Topographic Spreading Analysis of an Empirical Sex Workers’ Network

Johannes Bjelland, Geoffrey Canright, Kenth Engø-Monsen, Valencia P. Remple

2007
Topographic Spreading Analysis of an Empirical Sex Workers’ Network

Johannes Bjelland¹, Geoffrey Canright¹, Kenth Engø-Monsen¹, and Valencia P. Remple²

¹ Telenor R&I, 1331 Fornebu, Norway
  johannes.bjelland,geoffrey.canright,kenth.engo-monsen@telenor.com
² BC Centre for Disease Control Epidemiology, University of British Columbia, Vancouver, BC, Canada Valencia.Remple@bccdc.ca

Summary. We present an application of the ”topographic” analysis of spreading on networks to the empirically obtained network of female sex workers (FSW) and their customers in Vancouver, Canada. The topographic analysis uses eigenvector centrality (EVC) to define regions (subnetworks) in which spreading is fast and fairly predictable in its progression; this approach has been well supported by simulations. Here we present the first application to a sexual network—furthermore, one for which the ”starting” (currently) infected nodes are known. The data source is a network study conducted using interviews with indoor FSW in Vancouver, Canada. The network obtained is fully connected, with 553 nodes and 1500 edges. We present a novel generalization of the topographic approach to weighted graphs, and show that this approach is successful (as in the unweighted case) in predicting the progress of disease spreading. We find however that our model for link strengths for HIV gives a highly disconnected graph, for which the problem of HIV spreading essentially vanishes. For the other diseases (gonorrhea and chlamydia) we find that the obvious remedies suggested by our analysis are defeated by a combination of very dense connections and very central infected nodes.

1 Introduction

The problem of epidemic spreading over networks has received considerable attention in recent years, due both to its intrinsic intellectual challenge and to its practical importance. A good recent summary of such work may be found in Newman [1], while [2] gives an outstanding example of a nontrivial prediction which is obtained from explicitly modeling the network in the epidemic spreading. Most such work focuses on whole-graph properties, such as the percentage of infected nodes at long time. Two of us have in contrast focused on understanding the spread of an infection over time and space (the network) [3, 4, 5]. This work involves decomposing any given network into subgraphs called regions [3]. Regions are precisely defined, disjoint sugraphs
which may be viewed as coarse-grained units of infection—in that, once one node in a region is infected, the progress of the infection over the remainder of the region is relatively fast and predictable [4]. We note that this approach is based on the ‘SI’ model of infection, in which nodes, once infected, are never cured. This model is reasonable for some infections, such as HIV—which is one of the diseases studied here. We also study gonorrhea and chlamydia, for which a more appropriate model is SIS [6] (since nodes can be cured); we discuss the limitations of our approach for these cases below.

In this paper we apply the ‘topographic’ regions-analysis approach to an empirical sex network, built from interviews with female sex workers (FSW) in Vancouver, Canada. The network consists of the FSW’s themselves, plus their sex partners (paid and unpaid), as well as any partners of these partners which were known to the FSW. This method, beginning with 49 interviewed FSW, gave a highly connected network of 553 nodes [7]. Furthermore, STI (sexually transmitted infection) status was obtained for many of these nodes. In particular, two of the nodes were identified as being HIV-positive, while 11 other nodes have either gonorrhea, chlamydia, or both.

The aims of the present work are several. One goal is to extend our earlier topographic approach to a graph with weighted links. As we will see, this seemingly small change can have very large effects; but we will also see that the validity of our approach is confirmed, in spite of these large effects. This is because the modified approach (presented here for the first time) is consistent: we use the link weights to modify the graph’s adjacency matrix, and hence the nodes’ EVC values; and we use them again when we define the regions via the ‘steepest-ascent graph’ or SAG.

A second aim of this work is to try to exploit the insights gained from the topographic analysis, in order to find novel suggestions for preventive actions which may be taken to hinder the spread of the disease in question. We find that our progress towards this second goal is considerably more modest than that towards the first goal. We will show ‘thought experiments’, based on the empirical graph topology and link strengths, for which our analysis is extremely useful. However, we will not find practical suggestions which are immediately promising for the given Vancouver FSW graph. There are several reasons for this. First, the HIV graph is so thoroughly protected by condom use that we find little to add in terms of ideas for preventive measures. Secondly, the graphs for gonorrhea and for chlamydia are so thoroughly well connected, and also so well infected, that we do not find small topological changes which can make a large difference.

We note that our approach treats the network as static; hence any effects of network dynamics are not taken into account. We believe however that our qualitative results are fairly robust to the likely dynamics of this network, since its overall structure is thought to be fairly stable over time. Also, our analysis (once the network is mapped out—which can be time consuming!) is not computationally demanding, and so may be performed in essentially zero time compared to the time scale of epidemic spreading. Hence any suggestions
coming from the analysis may be implemented in something approaching real time.

2 Uniform transmission model

First we study the FSW graph without taking into account the link weights. That is, each sexual contact is given strength ‘1’ in the adjacency matrix. This is logically equivalent to giving each link the same probability of transmission per unit time.

Our purpose in doing this analysis is to be able to compare with the analysis using non-uniform link strengths (transmission probabilities). As we will see, the differences are large and important.

2.1 Visualization and biparteness

Our topographic analysis includes a novel approach to graph visualization: we group the nodes into their respective regions, and lay out the whole graph according to the ‘steepest-ascent graph’ [8]. We present the basic ideas here, and refer the reader to earlier papers [3, 4, 8] for details. We view the eigenvector centrality (EVC) of a node as a measure of its ‘well-connectedness’, and hence of its ‘spreading power’. Then we single out local maxima of the EVC as being particularly important in spreading; we call these nodes Centers. Also, since EVC (being recursively defined) is ‘smooth’, we can speak of ‘neighborhoods’ in the graph as having a typical EVC; and we conclude that spreading is fast in neighborhoods of high EVC, and slow in ‘lower’ neighborhoods. We then go on to define regions of the graph—one for each Center. Each node finds its region (mountain) by following a steepest-ascent path until it terminates at a local maximum (mountaintop, or Center). The set of steepest-ascent paths then forms a directed hierarchical tree graph (the steepest-ascent graph or SAG), which is useful both for visualizing the graph and for predicting the likely paths of fastest epidemic spreading.

The SAG for the unweighted FSW graph is shown in Figure 2. We note several interesting points from this visualization: (i) there are many regions (17). (ii) All the Centers (most central nodes in each region) are men. (iii) Many regions are small, ie, 1–3 nodes, while (iv) the bulk of the nodes (517/553) lie in one of the three largest regions (red, blue, dark grey). (v) Every region is well connected to the largest, red, region. Hence the red region is expected to play a dominant role in any epidemic spreading. (vi) One HIV-positive node is in the red region, and the other is (while in its own region) well connected to the central part of the red region.

Now we comment on these points. We believe that points (i)–(iii) derive from the fact that the graph is nearly bipartite. For example, for a bipartite graph composed of M and F nodes, if an M node is a Center (local maximum of centrality), then all of its neighbors are (a) female, (b) highly central, and (c)
Fig. 1. Regions visualization of the FSW network, with all links set to equal strength. Only the links in the steepest-ascent graph (SAG) are shown here for visual clarity. Each region has its own color, and the most central node in each region is enlarged.

Fig. 2. Same visualization as Figure 1, except that all links are shown. The yellow arrows mark the known HIV-positive nodes.
Topographic Spreading Analysis 5

automatically excluded from being a Center. Thus bipartiteness will tend to favor one gender over another. By the same token, highly central M nodes are never neighbors of other M Centers, and so are candidate Centers themselves. Hence there may be a tendency for more, and smaller, regions.

Points (iv)–(vi) tell us that this network is highly prone to infection: the many regions are not well isolated from one another, because of their common connection to the dense, infectious red region. Also, the two start nodes are in or near the central part of the red region, where spreading is fast.

2.2 Infectious spreading on the unweighted graph

We have simulated spreading on the uniform FSW network, by giving each link the same probability per unit time for spreading. The value used is thus arbitrary, as is the unit of time. We typically use a value of a few percent, since much larger values give very unsmooth time evolution (equivalent to poor time resolution). We report the results here because they are illustrative of the strengths and weaknesses of our method, for the case of multiple regions. (For reasons given below, these are the only multi-region simulations that we can perform with this graph.)

Taking the start (infected at $t = 0$) nodes as shown in Figure 2 above, we find, as expected, that the regions as we define them here are again valid coarse units of infection. We also find that it is difficult to stop or even retard the infection, because of the topology of the graph. The upper part of Figure 3 shows a typical epidemic progression, with the growth in the red, blue, and grey regions resolved. All three “take off” at about the same time, and the infection spreads rapidly. Measures to retard spreading in the red region—without resorting to large topological change—are not found to be effective.

We find however that protecting one node—the Center of the grey region—drastically weakens the red/grey connection. We see in the bottom part of Figure 3 the results when this is done: the red and blue regions take off as before, but the grey region’s takeoff is greatly retarded. This is an example of the kind of benefit that we believe can be obtained from our analysis.

We also considered the more promising problem of an infection starting in the grey region—again motivated by the observed red $\iff$ grey bottleneck in the topology. The top of Figure 4 shows that takeoff is retarded by a factor of about 3, compared to the former case (top of Fig. 3). It is retarded even further (about 7 times as slow) if we in addition protect the grey Center (bottom of Fig. 4).

3 Links weighted with transmission probabilities

In this section we add an important further element of realism by weighting the links of our FSW graph with transmission probabilities. We are forced in many cases to use rather crude approximations. Nevertheless we feel that
the resulting model is considerably closer to reality than the uniform model. Also (as we will see) it is strikingly different—in particular, each disease will have its own graph. That is, while the basic topology is the same as that in Figure 2, the set of link weights depends on the disease—because these weights represent transmission rates (probability/time). In fact, for the HIV case, the topology itself is changed, since we set some link strengths to exactly zero.

In practice, incorporating the link strengths into the analysis involves (1) building a weighted adjacency matrix $W$ using the link strengths, (2) finding the corrected EVC as the dominant eigenvector of this matrix $W$, and (3) redefining 'steepest ascent' to take account of the varying link strengths. The first two steps are clear; and we describe step (3) in Section 3.2.

Of course, before doing any of this, we must find the link strengths. We describe our procedure for doing so in the next section.

### 3.1 Estimating the probabilities

For each link we want a single weight (number) which gives the probability per unit time of transmission from an infected node to an uninfected node. This probability is based on a number of factors which must be estimated from limited data. We list these factors schematically as follows:
Transmission probability/unit time =

\[
\left[ \left( \text{unprotected probability/contact} \right) \times \left( \text{non-condom use prevalence} \right) \right] \\
+ \left[ \left( \text{protected probability/contact} \right) \times \left( \text{condom use prevalence} \right) \right]
\]

Now we discuss each factor in turn. For each disease (HIV, gonorrhea or 'NG', and chlamydia or 'CT') we estimate (unprotected probability/contact) from reference [9]. See Table 1. To correct for condom use, we must know the frequency of condom use for each link (condom use prevalence). For 256 links (about 17% of them) we have an estimate for (condom use prevalence) from survey data [7]. We know very little about the remaining links, except for whether they are a “client” relationship or a “non-client” relationship. We explain below how we generate link weights for the links for which we have no survey data.

Estimates for (contacts/time) were available (again) for those links for which we obtained survey information; however, here we have yet another source of uncertainty. That is, each interviewed FSW reported contacts with “regulars”, and also contacts with new or “non-regular” customers. We take the reported estimates of (contacts/time) for regulars as given. For the non-
Table 1. Transmission-probabilities/contact for NG (gonorrhea), CT (chlamydia) and HIV.

<table>
<thead>
<tr>
<th></th>
<th>NG</th>
<th>CT</th>
<th>HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unprotected</td>
<td>0.43</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>Protected</td>
<td>0.16</td>
<td>0.074</td>
<td>0</td>
</tr>
</tbody>
</table>

regulars, we assume that either (i) they will become regular in the future, or (ii) they will be replaced by other non-regular customers who play essentially the same role in the network. In short: we ignore the distinction between case (i) and (ii).

We still need a reasonable estimate of contacts/time for non-regulars. We proceed as follows: for each FSW, we define $T$ to be the total number of contacts per unit time (summed over all neighbors). Also we let $P$ be the percentage of contacts from regulars, and let $C$ be the number of contacts/time from regulars. Then clearly $C = PT$; and since we can estimate both $P$ and $C$ from the survey data, we get an estimate of $T = C/P$. We then estimate the total contacts/time $N$ for non-regulars to be $N = T - C$. Finally, we take, from the survey data, the expected number of non-regular neighbors (still for each FSW), and call this number $K$. We then (finally) get the expected contacts/time for each non-regular as $N/K$. Our model is clearly very crude, treating each non-regular in a very average way; but it enables us to move (as we will see) well beyond the equal-transmission-probability model, and so, we believe, moves us much closer to reality.

Now we come to the term due to protected sex. We estimate (protected probability/contact) by correcting the (unprotected probability/contact) data, using data for the correction due to condom use from [10]. We note here that we set (protected probability/contact) for HIV to be exactly zero. Not surprisingly, this will have dramatic effects on the spreading behavior—as we will see in Section 3.3.

This completes our prescription for estimating link weights for those links for which we have survey data. We then used a very simple approach—which we find appropriate to the high degree of uncertainty in our data—to estimate the remaining link weights (transmission probabilities/time). Our solution here is to first divide all links (surveyed and not surveyed) into two groups: client and non-client. Then, for each group, we simply reproduced the distribution over the “surveyed” links so as to also decorate all of the “non-surveyed” links. Since the survey data is discrete, the link-weight distribution obtained is never smooth. Hence we reproduced these discrete distributions by simply repeating (sampling) each value in the discrete distribution with a probability equal to its frequency in the distribution. That is: we do not attempt to create distributions for each parameter in the link-weight estimate; instead we simply copy the discrete link weight values obtained from the survey data onto the unknown links, with appropriate probabilities.
3.2 SAG*

Now we address another complication arising from the use of weighted links: we must reconsider the definition of the steepest-ascent graph (SAG), which is used both for assigning region membership and for visualization purposes. Our point here is simple, namely that the definition of steepest ascent should take account of the link strength. This rather obvious point has not been addressed in our earlier use of the SAG [3, 4, 5], because these earlier studies were applied to unweighted graphs. Hence we offer a brief account here of the modification used for weighted links.

We recall that region membership is assigned by in essence asking each node to find the steepest path to the “top”—i.e., to the “nearest” local maximum of the EVC. The notion of local maximum is independent of link strength. Suppose however that a node \( N \) has two local maxima (Centers, \( C_1 \) and \( C_2 \)) as neighbors: which region do we place \( N \) in? Since we want steepest-ascent paths to represent most likely spreading, it seems reasonable that a neighbor \( C_1 \) with a very weak link to \( N \) should not be assigned the steepest-ascent path—even if it is somewhat higher (in EVC) than \( C_2 \). In other words, if we retain the notion that steepest ascent gives the right answer, then we clearly want to define the slope as being

\[
\text{slope} = \frac{\Delta y}{\Delta x},
\]

with \( \Delta x \) (‘distance’) decreasing with increasing link strength.

Clearly, \( \Delta y \) is the EVC difference, as in earlier (unweighted) work; hence we simply need some reasonable definition for the ‘distance’ \( \Delta x \). We take here the simple heuristic \( \Delta x(i, j) = 1/W(i, j) \) with \( W(i, j) \) the link strength (transmission probability) between nodes \( i \) and \( j \). Our point here is then that node \( N \) may find that it is not simply in the region of its highest neighbor: instead, it will be placed in the same region as the neighbor \( N^* \) with the highest product \( \Delta y/\Delta x = [EVC(N^*) - EVC(N)][W(N, N^*)] \). In short: if its link to the highest neighbor is very weak, then (reasonably) it will be placed instead in the region of a neighbor with a stronger link. We believe this is consistent with our aim for defining regions—namely, that a region is a coarse grained unit of infection, such that infection within a region is relatively fast and predictable.

We call the resulting steepest-ascent graph SAG* (to distinguish it from the SAG, which does not take link strengths into account). We will see below that our spreading simulations can only give a limited test of our SAG* definition—since in one case (HIV) the weighted network breaks down, while in the other two (NG and CT) we only obtain a single region. Hence—while we retain a belief that our definition is promising—a thorough test will have to await application to a weighted graph which (i) has several regions, but yet (ii) is better connected than our HIV graph of Section 5.
3.3 HIV graph

The SAG* for our weighted HIV graph is shown in Figure 5. We see immediately that the contrast with Figure 2 is enormous.

Fig. 5. Regions analysis for the graph of Figure 2, corrected with the transmission probability on each link. Note that the graph breaks into very many small regions, due to the (assumed) zero transmission probability for reliable condom use. The two enlarged nodes are known to be HIV-infected.

In particular, the 17 regions of Figure 2 have multiplied many times. In addition (which is not so easily seen in the figure) some nodes are completely disconnected due to the zero-weight links, and hence do not appear in the figure at all. The apparently isolated nodes in the corner of the figure are one-node regions; such regions occur typically on the periphery of a graph, where all EVC values are small.

What is even more striking is that adding all nonzero links to the SAG* picture of Figure 5 makes very little change; that is, there are only six nonzero links which are not shown in the figure (four connecting the one-node regions to one other node each, and two other inter-region links). Hence we do not show the full graph: it is essentially that of Figure 5. This means in turn that HIV spreading—while seemingly unstoppable in the picture obtained from Figure 2—is in fact not a problem for this FSW network. In particular, the
two HIV-positive (male) nodes (marked with large squares in Figure 5) are each confined to an effective two-node network, consisting of themselves and their nonclient partner. Hence our expected picture of condom use for this empirical network implies that HIV spreading will be limited to the nonclient partner relationships of the two infected nodes, and so has effectively zero probability of reaching the rest of this dense sexual network.

Because the effective graph is so fragmented, and also because the HIV-infected nodes are effectively isolated, we have not performed spreading simulations on the weighted HIV graph. We note that the largest region in Figure 5 has 24 nodes, with a FSW as the most central node in the region. In fact the strongly bipartite picture obtained from the unweighted graph (Fig. 1) has also broken down here: both male and female Centers of the many regions are found. This is however not so surprising, given the fragmented nature of the effective graph.

### 3.4 Gonorrhea

Figure 6 shows the steepest-ascent (SAG$^*$) graph when we use link strengths appropriate to gonorrhea. Since 100% condom use does not give 100% protection [10], the effective gonorrhea graph has all the same links as were present in Figure 2; but they are reweighted. We see that the reweighting has still had a dramatic effect. In particular, the 17 regions found for the unweighted graph are now a single region for the weighted graph. Also, the Center of this one region (and so of the entire graph) is a FSW.

An interesting aspect of the gonorrhea SAG$^*$ is that one of the few existing homosexual (FSW $\iff$ FSW) links plays a very central role in the graph: the link between the Center and the head of the large red subregion is homosexual. This means that the two women involved are highly central in the weighted graph, and also that the link strength between them (transmission probability for gonorrhea) is not too small. One might then propose to remove this link—which (as it is certainly requested and paid for by a male customer) should be possible. As we will see below, however, removal of this link—or any single link—has little or no beneficial effect. (This conclusion is perhaps intuitively grasped from the fully linked visualization of Figure 7 below.)

Steepest-ascent graphs of either type are strict hierarchical structures—that is, they are directed trees, with links pointing strictly towards the root (Center). This means that, for any given region, one can readily define subregions in terms of branches of the tree. We have picked out the five largest branches of the gonorrhea SAG$^*$, and color coded them. We see that it is visually meaningful to think in terms of subregions for this region.

Figure 7 shows the NG-graph again, but with all links displayed. We note that presently infected nodes are marked yellow in Fig. 6 and in Fig. 7. From Fig. 6 we see two infected nodes lying at the heads of their (large) respective subregions, and hence only one hop from the Center. Also we see that every
major subregion is already infected. This immediately suggests that preventing the further spreading of gonorrhea on this graph will be quite difficult.

This pessimistic prognosis is also supported by the visualization of Figure 7. Here we see that all the major subregions are well connected to one another, with infected nodes lying in the heart of a dense cloud of links. We will test (and confirm) this pessimistic prediction via stochastic simulations—see Section 4 below.

3.5 Chlamydia

In Figure 8 we show the SAG∗ visualization of the chlamydia graph. Qualitatively we see much the same picture as for the NG graph: a single region, with a FSW at the Center of the region. In fact, the homosexual dyad that we found lying centrally in the NG graph is also central here—with the one difference that here the two FSW’s have exchanged roles (Center and subregion head).

Our SAG∗ visualizations suggest that the CT graph is perhaps even more well connected than the NG graph—in that there are very few subregions, and they are very large. And since (again) every major subregion is infected,
we arrive at the same qualitative prognosis for this graph: it will be difficult to hinder the further spreading of the disease.

We have also plotted the analog of Figure 7 for chlamydia—that is, the full graph with all nonzero links. The result is again qualitatively like that of Figure 7; hence we do not show it here.

4 Spreading on the gonorrhea-graph

For reasons already given above, we have not run spreading simulations on all three disease-graphs. The HIV-graph is so heavily disconnected by the many condom-use-induced zero links that we see no point in running simulations on it. These links of course, involving as they do real sexual contact, do not have exactly zero probability for infectious spreading, even with 100% condom use. Also the reported rates of 100% condom use are most likely overstated in many cases. Hence it would be of interest to set the strength of these “zero HIV links” to some small but positive value, and to examine the resulting graph. We reserve this idea for future work.

The remaining two graphs (NG and CT) are qualitatively very similar. Hence we have chosen to focus on one of them—the NG (gonorrhea) graph. We must emphasize immediately however that our simulations, being based on
SI dynamics [1], do not accurately model the long-time dynamics of diseases such as gonorrhea and chlamydia. A more appropriate model would be the SIS model [6] in which Infected nodes become again Susceptible after a variable time period.

We expect the SI model to give qualitatively correct results in the early stage of any infectious process—when few nodes are infected, and they have not had time to recover. Beyond this early stage the SI model can only over-estimate the degree of spreading. Hence we present simulation results in this section, based on the SI model, with two principal caveats:

- Takeoff of the disease will likely occur later for the more realistic SIS model than that we show here.
- The long-time infected fraction will not approach 100%, but rather a lower value.

With these caveats clearly in mind, we present some simulations on the gonorrhea graph. Our aim is to see what insights we can gain from our SAG+ picture. We will focus principally on when the infection takes off. Because we simply compare different scenarios (and their takeoff times) with one another, we feel that our (comparative) conclusions are not greatly weakened by the caveats given above.

**Fig. 8.** Region (SAG*) visualization for the chlamydia network CT. Yellow nodes are known STI-infected nodes.
Our procedure for simulation is the same as before: at each time step, each link $ij$ has a probability $p_{ij} = W(i,j)$ of transmitting the infection if exactly one of the pair $ij$ is already infected. Our link strength data, when the unit of time is one day, have values which vary from a few percent down to about $10^{-4}$. With these small values we can increment the simulator with a time step of one day, and get smooth results.

Our simulations differ from one another in three ways: (i) the choice of “start” nodes which are infected at $t = 0$; (ii) the choice of a set of “immune” nodes which cannot be infected; and (iii) sometimes, the choice of links which are to be blocked from transmission (removed). Choices (ii) and (iii) allow us to test various strategies for hindering spreading. In the real world of human sexual behavior, accomplishing either of these effects may be quite difficult; but we test them here simply to see what can be achieved.

First we simulate the reference case, in which those nodes which are known to be infected are the start nodes (see again Figs. 6 and 7), and we immunize no nodes or links. We find (Figure 9) that the infection takes off very fast—as anticipated in Section 3.4. Specifically, we see that the takeoff time is very short—just a few days. This is of course consistent with the fact that the infection has already reached three very central (as defined by EVC) nodes. This latter fact is consistent with two interpretations: either (i) the infection has recently come to this dense network, and it is on the verge of taking off, or (ii) the infection has been present for a long time, and has reached an equilibrium (and rather low) level.

We do not have sufficient empirical information to favor one of these interpretations over the other. If the first one is correct, it implies that one can expect a strong growth of infection rate in a relatively short time. If on the other hand the second is correct, then our model is likely inadequate, not only in the SI aspect but probably in other aspects as well. We remind the reader that our topographic analysis is most useful in understanding the spreading of new infections over fairly static networks; hence it may be useful in case (i), but has little to say about case (ii).

Now, in order to test our ideas further, we assume case (i). Based on our SAGS picture, we formulate various immunization strategies, and test them via simulation. We have tried (a) immunizing the Center node; (b) immunizing the Center and all nodes within one hop of the Center (subregion heads); (c) immunizing the two infected nodes in the large red subregion, plus that subregion’s head node; and (d) immunizing 50 nodes chosen at random.

Results for all of these cases are shown in Figure 9. A simple conclusion is starkly obvious: none of these immunization strategies are able to retard the takeoff. In fact, the only clear difference is the trivial and useless one: that the long-time infected fraction is reduced by the number of immunized nodes [for example, by 14 for scenario (b), and by 50 for scenario (d)].

In short: as strongly suggested by Figure 7, the NG network is sufficiently well connected, and sufficiently well infected, that we find no simple strategy which is at all effective in retarding the takeoff.
In order to investigate a different kind of test of the utility of our method of analysis, we next “cure” all infected nodes, and explore scenarios in which we can choose the start nodes freely. Our principal aim is to test the following hypothesis: that time to takeoff is strongly determined by distance from the Center of the SAG*.

Some simple tests of this hypothesis are shown in Figure 10. Here we show the progression of infection for three scenarios: (e) the Center is the only infected start node; (f) a node roughly halfway between the Center and the periphery is the start node; and (g) a very peripheral node is the start node.

The results of Figure 10 strongly support our hypothesis. Takeoff times vary from a few days to about 50 days to almost 150 days, as we move the start node outward in the SAG*.

We also see, in the bottom half of the figure, that our earlier picture [4, 5] of the movement of the infection “front” over the topography is confirmed here: the infection [assuming it doesn’t start at the top as in (e)] moves slowly at first, until it begins to reach more central nodes, at which point it speeds up, while moving “uphill” (towards the Center); subsequently it moves “downhill”, slowing down all the while. While we have seen this dynamic pattern many times before, this is the first time we have tested it on a graph with weighted

---

**Fig. 9.** Spreading simulations for gonorrhea, based on the SI model, and using various prevention strategies. “As-is” = known infected start nodes and no strategy; the other scenarios involve immunizing various nodes, as described in the text. The unit of time is one day.
Fig. 10. Three spreading simulations, based on three chosen scenarios, each with a single start node. We see that distance from the Center node (in a metric defined by the SAG∗) correlates strongly with time to takeoff. The lower part of the figure shows the average EVC of the newly infected nodes.

links (and with the EVC appropriately corrected via the weighted adjacency matrix).

While Figure 10 offers anecdotal evidence for our hypothesis, we also have statistical data. We have in fact run one-start-node simulations for each node on the graph, 10 times for each node, and recorded the average time needed to reach an infection number of 300 nodes (about 60%). To measure ‘distance’ from the Center, we define the dual notion of ‘closeness’: a node’s closeness to the Center is simply the product of the link strengths over the (unique) path to the Center in SAG∗. Thus many weak links gives low closeness, while few strong links gives high closeness; and both the number of hops and the link strengths of the hops affect the result.

Figure 11 gives a scatter plot for average infection time vs closeness, for all nodes in the graph except the Center node. We see a strong decreasing relationship: closer nodes need less time to infect the graph. Thus we find from these results further strong support for our hypothesis.
Fig. 11. Time needed for a single start node to infect 300 nodes, as a function of that start node’s ‘closeness’ to the graph’s Center (averaged over 10 experiments for each start node). Closeness is measured entirely in terms of the modified steepest-ascent graph $SAG^*$. We see a thorough statistical corroboration of the results of Figure 10.

5 Summary and Discussion

In this paper we have extended the topographic approach to the problem of epidemic spreading over networks to a problem involving two new features. First, the network is real: it is an empirical sex network, with some nodes known to be infected with the STI’s HIV, gonorrhea, and chlamydia. Secondly, we have data which allow us to assign non-uniform link strengths (transmission probabilities), and we have generalized the topographic approach to incorporate these link strengths.

To help in illuminating the effects of incorporating link strengths, we first performed the analysis by ignoring these weights. We visualized the resulting unweighted FSW network, and simulated the progress of HIV on this network (using uniform transmission probabilities). We found some interesting effects from the almost-bipartite nature of the unweighted network. We also found that the network is very highly connected—with the two HIV-infected nodes very close to the network’s Center—so that retarding the spread of HIV was difficult. Nevertheless we were able to show significant benefits to be obtained from our analysis, for some hypothetical cases involving start nodes placed elsewhere.
Incorporation of empirically obtained link strengths had large consequences. Each disease yielded a distinct weighted graph, by affecting the transmission probabilities. We found (using our assumption that perfect condom protection was possible) that the HIV graph broke down into many small components. While our visualization may still have some value, we saw no value in running simulations on these small components.

Simulations on the gonorrhea graph gave results much like those on the unweighted FSW graph: the graph was very well connected, and the already-infected nodes had rather central positions. The result was that we were unable to find simple topological fixes, inspired by our analysis, which could significantly retard spreading. However, we were able to find strong evidence confirming the basic applicability of our analysis to spreading. Specifically, we showed that our own notion of a node's distance from the Center of the graph correlated strongly with the time needed for that node to infect the graph.

We emphasize that this is the first application of the topographic approach to a weighted graph. Performing this analysis has required generalizing our earlier definition [4] of steepest ascent. The results we obtain here, based on this new, generalized definition, are very promising. Hence—even as we fail to come up with promising, concrete suggestions for hindering the spread of STI's in the Vancouver sex network—we feel that our results confirm the applicability of our approach to understanding spreading in the real-world case of a network with nonuniformly weighted links.

We see a clear need for two obvious extensions of this work. First, it would be useful to reconnect the HIV graph, by assigning small but nonzero probabilities to the '100%-condom-use' links. This would allow for a more meaningful regions analysis, and the accompanying testing by simulations (perhaps over a long time scale).

Secondly, our approach is most simply understood and applied for diseases for which SI spreading is appropriate (such as HIV). The application to gonorrhea or chlamydia would be greatly strengthened if one could generalize the method to the SIS and/or SIR case. This is an interesting challenge which remains for future work.

The data used arrives from self-reported infection status ([7]). To validate our model, empirically collected retrospective data on actual prevalence and incidence of the infections could be obtained. This is also recommended for future work.

Finally, we remind the reader of the motivation for this work. We believe that the topographic analysis, based on eigenvector centrality, is extremely useful for understanding epidemic spreading on a coarse scale. The analysis itself is not computationally demanding; hence it can be performed in essentially real time. Thus, we hope that our approach can be useful for disease prevention, in those cases for which the network can be mapped in reasonably short time—that is, short compared to both the time scale for infectious spreading, and the time scale for significant topology changes. The results presented here do not offer any immediate solution to the problem of STI's in
the Vancouver FSW network; but they do add further support to our belief
that this approach may be useful for this problem, and for others.

Acknowledgment

GC and KEM acknowledge partial support from the Future and Emerging
Technologies unit of the European Commission through Project DELIS (IST-
2002-001907). VPR acknowledges the financial and in-kind support, re-
spectively, of the BC Medical Services Fdn and HIV/STI Prevention and Control,
BC Centre for Disease Control.

References

2. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free net-
4. Geoffrey Canright and Kenth Engo-Monsen. Spreading on networks: a topo-
5. Geoffrey Canright and Kenth Engo-Monsen. Epidemic spreading over networks:
6. Kretzschmar M, van Duynhoven YTPH, and Severijnen AJ. Modelling preven-
tion strategies for gonorrhea and chlamydia using stochastic network simula-
Clients of indoor commercial sex workers: Heterogeneity in patronage patterns
and implications for HIV and STI propagation through sexual networks. *Sexu-
8. Geoffrey S. Canright and Kenth Engo-Monsen. Some relevant aspects of network
analysis and graph theory. In Jan Bergstra and Mark Burgess, editors, *Handbook
9. Jolly AM, Moffatt ME, Fast MV, and Brunham RC. Sexually transmitted
10. Holmes KK, Levine R, and Weaver M. Effectiveness of condoms in preventing