

# Efficient Mitigation Strategies for Epidemics in Rural Regions

Caterina Scoglio<sup>1\*</sup>, Walter Schumm<sup>2</sup>, Phillip Schumm<sup>1</sup>, Todd Easton<sup>3</sup>, Sohini Roy Chowdhury<sup>1</sup>, Ali Sydney<sup>1</sup>, Mina Youssef<sup>1</sup>

**1** K-State EpiCenter, Department of Electrical and Computer Engineering, Kansas State University, Manhattan, Kansas, United States of America, **2** Department of Family Studies and Human Sciences, Kansas State University, Manhattan, Kansas, United States of America, **3** Department of Industrial and Manufacturing Systems Engineering, Kansas State University, Manhattan, Kansas, United States of America

## Abstract

Containing an epidemic at its origin is the most desirable mitigation. Epidemics have often originated in rural areas, with rural communities among the first affected. Disease dynamics in rural regions have received limited attention, and results of general studies cannot be directly applied since population densities and human mobility factors are very different in rural regions from those in cities. We create a network model of a rural community in Kansas, USA, by collecting data on the contact patterns and computing rates of contact among a sampled population. We model the impact of different mitigation strategies detecting closely connected groups of people and frequently visited locations. Within those groups and locations, we compare the effectiveness of random and targeted vaccinations using a Susceptible-Exposed-Infected-Recovered compartmental model on the contact network. Our simulations show that the targeted vaccinations of only 10% of the sampled population reduced the size of the epidemic by 34.5%. Additionally, if 10% of the population visiting one of the most popular locations is randomly vaccinated, the epidemic size is reduced by 19%. Our results suggest a new implementation of a highly effective strategy for targeted vaccinations through the use of popular locations in rural communities.

**Citation:** Scoglio C, Schumm W, Schumm P, Easton T, Roy Chowdhury S, et al. (2010) Efficient Mitigation Strategies for Epidemics in Rural Regions. PLoS ONE 5(7): e11569. doi:10.1371/journal.pone.0011569

**Editor:** Thomas Mailund, Aarhus University, Denmark

**Received:** February 25, 2010; **Accepted:** June 17, 2010; **Published:** July 13, 2010

**Copyright:** © 2010 Scoglio et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by National Science Foundation - Award Number 0841112 SGER: Exploratory research on complex network approach to epidemic spreading in rural regions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: caterina@ksu.edu

## Introduction

Influenza A (H1N1), commonly known as swine flu, continues to be the dominant influenza virus in circulation across the globe with many countries and overseas territories reporting laboratory confirmed cases, including at thousands of deaths [1]. Factually, the origin of pandemic virus strains, such as the current H1N1, often trace back to rural regions. For example, the H1N1 2009 virus is suspected to have been originated in La Gloria, a small town near Veracruz, Mexico. Also, the previous strain of H1N1, commonly known as the Spanish Flu of 1918 that wrought devastation around the world, originated within the rural State of Kansas near Fort Riley. Other instances of epidemics originating in rural regions include the swine flu that originated in September 1988 at a hog barn in Walworth County, Wisconsin, the H5N3 virus that was identified at La Garnache farm in France in late January 2009, and the Asian flu that was a category 2 avian influenza in Ghizhou, China in 1956.

For analysis and containment purposes, large cities are generally considered to be infection hubs owing to the large population densities and mobility indices. Consequently, most spatio-temporal research on infectious human diseases focuses on large cities, such as Portland [2], Chicago [3], and Dresden [4], which respectively represent excellent examples of an agent-based model, a multi-scale meta-population model, and a social-structure model,

defining different levels of detail and complexity. Another approach to characterize the heterogeneous epidemic terrain of a human population is based on the construction of the network representing contacts among people. Studies of this type include [5], [6], and [7].

Various immunization strategies have been formulated for urban populations. Some of these strategies assumed that the human population distribution can be estimated as scale free. One such strategy was a targeted immunization wherein nodes having the highest connectivity were deemed to be the most critical for spreading the infection, and hence those highly connected people were chosen for vaccination [8]. However, global immunization strategies require the knowledge of the entire network of individuals and are complex from both computation and implementation stand points. Conversely, localized mitigation strategies, such as acquaintance immunization [9], randomly choose a subset of the entire population, and randomly select a set of their acquaintances to vaccinate. These strategies require a lower fraction of the population to be vaccinated than a random global immunization to dampen the impact of an epidemic. Other localized immunization methods include targeting acquaintances of randomly selected people by their estimated contact characteristics [10].

Disease dynamics in rural regions [11] [12] have received limited study. Since the population densities and human mobility

indices of rural regions are very different from those of cities, it is imperative to develop specific mitigation schemes to impede the spread of epidemics right from its likely source, the rural location. Furthermore, recent studies show that rural residents have a lower likelihood to obtain certain preventive health services than urban residents [13]. These factors necessitate research on predictive and optimally preventive strategies in rural regions.

This paper takes a unique look at rural regions, and presents mitigation strategies tailored for rural Clay county in Kansas, USA. We propose mitigation strategies that are based on a contact network model developed using data collected through a survey campaign conducted in rural Clay and Kearny counties in Kansas, USA. By characterizing the contact structure of rural regions, we are able to investigate the influences of this structure and various mitigation strategies on the speed, shape, and size of the outbreak. Our analysis shows that, although global targeted strategies are the most efficient in mitigating epidemics with a limited amount of resources, they can also be unfeasible due to partial knowledge of the population and conflicts with individual rights. Random vaccine distribution in selected popular location within a rural community offers the opportunity to indirectly reach the individuals who play a significant role in the epidemic propagation. We demonstrate with simulations that this location-based strategy can be 55% as effective as the best global target strategy.

## Methods

Our simulation results on epidemic spreading in rural communities are based on data collected through distributed surveys. This data is used to construct a contact network and to analyze the epidemic. Several random and targeted mitigation strategies are investigated through an SEIR model with parameters estimated in an analysis of the recent H1N1 influenza [3].

### 1. Survey Data

In Spring 2009, we surveyed residents of two rural Kansas counties through a visit to a county seat and mailed surveys, under our direct personal supervision. We obtained ethics approval in January 2009 for research protocols, survey forms, and informed consent procedures used, by the Institutional Review Board (IRB)

of Kansas State University, Human Subjects Committee, University research Compliance Office, 203 Fairchild Hall, Manhattan, Kansas. All potential participants were provided informed consent in a cover letter attached to their surveys; signatures were not required from the participants as a way of protecting their privacy.

The mailed surveys were well accepted with response rates of 64.8% and 41%, respectively for Clay County and Kearny County. The survey consisted of 30 short questions, a question concerning visits to local businesses and locations, a question concerning visits to cities within the surrounding region, and a set of contact questions. The spread of an epidemic in rural areas may be influenced by both the vulnerability of the population and the extent of their contacts with each other. Vulnerability includes their susceptibility to infection due to both poor health and a lack of preventive measures, such as vaccination. Once an epidemic has begun, the willingness of the population to comply with precautionary health measures can influence the rate and extent to which the epidemic spreads. In the survey, all these factors were assessed.

Survey results yielded four measures of risk factors important to the spread of epidemics: health risk, contact risk, prevention risk, and compliance risk. To what extent did these risks overlap? Each possible combination of risks was evaluated and summarized in Table 1. It is interesting to note that people with the most contacts tended to have the least preparedness for an epidemic, and people who were willing to visit others even during an epidemic were among those most at-risk because of their health status. Additionally, those who tended to visit friends and family members more often during normal times were also likely to retain this behavior even under epidemic conditions. This property of the rural communities is interesting, since it provides a given level of stability within the contact network and increases the accuracy of our epidemic analysis.

Survey results were also used to construct the weighted contact network. To this purpose, we used the survey responses about frequently visited locations and the levels of contacts. The respondents were asked to identify within a set of locations, those which they visit on a typical day. The responses were captured as a binary vector  $L_i$  for each respondent  $i$  with each element

**Table 1.** Possible Combinations of Risk Measures.

Health/Contact	Health/Prevention	Health/Compliance
As contact risk rose from low to medium to high, the percentage of respondents with one or more at-risk health conditions rose linearly from 37.7% to 45.0% to 56.3% ( $p < 0.18$ by chi-square test; $r = 0.15$ , $p < 0.07$ ).	Of those respondents from families in which all members had been vaccinated, only 34.1% had one or more at-risk health conditions, compared to 49.6% of those from families in which at least some of the members had not been vaccinated ( $p < 0.09$ ).	The percentage of respondents with one or more at-risk health conditions tended to rise as a function of their unwillingness to comply with a directive to stay at home during an epidemic: no visits (46.1%), one or two (38.8%), and three or more (75.0%) ( $p < 0.08$ ).
Those who would be most vulnerable or susceptible to an epidemic were actually most likely to be engaging in multiple contacts with friends, family, and guests.	Those who were most susceptible to an epidemic were least likely to be prepared for it in terms of anticipatory vaccination.	Those who were willing to visit others even during an epidemic were among those most at-risk because of their health status.
Contact/Prevention	Contact/Compliance	Prevention/Compliance
As contact levels rose from low to medium to high, the percentage of households with full vaccinations fell from 32.1% to 19.7% and 20.8%, respectively ( $p < 0.25$ by chi-square test; $r = 0.11$ , $p < 0.18$ ).	As contact levels rose from low to medium to high, the percentage of respondents who would not comply with health directives to remain at home rose from 29.4% to 40.0% to 44.7%, respectively ( $p < 0.28$ by chi-square test; $r = 0.13$ , $p < 0.12$ ).	Those from families that were fully vaccinated declined from 29.1% to 22.0% to 0.0% as respondents shifted away from no visits, one or two visits, or three or more visits during an epidemic ( $p < 0.08$ by chi-square test; $r = 0.17$ , $p < 0.04$ ).
Those with the most contacts tended to have the least preparedness for an epidemic.	Those who tended to visit friends and family members more often during normal times were also likely to retain this behavior even under epidemic conditions.	Those making the most visits were the least likely to be protected by vaccination.

doi:10.1371/journal.pone.0011569.t001

corresponding to a location. The contact questions asked the respondents to estimate the number of individuals with whom the respondent made contact for three different levels of contact. Contact levels were classified into *Proximity* contact (coming within 5 feet of another person, even if in passing), *Direct-Low* contact (directly touching another person for a short period of time in what most people would consider a low risk situation of being infected), and *Direct-High* contact (directly touching another person for an extended period of time or in what most people would consider a relatively high risk situation of being infected). The responses of the contact questions are quantified as values  $n_{x,i}$  that represent the number of individuals contacted by respondent  $i$  in a typical day according to each contact level  $x$ .

## 2. Contact Network Construction

With these responses, the rural community is represented as a weighted contact network where each of the survey respondents is represented as a node within the network that is connected together with links representing the contact between respondents. Each link has a weight that represents the normalized measure of contact between the connected pair of respondents or nodes. Each link's weight  $w_{ij}$  is taken as the average of three sub-weights that correspond to the interactions between node  $i$  and node  $j$  estimated for each contact level  $x$ . Values of weights  $w_{ij}$  range within the interval  $[0, 1]$ . We capture the location responses within the parameter  $\mu_{i,j} = (1 + l_{i,j}) / (1 + d)$ , where  $d$  is the total number of locations and  $l_{i,j}$  is the dot product of the respective location vectors  $L_i$  and  $L_j$  for nodes  $i$  and  $j$ . For a given type of contact  $x$ , the related sub-weight function depends on the node degree and the parameter  $\mu_{i,j}$ . When either node degree  $n_{x,i}$  or  $n_{x,j}$  is zero, the sub-weight should be equal to zero. The sub-weight should also increase monotonically with both  $n_{x,i}$  and  $n_{x,j}$ , approaching unity when both are large. When a pair of nodes visit all the locations,  $\mu_{i,j}$  is equal to unity and the sub-weight should be maximum. On the other hand,  $\mu_{i,j}$  has a small positive minimum, to allow for

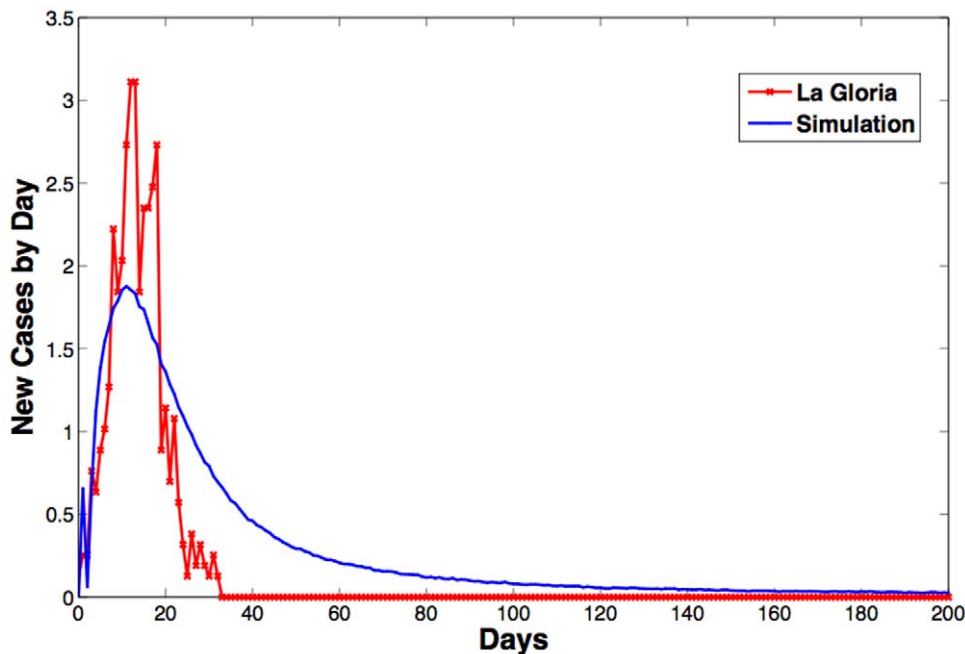
interactions outside the locations included in the survey. For given  $n_{x,i}$  and  $n_{x,j}$ , the sub-weight should be minimum when  $l_{i,j}$  is equal to zero, and should increase monotonically with increasing  $l_{i,j}$ , and consequently  $\mu_{i,j}$ .

For each contact level  $x$ , we compute the sub-weight  $w_{x,i,j}$  between node  $i$  and node  $j$  according to a simple function which follows the desired behavior:

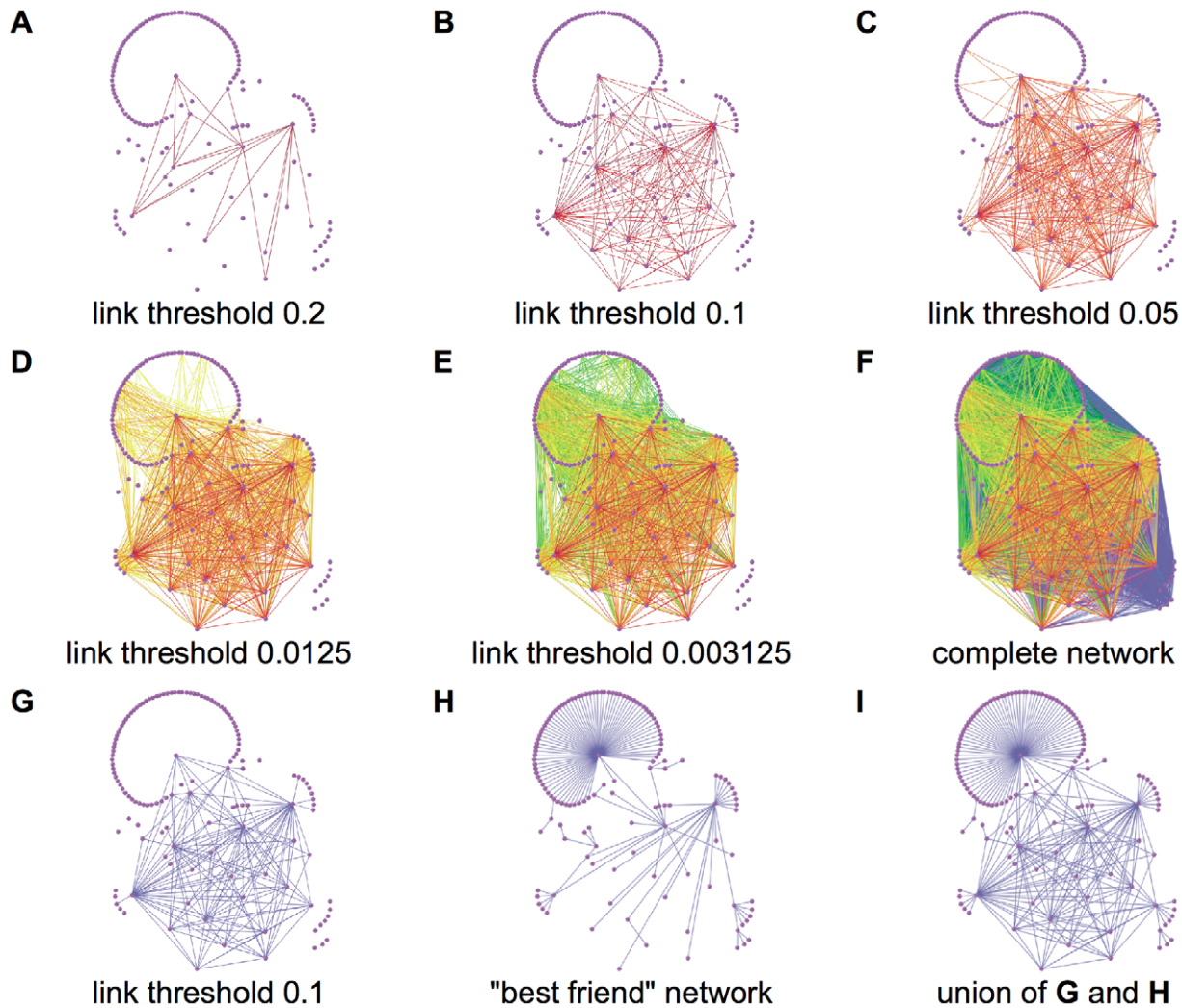
$$w_{x,i,j} = (1 - (1 - \mu_{i,j}\pi_x)^{n_{x,i}})((1 - (1 - \mu_{i,j}\pi_x)^{n_{x,j}})) \quad (1)$$

We selected the values of  $\pi_x$  such that they include the relative importance of each of the three contact categories from the survey responses, constraining  $\pi_{Proximity}$  to be less than  $\pi_{Direct-Low}$  and  $\pi_{Direct-Low}$  to be less than  $\pi_{Direct-High}$ . The values of  $\pi_x$  have been estimated by matching the epidemic curve of the H1N1 outbreak in La Gloria, Mexico [14], with the average epidemic curve obtained with the weighted network simulations for Clay Center, Kansas. Based on the minimum squared error, we found that the best contact levels for Proximity, Direct-High and Direct-Low contacts are  $\pi_{Proximity} = 0.0025$ ,  $\pi_{Direct-Low} = 0.015$ , and  $\pi_{Direct-High} = 1.0$ . Figure 1 reports the number of new infected individuals in La Gloria [14] with the corresponding simulated new infected individuals in Clay Center, Kansas given the best estimated values for the three contact levels. This parameter estimation was done through an SEIR epidemic model, which we describe in the following sections.

With the estimated model parameters, we have created the weighted contact network shown in Figures 2.(a)–2.(i). In Figure 2.(a), only links with strength greater or equal to 0.2 are depicted. In the successive figures more links are added by reducing the threshold, reaching the complete network in Figure 2.(f), where all links are depicted. The values for the thresholds considered in the other figures are respectively 0.10 for 2.(b), 0.05 for 2.(c), 0.0125 for 2.(d), and 0.003125 for 2.(e). Figure 2.(i) is the union of two networks: the network obtained for



**Figure 1. New infected individuals.** Number of new infected individuals in La Gloria, Mexico by day [14] with the corresponding new infected individuals in Clay Center, Kansas, given the best estimated values for the three contact levels, averaged over 10,000 simulations. doi:10.1371/journal.pone.0011569.g001



**Figure 2. Rural contact network.** The rural contact network is composed by nodes representing individuals and weighted edges representing contacts, displaying all edges with weights greater than the following thresholds: A) 0.20, B) 0.10, C) 0.05, D) 0.0125, E) 0.003125, and F) 0. In the last row, G) shows the edges with weights greater than 0.10, H) shows the highest weighted edge for each node, and I) shows the union of the previous two networks.

doi:10.1371/journal.pone.0011569.g002

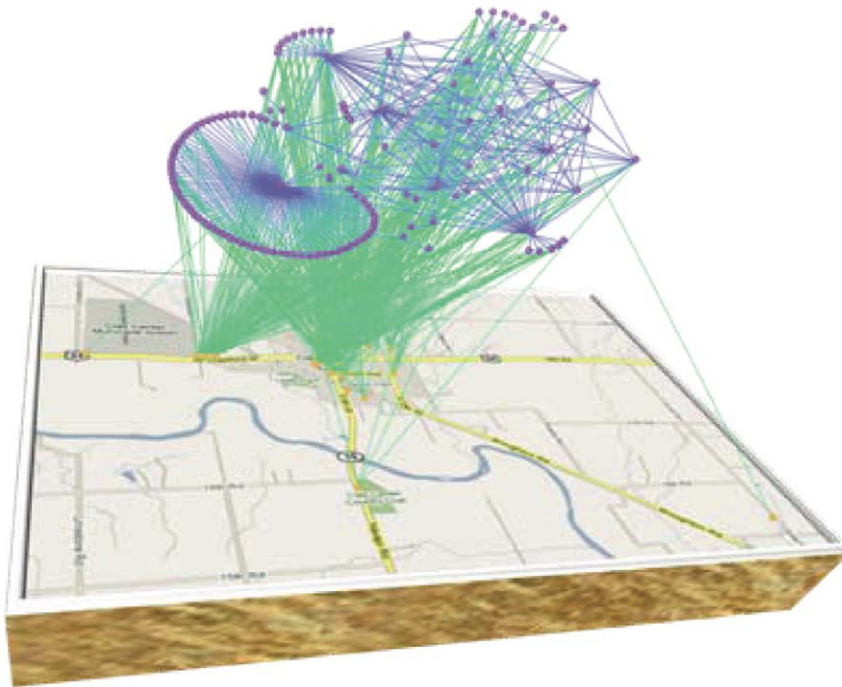
threshold 0.1 and the network of the *best friends* where for each node only the link with highest strength is depicted. Figure 2.(i) has been used to create Figure 3, where not only nodes representing people, but also nodes representing popular locations in Clay Center are shown.

We performed a sensitivity analysis on each  $\pi_{x_s}$  varying their values up to 15%. These variations, shown in Table 2, produce a maximum of 3.4% variation in total infection cases, with most changes resulting in a variation of less than 1% in total cases.

To compare our selection of the structure of the weight  $w_{x,i_j}$  shown in Eq. 1, we have constructed another weighted contact network based only on data about common visited locations. In this case, the weights  $w_{i,j}^l$  are computed as  $w_{i,j}^l = l_{i,j} / d$ . The network constructed in this way has 38% of nodes isolated, i.e., with node degree equal to zero, and when we simulated the SEIR model only 63% of the infected nodes coincided with the infected nodes obtained using the same model on our contact network. Consequently, the use of only location data produces not negligible differences in the results and is not considered in the following analysis.

### 3. Network Metrics

To describe in details the characteristics of the weighted contact network, we select some graph-theoretical metrics that reflect the local and global properties of the graph [15]. In Table 3, some relevant metrics for the contact networks are listed. The contact network is composed of 138 nodes ( $N$ ) and 9222 links. It is important to note that the network is not far from a fully connected network, which would have 9453 links. However, each link can have a very different importance due to the structure of the link weights. For this reason, we select the node strength as one metric to characterize a node. The strength  $s_i$  of node  $i$  is defined as the sum of the weights  $w_{i,j}$  of all links between node  $i$  and its neighbors  $N_i$ ,  $s_i = \sum_{j \in N_i} w_{i,j}$ . The node strength is analogous to the node degree in the binary network, which measures the number of contacts or neighbors of a node. The second metric we compute is the average shortest path. To compute shortest path properly, we define the distance  $d_{i,j}$  between any neighbor nodes  $i$  and  $j$  as  $d_{i,j} = 1 - w_{i,j}$ . The distance defined in this way is always non-negative and reveals a short distance separating node  $i$  and node  $j$  when their link weight  $w_{i,j}$  is high. The third metric, the network



**Figure 3. Popular location network.** The network of people and popular locations in Clay Center, Kansas, where the nodes (survey respondents) in the cloud network are connected via green edges to the locations in Clay Center according to the survey responses. The map is courtesy of Google. doi:10.1371/journal.pone.0011569.g003

diameter is defined as the longest of the shortest paths. As a centrality measure, we compute the betweenness  $b_i$  of a node  $i$ . It is defined as the measure of the number of shortest paths between any pair of nodes passing through node  $i$ .

$$b_i = \sum_{h,j} \frac{\sigma_{h,j,i}}{\sigma_{h,j}} \quad (2)$$

where  $\sigma_{h,j}$  is the total number of shortest paths from node  $h$  to  $j$  and  $\sigma_{h,j,i}$  is the number of those shortest paths that pass through the node  $i$ . A node that appears in many shortest paths has high betweenness. Each  $b_i$  is normalized by the maximum number of shortest paths that can pass through a node  $(N-1)(N-2)/2$ . Another measure of node centrality is the clustering coefficient  $c_i$  of a node  $i$ , which measures the level of connection among the neighbors of node  $i$ .

$$c_i = \frac{1}{k_i(k_i-1)} \sum_{j,k} (\hat{w}_{i,j} \hat{w}_{i,k} \hat{w}_{j,k})^{\frac{1}{3}} \quad (3)$$

where  $\hat{w}_{i,j} = w_{i,j} / \max(w_{i,j})$  and  $k_i$  is the degree of node  $i$  in the binary version of the weighted contact network [16]. By averaging

over all individual clustering coefficients, we obtain the average clustering coefficient of the contact network. The node coreness is the maximum value  $k$  such that the node still exists in the network, before being removed in the  $k+1$  core. The  $k$ -core of a graph is a maximal subgraph in which each vertex has at least strength  $k$ . The coreness measures the deepness of a node in the core of the network where a higher value indicates that the node is deeper in the core. The discrete values for the strength classes are obtained by a fine quantization of the node strength (step size on the order of  $10^{-6}$ ). From the spectral domain, the maximum eigenvalue is the largest eigenvalue of the weighted adjacency matrix  $W$  representing the network. The elements of  $W$  are the weights  $w_{i,j}$ , and the matrix in this case is symmetric and has zeros in the main diagonal. A large maximum eigenvalue corresponds to a small epidemic threshold in the Susceptible-Infected-Susceptible model [17].

Networks often display some level of grouping of nodes in an organized fashion that allow them to be divided into different clusters or communities. One popular method of detecting communities is to maximize a parameter known as modularity [18]. Modularity is a measure of the difference between the edges within each community and the expected number of edges in the same community, summed over all communities within the graph.

**Table 2. Sensitivity analysis showing percentage differences from original number of total cases.**

Percentage Difference	-15%	-10%	-5%	+5%	+10%	+15%
$\pi_{Proximity}$	-0.29334	-0.22053	-0.10083	-0.00539	0.15957	0.22018
$\pi_{Direct-Low}$	-0.70028	-0.49687	-0.26954	0.20365	0.50449	0.77051
$\pi_{Direct-High}$	-3.44748	-2.02815	-0.95024	X	X	X

doi:10.1371/journal.pone.0011569.t002

**Table 3.** Network Metrics for Contact Network.

Network Metric	Value	Network Metric	Value
Links	9222	Aver. Link Weight	0.0064540
Diameter in # of Hops	2	Aver. Node Coreness	0.49579034
Aver. Clustering Coefficient	0.0037	Aver. Node Betweenness	0.000179682
Aver. Node Strength	0.8626	Aver. Shortest Path by Distance	1.01699
Diameter by Distance	1.99996	Max Eigenvalue of Weighted Matrix	480.5959

doi:10.1371/journal.pone.0011569.t003

Using the weighted version of the algorithm described in [19] [18], we found two communities within our contact network. With a modularity value of 0.1087, 61% of the population fell into community 1 and the remaining 39% of the population composed community 2.

#### 4. SEIR Model on the Contact Network

We expand the weighted compartmental model [17] to represent the different disease states of individuals: Susceptible ( $S$ ), Exposed ( $E$ ), Infectious ( $I$ ), and Recovered ( $R$ ). The states and the transitions between states are unique to each disease and its characteristics, requiring customization to each disease. In this model, we selected  $\beta = 0.4$ , where  $\beta$  is the rate of infection across a link between a susceptible individual and infected individual,  $\varepsilon \approx 0.909$ , where  $\varepsilon$  is the transition rate parameter between the Exposed and Infected compartments, and  $\delta = 0.4$ , where  $\delta$  is the transition rate parameter between the Infected and Recovered compartments [20].

The network topology plays an important role in the spreading process in the transition from  $S$  to  $E$  when an  $I$  individual contacts an  $S$  individual and successfully infects him/her. The probability that node  $i$  is not infected at time  $t$  depends on the probabilities that a neighbor node  $j$  is previously infected  $p_{j,t-1}$ , is in contact with node  $i$  ( $w_{i,j}$ ), and successfully infects node  $i$  ( $\beta$ ) [17].

$$\zeta_{i,t} = \prod_{j \in N_i} (1 - w_{i,j} \beta p_{j,t-1}) \quad (4)$$

The probability that a node is infected (transition from  $S$  to  $E$ ) at time  $t$  is then  $1 - \zeta_{i,t}$ . The remaining transitions are topology independent and only depend on the rate parameters,  $\varepsilon$  and  $\delta$ , of the disease model. When an individual has contracted the disease and has transitioned into the exposed compartment, the next transition to the state  $I$  occurs with rate  $\varepsilon$ . Once infected, a node attempts to infect its susceptible neighbors until it transitions to the recovered state. Each infected node recovers with recovery rate  $\delta$ . Once a node is recovered ( $R$ ), it remains recovered for the remainder of the simulation. The recovered compartment serves as an accumulator of all the cases, thus the number of recovered individuals  $R$  at the end of a simulation is a good approximation of the total number of cases caused by the outbreak. The blue curve in Figure 1 has been computed using the above model.

#### 5. Epidemic Simulations

The analysis of the epidemic evolution and the evaluation of multiple mitigation strategies are performed using an SEIR model on the contact network. We propose different immunization strategies that can be implemented as vaccinations or antiviral treatments. The immunization strategies are classified in three categories based on individuals, locations, and communities. In

each individual immunization strategy, nodes are chosen either deliberately, based on a node metric or randomly.

The random selection of nodes as recipients of an immunization represents an unbiased distribution of resources and is the simplest method for distribution. The node metrics selected for the targeting strategies include node strength, node coreness, and node betweenness. Node strength, as a measure of how well an individual is connected with the rural population, is an intuitive measure of how likely a node is to be infected by other nodes as well as how likely the node is to pass the infection on to others. Therefore to mitigate the infection while using node strength to select nodes, we target the nodes with the highest strength. The node coreness is a measure of how *deep* a node is in the core of a network. This depth is a measure of the maximum strength of the nodes iteratively removed from the network periphery before the node is removed. From a topological perspective, the core of the network facilitates connectivity and is vital for it. Therefore a targeted removal or immunization of the core nodes serves to hinder and disrupt the connectivity that allows the spread of the infection. The betweenness of a node measures how many shortest paths between all pairs of nodes choose to route through the node. Thus targeting nodes with highest betweenness serves to disrupt the shortest paths that the virus can take, forcing it to longer routes. We applied these different targeting strategies globally on the entire network and then within the communities and selected locations. The immunization of a node is implemented by forcing the immunized nodes to remain susceptible throughout the epidemic.

In Table 4 the reduction in the number of cases by percentage with respect to the unmitigated epidemic is shown, for different criteria for the selection of the 10% of immunized people among the global population. The most effective strategy is the one where the 10% of nodes with highest strength are selected, in line with previous results, followed by the one based on the selection of 10% of the nodes with the deepest coreness. However, these types of strategies have an inherent problem: how can we practically detect those special nodes? Fortunately, the data collected on the location popularity, can help to solve this problem. The survey respondents

**Table 4.** Global Mitigation Strategies.

Mitigation Strategy	% Reduction of Total Cases	Cases Prevented per Vaccine
Random	11.40	0.69
Highest Strength Nodes	34.57	2.11
Highest Coreness Nodes	25.18	1.53
Highest Betweenness Nodes	16.27	0.99

doi:10.1371/journal.pone.0011569.t004

**Table 5.** Location-based Mitigation Strategies.

Mitigation Strategy 10% Immunization	% Reduction of Tot. Cases	Cases Prevented per Vaccine
Random in Most Popular Location	14.25	0.87
Random in Highest Strength Location	18.97	1.16
Highest Strength in Most Popular Location	34.57	2.11
Highest Coreness in Most Popular Location	25.17	1.53
Highest Coreness in Highest Coreness Location	24.31	1.48

doi:10.1371/journal.pone.0011569.t005

associated themselves with various locations in the county by indicating which ones they typically visit. We used two criteria to select the locations for targeting and random strategies. To select locations for the random strategies, we chose the locations having the highest average value of the desired metric and being associated with at least 10% of the population. For the targeting strategies, we chose the locations visited by more than 10% of the population, and we immunized 10% of the population by selecting nodes with the highest combined sum for the desired metric within those locations. In Table 5, the reduction in the number of cases by percentage is shown, when the immunization of the selected people is performed among the group visiting a particular location.

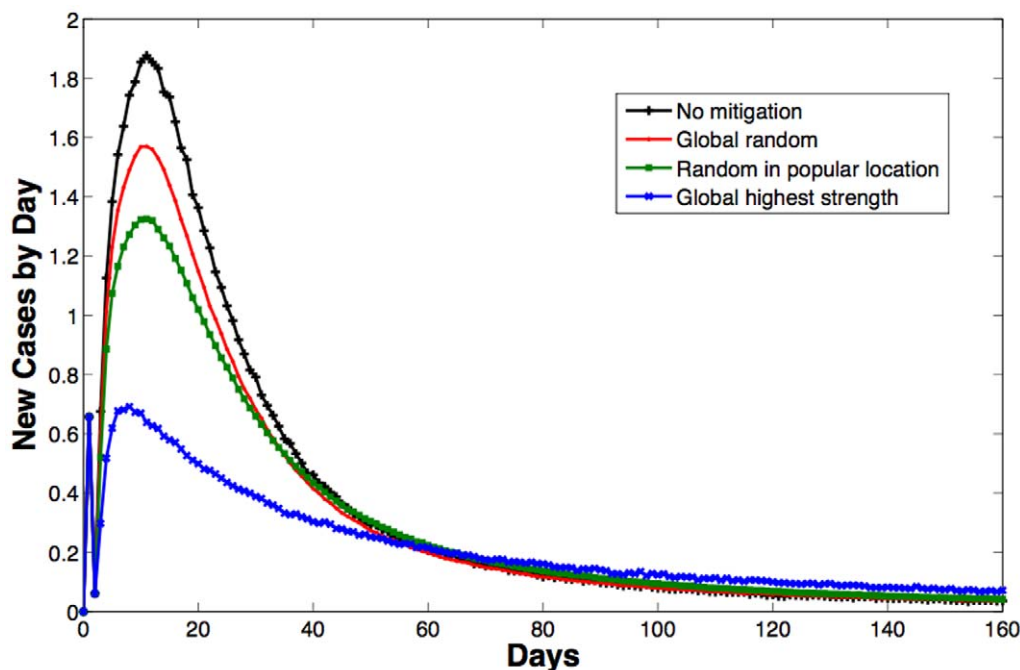
## Results and Discussion

Obtained results span the two investigated areas, namely risk assessment and mitigation strategy evaluations. Concerning risk assessment, very few rural respondents (2%) did not have a high

level of risk in at least one of four areas assessed: health risk, contact risk, prevention risk, and compliance risk. Over 75% of households did not have complete uptake of flu vaccine, nearly half of respondents had at least one major health risk, and nearly two-fifths of respondents said they would not comply with directives to stay at home during an epidemic. Risk levels were positively associated, suggesting that risks were compounded with each other, a situation posing greater problems for any attempt to predict or reduce the spread of epidemics in rural areas. Married respondents were much less likely to report selected health risks by a substantial margin (38% vs. 69%). Other demographics factors had relatively small associations with health, compliance, prevention, or contact risks, although some nonlinear associations between income and the risk factors were noted, with middle-income respondents having the lowest risk levels compared to lower or higher-income respondents.

Concerning mitigation strategies evaluation, Table 4 shows that the random immunization of 10% of the population (first strategy) reduces the epidemic size by 11.40%, with no substantial gain. However, if 10% of the nodes with highest node strength are immunized (second strategy), the epidemic size is reduced by 34.57%, more than three times the size of the random immunization campaign. In the interesting case where the 10% of the immunized nodes are randomly selected within the group of people frequently visiting a specific popular location (third strategy), an intermediate benefit, of about 19% epidemic size reduction, is obtained. The identification of specific locations visited by highest strength nodes has the clear benefit of improving the efficiency of a random immunization campaign, when this campaign is conducted in specific locations. Figure 4 shows the curves of new infected nodes with time under free evolution and for the discussed three mitigation strategies.

Our simulations suggest that information and immunization activities for rural communities should be carried out in specific



**Figure 4. Simulation of infection evolutions in different mitigation scenarios.** Newly infected nodes by day as a percentage of the population without and with mitigations strategies, including random vaccination throughout the population, random vaccination among nodes associated with a selected location, and targeted vaccination of a set of nodes having the highest node strength within the rural contact network. doi:10.1371/journal.pone.0011569.g004

locations, called key locations, which not only most people but also the most key people (highest strength nodes) often visit. Detecting key locations requires some amount of data collection and analysis. However, detecting key locations is much easier than identifying highest strength nodes (individuals with high levels of contact). In other words, the probability of immunizing a highest strength node given a node random selection in a key location is much higher than the probability of immunizing a highest strength node given a node random selection in the entire population.

In the presence of limited anti-viral and vaccination resources, government health agencies should seek to use the most effective methods of distribution for mitigation of the threat. Here, we have investigated the distribution of immunizations to 10% of the population through various targeting strategies. This work is of particular interest to rural regions, as they are more likely to face resource shortages due to smaller budgets than urban areas.

## References

- Centers for Disease Control and Prevention (2009) H1N1 Flu: International Situation Update.
- Eubank S, Guclu H, Anil Kumar V, Marathe M, Srinivasan A, et al. (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429: 180–184.
- Vespignani A (2009) Global Epidemic and Mobility Modeler. <http://www.gleamviz.org/>. Accessed 2010 Jan 11.
- Perminov V, Kornilina M (2007) Individual-based model for simulation of urban epidemic. *Matematicheskoe Modelirovanie* 19: 116–127.
- Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Physical Review Letters* 86: 3200–3203.
- Boguna M, Pastor-Satorras R, Vespignani A (2003) Epidemic Spreading in Complex Networks with Degree Correlations. *LNP Springer-Verlag* 625: 127–147.
- Newman MEJ (2002) The spread of epidemic disease on networks. *Physical Review E* 66: 016128.
- Pastor-Satorras R, Vespignani A (2002) Immunization of complex networks. *Physical Review E* 65: 036104.
- Cohen R, Havlin S, Ben-Avraham D (2003) Efficient Immunization Strategies for Computer Networks and Populations. *Physical Review Letters* 91: 247901.
- Gallos L, Liljeros F, Argyrakis P, Bunde A, Havlin S (2007) Improving immunization strategies. *Physical Review E* 75: 045104.
- Retailliau HF, Storch GA, Curtis AC, Horne TJ, Scally MJ, et al. (1979) The Epidemiology of Influenza B in a rural setting in 1977. *Am J Epidemiol* 109: 639–49.
- Zimmerman RK, Santibanez TA, Janosky JE, Fine MJ, Raymund M, et al. (2003) What affects influenza vaccination rates among older patients? An analysis from inner-city, suburban, rural, and veterans affairs practices. *The American Journal of Medicine* 114: 31–38.
- Casey MM, Thiede Call K, Klingner JM (2001) Are rural residents less likely to obtain recommended preventive healthcare services? *American Journal of Preventive Medicine* 21: 182–188.
- Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, et al. (2009) Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science* 324: 1557–1561.
- Barrat A, Barthelemy M, Vespignani A (2008) *Dynamical processes on complex networks* Cambridge University Press. 347 p.
- Saramaki J, Kivela M, Onnela JP, Kaski K, Kertesz J (2007) Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E* 75: 027105.
- Schumm P, Scoglio C, Gruenbacher D, Easton T (2007) Epidemic spreading on weighted contact networks. *Information and Computing Systems 2007, Bionetics 2007*. 2<sup>nd</sup>. pp 201–208.
- Newman MEJ (2006) Modularity and community structure in networks, *PNAS USA*, vol. 103.
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Physical Review E* 69: 026113.
- Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc Roy Soc Lond* 115: 700–721.

Additionally, rural regions are more likely to have a small set of local businesses and locations than urban areas due to lower population densities. Therefore classification and analysis of popular locations to be targeted for vaccine distribution is a feasible task. This work has shown the benefit of being able to select proper distribution locations; a strategy that can be implemented without having full knowledge of every individual within the rural population.

## Author Contributions

Conceived and designed the experiments: CS WS PS TE. Performed the experiments: CS WS PS. Analyzed the data: CS WS PS AS. Contributed reagents/materials/analysis tools: CS WS PS TE SRC AS MY. Wrote the paper: CS WS PS SRC MY.