Data Mining Critical Infrastructure Systems: Models and Tools

Anika Tabassum*, Supriya Chinthavali†, Liangzhe Chen‡ and B. Aditya Prakash*

***Abstract*—Critical infrastructures (CIs) such as power, water, transportation, and telecommunication are highly complex interacting systems that are vital to national security, economy and public life. They play an important role in several core urban computing challenges. Advances in computing resources and techniques have led to enormous progress in developing intelligent frameworks for analyzing these large heterogeneous systems. In this article, we survey state-of-the-art and foundational work in this upcoming area from a data mining perspective. We discuss basic concepts of CIs, their properties, impacts on them due to natural or human-caused disturbances and different computational methodologies used for modeling and understanding their behavior. We also discuss recent work that specifically deals with two critical sectors of CIs, namely power and transportation systems. Finally, we also describe several existing tools and methods that are used to facilitate decision making for domain operators, enable efficient and faster disaster response for federal and state agencies and help improve the security and resiliency of these CIs.**

***Index Terms*—Critical Infrastructures, Urban Computing, Data Mining**

I. Introduction

Urban computing is a process of measuring, modeling, analyzing and integrating complex heterogeneous data gathered from urban spaces [1]. The rapid growth of urbanization and modernization of people's lives have led to serious sustainability issues and hence there is a critical need for designing efficient and environment friendly systems, utilizing traffic flow in the city, situational awareness during extreme events to improve resiliency, public health, air pollution, etc. These complex and dynamic requirements give rise to several computational challenges in transportation, environment, cyber physical systems and internet of things [2], [3]. Some of these challenges include predicting power consumption [4], measuring air quality [5] and predicting, utilizing and controlling crowd/traffic flow in a city [6]–[8]. With the advent of social media, localizing and visualizing disaster events using such data has also proven to be a fertile ground for computational problems. For example, Sakaki et al. [9] proposed a classifier to monitor and detect earthquakes from Twitter and Yang et al. [10] designed a visualization of a four phase model using Twitter data (that represents content (what), location (where), time(when) and the user network (who) responding to the disaster). Public health issues like syndromic surveillance of a disease like influenza [11], and then controling it [12] are important problems as well.

Critical Infrastructure Systems: Critical infrastructure refers to systems, facilities, technologies and networks that are vital to security, public health and socio-economic well being of people. Clearly, they play an important role in many urban computing challenges. For example communication networks are inherently crucial for disaster response and controlling traffic flow. Hence, strengthening and maintaining secure and resilient Critical Infrastructure Systems (CISs) is a primary US national goal (even addressed through a presidential policy directive (PPD-21) [13]), and it requires proactive and coordinated efforts among federal, state, local, public and private owners and operators of CISs. Critical infrastructure networks (CIs) such as power, water, transportation, etc. are highly interdependent, and failure of one has a cascading effect on another which affect national security, economy and public health. Infact, a single vulnerable network can have a huge impact due to interdependencies. For example, the well-known 2003 Northeast blackout in U.S. [14] impacted multiple CIs. The massive power outage affects water and waste systems, transportation, communication, financial services, which cascaded to impact public health and food industries. Nearly 50 million people were affected and cause a loss of $5 billion to U.S. national economy [15]. Smart cities are extensively leveraging telecommunication technologies and cyber physical systems to provide a safe and a sustainable environment for increasing urban populations [16]. This also leads to an increasing risk of triggering cascading failures due to complex interactions and inter dependencies leading to debilitating impacts and creating urban computing challenges of cyber security, real-time situational awareness, handling traffic flow, meeting electricity demand, managing public health etc.

Data mining challenges: Modeling and analyzing such CISs gives rise to a rich and fruitful space of data mining challenges.

1) *Complexity.* CIs networks have a complex structure. As mentioned above, even a single CI network, e.g., electrical grid system consists of many underlying subsystems, *i)* power generator generates power using different types of fuel, *ii)* transmission network transfer power to different distribution substations over long distances, *iii)* distribution substations transfer power to local facilities and residential areas over the distribution grid, *iv)*Oil and Natural gas (NG) pipeline networks carry fuel to power generation stations. These pipeline networks consist of natural gas compressors,gas processing plants, NG terminals and other subsystems. These complex subsystems consist of large-scale data which is very useful for

*Department of Computer Science, Virginia Tech. Email: {anikat1,badityap}@cs.vt.edu

†Oak Ridge National Laboratory. Email: chithavalis@ornl.gov

‡Pinterest. Email: liangzhechen@pinterest.com

This article is based on our recent tutorial at SIAM Data Mining (SDM) 2018.

analyzing CIs.

2) *Heterogeneity.* CI networks are also extremely heterogeneous. They consist of many interdependencies like *i)* physical, where one infrastructure is physically connected or interdependent on another infrastructure e.g. power lines connected to water pumping stations, *ii)* geographical, where changes caused by local environmental events impact all CI components that are co-located, *iii)* cyber, state of infrastructure depends on information transmitted through information infrastructure, e.g., electronic and informational linkages. This heterogeneity and dependencies give rise to different types of nodes, edges, links, and multiple sources of information in the network [17].
3) *Dynamics.* CISs are also highly dynamic. Multiple incidents can cause failure of CI networks, e.g., loss of power, natural or human-made disaster which affect network in different states of operation varying with time. This dynamic property makes the system modeling more challenging.
4) *Scale.* For all these situations, designing scalable algorithms is a fundamental goal. These systems are in large scale and hence naturally give rise to 'big-data' problems.

Overview: In this article, we present the state-of-the-art research on models and tools used within CISs. This area of research is highly interdisciplinary, with connections to high-impact areas, like public safety/security, national economy, physics and power engineering and social media. We will first discuss various approaches to CI modeling. Then we mainly focus on two specific CI networks: power and transportation systems. We chose these networks since they are the heart of the sixteen CIs defined by the Department of Homeland Security (DHS) which have the potential to impact every other CIs (as discussed before, see the 2003 blackout example). In addition, these two areas have also seen a spate of recent work from a data mining perspective. We finally look into the methods that help improve situational awareness during disasters and then present some existing CI tools designed for helping domain experts in decision making and used by agencies such as national labs.

II. Modeling

CIs are highly interdependent and complex — failure of a single CI network can severely affect other CIs. Hence, in order to understand critical infrastructures to identify vulnerabilities, to protect them against threats and to support decision making, modeling and simulating them are essential. Modeling and simulating the interdependencies of CIs can be categorized into system dynamics-based, agent-based, network-based approaches [15].

System dynamics based approaches typically model CIs utilizing a *top-down* method to manage and analyze complex interdependencies based on domain knowledge of the particular system [15]. The main philosophy of such approaches is that for modeling a system, it is necessary to understand the behavior of a system. Different system-dynamics models have been proposed different sectors of CIs like telecommunication, petroleum, natural gas, and electric supply systems [18]–[20]. As an example, different models for electric grids have been proposed which represent system behavior using physical power equations and flow of electricity [20]–[22]. These models exhibit high-fidelity, and due to their complexity and highly detailed structure, they are also computationally expensive.

In contrast, agent-based modeling adopts a *bottom-up* approach which considers complex system behavior arising from the interaction of individual autonomous agents. For example, several agent-based systems [22]–[24] model power systems by considering the interaction and impacts of electric power markets on interdependencies of CIs. Similarly, to control and simulate traffic systems, different agent-based models have been proposed from a civil engineering perspective which consider environmental factors and vehicle crashes [25]–[27]. Balmer et al. [28] generate strategies for a traffic model by simulating the movement of agents, avoiding obstacles and generating congestion.

Finally, network-based approaches view CIs as *networks* where nodes represent different CIs components, and the links represent connections or edges among them. This approach requires less domain knowledge for modeling than the former, and hence they can be generalized to different systems. As a result while they are not high-fidelity and can not model all behaviors of the systems, they are great at modeling specific aggregate aspects. Hence they are also closer to data mining methodologies and problems. For instance, to ensure reliable broadcast in communication networks Duan et al. [29] studied interdependencies between communication and power grid networks and proposed an algorithm to handle both crash failures in communication and cascading failures in the power grid. As another example, Lee et al. created a system to generate CI networks [30] and model a cascade of failures among different CIs based on their physical topology and temporal dependencies. Chen et al. [31] proposed an optimization algorithm named *FASCINATE* to infer cross-layer dependencies in multilayered CI networks where each layer consists of a different CI network. To infer the cross-layer relation between two different networks, they viewed multilayered connections as a collective collaborative filtering problem. The overall approach is shown in Fig 1.

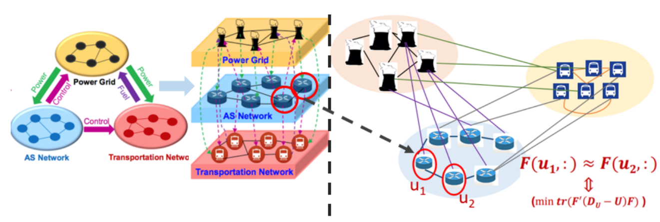


Fig. 1. An example of a network-based approach to system modeling. The goal of *FASCINATE* proposed by Chen et al. [31] is to infer dependecies among the CI networks. Left figure shows how a CI network is converted into a cross-layered network. Right figure shows how they find out the hidden dependencies across the transportation and power grid networks by viewing it as a collective collaborative filtering problem.

III. Example CI: Power Systems

In this section, we provide a few examples of work on CI systems dealing with power and energy, from a network modeling perspective. As representative work, we focus on methods that identify vulnerable facilities to protect against unknown natural disasters (non-adversarial) and to protect the system against adversarial attacks with known patterns and strategies.

A. *Identifying vulnerable facilities*

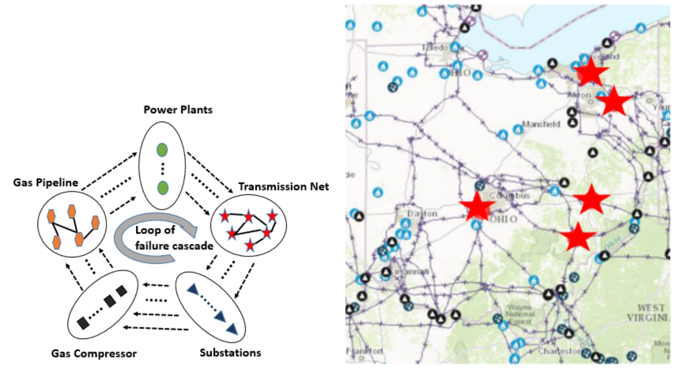
Identifying vulnerable critical facilities in a power system is necessary to protect and enhance them against unknown natural disasters. The state-of-the-art can be divided into mainly two different techniques: using only the network structure and/or incorporating failure cascade dynamics as well.

Several algorithms have been proposed in order to identify critical nodes using only the network structure [32]–[34]. Arianos et al. [32] introduced the concept of using geodesic distance for power flow to calculate resiliency of power grids. Chen et al. [33] proposed an algorithm *OPERA* for the connectivity control problem. It aims to find a set of optimal nodes that maximizes the impact on an interdependent CI network consisting of physical, control and communication layers. In addition, they developed *SUBLINE*, to unify a family of prevalent subgraph connectivity measures to quantify network dependency of the graph and subsequently use it for identifying the optimal nodes.

Additionally, incorporating failure dynamics can help prevent catastrophic failures of the whole system. The 2003 blackout shows an instance (see Section I for detail) where a transmission line failure cascaded to failures of water, waste-treatment, and communication systems. Buldyrev et al. [34] developed a framework based on mutually connected clusters to study cascading failures in interdependent networks. They analyzed the presence of a giant connected cluster under simple random failures on Erdos-Renyi networks and show a phase transition and a critical threshold. In contrast, Chen et al. [35] developed an algorithm *HOTSPOTS* to model more complicated failure cascades and identify critical nodes that may lead to substantial failures. Their heterogeneous network consists of power plants, substations, transmissions and gas compressors (see Fig. 2). First, they propose a 'path-based' failure cascade model on this complex system representing how every component of a CI network interacts with each other. Instead of the typical neighbor-based cascade models they propose a novel path-based failure cascade (*F-CAS*). Second, given the *F-CAS* model, they formulate an optimization algorithm to identify a set of critical transmission nodes whose failure will maximize the number of failed substations (another CI network). Finally, they propose a dominator-tree based approach to solve this problem efficiently.

B. *Protecting power system against attacks*

Next we discuss several works that studied how to detect an attack, and protect and/or to reduce the effect of an adversarial attack on power grid. To detect an attack or a node failure in the network, Hooi et al. [36] proposed an online anomaly detection algorithm *GRIDWATCH* that can help a sensor to detect a failure in the electrical grid. Using *GRIDWATCH* they also suggested an optimization which can maximize the probability of detecting a failure in power grid within a given budget.



(a) Heterogeneous network (b) Critical transmission nodes

Fig. 2. An example of identifying vulnerable facilities in power system using both network structure and failure dynamics (Section III). Fig. (a) is an overview of *HOTSPOTS* proposed by Chen et al. [35] which shows a heterogeneous network in a power system which consists of power plants, transmission networks, substations, gas compressor, and gas pipelines. Fig. (b) shows the maximum failed transmission nodes in OH that the algorithm identifies and that these nodes are close to the failed nodes during 2003 blackout [14].

To enhance grid resiliency it is necessary to understand which nodes should be protected first under targeted attacks. Using the giant connected cluster formulation of Buldyrev et al. [34] discussed above, Huang et al. [37] developed a framework to understand robustness in interdependent networks under degree-based targeted attacks. Their main idea was to map a targeted network to a random one (on a different network). Their findings show that protecting the higher degree nodes with low probability to fail can significantly improve robustness of interdependent networks.

Finally, if an attack happens it is necessary to reduce its impact on the interdependent network. Strategies for this task have been proposed by many papers [38], [39]. Wang et al. [38] proposed a load redistribution approach where a failure at node i redistributes its load to its neighboring nodes to reduce the impact of an attack. Ouyang et al. [39] proposed a tri-level optimization problem which maximizes the resiliency of the system and also minimize loss. The inner level optimizes the damaged components to repair, middle level identifies the most disruptive attack, and outer level optimizes the defense decision.

IV. Example CI: Transportation Systems

In transportation systems, research has been done on predicting traffic on roads, traffic states like accidents, road constructions as well as several approaches to improve congestion control.

A. *Traffic flow and state prediction*

Traffic flow is the study of interactions of vehicles with the traffic infrastructure, such as traffic control devices, highways and traffic signs. Predicting traffic flow and states can help improve an intelligent transportation system and prevent congestion. In order to predict traffic flow on roads, it is necessary to understand the influence of road segments based on propagating congestion. Anwar et al. [40] developed an algorithm to identify the most congested areas on traffic using the road intersection network, the number of vehicles passing during green lights and the ratio of effective usage of green light time on each road segment. Many different algorithms have also been proposed in order to forecast traffic flow at time $t+k$ from the recent traffic flow data up to time t [41]–[43]. Wu et al. [41] suggest a random effect model integrating the temporal factors while Moretti et al. [42] developed a hybrid ensemble technique using an artificial neural network along with a statistical regression approach. As another example, Zheng et al. [43] predicted traffic flow using occupancy data from buildings, instead of using traffic data directly.

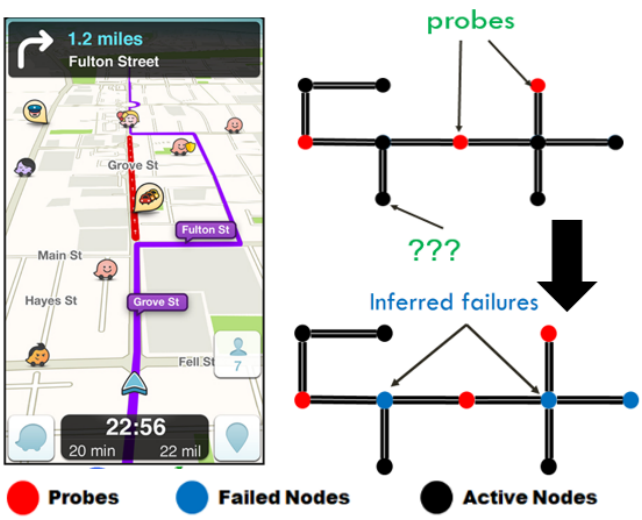


Fig. 3. Traffic state inference using partial data. *GRAPHSTATEINF* proposed by Adhikari et al. [44] dynamically identifies all the failed nodes in the road network by leveraging a subset observations of failed nodes. The left figure shows an example of a crowd-sourced app using which the user reports incidents (the so-called probes of failed nodes) (left figure). In the right figure, a toy road network showing the observed partially failed nodes (marked as red) and inferred unobserved failed nodes (marked as blue).

Traffic states govern traffic flow and can be categorized into 1) traffic infrastructure states (TIS), e.g., weather, presence of accidents, roadworks and 2) traffic flow state (TFS), e.g., flow rate, density, speed. From the partial observations of TFS and TIS at several periods of different traffic links Gu et al. [45] designed a model to estimate the traffic states using expectation maximization and kalman filters. Instead, Adhikari et al. [44] formulate the network state detection problem as an inference problem given partial state data (from a crowd-sourced app like Waze). They develop a near-optimal efficient algorithm, *GRAPHSTATEINF*, which tries to find a set of failed nodes in a network using observed failed nodes and correlations among failure nodes in the network. Their main idea was to leverage the information-theoretic MDL (minimum description length) principle, which searches for the 'best' set of failures which 'explains' the given partial set with the minimum encoding cost. This approach is useful in predicting the impact of networks due to congestion (shown in Fig. 3).

B. Congestion tracking and control

To build an improved traffic system it is necessary to control or reduce traffic congestion. For this, detecting any congestion is the first step. Anwar et al. [46] model congestion on road networks as partitions of these networks such that road blocks in each partition are homogeneous and have similar congestion (see Fig. 4, where roads in the same round block have similar congestion). However, traffic congestion varies with time and re-partitioning them at each time-step is computationally expensive. Hence they also develop an algorithm to incrementally update the congested partitions at a new time point based on previous time and current traffic data. Their main idea is to find the unstable nodes and assign these nodes to a block which maximizes the number of bounded cycles in a block.

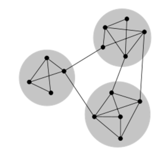


Fig. 4. Congestion Example

In order to control congestion, Sundar et al. [47] designed an automatic traffic signal control system from the traffic density in the route. They used signals from RFID sensors inserted in every vehicle, which transmits messages to the reader. Along with congestion control, this system is also able to detect stolen vehicles and clear traffic signal automatically for the emergency vehicles.

V. Facilitating Decision Making

Various CI systems and tools have been developed to aid the domain experts in taking multifaceted decisions for emergency management, improve situation awareness as well as for budget planning purposes. Next we give an overview of this space.

A. Improving situation awareness

The idea of situation awareness is to predict emergency and differentiate casual events from non-casual events, and many approaches use crowdsourcing in some way for this along with other techniques. Liang et al. [48] built a classifier to distinguish flooded areas from non-flooded areas from satellite images. The authors developed a semi-supervised learning algorithm which divides the satellite image into several patches based on the proximity and intensity of the pixels. A user is asked to label a few patches and based on that the classifier automatically classifies all other patches. Fig 5 shows the steps of the algorithm. They used crowd-sourced knowledge for getting user feedback for their classifier instead of using domain expert to label the image patches.

Similarly, in the context of traffic data, Hooi et al. [49] proposed an algorithm to find out traffic accidents by detecting change points from sensor data. Huang et al. [50] designed a crowd-sourcing based anomaly prediction system which allows a user to report urban anomalies they encountered. Based on these historical anomaly data they developed a Bayesian inference model to understand dependency among regions regarding the anomaly distribution. Next, they built a Markov model to learn state transitions between normal and

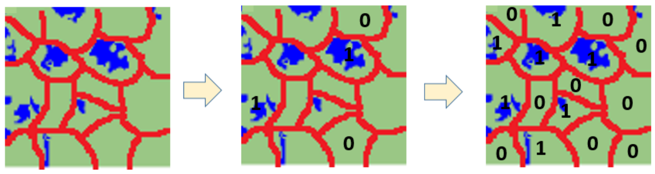


Fig. 5. An example of human-guided flood detection technique from satellite images proposed by Liang et al. [48] for improving situation awareness (see Sec V). In Step 1 the algorithm divides their images into several patches. In Step 2, some of the patches of the image are labeled using crowdsourced feedback. In Step 3, based on the user feedback, the algorithm labels other patches as flood or non-flooded areas.

anomaly data and predict the state of the next time slot. Muralidhar et al. [51] also developed an online monitoring system *ILLIAD* for anomaly detection and state estimation in cyber physical systems (CPS), e.g., wireless and wired networks. They combined model-based and data driven approaches to learn invariant functional relationships between components of the CPS (shown to represent the underlying network structure of the CPS). Next they checked for the violation of any of these invariant relationships over time which was then treated as an anomaly (see Fig. 6).

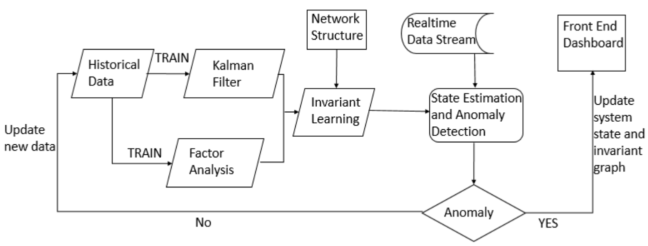


Fig. 6. An example to improve situation awareness in CPS (wireless and wired networks). The framework of *ILLIAD* proposed by Muralidhar et al. [51]. They used Kalman filters (model based) and autoregression and latent factor based (data-driven) methods to learn the invariant functional relationship between the components and used that for state estimation and anomaly detection.

Social media data can also be leveraged to detect disaster-prone areas. Mcclendon et al. [52] show how social media data can support the decision for emergency management by categorizing disaster-affected areas. Using Twitter users as sensors Sakaki et al. [9] designed a classifier to detect target events and a probabilistic spatiotemporal model to find the center and trajectory of the event. Zhao et al. [53] proposed a model to predict spatial events in social media considering different spatial locations and various spatial relationship with the task. Several frameworks have been built to find disaster location and events from web pages as well. Farag et al. [54] developed an event model that can automatically capture the event information and incorporated the model into a focused crawler algorithm which can identify the web pages relevant to that event. To detect the disaster prone-areas and for fast emergency response using social media, it is also necessary to identify trustworthy users whose contents are influential. Vedula et al. [55] developed an unsupervised algorithm to identify the trustworthy and influential users from the network during a crisis.

B. Other CIS systems and tools

Several CIs tools have been developed to support the decision-making process. Argonne National Lab has developed a risk-based decision support system and a simulation model named Critical Infrastructure Protection and Decision Support System (CIPDSS) [56] to protect CI systems against vulnerabilities, natural or human-made disasters, etc. For studying energy development and impacts of climate Los Alamos National Lab (LANL) designed Climate-Energy Assessment for Resiliency (*CLEAR*) model which enables to assess the interdependency of CIs regarding their relationship with climate. Figure 7 shows an example of *CLEAR* [57] assessing CO_2 emission in transportation and energy sectors.

In addition, a toolkit *URBAN-NET* developed by Lee et al. [30] integrates network construction, visualization, failure cascade modeling, and a simulator to identify critical facilities. They generated the physical CI networks from disparate data sources which contains location and information of CI components. For modeling and simulation, they considered the system into three different categories: topology-based, simulation-based and monitoring based analysis. In the topology-based analysis, they consider only physical CI interdependencies and compute the importance of each node and link based on their interdependencies. In simulation-based analysis, they also incorporated temporal dependencies such as the restoration period of network failure, and capacity of a CI component to handle the failure of its interdependent network. They showed an example with a road-gas network where each node importance is based on its efficiency of transportation as well as how well it is reachable to gas stations. Finally they also created a visualization to show the critical nodes and edges (shown in Fig. 8).

There are other such tools have been developed by Oak Ridge National Lab (ORNL) to facilitate decision making for urban infrastructures. *URBAN-CAT* [58] has been developed to understand the impacts of climate change, *URBAN-MET* [59] has been designed to study interactions between urban and environmental systems, *LANDSCAN* [60] has been developed to model the distribution of population and settlement based on demographic and remote sensing imagery data. Besides, several CIs datasets are also available to aid research and decision making: a geospatial and US domestic infrastructure HSIP gold data [61], NHDPlus and USGS hydrology data [62], [63], and EIA energy data [64].

VI. Conclusion

In this survey, we presented an overview of state-of-the-art in CI research from an urban computing and data mining perspective. First, we describe the state-of-the-art approaches that have been used in CIs for modeling dynamics in the system. First we discussed some of the different frameworks for modeling CI systems in general. Second, we discuss the data mining problems related to two vital sectors in CIs, electric grid systems, and transportation systems. Within the power grid systems, we showcase existing techniques to identify vulnerable facilities and to protect the system from adversarial attacks. Within the transportation domain, we

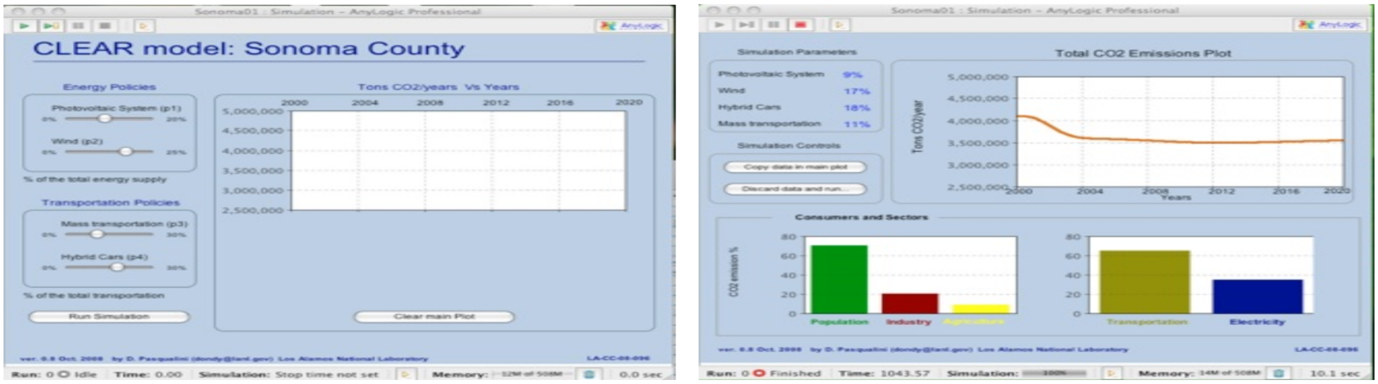
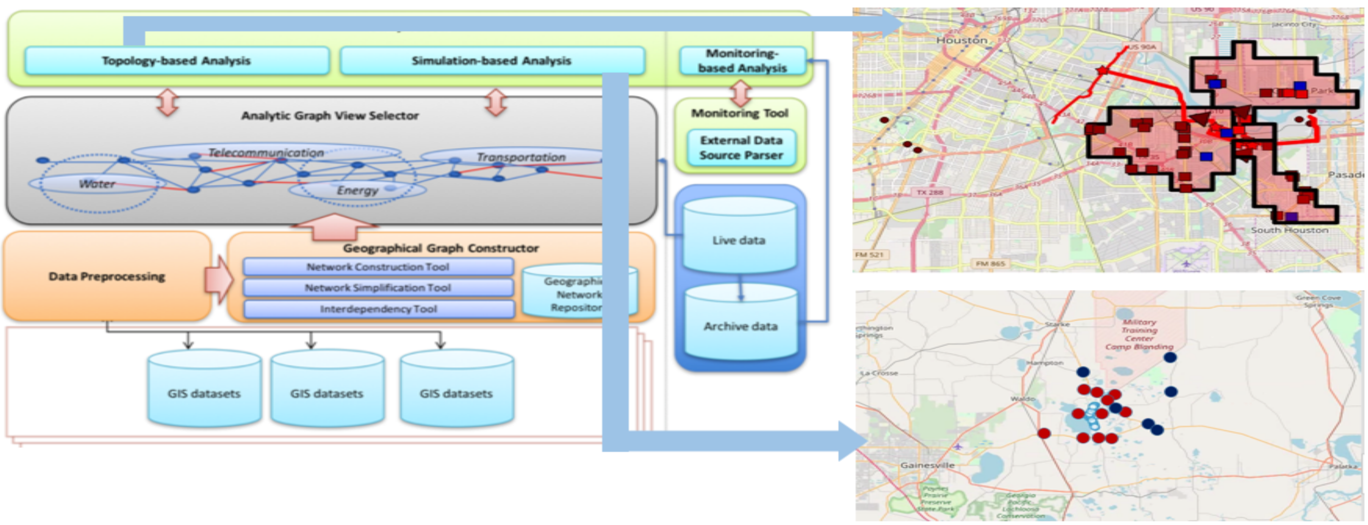


Fig. 7. A screenshot of the interface of *CLEAR* designed by LANL [57] showing emission of CO_2 in typical sectors of CIS. In the left figure a user can choose the energy and transportation policy, whereas the right figure shows the simulation result, i.e., the amount of CO_2 emission in different sectors based on the chosen policy.



s

Fig. 8. Overview of *URBAN-NET* by Lee et al. [30]. Left fig. shows how *URBAN-NET* is collecting GIS data and preprocessing then to analyze CI networks. Top right fig. shows a topology-based example where the system identifies the vulnerable CI networks which fails due to the failure of some initial seed nodes. Bottom right fig. shows a simulation-based analysis where it identifies the critical nodes which fail after a certain time due to interdependency.

presented algorithms for traffic states and flow prediction and to control congestion. Finally, we also described some popular tools and techniques, that have been developed to ease the decision-making process for domain experts.

There are several open problems and this is a rich domain with high potential for interdisciplinary impact. For example, in context of energy systems, some of these problems include: *1)* Interpreting or explaining the behavior of CIs models, for e.g., are the algorithms able to identify critical facilities from the system; *2)* In terms of modeling, federal entities such as Department of Energy (DoE) focus on improving understanding of how extreme events (such as hurricanes and wildfires) impact the production of electricity and power equipment (such as flooded substations, downpoles etc.) [65]. Since renewable generation such as solar and wind tends to be highly intermittent, there is a lot of interest to resolve this issue to aim towards uninterrupted electricity supply and improve sustainability [66]; *3)* Impacts of electricity production caused by outages in gas pipe lines is also being heavily pursued due to growth in natural gas fuel production within the US; *4)* Another area of investigation is to understand the impacts of power restoration due to disruptions in the transportation infrastructure.

Acknowledgements: This article is based on work partially supported by the NSF (IIS-1353346, CAREER IIS-1750407), the NEH (HG-229283-15), ORNL, the Maryland Procurement Office (H98230-14-C-0127), and a Facebook faculty gift.

References

[1] "Urban dynamics institute." [Online]. Available: https://udi.ornl.gov/

[2] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.

[3] F. Salim and U. Haque, "Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and internet of things," *International Journal of Human-Computer Studies*, vol. 81, pp. 31–48, 2015.

[4] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1956–1967, 2017.

[5] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1436–1444.

[6] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction." in *AAAI*, 2017, pp. 1655–1661.

[7] Y.-J. Wu, F. Chen, C. Lu, B. Smith, and Y. Chen, "Traffic flow prediction for urban network using spatio-temporal random effects model," in *91st Annual Meeting of the Transportation Research Board (TRB)*, 2012.

[8] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu, "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 1–10, 2017.

[9] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 851–860.

[10] S. Yang, H. Chung, X. Lin, S. Lee, L. Chen, A. Wood, A. L. Kavanaugh, S. D. Sheetz, D. J. Shoemaker, and E. A. Fox, "Phasevis1: What, when, where, and who in visualizing the four phases of emergency management through the lens of social media." in *ISCRAM*, 2013.

[11] L. Chen, K. T. Hossain, P. Butler, N. Ramakrishnan, and B. A. Prakash, "Syndromic surveillance of flu on twitter using weakly supervised temporal topic models," *Data mining and knowledge discovery*, vol. 30, no. 3, pp. 681–710, 2016.

[12] Y. Zhang, A. Ramanathan, A. Vullikanti, L. Pullum, and B. A. Prakash, "Data-driven immunization," in *Data Mining (ICDM), 2017 IEEE International Conference on*. IEEE, 2017, pp. 615–624.

[13] "Presidential policy directive – critical infrastructure security and resilience." [Online]. Available: https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil

[14] B. Liscouski and W. Elliot, "Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations," p. 86, 2004.

[15] M. Ouyang, "Review on modeling and simulation of interdependent critical infrastructure systems," *Reliability engineering & System safety*, vol. 121, pp. 43–60, 2014.

[16] H. Boyes, R. Isbell, and T. Watson, "Critical infrastructure in the future city," in *International Conference on Critical Information Infrastructures Security*. Springer, 2014, pp. 13–23.

[17] P. Pederson, D. Dudenhoeffer, S. Hartley, and M. Permann, "Critical infrastructure interdependency modeling: a survey of us and international research," *Idaho National Laboratory*, vol. 25, p. 27, 2006.

[18] G. P. O'Reilly, A. Jrad, A. Kelic, and R. LeClaire, "Telecom critical infrastructure simulations: Discrete-event simulation vs. dynamic simulation how do they compare?" in *Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE*. IEEE, 2007, pp. 2597–2601.

[19] T. Brown, "Multiple modeling approaches and insights for critical infrastructure protection," *NATO Security through Science Series D-Information and Communication Security*, vol. 13, p. 23, 2007.

[20] X. Xu, H. Jia, H.-D. Chiang, D. Yu, and D. Wang, "Dynamic modeling and interaction of hybrid natural gas and electricity supply system in microgrid," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1212–1221, 2015.

[21] S.-K. Kim, J.-H. Jeon, C.-H. Cho, J.-B. Ahn, and S.-H. Kwon, "Dynamic modeling and control of a grid-connected hybrid generation system with versatile power transfer," *IEEE transactions on industrial electronics*, vol. 55, no. 4, pp. 1677–1688, 2008.

[22] H. A. Song, B. Hooi, M. Jereminov, A. Pandey, L. Pileggi, and C. Faloutsos, "Powercast: mining and forecasting power grid sequences," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2017, pp. 606–621.

[23] D. C. Barton, E. D. Eidson, D. A. Schoenwald, K. L. Stamber, and R. K. Reinert, "Aspen-ee: an agent-based model of infrastructure interdependency," *SAND2000-2925. Albuquerque, NM: Sandia National Laboratories*, 2000.

[24] K. Hopkinson, X. Wang, R. Giovanini, J. Thorp, K. Birman, and D. Coury, "Epochs: a platform for agent-based electric power and communication simulation built from commercial off-the-shelf components," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 548–558, 2006.

[25] H. A. Aziz and S. V. Ukkusuri, "Integration of environmental objectives in a system optimal dynamic traffic assignment model," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 7, pp. 494–511, 2012.

[26] H. A. Aziz, S. V. Ukkusuri, and S. Hasan, "Exploring the determinants of pedestrian–vehicle crash severity in new york city," *Accident Analysis & Prevention*, vol. 50, pp. 1298–1309, 2013.

[27] F. Zhu, H. A. Aziz, X. Qian, and S. V. Ukkusuri, "A junction-tree based learning algorithm to optimize network wide traffic control: A coordinated multi-agent framework," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 487–501, 2015.

[28] M. Balmer, N. Cetin, K. Nagel, and B. Raney, "Towards truly agent-based traffic and mobility simulations," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*. IEEE Computer Society, 2004, pp. 60–67.

[29] S. Duan, S. Lee, S. Chinthavali, and M. Shankar, "Best effort broadcast under cascading failures in interdependent critical infrastructure networks," *Pervasive and Mobile Computing*, vol. 43, pp. 114–130, 2018.

[30] S. Lee, L. Chen, S. Duan, S. Chinthavali, M. Shankar, and B. A. Prakash, "Urban-net: A network-based infrastructure monitoring and analysis system for emergency management and public safety," in *Big Data (Big Data), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2600–2609.

[31] C. Chen, H. Tong, L. Xie, L. Ying, and Q. He, "Fascinate: Fast cross-layer dependency inference on multi-layered networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 765–774.

[32] S. Arianos, E. Bompard, A. Carbone, and F. Xue, "Power grid vulnerability: A complex network approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 19, no. 1, p. 013119, 2009.

[33] C. Chen, J. He, N. Bliss, and H. Tong, "On the connectivity of multi-layered networks: Models, measures and optimal control," in *2015 IEEE International Conference on Data Mining*. IEEE, 2015, pp. 715–720.

[34] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, p. 1025, 2010.

[35] L. Chen, X. Xu, S. Lee, S. Duan, A. G. Tarditi, S. Chinthavali, and B. A. Prakash, "Hotspots: Failure cascades on heterogeneous critical infrastructure networks," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 1599–1607.

[36] B. Hooi, D. Eswaran, H. A. Song, A. Pandey, M. Jereminov, L. Pileggi, and C. Faloutsos, "Gridwatch: Sensor placement and anomaly detection in the electrical grid," 2018.

[37] X. Huang, J. Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley, "Robustness of interdependent networks under targeted attack," *Physical Review E*, vol. 83, no. 6, p. 065101, 2011.

[38] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the us power grid," *Safety Science*, vol. 47, no. 10, pp. 1332–1336, 2009.

[39] M. Ouyang and Y. Fang, "A mathematical framework to optimize critical infrastructure resilience against intentional attacks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 11, pp. 909–929, 2017.

[40] T. Anwar, C. Liu, H. L. Vu, and M. S. Islam, "Roadrank: Traffic diffusion and influence estimation in dynamic urban road networks," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 1671–1674.

[41] Y.-J. Wu, F. Chen, C.-T. Lu, and S. Yang, "Urban traffic flow prediction using a spatio-temporal random effects model," *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 282–293, 2016.

[42] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, 2015.

[43] Z. Zheng, D. Wang, J. Pei, Y. Yuan, C. Fan, and F. Xiao, "Urban traffic prediction through the second use of inexpensive big data from buildings," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 1363–1372.

[44] B. Adhikari, P. Rangudu, B. A. Prakash, and A. Vullikant, "Near-optimal mapping of network states using probes," in *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 2018, pp. 108–116.

[45] Y. Gu, Z. Qian, and G. Zhang, "Traffic state estimation for urban road networks using a link queue model," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2623, pp. 29–39, 2017.

[46] T. Anwar, C. Liu, H. L. Vu, and M. S. Islam, "Tracking the evolution of congestion in dynamic urban road networks," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 2323–2328.

[47] R. Sundar, S. Hebbar, and V. Golla, "Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1109–1113, 2015.

[48] J. Liang, P. Jacobs, and S. Parthasarathy, "Human-guided flood mapping: From experts to the crowd," in *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018, pp. 291–298.

[49] B. Hooi, L. Akoglu, D. Eswaran, A. Pandey, M. Jereminov, L. Pileggi, and C. Faloutsos, “Changedar: Online localized change detection for sensor data on a graph,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018, pp. 507–516.

[50] C. Huang, X. Wu, and D. Wang, “Crowdsourcing-based urban anomaly prediction system for smart cities,” in *Proceedings of the 25th ACM international on conference on information and knowledge management*. ACM, 2016, pp. 1969–1972.

[51] N. Muralidhar, C. Wang, N. Self, M. Momtazpour, K. Nakayama, R. Sharma, and N. Ramakrishnan, “illiad: Intelligent invariant and anomaly detection in cyber-physical systems,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 9, no. 3, p. 35, 2018.

[52] S. McClendon and A. C. Robinson, “Leveraging geospatially-oriented social media communications in disaster response,” *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, vol. 5, no. 1, pp. 22–40, 2013.

[53] L. Zhao, F. Chen, C.-T. Lu, and N. Ramakrishnan, “Multi-resolution spatial event forecasting in social media,” in *Data Mining (ICDM), 2016 IEEE 16th International Conference on*. IEEE, 2016, pp. 689–698.

[54] M. M. Farag, S. Lee, and E. A. Fox, “Focused crawler for events,” *International Journal on Digital Libraries*, vol. 19, no. 1, pp. 3–19, 2018.

[55] N. Vedula, S. Parthasarathy, and V. L. Shalin, “Predicting trust relations within a social network: A case study on emergency response,” in *Proceedings of the 2017 ACM on Web Science Conference*. ACM, 2017, pp. 53–62.

[56] “Critical infrastructure protection decision support system (cipdss).” [Online]. Available: http://www.bwbush.io/projects/cipdss.html

[57] “Climate-energy assessment for resiliency (clear).” [Online]. Available: https://public.lanl.gov/dp/CLEAR.html

[58] “Urban climate adaptation tool (urban-cat).” [Online]. Available: https://udi.ornl.gov/content/urban-climate-adaptation-tool-urban-cat

[59] “Urban-met: Modeling urban energy savings scenarios using earth system microclimate and urban morphology.” [Online]. Available: https://studylib.net/doc/11896703/urban-met--modeling-urban-energy-savings-scenarios-using

[60] “Landscan.” [Online]. Available: https://landscan.ornl.gov/

[61] “Hsip gold dataset.” [Online]. Available: https://gii.dhs.gov/HIFLD/hsip-guest

[62] “Nhdplus.” [Online]. Available: http://www.horizon-systems.com/nhdplus/

[63] “Usgs water data.” [Online]. Available: https://waterdata.usgs.gov/nwis/rt

[64] “Us energy information administration.” [Online]. Available: https://www.eia.gov/

[65] E. S. Blake, T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven Ii, “Tropical cyclone report: Hurricane sandy,” *National Hurricane Center*, vol. 12, pp. 1–10, 2013.

[66] “Us department of energy (doe) advanced modeling grid research program.” [Online]. Available: https://www.energy.gov/sites/prod/files/2018/07/f53/1.1.1%20NA%20Grid%20Resil%20Modeling%20Panel%20-%20Ghassemian%2C%20DOE.pdf