Critical Infrastructure Data Analytics
Models and Tools

B. Aditya Prakash

Department of Computer Science

Virginia Tech.

http://www.cs.vt.edu/~badityap

Tutorial at NSF Smart Grid Analytics Workshop
Georgia Tech, June 4, 2019

Thanks!

- To all the organizers
  - Santiago Grijalva
  - Zoran Obradovic
  - Mladen Kezunovic
  - Renata Rawlings-Goss

Part of a longer tutorial

- http://people.cs.vt.edu/~badityap/TALKS/18-sdm-tutorial
- Given at SIAM Data Mining 2018
- All Slides are posted there.
- Also, invited article at IEEE Intelligent Informatics Bulletin Dec, 2018.

  http://people.cs.vt.edu/~badityap/papers/cis-ieeeiib18.pdf
- **License**: for education and research, you are welcome to use parts of this presentation, for free, with standard academic attribution. For-profit usage requires written permission by the authors.

Outline

- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Outline

- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Urban computing

- Many broader problems and challenges in big cities

Q1: Smart grid

Q2: Urban flow

Q3: Situation awareness

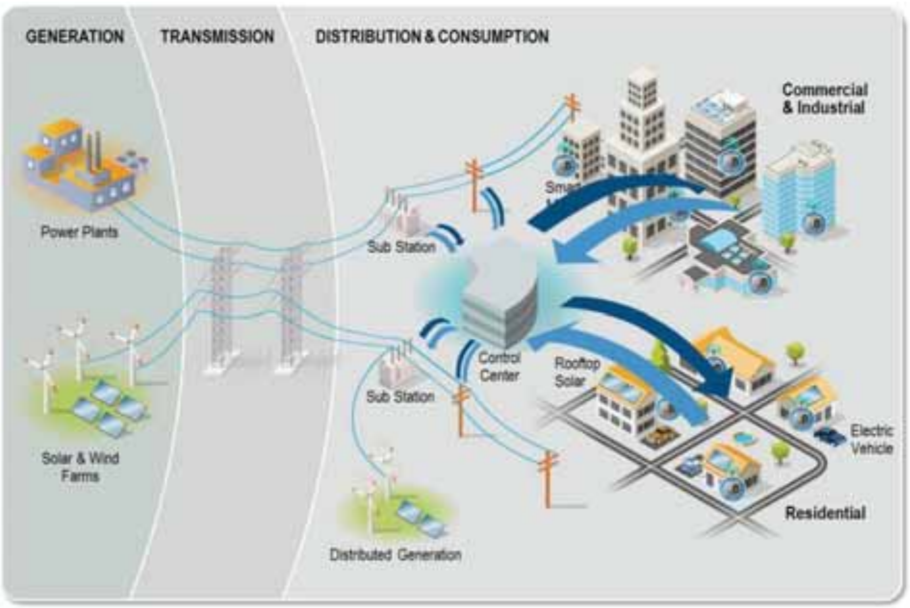
Q4: Robustness & Evolution

Q5: Public Health

Q6: Air pollution

Q1: Smart grid

- How to design more efficient and environment friendlier systems for managing electric grids.



Use digital communication techniques to detect and react to local changes in usage

Challenges:
Huge Data Processing from Sensors
Protection from Cyberterrorism
Privacy Concern

Q2: Urban flow

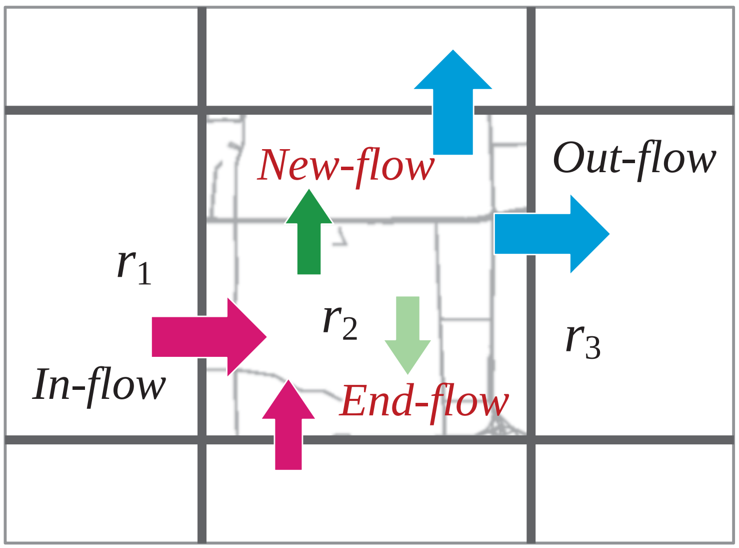
- Traffic flow is the study of interactions between vehicles, drivers, crowds and infrastructure (including highways, signage and traffic control devices)
  - How to predict and utilize the traffic flow in a city.

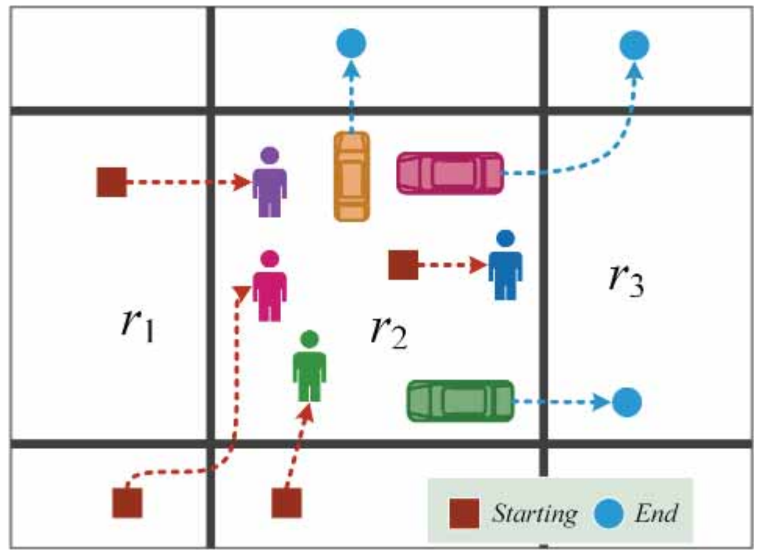
Predicting the traffic flow

Use traffic flow information for finding good billboard locations

Crowd flow prediction

- Predicting the crowd flow in a region at a specific time (ST-ResNet [Zhang et al., AAAI'17])





- Important for:
  - Traffic management
  - Risk assessment
  - Public safety

Microsoft Research

Traffic flow prediction

- Predicting traffic on an urban traffic network (ITS [Wu et al., TRB'12])

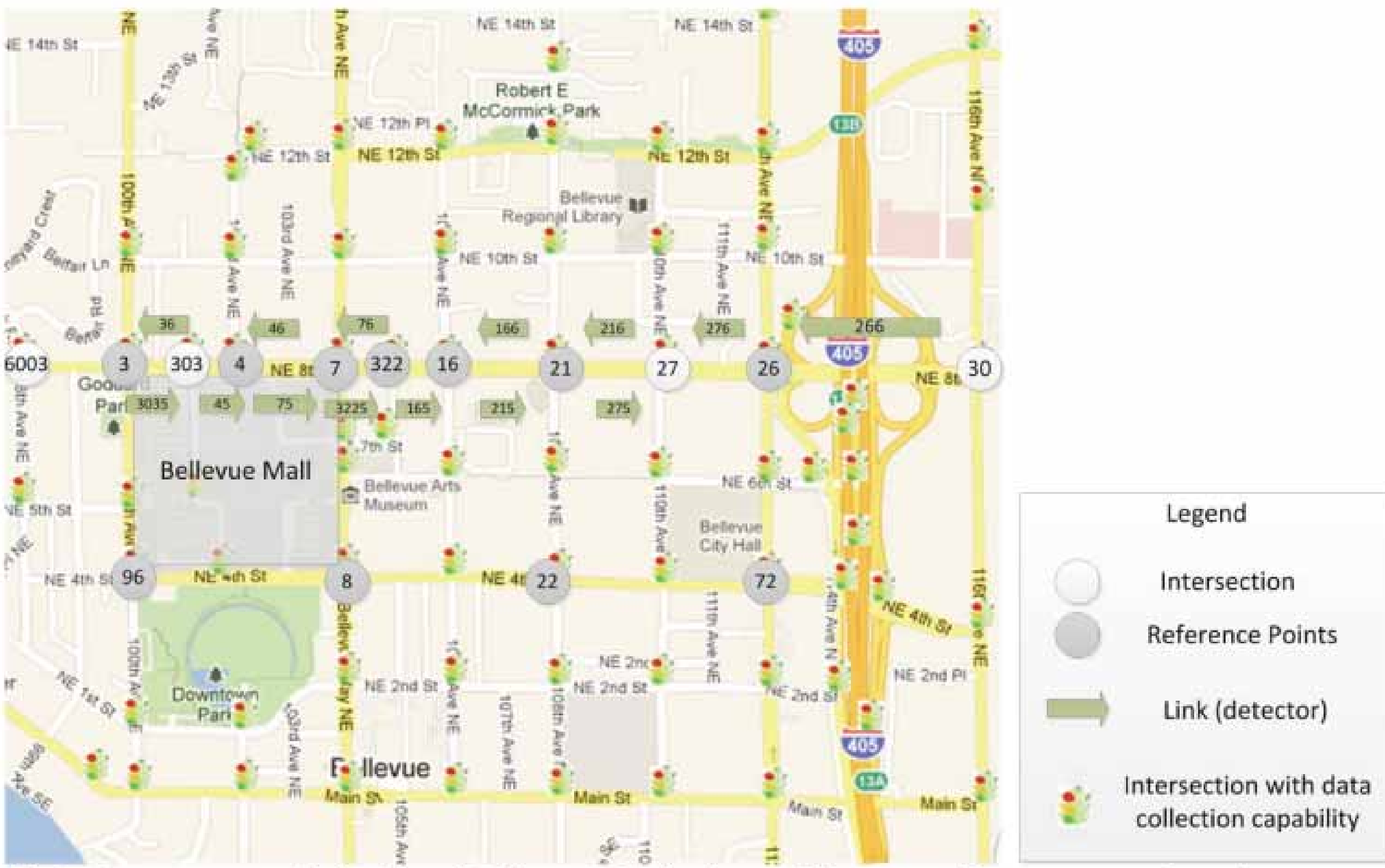
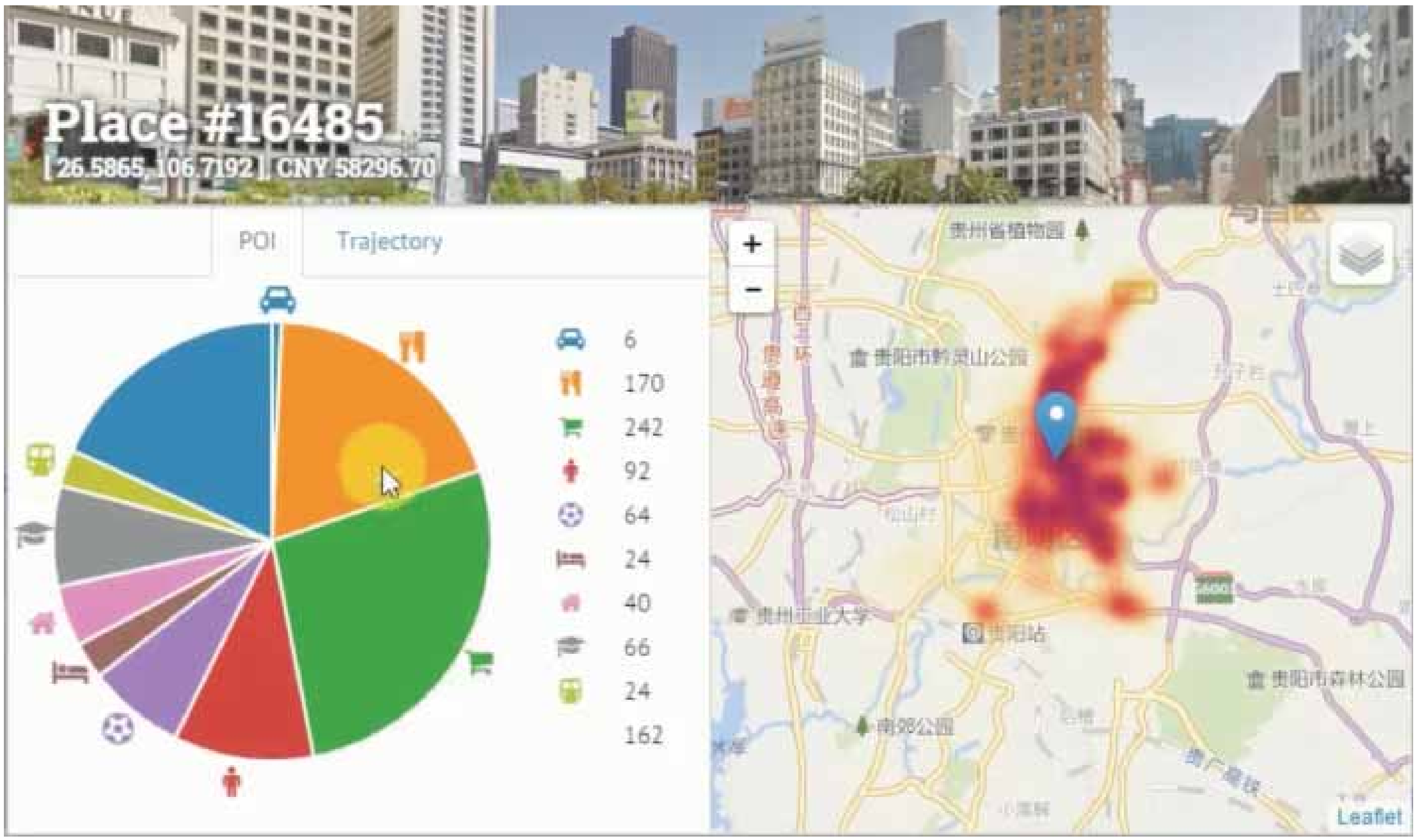


Fig. 1. Downtown area in the City of Bellevue, WA (background images are from maps.google.com)

Smart billboard locations

- Finding the most influential Locations to place billboards using the traffic flow (SmartAdP [Liu et al., IEEE Trans. Vis. Comput. Graph'16])



Guizhou, China

Q3: Situation awareness

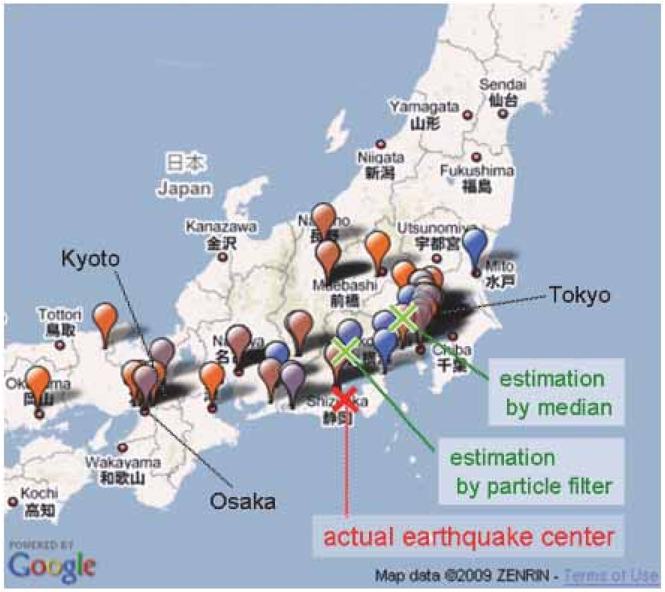
- Improving situation awareness during extreme events using social media
  - Using Twitter to locate events during Earthquakes
  - Detecting fire emergencies

VirginiaTech

Event detection

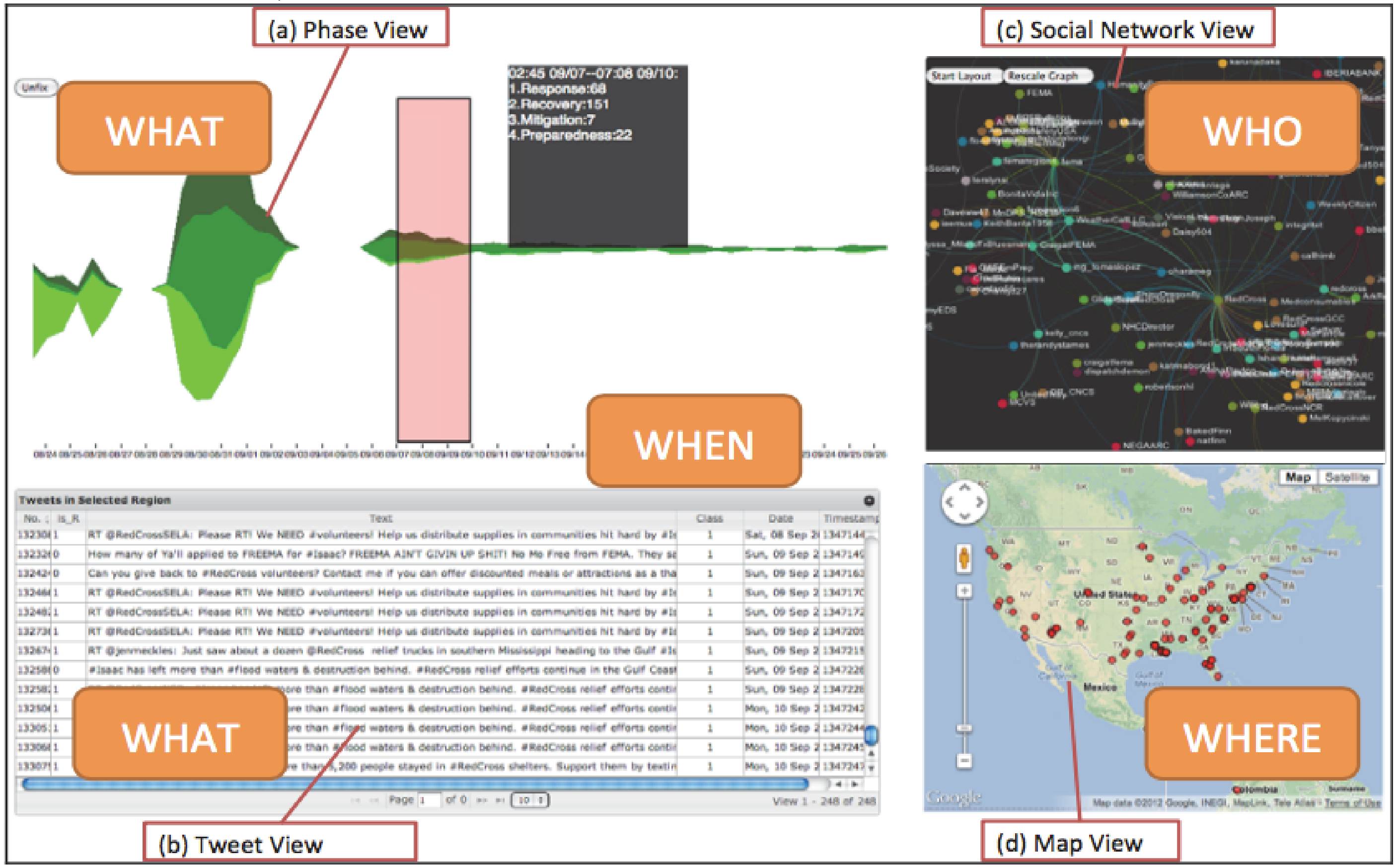
- Detecting earthquake locations using social media (JMA [Sakaki et al., WWW'10])

Published	Location	Title	Screen_name	URL
2009-08-11 05:08:57	Saitama, Japan	地震おおいわー	tondol	http://twitte
2009-08-11 05:08:56	unknown	地震。	Tiroly	http://twitte
2009-08-11 05:08:53	iPhone: 35.509506,139.615601	揺れたね	Hakkan	http://twitte
2009-08-11 05:08:53	Mie Prefecture	すごい地震だ [mb]	narude531masu	http://twitte
2009-08-11 05:08:52	Kawasaki city	地震だ!!	yakefasamma	http://twitte
2009-08-11 05:08:52	unknown	地震こわいですかんべん	wzzc	http://twitte
2009-08-11 05:08:52	Kansai	あら、地震?	Haru_ro	http://twitte
2009-08-11 05:08:52	Sakado, Saitama, Japan	地震だ	d_wackys	http://twitte
2009-08-11 05:08:51	unknown	愛知も揺れたw	edomain	http://twitte
2009-08-11 05:08:51	unknown	また地震 長いな…	laukaz	http://twitte
2009-08-11 05:08:51	JP	地震なう	echomum	http://twitte



Disaster phase detection

- Tweet classification and visualization for disaster phase detection (PhaseVis [Yang et al., ISCRAM13])



Urban computing

- Many problems and challenges in big cities

Q1: Smart grid

Q2: Urban flow

Q3: Situation awareness

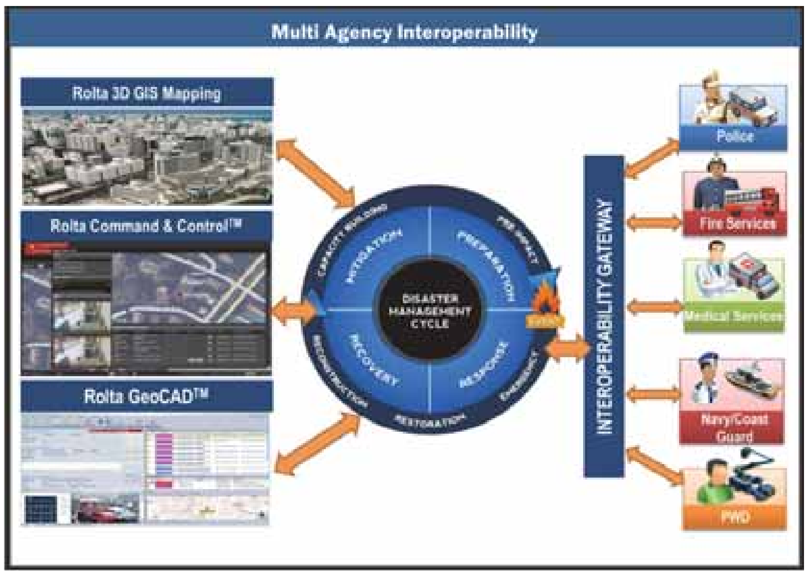
Q4: Robustness & Evolution

Q5: Public Health

Q6: Air pollution

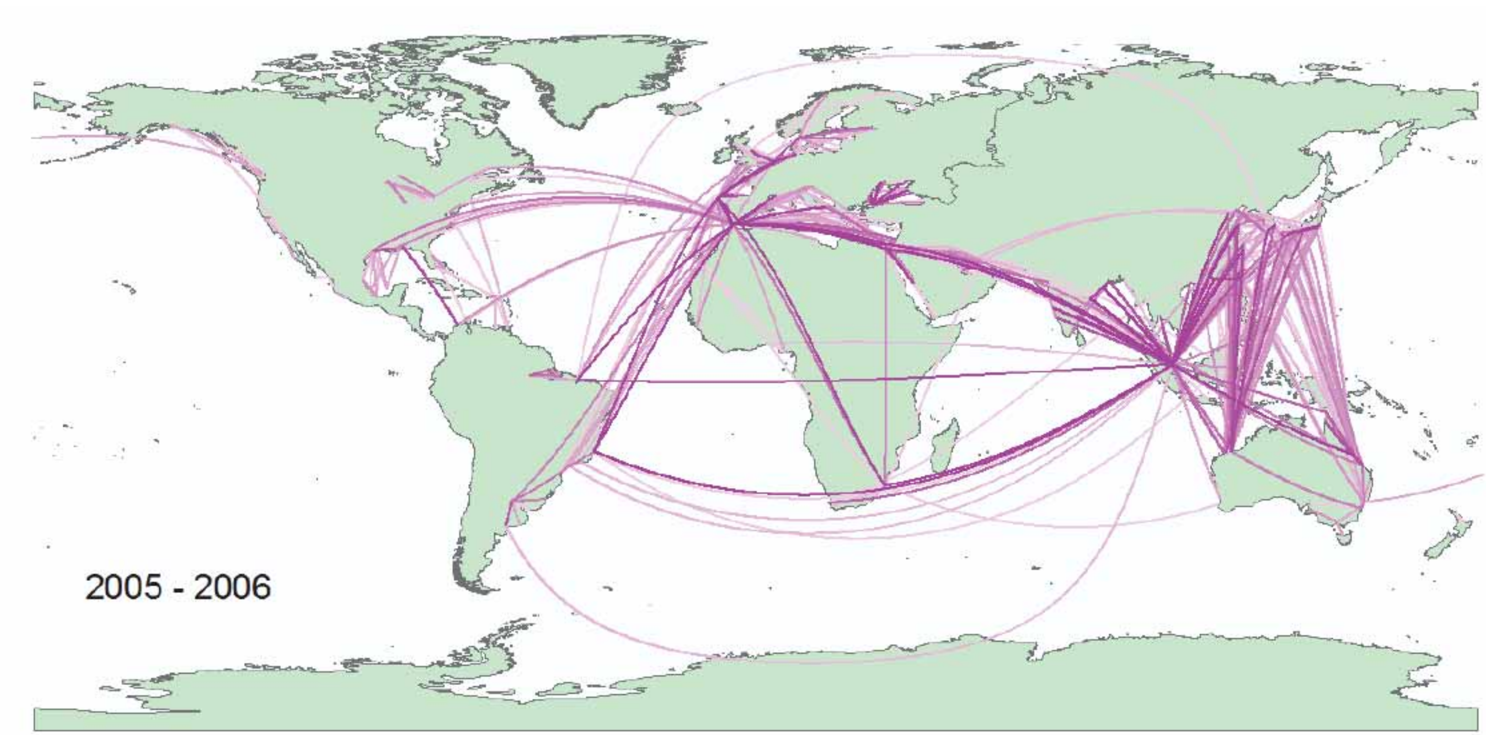
Q4: Robustness & Evolution

- Improving system robustness and modeling system evolutions
  - Failure simulation and prediction system



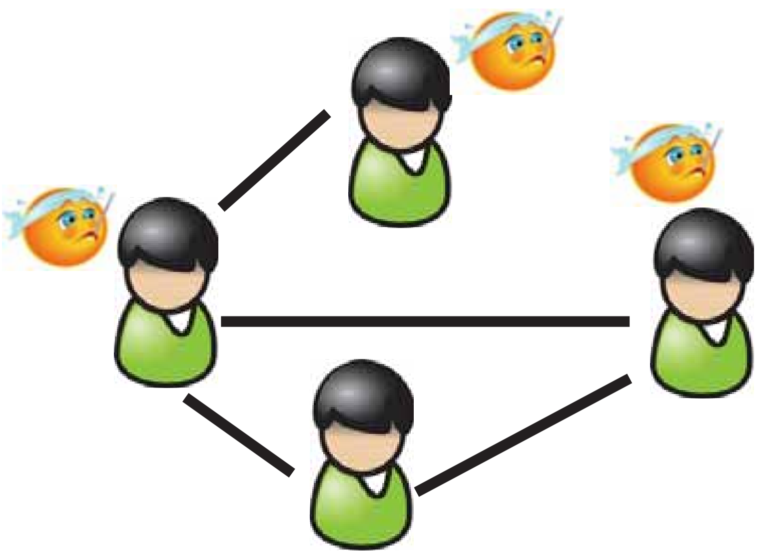
VirginiaTech

Analyzing species flow and invasion risk [Xu et al., KDD'14]

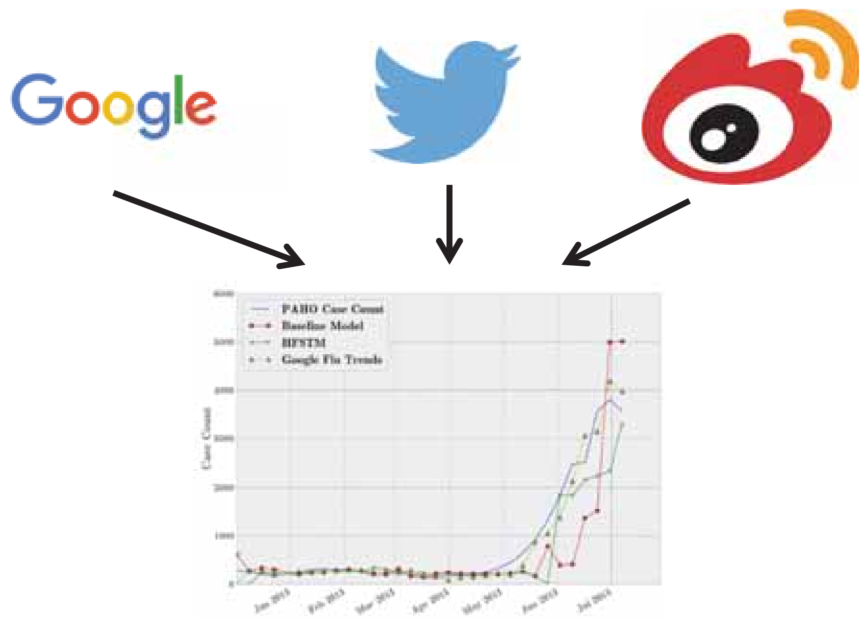


Q5: Public health

- To assure the condition in which people can be healthy
  - Immunization/vaccination
  - Health surveillance



Controlling disease spread over a contact network



Syndromic surveillance of flu

Q6: Air pollution

- Infer real-time and fine-grained air quality throughout a city using Big Data (U-Air [Zheng et al., KDD'13])

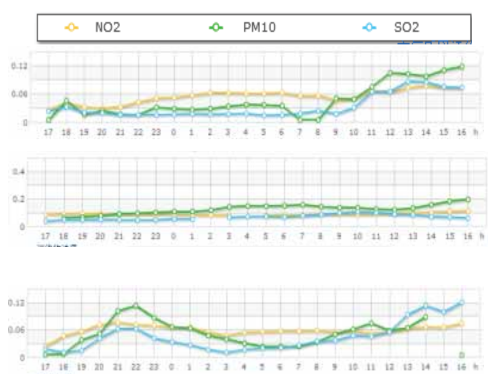
Meteorology

Traffic

Human Mobility

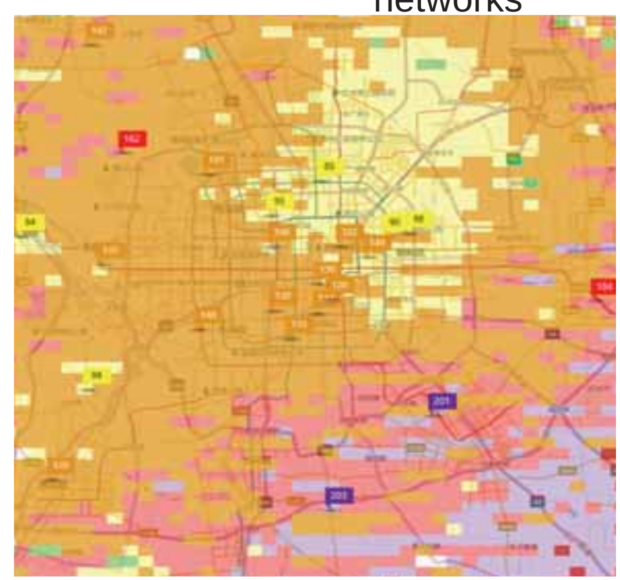
POIs

Road networks



Historical air quality data

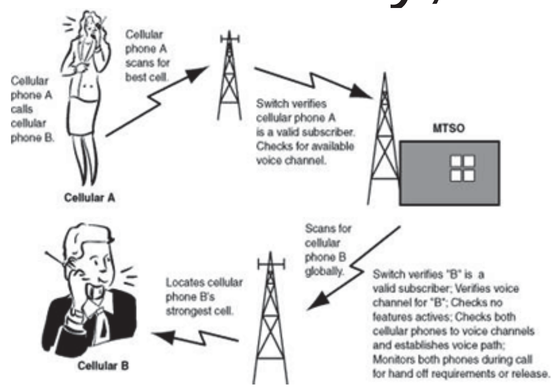
Real-time air quality reports



Critical Infrastructure Systems → Urban Computing

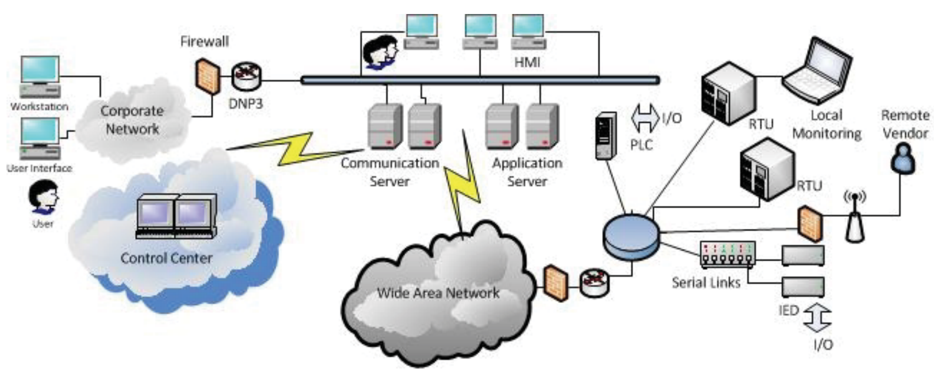
- Vital to our national security, economy.

Transportation System

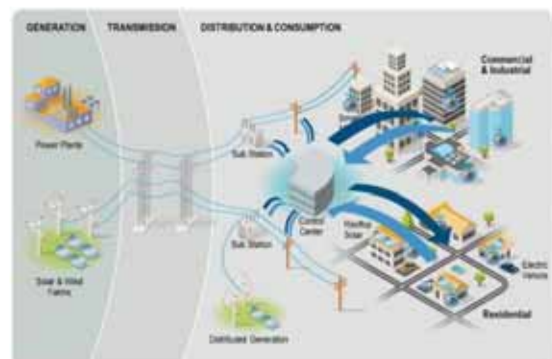


Cellular System

Water System



Cyber System



Electric Grid System

VirginiaTech

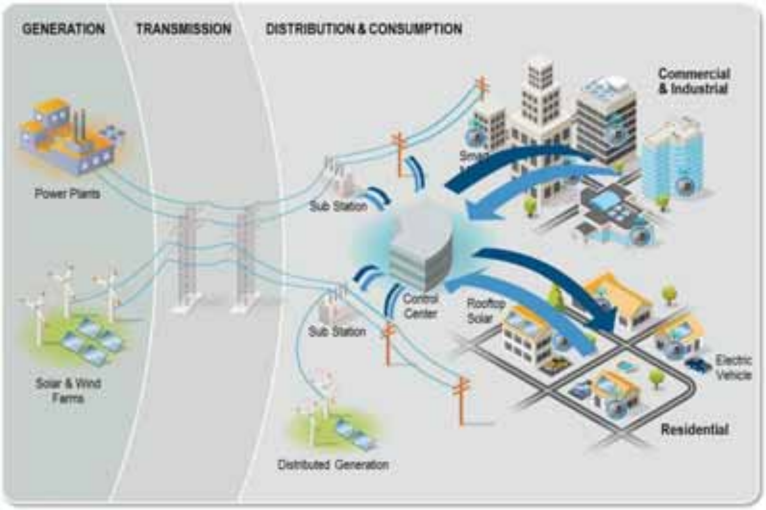
Importance of CIS

- CIS are the fundamental for many of the urban computing problems

Q5 (public health)

Q2 (urban flow)

Q3 (situation awareness)



Electric grid system

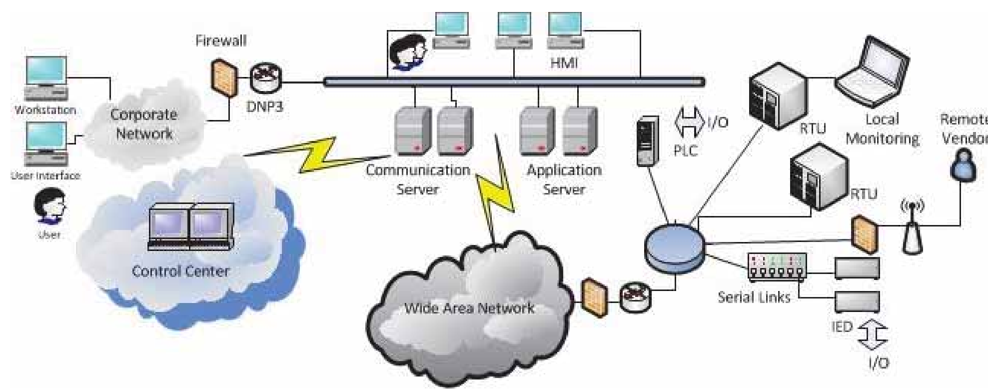
Provide power to →

Importance of CIS

- CIS are the fundamental for many of the urban computing problems

Traffic flow analysis based on the transportation system

Q2 (urban flow)

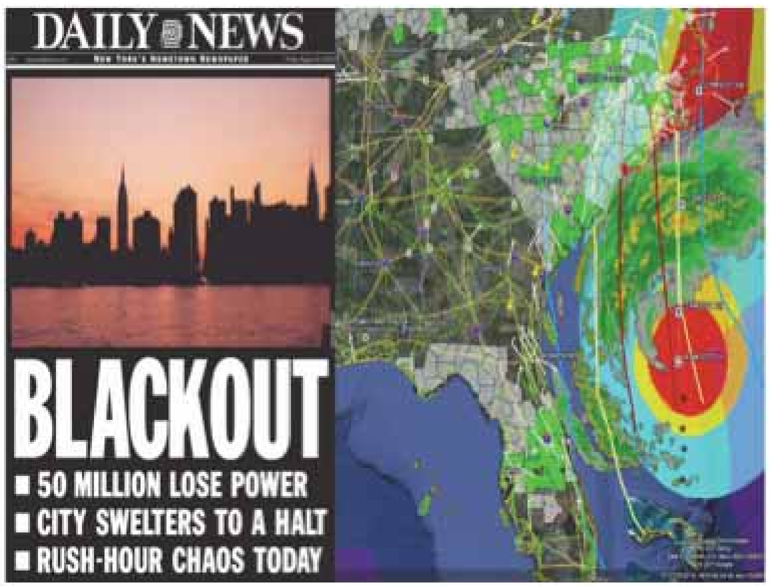


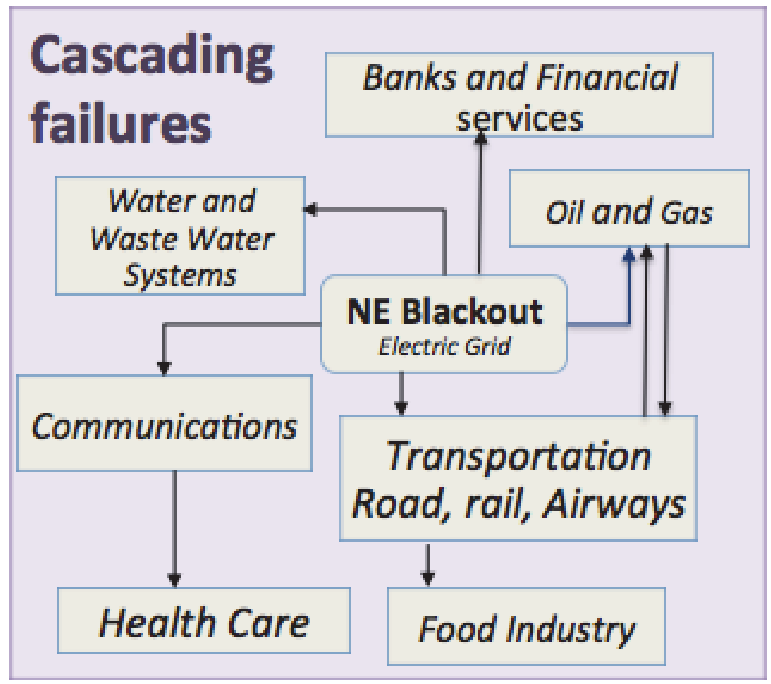
Cyber system facilitates online communications

Q1 (smart grid)

Ex: 2003 Northeast Blackout

- Cascading failures on CIS lead to huge loss in different areas





Data mining problems

- What are the underlying data mining problems in CIS?

Aim 1: System modeling

- Modeling the regular and anomalous dynamics of the system

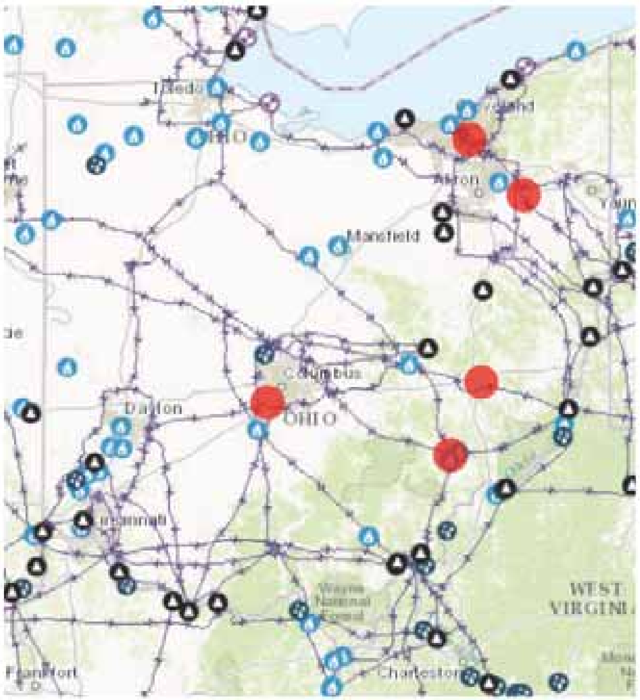
Traffic flow modeling

System modeling during a hurricane

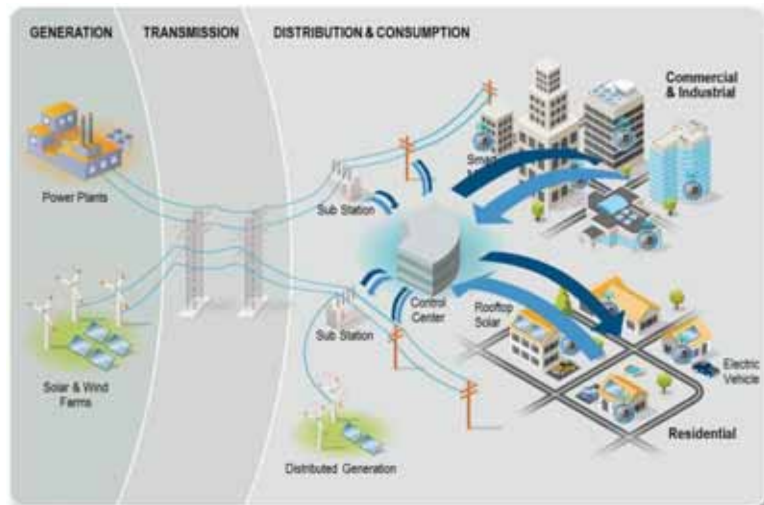
VirginiaTech

Aim 2: Vulnerability analysis

- Analyze system vulnerability and resilience



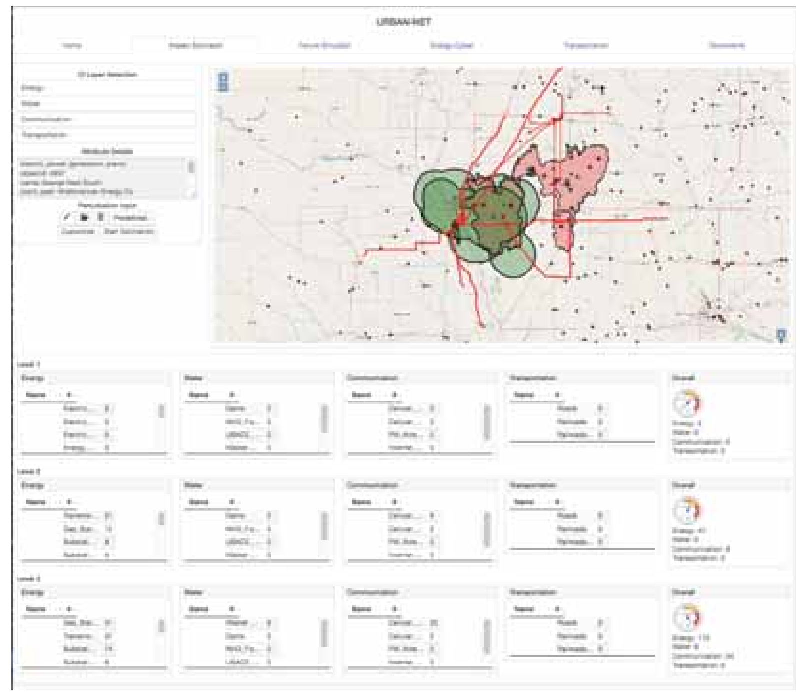
Identifying critical facilities in the system



Quantifying system vulnerability

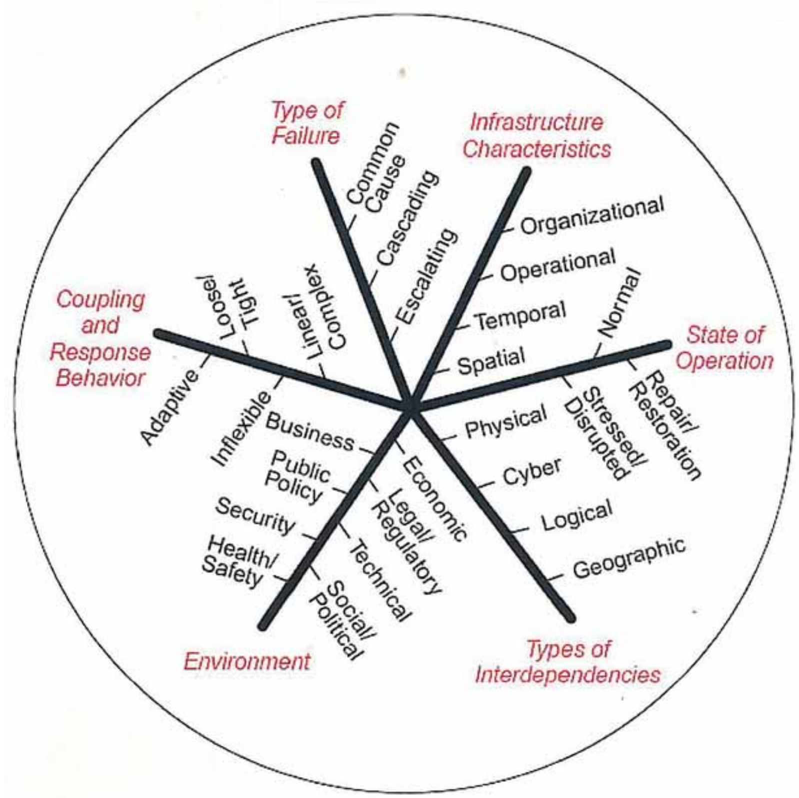
Aim 3: Supporting tools & systems

- Tools & systems to facilitate decision makings



Challenges in working with CIS

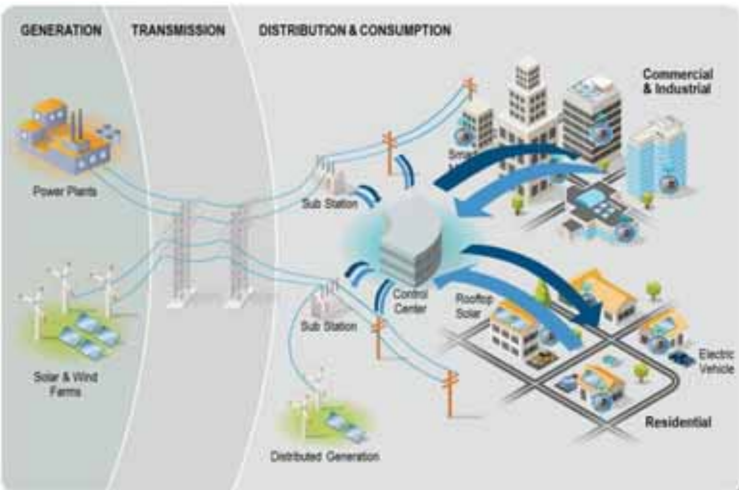
- Why are CIS hard to work with:
  - 1. Complexity
    - Hierarchy of subsystems
  - 2. Heterogeneity
    - Types of interdependencies
  - 3. Dynamics
    - Different types of failures
    - State of operation
    - Coupling behavior



Prakash 2019
VirginiaTech

Challenge 1: Complexity

- Many underlying subsystems



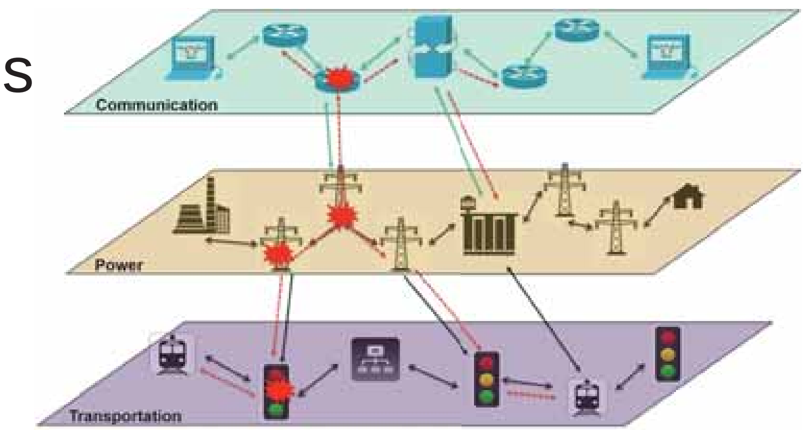
Electric Grid System

- Power generator: generates power using different types of fuel.
- Transmission network: transfer power to different areas.
- Distribution plant: distribute power to local facilities.
- Pipeline network: transfer resources such as natural gas

Even more subsystems inside these subsystems (like natural gas compressors inside the pipeline network)

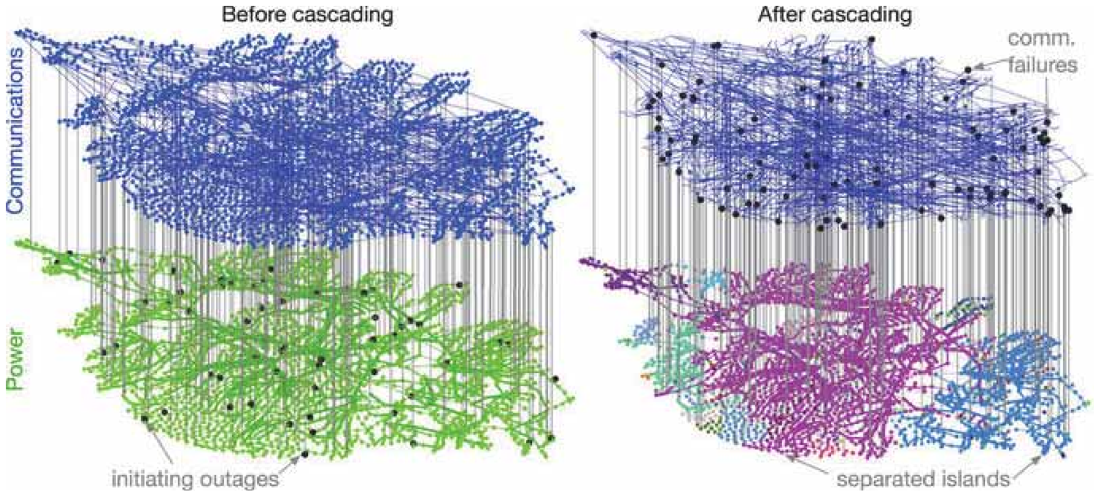
Challenge 2: Heterogeneity

- Different types of interdependencies
  - Physical: the state of an infrastructure depends on the material output of another
  - Geographical: changes caused by local environmental events
  - Cyber: the state of an infrastructure depends on the information transmitted through the information infrastructure
  - Logical: other dependencies



Challenge 3: Dynamics

- Different types of incidents can cause the failure of a facility
  - Loss of dependencies (power, fuel, etc)
  - Malfunctioning (due to natural or man-made disasters)
- The system has different states of operations
  - Normal, repair, stressed



Prakash 2019

VirginiaTech

From a Data Analytics viewpoint...

- Highly **heterogeneous** data. E.g.:
  - Networks with many different types of nodes and links.
  - Multiple networks
  - Multiple sources of information
- Complex system dynamics
  - **Temporal behavior** is important
- **Large scale** big-data for analysis.
- Require **actionable results** and intelligent systems.

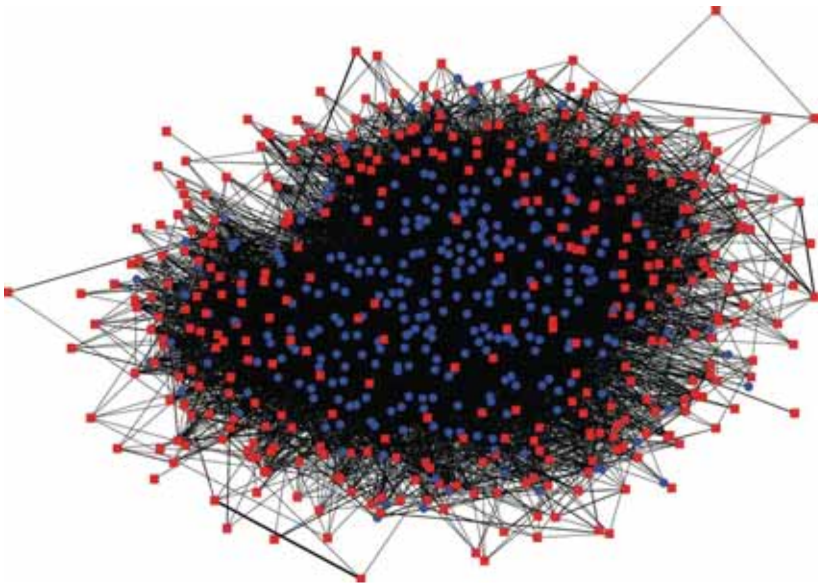
Four V's of Big-data: Variety, Volume, Velocity, Veracity

Modeling CIS

- How to model the dynamics of the system? [Ouyang'13]
  - System dynamics based
  - Agent based
  - Network based
  - Empirical
  - Economic theory based

Network based approaches

- Describe CIS as networks with nodes representing different CIS components and links mimicking the physical and relational connections among them

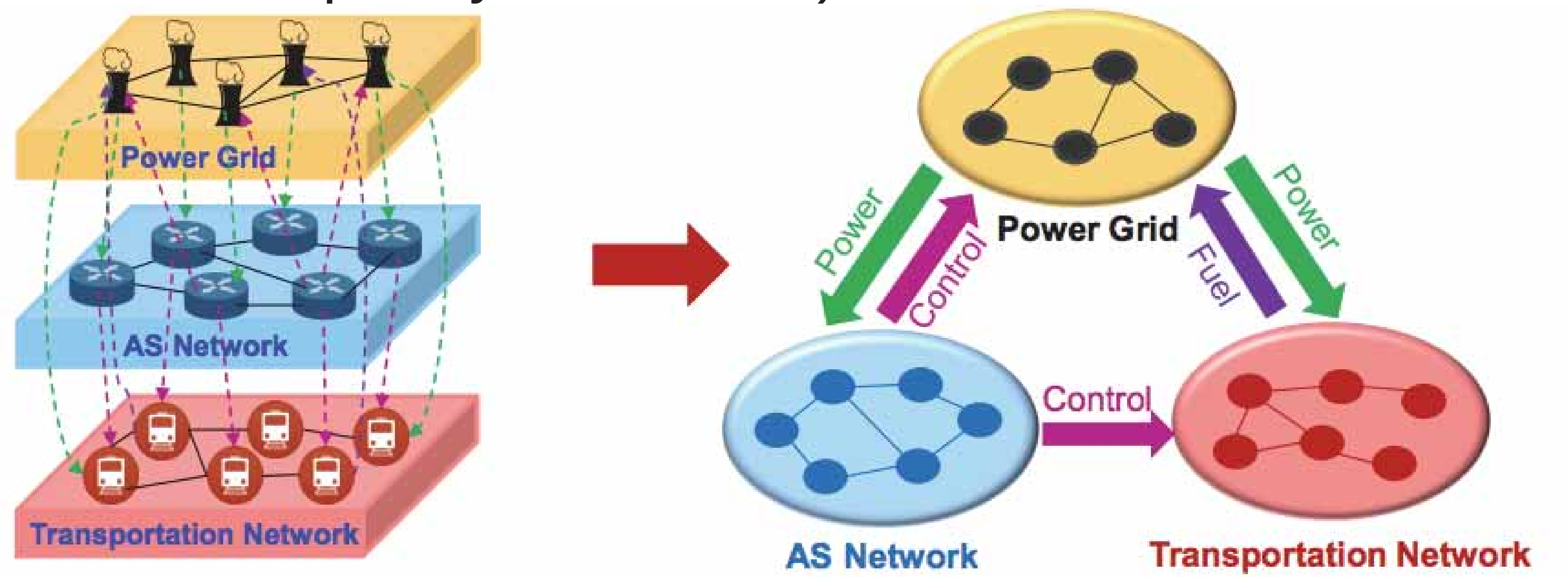


- Less ‘realistic’: require less domain knowledge
- More general: can be applied for different systems

Infer across-layer connections

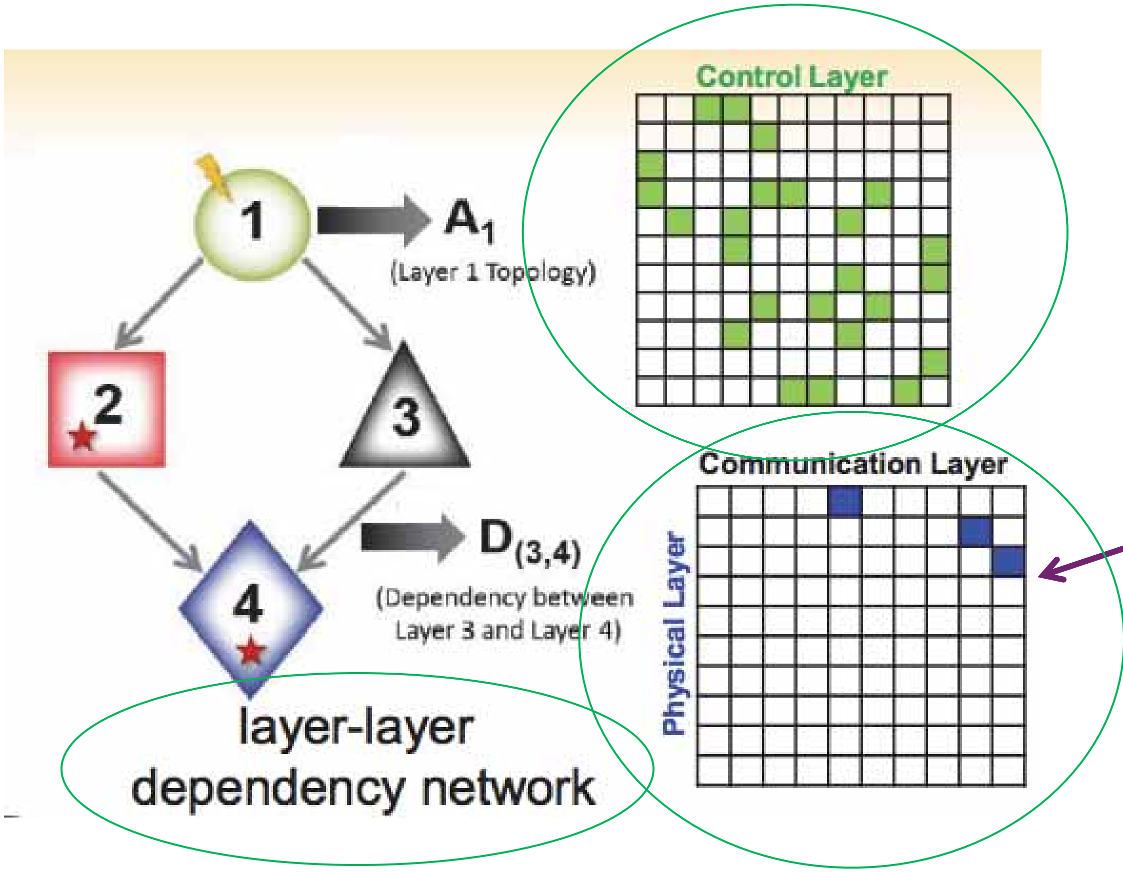
[Chen, KDD'16]

- Different CIS layers can be connected due to their complex interdependencies (not completely observable)



Some slide materials taken from: http://www.public.asu.edu/~cchen211/FASCINATE_KDD.pdf

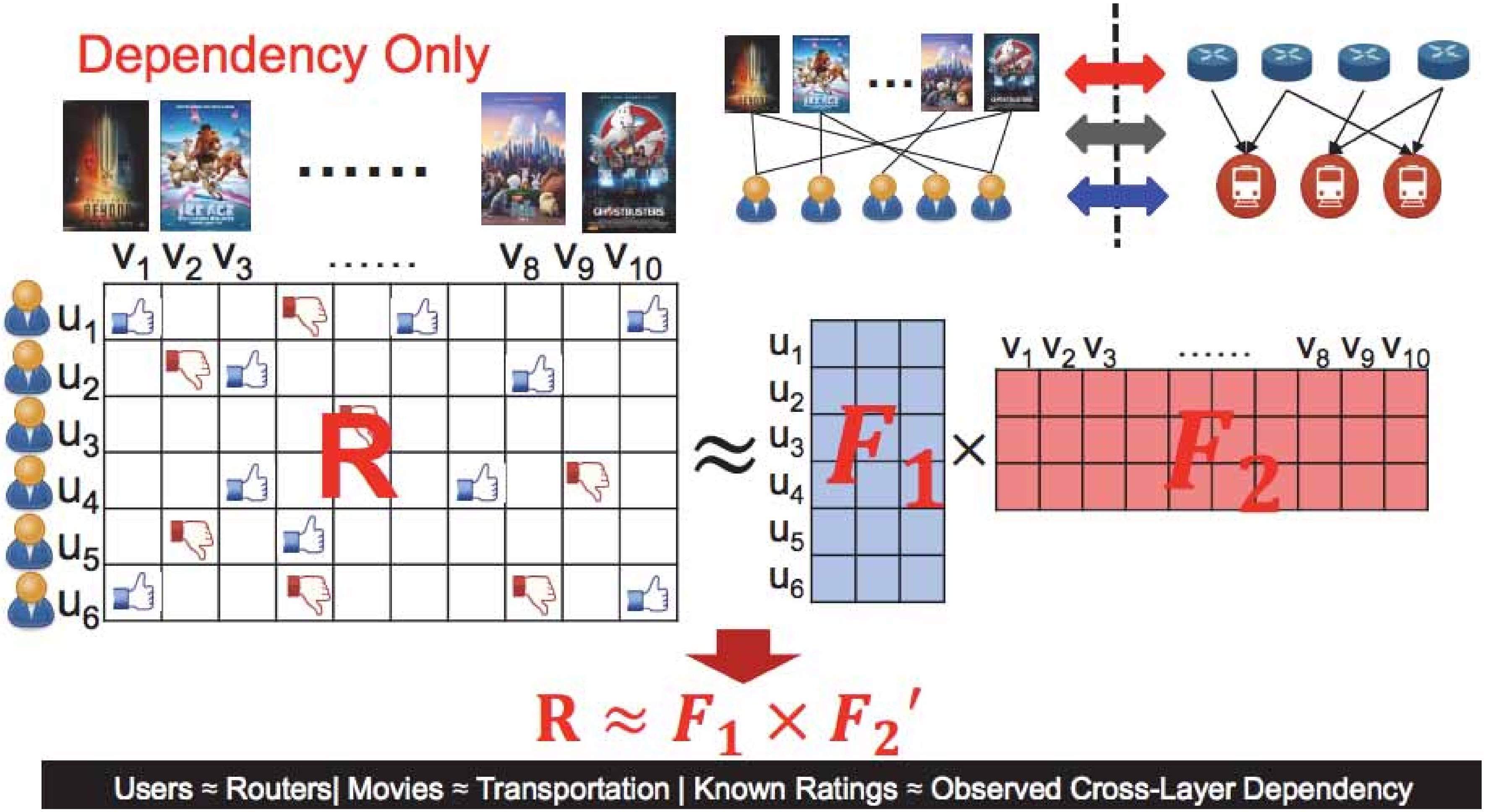
Infer across-layer connections



Infer the incomplete across-layer relations

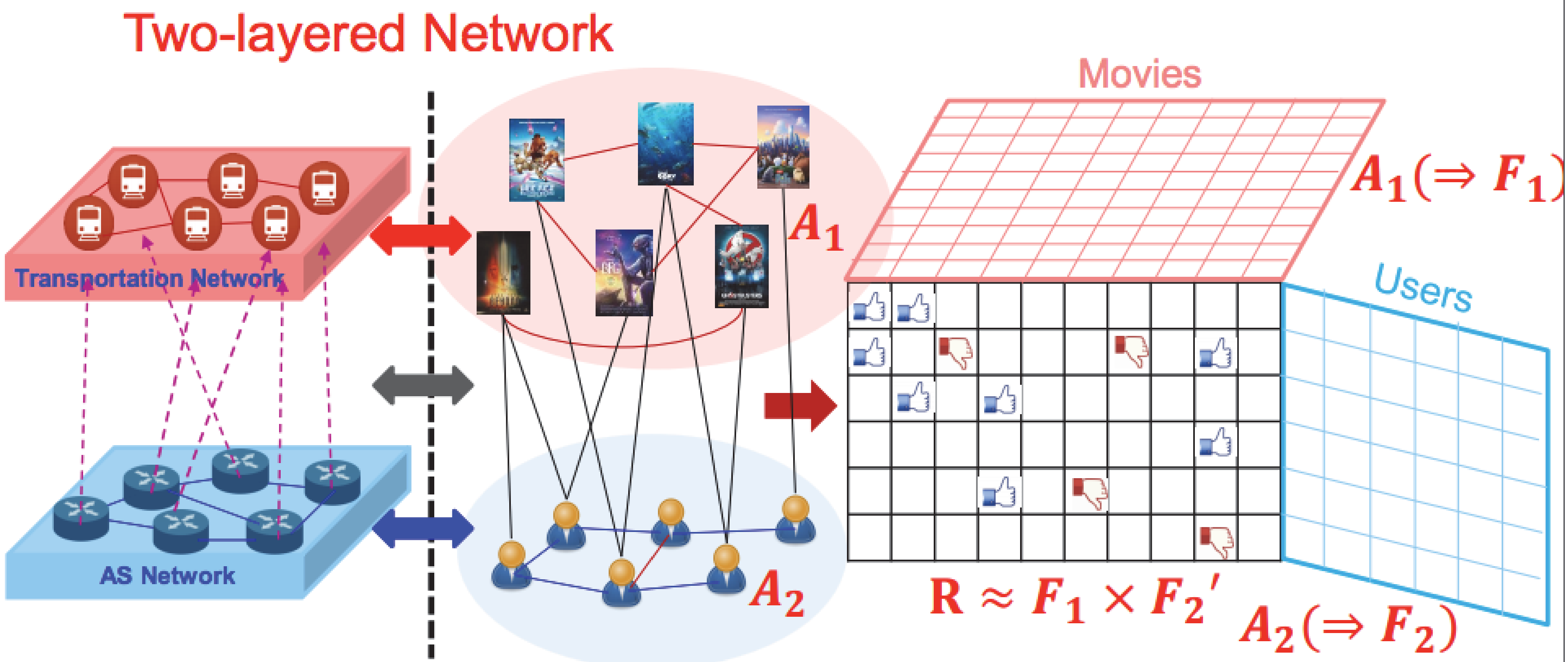
Key idea 1

- Collaborative filtering



Key idea 2

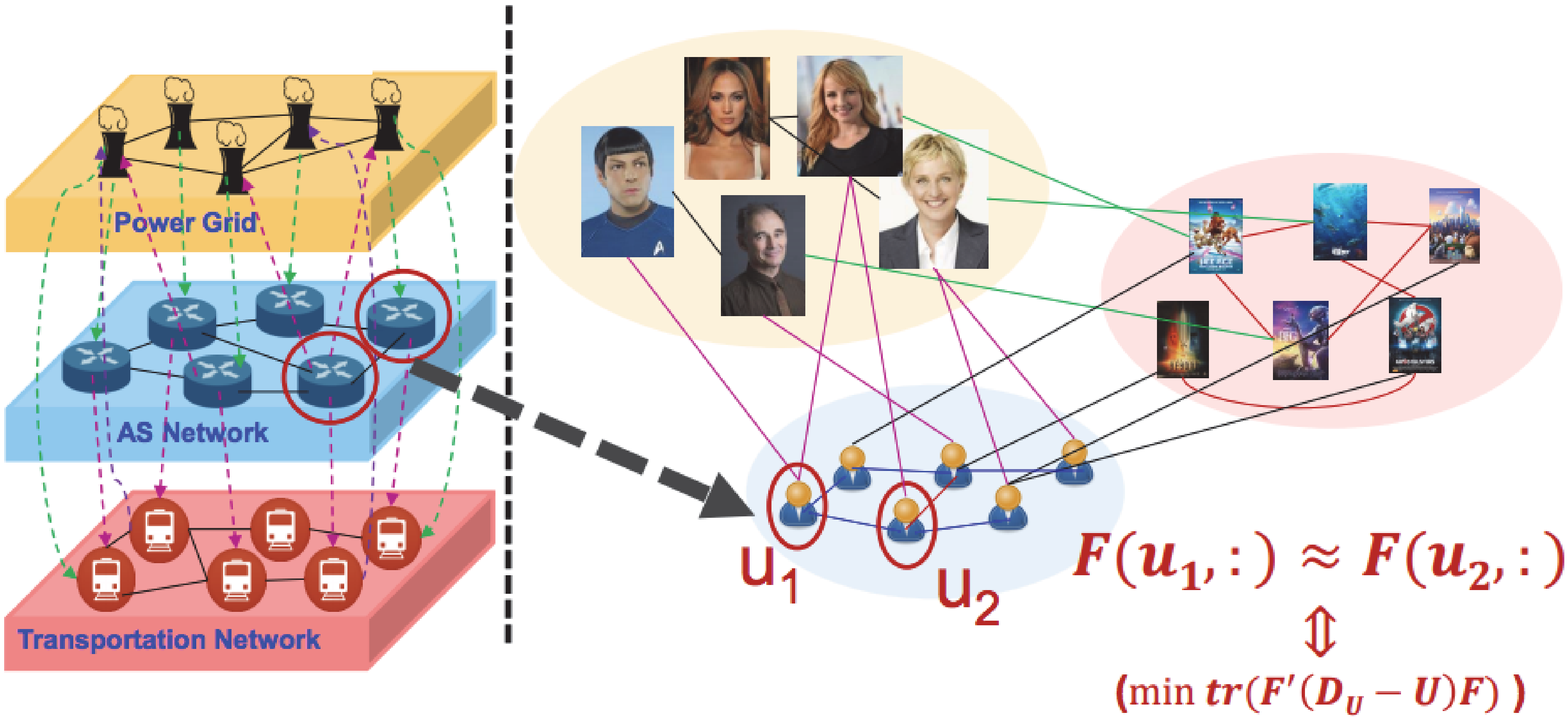
- Collaborative filtering with side information



Movie-Movie Similarity ≈ Transportation Network | Social Network ≈ AS Network
Known Ratings ≈ Support from Routers to Transportation Network

Key idea 3

- Node homophily: closely connected entities tend to have similar latent profiles



Celebrities ≈ Power Plants | Users ≈ Routers | Movies ≈ Transportation
Known Ratings, Movie Cast, Fans ≈ Observed Cross-Layer Dependencies

FASCINATE

- A collaborative filtering based optimization algorithm

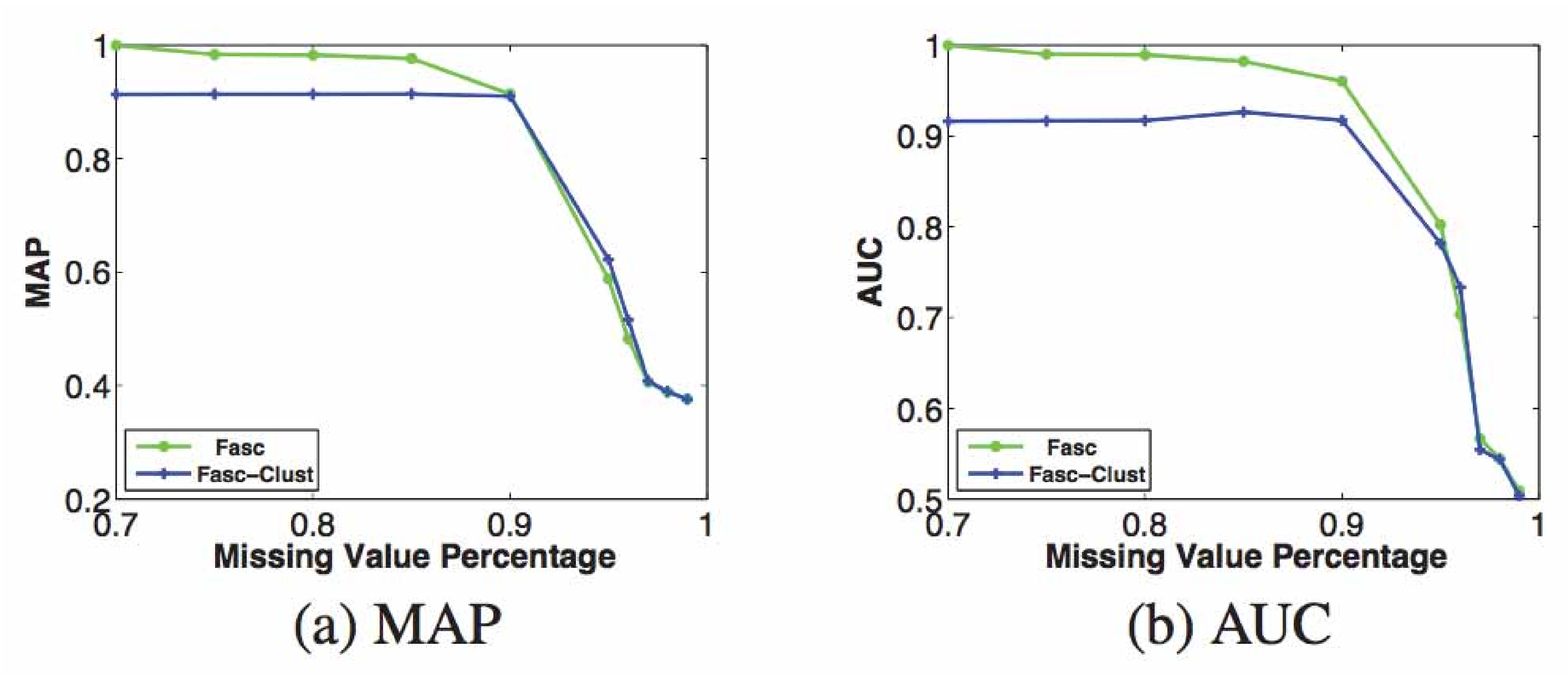
$$\min_{F_i \geq 0 (i=1,\ldots,g)} J = \underbrace{\sum_{i,j:G(i,j)=1} \| W_{i,j} \odot (D_{i,j} - F_i F_j') \|_F^2}_{\text{Matching observed cross-layer dependencies}} +$$

$$\alpha \sum_i \underbrace{tr(F_i'(T_i - A_i)F_i)}_{\text{Node homophily}} + \beta \sum_i \underbrace{\| F_i \|_F^2}_{\text{Regularization}}$$

VirginiaTech

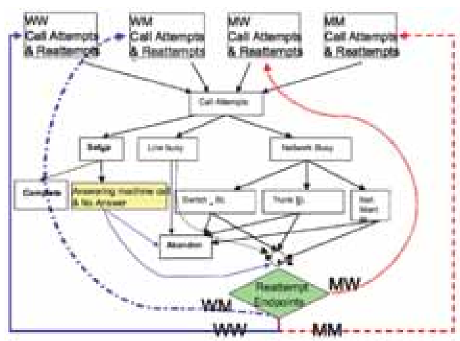
Performance

- Achieving good performance in inferring the across-layer dependencies



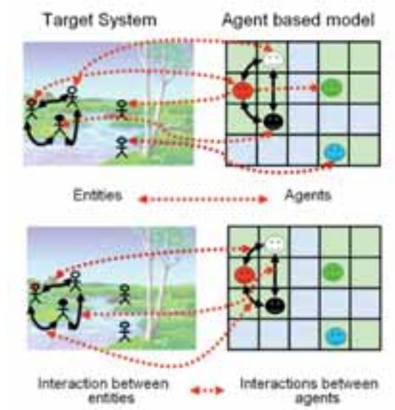
Modeling CIS

System dynamics based



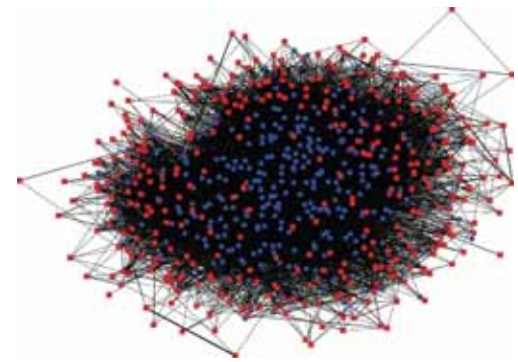
- Good for realistic and precise modeling of a single system
- Typically require domain knowledge
- Can be time consuming

Agent based



- Assumptions for agents' behaviors/policies
- Very system specific
- Useful for testing policies or strategies

Network based



- Less realistic modeling
- Require less domain knowledge
- Can work for multiple systems together
- More general

We will focus more on the network based methods

VirginiaTech

Outline

- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- Protecting CIS against attacks

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- Protecting CIS against attacks

Find facilities to protect/enhance against unknown natural disasters (non-adversarial)

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- Protecting CIS against attacks

Using mainly the static topology/structure of the network to find crucial nodes

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- Protecting CIS against attacks

Integrating the failure cascade dynamic into the analysis

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- Protecting CIS against attacks

Protect the system against adversarial attacks with known patterns and strategies

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - **Network structure based**
  - Dynamics based
- Protecting CIS against attacks

Power grid resilience metric

[Arianos, Chaos'09]

- Power transmission grid network
  - Transmission network connecting power generators and load nodes

Power grid resilience metric

[Arianos, Chaos'09]

- Generalize the geodesic (shortest) distance to account for the flow capacity between power grid nodes

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$

Global efficiency score

- Power does not flow from one node to another in a single path
- The power only flows from the generator nodes to the load nodes
- Capacity C_{ij} from generator i to load j (the maximum power injection)

Net-ability

- Generalize the geodesic (shortest) distance to account for the flow capacity between power grid nodes

Adapt to power grid

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \longrightarrow A = \frac{1}{N_{\mathcal{G}} N_{\mathcal{D}}} \sum_{i \in \mathcal{G}} \sum_{j \in \mathcal{D}} C_{ij} \sum_{k \in \mathcal{H}_{ij}} p_{ij}^{k} \frac{1}{d_{ij}^{k}}$$

Global efficiency score

Generator nodes

Load nodes

Power share of paths

Net-ability

- Generalize the geodesic (shortest) distance to account for the flow capacity between power grid nodes

Adapt to power grid

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \longrightarrow A = \frac{1}{N_G N_D} \sum_{i \in \mathcal{G}} \sum_{j \in \mathcal{D}} C_{ij} \sum_{k \in \mathcal{H}_{ij}} p_{ij}^k \frac{1}{d_{ij}^k}$$

Global efficiency score

Distance based on economic and technical cost

$$d_{ij}^k = \sum_{l \in k} f_k^l Z_l$$

Impedance

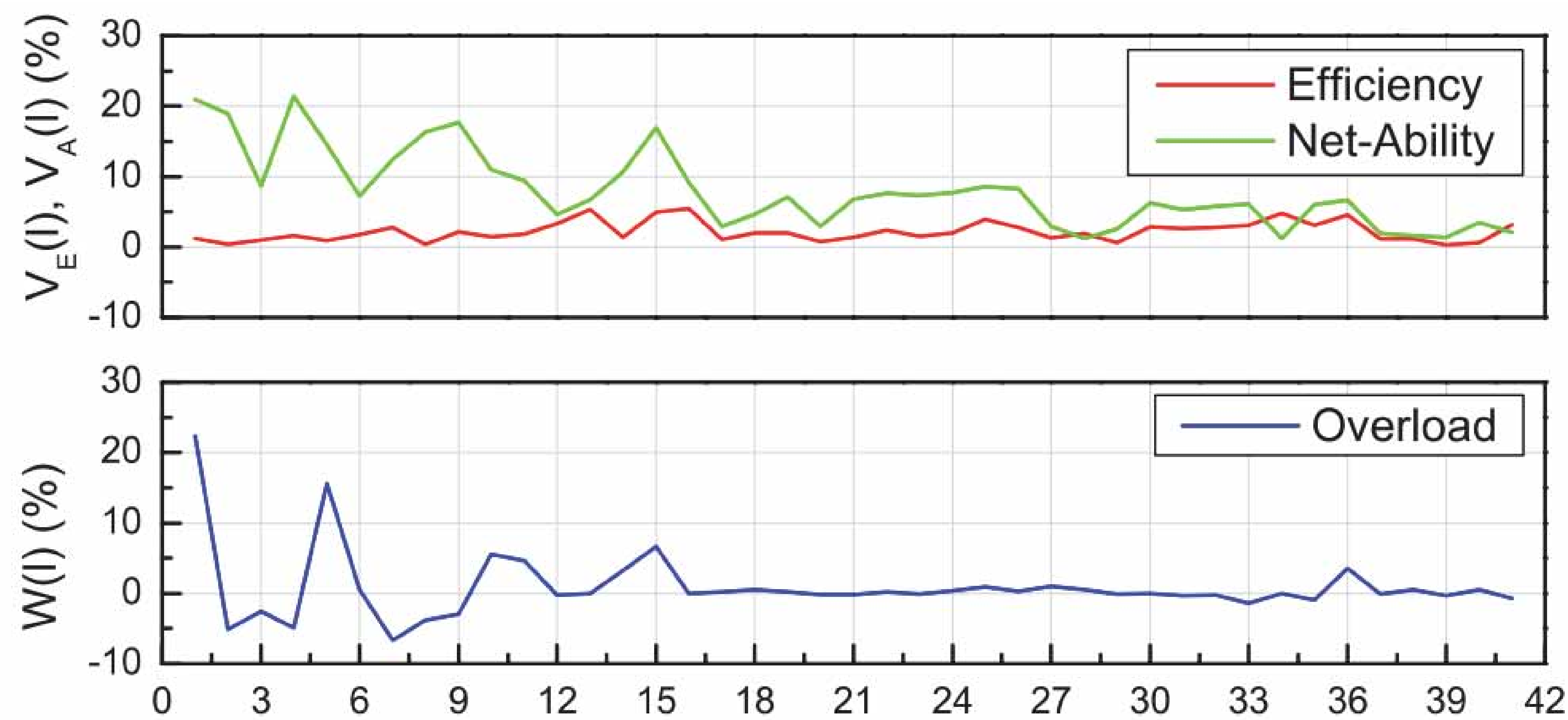
Power distribution factor

Prakash 2019

VirginiaTech

Global performance evaluation

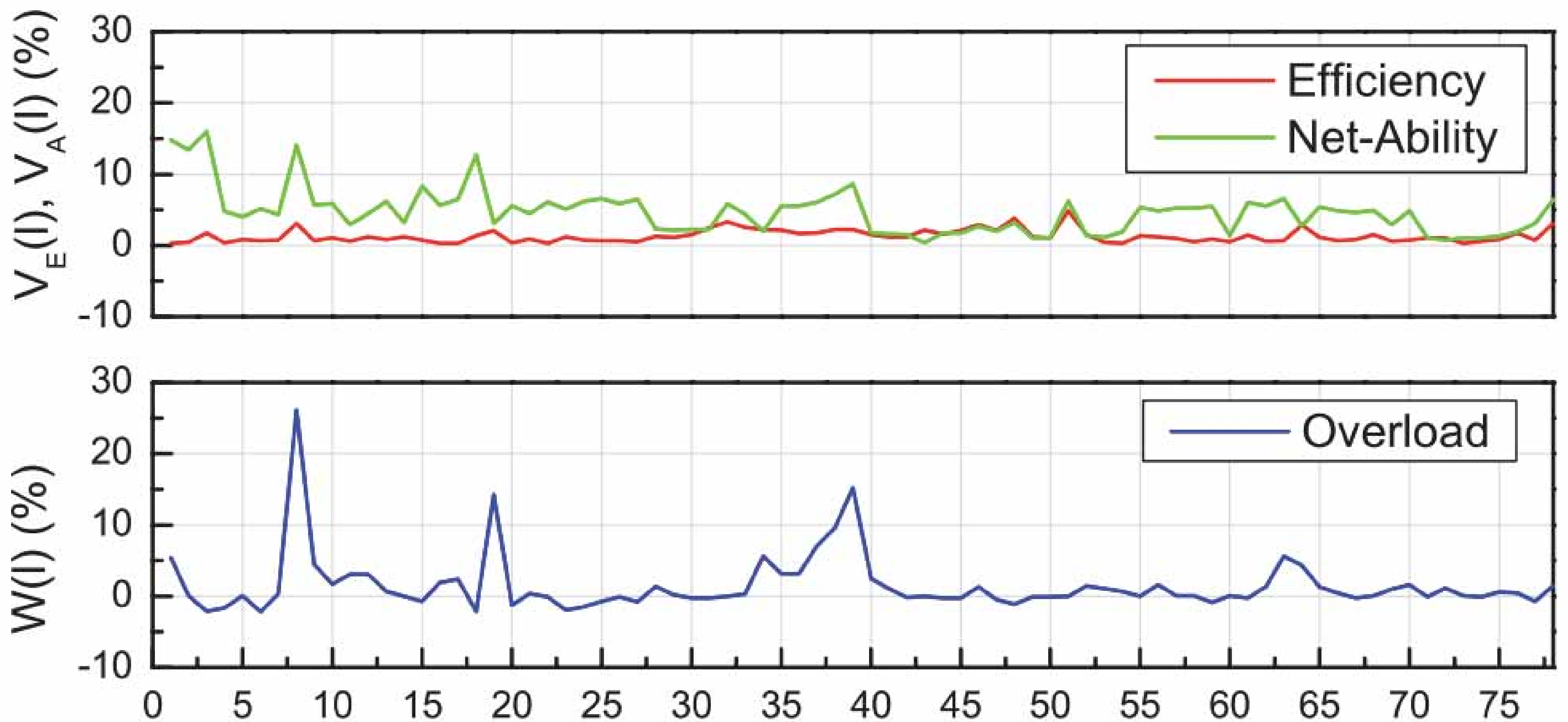
- Comparing with overload rate upon line removal



VirginiaTech

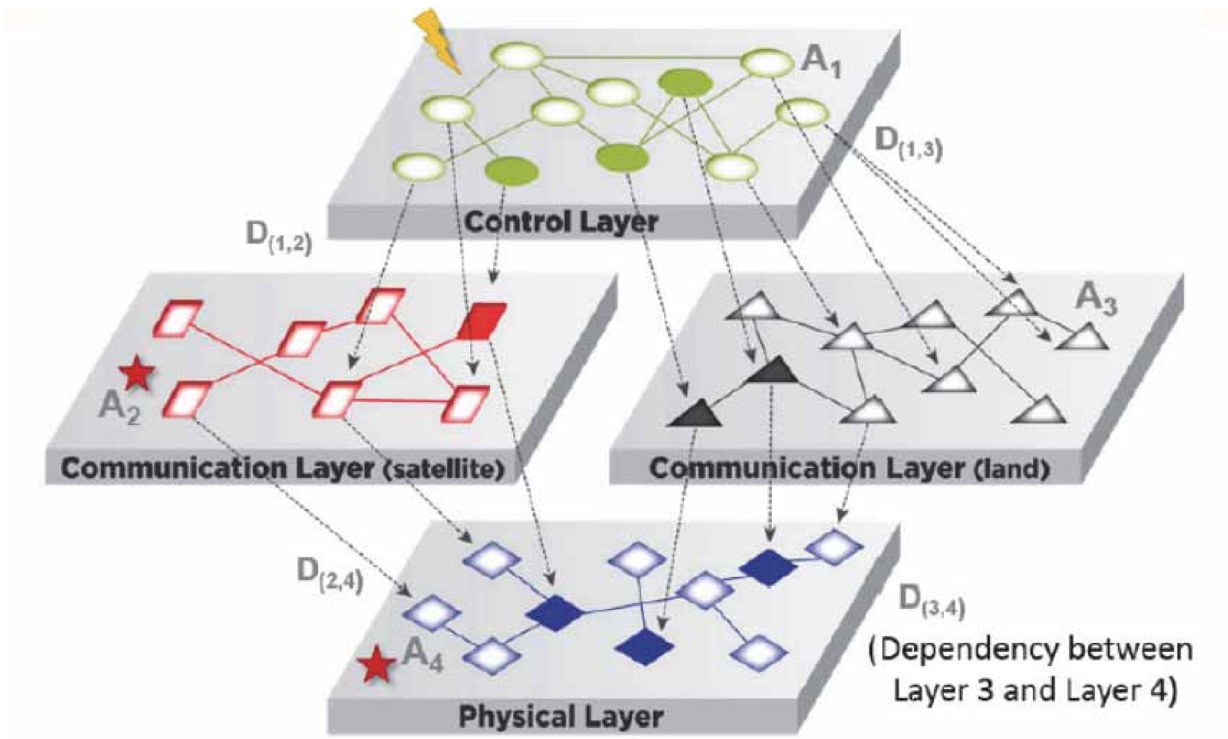
Global performance evaluation

- Comparing with overload rate upon line removal



Connectivity measures on multi-layered networks [Chen, ICDM'15]

- Multi-layered networks



A four-layered network

Some slide materials taken from: http://www.public.asu.edu/~cchen211/ICDM15_Mulan.pdf

Connectivity unification (SUBLINE family)

- Key idea: graph connectivity as an aggregation over the subgraph connectivity

$$C(\mathbf{A}) = \sum_{\pi \subseteq \mathbf{A}} f(\pi)$$

  - A: adjacency matrix of the graph
  - π: a non-empty subgraph in A
  - $f(\pi)$: connectivity of the subgraph π
  - $C(A)$: connectivity of graph A

Connectivity unification (SUBLINE family)

- Key idea: graph connectivity as an aggregation over the subgraph connectivity

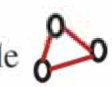
$$C(\mathbf{A}) = \sum_{\pi \subseteq \mathbf{A}} f(\pi)$$

Examples

- Path Capacity: $f(\pi) = \begin{cases} \beta^{len(\pi)} & \text{if } \pi \text{ is a valid path of length } len(\pi) \\ 0 & \text{otherwise.} \end{cases}$

- Loop Capacity: $f(\pi) = \begin{cases} 1/len(\pi)! & \text{if } \pi \text{ is a valid loop of length } len(\pi) \\ 0 & \text{otherwise.} \end{cases}$

- Triangle Capacity: $f(\pi) = \begin{cases} 1 & \text{if } \pi \text{ is a triangle} \\ 0 & \text{otherwise.} \end{cases}$



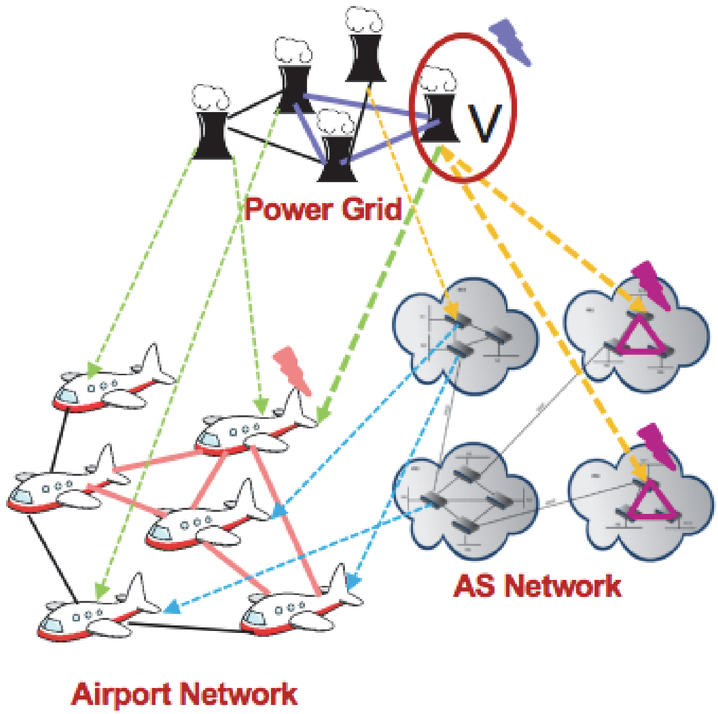
- …

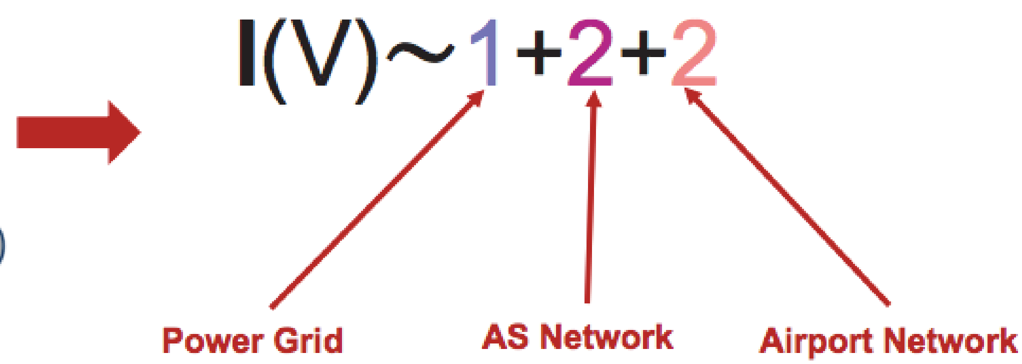
VirginiaTech

Connectivity control

- Define $\mathbf{I}(S_i) = \sum_{i=1}^{g} \alpha_j (C(A_j) - C(A_j \backslash S_{i \to j}))$
  - Example: C = Triangle Capacity

I(V) ~ #Triangles in which A & its dependencies participate





Optimal control

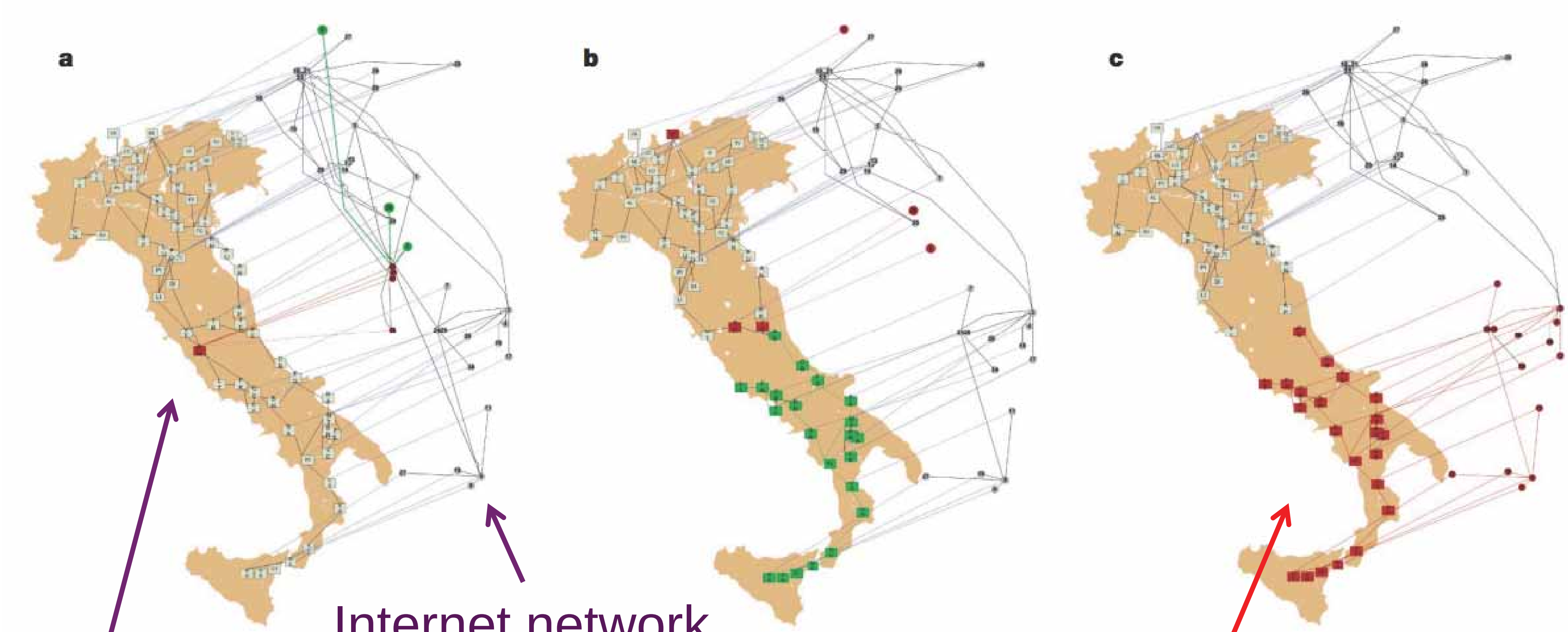
- Goal
  - Find an optimal node set in the control layer to maximize its impact on the target layers
- Theorem
  - The SUBLINE family enjoy the **diminishing returns property**
- Solutions (**OPERA**)
  - Greedy algorithm (linear)

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - **Dynamics based**
- Protecting CIS against attacks

Robustness under failure cascade [Buldyrev, Nature'10]

- Failure cascade between different layers



Internet network

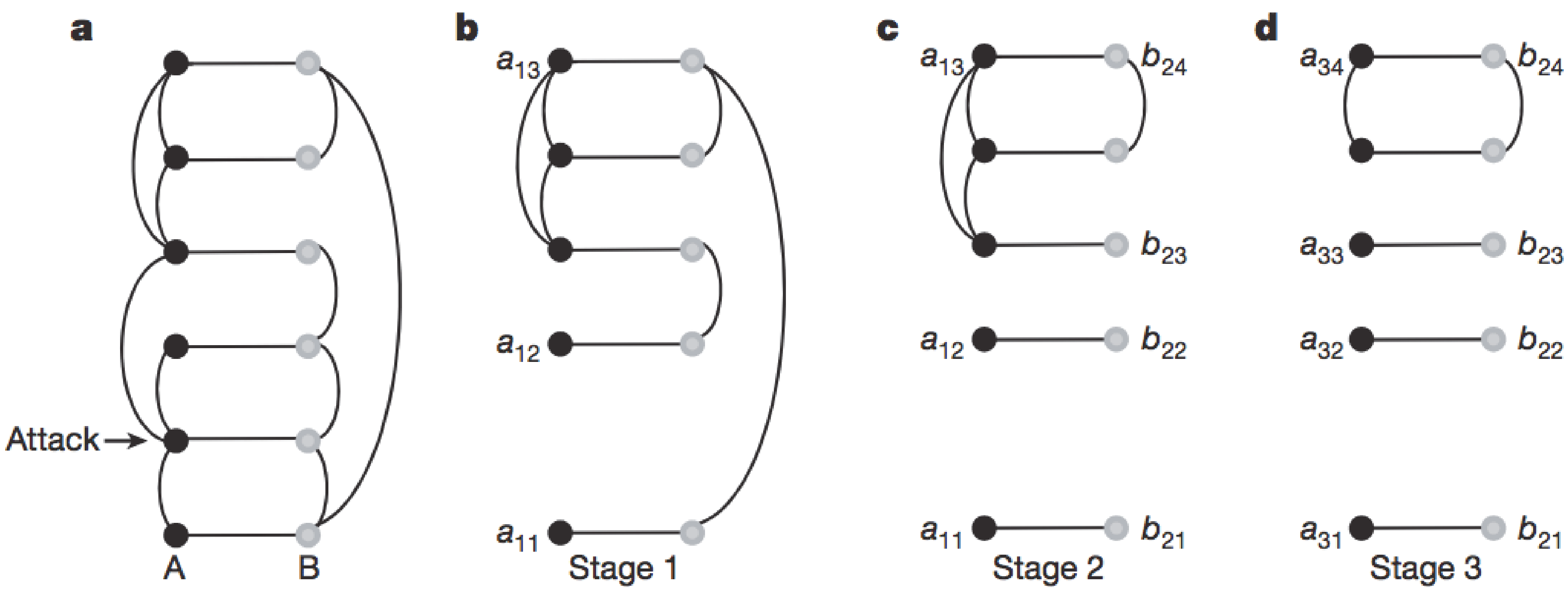
Power network

Disconnection from the giant cluster

VirginiaTech

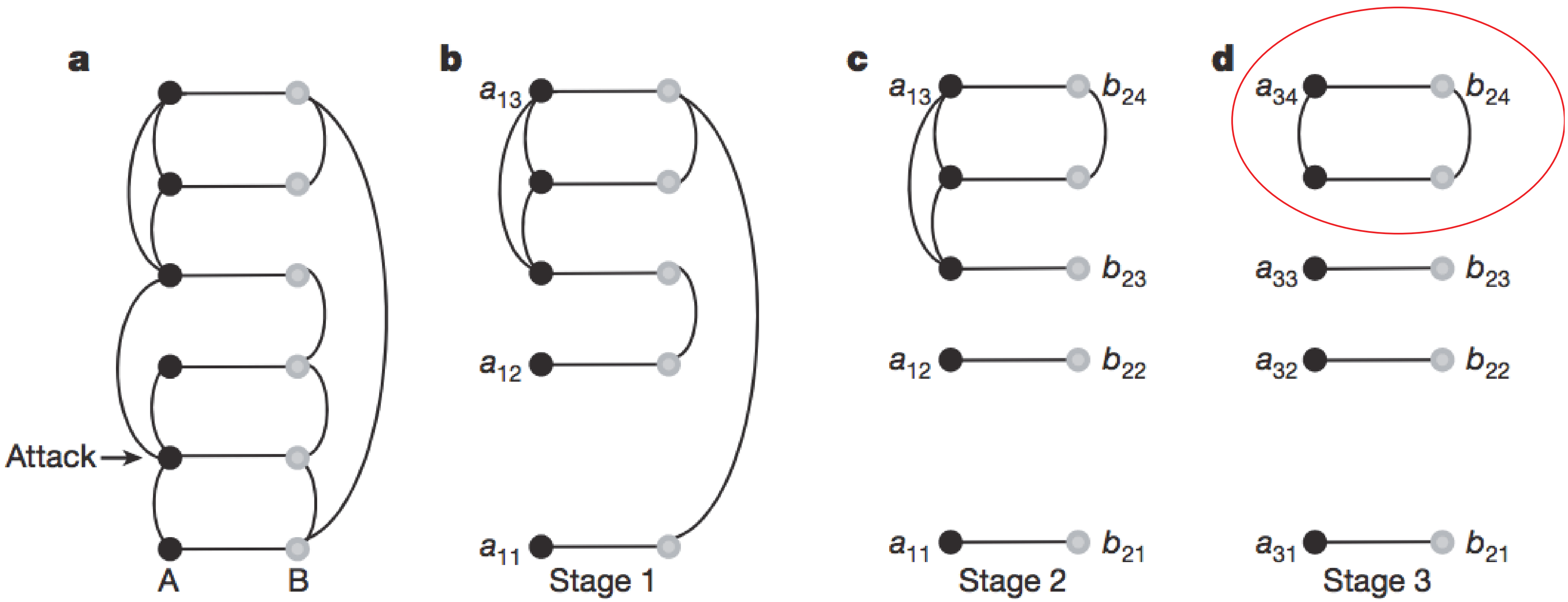
Failure cascade based on mutually connected clusters

- Only the mutually connected clusters are functional



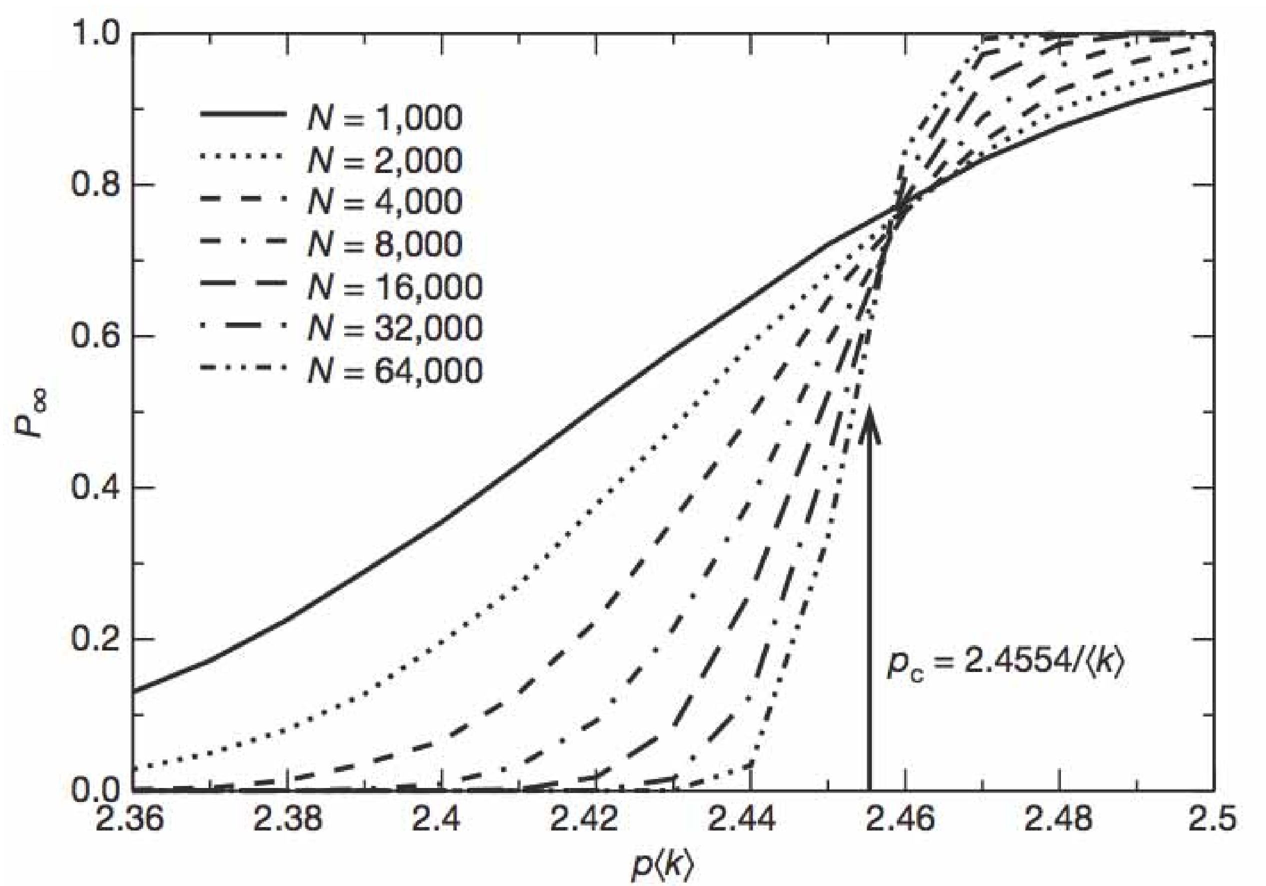
Failure cascade based on mutually connected clusters

- At the end, we study the size of the giant (largest) mutually connected clusters.



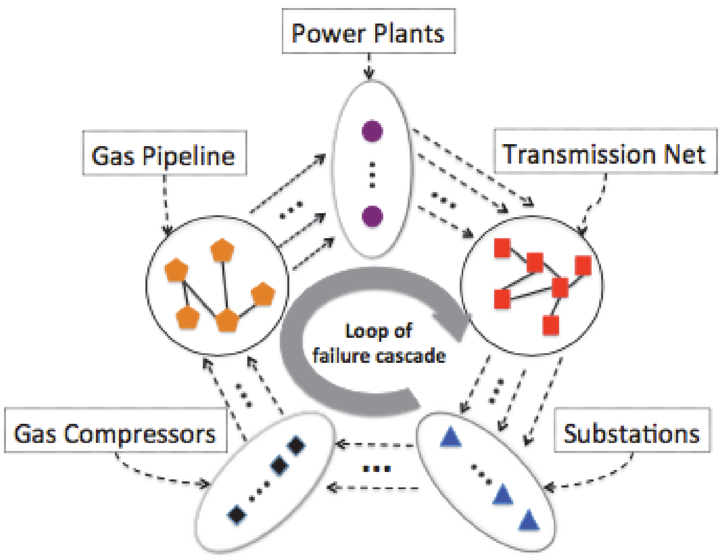
Analysis on ER networks

- A critical threshold p_c to maintain a giant mutually connected cluster at the end.



HotSpots [Chen, CIKM'17 and KDD'19]

- Given a heterogeneous interconnected CIS network

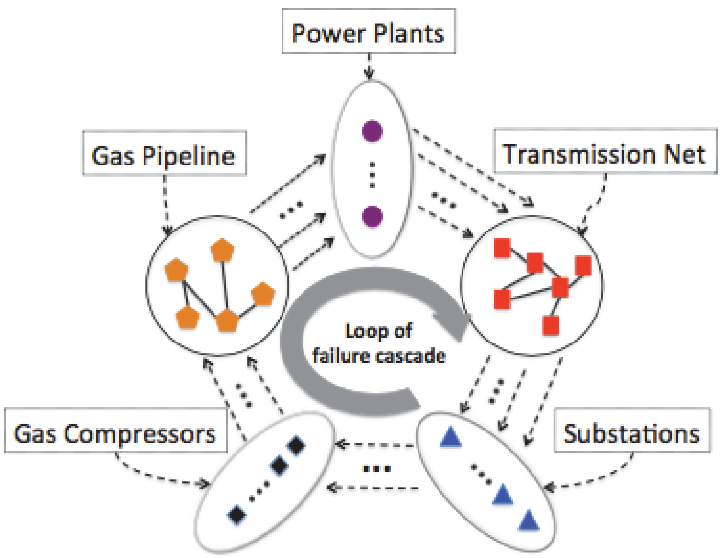


Infrastructure Type	Node Type	Description
Power	Electrical power plants (g)	Generate electrical power which is transmitted to substations through the transmission network.
	Transmission nodes (t)	Move electrical power from power plants to substations.
	Electrical substations (s)	Transform voltage and distribute electrical powers to consumers.
Natural gas	Natural gas compressors (c)	Increase the pressure of a gas to transport it through pipelines.
	Pipelines (p)	Transport natural gas to consumers.

OAK RIDGE National Laboratory

HotSpots

- Given a heterogeneous interconnected CIS network



Inter-connections:

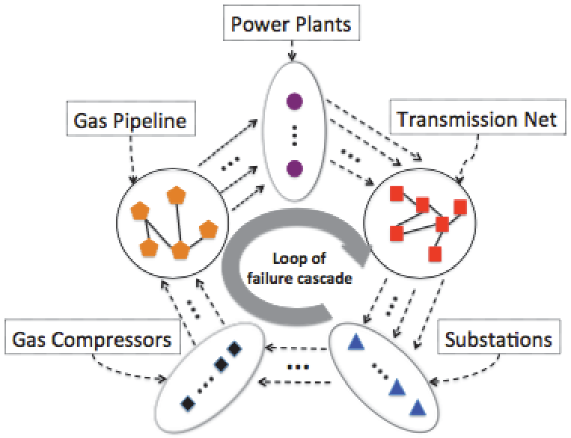
- Power plants are connected to the closest transmission line, and gas pipeline
- Substations are connected to the closest transmission line
- Gas compressors are connected to the corresponding substations that provide power to them; and to the closest gas pipeline
- Gas pipeline and transmission network themselves are networks with connections

Use the **Urbannet toolkit** to automatically construct heterogeneous CI networks from original raw shapefiles.

More details later in the tutorial

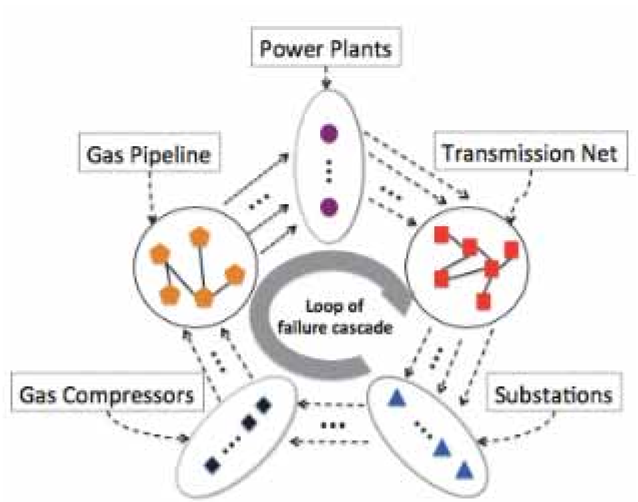
HotSpots

- Given a heterogeneous interconnected CIS network
- Goal 1: Model the failure cascade among multiple CIS
- Goal 2: Identify critical facilities that may lead to large failure spread over the entire system



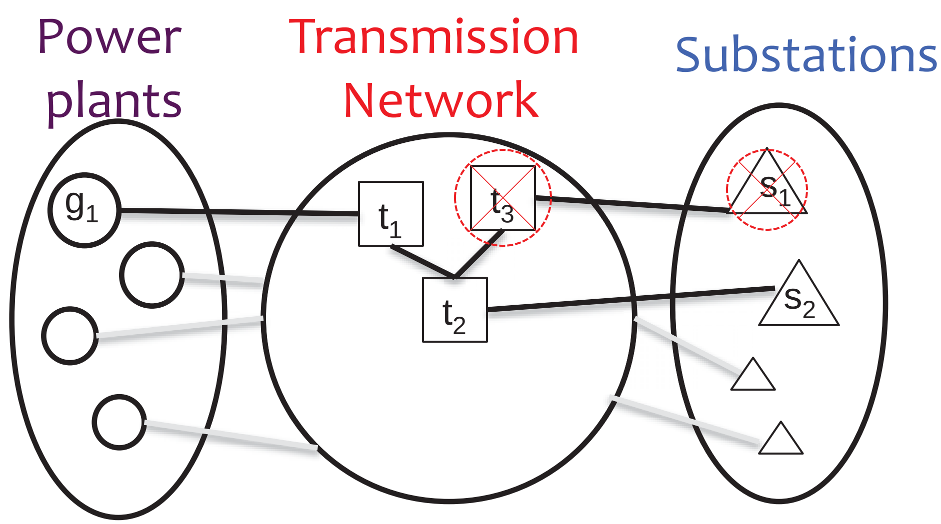
Goal 1: Failure Cascade Model

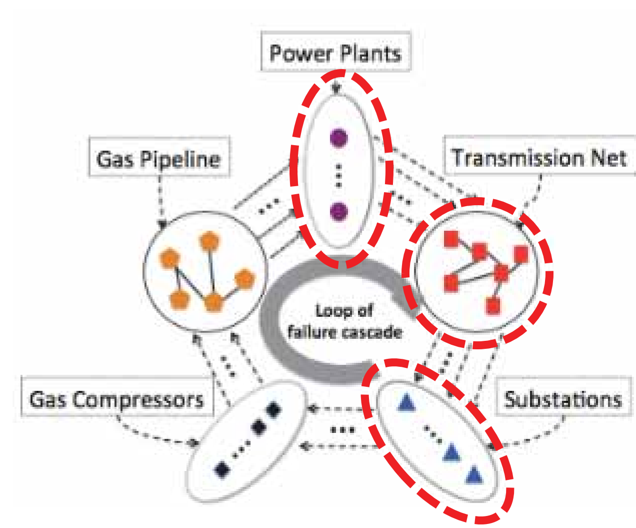
- Propose F-CAS model:



Goal 1: Failure Cascade Model

- Propose F-CAS model:
  - If a substation has no path in the trans. network to any power plant, it fails.

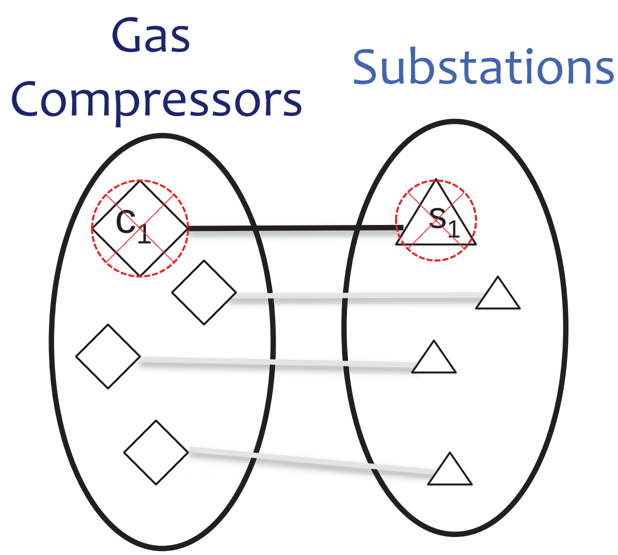


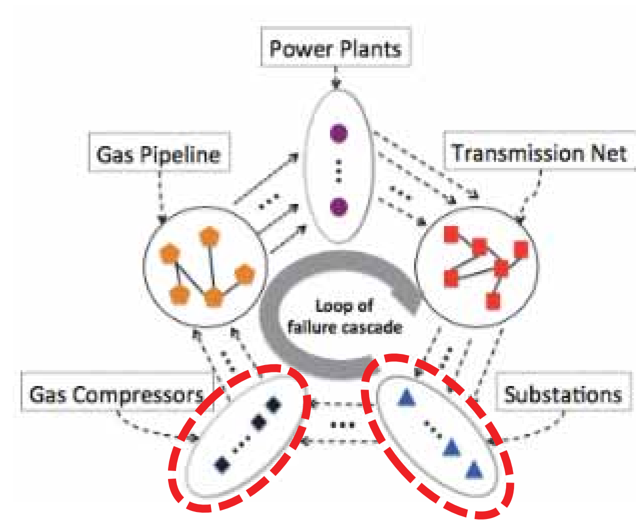


VirginiaTech

Goal 1: Failure Cascade Model

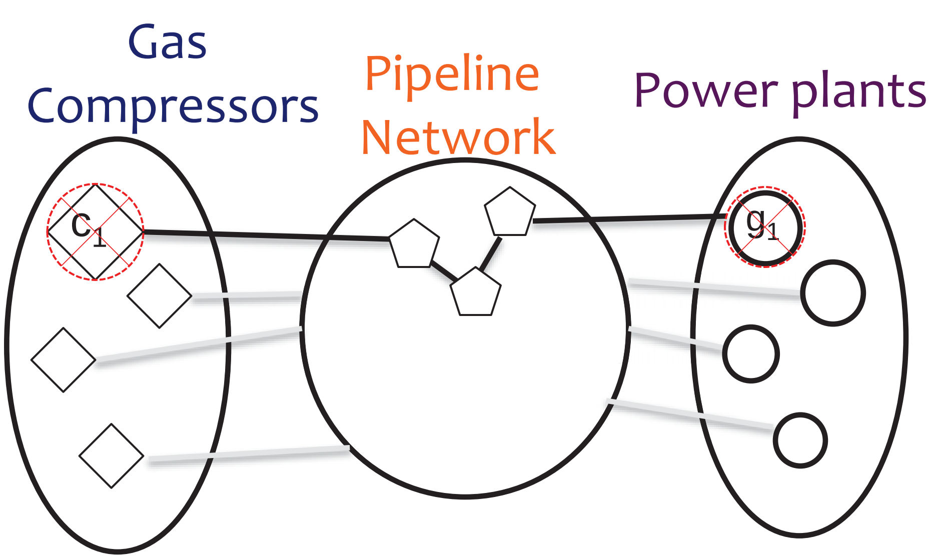
- Propose F-CAS model:
  - If a substation has no path in the trans. network to any power plant, it fails.
  - If a natural gas compressor's associated substation fails, it fails.

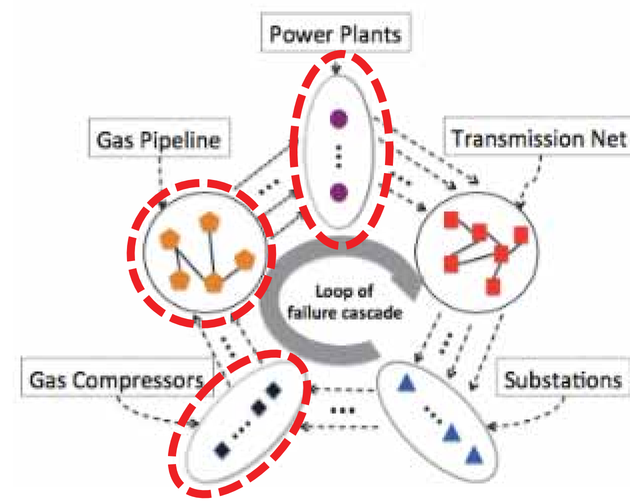




Goal 1: Failure Cascade Model

- Propose F-CAS model:
  - If a substation has no path in the trans. network to any power plant, it fails.
  - If a natural gas compressor's associated substation fails, it fails.
  - If a power plant's reachable natural gas compressor fails, it fails



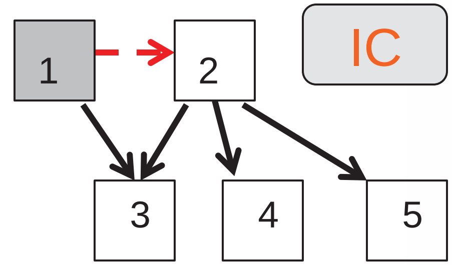


Goal 1: Failure Cascade Model

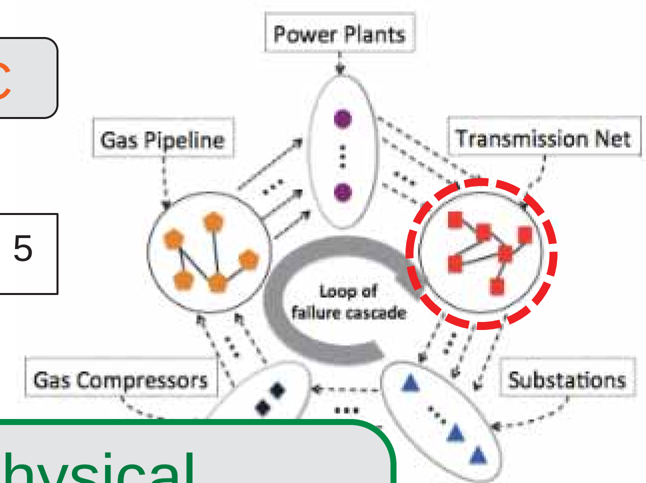
- Propose F-CAS model:
  - If a substation has no path in the trans. network to any power plant, it fails.
  - If a natural gas compressor's associated substation fails, it fails.
  - If a power plant's reachable natural gas compressor fails, it fails
  - In the trans. network, we propose two Independent Cascade (IC) [Kempe 2003] based models

Trans-naive

$$e_{ij} = \begin{cases} c & \text{if } t_i \text{ and } t_j \text{ share a child} \\ 0 & \text{otherwise} \end{cases}$$

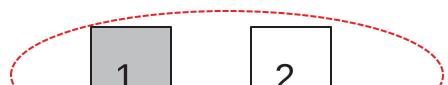


IC



Trans-real

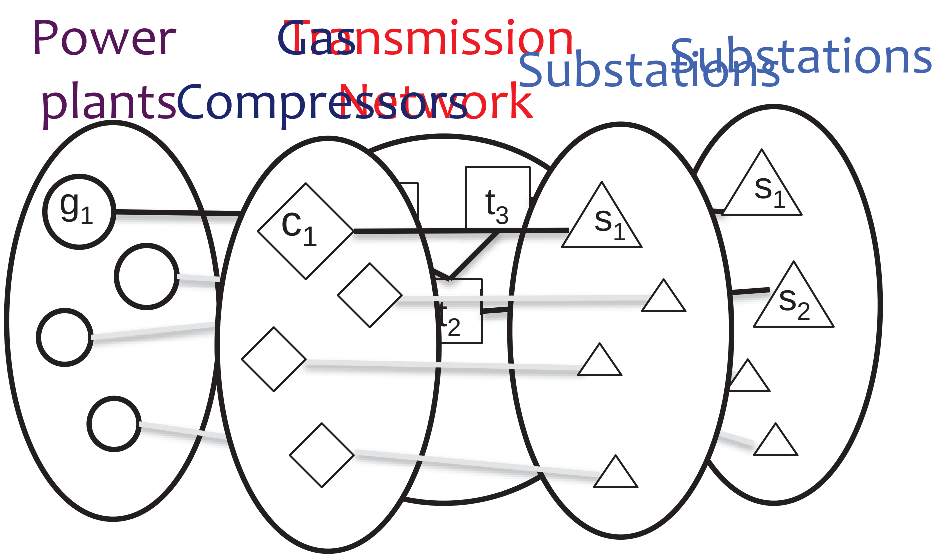
$$e_{ij} = \frac{\sum_{x \in Cs(Par(t_j) \setminus t_i)} Load(x)}{\sum_{x \in Par(t_j) \setminus t_i} Capacity(x)}$$

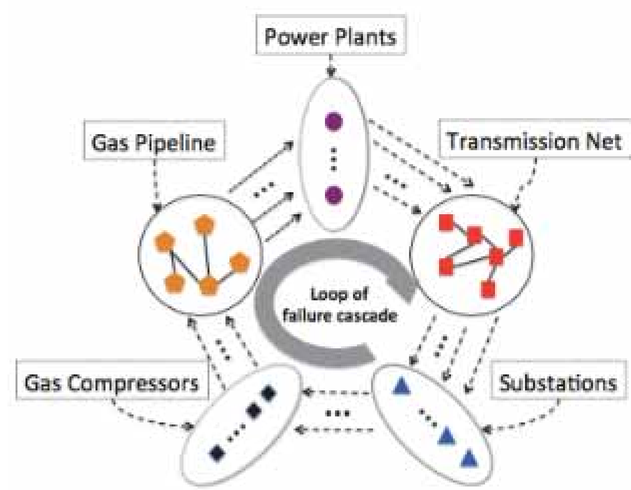


Based on physical interdependencies, domain constraints, and tractability

F-CAS: Novelty

- Neighbor-based failure cascade
  - Substation to gas compressors
- (New) Path-based failure cascade
  - Not handled by traditional cascade models such as the IC model, epidemiological models (SI, SIR, etc.)





Goal 2: Find Critical Nodes

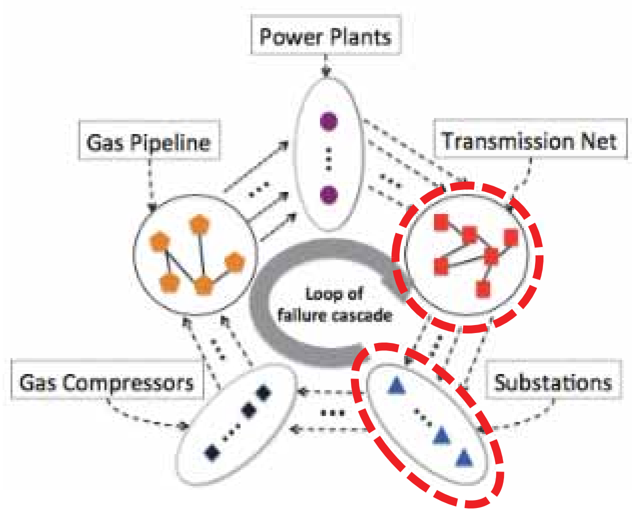
- Problem 1 (Max-Sub)
  - ***Given***: a heterogeneous network G, the failure cascade model F-CAS, and a budget k,
  - ***Find***: the critical set ***S**** of k transmission nodes, the failures of which maximize the expected number of failed substations.

$$S^* = \arg\max_S \mathbb{E}[\#s|S]$$

Expected number of failed substations in the end given S

$$\mathbb{E}[\#s|S] = \sum_{s_i} \Pr(s_i|S)$$

Probability of a substation s_i failing given S as the initial failures



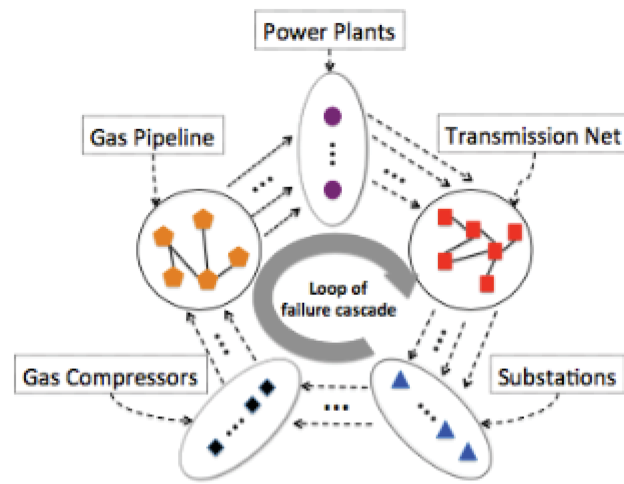
Prakash 2019

Challenge

- E[#s|S] is hard to directly optimize
  - For each s, we need to check the connectivity of the entire transmission network
- NP-Hard

Idea: Quickly estimate if a substation node would fail without re-checking the full network

new data structure
Dominator trees

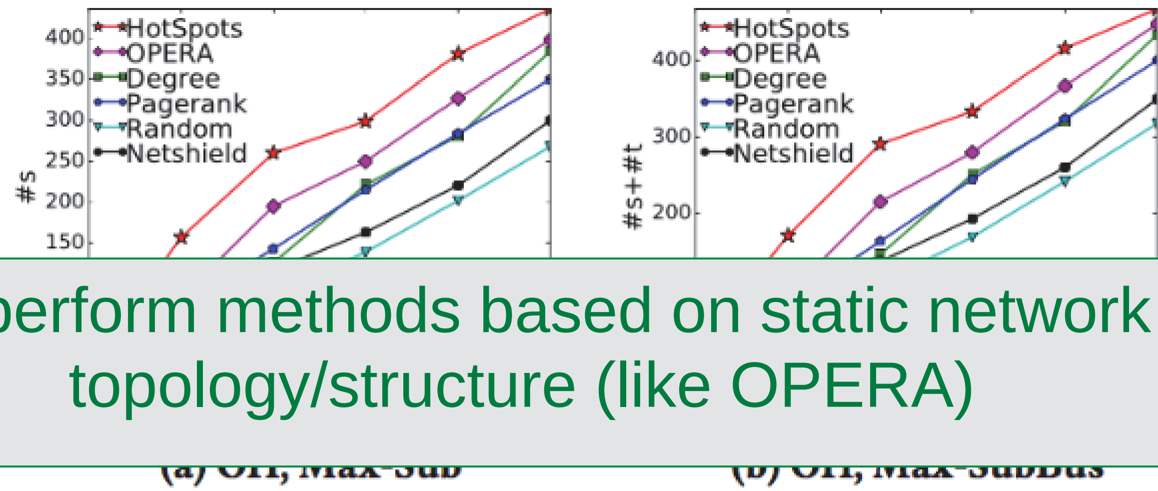


VirginiaTech

Effectiveness

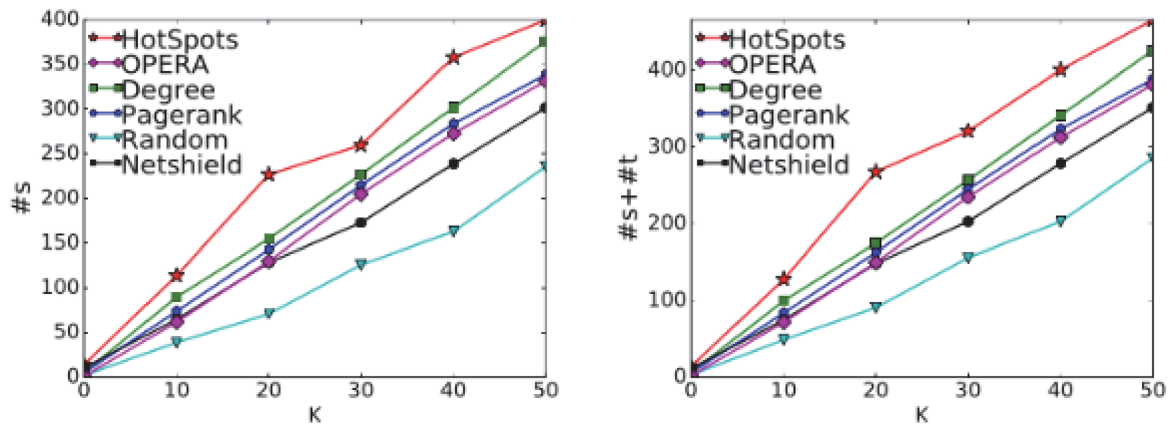
- Simulate the final failure spread of the selected critical nodes

Trans-naive



Outperform methods based on static network topology/structure (like OPERA)

Trans-real

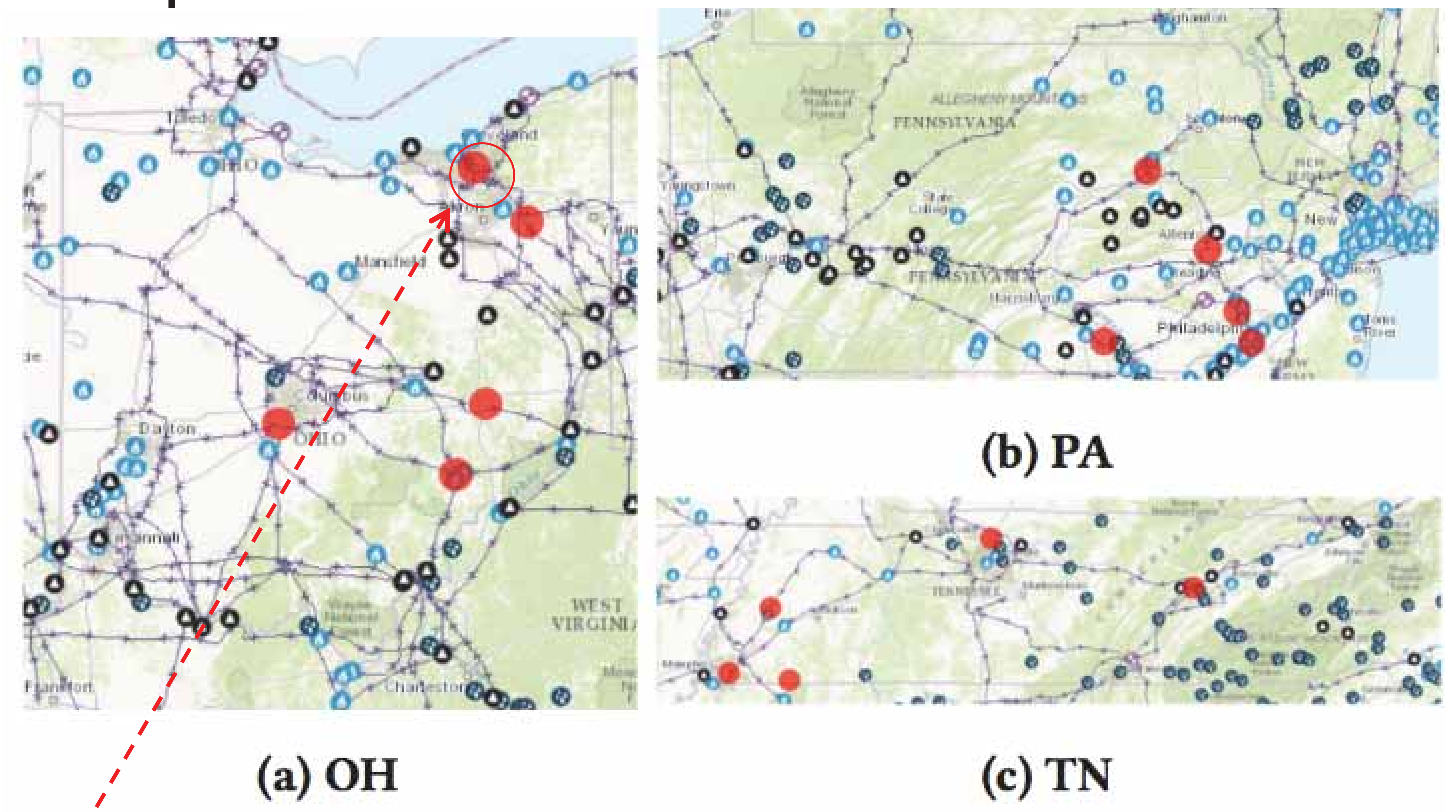


(c) OH, Max-Sub

(d) OH, Max-SubBus

Case Study: 2003 Blackout

- Evaluating critical nodes detected by HotSpots



(a) OH

(b) PA

(c) TN

Close to the transmission lines that initially failed during the blackout

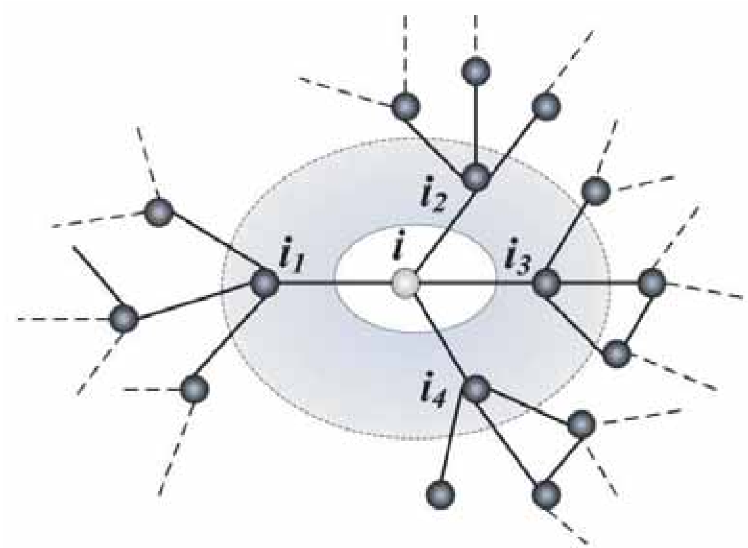
All detected nodes are on or close to the high voltage transmission lines

Part 1: Power systems

- Identifying critical/vulnerable facilities
  - Network structure based
  - Dynamics based
- **Protecting CIS against attacks**

Cascade-based attack vulnerability [Wang, Safety Science'09]

- Study the effect of two different attacks for the network robustness against failure cascade
- Load redistribution in the network cascading the failure

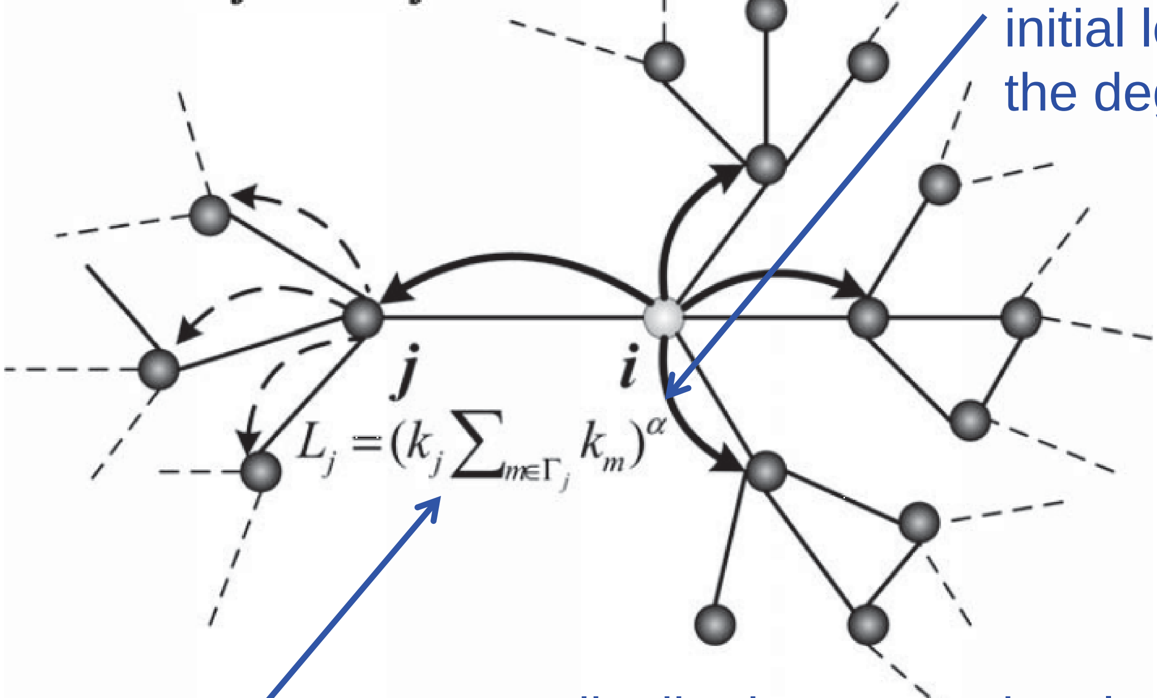


An attack on node i redistributes its load to the neighboring nodes

Model for load and load redistribution

Capacity of a node $C_j = TL_j$

α: Control how much the initial load depends on the degrees



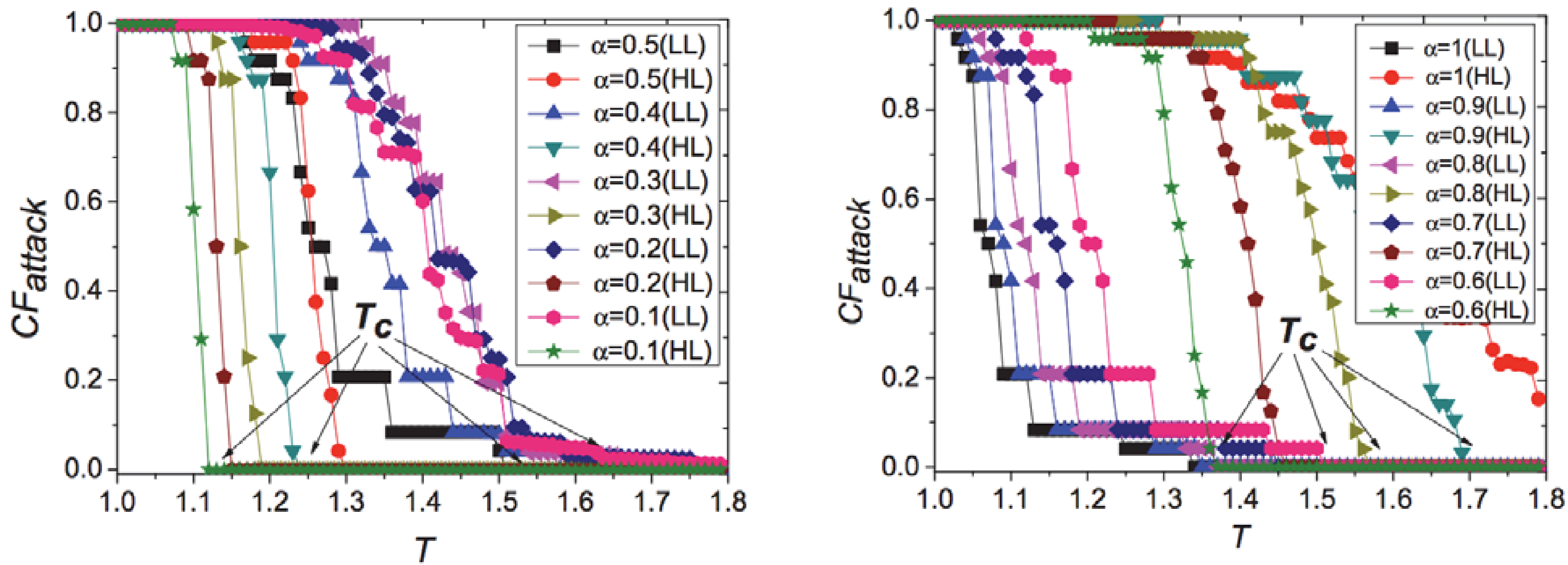
Initial load: product of degree and the summation of the neighboring degrees

Redistribution proportional to the initial loads

$$\Delta L_{ji} = L_i \frac{L_j}{\sum_{n\in\Gamma_i} L_n} = L_i \frac{\left[k_j \Sigma_{m\in\Gamma_j} k_m\right]^\alpha}{\sum_{n\in\Gamma_i} [k_n \Sigma_{f\in\Gamma_n} k_f]^\alpha}$$

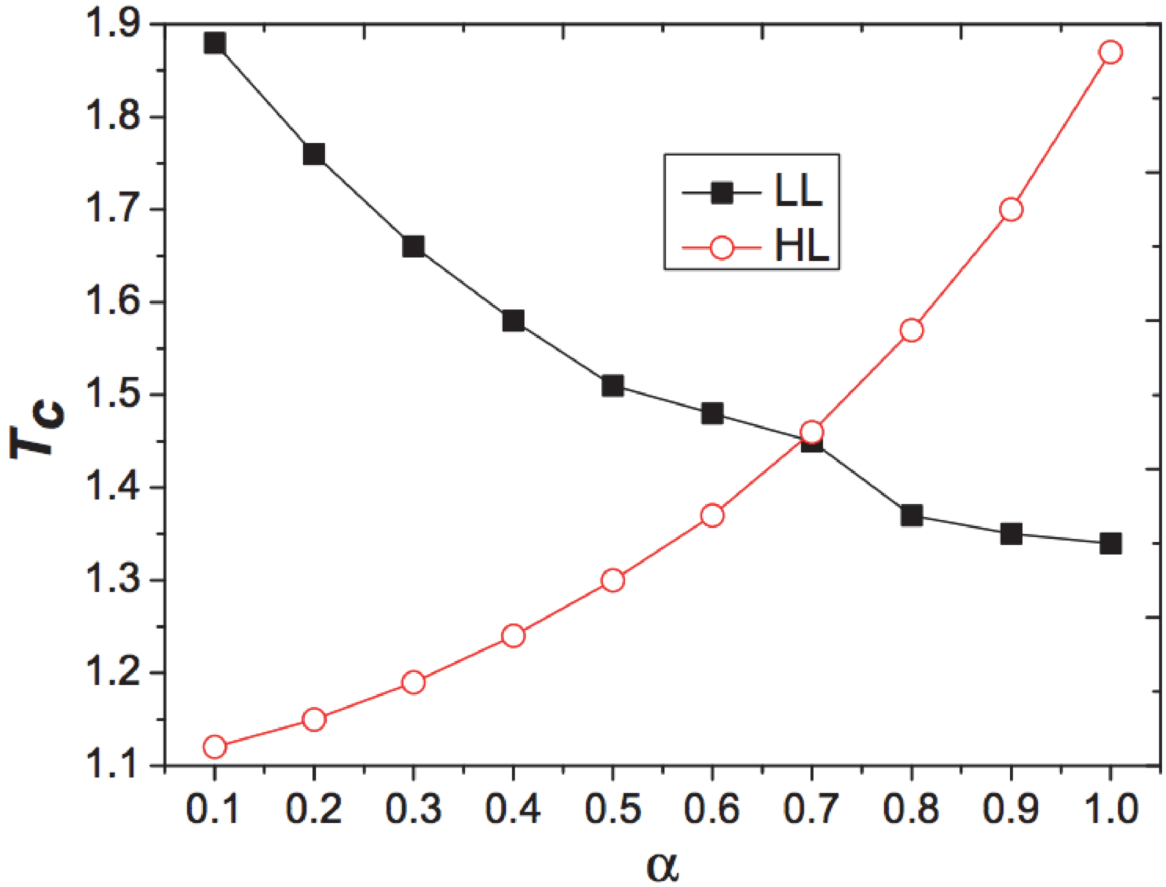
Comparison of two different attacking strategies

- HL: selecting 50 nodes with the highest loads
- LL: selecting 50 nodes with the lowest loads



T: ratio between capacity and initial load
CFattack: impact of the attack

Comparison of two different attacking strategies

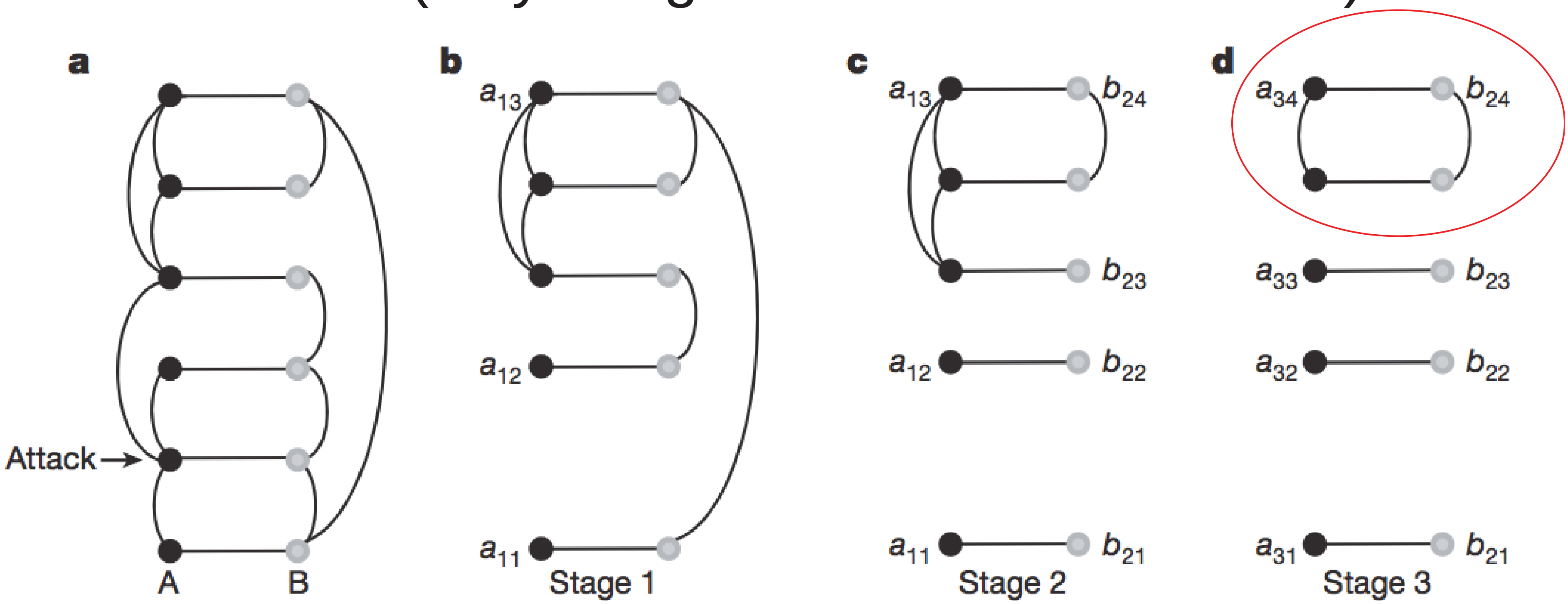


Different optimal strategies with different α setting

Robustness of interdependent networks under targeted attack

[Huang, Physical Review'11]

- Failure cascade based on mutually connected clusters (only the giant cluster is functional)



Based on the non-adversarial work [Buldyrev, Nature'10], but integrating adversarial attacking strategy

Degree based attacks

- Use α to adjust whether to target high degree nodes or low degree nodes

Degree of the node

$$W_{\alpha}(k_i) = \frac{k_i^{\alpha}}{\sum_{i=1}^{N} k_i^{\alpha}}, \qquad -\infty < \alpha < +\infty$$

Probability of a node being attacked

$\alpha>0$: target high degree nodes
$\alpha=0$: random selection
$\alpha<0$: target low degree nodes

Main conclusion

Degree-based targeted attack on interdependent network A and B	converted to →	Random attack on interdependent network A' and B

Idea: Removal of edges connecting a deleted edge is equivalent to randomly removing a portion of edges of the remaining nodes

Optimize resilience against attacks [Ouyang, 2017]

- Integrate the arranging of the repair sequence of damaged components under limited repair resources into protection planning

Maximizing the resilience of the system

$$R(T) = 1 - \frac{\int_{t_0}^{t_0+t_r} [P_T - P_R(t)]dt}{TP_T}$$

↓

Minimizing the resilience loss

$$RL = \int_{t_0}^{t_0+t_r} [P_T - P_R(t)]dt = \int_{0}^{t_r} [P_T - P_R(t)]dt$$

CIRO-IA model

Attacking strategy

Repair optimization

Protection strategy

$$RL\left(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{r}, \boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{P}^S, \boldsymbol{f}, \Delta \boldsymbol{P}^D\right)$$

$$= \min_{\boldsymbol{w}} \max_{\boldsymbol{v}} \min_{\boldsymbol{x},\boldsymbol{r},\boldsymbol{\eta},\boldsymbol{\theta},\boldsymbol{P}^S,\boldsymbol{f},\Delta\boldsymbol{P}^D} \sum_{t=1}^{T_P-1} \left(\sum_{n \in V^D} \Delta P_{nt}^D \right)$$

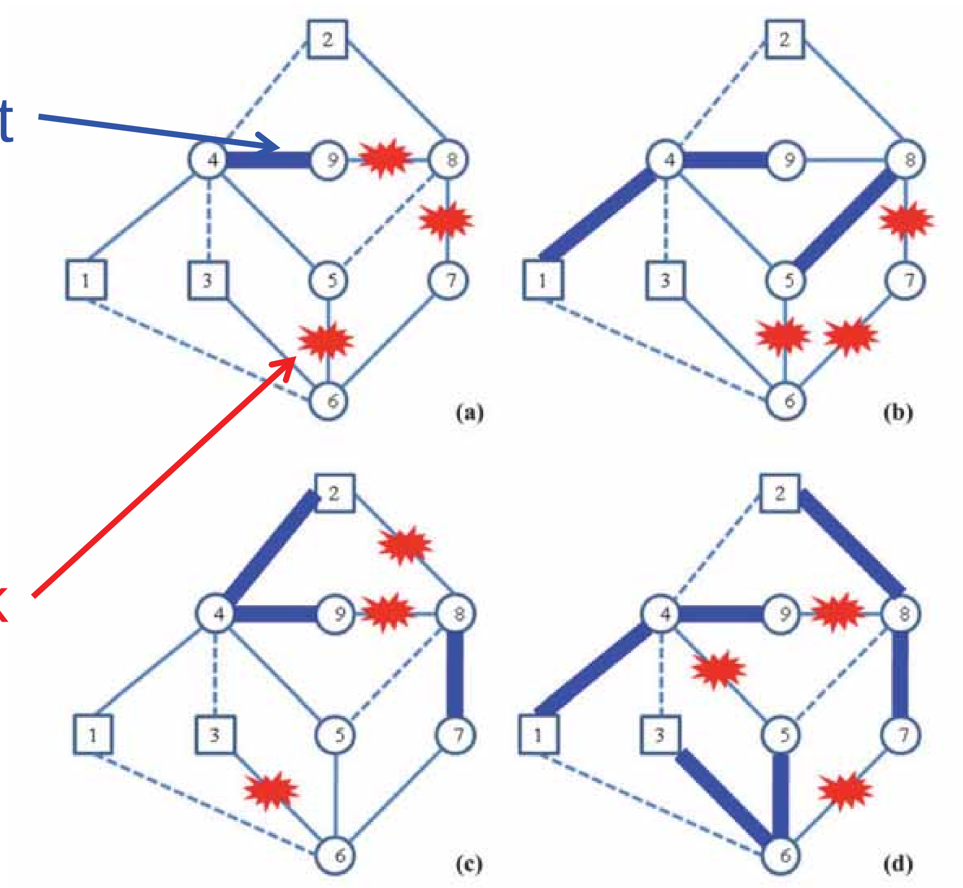
$$\times \left[\sum_{k=1}^{RR} \eta_{k(t+1)} \left(\sum_{n \in V^*} \sum_{s=1}^{t+1} r_{kns} \tau_n + \sum_{l \in L^*} \sum_{s=1}^{t+1} r_{kls} \tau_l \right) \right.$$

$$\left. - \sum_{k=1}^{RR} \eta_{kt} \left(\sum_{n \in V^*} \sum_{s=1}^{t} r_{kns} \tau_n + \sum_{l \in L^*} \sum_{s=1}^{t} r_{kls} \tau_l \right) \right] \quad (4$$

Case study

Edges to protect

Nodes to attack



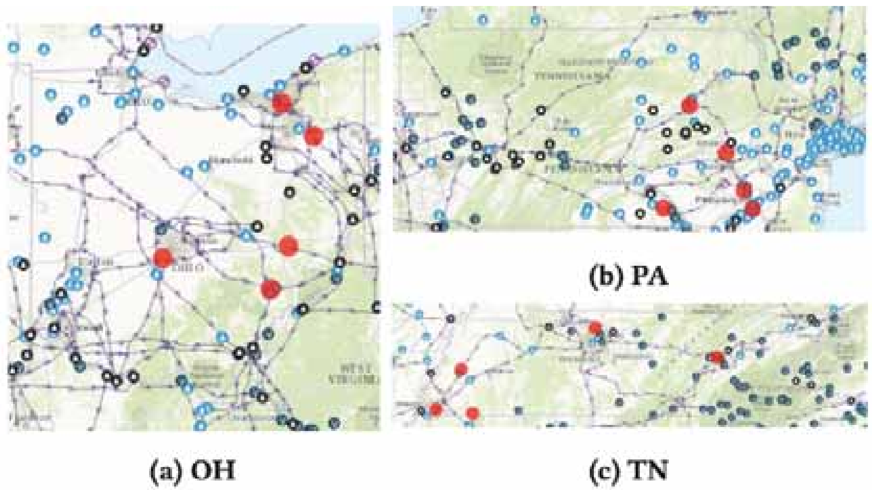
VirginiaTech

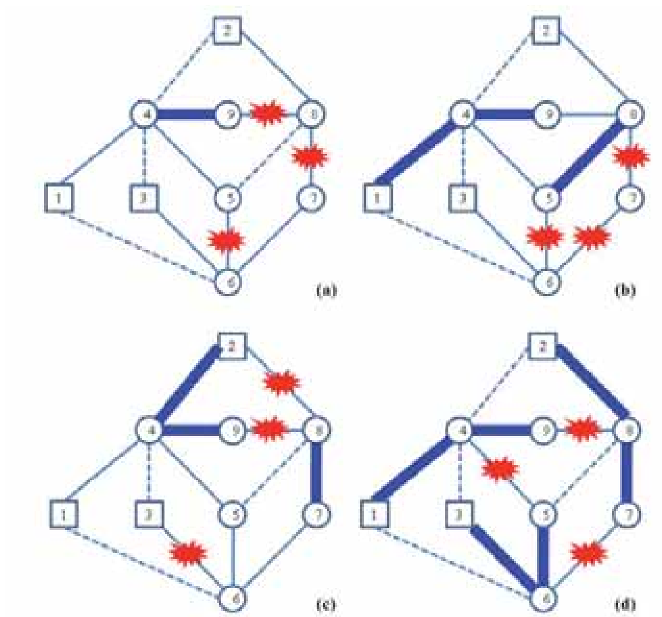
Summary: Part 1

Improving power system robustness

Non-adversarial

Adversarial





Improvement based on network topology or failure cascade dynamics

Protection against adversarial attacks

Outline

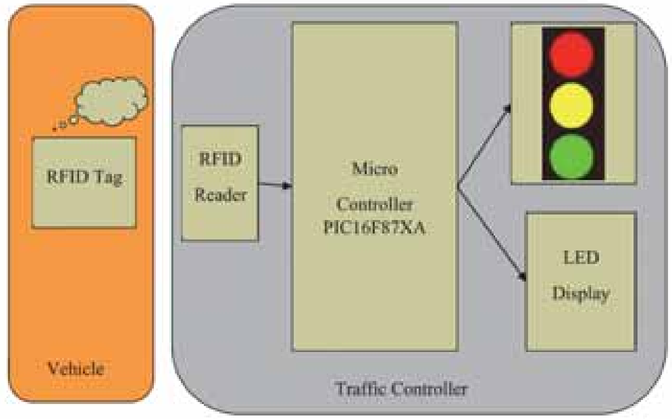
- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Transportation Systems

- Example problems:

Predicting traffic flow

Predicting different traffic states such as weather, accidents, etc.



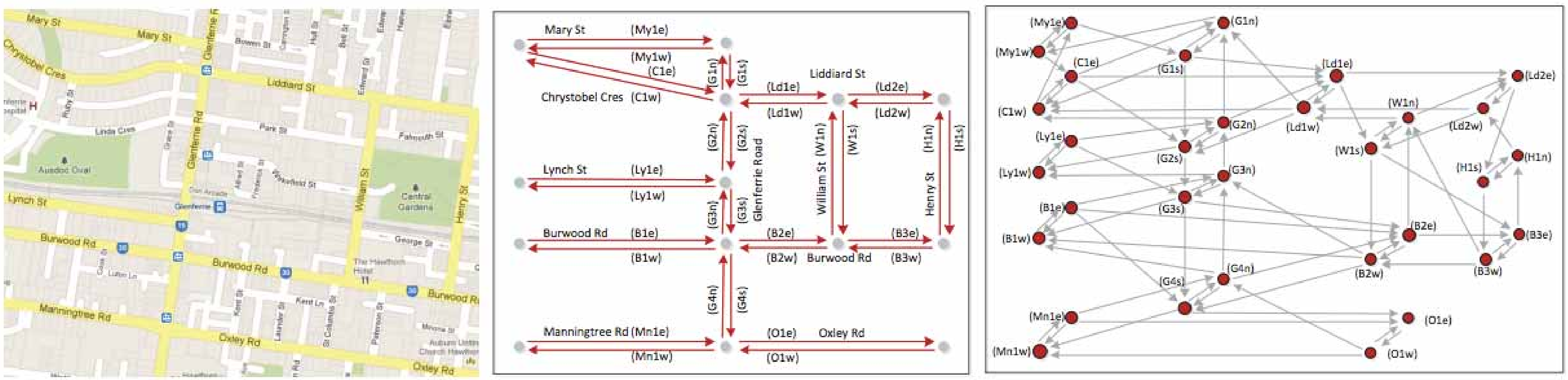
Congestion tracking and control

Influence estimation for traffic diffusion [Anwar+, CIKM'15]

- Given
  - Traffic data:
    - Traffic volume: the count of vehicles crossing a road segment during the green light time
    - Degree of saturation: the ratio of effectively used green light time and the total green light time
  - Road network:
    - A network of road intersections, connected by road segments with features
- Compute the influence score for each road segment (how much the traffic on this segment influence that on the global network)

RoadRank

- Detect the influence between road segments in terms of propagating congestions.



(a) Actual road map (b) Road network (c) Influence graph

Defining traffic diffusion probabilities

$$tdp(r_i \rightarrow r_j) = \frac{td(r_i \rightarrow r_j)}{\sum\limits_{\forall k: r_i \xrightarrow{inf} r_k} td(r_i \rightarrow r_k)}$$

Calculating PageRank-based influence score

Finding most influential roads

- Detecting congestion areas

Rank	Road segment	RR Score
	03-02-2012 08:05 AM	
1.	Hoddle St (Victoria Parade to Elizabeth St)	02.57095
2.	Hoddle St (Elizabeth St to Albert St)	02.00645
3.	Mills St (Canterbury Rd to Danks St)	01.89356
4.	Heidelberg Rd (Hoddle St to The Esplanade)	01.83253
5.	Heidelberg Rd (The Esplanade to Hoddle St)	01.81797
	03-02-2012 10:05 AM	
1.	Hoddle St (Victoria Parade to Elizabeth St)	03.49890
2.	Hoddle St (Elizabeth St to Albert St)	02.50019
3.	Heidelberg Rd (Hoddle St to The Esplanade)	02.38343
4.	Heidelberg Rd (The Esplanade to Hoddle St)	02.35427
5.	Hoddle St (Elizabeth St to Victoria Parade)	02.31045

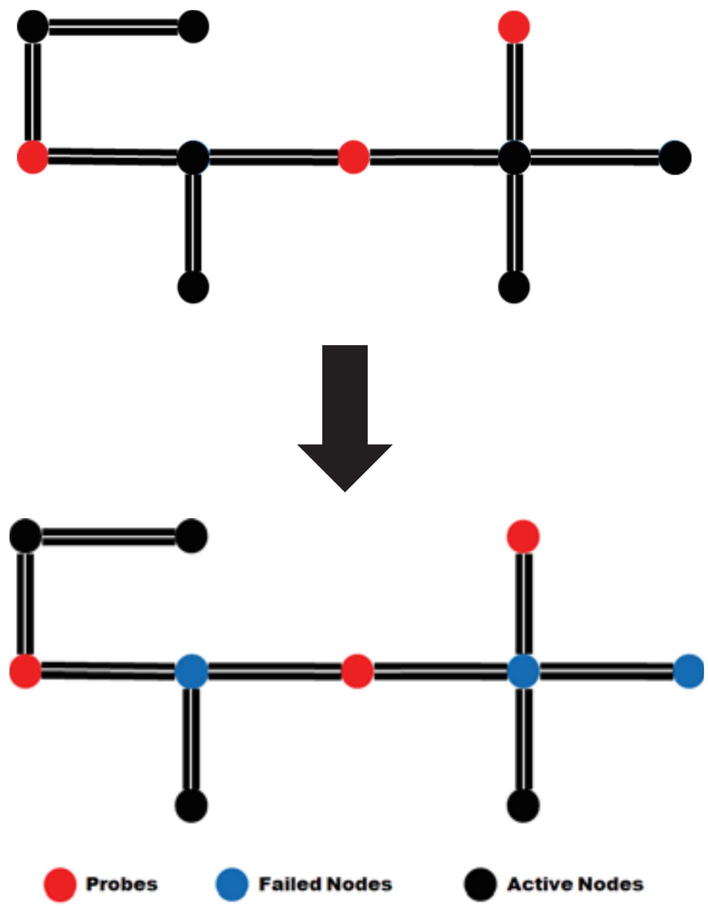
State estimation using crowd sourced apps [Adhikari+, SIAM Data Mining '18]

- Crowd sourced application
  - Navigation
  - Reporting incidents on road
- Users report incidents like
  - Accidents
  - Traffic Jam
  - Stranded Vehicles
  - …

Waze app

Problem formulation

- **Given**
  - A network $G(V,E)$ with $I \subset V$ which have failed
  - Probes: nodes observed to have been failed $Q \subset I$
- **Infer**
  - Most likely unobserved nodes which have failed $I - Q$

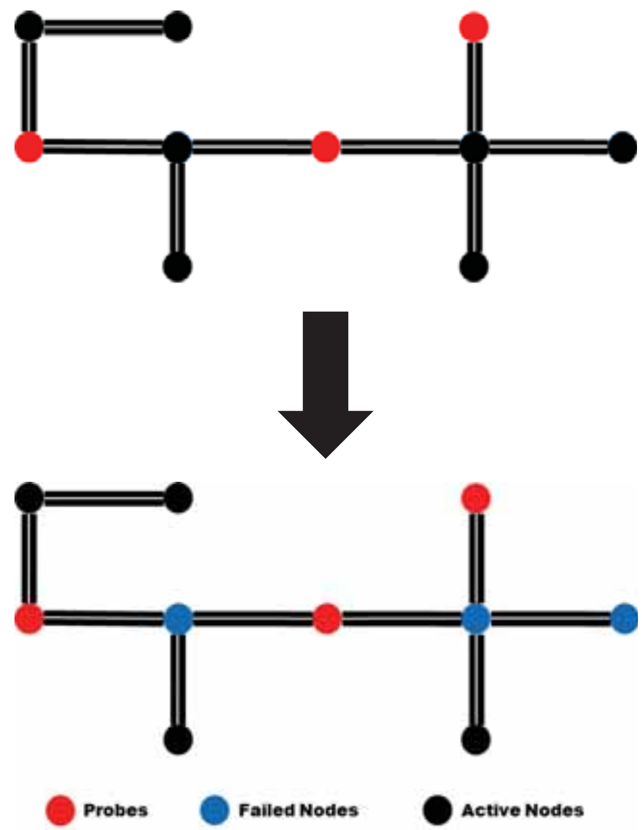


Road Network Failure

Failures are geographically correlated
[Agarwal et al., IEEE/ACM ToN 2013]

GRAPHSTATEINF based on MDL

- **Given:**
  - Graph $G(V,E)$
  - Probes $Q \subset V$
  - Probability Dist. p_s and F
- **Find:**
  - The failure set I
- **Such that:**
  - MDL cost is minimized



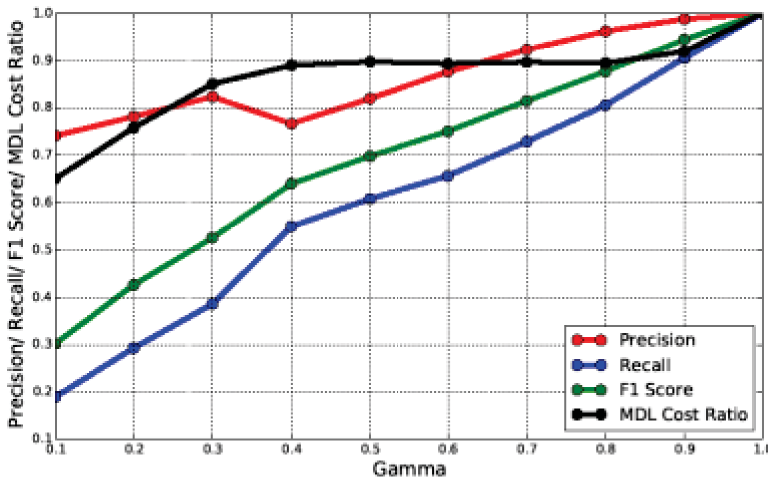
$$\mathcal{L}(|\mathcal{Q}|, |I|, I, \mathcal{Q}) = -\log \binom{|I|}{|\mathcal{Q}|} - \log \left(\sum_{s \in V} p_s(s) \prod_{v \in I} F(v \mid s) \prod_{v' \notin I} \left(1 - F(v' \mid s)\right)\right)$$

$$-2|\mathcal{Q}| \log(\gamma) - 2(|I| - |\mathcal{Q}|) \log(1-\gamma)$$

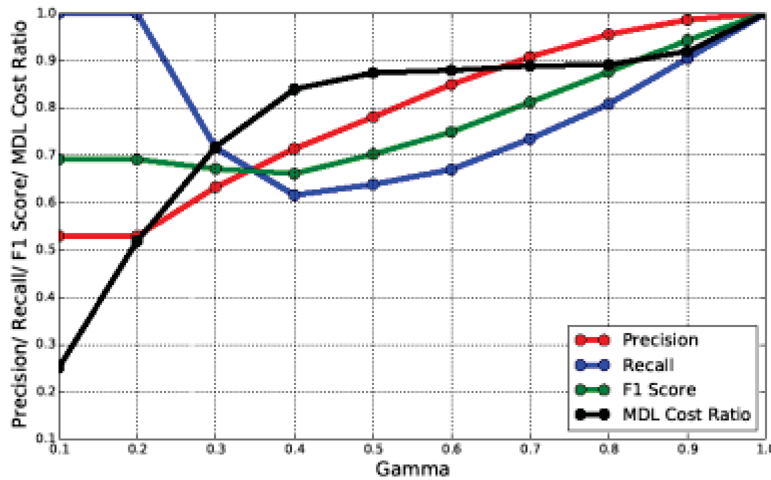
Near optimal

Performance

Higher is better



LocalSearch



GraphMap

GraphMap results in higher F1 Score

Outline

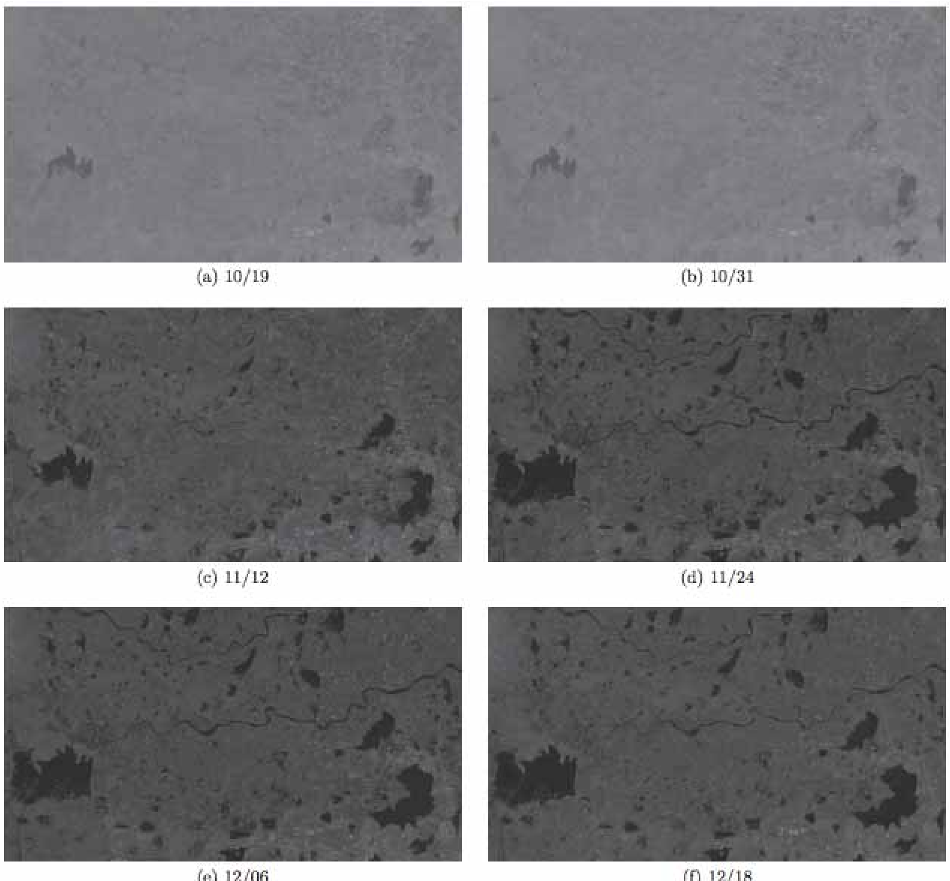
- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Facilitating decision making

- Improving situation awareness
  - **Ex 1: Finding flooding area**
  - Ex 2: Spatial event discovery
- Other CIS systems & tools
  - Critical Infrastructure Protection/Decision Support System (CIP/DSS)
  - Urbannet toolkit & web interface
  - Other resources

Flood mapping on satellite images [Liang, WWW'18]

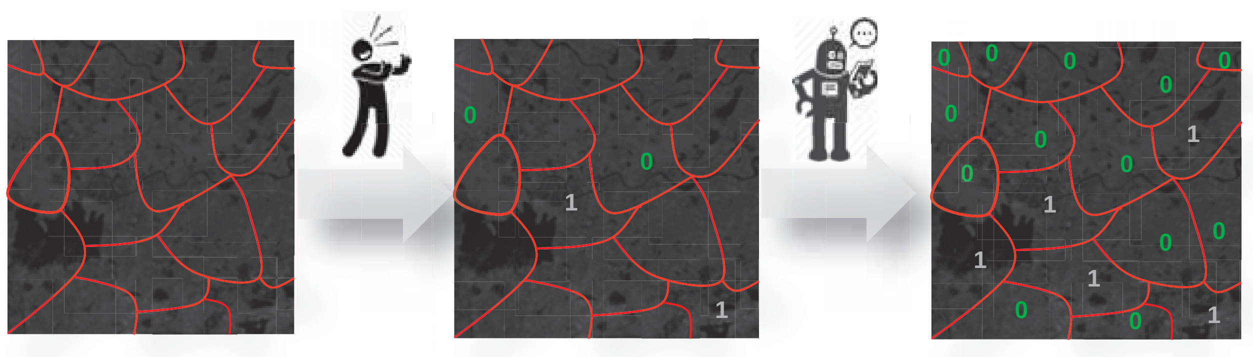
- Distinguish flooded areas from non-flooded areas using image segmentation techniques



Satellite images of Chennai

Human guidance

- A semi-supervised learning algorithm
  - Divide the satellite image into patches using a graph-based approach depending on the proximity and intensity of the pixels
  - Classify each patch: each time the user is asked to label a few patches, and then learn a classifier to automatically classify the other patches

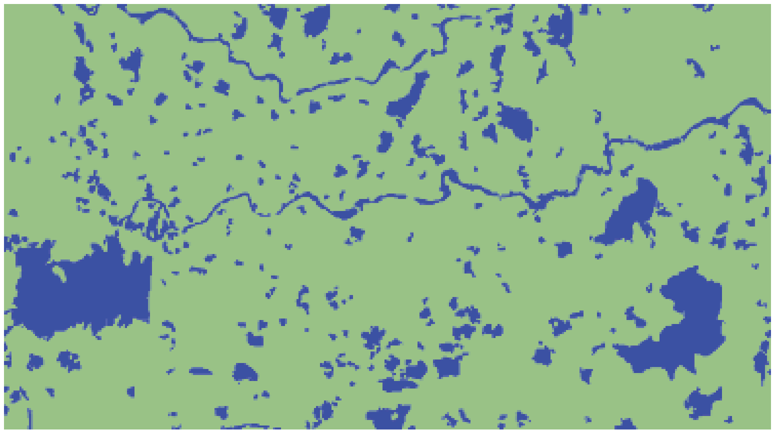


Prakash 2019

VirginiaTech

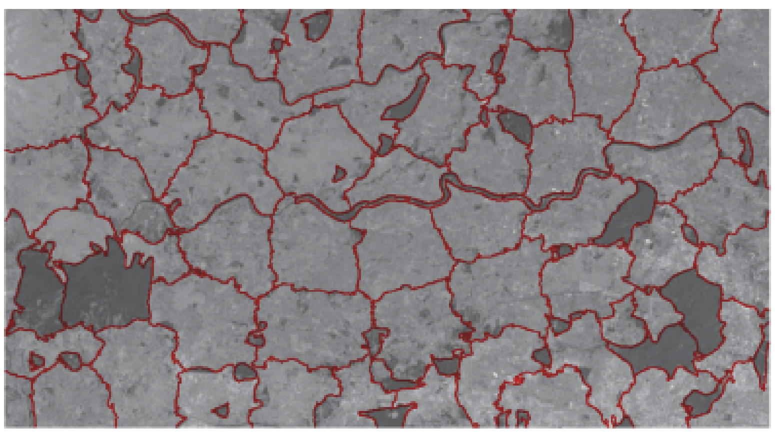
Performance

- Identify flooding areas with high accuracy

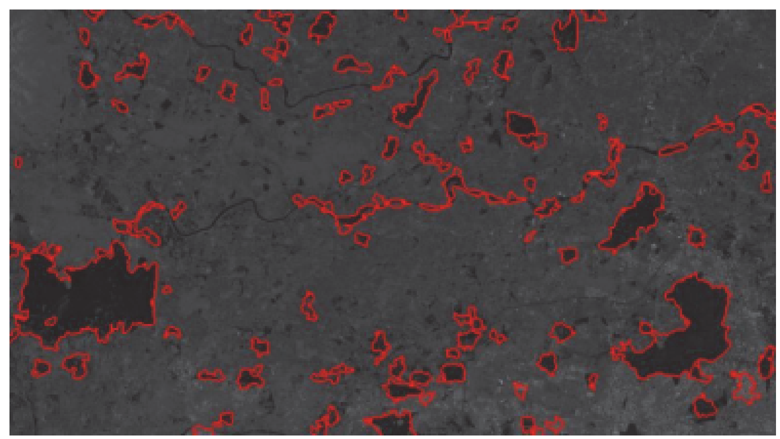


(a) Our Method

(b) Watershed Algorithm



(c) Normalized Cuts Algorithm (100 partitions)



(d) Graph-based Clustering with Post-processing (100 partitions)

Facilitating decision making

- Improving situation awareness
  - Ex 1: Finding flooding area
  - **Ex 2: Spatial event discovery**
- Other CIS systems & tools
  - Critical Infrastructure Protection/Decision Support System (CIP/DSS)
  - Urbannet toolkit & web interface
  - Other resources

Real-time event detection

[Sakaki, WWW'10]

- Using Twitter users as sensors
- Design
  - A classifier for detecting target events
  - A probablistic spatial-temporal model that finds the center and trajectory of the event

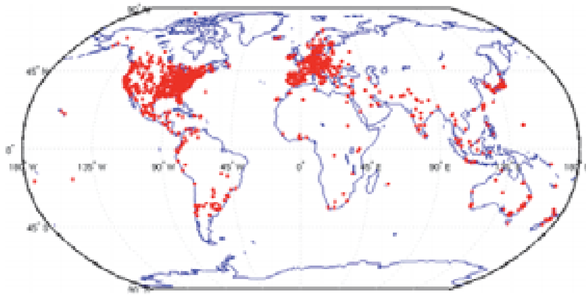


Figure 1: Twitter user map.

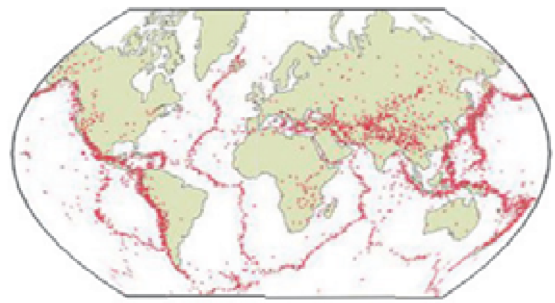
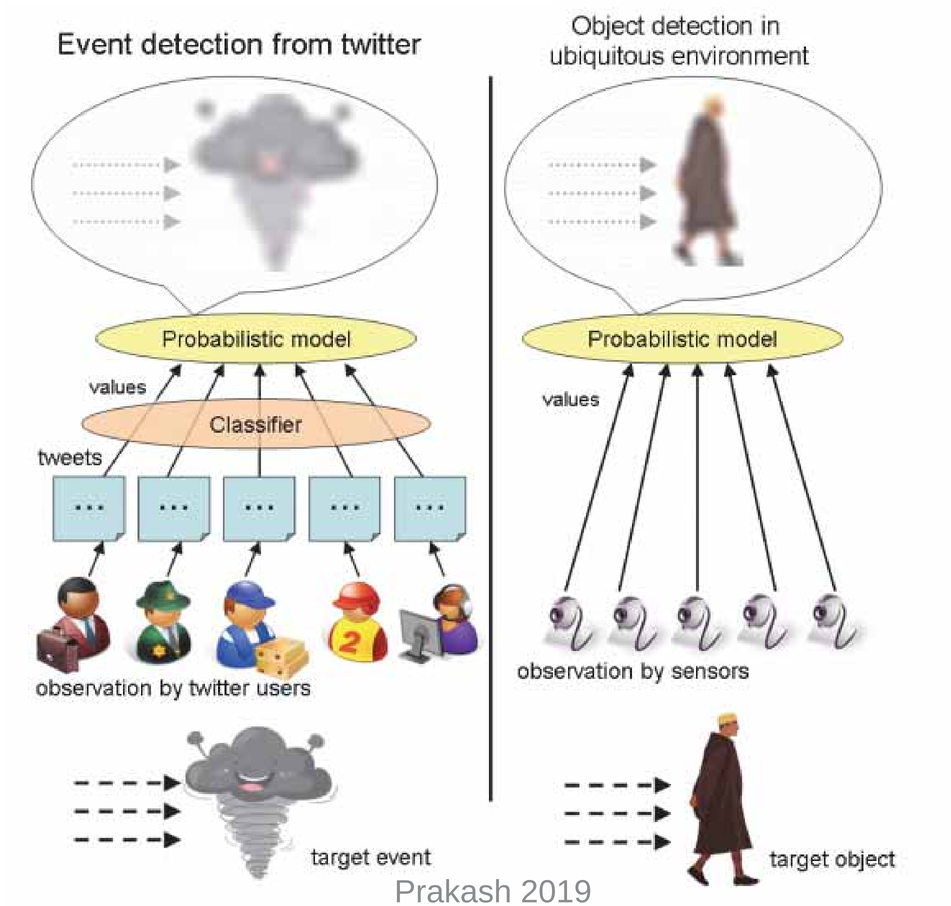


Figure 2: Earthquake map.

VirginiaTech

Summary of the framework

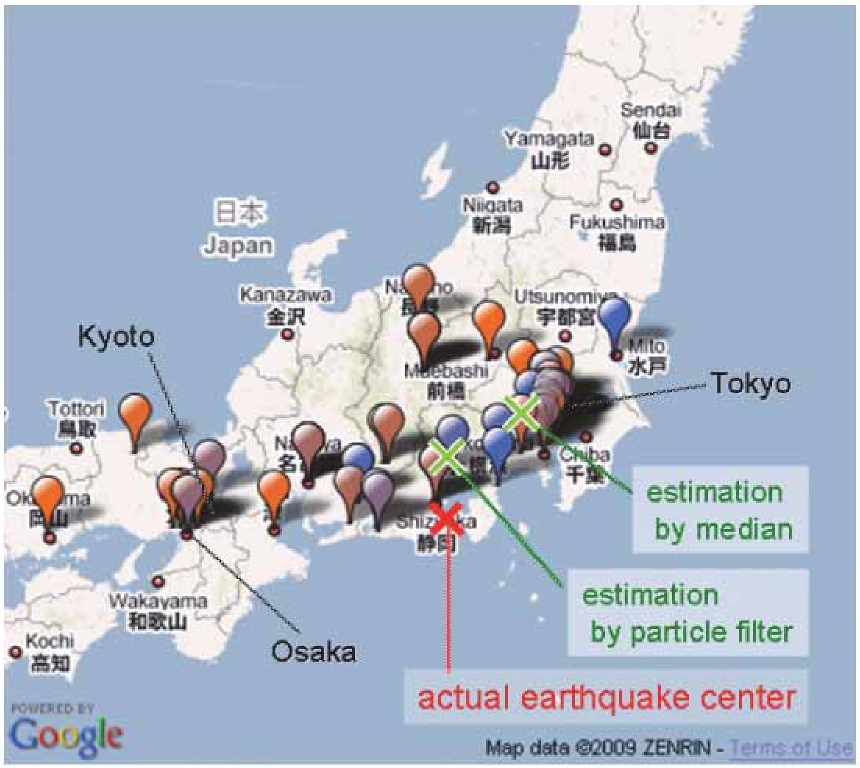
- Correspondence between event detection from Twitter and object detection in a ubiquitous environment

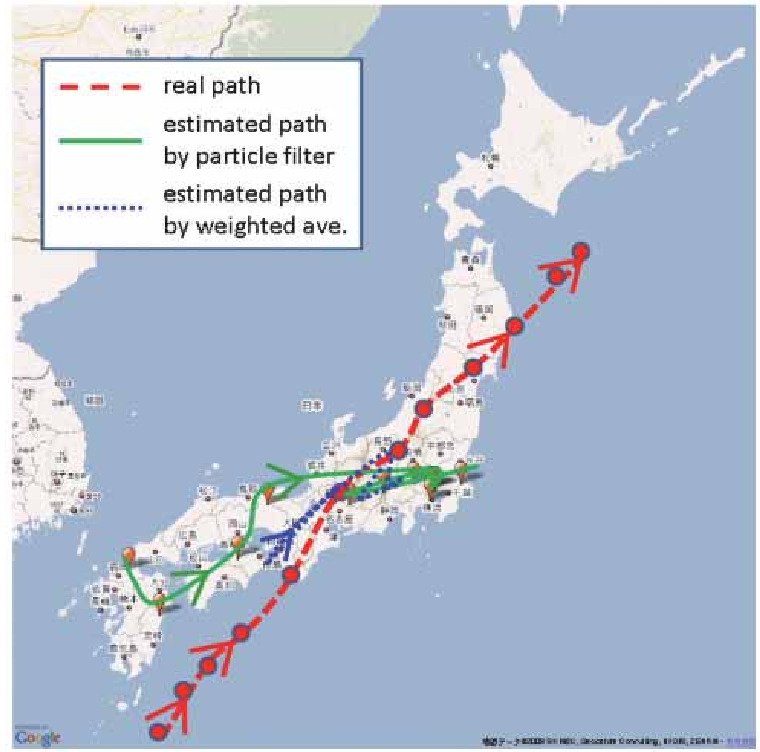


Prakash 2019

VirginiaTech

Results





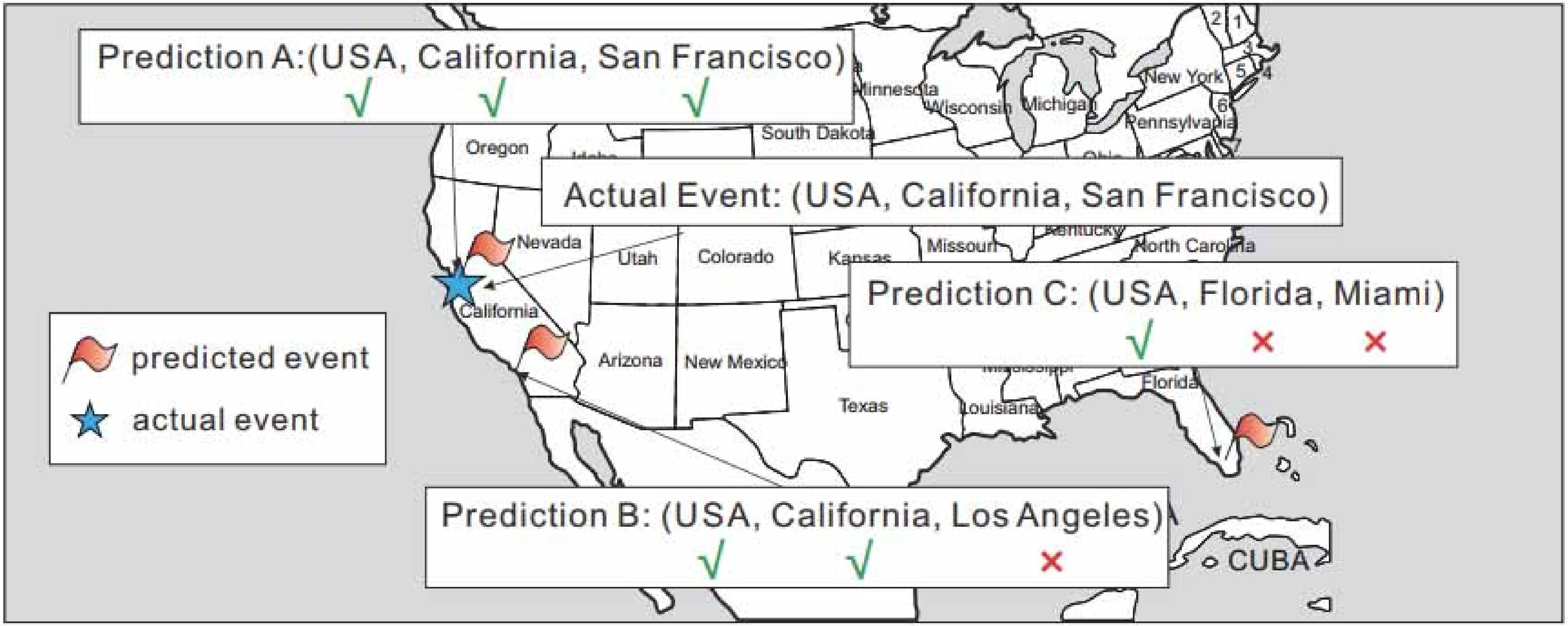
Earthquake location estimation

Typhoon trajectory estimation

Multi-resolution spatial event forecasting in social media

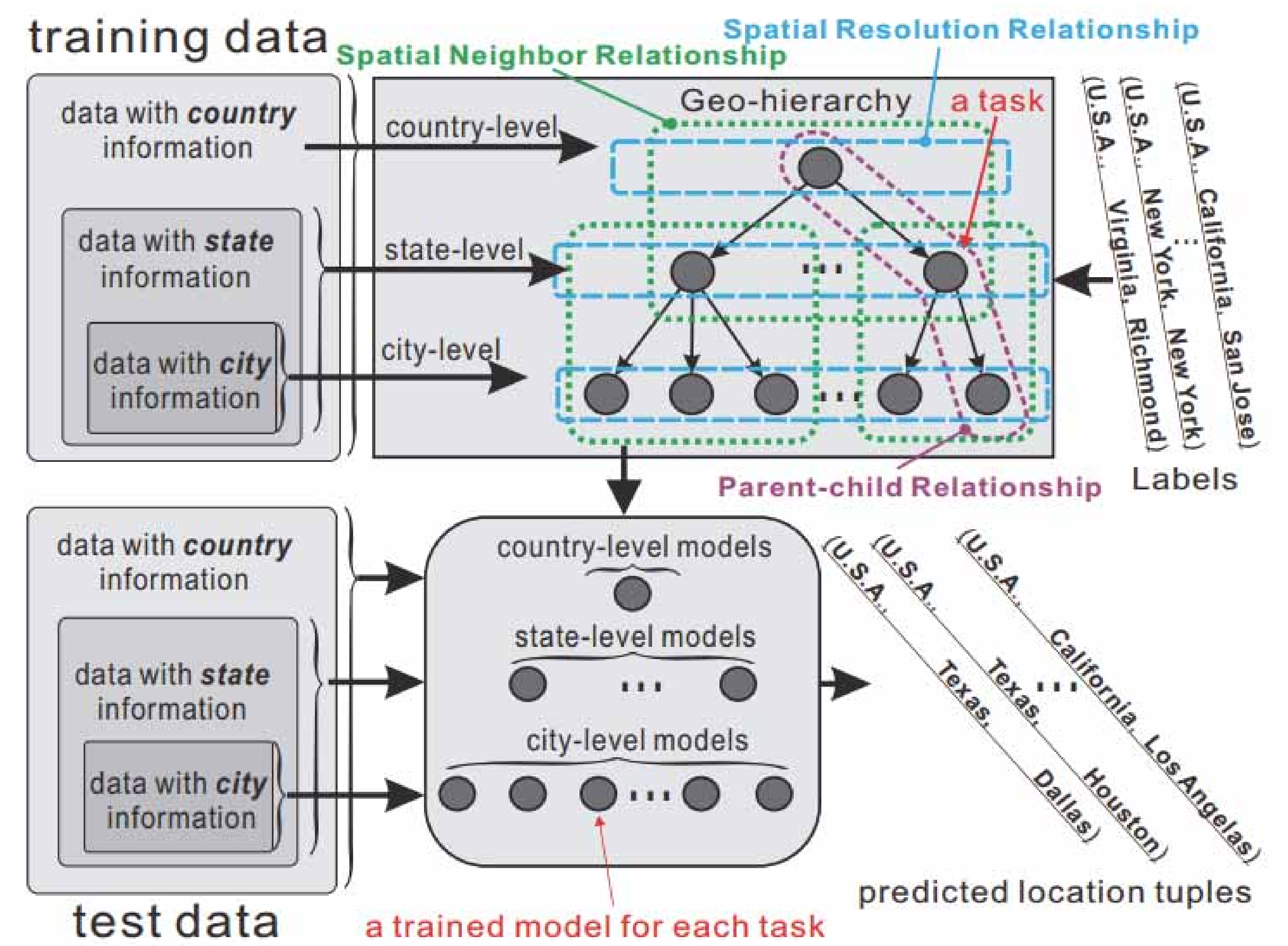
[Zhao, ICDM'16]

- Trade-off between accuracy and discernibility



Three different predictions, correct in different discernibilities

MREF model



Prakash 2019

VirginiaTech

Performance

Method	Brazil	Colombia	Ecuador	El Salvador	Mexico	Paraguay	Uruguay	Venezuela
ARX	0.63,0.47,0.54	0.30,0.40,0.35	0.33,**0.47**,0.39	0.44,0.42,0.43	**0.43**,0.20,0.27	0.52,0.27,0.36	0.53,0.60,0.56	0.51,0.23,0.32
LR	0.43,0.41,0.42	0.33,0.38,0.36	0.37,0.39,0.38	0.50,0.34,0.41	0.30,0.11,0.16	0.52,0.23,0.32	0.48,0.47,0.48	0.40,0.33,0.36
KDE-LR	0.99,0.01,0.02	**0.68**,0.01,0.01	0.16,0.13,0.15	0.28,0.09,0.14	0.02,0.15,0.04	0.04,0.35,0.07	0.13,**0.93**,0.22	0.69,0.03,0.06
LDA-LR	**1.00**,0.01,0.02	0.01,**0.63**,0.02	0.16,0.13,0.15	0.26,0.09,0.13	0.01,0.19,0.02	0.04,0.36,0.07	0.14,**0.93**,0.24	**0.99**,0.04,0.07
LASSO	0.74,0.45,0.56	0.40,0.41,0.40	0.34,0.42,0.38	**0.62**,0.36,0.46	0.18,**0.42**,0.25	0.72,0.25,0.37	0.61,0.46,0.52	0.19,**0.80**,0.31
MTL	0.68,**0.48**,0.56	0.37,0.44,**0.41**	0.24,0.55,0.34	0.42,**0.45**,0.43	0.42,0.24,**0.31**	0.57,0.29,0.38	0.60,0.54,0.56	0.37,0.45,0.41
TMTL	0.46,0.42,0.44	0.36,0.34,0.35	0.37,0.43,**0.40**	0.57,0.43,0.49	0.29,0.25,0.27	0.25,**0.42**,0.31	0.60,0.64,0.62	0.41,0.58,**0.48**
MREF	0.79,0.47,**0.59**	0.37,0.39,0.38	**0.38**,0.43,**0.40**	0.58,0.43,**0.50**	0.29,0.30,0.29	**0.75**,0.26,**0.39**	**0.66**,0.60,**0.63**	0.24,0.49,0.33

State Level (precision, recall, F-measure)

Method	Brazil	Colombia	Ecuador	El Salvador	Mexico	Paraguay	Uruguay	Venezuela
ARX	0.73,0.63,0.67	0.35,0.41,0.38	0.34,0.51,0.41	0.53,0.55,0.54	0.55,0.39,0.46	0.48,0.42,0.45	0.33,0.57,0.42	0.63,0.41,0.50
LR	0.53,0.56,0.55	0.34,**0.54**,0.41	0.21,0.69,0.32	0.51,0.53,0.52	0.30,**0.89**,0.45	0.58,0.37,0.45	0.49,0.45,0.47	0.55,0.48,0.51
KDE-LR	**1.00**,0.08,0.16	0.02,0.18,0.04	0.10,0.38,0.16	0.10,0.29,0.14	**0.93**,0.23,0.37	**1.00**,0.12,0.21	0.23,0.20,0.21	0.37,0.37,0.37
LDA-LR	**1.00**,0.08,0.16	**0.99**,0.05,0.09	0.08,**0.79**,0.15	0.08,0.32,0.12	0.94,0.23,0.37	**1.00**,0.12,0.21	0.19,0.21,0.20	0.41,0.40,0.41
LASSO	0.70,**0.67**,0.68	0.43,0.43,**0.43**	0.34,0.50,0.40	0.64,0.44,0.52	0.41,0.69,**0.52**	0.31,**0.77**,0.44	0.52,0.49,0.50	**0.64**,0.40,0.49
MTL	0.60,0.72,0.66	0.40,0.50,0.45	**0.39**,0.51,**0.44**	0.55,0.51,0.53	0.70,0.30,0.42	0.65,0.37,0.47	0.58,0.55,0.56	0.57,**0.54**,**0.55**
TMTL	0.61,0.36,0.45	0.37,0.38,0.37	0.36,0.49,0.41	**0.61**,0.51,**0.56**	0.42,0.34,0.38	0.43,0.50,0.46	0.52,0.52,0.52	0.54,0.37,0.44
MREF	0.75,0.64,**0.69**	0.36,0.51,**0.43**	0.37,0.49,0.42	0.27,**0.59**,0.37	0.35,0.77,0.49	0.58,0.41,**0.48**	**0.63**,**0.58**,**0.61**	0.53,0.42,0.47

Country Level (precision, recall, F-measure)

Method	Brazil	Colombia	Ecuador	El Salvador	Mexico	Paraguay	Uruguay	Venezuela
ARX	0.93,**1.00**,0.96	0.73,**0.97**,0.83	0.53,0.87,0.65	0.66,0.97,0.78	0.99,**1.00**,**1.00**	0.90,0.87,0.88	0.60,0.90,0.72	0.90,0.98,0.94
LR	0.95,**1.00**,0.97	0.79,**0.97**,0.87	0.56,**0.95**,0.70	0.78,0.82,0.80	**1.00**,0.98,0.99	0.89,0.97,0.93	0.63,0.93,0.75	0.92,0.96,0.94
KDE-LR	0.97,0.96,0.97	0.93,0.80,0.86	0.88,0.59,0.70	**0.85**,0.76,0.80	**1.00**,0.99,**1.00**	**1.00**,0.85,0.92	**0.97**,0.69,0.80	**1.00**,0.91,0.95
LDA-LR	0.96,0.96,0.96	**0.95**,0.82,0.88	**0.95**,0.57,0.71	0.82,0.78,0.80	0.93,**1.00**,0.96	0.91,0.92,0.91	0.94,0.70,0.80	**1.00**,0.91,0.95
LASSO	0.95,0.99,0.97	0.81,0.95,0.87	0.59,0.93,0.72	0.75,0.86,0.80	0.99,0.99,0.99	0.90,**0.99**,0.94	0.54,**0.99**,0.70	0.93,0.99,0.96
MTL	**0.98**,0.97,0.97	0.83,0.94,0.88	0.58,0.88,0.70	0.79,0.87,0.83	0.99,0.99,0.99	0.92,0.94,0.93	0.68,0.75,0.71	0.95,0.95,0.95
TMTL	0.82,0.98,0.89	0.88,0.92,**0.90**	0.67,0.87,**0.76**	0.70,0.87,0.78	**1.00**,**1.00**,**1.00**	0.94,0.98,**0.96**	0.67,0.72,0.70	0.88,**1.00**,0.94
MREF	0.97,**1.00**,**0.98**	0.86,0.94,**0.90**	0.66,0.91,**0.76**	0.76,**0.98**,**0.86**	**1.00**,**1.00**,**1.00**	0.93,**0.99**,**0.96**	0.69,0.97,**0.81**	0.96,**1.00**,**0.98**

Achieving good precision, recall and F1 score performance

Facilitating decision making

- Improving situation awareness
  - Ex 1: Finding flooding area
  - Ex 2: Spatial event discovery
- Other CIS systems & tools
  - **Critical Infrastructure Protection/Decision Support System (CIP/DSS)**
  - Urbannet toolkit & web interface
  - Other resources

CIP/DSS [Bush, 2005]

- Main focus: develop a risk-based decision support system to provide insights for making critical infrastructure protection decisions
- Covering different problems in a wide range of infrastructures:
  - Transportation, water distribution system, agriculture, banking and finance, etc.
- URL: http://public.lanl.gov/dp/CIP.html

Los Alamos NATIONAL LABORATORY

Example system: CLEAR$_{CO2}$

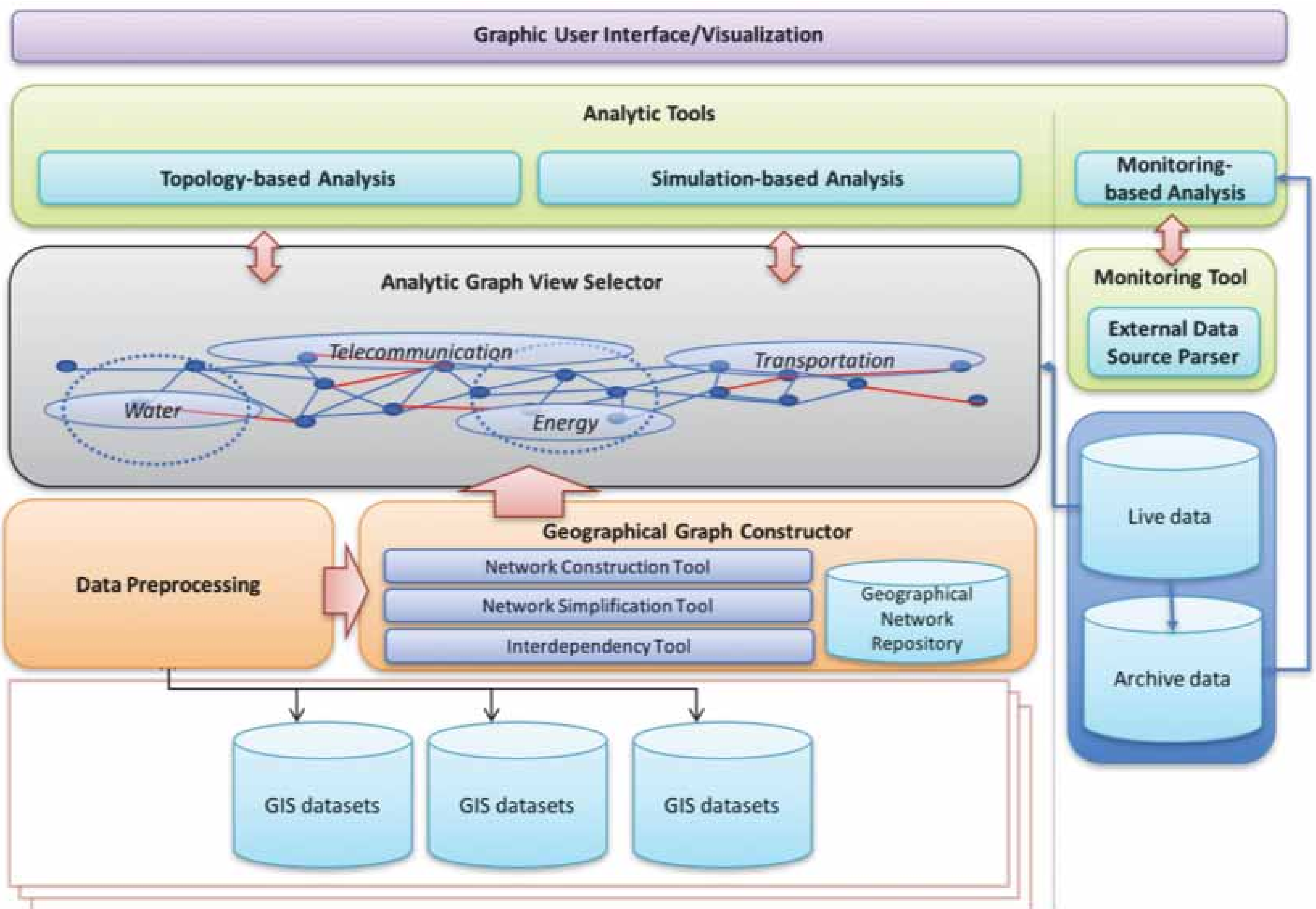
CLEAR$_{CO2}$ model interface; users can choose different energy and transportation policies in the main view (**left figure**), run the chosen scenario and look at the simulation in real time (**right figure**); the main view also allows to compare different scenarios.

Facilitating decision makings

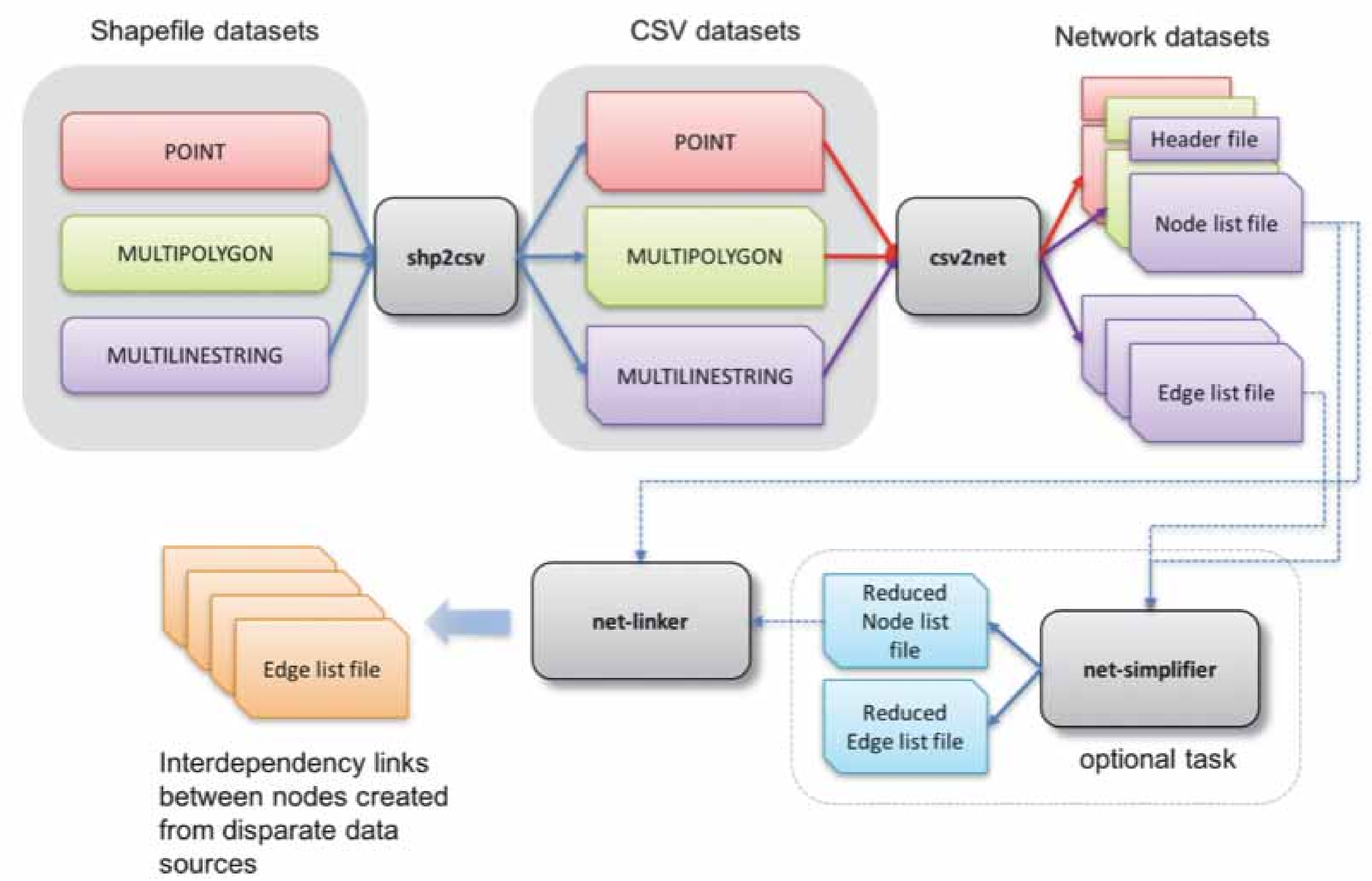
- Improving situation awareness
  - Ex 1: Finding flooding area
  - Ex 2: Spatial event discovery
- Other CIS systems & tools
  - Critical Infrastructure Protection/Decision Support System (CIP/DSS)
  - **Urbannet toolkit & web interface**
  - Other resources

URBANNET [Lee+ Big Data'16, Chen+ CIKM 2017]

- A system to generate networks for CIS



Data processing pipeline

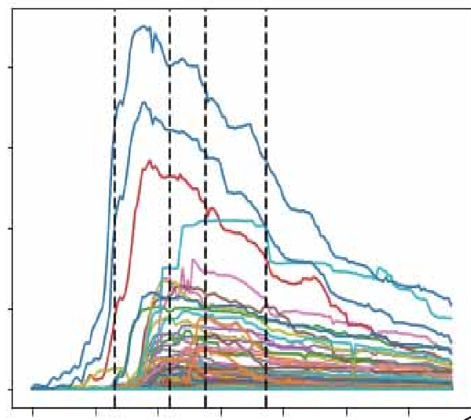


Urbannet

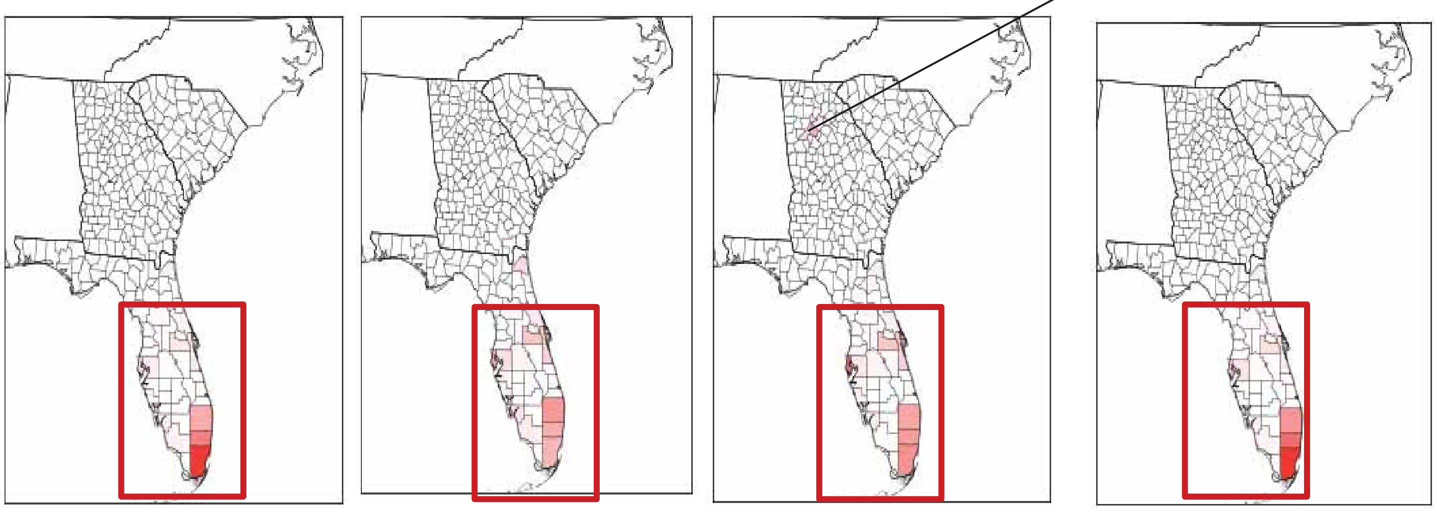
- A licensed (ORNL and VT) toolkit that integrates
  - Network construction
  - CIS visualization
  - Failure cascade modeling
  - HotSpots algorithm to identify critical facilities
  - Scenario generator & simulator

Additions: Provide actionable insights in emergency management

Hurricane Irma power outage data



Get the middle counties which faced damage because of a storm (not hurricane)



***Rationalizing* AI**

Red rectangle shows the affected counties due to hurricane and their restoration period

Facilitating decision makings

- Improving situation awareness
  - Ex 1: Finding flooding area
  - Ex 2: Spatial event discovery
- Other CIS systems & tools
  - Critical Infrastructure Protection/Decision Support System (CIP/DSS)
  - Urbannet toolkit & web interface
  - **Other resources**

Other resources

- **HSIP gold**
  - A unified infrastructure geospatial data inventory, which includes domestic infrastructure datasets collected from various government agencies and partners
  - URL: https://gii.dhs.gov/HIFLD/hsip-guest
- NHDplus
- EIA
- USGS water data
- …

Other resources

- HSIP gold
- **NHDplus**
  - A dataset created by the US Environmental Protection Agency (EPA), which includes information about the nation's hydrological framework
  - URL: http://www.horizon-systems.com/nhdplus/
- EIA
- USGS water data
- …

Other resources

- HSIP gold
- NHDplus
- **EIA**
  - Open source Energy datasets from US Energy Information Administration
  - URL: https://www.eia.gov/
- USGS water data
- …

Other resources

- HSIP gold
- NHDplus
- EIA
- **USGS water data**
  - Provide real time stream flow data across the nation
  - URL: https://waterdata.usgs.gov/nwis/rt
- …

Outline

- Introduction
  - Data (network and sequence) mining challenges in CI systems
- Part 1: Power Systems
  - Identifying and protecting against vulnerability in power networks
- Part 2: Transportation Systems
  - Traffic states/flow prediction and control
- Part 3: Decision Making
  - Tools for facilitating decisions
- Conclusion

Urban computing

- Many problems and challenges in big cities

Q1: Smart grid

Q2: Urban flow

Q3: Situation awareness

Q4: Robustness & Evolution

Q5: Public Health

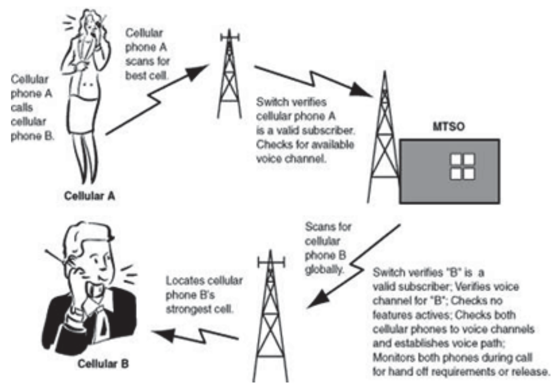
Q6: Air pollution

VirginiaTech

Critical Infrastructure Systems

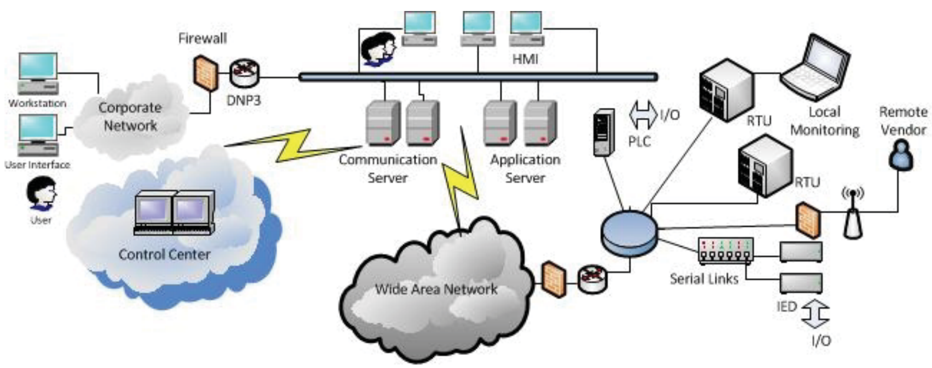
- Vital to our national security, economy.

Transportation System

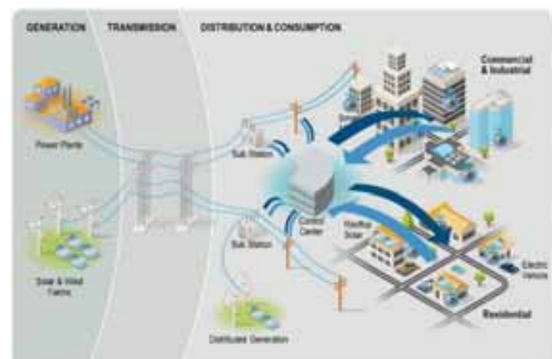


Cellular System

Water System



Cyber System

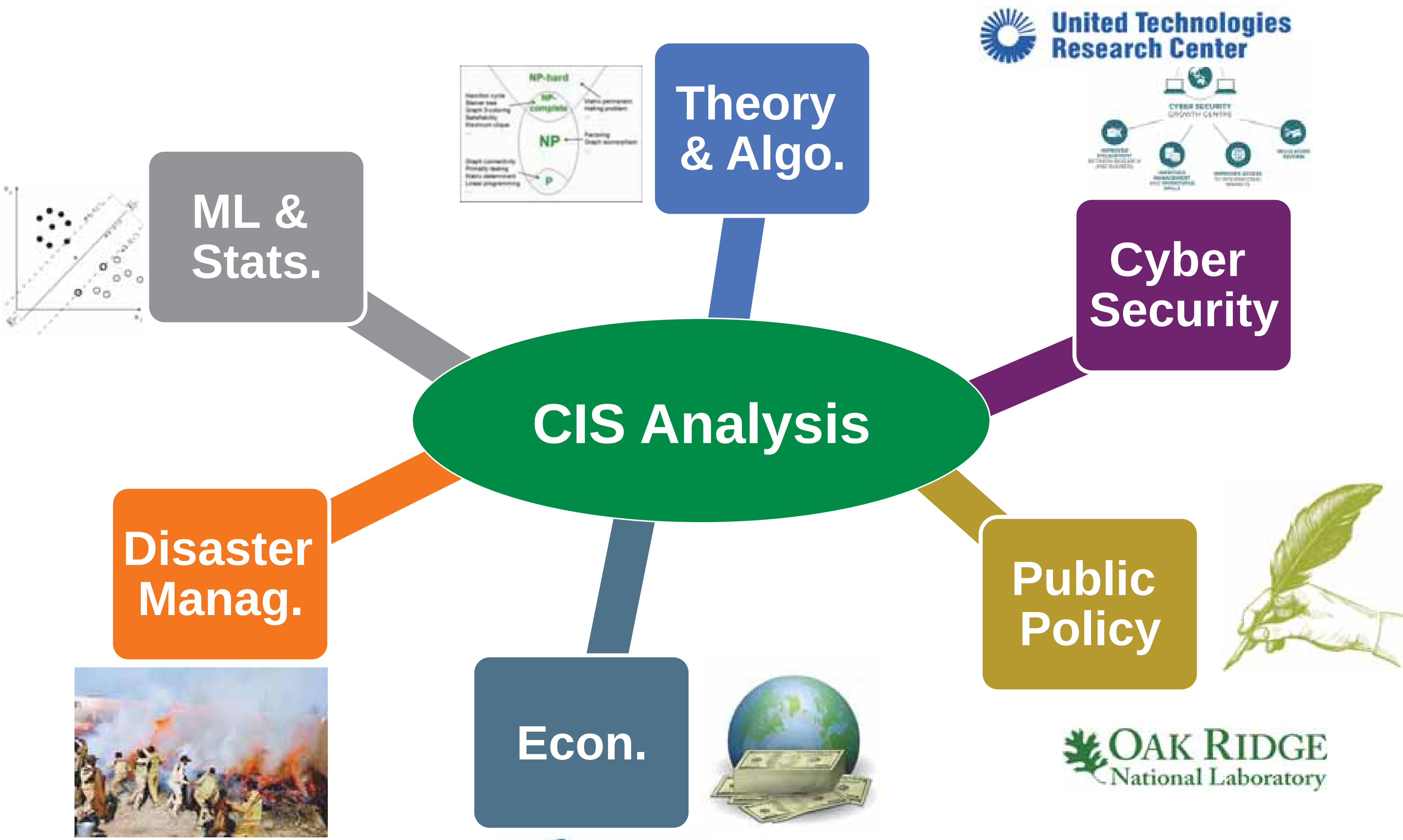


Electric Grid System

VirginiaTech

Conclusions

- Many important problems in CIS for data miners
- Complex system dynamics, unknown system interdependencies pose huge challenges to traditional approaches
- An open domain with many opportunities!



Microsoft Research

Los Alamos NATIONAL LABORATORY EST. 1943

VirginiaTech

Acknowledgements

Funding

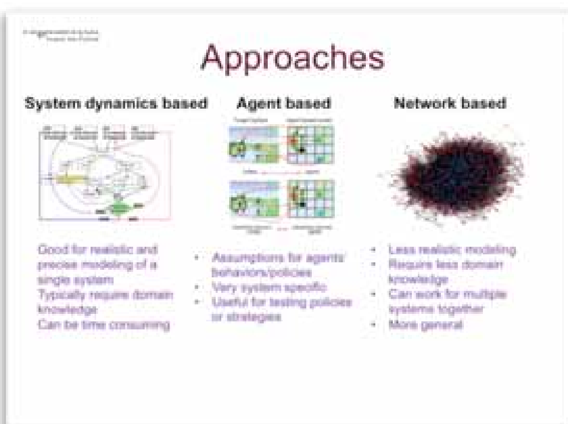
Critical Infrastructure Data Analytics: Models and Tools

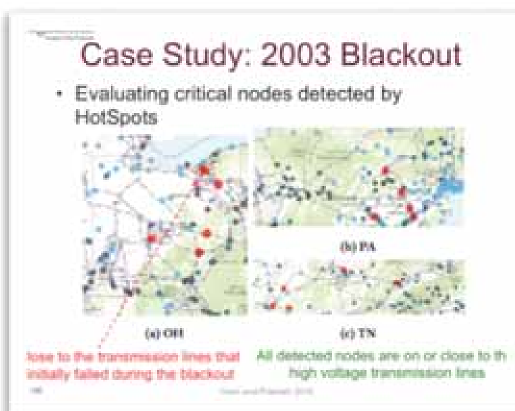
B. Aditya Prakash

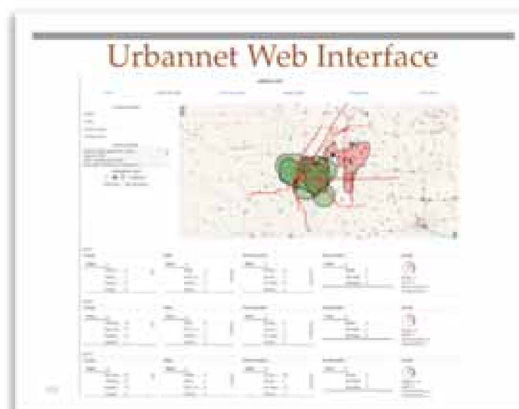
Modeling

Algorithms

Tools







References: Intro (1)

1. Weng, Yang, et al. "Robust data-driven state estimation for smart grid." IEEE Transactions on Smart Grid 8.4 (2017): 1956-1967.
2. Zhang, Junbo, Yu Zheng, and Dekang Qi. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction." AAAI. 2017.
3. Wu, Yao-Jan, et al. "Traffic flow prediction for urban network using spatio-temporal random effects model." 91st Annual Meeting of the Transportation Research Board (TRB). 2012.
4. Liu, Dongyu, et al. "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations." IEEE transactions on visualization and computer graphics 23.1 (2017): 1-10.
5. Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors." Proceedings of the 19th international conference on World wide web. ACM, 2010.

References: Intro (2)

6. Yang, Seungwon, et al. "PhaseVis1: What, when, where, and who in visualizing the four phases of emergency management through the lens of social media." ISCRAM. 2013.
7. Xu, Jian, et al. "Improving management of aquatic invasions by integrating shipping network, ecological, and environmental data: data mining for social good." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.
8. Zheng, Yu, Furui Liu, and Hsun-Ping Hsieh. "U-air: When urban air quality inference meets big data." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

References: Part 1 (1)

1. Ouyang, Min. "Review on modeling and simulation of interdependent critical infrastructure systems." Reliability engineering & System safety 121 (2014): 43-60
2. O'Reilly, Gerard P., et al. "Telecom Critical Infrastructure Simulations: Discrete-Event Simulation vs. Dynamic Simulation How Do They Compare?." Global Telecommunications Conference, 2007. GLOBECOM'07. IEEE. IEEE, 2007.
3. Brown, Theresa. "Multiple modeling approaches and insights for critical infrastructure protection." NATO Security through Science Series D-Information and Communication Security 13 (2007): 23.
4. Kim, Seul-Ki, et al. "Dynamic modeling and control of a grid-connected hybrid generation system with versatile power transfer." IEEE transactions on industrial electronics 55.4 (2008): 1677-1688.
5. Xu, X., et al. "Dynamic modeling and interaction of hybrid natural gas and electricity supply system in microgrid." IEEE Transactions on Power Systems 30.3 (2015): 1212-1221.

References: Part 1 (2)

6. Song, Hyun Ah, et al. "PowerCast: Mining and forecasting power grid sequences." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2017.
7. Basu, Nipa, R. Pryor, and Tom Quint. "ASPEN: A microsimulation model of the economy." Computational Economics 12.3 (1998): 223-241.
8. Barton, Dianne C., et al. "Aspen-EE: an agent-based model of infrastructure interdependency." SAND2000-2925. Albuquerque, NM: Sandia National Laboratories (2000).
9. Hopkinson, Kenneth, et al. "EPOCHS: a platform for agent-based electric power and communication simulation built from commercial off-the-shelf components." IEEE Transactions on Power Systems 21.2 (2006): 548-558.
10. Balmer, Michael, et al. "Towards truly agent-based traffic and mobility simulations." Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1. IEEE Computer Society, 2004.

References: Part 1 (3)

11. Lee, Sangkeun, et al. "URBAN-NET: A network-based infrastructure monitoring and analysis system for emergency management and public safety." Big Data (Big Data), 2016 IEEE International Conference on. IEEE, 2016.
12. Chen, Chen, et al. "FASCINATE: Fast cross-layer dependency inference on multi-layered networks." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

References: Part 2

1. Arianos, Sergio, et al. "Power grid vulnerability: A complex network approach." Chaos: An Interdisciplinary Journal of Nonlinear Science 19.1 (2009): 013119.
2. Chen, Chen, et al. "On the connectivity of multi-layered networks: Models, measures and optimal control." Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015.
3. Chen, Liangzhe, et al. "HotSpots: Failure Cascades on Heterogeneous Critical Infrastructure Networks." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017.
4. Buldyrev, Sergey V., et al. "Catastrophic cascade of failures in interdependent networks." Nature 464.7291 (2010): 1025.
5. Wang, Jian-Wei, and Li-Li Rong. "Cascade-based attack vulnerability on the US power grid." Safety Science 47.10 (2009): 1332-1336.
6. Huang, Xuqing, et al. "Robustness of interdependent networks under targeted attack." Physical Review E 83.6 (2011): 065101.
7. Ouyang, Min, and Yiping Fang. "A mathematical framework to optimize critical infrastructure resilience against intentional attacks." Computer-Aided Civil and Infrastructure Engineering 32.11 (2017): 909-929.

References: Part 3 (1)

1. Anwar, Tarique, et al. "Roadrank: Traffic diffusion and influence estimation in dynamic urban road networks." Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015.
2. Wu, Yao-Jan, et al. "Urban traffic flow prediction using a spatio-temporal random effects model." Journal of Intelligent Transportation Systems 20.3 (2016): 282-293.
3. Moretti, Fabio, et al. "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling." Neurocomputing 167 (2015): 3-7.
4. Zheng, Zimu, et al. "Urban traffic prediction through the second use of inexpensive big data from buildings." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
5. Gu, Yiming, Zhen Qian, and Guohui Zhang. "Traffic State Estimation for Urban Road Networks Using a Link Queue Model." Transportation Research Record: Journal of the Transportation Research Board 2623 (2017): 29-39.

References: Part 3 (2)

6. Adhikari, Bijaya, et al. “Near-optimal Mapping of Network States using Probes.” SIAM International Conference on Data Mining (SDM 2018)
7. Anwar, Tarique, et al. "Tracking the evolution of congestion in dynamic urban road networks." Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016.
8. Sundar, Rajeshwari, Santhoshs Hebbar, and Varaprasad Golla. "Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection." IEEE Sensors Journal 15.2 (2015): 1109-1113.

References: Part 4 (1)

1. Liang, Jiongqian, Peter Jacobs, and S. Parthasararthy. "Human-guided flood mapping on satellite images." WWW 2018, Lyon.
2. Huang, Chao, Xian Wu, and Dong Wang. "Crowdsourcing-based urban anomaly prediction system for smart cities." Proceedings of the 25th ACM international on conference on information and knowledge management. ACM, 2016.
3. McClendon, Susan, and Anthony C. Robinson. "Leveraging geospatially-oriented social media communications in disaster response." International Journal of Information Systems for Crisis Response and Management (IJISCRAM) 5.1 (2013): 22-40.
4. Muralidhar, Nikhil, et al. "illiad: InteLLigent Invariant and Anomaly Detection in Cyber-Physical Systems." ACM Transactions on Intelligent Systems and Technology (TIST) 9.3 (2018): 35.
5. Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors." Proceedings of the 19th international conference on World wide web. ACM, 2010.

References: Part 4 (2)

6. Zhao, Liang, et al. "Multi-resolution spatial event forecasting in social media." Data Mining (ICDM), 2016 IEEE 16th International Conference on. IEEE, 2016.
7. Lee, Sangkeun, et al. "URBAN-NET: A network-based infrastructure monitoring and analysis system for emergency management and public safety." Big Data (Big Data), 2016 IEEE International Conference on. IEEE, 2016.
8. Chen, Liangzhe, et al. "HotSpots: Failure Cascades on Heterogeneous Critical Infrastructure Networks." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017.