

Finding Spatial Clusters Susceptible to Epidemic Outbreaks due to Undervaccination

Extended Abstract

Jose Cadena
Lawrence Livermore National
Laboratory
Livermore, CA, USA
cadenapico1@llnl.gov

Achla Marathe
Biocomplexity Institute, Dept. of
Public Health Sc., Univ. of Virginia
Charlottesville, VA, USA
achla@virginia.edu

Anil Vullikanti
Biocomplexity Institute, Dept. of
Computer Science, Univ. of Virginia
Charlottesville, VA, USA
vsakumar@virginia.edu

ABSTRACT

Geographical clusters of undervaccinated populations have emerged in various parts of the United States in recent years. Public health response involves surveillance and field work, which is very resource intensive. Given that public health resources are often limited, identifying and rank-ordering critical clusters can help prioritize and allocate scarce resources for surveillance and quick intervention. We quantify the criticality of a cluster as the additional number of infections caused if the cluster is underimmunized. We focus on finding clusters that maximize this measure and develop efficient approximation algorithms for finding critical clusters by exploiting structural properties of the problem. Our methods involve solving a more general problem of maximizing a submodular function on a graph with connectivity constraints. We apply our methods to the state of Minnesota, where we find clusters with significantly higher criticality than those obtained by heuristics used in public health.

ACM Reference Format:

Jose Cadena, Achla Marathe, and Anil Vullikanti. 2020. Finding Spatial Clusters Susceptible to Epidemic Outbreaks due to Undervaccination. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020)*, Auckland, New Zealand, May 9–13, 2020, IFAAMAS, 3 pages.

1 INTRODUCTION

Many highly contagious childhood diseases, such as measles, can be prevented by vaccination. Thus, it is worrisome that large outbreaks of such diseases have occurred in recent years. One of the reasons for the emergence of underimmunized geographical clusters, such as in California [7] and Minnesota [4], is misperceptions about the side effects of vaccines [2]. The typical response by public health agencies is to monitor clusters where immunization rates are falling, run active information campaigns, and engage community leaders. However, implementing public health interventions in all these clusters would be costly and time-consuming, which motivates the following question: *when do undervaccinated clusters pose significant risk (i.e., become critical) for the broader community?* We develop a method to address this important public health policy question. Our contributions are summarized below.

1. Formalizing criticality. We formalize the notion of *criticality* of a subset $S \subseteq V$ in a social contact network $G = (V, E)$, as

the *expected number of additional infections* that would occur if the immunization rate within S is “low”. Extending this notion, we introduce the MaxCrit problem: find the spatial cluster with maximum criticality in a population.

2. Rigorous algorithms for MaxCrit. We show that MaxCrit is NP-hard and design algorithm APPROXMAXCRIT, which has a worst case approximation guarantee of $\Omega(1/k^{(d-1)/(2d-1)})$, relative to the optimum, for clusters of size k . Here, d is the *doubling dimension* of the graph.

3. Improved algorithms for submodular function maximization with connectivity. We show that the criticality function is submodular, which implies that MaxCrit is a special case of maximizing a non-negative monotone submodular function over connected subgraphs of size k . *Our algorithm is the first improvement over the best known bound of [6] for submodular function maximization with connectivity constraints.*

4. Social impact. Our method for finding critical sets, applied to detailed population and contact network models, provides an operational tool for public health agencies to prioritize limited surveillance and outreach resources towards the most critical clusters.

2 PRELIMINARIES

Let V denote a population, and let $G = (V, E)$ be a contact graph on which a disease can spread. A person or node $v \in V$ can propagate the disease to its neighbors. Each person v is associated with a geographical location—i.e., their place of residence—denoted by $\text{loc}(v)$. Let \mathcal{R} denote the geographical area where the nodes V are located—e.g., Minnesota—and let $\mathcal{R} = \{r_1, \dots, r_N\}$ be a decomposition of \mathcal{R} into census block groups. For a block group $r_i \in \mathcal{R}$, we use $V(r_i)$ to denote the set of nodes associated with location r_i ; that is, those with $\text{loc}(v) \in r_i$. Analogously, for a set of block groups or *region* $R \subset \mathcal{R}$, let $V(R) = \cup_{r_i \in R} V(r_i)$ be the set of nodes located within R . We consider a graph $H_{\mathcal{R}} = (\mathcal{R}, E_{\mathcal{R}})$ on the set of block groups, where two block groups are connected if they are geographically contiguous, i.e., adjacent on a map. We use $\text{Conn}(\mathcal{R})$ to denote all the subsets $R \subset H_{\mathcal{R}}$ that are spatially connected.

Disease model. We use an SEIR model for diseases like measles [1]. Let γ denote the average region-wide vaccination rate—around 0.97 in Minnesota. Let \mathbf{x} be a *vaccination vector*: $x_i \in [0, 1]$ denotes the probability that node i is vaccinated (so $x_i = \gamma$, by default). Let Src_A denote the source of the infection: this could be one node or a small number of nodes from a region $A \subset \mathcal{R}$, which initially get infected. We use $\#\text{inf}(\mathbf{x}, \text{Src}_A)$ to denote the expected number of infections given an intervention \mathbf{x} and initial conditions Src_A .

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand. © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

2.1 Criticality

For a vaccination vector \mathbf{x} , let \mathbf{x}^S denote the corresponding intervention where a subset $S \subset V$ of nodes is undervaccinated. That is, $\mathbf{x}_i^S = \mathbf{x}_i$ for $i \notin S$ and $\mathbf{x}_i^S = \gamma'$ for $i \in S$, where $\gamma' \ll \gamma$.

We define the **criticality** of a region R as the *expected number of additional infections that occur if nodes in R are undervaccinated*, with respect to source Src_A :

$$\text{crit}(R, \mathbf{x}, \text{Src}_A) = \#\text{inf}(\mathbf{x}^{V(R)}, \text{Src}_A) - \#\text{inf}(\mathbf{x}, \text{Src}_A),$$

Finding critical clusters. In practice, public health interventions involve intensive field work, which is most effective within small, localized geographical regions. Therefore, we focus on finding regions that have high criticality *and* small size.

PROBLEM 1 (MAXCRIT($G, H_{\mathcal{R}}, k$)). *Given an instance ($G, H_{\mathcal{R}}, k$), find a connected region $R \in \text{Conn}(\mathcal{R})$ of size at most k that maximizes criticality over all choices of source:*

$$R = \text{argmax}_{R' \in \text{Conn}(\mathcal{R}), |R'| \leq k} \text{crit}(R', \mathbf{x}, \text{Src}_{R'})$$

In words, the MaxCrit problem involves maximizing over *all* possible choices of the sources $\text{Src}_{R'}$ in the cluster R' .

3 PROPOSED METHODS

Our strategy involves showing that the crit function is submodular. Intuitively, crit satisfies a *diminishing returns* property because an unvaccinated block group r causes more additional infections in the context of a smaller region. Thus, the MaxCrit problem reduces to maximizing a submodular function over connected subgraphs R of $H_{\mathcal{R}}$ with $|R| \leq k$.

We present our algorithm for MaxCrit in Algorithm 1. For submodular maximization with connectivity constraints, Kuo et al. [6] give an $\Omega(1/\sqrt{k})$ approximation guarantee. We derive an improved algorithm by exploiting the geometric structure of the graph. As in [6], the main idea is to first solve the relaxed problem where we ignore the connectivity constraint for each location r and surrounding locations at distance ℓ —i.e., the *ball* $B(r, \ell)$ in line 11. Then, we make the solution connected via a Steiner tree (line 12). Our algorithm has an approximation guarantee of $\Omega(1/k^{(d-1)/(2d-1)})$ for a graph with doubling dimension d , which improves the bound of [6], since d is finite for any graph. We also solve a Budgeted Steiner Tree (BST) problem [?] as a subroutine in our algorithm (line 5). BST is a relaxed version of the problem where the function is modular (instead of submodular) on the criticality of the nodes. This step is not needed for the approximation guarantee, but we find that it improves optimization power in practice.

4 APPLICATION TO MINNESOTA

Experimental setup. We use a realistic contact network model of Minnesota [5] with 5, 048, 920 individuals in total, aggregated into 4,082 census block groups from the 2010 U.S. census. We consider an SEIR stochastic model for measles, as described in Section 2. The criticality of a region R of block groups is assessed by leaving every individual inside R unvaccinated; everybody else in the population is vaccinated with probability 0.97, which is the statewide vaccination rate. We compare our algorithms with two heuristics used in public health. The **POPULATION** heuristic finds a cluster of size k with the largest total population. The **VULNERABILITY** heuristic

Algorithm 1 APPROXMAXCRIT($H_{\mathcal{R}}, k$)

```

1: Run MAXSUBMODCONN( $H_{\mathcal{R}}, k, d$ ), and let  $\hat{T}$  be the subgraph returned
2: for  $r \in H_{\mathcal{R}}$  do
3:   Let  $\text{wt}_r = \text{crit}(r)$ 
4: end for
5: Let  $T' = k\text{-MAXST}(H_{\mathcal{R}}, k)$  (using the approach of [3])
6: return  $\text{argmax}\{\text{crit}(T_r), \text{crit}(T')\}$ 
7:
8: procedure MAXSUBMODCONN( $H_{\mathcal{R}}, k, d$ )
9:   Let  $\ell = k^{\frac{d}{2d-1}}$ 
10:  for  $r \in H_{\mathcal{R}}$  do
11:    Run GREEDY( $B(r, \ell), \text{crit}, \ell$ ) to get subset  $S \subseteq B(r, \ell)$  of size  $\ell$ 
12:    Construct a minimum steiner tree  $T(r, \ell)$  of  $S$  in graph  $H_{\mathcal{R}}$ 
13:  end for
14:  return  $\text{argmax}_r\{\text{crit}(T(r, \ell))\}$ 
15:
16: procedure GREEDY( $V, f, k$ ) [8]
17:   Let  $S = \emptyset$ 
18:  for  $i = 1$  to  $k$  do
19:     $x = \text{argmax}_{x' \in V \setminus S} f(S \cup \{x'\}) - f(S)$ 
20:     $S = S \cup \{x\}$ 
21:  end for
22:  return  $S$ 

```

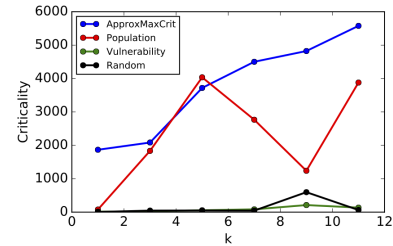


Figure 1: Comparison of algorithms for the MaxCrit problem as a function of the solution size k

prioritizes individuals who are most likely to get infected when no one is vaccinated. We also compare to a **RANDOM** baseline, which finds a connected cluster of size k by doing a random walk on the graph $H_{\mathcal{R}}$.

Optimization power. In Figure 1, we show the criticality obtained by **APPROXMAXCRIT** and the baselines as a function of cluster size k . **APPROXMAXCRIT** exhibits notably better performance than the heuristics. **RANDOM** performs poorly and results in almost no additional infections. Surprisingly, **VULNERABILITY** does not perform much better than **RANDOM**. Overall, the **POPULATION** heuristic has better performance among the baselines but non-monotonic growth with cluster size.

Demographics of critical clusters. We compare the distribution of age and income in the cluster discovered by **APPROXMAXCRIT** to that of the entire state (results not shown). The most critical cluster has significantly more households of low income (below \$25,000) compared to the entire state—19.6% to 34.9%. Similarly, minors are over-represented. 26.6% of the population is between 5–18 years old compared to the average of 18.7%.

When we focus on the Minneapolis area instead of the entire state, the most critical cluster covers Brooklyn Park, where measles outbreaks occurred in 2017 and 2019.

Acknowledgements: Work is supported by NIH grant 1R01GM109718, NSF BIG DATA IIS-1633028, NSF DIBBS ACI-1443054. Prepared by LLNL under Contract DE-AC52-07NA27344. LLNL-ABS-805897.

REFERENCES

- [1] R.M. Anderson and R.M. May. 1991. *Infectious Diseases of Humans*. Oxford University Press, Oxford.
- [2] Jessica E. Atwell, Josh Van Otterloo, Jennifer Zipprich, Kathleen Winter, Kathleen Harriman, Daniel A. Salmon, Neal A. Halsey, and Saad B. Omer. 2013. Nonmedical Vaccine Exemptions and Pertussis in California, 2010. *Pediatrics* (2013).
- [3] Jose Cadena, Feng Chen, and Anil Vullikanti. 2017. Near-Optimal and Practical Algorithms for Graph Scan Statistics. In *SIAM Data Mining (SDM)*.
- [4] Jose Cadena, David Falcone, Achla Marathe, and Anil Vullikanti. 2019. Discovery of under immunized spatial clusters using network scan statistics. *BMC Medical Informatics and Decision Making* 19, 1 (2019), 28.
- [5] S. Eubank, H. Guclu, V. S. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczka, and N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429 (2004), 180–184. Issue 6988.
- [6] Tung-Wei Kuo, Kate Ching-Ju Lin, and Ming-Jer Tsai. 2015. Maximizing Submodular Set Function With Connectivity Constraint: Theory and Application to Networks. *IEEE/ACM Transactions on Networking* 23, 2 (2015), 533–546.
- [7] Tracy A Lieu, G Thomas Ray, Nicola P Klein, Cindy Chung, and Martin Kulldorff. 2015. Geographic clusters in underimmunization and vaccine refusal. *Pediatrics* 135, 2 (2015), 280–289.
- [8] GL Nemhauser, LA Wolsey, and ML Fisher. 1978. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming* 14, 1 (1978), 265–294.