Results and Findings

Task 1a: Learning complex models

We demonstrated both theoretically and empirically that our graph learning framework requires fewer calls to BestSubgraph than a naive greedy approach, $O(N^3)$ as compared to $O(N^4)$ for exact greedy search, and O(N log N) as compared to $O(N^2)$ for approximate greedy search, resulting in 2 to 3 orders of magnitude speedup in practice. All five methods were able to learn the structure of graphs with 100 nodes, typically requiring less than an hour of run time, and the faster methods (using the ULS algorithm for event detection) were able to scale to graphs with 500-1000 nodes. It is still an open question whether these methods can be extended to scale up to massive graphs with millions of nodes, as necessary for evaluation of online social network data, and we hope to address this in future work.

We tested the detection power and spatial accuracy of these approaches for detecting various types of simulated disease outbreaks, including outbreaks which spread according to spatial adjacency, adjacency plus simulated travel patterns, and random graphs (Erdos-Renyi and preferential attachment), and observed tradeoffs between scalability and detection performance.

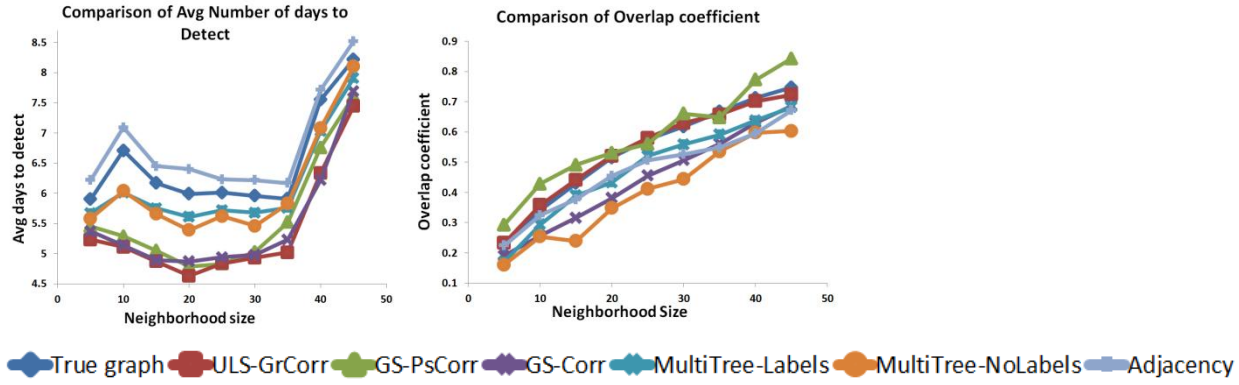


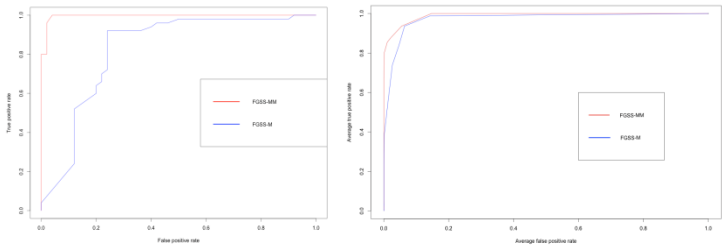
Figure 3 from Somanchi and Neill (2013): Comparison of detection performance of the true, learned, and adjacency graphs for injects based on adjacency with simulated travel patterns.

We demonstrated that the graph structures learned by our framework can then be used by graph-based event detection methods such as GraphScan, enabling more timely and more accurate detection of events (such as disease outbreaks) which spread based on that latent structure. For example, the above figure compares performance of our methods on detecting outbreaks which spread via spatial adjacency plus (simulated) travel patterns, demonstrating improved detection power and spatial accuracy, as compared to the true graph (spatial adjacency + travel) and the spatial adjacency graph without travel patterns. Most recently, we compared the results to an existing, state-of the art algorithm (MultiTree) and demonstrated significant improvements in detection performance (though MultiTree can scale to much larger graph sizes). For example, in the above graphs, our methods consistently achieved more timely and more accurate event detection than MultiTree, even when MultiTree was provided with labeled data not available to our algorithms.

In general, the learned graph structures tend to be similar to the true underlying graph structure, capturing nearly all of the true edges but also adding some additional edges. The resulting graph achieves similar spatial accuracy to the true graph, as measured by the overlap coefficient between true and detected clusters. Interestingly, the learned graph often has better detection power than the true underlying graph, enabling more timely detection of outbreaks or other emerging events. This result can be better understood when we realize that the learning procedure is designed to capture not only the underlying graph structure, but the characteristics of the events which spread over that graph. By finding graphs where the highest connected subgraph score is consistently close to the highest unconstrained subset score when an event is occurring, we identify a graph structure which is optimized for timely and accurate event detection.

Task 1d: Learning new event models

In McFowland and Neill (2013, in preparation), we evaluated our extensions to the Fast Generalized Subset Scan (FGSS) algorithm, FGSS-Mixture (FGSS-MIX) and FGSS-Multiple Models (FGSS-MM), on their ability to discover novel anomalous patterns in data from two real-world application domains: network intrusion detection and masquerade detection (i.e., rapidly identifying individuals using a computer system who have legitimate credentials but are not who they claim to be and are likely involved in harmful activities, based on their observed actions). We evaluated the methods on the KDD Cup 1999 Network Intrusion dataset (which contains a variety of simulated network intrusions on a military network environment) and the R-U-U Masquerade dataset (which contains low-level activity patterns of a computer system's legitimate users, in addition to that of masqueraders). We used two main evaluation metrics: ROC curves (to measure how well the method can distinguish between datasets containing anomalies and those without anomalies) and precision versus recall curves (to measure how well the methods can distinguish between anomalous and normal records in a dataset containing anomalies). We observe that FGSS-MM experiences large improvements in detection performance, in both domains, as compared to the simpler FGSS-MIX approach (note that the original FGSS algorithm does not provide a mechanism for discovery of novel patterns). For example, the two graphs below show ROC curves for the KDD Cup (intrusion type: "Guess Password") and R-U-U datasets respectively.



We also show that FGSS-MM enables effective iterative discovery of novel attacks in the intrusion detection domain, identifying previously unseen attack types given the existing set of known attack types. We are in the process of evaluating the cumulative number of false positives (records reported as anomalous that correspond to normal activity or a previously known attack type) before each algorithm identifies examples of all known attack types, assuming an iterative discovery process in which we learn

a model of each newly discovered attack type once it is found, add it to the set of known models, and re-run the algorithm with the new set of models. One simplifying assumption in this evaluation is that, once a new attack type is discovered, a training dataset containing many examples of that attack type is provided to the algorithm, allowing it to model the new attack type using standard Bayes Net structure learning algorithms. The next stage of our work will address the more difficult and more realistic case where active learning approaches must be used to find other examples of that attack type and learn the model starting from the single discovered training example.

Task 2a: Extend LTSS to general multivariate datasets

In McFowland, Speakman, and Neill (2013), we evaluated the ability of our Fast Generalized Subset Scan (FGSS) method to detect anomalies in general datasets from three real-world domains. First, we considered the 1999 KDD Cup network intrusion dataset, which contains a variety of simulated network intrusions on a military network environment. Second, we considered synthetic anomalies injected into real container shipment data, simulating the anomalous patterns which may result from illicit activities such as smuggling or terrorism. Finally, using a state-of-the-art simulator (Hogan et al., 2007), we simulated the effects of an airborne anthrax release on the number and spatial distribution of respiratory ED cases. We compared FGSS to several previously proposed methods: Anomaly Pattern Detection (Das et al., 2008), Anomalous Group Detection (Neill et al., 2009), and the Bayesian Network Anomaly detector (which identifies individually anomalous records). We evaluated each method's ability to a) distinguish between anomalous and normal records in a dataset which contains anomalies (as measured by the precision vs. recall curve), b) distinguish between datasets containing anomalies and those without anomalies (as measured by the ROC curve), c) identify which attributes are anomalous for the affected group of records (as measured by the overlap coefficient between true and reported sets of attributes), and d) scale to large and high-dimensional data (as measured by runtime as a function of the number of records N and number of attributes M). Our results demonstrate that FGSS can successfully detect and characterize useful anomalous patterns in all three of these application domains, and can scale up to huge and high-dimensional datasets. For example, the following table (Table 1 from McFowland, Speakman, and Neill, 2013) demonstrates substantial improvements in performance for the container shipment dataset as measured by area under the precision vs. recall curve.

Table 1: PIERS Container Shipment Data: Average area (in percent) under the PR curve, with standard errors. For each row, the method which demonstrates the best performance, and those methods with performance not significantly different at significance level $\alpha = 0.05$, are bolded.

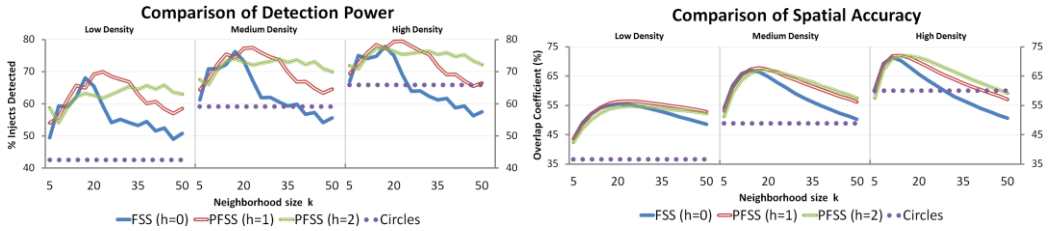
N	k_{inj}	s_{inj}	m_{inj}	$FGSS-BJ$	$FGSS-HC$	AGD	APD	BN
1000	1	10	1	**76.9±3.9**	52.3±4.7	62.2±4.2	47.7±4.3	18.8±2.7
1000	1	10	2	**80.9±3.2**	67.6±4.0	64.9±4.1	65.5±4.1	38.9±3.7
1000	4	25	1	**94.2±1.0**	61.7±2.9	**93.0±1.2**	52.9±2.0	43.5±2.2
1000	4	25	2	**97.3±1.0**	87.3±1.6	94.3±0.7	77.3±1.5	73.1±1.8
1000	10	10	1	**90.8±1.2**	62.0±1.9	80.4±1.5	52.9±1.6	39.6±1.5
1000	10	10	2	**91.5±0.8**	85.4±1.0	83.5±1.0	75.7±1.3	71.4±1.2
10,000	4	25	1	**79.7±2.5**	40.2±3.5	These	41.8±2.9	8.0±1.0
10,000	4	25	2	**71.2±1.8**	65.2±2.5	runs	64.2±2.6	26.4±1.9
10,000	10	10	1	**54.3±2.4**	41.5±2.7	did not	16.2±1.4	6.8±0.8
10,000	10	10	2	51.6±2.0	**65.4±1.7**	complete	40.2±2.3	26.7±1.4

In general, FGSS consistently outperformed APD and BN with respect to detection power, with comparable or better performance as compared to AGD. Additionally, FGSS was able to scale to datasets consisting of 100,000 data records, while AGD fails to scale even to 10,000 records, and FGSS

was also able to much more accurately characterize the affected subset of attributes as compared to the competing methods. Full results are available in the published paper (McFowland, Speakman, and Neill, 2013).

Task 2b: Extend LTSS to constrained subset scans

In Speakman, Somanchi, McFowland, and Neill (2013), we present an empirical comparison of detection power and spatial accuracy for a large set of simulated respiratory disease outbreaks injected into real-world Emergency Department data from Allegheny County, PA. Three methods are compared: Kulldorff's spatial scan statistic (Circles) which returns the highest scoring circular region, Fast Subset Scanning (FSS), which returns the highest scoring unpenalized subset within a region consisting of a center location and its k −1 nearest neighbors, and Penalized Fast Subset Scanning (PFSS), which returns the highest scoring *penalized* subset within a region consisting of a center location and its k −1 nearest neighbors. The soft proximity constraints reward spatial compactness while penalizing sparse regions. We provide results for both weaker (h = 1) and stronger (h = 2) constraints, and note that the FSS method is equivalent to PFSS with h = 0.

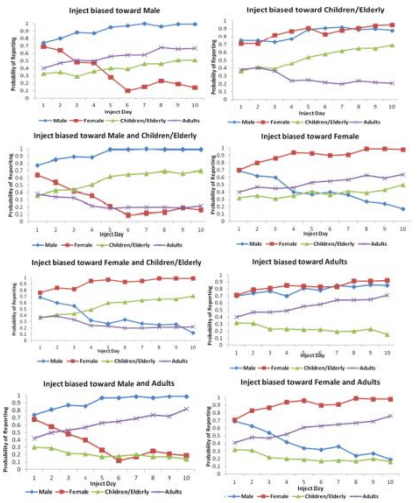


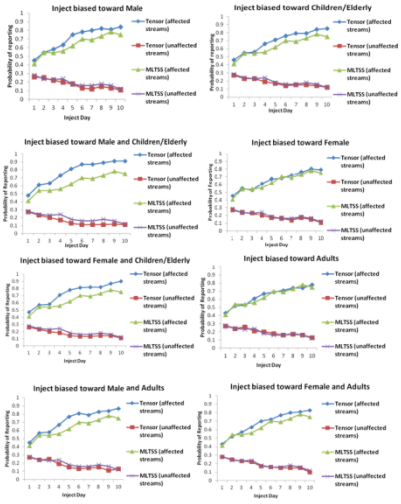
The left figure above provides a comparison of detection power (proportion of outbreaks detected at 1 false positive per year) across multiple methods, neighborhood sizes k, and outbreak densities. For "medium density" injects, the detection power of FSS peaked at 76.2% for k = 17 but was much lower for large k, dropping to 54.1% for k = 47. Although peak performance for PFSS was similar at 77.5% for k = 23 and h = 1, we note that PFSS was much more robust to parameter selection, with detection power only decreasing to 63.2% (k = 47) and 65.9% (k = 8) for h = 1 and h = 2 respectively. As expected, this pattern holds for high density outbreaks as well. PFSS also performed surprisingly well in the low density outbreaks: despite the lack of spatial structure, our method with soft proximity constraints outperforms the unpenalized method across a wide range of neighborhood sizes k.

The right figure above provides a comparison of spatial accuracy across multiple methods, neighborhood sizes k, and outbreak densities. For medium density outbreaks, FSS and PFSS have similar peak overlap coefficients of 67.0% (k = 14) and 67.9% (k = 17, h = 1) respectively. However, both PFSS methods maintain high performance for a wide range of neighborhood sizes from k = 17 to k = 50. Comparing the weaker and stronger penalties for PFSS, we note that h = 2 lags slightly behind h = 1 for low density injects, but surpasses h = 1 for high density outbreaks. This shift reflects the role that stronger soft proximity constraints have when detecting outbreaks with a strong, tight spatial structure.

Task 2b, continued: Extension of LTSS to multivariate and tensor datasets

We evaluated MD-Scan using simulated disease outbreaks injected into real-world Emergency Department data from Allegheny County, PA. In addition to demonstrating substantial improvements in detection time and spatial accuracy for outbreaks which have differing impacts on different subpopulations, MD-Scan was able to accurately identify the characteristics (including, but not limited to, demographics such as age, gender, race/ethnicity, and socioeconomic status) of the affected subpopulation. Additionally, detection performance was further enhanced by incorporating additional constraints such as spatial proximity and graph connectivity into the iterative detection procedure.





For example, the set of graphs on the left (Figure 3 from Neill and Kumar, 2013) demonstrates that MD-Scan is able to correctly report the affected demographic groups, while the set of graphs on the right (Figure 2 from Neill and Kumar, 2013) demonstrates that MD-Scan outperforms our previous multivariate LTSS approach in terms of correctly reporting the three affected data streams (diarrheal, respiratory, and fever Emergency Department visits) without reporting the unaffected streams.

Task 2c: Fast graph scanning

Our previous annual report presented the evaluation of GraphScan's detection time and spatial accuracy, using a set of simulated respiratory disease outbreaks injected into real-world Emergency Department data from Allegheny County, Pennsylvania. GraphScan demonstrated substantial improvements in detection time and spatial accuracy as compared to LTSS without connectivity constraints, the previously proposed Upper Level Set scan statistic (Patil and Taillie, 2004), and the unconstrained LTSS approach.

The figure below (Figure 7 from Speakman, McFowland, and Neill, 2013) shows GraphScan's huge increase in speed over previous methods that searched for the highest scoring connected subset. The previously proposed FlexScan method (Tango and Takahashi, 2005) required a week of computation time to search over neighborhoods of size 30 for a single day's worth of data. Through the speed improvements of the LTSS property, GraphScan is able to perform this search in less than one second. Searching over all connected subgraphs (without neighborhood constraints) for the 97 Allegheny County

zip codes requires about 2.5 seconds for the latest version of GraphScan (as compared to about 2 minutes for the previous version) and would require hundreds of millions of years for FlexScan.

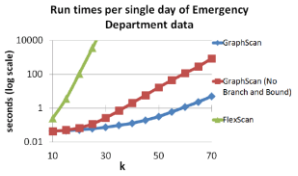


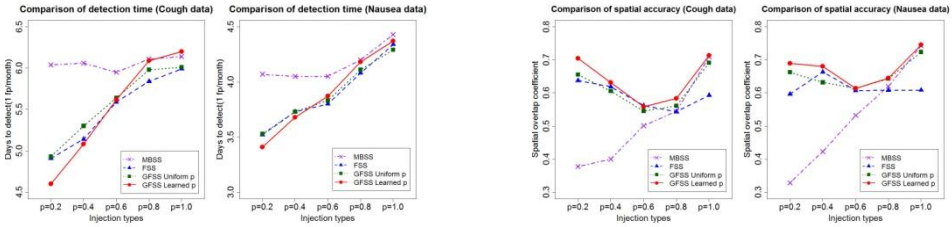
Figure 7: Run time analysis for FlexScan and GraphScan with and without Branch and Bounding. The x-axis denotes the "neighborhood size" as various values of k.

Task 2c, continued: Event detection in heterogeneous social media graphs

As a case study of our new approach to event detection in heterogeneous social media graphs, we consider civil unrest event detection using Twitter data, and present empirical evaluations illustrating the effectiveness and efficiency of our proposed approach. The dataset used was randomly sampled from 10% of the raw Twitter data from July 2012 to December 2012 across 6 countries: Argentina, Chile, Colombia, Ecuador, Mexico, and Venezuela. For verification, we validated our detection results against labeled sets collected from reliable news outlets. For each country, three top newspapers were selected from the 2012 top 100 newspapers in Latin America, as provided by International Media & Newspapers. In addition, news was also collected from the most influential international news outlets and with subject matter expert input. We labeled an event as "significant" by checking whether it was reported by any of these news outlets to format our Gold Standard Report (GSR). The GSR events were then classified to two categories: violent and non-violent. Given an event warning generated by our approach, if the detected date is within the time window that starts seven days before a true event date and ends at the true event date, it is counted as a positive match; otherwise, it is counted as a negative match or a false alarm. We then calculated the Receiver Operating Characteristic (ROC) curves for the six countries for both violent and non-violent event types. The results demonstrate the effectiveness of our approach. We also observe that performance on detection of violent events is better than that on non-violent events. One potential reason is due to the fact that violent events tend to attract a higher volume of attention in Twitter than non-violent events. As a result, the signal of violent events could be stronger than that of non-violent events and hence can be more accurately detected. We intend to finish the empirical evaluation shortly and submit the full paper to ICDM 2013 this summer.

Task 3b: Scalable event detection and visualization

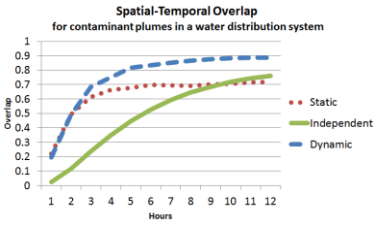
In Shao, Liu, and Neill (2011), we demonstrate that the Generalized Fast Subset Sums (GFSS) method with learned distribution of the sparsity parameter p outperforms our previously proposed Multivariate Bayesian Scan Statistic (MBSS) and Fast Subset Sums (FSS) methods, as well as the GFSS method assuming a uniform distribution of p. The figures below (Figures 7 and 8 of Shao, Liu, and Neill, 2011) compare these methods with respect to detection power and spatial accuracy respectively, demonstrating the value of learning the model parameters from data. More detailed results are provided in the full, published paper.



Task 3c: Automated event investigation and tracking

In Speakman, Zhang, and Neill (2013), we describe a novel graph-based event detection approach which can accurately identify and track dynamic outbreaks (where the affected region changes over time). Our approach enforces soft constraints on temporal consistency, allowing detected regions to grow, shrink, or move while penalizing implausible region dynamics. Using simulated contaminant plumes diffusing through a water distribution system, we demonstrate that our method enables both earlier detection of contaminant events and more accurate tracking of the spread of contamination.

We compared the performance of our novel dynamic GraphScan method to two competing methods on a detection task derived from the "Battle of the Water Sensor Networks" data, requiring the methods to identify which nodes are affected over time as the contamination spreads through the network. Outbreak plumes were simulated in a water distribution system for 12 one-hour periods. We assumed noisy binary sensors (with 10% false positive and 90% true positive rates) observed hourly at each pipe junction. Our method ("Dynamic") was compared to the "Static" method, which assumes that the affected region does not change over time and therefore is constrained to only return spatial-temporal "cylinders", and the "Independent" method, which separately optimizes the spatial subset for each time slice without taking temporal consistency into account. The methods were evaluated on spatial-temporal overlap, defined as the number of sensors contained in both the detected and affected space-time regions divided by the number of sensors in either the detected or affected space-time regions. A measure of 1 is a perfect match of spatial subsets across each time window and 0 would reflect disjoint space-time regions. Additionally, average time to detect an outbreak (at a fixed false positive rate of 1/month) was 4.24, 4.56, and 6.65 hours for the Dynamic, Static, and Independent methods respectively.



We observe that the Static method performs well in the early stages of the spreading plume. However, its performance levels off because it is unable to capture the dynamics of the pattern as it spreads through the system. Independent struggles early because the signal from an individual time slice may be difficult to detect; as the plume grows, it is more easily identified without additional temporal information. The Dynamic method shows strong detection performance in the early stages of the

contamination event while also being able to accurately track the propagation of the plume beyond that of the Static method.

Task 4a: Prediction using leading indicator data

In Flaxman and Neill (2012), we demonstrated that our Iterative Average Dot Product (IADP) method can accurately maximize cross-correlation over subsets of the monitored data streams and proximity-constrained subsets of locations. For jointly maximizing cross-correlation over streams and locations, we compared to ground truth for small problem sizes (for which we can compute the true maximum by exhaustive search over all subsets) and performed exploratory analysis for large subsets. For maximizing cross-correlation over streams for a given set of locations, we also compared to previous methods in the literature (LASSO and Google Correlate). Our results demonstrate that IADP has similar run time to Google Correlate and much faster run time than LASSO, while finding subsets with significantly higher correlation than either competing method. For our exploratory analysis, IADP found a spatial area consisting of 10 census tracts in the West Englewood neighborhood of Chicago, and a subset of 12 of the 28 monitored data streams, for which the correlation between this week's leading indicator count and next week's violent crime count was 0.786 (time series shown below); the average correlation of the regions it found, averaged over all 77 community areas, was 0.546.

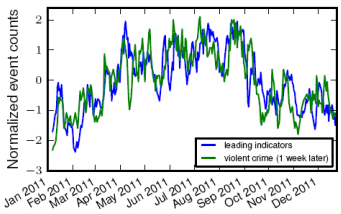


Figure 3: Best correlation discovered in table 6 between LIs and VC is $r = 0.786$, occurring in community area 67 (West Englewood).

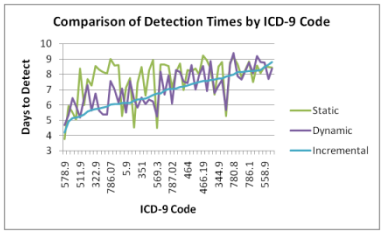
In comparison, the LASSO and Google Correlate methods (which could not search over subsets of census tracts, and thus simply aggregated counts for each community area), found much lower maximum correlations (0.660 and 0.652 for West Englewood respectively) and much lower average correlations (0.325 and 0.404 respectively).

In Flaxman and Neill (2013), we evaluated our new, kernel-based test for space-time interaction on both simulated data and real-world crime data (shootings/homicides and 911 calls) from Chicago. Our power analysis on simulated data follows that of Diggle et al. (1995), generating points from a cluster process. The power of our method was comparable to or better than the best set of thresholds for Diggle's test, which represents the current state of the art in space-time interaction testing. Based on conversations with city officials, we focused on a very specific and relevant question for the Chicago data: which types of calls to 911 exhibit space-time interaction with homicides and aggravated battery with a handgun. The goal is to respond more proactively to these leading indicators so as to prevent homicides. We used a dataset provided by the Chicago Police Department of geocoded, date- and time-stamped calls to 911 and another dataset of crime incident reports. Data were from 2007 through the first half of 2010. There were just under 9 million calls for service (911 calls, plus calls to the dispatcher made by police for a

variety of reasons) and 9,087 homicides and shootings reported in this time period. There were 271 different types of calls to 911. We considered each separate type of 911 call as one point process and homicides / shootings as the other point process, and we compared our new test for leading indicators (based on directional, bivariate space-time interaction) with the bivariate Knox test, reporting all leading indicators with significant space-time interaction ($\alpha < .01$) for each test. As an interesting (but very preliminary) evaluation, we had a representative from the CPD evaluate the plausibility of each leading indicator on a 1-5 scale, and our test was found to return a more plausible set of leading indicators than bivariate Knox. Of course, a much larger sample of evaluators would be necessary to make this subjective evaluation more comprehensive, and we are also performing an objective evaluation by comparing the prediction performance of CityScan using the different sets of leading indicators found by each method.

Task 4c: Incorporating rich text data

In Liu and Neill (2011), we compared the performance of the semantic scan statistic to the traditional spatial scan approach on the outbreak detection task, using Emergency Department data from Allegheny County, PA. Our data consisted of 612,713 cases from 1/1/2004-12/31/2005, with a free-text chief complaint, International Classification of Diseases (ICD-9) code, and general syndrome classification (prodrome) for each case. We generated synthetic outbreaks consisting of cases drawn at random using a given ICD-9 code, and compared the results of the standard spatial scan (maximizing over the ten prodrome categories: respiratory, gastrointestinal, etc.) and the semantic scan (generating topics from the free text and maximizing over the topics). We used average detection time (in days) as well as proportion of outbreaks detected, at a fixed false positive rate of 1 per month, as our evaluation metrics. The figure below compares the detection performance of the static, dynamic, and incremental subset scan methods, demonstrating that the dynamic and incremental methods tend to achieve more timely detection than the static method across a wide range of disease syndromes (as grouped by International Classification of Diseases (ICD)-9 codes).

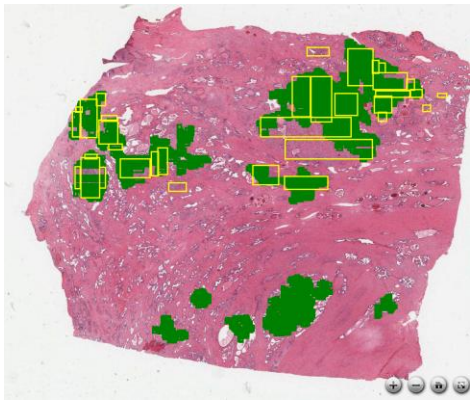


Additionally, all three semantic scan methods were able to accurately identify synthetically-generated unprecedented outbreaks (e.g. a disease which makes the patient's nose turn green), detecting in an average of 5.3 to 6.4 days, while the typical prodrome-based method had low power to detect such novel outbreaks, requiring an average of 10.9 days to detect.

Task 4d: Incorporating society-scale data

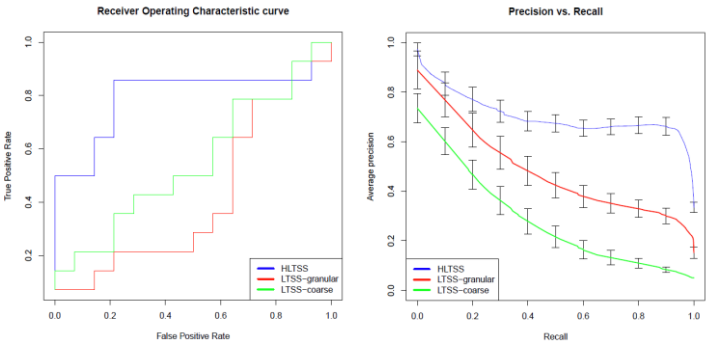
We demonstrate the performance of our novel methods for event detection in massive multi-scale images, on the task of identifying cancerous locations on digital slides of prostate biopsy samples. Our

preliminary evaluations show that our methods detect regions of cancer in a few minutes with high accuracy, both as measured by the ROC curve (measuring ability to distinguish between benign and cancerous slides) and by the spatial precision-recall curve (measuring ability to pick out the malignant areas on a slide which contains cancer). The initial dataset consisted of 28 prostate biopsy samples from the Department of Pathology, University of Pittsburgh Medical Center: 14 images which contained cancerous locations and 14 images which were examples of benign prostate biopsy samples. The former 14 cancerous images were also annotated by a pathologist indicating the locations of cancer. These annotations were used for evaluating the performance of our method in detecting regions of interest within a digital image. Each image contained about 5 billion pixels at the most granular level and required around 10GB of space in uncompressed form. We are in the process of acquiring more images with annotations to show the effectiveness of our method on a large number of examples.



From Somanchi and Neill (2013): An example of our detected regions as compared to ground truth. The green regions were detected by our HLTSS method, and the yellow rectangles are ground truth boxes drawn by a pathologist.

The above figure shows an example of our detected regions as compared to the ground truth results. We can see that we have identified most of the regions with high accuracy though there were a few false positives and false negatives. The two figures below respectively show the Receiver Operating Characteristic curve for differentiating cancerous images, and the Precision and Recall curve for finding cancerous regions within a cancerous image. We observe that our hierarchical linear-time subset scanning method is able to achieve high performance on both tasks, while standard linear-time subset scanning (without taking the hierarchy into account) performs poorly.



Results of Educational Initiatives

As described in detail in "Research and Educational Activities", this research project has directly resulted in the mentoring of fifteen graduate students (five current, eight graduated, and two left the lab), including the completion of seven First and Second Heinz Research Papers (graduation requirements for the Ph.D. in Heinz College at CMU), one master's thesis, and three MSIT capstone projects. Additionally, Dr. Neill has developed a new curriculum in Machine Learning and Policy at CMU, including three new courses (described in "Research and Educational Activities") and a new joint Ph.D. program in Machine Learning and Public Policy, with the first student joining the program in Fall 2011. The special topics course, "Special Topics in Machine Learning and Policy: Mining Massive Datasets", was taught for the first time in Spring 2013, with high relevance to future work incorporating massive and societal-scale data on this project (Task 4d).

References

H. S. Burkom, S. P. Murphy, J. Coberly, and K. Hurt-Mullen, K. Public health monitoring tools for multiple data streams. *Morbidity and Mortality Weekly Report* , 54 (Supplement):55–62, 2005.

K. Das, J. Schneider, and D. B. Neill. Anomaly pattern detection in categorical datasets. *Proc. 14th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 169-176, 2008.

S. Flaxman and D. B. Neill. Detecting spatially localized subsets of leading indicators for event prediction. Submitted for publication, 2012.

S. Flaxman and D.B. Neill. New tests for space-time interaction in spatio-temporal point processes. In *Proceedings of the 2nd Spatial Statistics Conference*, 2013, accepted.

W. R. Hogan, G. F. Cooper, G. L. Wallstrom, M. M. Wagner, and J. M. Depinay. The Bayesian aerosol release detector: an algorithm for detecting and characterizing outbreaks caused by an atmospheric release of Bacillus anthracis. *Statistics in Medicine* 26(29): 5225-52, 2007.

M. Kulldorff. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26(6):1481-1496, 1997.

M. Kulldorff, F. Mostashari, L. Duczmal, W. K. Yih, K. Kleinman, and R. Platt. Multivariate scan statistics for disease surveillance. *Statistics in Medicine*, 26: 1824-1833, 2007.

T. Kumar and D. B. Neill. Fast tensor scan for event detection and characterization. Submitted for publication, 2012.

Y. Liu and D.B. Neill. Detecting previously unseen outbreaks with novel symptom patterns. *Emerging Health Threats Journal* 4: 11074, 2011.

E. McFowland III, S. Speakman, and D.B. Neill. Fast generalized subset scan for anomalous pattern detection. *Journal of Machine Learning Research*, in press, 2013.

D. B. Neill. Detection of Spatial and Spatio-Temporal Clusters. Ph.D. thesis, Carnegie Mellon University, School of Computer Science, 2006.

D. B. Neill. Fast Bayesian scan statistics for multivariate event detection and visualization. *Statistics in Medicine* 30(5): 455-469, 2011.

D. B. Neill. Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society (Series B: Statistical Methodology)* 74(2): 337-360, 2012.

D.B. Neill. New directions in artificial intelligence for public health surveillance. *IEEE Intelligent Systems* 27(1): 56-59, 2012.

D.B. Neill. Using artificial intelligence to improve hospital inpatient care. *IEEE Intelligent Systems*, 2013, in press.

D. B. Neill and G. F. Cooper. A multivariate Bayesian scan statistic for early event detection and characterization. *Machine Learning* 79: 261-282, 2010.

D. B. Neill, G. F. Cooper, K. Das, X. Jiang, and J. Schneider. Bayesian network scan statistics for multivariate pattern detection. In J. Glaz, V. Pozdnyakov, and S. Wallenstein, eds., *Scan Statistics: Methods and Applications*, 221-250, 2009.

D.B. Neill and T. Kumar. Fast multidimensional subset scan for outbreak detection and characterization. *Online Journal of Public Health Informatics* 5(1), 2013. Presented at the 2012 International Society for Disease Surveillance Annual Conference.

D. B. Neill and J. Lingwall. A nonparametric scan statistic for multivariate disease surveillance. *Advances in Disease Surveillance*, 4:106, 2007.

D. B. Neill, E. McFowland III, and H. Zheng. Fast subset scan for multivariate event detection. *Statistics in Medicine*, 32: 2185–2208, 2013.

D. B. Neill, A. W. Moore, and G. F. Cooper. A Bayesian spatial scan statistic. In *Advances in Neural Information Processing Systems 18*, 1003–1010, 2006.

D. B. Neill, A. W. Moore, M. R. Sabhnani, and K. Daniel. Detection of emerging space-time clusters. In *Proc. 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 2005.

K. Shao, Y. Liu, and D.B. Neill. A generalized fast subset sums framework for Bayesian event detection. *Proceedings of the 11th IEEE International Conference on Data Mining*, 617-625, 2011.

S. Somanchi and D.B. Neill. Fast graph structure learning from unlabeled data for outbreak detection. *Emerging Health Threats Journal* 4: 11017, 2011.

S. Somanchi and D.B. Neill. Fast graph structure learning from unlabeled data for event detection. Submitted for publication, 2012.

S. Speakman and D. B. Neill. Fast graph scan for scalable detection of arbitrary connected clusters. In *Proc. International Society for Disease Surveillance Annual Conference*, 2010.

S. Speakman, E. McFowland III, and D.B. Neill. Scalable detection of anomalous patterns with connectivity constraints. Submitted to *Journal of Computational and Graphical Statistics*, 2013.

S. Speakman, E. McFowland III, S. Somanchi, and D.B. Neill. Scalable detection of irregular disease clusters using soft compactness constraints. *Emerging Health Threats Journal* 4: 11121, 2011.

S. Speakman, S. Somanchi, E. McFowland III, and D.B. Neill. Disease surveillance, case study. In R. Alhajj and J. Rokne, eds., *Encyclopedia of Social Network Analysis and Mining*. Springer, 2013, in press.

S. Speakman, S. Somanchi, E. McFowland III, and D.B. Neill. Penalized fast subset scanning. Submitted to KDD 2013.

S. Speakman, Y. Zhang, and D.B. Neill. Tracking dynamic water-borne outbreaks with temporal consistency constraints. *Online Journal of Public Health Informatics* 5(1), 2013. Presented at the 2012 International Society for Disease Surveillance Annual Conference.

T. Tango and K. Takahashi. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 4:11, 2005.