals have frequent interactions, while there are less frequent interactions with nearby districts such as when an individual travels from their own neighbourhood

to another. The study thus considers how the length of the epidemic is affected when an epidemic starts in one district and then with some probability “jumps” to another district.

A. The Epidemic Duration Problem

The length of an epidemic is the number of time steps, after the initial infection, until there are zero infected individuals within a community. The epidemic duration problem [4] seeks to find graphs which promote epidemics with a long duration. Originally, this was determined by averaging the length of 50 simulated epidemics on a candidate graph. In contrast, this paper uses a single sample epidemic to evaluate the fitness of a solution; this is combined with a selection method known as *skeptical tournament selection* [21], detailed in Section III-A.

B. The Model of Infection

The basic Susceptible-Infected-Removed (SIR) model of infection [16] [18] provides a simple model for the simulation of epidemics. In this model a population is divided into three mutually exclusive groups. The *susceptible* group consists of those not yet infected, but who may still become infected by the epidemic. The *infected* group consists of those currently infected. The *removed* group consists of those previously infected but no longer susceptible; conceptually, they have been removed due to either immunity or death. Unfortunately, the basic SIR model scales poorly with the number of individuals in the population: in the basic SIR model all individuals can infect all others, so that a large population makes it very likely that the epidemic will explode. To remedy this problem, social contact networks are used instead of completely connected graphs; thus constraining the epidemic to spread in a more realistic fashion. The basic SIR model is useful in small, well-mixed groups such as families or workplaces where an epidemic can spread between any two individuals with sufficiently equal probability.

An epidemic, as realized in this study, begins by choosing one individual within a community to be infected. From there, the epidemic is permitted to spread along edges of the network. The epidemic disease lasts a single time step within an infected individual; after this time they are no longer able to spread the epidemic and now belong to the removed group.

In the current study, the overall community is comprised of a fixed number k of smaller districts. To simplify analysis, all districts have the same size and same structure. Essentially, a personal contact network (graph) of the required district size is evolved and then k copies are used to construct the overall community. Each of the k districts is connected to each of the other districts. A district yet to experience the epidemic has a probability α' that one of its members will become infected by a member of another district that is infected, calculated independently for each neighbouring infected district. Fig. 1 demonstrates how a community of districts is created using a personal contact network. For simplicity, the first individual infected within a district is the individual represented by the vertex with the lowest index within the district, and is called

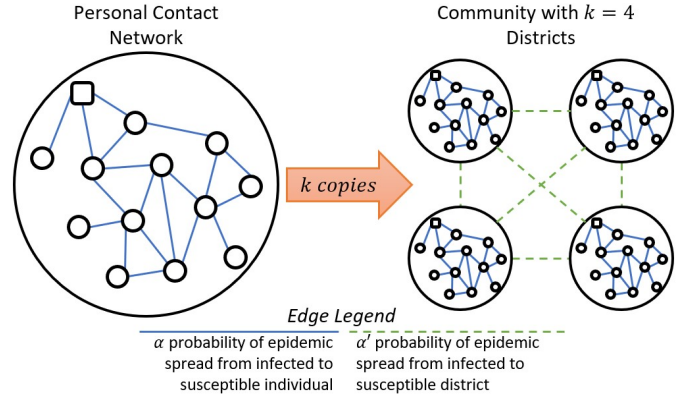


Fig. 1. A community is comprised of k identical personal contact networks which are completely connected to one another. The rectangular vertex represents the vertex with the lowest index within the district (patient zero).

patient zero. Further details on the experimental design are provided in Section III.

The remainder of this paper is organized as follows. Section II formally defines the representation. Section III reviews the design of the experiments, how parameters are obtained and their values. The results from the experiments will be presented and analyzed in Section IV. Lastly, Section V will conclude the paper as well as present possibilities for future research.

II. THE LOCAL THADS-N REPRESENTATION

The Local THADS-N representation does not directly specify a solution to a problem. Instead, the representation provides several *editing operations* that can be used to modify a graph. A solution to the problem is then specified by applying a sequence of edits to the initial graph.

Editing an initial configuration is a generative solution to the network induction problem. This type of solution is used because it allows domain information to be included in the original graph (e.g. a reasonable number of edges). The initial graph may be used to put the evolutionary search in a good part of network space. It also allows graphs to be evolved using an evolutionary algorithm that uses a simple linear structure, namely, the list of editing commands. The editing commands are represented as large integers, sliced with **div** and **mod** operations to specify specific commands and their parameters.

Some existing applications of generative representations include [10], [15], [19]. The potential effectiveness of generative solutions is demonstrated in [17]. The operations in this representation specify changes to the edges of a graph.

In this study, the strings of operations are applied to an initial graph that has the general structure shown in Fig. 2. The example shown has 128 vertices but it is generalized to graphs with the desired number of vertices for a given district. This structure was chosen as previous research [4], [6] demonstrated that graphs having vertices with four or five edges are desirable for the test problem at hand.

One possible representation for directly evolving graphs with n vertices would be a binary gene with $\binom{n}{2}$ loci, each

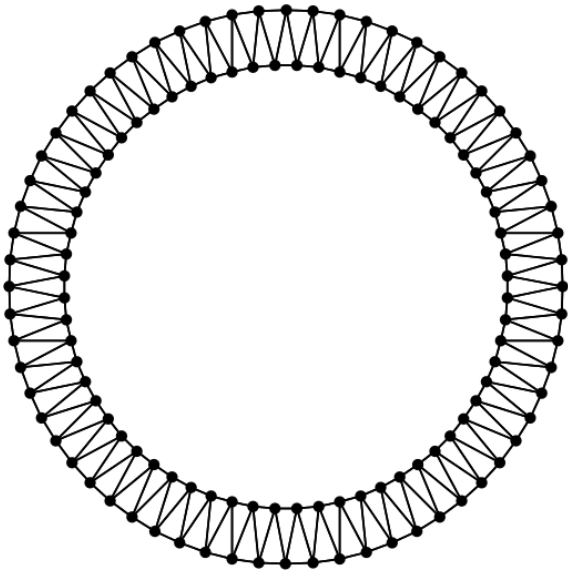


Fig. 2. Initial graph: each vertex has edges to the two vertices before and after that vertex in the ring. The example shown has 128 vertices.

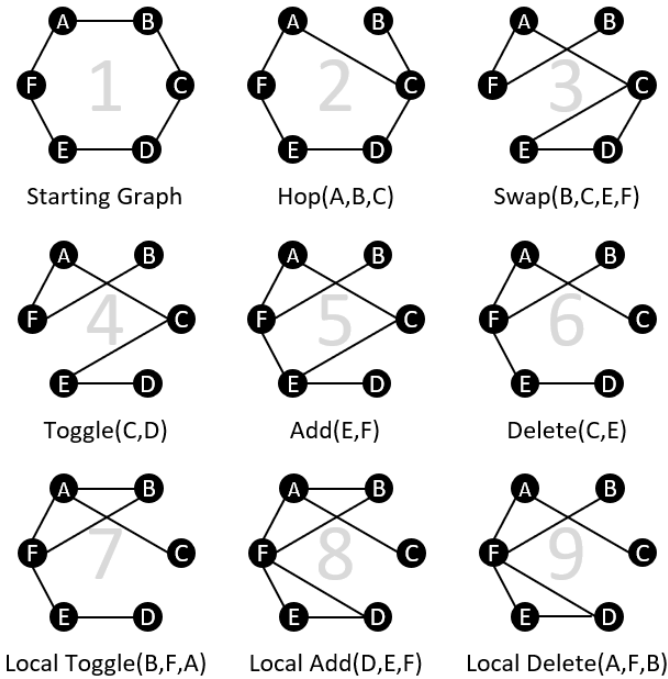


Fig. 3. Examples of operators included in the Local THADS-N representation. The figure shows the eight of the nine operations being applied sequentially to an initial six cycle graph. The operations are applied in reading order such that the final graph is the result of applying all eight operations. The null operation is omitted as it does not change the graph.

of which specifies the presence or absence of a possible edge. This representation seems natural. However, it was found to perform badly on many test problems [5], as were its subsequent modifications. The reason for this poor performance relates to the fact that almost all interesting graphs are sparse,

while very few graphs represented in this manner are sparse. Although it is possible to adjust the primitive probability of an edge existing, this helps only in a limited way because in this representation crossover has a high probability of increasing the number of edges when applied to two sparse graphs.

The generative representation used in this study also searches a relatively restricted portion of network space, namely a sphere, in the space, whose centre is the starting graph and with radius equal to the number of allowed edits.

Some of the operations within the generative local THADS-N representation are *local*. Local operations are only applied to triples of vertices for which any pair of the vertices are at distance two or less before and after the operation is applied.

A. Local THADS-N Edge Operations

Given a graph $G(V, E)$ and vertices p, q, r , and s from V the allowed operations are as follows:

- **Toggle(p, q):** If edge $\{p, q\}$ is in E then remove $\{p, q\}$ from E , otherwise add $\{p, q\}$ to E .
- **Local Toggle(p, q, r):** If edge $\{p, q\}$ and $\{q, r\}$ are in E then **Toggle(p, r)**.
- **Hop(p, q, r):** If edge $\{p, q\}$ and $\{q, r\}$ are in E and edge $\{p, r\}$ is not in E then remove edge $\{p, q\}$ from E and add edge $\{p, r\}$ to E .
- **Add(p, q):** If $\{p, q\}$ is not in E then add $\{p, q\}$ to E , otherwise do nothing.
- **Local Add(p, q, r):** If edge $\{p, q\}$ and $\{q, r\}$ are in E then **Add(p, r)**.
- **Delete(p, q):** If $\{p, q\}$ is in E then remove $\{p, q\}$ from E , otherwise do nothing.
- **Local Delete(p, q, r):** If edge $\{p, q\}$ and $\{q, r\}$ are in E then **Delete(p, r)**.
- **Swap(p, q, r, s):** If $\{p, q\}$ and $\{r, s\}$ are the only edges between p, q, r and s then remove $\{p, q\}$ and $\{r, s\}$ from E and add $\{p, s\}$ and $\{q, r\}$ to E .
- **Null():** Do nothing.

B. Past solutions to the epidemic duration problem

In earlier research, networks that yielded a long epidemic duration were found to be “banana” shaped, with patient zero at the end of the banana, and the epidemic proceeding down the length of the banana – the thickness of the banana is just enough to prevent the epidemic from burning out early. An example of this shaped network is included in Fig. 7(a). If the probability of epidemic spread is high, then longer, thinner bananas work. In this work we would expect individual districts to become banana shaped.

III. EXPERIMENTAL DESIGN

A. Evolutionary Computation

The representation uses evolutionary computation to generate several potential solutions to the problem using a steady state evolutionary algorithm. The variables with respect to system design were determined empirically.

A population of 1000 chromosomes each containing a string of 256 Local THADS-N operations is used, where each

operation is chosen according to the probabilities given in Table I. When a single string of operations is applied to the graph in Fig. 2, a district is produced. The chromosomes are initially generated at random based upon the probabilities provided to each of the operations via the program parameters. From there, the chromosomes undergo 500,000 mating events, with statistical output being recorded every 5,000 mating events. Each mating event consists of a round of tournament selection, crossover and mutation. Tournament selection selects 7 chromosomes at random from the population. The two chromosomes with the worst fitness are replaced with copies of the two chromosomes with the best fitness. Next, the two copies undergo two-point crossover, and mutation occurs on 1-3 of the operations within that chromosome, replacing them with new commands chosen by the same probability distribution discussed above. The choice of 1-3 mutations is randomly determined with each choice being equiprobable. Lastly, the fitness of the children is recalculated. After evolution the candidate solution with the best fitness from the whole population is saved. This process is repeated 30 times for each parameter setting being tested.

Prior to initiating an epidemic the size and number of districts, k , is used to generate the community; recall that the overall community is modelled as a set of k identical districts which are fully connected to one another as shown in Fig. 1. The connections between the districts provide the ability for an epidemic to spread between districts. To commence each epidemic, the vertex with the lowest index, *patient zero*, from the district with the lowest index, *district zero*, is marked infected and the epidemic is permitted to spread along edges from vertex to vertex within a district. This is known as a *within-district infection*. Within-district infections have probability α of spreading from infected to susceptible individuals within the same district via edges in the graph; each of these probabilities is calculated independently.

The epidemic can also spread from any district with infected individuals to any district yet to experience any infections. This is a *between-district infection* and has probability α' of occurring at each time step, as discussed in Section I-B. When a new district is first infected, it is again *patient zero* who is the first individual infected within that district. This decision allows for successive evolution to capitalize on a known starting location rather than a new location every fitness evaluation, on a per district basis. It is important to note that the fitness measure does not indicate the absolute quality of a network, but rather its relative quality. This permits successive candidate solutions to converge to networks which are more likely to create epidemics satisfying the problem.

The epidemic duration problem calculates fitness by simulating a single epidemic on a candidate solution. The length of the epidemic is measured as the number of time steps until there are zero infected individuals in the (overall) community. This single sample epidemic is used to estimate the fitness of a solution, implying that a mediocre community may attain a fitness value for this single epidemic that is far greater than its mean fitness across epidemics. To circumvent a scenario in

which a solution is provided undeserved reproductive rights *skeptical tournament selection* is used [21]. This modifies traditional tournament selection by recalculating the fitness of the parents after a mating event. This reduces the probability of re-selecting a solution which once had a sample fitness far above its average fitness, although this situation will still occur. The use of skeptical tournament selection has been shown to yield better solutions for the epidemic duration problem [21], hence its use. Additionally, this process favours graphs with less variance as solutions which are chosen for reproduction with increased frequency also have their fitness recalculated more frequently.

B. Point Packing Parameter Selection

When implementing evolutionary algorithms, a significant concern is how to appropriately set the values for the numerous parameters. Frequently, a researcher performs pragmatic preliminary experimentation with various parameter settings to determine which combinations work best. Alternatively, parameters such as population size, mutation rate, crossover rate, etc. may be chosen based upon appropriate ranges that are altered one at a time to determine that parameter's impact on the ability of the algorithm to find good solutions. At the extreme, a full factorial exploration of the parameter space is also an option as evolutionary algorithms commonly interact in a non-linear manner.

The representation used in this study has nine parameters, corresponding to the probabilities associated with each of the nine possible editing operations. A full factorial exploration would grow as the eighth power of the sampling density as the nine probabilities must sum to 1, removing one of the degrees of freedom. Anything less than a full factorial exploration would allow for the optimal parameter setting to never be discovered.

The current study relies on sets of parameters from an earlier study [13]. In that study, the parameters were determined by *point packing*, a mechanism which thoroughly explores the parameter space without the overhead of a full factorial exploration. Both the point packing [2] and full factorial approach result in a grid of points throughout the parameter space, however the point packing results in far fewer points than the full factorial exploration. Point packing also allows the researcher to select the minimum spacing between points in the parameter space, which in turn determines the number of parameter sets to be tested. While the number of parameter sets tested correlates with the degree to which the parameter space is explored, clearly running more parameter sets has a higher cost. Importantly, the costs associated with point packing grow significantly more slowly than the costs associated with the full factorial approach. As with the full factorial design, point packing ensures thorough exploration of the parameter space and objective design removing any researcher bias. It achieves this with far fewer points than a fixed grid.

Point packing is described fully in [2]. In brief, it requires the researcher to choose a value for the minimum allowable distance between parameter sets. The parameter sets used

TABLE I
THE SETS OF LOCAL THADS-N OPERATION DENSITIES USED FOR CONSTRUCTING SOLUTIONS TO THE EPIDEMIC DURATION PROBLEM

| Set | Toggle | Hop | Add | Delete | Swap | L-Toggle | L-Add | L-Delete | Null | Justification for Choice |
|-----|---------|---------|---------|---------|---------|----------|---------|----------|---------|--------------------------|
| 1 | 0.04381 | 0.01885 | 0.02812 | 0.04993 | 0.00027 | 0.43224 | 0.00508 | 0.01696 | 0.40473 | High Fitness |
| 2 | 0.00382 | 0.01564 | 0.02205 | 0.00214 | 0.02848 | 0.02672 | 0.41832 | 0.03961 | 0.44321 | High Fitness |
| 3 | 0.03415 | 0.01919 | 0.00443 | 0.20950 | 0.05859 | 0.33277 | 0.32325 | 0.00469 | 0.01343 | High Fitness |
| 4 | 0.01763 | 0.37651 | 0.01251 | 0.01565 | 0.36368 | 0.00940 | 0.17572 | 0.00993 | 0.01898 | Low Standard Deviation |
| 5 | 0.35718 | 0.20156 | 0.00448 | 0.00787 | 0.00263 | 0.00880 | 0.35559 | 0.05101 | 0.01090 | Low Standard Deviation |
| 6 | 0.01221 | 0.01963 | 0.38384 | 0.09418 | 0.00695 | 0.00673 | 0.03030 | 0.01730 | 0.42884 | Low Standard Deviation |
| 7 | 0.00377 | 0.04566 | 0.40664 | 0.07685 | 0.02258 | 0.00606 | 0.39520 | 0.02680 | 0.01644 | Both |
| 8 | 0.00197 | 0.07271 | 0.00097 | 0.04354 | 0.03551 | 0.01226 | 0.79455 | 0.02013 | 0.01836 | Both |

TABLE II
THE VARIABLES ASSOCIATED WITH THE VARIOUS EPIDEMIC COMMUNITIES FOR THE EPIDEMIC DURATION PROBLEM

| Community | k | district Size | Total Vertices | α' |
|-----------|-----|---------------|----------------|-----------|
| A | 1 | 512 | 512 | N/A |
| B | 4 | 128 | 512 | 0.10 |
| C | 4 | 128 | 512 | 0.05 |
| D | 8 | 64 | 512 | 0.10 |
| E | 8 | 64 | 512 | 0.05 |
| Original | 1 | 128 | 128 | N/A |

in the current study were selected as they were the top-performing from amongst those generated by the point packing in [13], which also considered the problem of maximizing epidemic duration, albeit for a different epidemic model and only for single communities of a fixed size. In that study, a minimum distance of 0.535 was specified, generating some 29 parameter sets.

C. Description of Experiments

Eight sets of parameters are used in the current study. Of the 29 parameter sets from the previous study, we selected the five that produced results with the highest fitness, and the five that produced results with the lowest standard deviation resulting from 30 runs in the previous study. Listed in Table I, it can be seen that two parameter sets satisfied both of these requirements, therefore there are eight parameter sets in total.

The densities for a given parameter set are designed to add up to 1.0. It may appear that there are some rather unusual choices of values for parameters in comparison to a traditional parameter sweep. For example, in set 1 toggle has a probability (density) of 4.381% of being selected for any given entry in the chromosome. A more traditional parameter sweep would be more likely to choose probabilities that were (for example) multiples of 5 or 10%. This type of situation, however, is a feature of point packing and it is one of the reasons it is able to explore the parameter space so well.

Each of the eight parameter settings is applied to five communities, listed in Table II as communities A–E. The overall number of individuals in a community is held constant at 512. Each of these individuals belongs to one of 8 separate districts each of size 64 (communities D and E), 4 separate districts each of size 128 (communities B and C), or a single district of size 512 (community A). For all communities, the probability of within-district infection is held constant at

$\alpha = 50\%$. The probability of between-district infection is either $\alpha' = 5\%$ or 10% for 4 or 8 districts, and does not apply in community A as this consists of only one district. For comparison purposes, we also consider the problem of epidemic duration on a community with a single district of size 128 (listed as “original” in Table II).

IV. RESULTS AND DISCUSSION

The results shown in Fig. 4 represent the box and whisker plots of the maximum epidemic lengths achieved by 30 runs of the evolutionary algorithm. Along with these, the results achieved from epidemics on a community comprised of a single district with 128 vertices is included within Fig. 4(f) and labelled as “original”. This is provided as a comparison against the five communities totalling 512 individuals each (communities A–E) implemented in this study, allowing some insight as to what happens within a single district which is of average size in comparison to the others.

A rough sine wave pattern can be seen between parameter sets across the communities, although the variance from this pattern is observed most dramatically when four districts are used, as in communities B and C. In contrast, using eight districts flattened the observed pattern between parameter sets. One outlier, dominated by the hop and swap operations, is parameter set 4; in this case, the number of districts has a negative impact on the fitness achieved regardless of the value of α' . Parameter sets 2 and 8 each result in the best values for three of the communities; Original, D, and E yielded the best results with parameter set 2 and the rest did so with parameter set 8. As parameter set 2 relies on local add and null a large number of local connections will exist between district members when compared to those without these conditions. Parameter set 8 uses mostly local add, with marginal probability attributed to the other edge operations. Parameter settings with a sizeable use of the local add operation have been among point packing’s most fit results since its introduction to the representation in [11], [13].

The epidemic data from the communities is combined into Fig. 5. Most glaringly, it is apparent that, of the communities tested, the optimal one for maximizing epidemic duration placed all individuals into a single district. This is based on the fact that community A clearly outperforms all the other communities tested. The addition of separate districts was expected to allow the epidemic to spread more slowly; ideally, the epidemic would spread to another district just as

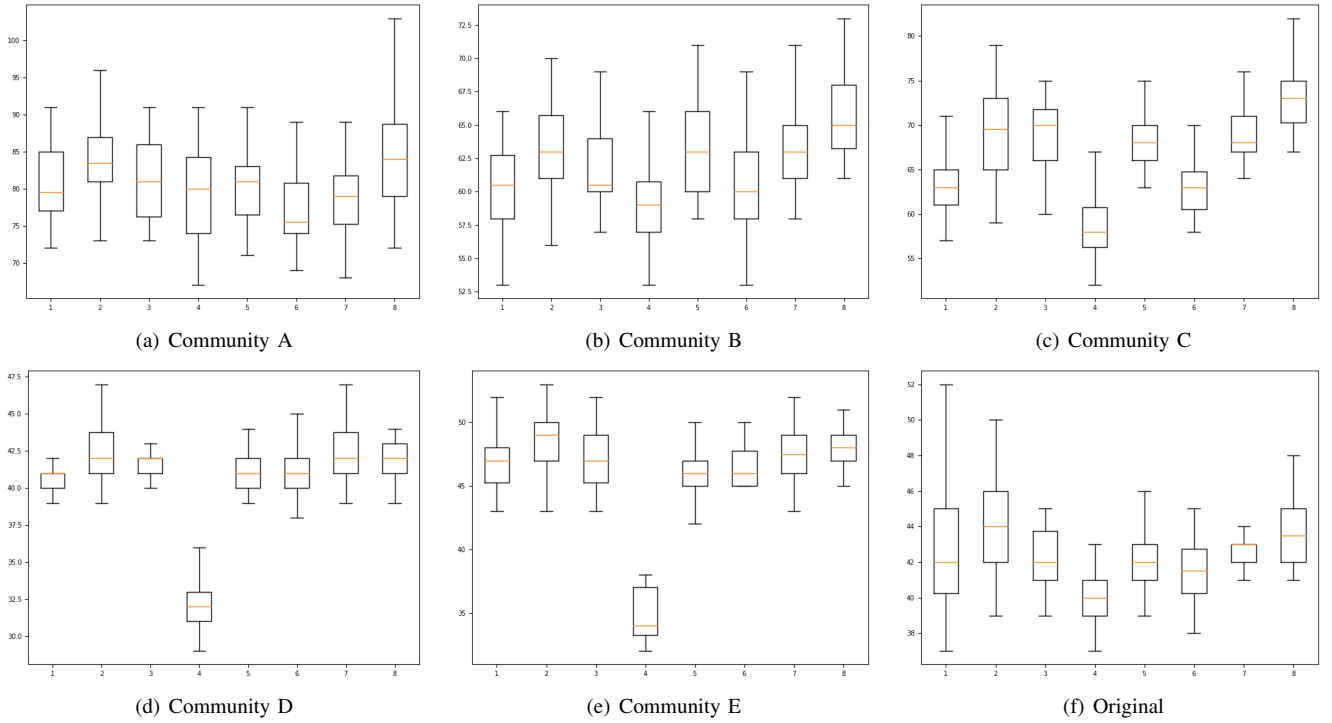


Fig. 4. Box and whisker plots of the maximum epidemic duration from 30 runs of the evolutionary algorithm across all communities and all operation probability densities.

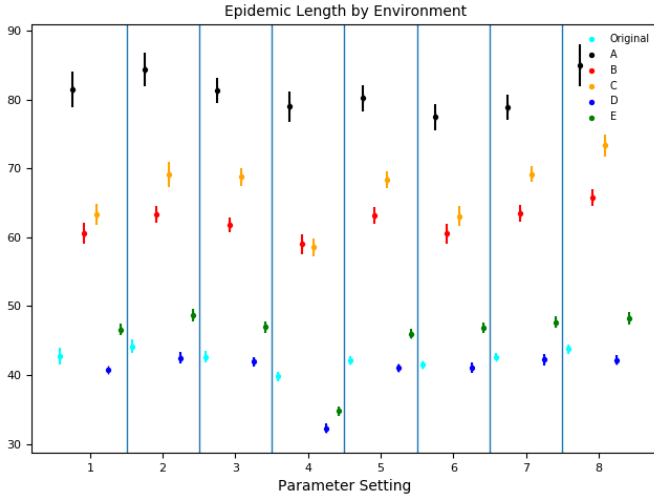


Fig. 5. The mean of the maximum epidemic duration, with 95% confidence interval, from 30 runs of the evolutionary algorithm for all communities.

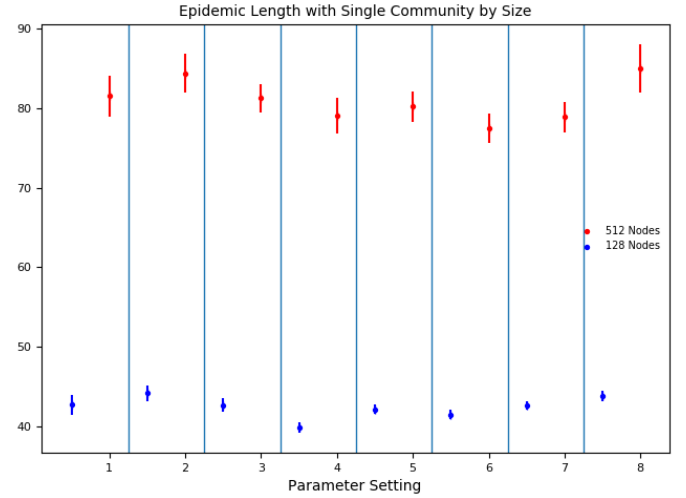


Fig. 6. The mean of the maximum epidemic duration, with 95% confidence interval, from 30 runs of the evolutionary algorithm on communities with a single district.

it had infected the last members of its current district. This situation, however, did not materialize; likely this was due to a value of α' which was too high, resulting in the epidemic spreading too quickly between districts. This is evidenced by the fact that decreasing the value of α' provided significant fitness increases for all but one (parameter set, community) pair, namely parameter set 4 when comparing communities B (with $\alpha' = 0.10$) and C (with $\alpha' = 0.05$ and all other settings identical to B). The significant impact observed by changing

the values of α' , the number of districts, and the district size indicate that further optimization is required in order to best pair community settings with probability densities.

Furthermore, Fig. 6 features two communities, each comprised of a single district with either 128 or 512 vertices. This was included to demonstrate how a change to only the district size can impact the fitness evaluation. Obviously, the larger district is able to sustain a longer epidemic as there are more

individuals to infect, so the increase is unsurprising. However, note that the four-fold increase in the number of vertices, from 128 to 512, only results in an approximately two-fold increase in epidemic duration, from 39.83–44.17 for 128 vertices to 77.5–85.03 for 512 vertices. It is also important to note that any increase in duration also coincides with an increase in standard deviation, as a wider range of possibilities are present regardless of other factors being considered.

Lastly, Fig. 7 provides visual representations of three districts which achieved the best fitness for communities with districts of size 128 using parameter setting 8. Recall that the three districts represented in the figure differ in both their α' values and fitness evaluations. The original community, with a single district, will have evolved differently than those within communities B and C as the fitness value depends on the length of an outbreak within one district as compared to four in communities B and C. The personal contact network from previous work appears to have diverged more from the initial ring structure than the districts from B and C; these networks have a significant area of their networks roughly resembling the starting ring. In contrast, the original district resembles the “banana” shape which has been a common trend within previous work, though there still exists a section of the network with a ring shape. Looking closely at the networks from the districts for communities B and C can provide some insight into how α' , which is the probability of transmission between districts, impacts network evolution. Community B’s district, with $\alpha' = 0.10$, features a smaller tail and a larger proportion of the population still within the original ring structure as compared to community C with $\alpha' = 0.05$. Looking again at the original community, which has no α' value as it is the entire community, a pattern emerges: lacking an α' value is effectively the same as an α' value of 0.00. Therefore, as α' increases the districts resemble the original structure more closely. It is also interesting to note that “patient zero”, the first person infected in a given district, is at the end of the tail in all cases.

V. CONCLUSIONS AND FUTURE WORK

The inclusion of districts with the model has allowed for the Graph Evolution Tool (GET) to increase its functionality when it comes to the goal of being able to use the model to simulate epidemics on human populations. However, there are several additional areas of development available to further improve the model.

A. Between-district and within-district infection rates

A logical starting point would be to further investigate the effect of modifying the values of α' , the between-district infection rate and α , the within-district infection rate. In the current study α always had a fixed value of 0.5. Meanwhile, this study demonstrated that a lower value for α' had a positive effect upon epidemic duration. A more thorough investigation of the settings for these two parameters would provide greater insight on their impacts across the communities investigated here as well as other, larger, social contact networks. A further



Fig. 7. The personal contact networks of size 128 representing the district which resulted in the longest epidemic with parameter set 8 within the designated community. Patient zero is labelled for clarity.

possibility is the implementation of α' as a function based on the number of infected individuals or some combination of properties within a community or its districts.

B. Improved point packing

In this study we used point packings that were generated uniformly at random and then normalized to generate probabilities for the various graph editing operations that add to 1.0. It turns out that this method of finding point packings can be problematic, as it has a bias toward the probability $1/n$ where n is the number of parameters being set. A more complex form of normalization permits uniform sampling of the simplex of points with positive coordinates summing to 1.0. While the parameter setting performed here is not incorrect, a more complete sampling might be possible with the new technique.

C. Epidemic Model and Community Properties

Previous work included the use of two epidemic models, the SIR model used here, and the SIRS model where an individual once again becomes susceptible some number of time steps after recovering from infection. It would be worthwhile to study the impact of the SIRS model when used with districts. Another modification could include the ability to represent voluntary vaccination rates based on the severity of the epidemic among other factors as presented in [20]. Additionally, it would be beneficial to thoroughly investigate the number of districts and their sizes, along with the possibility of connecting the districts in an adaptive manner, or at least not simply as a fully connected graph.

D. Algorithm Complexity

As the representation has grown and evolved, so too has the running time. A necessity of epidemics lasting longer is that the fitness calculations, which are a simulated epidemic, take longer to calculate. Therefore, as the model becomes more complex, complications will compound and further iteration, without simplification, will become increasingly challenging. Furthermore, large increases to community size would allow for the model to be used on larger human populations, although this would also dramatically impact the computation time.

E. Comparing Solutions and Visualization of Graphs.

There exist other tools which can generate graphs to test against their ability to model epidemic behavior. Therefore, a future area of research could be to compare the Local THADS-N representation against other solutions. This work could include examining the success of these methodologies in modeling real world epidemic behavior. Furthermore, it would be interesting to examine the actual graphs generated, including possibly by using graph visualization tools in order to view epidemic behaviour on a community.

REFERENCES

- [1] D. Ashlock and L. Barlow. A class of representations for evolving graphs. In *Proceedings of the 2015 Congress on Evolutionary Computation*, pages 1295–1302, 2015.
- [2] D. Ashlock and S. Graether. Conway crossover to create hyperdimensional point packings, with applications. In *Proceedings of the 2016 Congress on Evolutionary Computation*, pages 1570–1577, Piscataway, NJ, 2016. IEEE Press.
- [3] D. Ashlock and F. Jafarholi. Representations for the evolution of extremal epidemic networks. In *Proceedings of the 2008 World Congress on Computational Intelligence*, pages 660–667, Piscataway NJ, 2008. IEEE Press.
- [4] D. Ashlock and C. Lee. Characterization of extremal epidemic networks with diffusion characters. In *Proceedings of the 2008 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 264–271, 2008.
- [5] D. Ashlock, J. Schonfeld, L. Barlow, and C. Lee. Test problems and representations for graph evolution. In *Proceedings of the IEEE Symposium on the Foundations of Computational Intelligence*, pages 38–45, 2014.
- [6] D. Ashlock and E. Shiller. Fitting contact networks to epidemic behavior with an evolutionary algorithm. In *Proceedings of the 2011 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 1–8, Piscataway NJ, 2011. IEEE Press.
- [7] D. Ashlock and M. Timmins. Adding local edge mobility to graph evolution. In *Proceedings of the 2016 IEEE World Congress on Computational Intelligence*, Piscataway NJ, 2016. IEEE Press.
- [8] D. A. Ashlock and F. Jafarholi. Behavioral regimes in the evolution of extremal epidemic graphs. In *Proceedings of the 2008 Congress on Evolutionary Computation*, pages 660–667, 2008.
- [9] MM Telo Da Gama and A Nunes. Epidemics in small world networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1-2):205–208, 2006.
- [10] A. Devert, T. Weise, and K. Tang. A study on scalable representations for evolutionary optimization of ground structures. *Evolutionary computation*, 20(3):453–472, 2012.
- [11] M. Dubé. The local thads-n generative representation for epidemic profile matching. Undergraduate thesis, Brock University, 05 2018.
- [12] M. Dubé, S. Houghten, and D. Ashlock. Parameter selection for modeling of epidemic networks. In *Proceedings of the 2018 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2018.
- [13] M. Dubé, S. Houghten, and D. Ashlock. Representation for evolution of epidemic models. Submitted for Publication, 01 2019.
- [14] Yuanzheng Ge, Zhichao Song, Xiaogang Qiu, Hongbin Song, and Yong Wang. Modular and hierarchical structure of social contact networks. *Physica A Statistical Mechanics and its Applications*, 392:4619–4628, 10 2013.
- [15] P. F. Hingston, L. C. Barone, and Z. Michalewicz. *Design by Evolution: Advances in Evolutionary Design*. Springer, New York, NY, 2008.
- [16] F.C. Hoppensteadt and C.S. Peskin. *Mathematics in medicine and the life sciences*, volume 10. Springer Science & Business Media, 2013.
- [17] G.S. Hornby. Improving the scalability of generative representations for opened design. In R. Riolo, T. Soule, and B. Worzel, editors, *Genetic Programming Theory and Practice V*, Genetic and Evolutionary Computation Series, pages 125–142. Springer, Boston, MA, 2008.
- [18] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- [19] T. Kowaliw, N. Bredeche, and R. Doursat. *Growing Adaptive Machines*. Springer, 2014.
- [20] Ana Perisic and Chris Bauch. Social contact networks and disease eradicability under voluntary vaccination. *PLoS computational biology*, 5:e1000280, 03 2009.
- [21] M. Timmins and D. Ashlock. Network induction for epidemic profiles with a novel representation. *Biosystems*, 162:205–214, 2017.

ACKNOWLEDGMENTS

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).