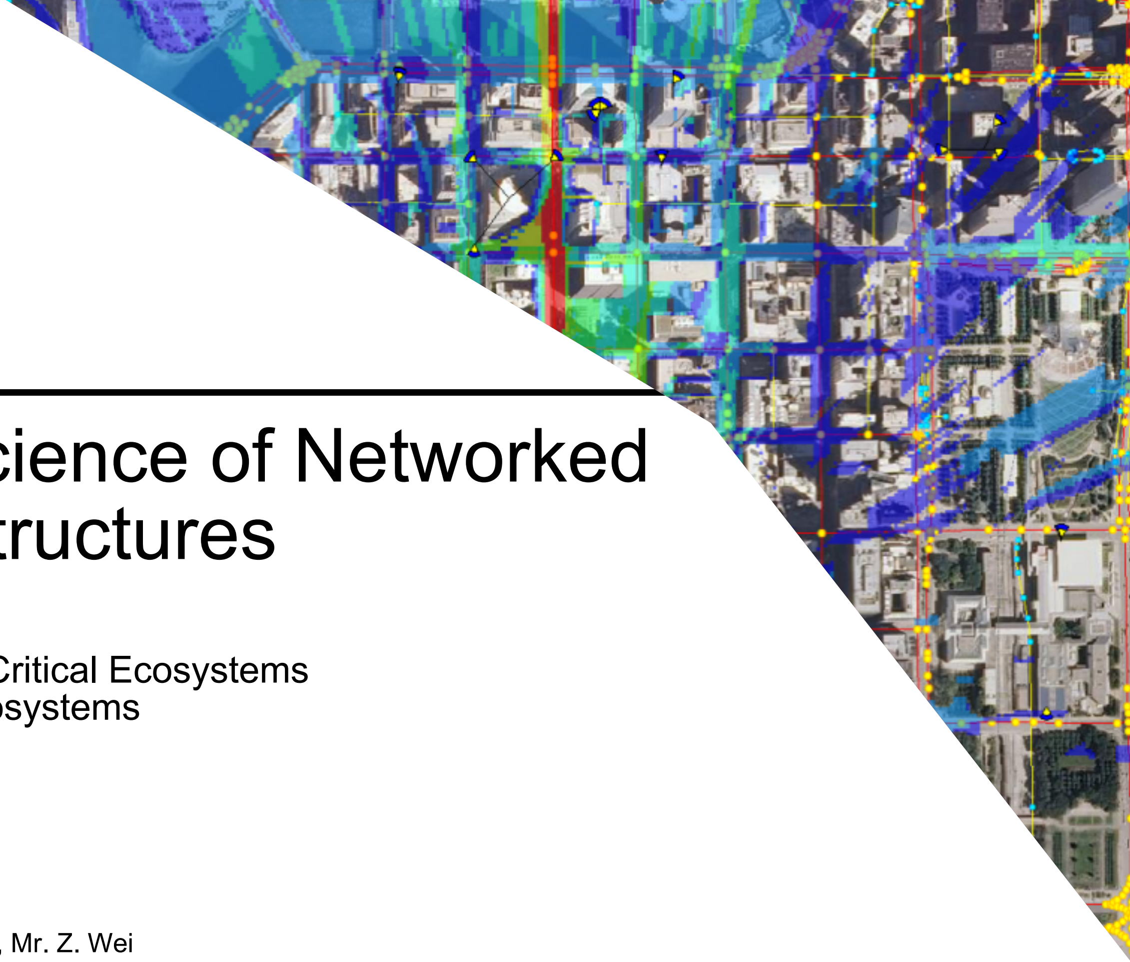
Resilience & Data Science of Networked Cyberphysical Infrastructures



CHANCE - Coupled Human And Natural Critical Ecosystems
CoTRE - Complexity Twin to Resilient Ecosystems

Dr. Weisi Guo
Alan Turing Institute and University of Warwick
Cranfield University (fall 2019)

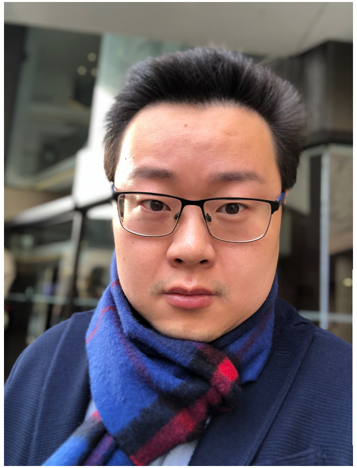
Thanks to team members: Dr. G. Moutsinas, Dr. A. Pagani, Mr. Z. Wei

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Brief Introduction

Background

- Cambridge: MEng General Engineering (2001-05)
- Cambridge: PhD Computer Science (2007-10)
- T-Mobile International: Radio Network Engineer (2005-07)
- Sheffield Uni: Post-Doc (2010-12)
- Warwick Uni: Assistant Prof (2012-17) Associate Prof (2017-19)
- Cranfield Uni: Chair in Human Machine Intelligence (2019-)



DEN Lab
Data Embedded Networks Lab

The Alan Turing Institute

NYU

CUSP
CENTER FOR URBAN SCIENCE+PROGRESS

Interests:

- Networked Dynamics, Machine Learning for 5G, Molecular Signal Processing

Funders & Stakeholders:

- EPSRC, H2020, InnovateUK, DSTL/MOD/GCHQ, USAF, LRF, Royal Society, British Council
- National Infra. Comm., LEP, IEEE Standardization
- UK and EU SMEs

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Outline of Talk

1. Motivation & Review of Networked Resilience
2. Functional Analysis of Resilience with Uncertainty
3. Data Driven Analysis of Resilience
4. Optimal Sampling of Networked Dynamics via Graph Fourier Transform

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Critical Infrastructure & Ecosystems

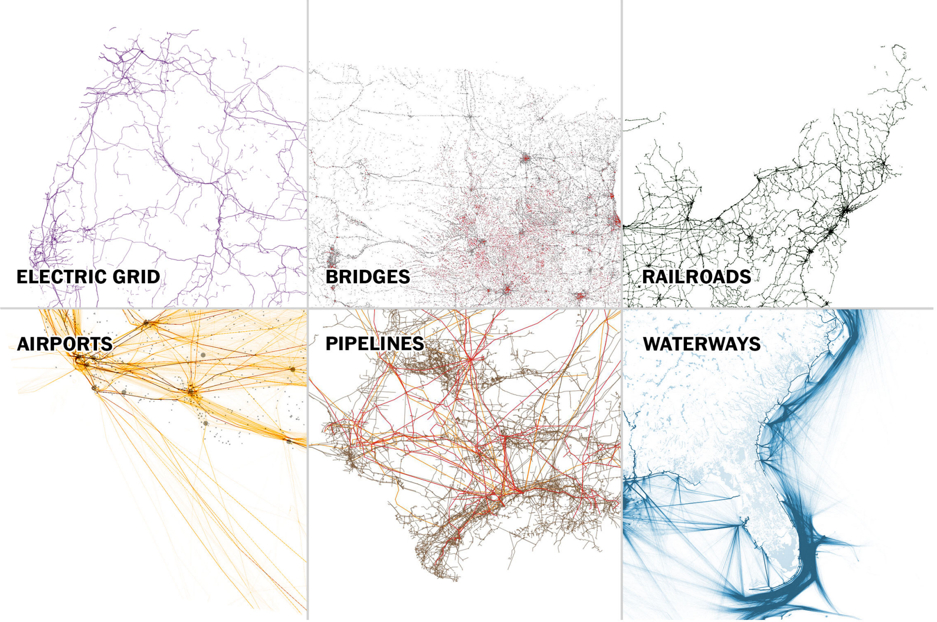
Many of our built infrastructures are networked together:

- Water Supply
- Transportation
- Electricity Supply
- Telecommunications

They combine local **functional elements** with interdependent **coupling elements**. Together, they form the backbone of our modern civilization, providing services to billions.

These often sit alongside natural ecosystems with network dimensions:

- Food Webs
- Organizational Structure
- Gene Regulation



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Part 1/4.
Networked Resilience arises from both Local Dynamics & Global Topology

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Challenge of Cascades on Networks

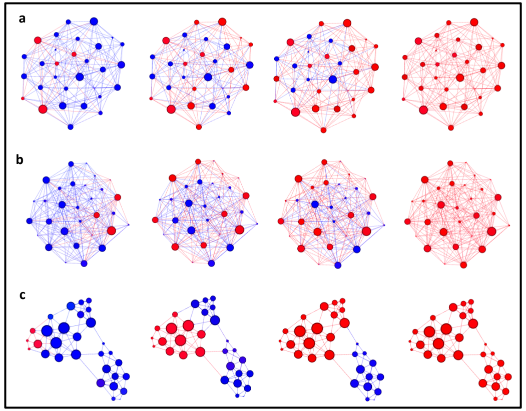
Background: Cascade effects (undesirable, not designed) are a common problem on networked systems.

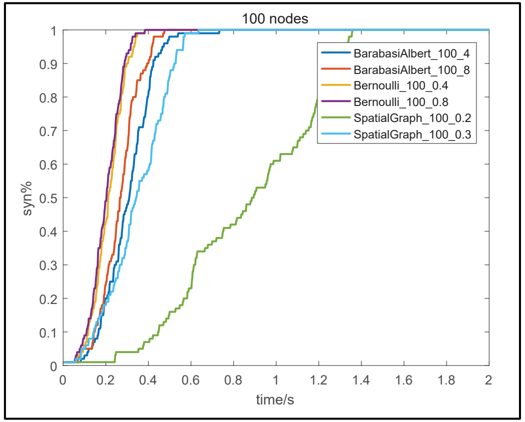
Examples: virus (cyber and epidemic), pollution, false information, circadian clock (see right)…etc.

Literature: we know how effects spread, but we don't know that well how it affects and is affected by the performance and complex behaviour of the individual components.

Goal: What we want to understand is their individual resilience behaviour and their coupled resilience behaviour.

Challenge: The dimensionality of the problem is huge (networks with millions of nodes), and the behaviour can be complex models. How do we gain meaningful insight?





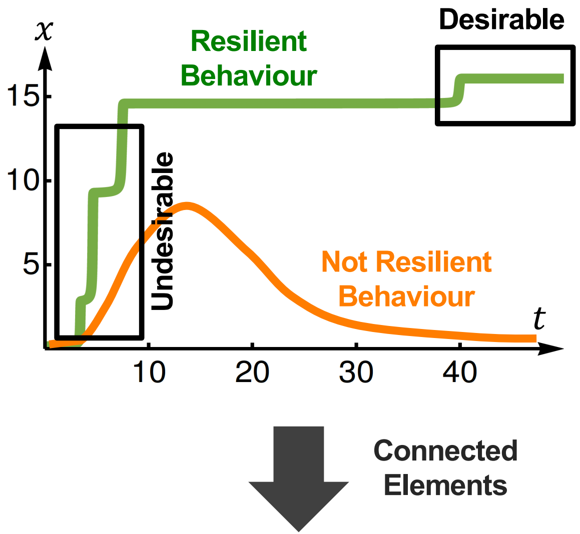
Examples of Cascades & Failures

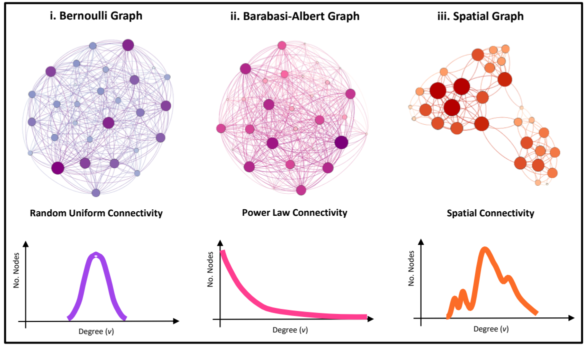
Continuous and Discrete: Cascade effects can be continuously affecting the network (pollution) or discretely breaking down the network (attacks cause failure).

Node Centrality: we know node centrality (eigenvector centrality, PageRank…etc.) identifies nodes with greatest influence to spread cascades.

Graph Structure: we also know that targeted attacks against high degree nodes is effective, and random attacks against small-world networks is ineffective.

Rewiring: more recently, we know how mesoscale core-periphery structure can best preserve integrity of a network when subject to attacks and under a finite energy budget.





CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

What is Networked Resilience & Robustness

1. Simple Graphs
2. Dynamical Systems

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

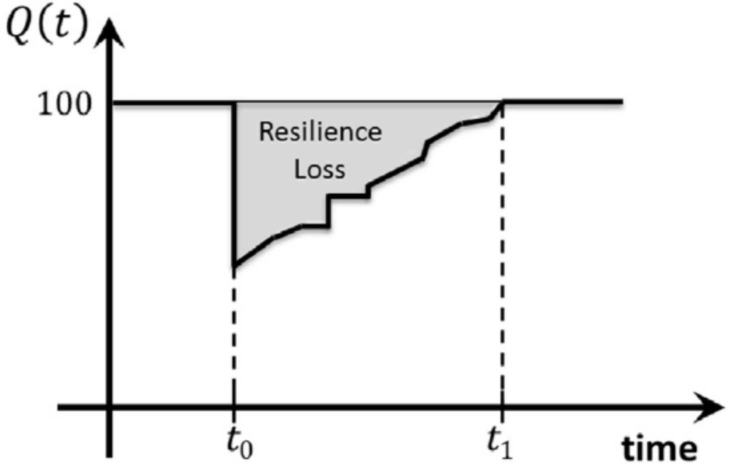
Brief Review

A review of definitions and measures of system resilience

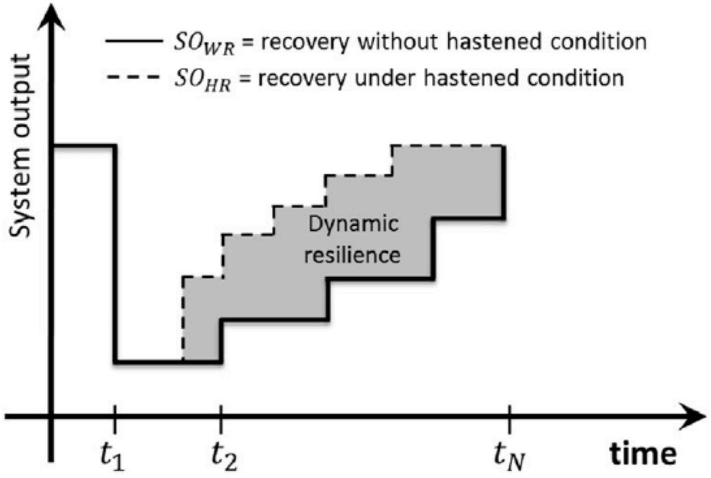
Seyedmohsen Hosseini [a], Kash Barker [a,*], Jose E. Ramirez-Marquez [b,c]

[a] School of Industrial and Systems Engineering, University of Oklahoma, United States
[b] School of Systems and Enterprises, Stevens Institute of Technology, United States
[c] Tec de Monterrey, School of Science and Engineering, Zapopan Guadalajara, Mexico

Bouncing Back: systems that return to their operating state after a negative shock is the commonality in resilient behaviour. Rate of return, asymptotic convergence, are all important metrics.



Hasten & Bounce Forward using Networks: couplings on the network can help individuals bounce back faster, as well as move forward by connecting them (rewiring) to new elements that can improve their resilience in the future.

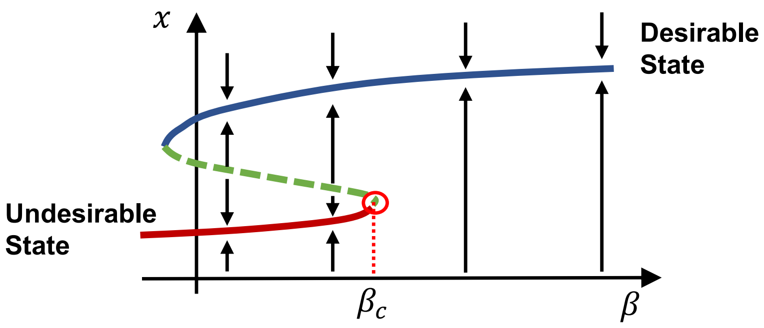


CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Resilience

Resilience is important as systems constantly face stressors (demand spikes) and perturbations.

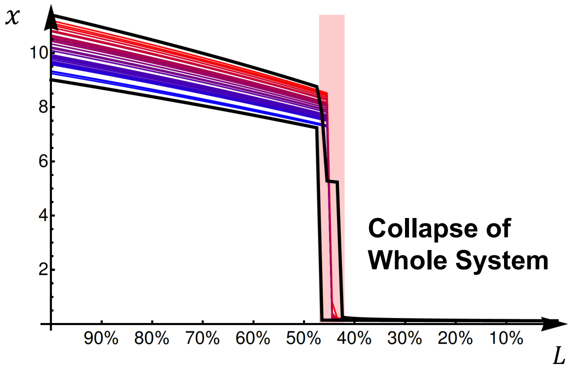
Each system has a desirable equilibrium state and want to avoid undesirable states. Cascades can cause system wide poor performance.



Robustness

Robustness is important as systems constantly face complete failures at the sub-system or connector level.

Each sub-system requires connectivity to function properly. Failures can lead to cascade failures.



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Brief Review of Resilience

Stability: defined as return to a desirable equilibrium after some perturbation or shock.

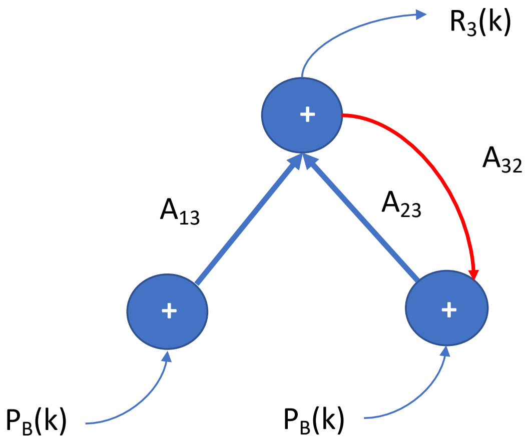
Random Graphs: In random graphs, it was shown (May 1972) that instability (largest eigenvalue) scales with size of the network (*N*) and average connectivity (*C*):

$$\propto \sqrt{NC}$$

Later expanded to random graphs with delays.

Linear Dynamics on Small Structured Graphs with Defined I-O: we know that linear stability is defined by the largest root of the transfer function.

Large Structured Graphs with No I-O: we don't know. So we currently check no. loops / trophic coherence (2^{nd} part of talk), but we can also develop some new theories (1^{st} part).



Trophic coherence determines food-web stability

Samuel Johnson, Virginia Domínguez-García, Luca Donetti, and Miguel A. Muñoz

PNAS December 16, 2014 111 (50) 17923-17928; published ahead of print December 2, 2014
https://doi.org/10.1073/pnas.1409077111

Looplessness in networks is linked to trophic coherence

Samuel Johnson and Nick S. Jones

PNAS published ahead of print May 16, 2017 https://doi.org/10.1073/pnas.1613786114

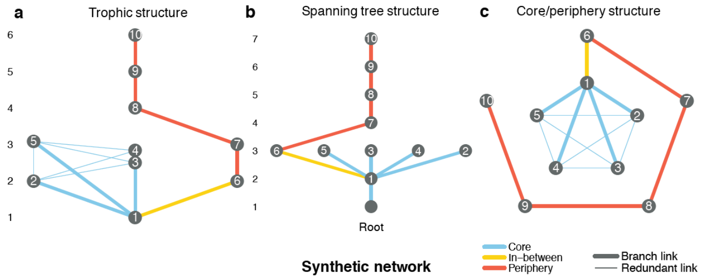
Brief Review of Robustness

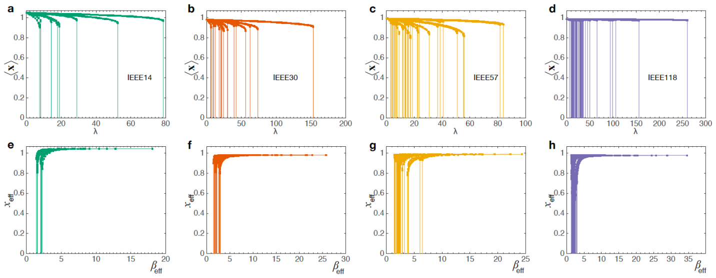
Attack/Removal Process: random removal/targeted removal leads to slow destruction of network.

Failure Process: loss of complete connectivity = loss of functionality (assumes coupling determines function).

Relation to Network Structure: easy to see the role of network topology on the overall performance (e.g. small world network robust to random attacks, vulnerable to targeted attacks).

Cascades on Electricity Networks: cascade failure on electricity grid networks (Gao et al. Nature: SI 2016).





CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

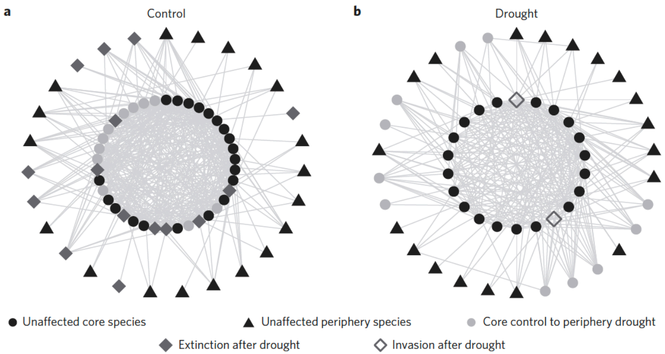
Rewiring for Improvement

Rewiring: preserving mesoscale community structure can preserve continuous robustness against further removal (Lu et al. Nature Climate Change 2016).

Rich club of connections makes unforeseen attacks less likely to lead to cascade failures.

Limited Capture of Local Connections: a pure connectivity analysis leads to conclusions that weakly entities are not important (e.g. Malaria fly can be eliminated from ecosystem without cascade damage)…

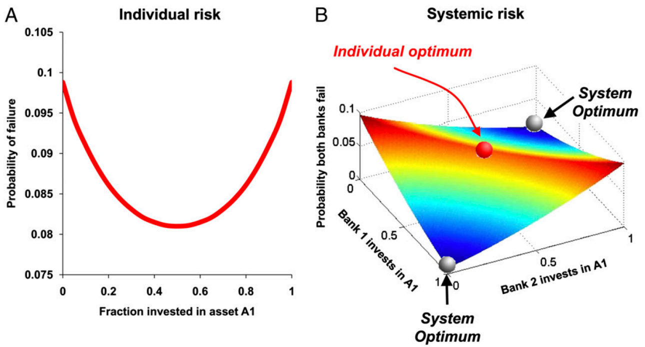
Regulatory Paradox: a small at a peripheral part can upset the whole ecosystem in the <u>opposite way</u>, e.g. saving a small bank can collapse the the whole financial system (May et al. PNAS 2011).



Individual versus systemic risk and the Regulator's Dilemma

Nicholas Beale[a,1], David G. Rand[b,1], Heather Battey[c], Karen Croxson[d,2], Robert M. May[e], and Martin A. Nowak[b,f,3]

[a]Sciteb, London W1B 4BD, United Kingdom; [b]Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138; [c]Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom; [d]New College and Oxford–Man Institute of Quantitative Finance, University of Oxford, Oxford OX1 3DW, United Kingdom; [e]Department of Zoology, University of Oxford, Oxford OX1 3DW, United Kingdom; and [f]Department of Mathematics and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Gap in Knowledge

1. Just Networks: Current knowledge is mostly limited to weighted or unweighted graphs (they may change over time).

2. The nodes and edges do not have dynamical behaviours.

3. Dynamics Matter: many (most?!) dynamical systems have non-linear behaviours and multiple equilibrium states.

4. In stable equilibriums, they cannot be easily disturbed, whereas in unstable equilibriums, they can be easily disturbed.

5. This already allows you to understand how modeling dynamics is important.

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

What do we know already?

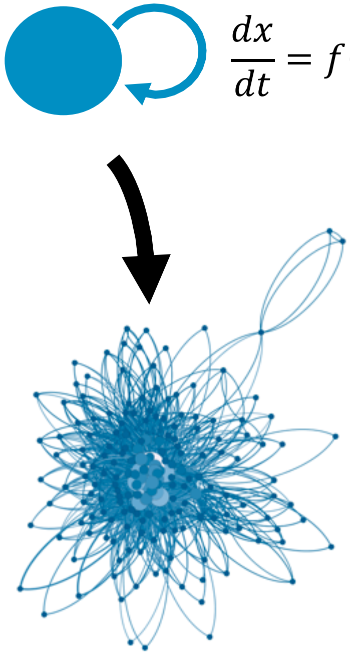
Engineers and ecologists have a good understanding of:

- Local Dynamics & Governing Equations
- Stability and Control
- Data and Experience
- Stressors & Perturbations

But, as we connect systems in increasingly larger networks, we **don't know**:

- Gives Insight over Pure Predictive Approaches (ARIMAs, HMMs, DGPs, CNNs).
- Relationship between: Topology and Dynamics
- How Local Effects affect Network Wide Cascades

$$\frac{dx}{dt} = f(x, \beta)$$



$$\frac{dx_i}{dt} = f(x_i) + \sum_{j=1}^{N} a_{ij}\, g(x_i, x_j)$$

Open Questions:

Is system resilience more sensitive to <u>network topology</u> or <u>component dynamics</u>?
How can this knowledge inform the design of new critical infrastructure systems?
What are the wider applications of this framework (ecology, biology, society)?

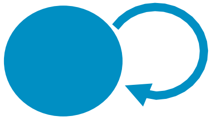
CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Part 2/4: Functional Analysis with Uncertainty Quantification

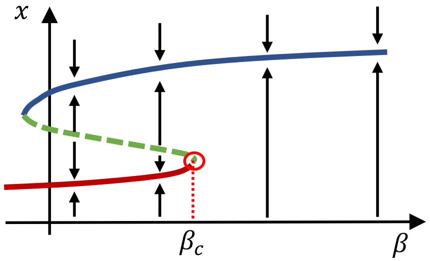
1. Telecommunications (Engineered)
2. Bee Pollination (Natural)
3. Gene Regulation (Natural)

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

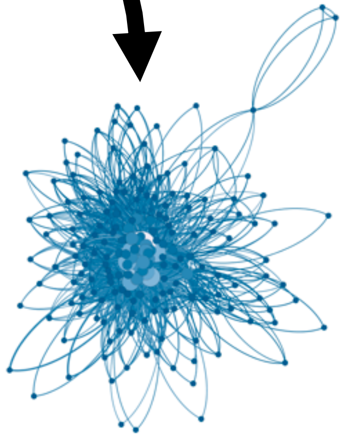
(i) Coupled Dynamics in a Complex Network



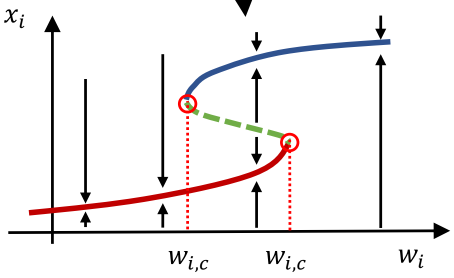
$$\frac{dx}{dt} = f(x, \beta)$$



a. Resilience Function of an Isolated Node in Terms of Parameter β



b. Coupling Individual Dynamics Shifts their Resilience Functions



c. Approximating the Resilience Function of a Connected Node in Terms of its Weighted Degree w_i

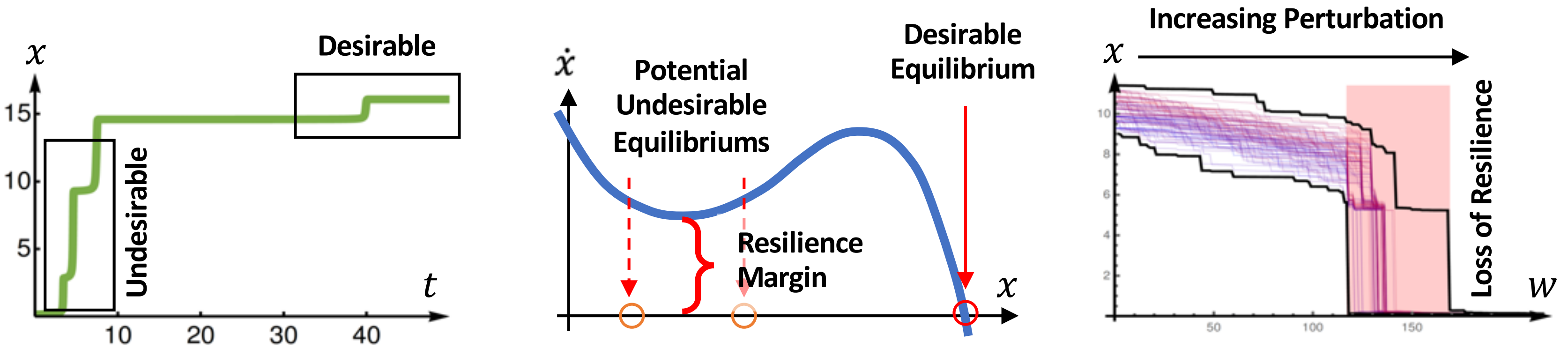
$$\frac{dx_i}{dt} = f(x_i) + \sum_{j=1}^{N} a_{ij}\, g(x_i, x_j)$$

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Mapping Resilience to Robustness

Familiar notions of dynamic time response (bounce back) is mapped to changing equilibrium states and a resilience margin. Networked systems can have cascade dynamics (resilience), but when it causes cascade unrecoverable failures (robustness).

(ii) Characteristic Functions



a. Dynamic Response: Recovery of Resilience

b. Rate Dynamics: Equilibriums Shift with Perturbations

c. Resilience Function: Under Perturbations (*w*)

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Current Literature

Discovering the explicit relationship between:

- **Average Network Dynamics** as a function:
- Local Node Dynamics
- Network Topology

This was applied to a variety of ecological and biological dynamics in 2 key papers (Nature Physics 2013) and (Nature 2016).

Basic idea is to develop a mean field approximation.

LETTER

doi:10.1038/nature16948

Universal resilience patterns in complex networks

Jianxi Gao[1*], Baruch Barzel[2*] & Albert-László Barabási[1,3,4,5]

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F(x_i) + \sum_{j=1}^{N} A_{ij} G(x_i, x_j)$$

↓

$$f(\beta_{\mathrm{eff}}, x_{\mathrm{eff}}) = F(x_{\mathrm{eff}}) + \beta_{\mathrm{eff}} G(x_{\mathrm{eff}}, x_{\mathrm{eff}})$$

Topology Mapped to Dynamics (Network Average)

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

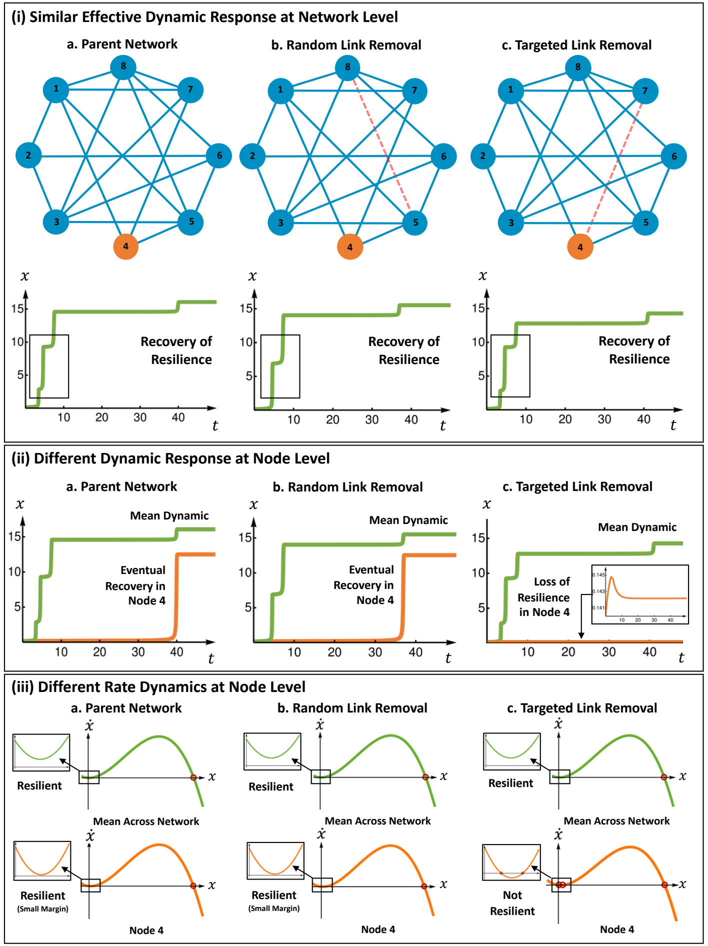
Mean Network Dynamics Hides Node Level Behaviour

Near identical networks and dynamics can hide different node level dynamics.

Here we show how the resilience of node 4 can vary between being resilient (bouncing back) to collapsing.

The overall network dynamics (Nat. Phy.13 & Nat16 papers) predicts would be the same. Indeed, one expects mean field to give similar expectations.

What we wish to do is to improve on this and give node level accurate predictions, because most interventions are made at the node level.



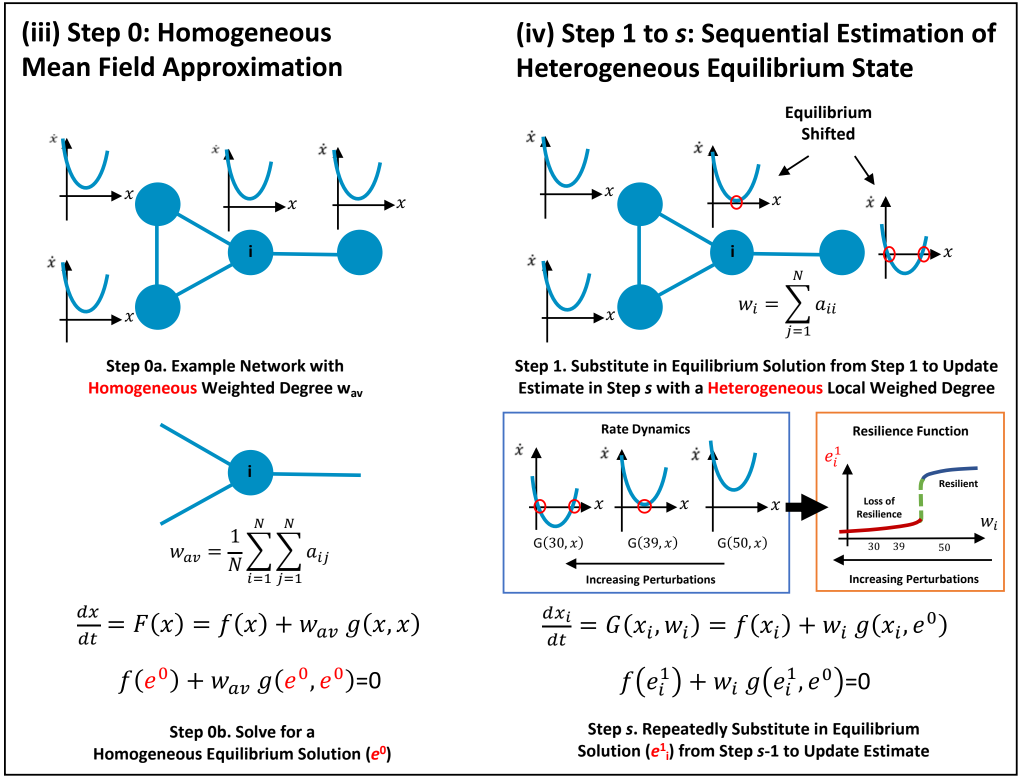
CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Sequential Estimation: Heterogeneous Mean Field

We first take a homogeneous estimate to give a mean field understanding of equilibrium states. The trick here is finding a network wide topological measure.

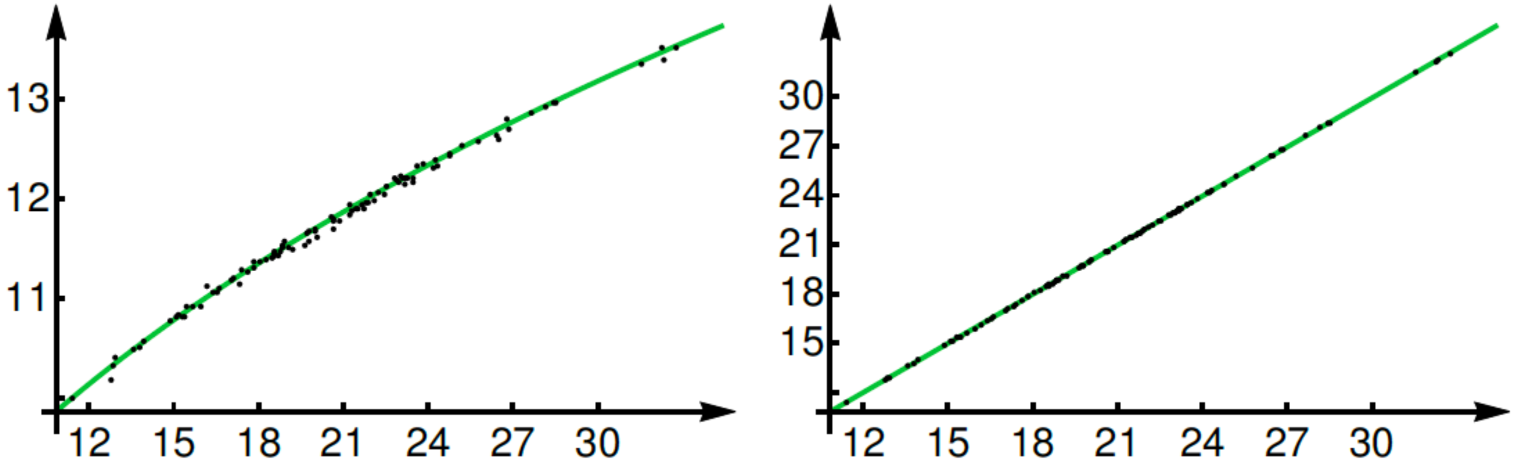
We then iteratively substitute this back into the network using local network measures to create heterogeneous solutions.

Improves over current methods [1] by giving node level prediction, which helps to inform action [2].



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Accuracy of Equilibrium on Two Networks



(a) Mutualistic interactions dynamics. (b) Gene regulatory dynamics.

Fig. 3: The first order approximation of two dynamical systems on the same Erdös-Rényi graph with 100 vertices and $p = 0.2$. The horizontal axis is the weighted in-degree of a vertex, w_i, and the vertical axis is the value of the equilibrium at this vertex. The equilibrium computed numerically is shown in black and the blue line is the graph of the function $\chi^{\{1\}}$.

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Case Study: Telecommunications (Wireless Load Balancing)

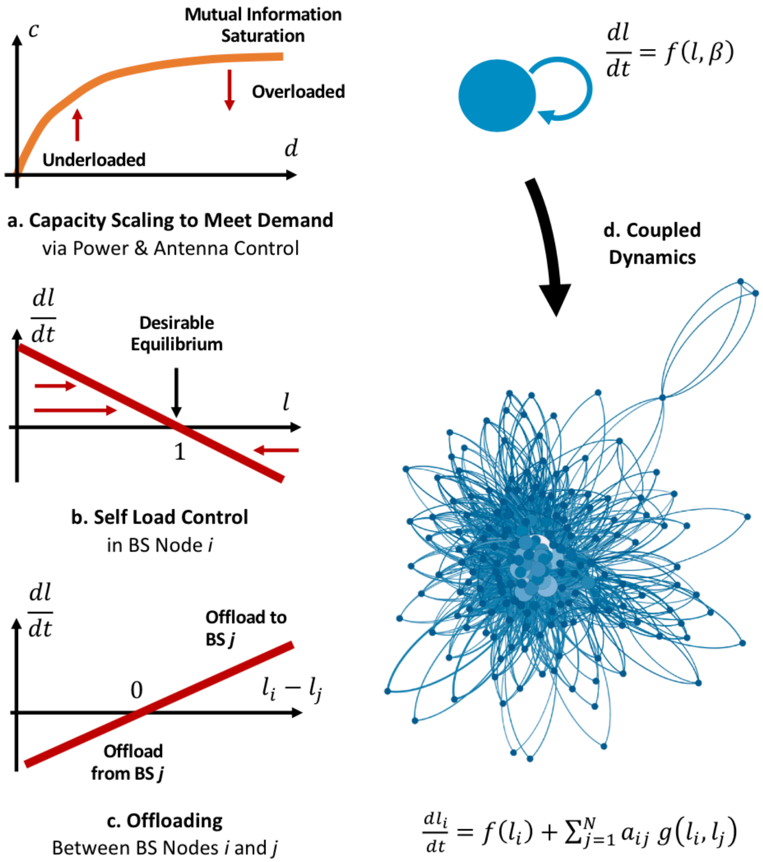
Here we study a mobile network, whereby the load demand (*l*) dynamics is governed by [7]:

- Load balancing inside a cell (capacity scaling using adaptive modulation coding schemes / power control / antenna switching)
- Load balancing between coupled cells
- Data demand from consumers

The load dynamics in each cell can be described by its own attempt to satisfy demand (RHS 1st term) and the coupling with other cells (RHS 2nd term).

$$\dot{l}_i = f(l_i) + \sum_{j=1}^{N} a_{ji} g(l_j - l_i),$$

$$\dot{l}_i = \beta(1 - l_i) + \sum_{j=1}^{N} a_{ji} \alpha(l_j - l_i),$$



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Stability Criteria for Load and Capacity Dynamics

A. Stability

In order to determine the stability of the equilibrium we compute the eigenvalues of the Jacobian at the equilibrium. Let F_i be the i-th component of the function F of equation (4), then we have

$$\begin{aligned}\frac{\partial}{\partial l_i} F_i(L)\Big|_{L=r\mathbf{1}} &= f'(r) - \sum_{j=1}^{N} a_{ji} g'(0) \\ &= f'(r) - \alpha \sum_{j=1}^{N} a_{ji} \\ &= f'(r) - \alpha w_i.\end{aligned} \tag{6}$$

where we define $w_i = \sum_{j=1}^N a_{ji}$.

When $k \neq i$ we have

$$\frac{\partial}{\partial l_k} F_i(L)\Big|_{L=r\mathbf{1}} = \sum_{j=1}^{N} \delta_{jk} a_{ji} g'(0) = \alpha a_{ki},$$

where δ_{ki} is the Kronecker delta. This equation together with equation (6) shows that the Jacobian has the form

$$J(r\mathbf{1}) = f'(r)\mathrm{Id} - \alpha D + \alpha A = f'(r)\mathrm{Id} - \alpha \Lambda,$$

where Id is the identity matrix, D is the weighted in-degree matrix and Λ the weighted in-Laplacian of the graph. Notice that the spectrum of $J(r\mathbf{1})$ is a spectral shift of the spectrum of $\alpha\Lambda$.

We assume that $\phi_i(l_i) = d_i/l_i$. This implies that $\phi_i^{-1}(c_i) = d_i/c_i$ and $\phi_i'(l_i) = -d_i/l_i^2$. Then the system (7) becomes

$$\begin{aligned}\dot{c}_i &= -\frac{c_i^2}{d_i}\left(\beta\left(1 - \frac{d_i}{c_i}\right) + \sum_{j=1}^{N} a_{ji}\alpha\left(\frac{d_j}{c_j} - \frac{d_i}{c_i}\right)\right) \\ &= \beta c_i\left(1 - \frac{c_i}{d_i}\right) + \sum_{j=1}^{N} \alpha\, a_{ji} c_i \left(1 - \frac{c_i d_j}{c_j d_i}\right).\end{aligned} \tag{8}$$

At first glance it seems that the above equation implies that the self-dynamics of a BS is given by $f(c_i) = \beta c_i(1 - c_i/d_i)$ and it has two equilibria, d_i which is stable and 0 which is unstable. The equilibrium $(d_i, \dots, d_N)$ corresponds to the stable equilibrium of the system (3) and from this we deduce that it is not just asymptotically stable but also a global attractor of the system. The equilibrium 0, however, is not an admittable one because it appears also as a denominator and in this case the right-hand side of (8) cannot be evaluated. In a sense the 0 "equilibrium" of the system (8) corresponds to infinity in the system (3).

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

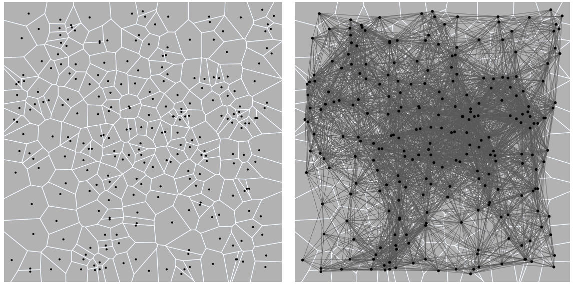
Stochastic Geometry Networks with Eigenvalues Bounded by Gershgorin Circle

Eigenvalues are bounded by disks centred on the diagonal of the matrix:

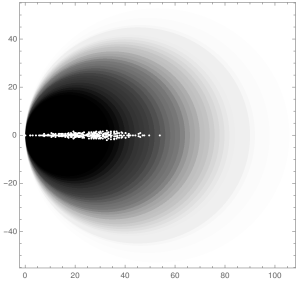
- Sum the absolute value of other row or column members
- Smallest determines radius of circle
- Each circle contains one eigenvalue.

In our case, all the Laplacian eigenvalues are positive. We show with any PPP & PCP networks.

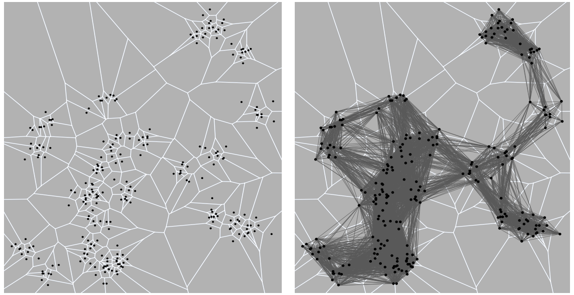
As such, load balancing is always stable, irrespective of: dynamics, and topology; provided that our measurements are accurate.



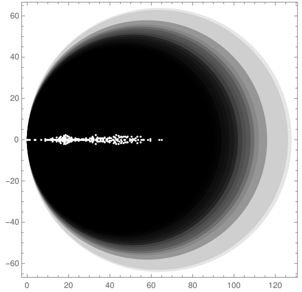
a1. PPP Distributed Cells & Random Neighbor Association



a2. Eigenvalue Distribution



b1. PCP Distributed Cells & Random Neighbor Association



a2. Eigenvalue Distribution

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Probabilistic Uncertainty under Measurement Noise

In the case of many systems, noise can arise from:

- Real stochasticity in the environment
- Noisy behaviour in the components / sub-system
- Measurement noise in the sensor

In our case, we assume that there is both noise in the system measurement of load balancing data flow and an underlying stochasticity in the process.

$$(J)_{ii} = -\beta - \zeta_i - \sum_{j=1}^{N} a_{ji}(\gamma + \xi_{ji})$$

$$(J)_{ij} = a_{ji}\gamma + a_{ji}\xi_{ji}.$$

We elegantly show that provided the measurement noise has a smaller variance than the underlying stochastic process, then the system is always stable for: (1) all dynamics, and (2) all network topologies.

$$f_Y(x) = \frac{1}{2^N} \sum_{n=1}^{N} \binom{N}{n} (1-\gamma/c)^n (1+\gamma/c)^{N-n} \times \frac{1}{(n-1)!} \sum_{k=0}^{\left\lfloor \frac{x}{2(c-\gamma)} \right\rfloor} (-1)^k \binom{n}{k} \left(\frac{x}{2(c-\gamma)} - k \right)^{n-1}.$$

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Case Study: Bee Pollination

Here we study bee pollination network, whereby the population dynamics is governed by:

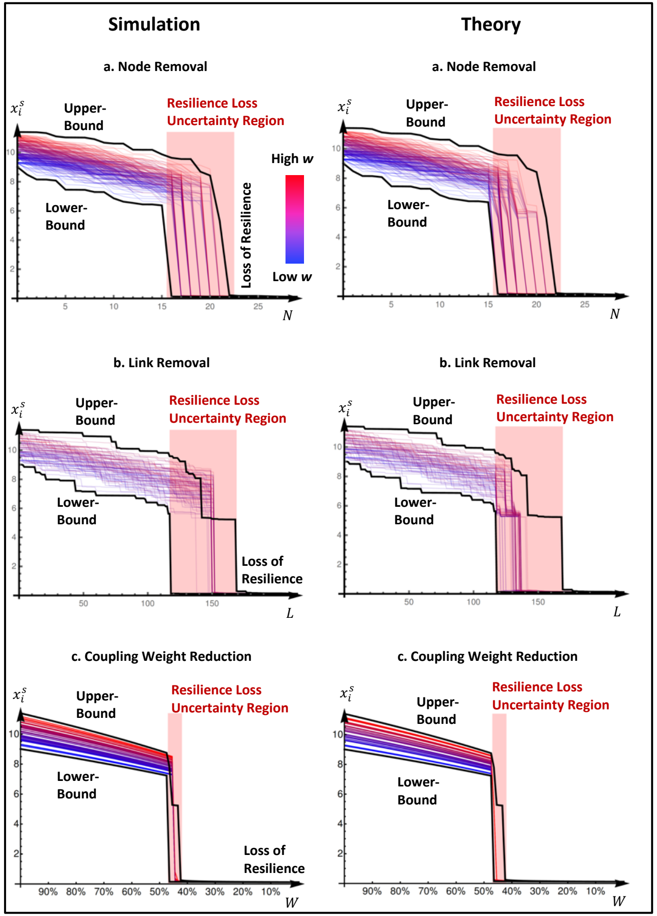
- Carrying Capacity of Bees (*K*)
- Allee Effect (critical hive threshold, *C*)
- Mutualistic Interactions

$$\frac{dx_i}{dt} = x_i(1 - \frac{x_i}{K})(\frac{x_i}{C} - 1) + \sum_j^N a_{ji} \frac{x_i x_j}{D_i + E_i x_i + H_j x_j} + B_i,$$

We show excellent predictability of both the resilience collapse subject to 3 standard perturbation simulations:

- Node Removal: dying of bee colonies
- Link Removal: cut-off from interactions/migration
- Weight Reduction: lowering of interactions

Gives insight into colony collapse disorder.



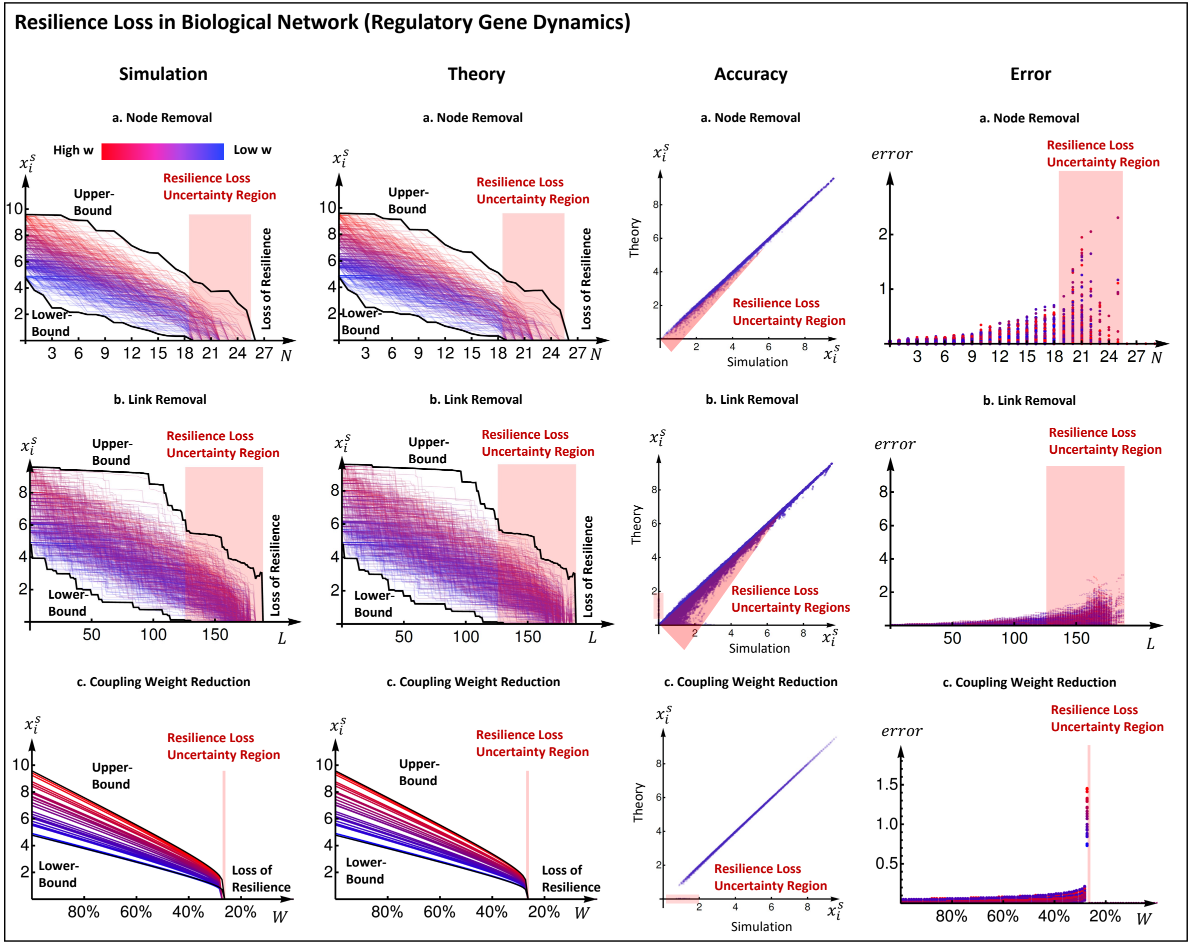
CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Case Study: Gene Regulation

Michaelis-Menten Kinetics infer gene regulatory network.

Smoother collapse profile leads to more accurate predictions.

Error increases towards collapse regime (as network becomes very small and mean field estimation becomes less meaningful).



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Part 3/4: Data-Driven* Analysis

1. Rail Transportation (Engineered)
2. Water Networks (Engineered)

*Suitable for High-Dimensional Dynamics

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Resilience

In absence of well defined measures of resilience, we must use data to help us find proxy measures. We use hierarchical coherence as a proxy for measuring the stability of feedback loops on large complex networks. The **hierarchical level*** (trophic level) is defined as:

$$s_i = 1 + \frac{1}{k_i^{in}} \sum_j a_{ij} s_j .$$

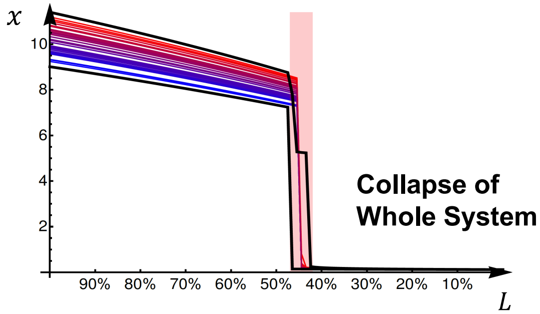
The **incoherence** of the network (instability) is defined as [3]:

$$q = \sqrt{\frac{1}{L} \sum_{ij} a_{ij} x_{ij}^2 - 1}$$

Robustness

Robustness can be well simulated using random and targeted node/link removal.

Identifying the average number of steps until collapse or decay to 50% is quite common. Other mesoscopic proxies such as core-periphery size and rich club coefficient can also be used [4].



CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

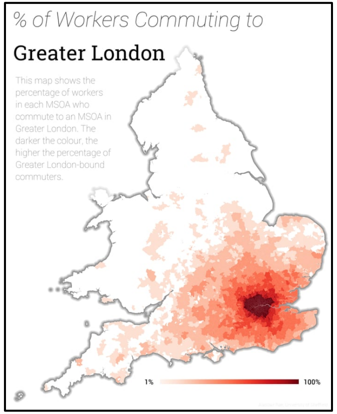
Case Study: Transport (Rail)

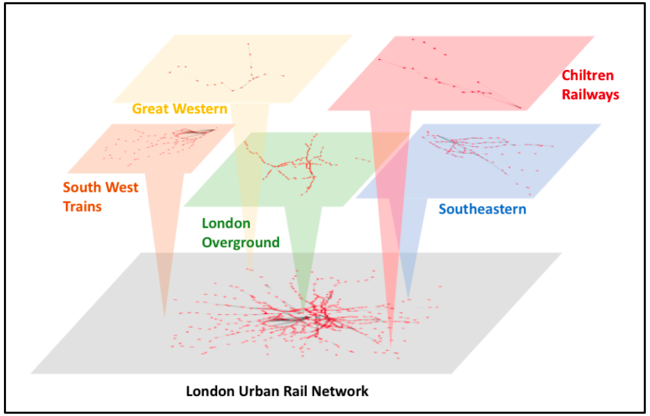
Here we study morning commuter rail travel, using census data and transport planning API to examine [5]:

- Which railway route I(include which train and what service) people will take (if any)
- How long it will take to get there

We construct a hierarchical multi-scale graph, where:

- Multiple transport links overlap on common stations
- Minor flows are removed (counter commuter flow <30 passengers).





Data Processing

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Check for updates

Cite this article: Pagani A, Mosquera G, Alturki A, Johnson S, Jarvis S, Wilson A, Guo W, Varga L. 2019 Resilience or robustness: identifying topological vulnerabilities in rail networks. *R. Soc. open sci.* **6**: 181301.
http://dx.doi.org/10.1098/rsos.181301

Received: 8 August 2018
Accepted: 3 January 2019

Subject Category:
Mathematics

Subject Areas:
graph theory

Keywords:
complex networks, resilience, robustness, trophic coherence, rich-core club

Author for correspondence:
Alessio Pagani
e-mail: apagani@turing.ac.uk

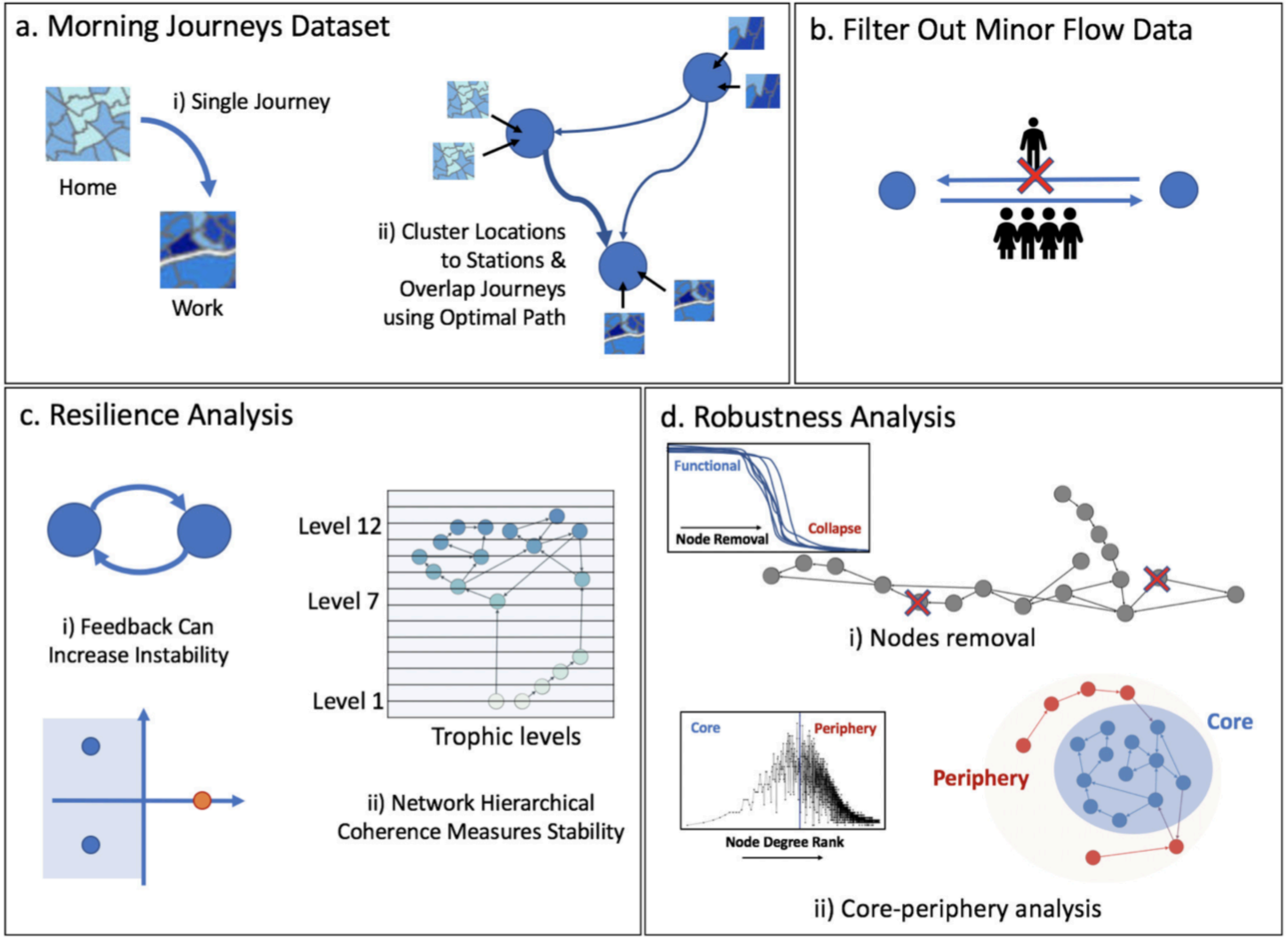
Resilience or robustness: identifying topological vulnerabilities in rail networks

Alessio Pagani[1], Guillem Mosquera[1,2], Aseel Alturki[3], Samuel Johnson[5], Stephen Jarvis[3], Alan Wilson[1], Weisi Guo[1,4] and Liz Varga[6]

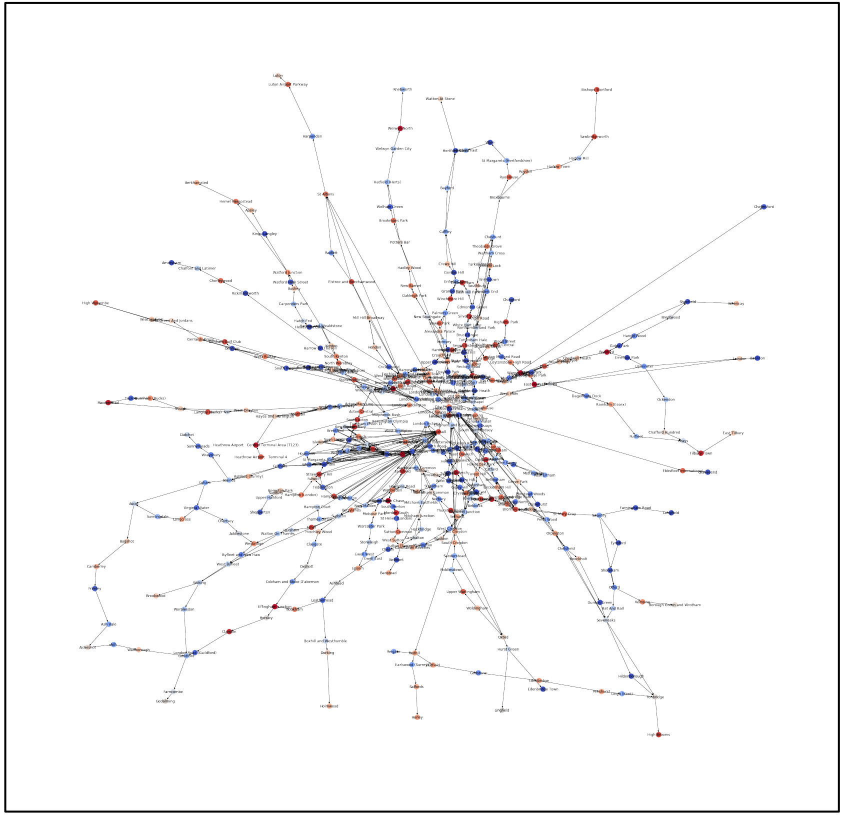
[1]The Alan Turing Institute, London, UK
[2]Mathematics Institute, [3]Department of Computer Science, and [4]School of Engineering, University of Warwick, Coventry, UK
[5]School of Mathematics, University of Birmingham, Birmingham, UK
[6]School of Management, Cranfield University, Cranfield, UK

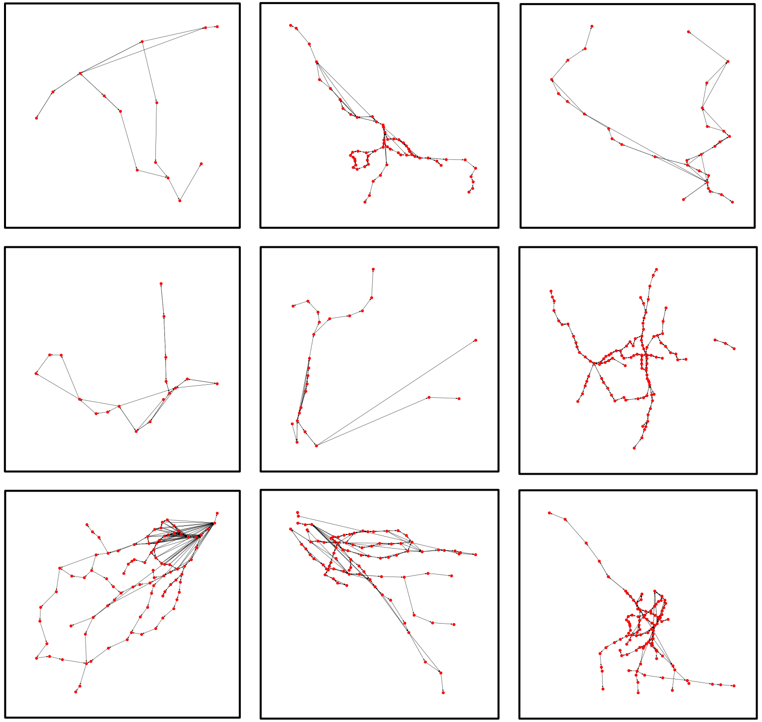
AP, 0000-0002-8404-8890; WG, 0000-0003-3524-3953

Many critical infrastructure systems have network structures and are under stress. Despite their national importance, the complexity of large-scale transport networks means that we do not fully understand their vulnerabilities to cascade failures. The research conducted through this paper examines the interdependent rail networks in Greater London and surrounding commuter area. We focus on the morning commuter hours, where the system is under the most demand stress. There is increasing evidence that the topological shape of the network plays an important role in dynamic cascades. Here, we examine whether the different topological measures of resilience (stability) or robustness (failure) are more appropriate for understanding poor railway performance. The results show that resilience, not robustness, has a strong correlation with the consumer experience statistics. Our results are a way of describing the complexity of cascade dynamics on networks without the involvement of detailed agent-based models, showing that cascade effects are more responsible for poor performance than failures. The network science analysis hints at pathways towards making the network structure more resilient by reducing feedback loops.

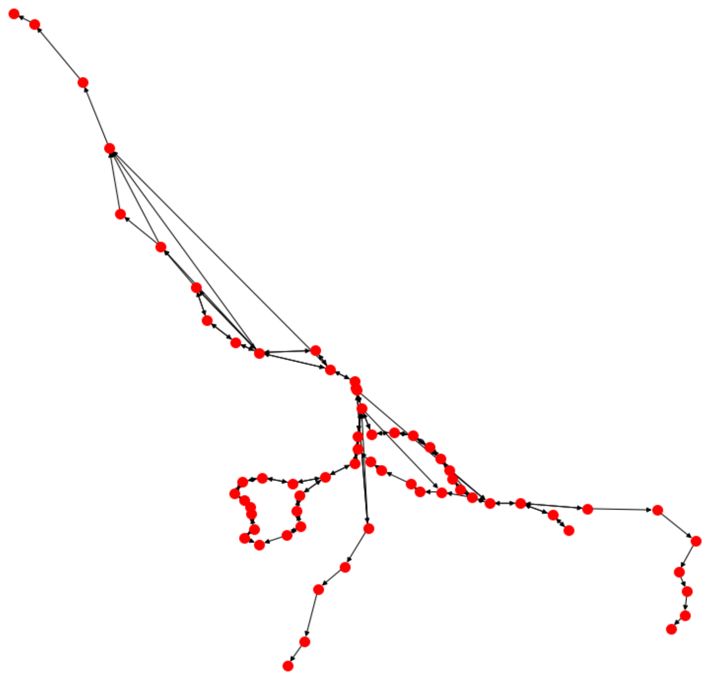


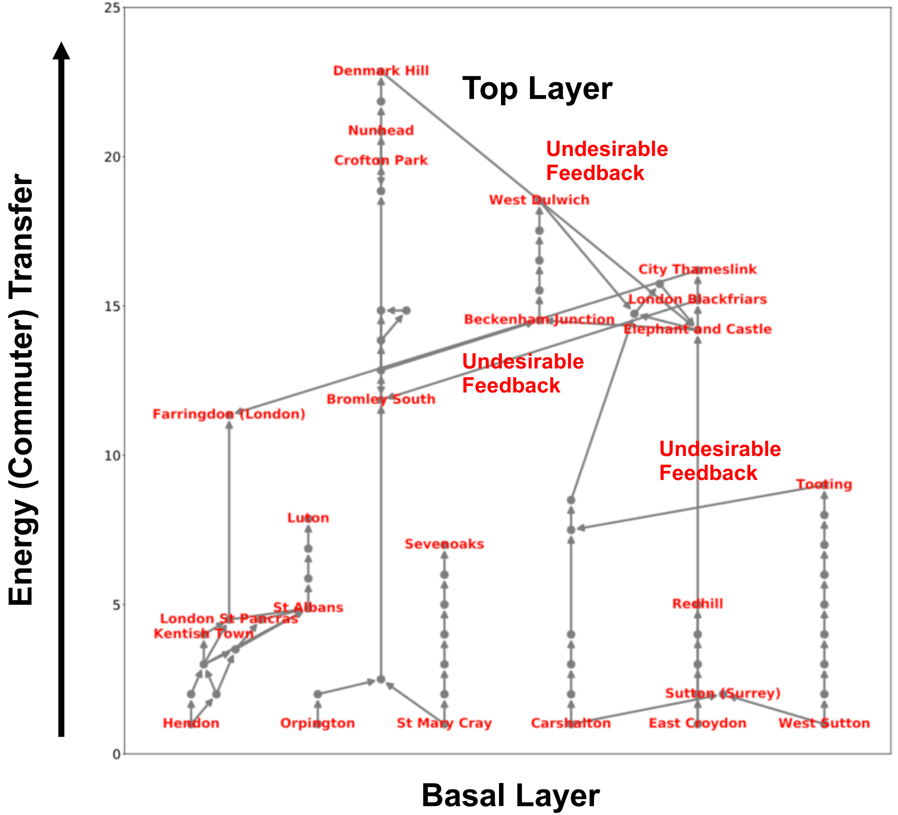
Overall and Individual Rail Network Topologies





Hierarchical Graph (Example: Thameslink)





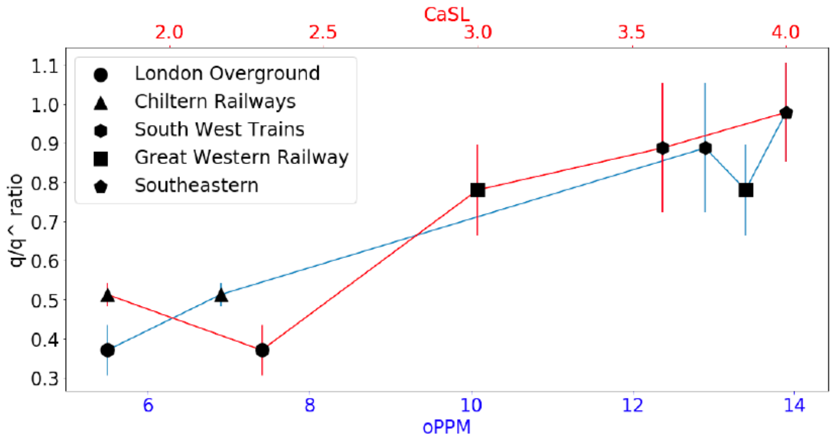
Results

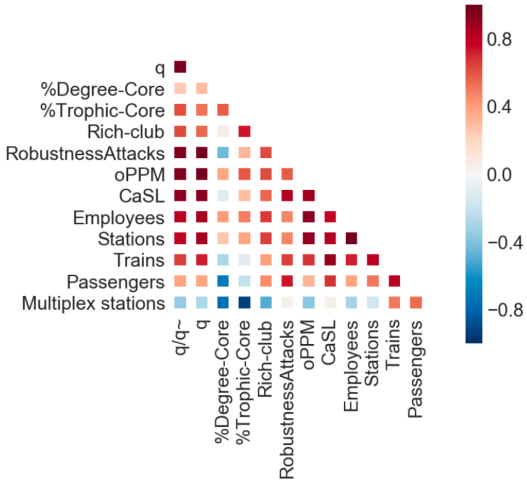
We analyse [5]:

- Resilience vs. Robustness against consumer satisfaction & late train data
- Greater London ~1 Hour Commuter Range

As a pure data-driven study, we show:

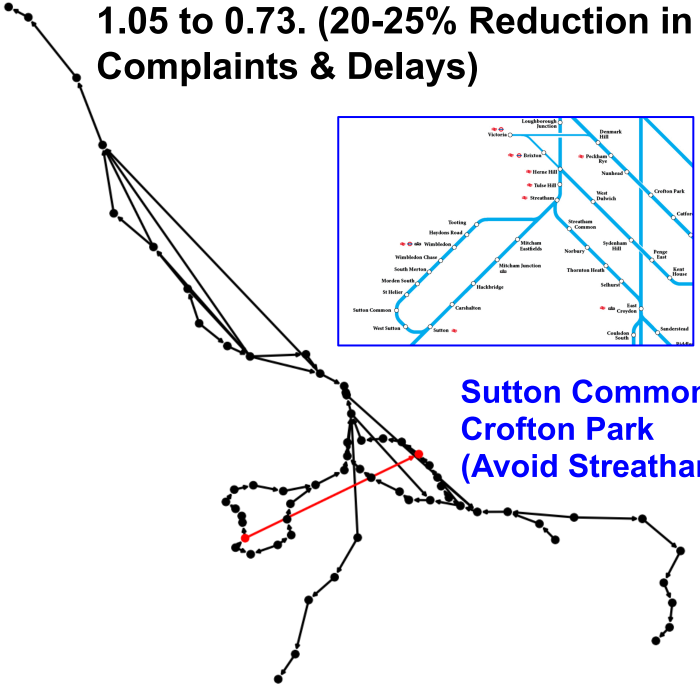
- Rail performance is strongly correlated to resilience, but not robustness;
- Pointing towards a pathway to reduce interdependency between rail services to reduce cascade effects.





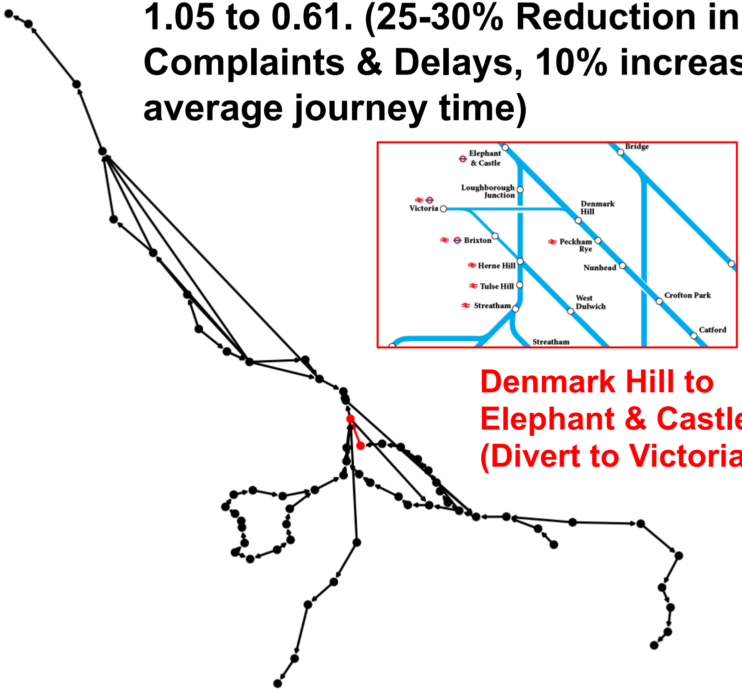
Minimum Change for Improvement (Example: Thameslink)

New Link Reduces Incoherence from 1.05 to 0.73. (20-25% Reduction in Complaints & Delays)



Sutton Common & Crofton Park (Avoid Streatham)

Remove Link Reduces Incoherence from 1.05 to 0.61. (25-30% Reduction in Complaints & Delays, 10% increase in average journey time)



Denmark Hill to Elephant & Castle (Divert to Victoria)

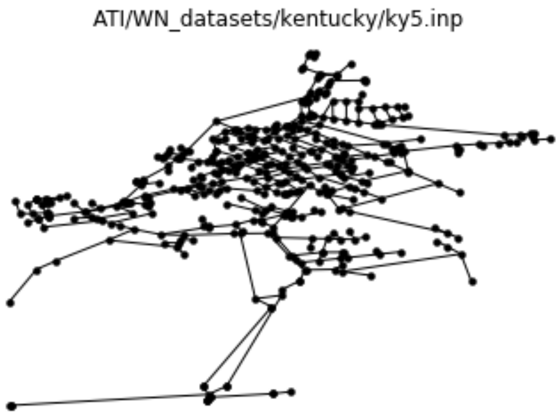
Case Study: Water Distribution Network

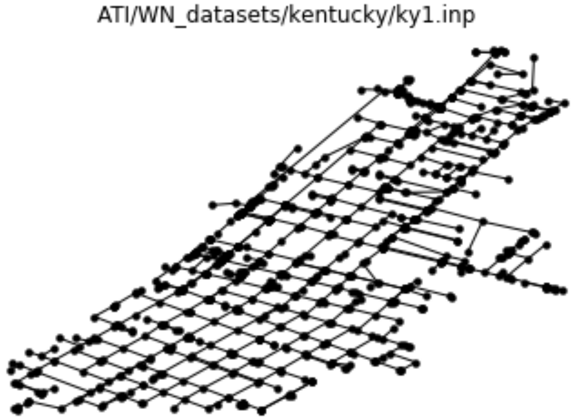
Here we study different urban and rural water distribution networks (WDN) from the world with data from [6] and using EPANET to simulate the WDN performance:

- WDN topology and units
- Simulate demand variation across WDN
- Define failures in terms of pressure in pipes.

<u>Nodes:</u>
Junctions have water demand, **Reservoirs** provide water, **Pumps** increase pressure, **Valves** manage flows.

<u>Edges:</u>
Pipes that connect nodes. Pipe properties include diameter, length, roughless and minor loss.

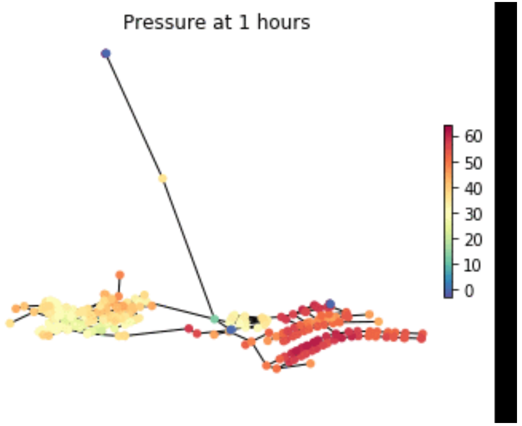


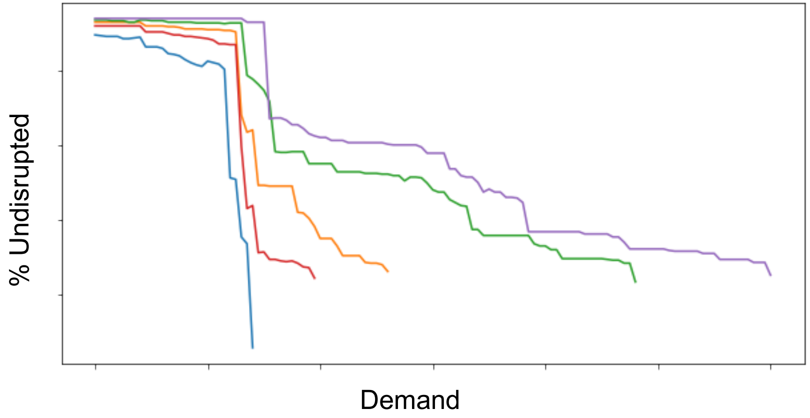


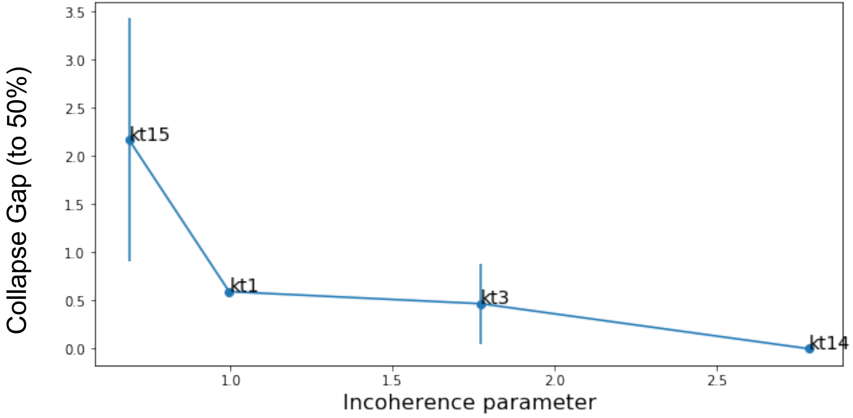
Water Distribution Cascade Failure

As a pure data-driven study, we show:

- Cascade failure performance is strongly correlated to resilience (data driven structural parameter).
- Pointing towards a pathway to improve topological structure by increasing WDN structural coherence.
- This can be achieved using **dynamic topology reconfiguration**.

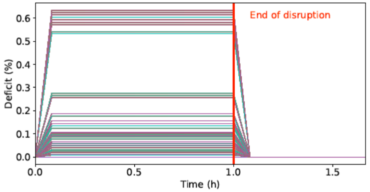




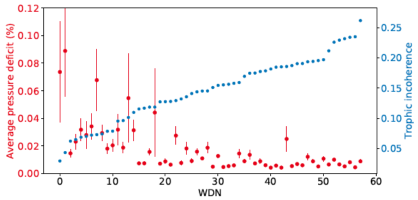


Water Distribution Pressure Deficit & Pollution

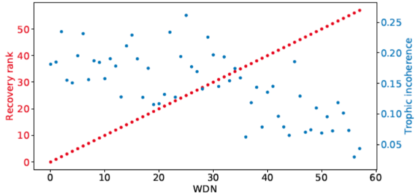
Trophic incoherence impacts a variety of dynamics on WDNs, more so many other recognised resilience measures.



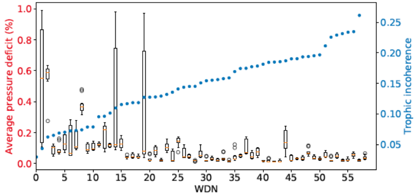
(a) Average pressure difference compared with standard condition. Each line is a different junction closure.



(b) Average pressure deficit VS trophic incoherence.

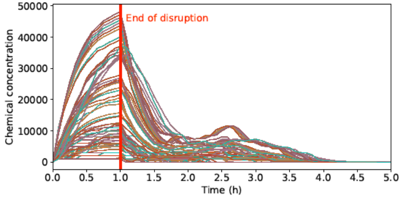


(c) Pressure deficit rank VS trophic incoherence.

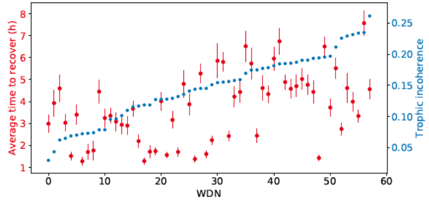


(d) 10 most significant pressure deficits VS trophic incoherence.

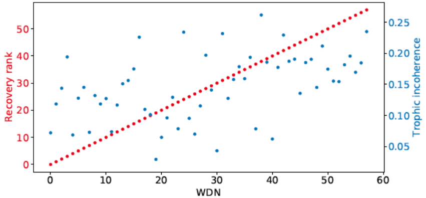
Figure 2: Pressure deficit



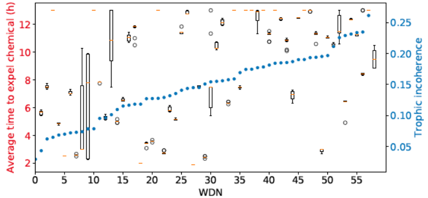
(a) Average expulsion time of a chemical. Each line is a different injection junction.



(b) Average chemical expulsion time VS trophic incoherence.



(c) Expulsion time rank VS trophic incoherence.



(d) 10 most disruptive nodes VS trophic incoherence.

Figure 4: Chemical

Impact of Trophic Incoherence on Variety of WDNs with Different Demand Scenarios

	Synthetic WDNs SD	Synthetic WDNs VD	Kent. WDNs
Junction breakage			
Mean % nodes with deficit >25%	-0.66	-0.65	-0.55
Mean % nodes with deficit >50%	-0.66	-0.66	-0.56
Mean % pop. impacted (deficit >25%)	-0.66	-0.65	-0.39
% Pop. impacted (deficit >50%)	-0.66	-0.66	-0.39
Mean time to recover	0.62	0.39	0.44
Chemical injection			
Mean Time to recover	0.52	0.48	0.21
Mean % chemical Extent	-0.29	-0.30	-0.52
Mean % pop. impacted	-0.40	-0.37	-0.46

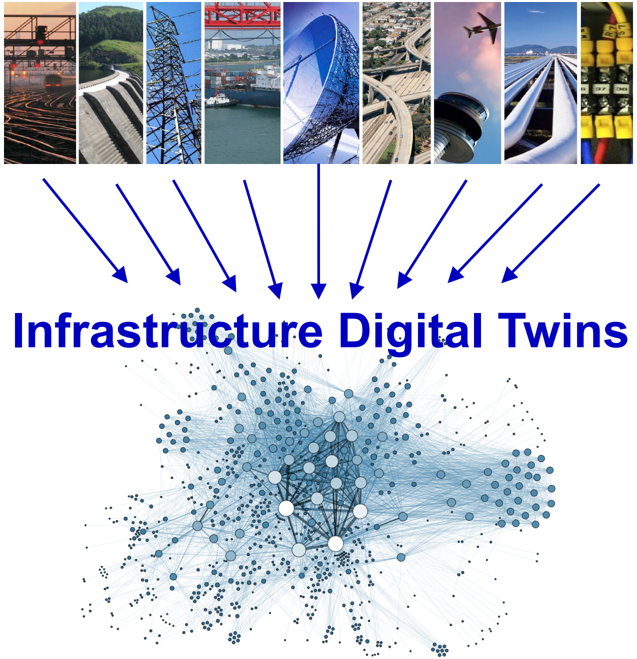
Table 1: Summary of the results. Synthetic networks are tested with static (SD) and variable (VD) demand patterns. Kentucky WDNs are tested with the pattern provided. In variable demand mode, the disruptions are created during the time of the day with peak water demand.

Part 4/4: Optimal Data Collection

1. New Sampling Theory for Graphs with Explicit Nonlinear Dynamics
2. Fourier Basis Sampling on Unknown Nonlinear Dynamic Graphs
3. Some Examples of Case Studies

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

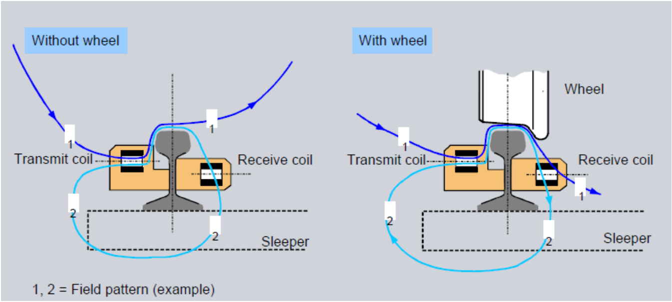
Digital Twins

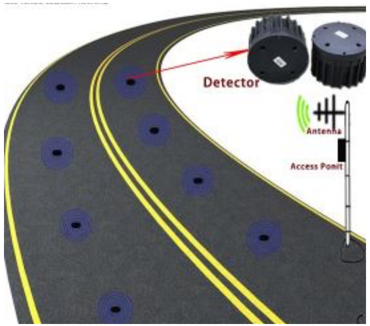


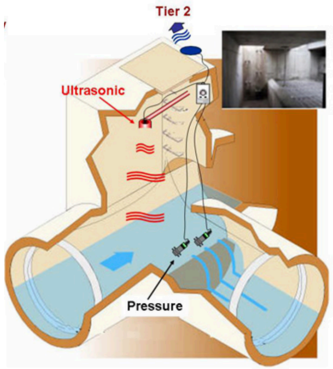
- Increased affordability
- Predictive maintenance
- Environmental protection
- Reduced operational expenditure
- Increased resilience
- Increased revenue

Current Limitations

- Sensor installation
- Data transmission and processing
- Maintenance and replacement
- Process information





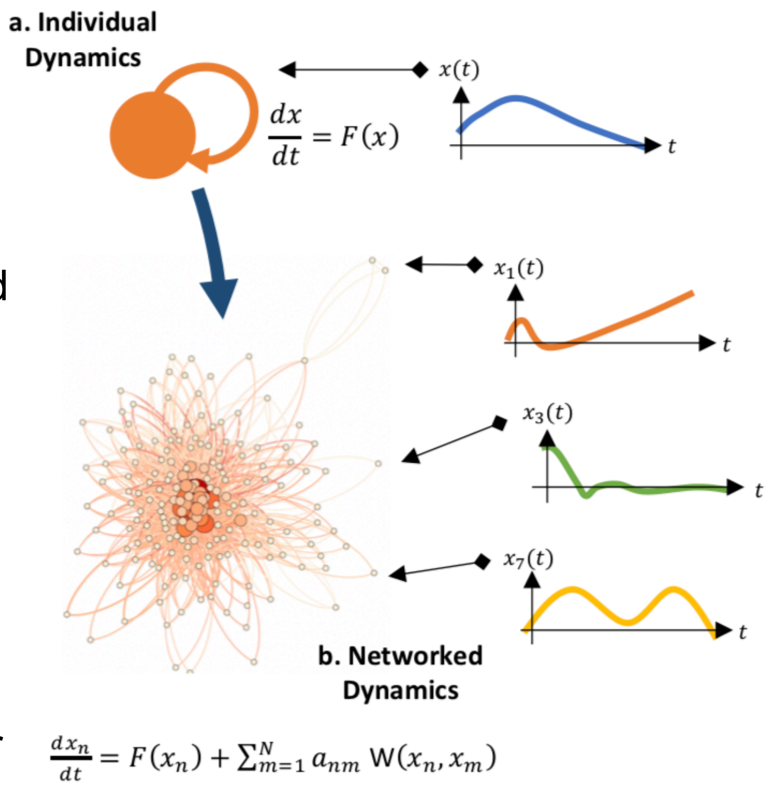


New Sampling Theorem

A key challenge faced by stakeholders is:
- Where (graph) do we collect data?
- What sampling rate (node) do I need?

Whilst individual graph (spectral properties) and dynamic sampling (Nyquist rate) are well governed by established theorems, we do not have a joint dynamic graph sampling theorem. This may seem similar to the problem of compressed sensing (tensor), but here we have explicit non-linear dynamics (causal relations between data).

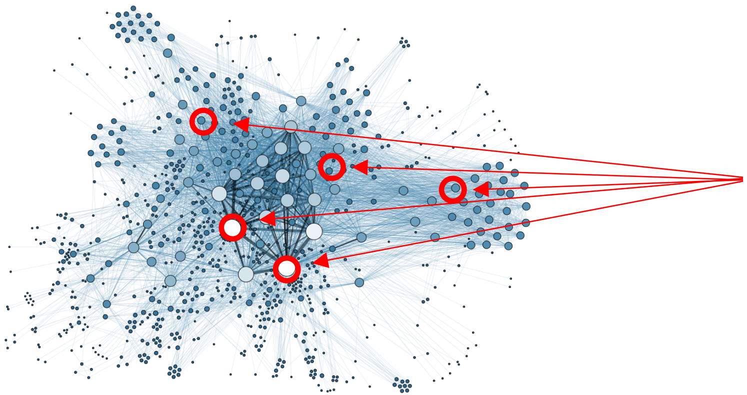
Here we create a joint optimal sampling theorem mapping frequency of graph, frequency of dynamics, to the graph structure and the nonlinear dynamics.



Infrastructure Networks with Nonlinear Dynamics

Dynamical elements connected via a large-scale networks.

1) Topology



How to optimally sample networked **dynamical** elements?

2) Dynamics

$$\frac{dx_n(t)}{dt} = f_n(x_1(t), \cdots, x_N(t)) = F_n(x_n(t)) + \sum_{m=1}^{N} \alpha_{n,m} \cdot W_{n,m}(x_n(t), x_m(t)) + \eta_n$$

node

connected nodes

Case Study: Digital Water

Remote sensing and digital twin technologies provide connectivity between an utility and its diversified water supply.

- Water resource planning
- Real-time water network pump scheduling
- Water and wastewater network control
- Capital and operational intervention planning

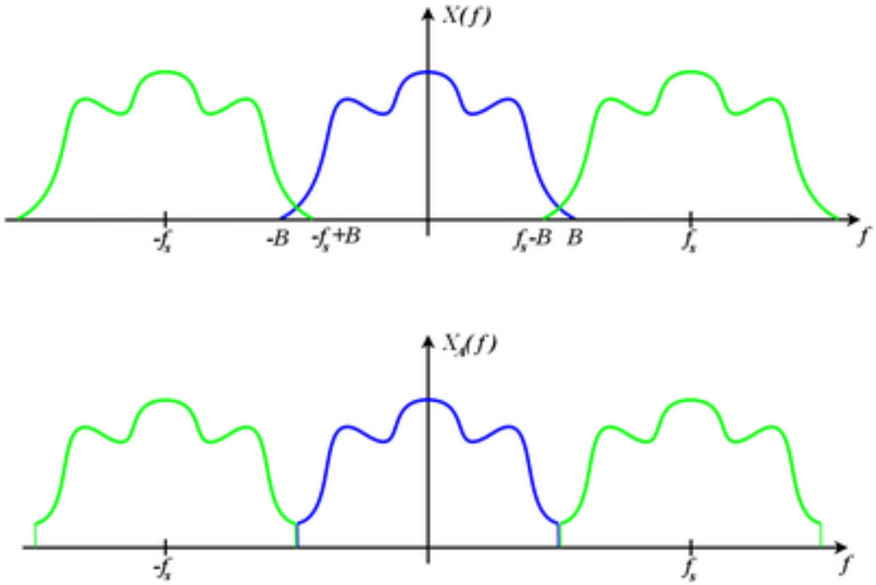
UK water distribution network:

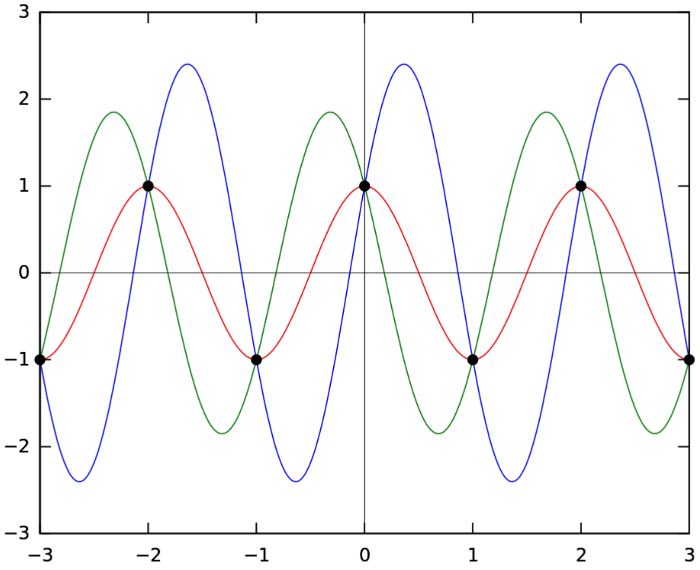
- 1,433 water treatment works
- 5,950 service reservoirs
- 9,000 sewage treatment works
- 416,175 km of water mains
- 393,460 km of sewers

- Digital water will represent a £9.7 billion to 14 billion market opportunity by 2019 to 2020
- 20 to 40% - the reduction in water pipe leakage through smart pressure monitoring

Data source: https://www.gov.uk/government/publications/water-and-treated-water/water-and-treated-water

Classical Method (Time-Series): Nyquist-Shannon





Problems: 1) How can we sample at sub-Nyquist, 2) How does it work on a graph?

Classical Method (Graph): Laplacian

$$L = D - A$$

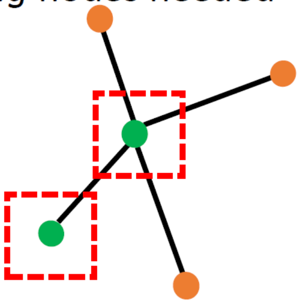
D degree matrix
A adjacency matrix

$>r, \leq N$ *sampling* nodes needed

$$\left[\ \mathcal{L}_{N\times N}\ \right] \longrightarrow [\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_N]$$

Laplacian matrix

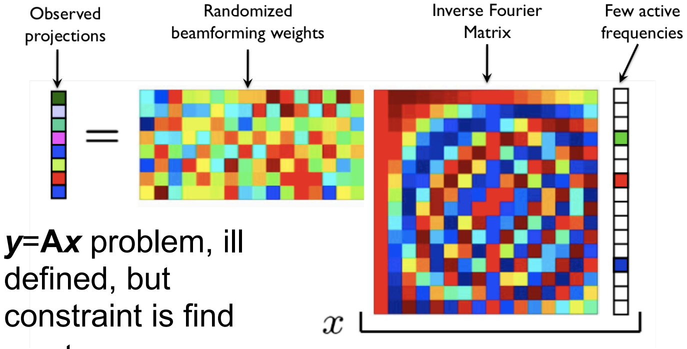
eigenvalues



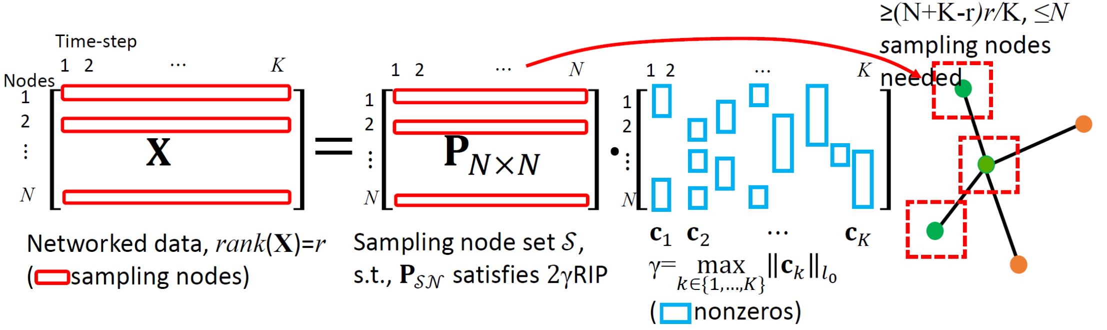
Problem: constructed Laplacian Matrix does not consider Dynamics

[1] I. Pesenson, "Sampling in paley-wiener spaces on combinatorial graphs," Transactions of the American Mathematical Society, vol. 360, no. 10, pp. 5603–5627, 2008.
[2] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," IEEE Transactions on Signal Processing, vol. 64, no. 14, pp. 3775–3789, 2016

Dynamic Graphs: Compressed Sensing



y*=*Ax problem, ill defined, but constraint is find most sparse x.



restricted isometry property R. I. P. (near orthonormal in sparse vectors):

$$1-\delta_{2\gamma} \leq \frac{\|\mathbf{P}_{\mathcal{SV}}\cdot\mathbf{c}\|_{l_2}^2}{\|\mathbf{c}\|_{l_2}^2} \leq 1+\delta_{2\gamma},\ \ \gamma=\max_{k\in\mathcal{K}}\|\mathbf{c}_k\|_{l_0}$$

[1] R. Du, L. Gkatzikis, C. Fischione, and M. Xiao, “Energy efficient sensor activation for water distribution networks based on compressive sensing,” IEEE Journal on Selected Areas in Communications, vol. 33, no. 12, pp. 2997–3010, 2015.

[2] Xu, X. Qi, Y. Wang, and T. Moscibroda, “Efficient data gathering using compressed sparse functions,” in 2013 Proceedings IEEE INFO-COM 2013, pp. 310–314.

[3] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, “Sensing, compression, and recovery for wsns: Sparse signal modeling and monitoring framework,” IEEE Transactions on Wireless Communications, vol. 11, no. 10, pp. 3447–3461, 2012

Compressed Sensing Approaches

Discrete Cosine Transform (DCT)

- Approximates the dynamics using sum of cosine functions oscillating at different frequencies.

Principal Component Analysis (PCA)

- Reconstruct the signal using correlation.

Dictionary Learning

- Reconstruct the signal using linear combination of basic elements.

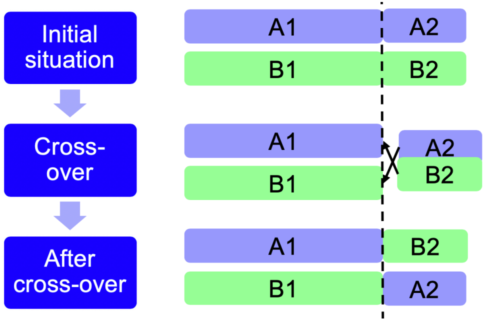
Benefit:

- Universal approach to compressing tensors

Problems:

- Compression is limited because RIP is strict and WDNs are relatively high rank.

Practical Methods used in industry

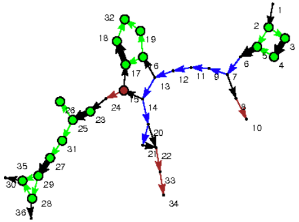


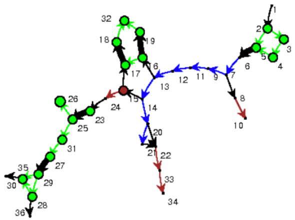
Genetic algorithms

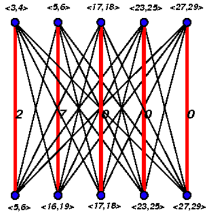
- Community detection
- Numerical optimization (genetic algorithm, ant colony)

Numerical optimization and experience

- For example, optimizing the number of sensors that minimize an objective (energy consumption, population impacted).





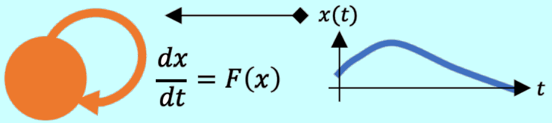


[1] M. Guerrero, F.G. Montoya, R. Baños, A. Alcayde, C. Gil “Adaptive community detection in complex networks using genetic algorithms”
[2] G. Morcous, Z. Lounis “Maintenance optimization of infrastructure networks using genetic algorithms”
[3] J.W. Berry, L. Fleischer, L W.E. Hart, C.A. Phillips, J.P. Watson “Sensor Placement in Municipal Water Networks”

Equation-Driven vs Data-Driven

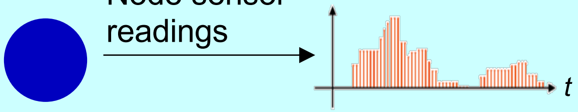
Explicit mapping: optimal sampling locations and rates to the graph properties and the governing dynamics.

Node dynamics

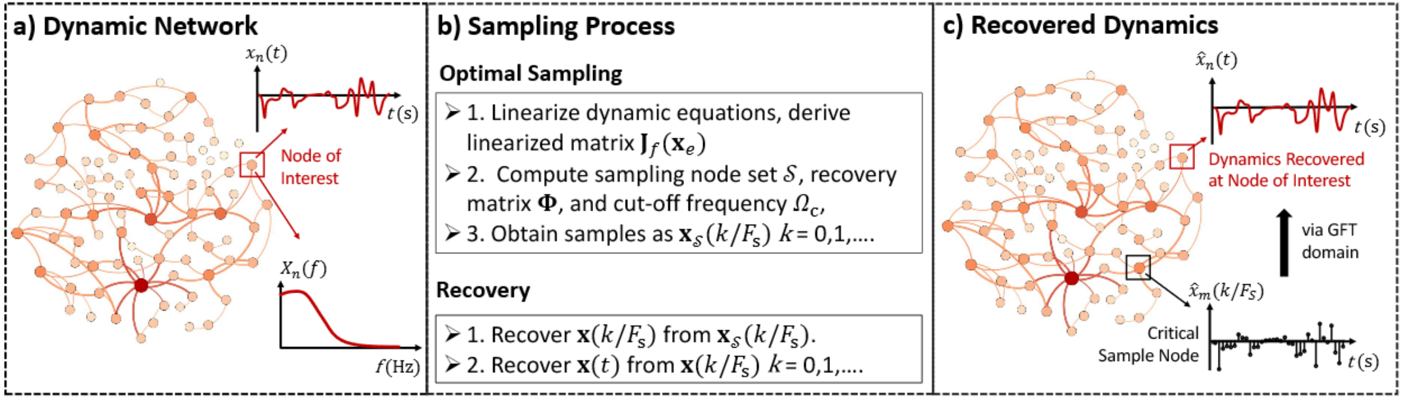


Implicit mapping: determine which set of nodes are optimal to recover the full network's dynamics.

Node sensor readings



Equation-Driven



$$\mathbf{y}(t) = \mathbf{x}(t) - \mathbf{x}_e$$

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{J}_f(\mathbf{x}_e) \cdot \mathbf{y}(t) + o(\|\mathbf{y}(t)\|) \approx \mathbf{J}_f(\mathbf{x}_e) \cdot \mathbf{y}(t)$$

$$\mathbf{J}_f(\mathbf{x}_e) \triangleq \begin{bmatrix} \frac{\partial f_1(t)}{\partial x_1(t)} & \cdots & \frac{\partial f_1(t)}{\partial x_N(t)} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N(t)}{\partial x_1(t)} & \cdots & \frac{\partial f_N(t)}{\partial x_N(t)} \end{bmatrix} \Bigg|_{\mathbf{x}(t)=\mathbf{x}_e}$$

[1] Zhuangkun Wei, Bin Li, and Weisi Guo. "Optimal Sampling in Joint Time-and Graph-Domains for Dynamic Complex Networks." arXiv preprint arXiv:1901.11405 (2019).

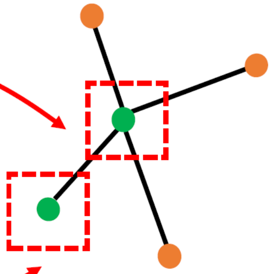
Equation-Driven

$r = \text{rank}(\mathbf{J}_f(\mathbf{x}_e)) = N\text{-}d$
sampling nodes needed

$$\left[\ \mathbf{J}_f(\mathbf{x}_e)\ \right] \longrightarrow [\lambda_1\ \ \lambda_2\ \ \lambda_3\ \cdots\ \lambda_d\ \ \lambda_{d+1}\ \ \cdots\ \ \lambda_N]$$

Linearization matrix

Find approximated eigenvalues
e.g., $\exp(\lambda_1) \approx \exp(\lambda_2) \approx \exp(\lambda_3) \approx \cdots \approx \exp(\lambda_{d+1})$



$$\mathbf{J}_f(\mathbf{x}_e) = \mathbf{W} \circ \mathbf{J}_W(\mathbf{x}_e) + diag\left\{\frac{\partial F_1}{\partial x_1}, \cdots, \frac{\partial F_N}{\partial x_N}\right\}\Bigg|_{\mathbf{x}_e}$$

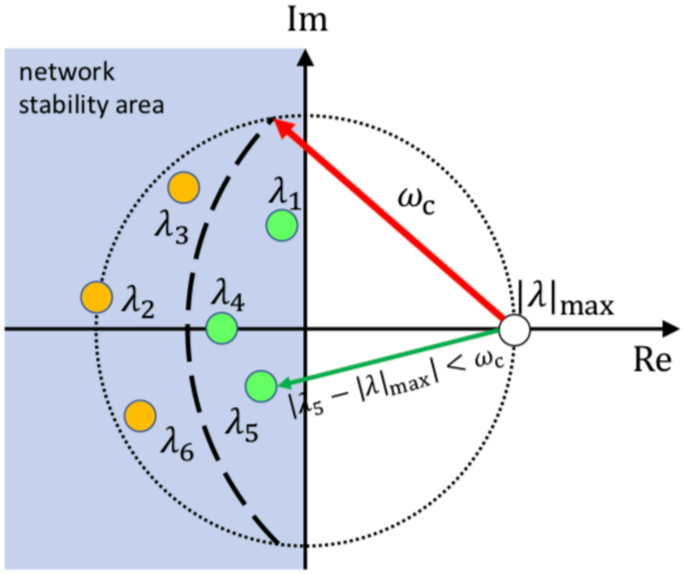
topology

Dynamic equations

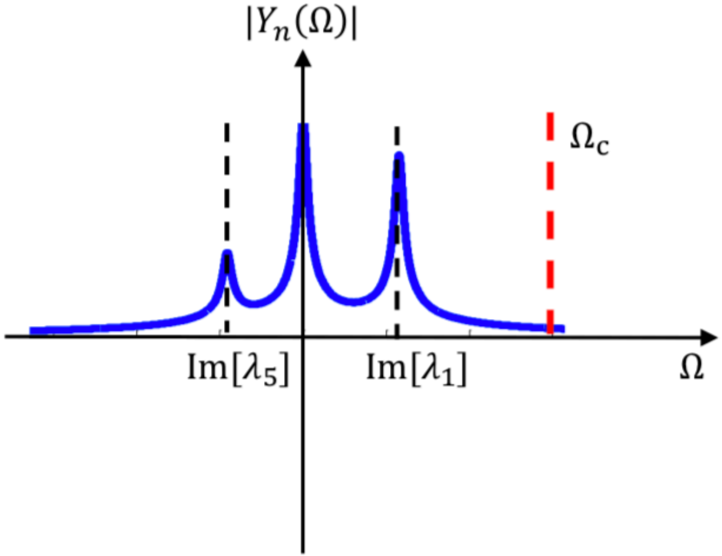
Explicit Mapping of Graph and Dynamics

Topology (Eigenvalues)
Mapped to Dynamics
(Sampling Freq.)

$$Y_n(\Omega) = \sum_{\lambda_j \in \lambda_{+\infty}(\mathbf{A})} \frac{\lambda_{n,j} \cdot \tilde{y}_j(0)}{-\mathrm{Re}[\lambda_j] + i\,(\Omega - \mathrm{Im}[\lambda_j])}.$$

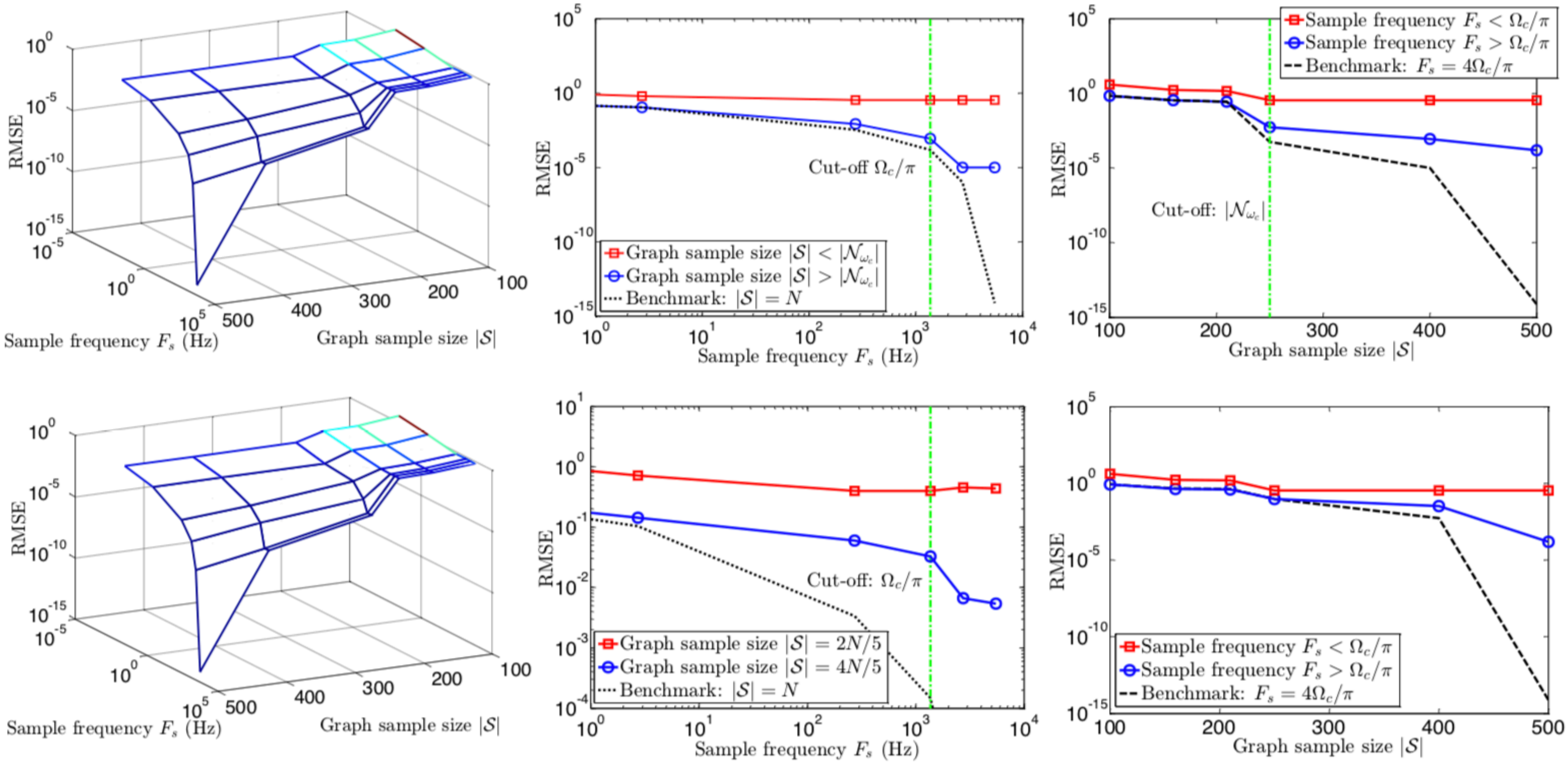


(a)



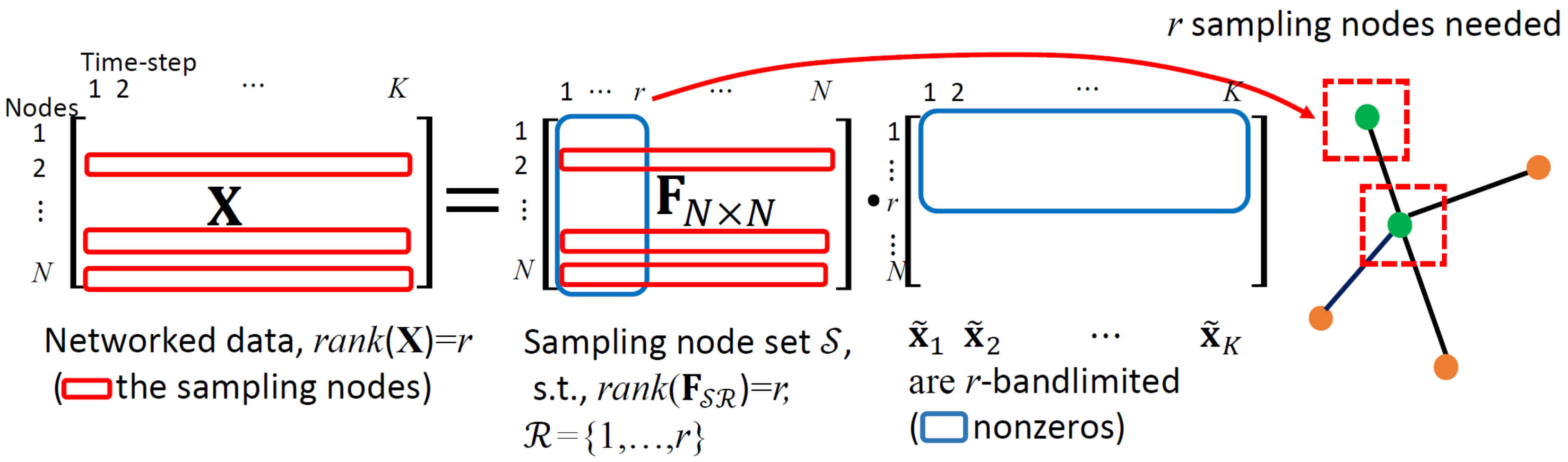
(b)

Non-Linear Dynamics with Different Disturbances



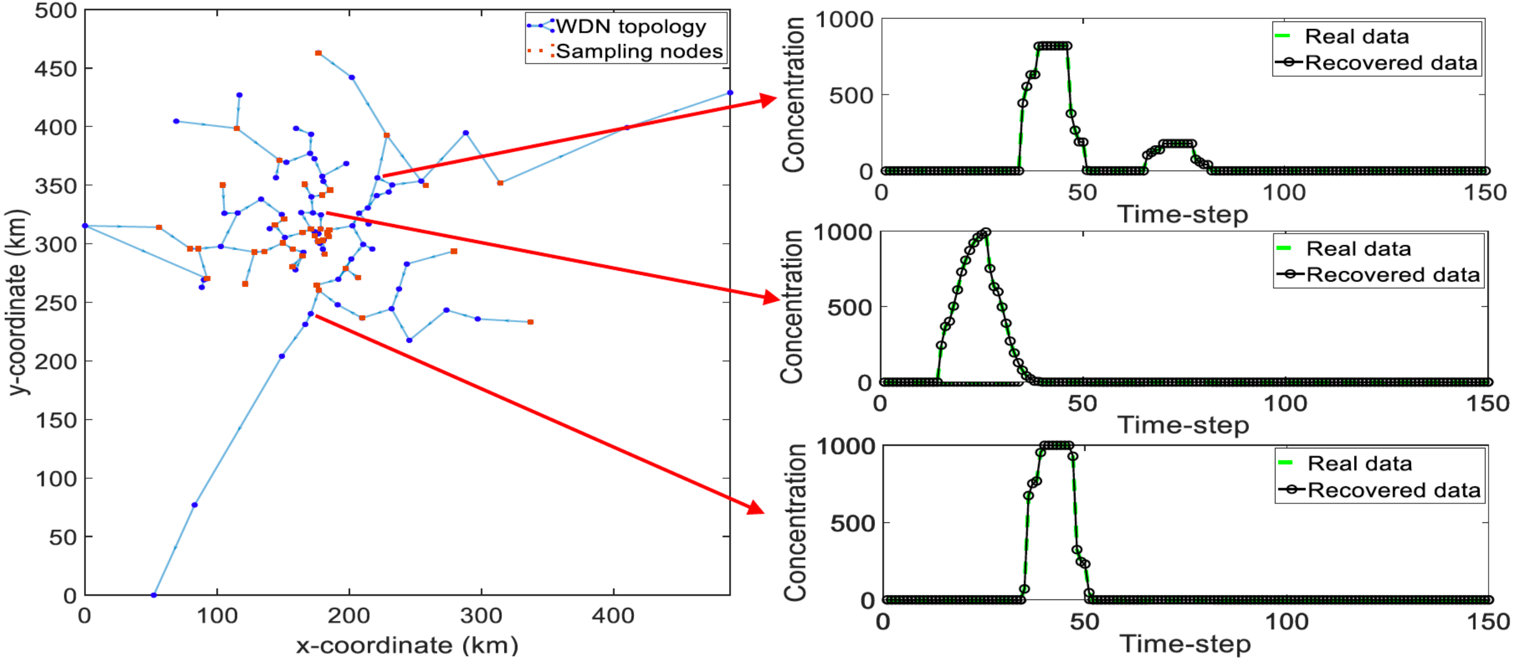
(a) (b) (c)

Data Driven Graph Fourier Transform



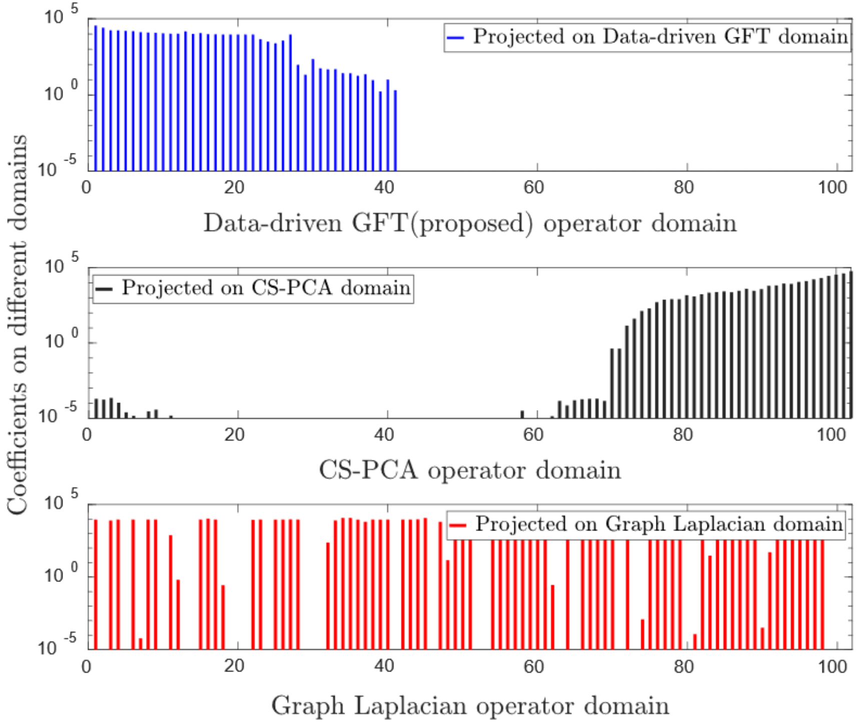
[1] Zhuangkun Wei, Alessio Pagani, Guangtao Fu, Ian Guymer, Wei Chen, Julie McCann, and Weisi Guo. "Optimal Sampling of Water Distribution Network Dynamics using Graph Fourier Transform." arXiv preprint arXiv:1904.03437 (2019).

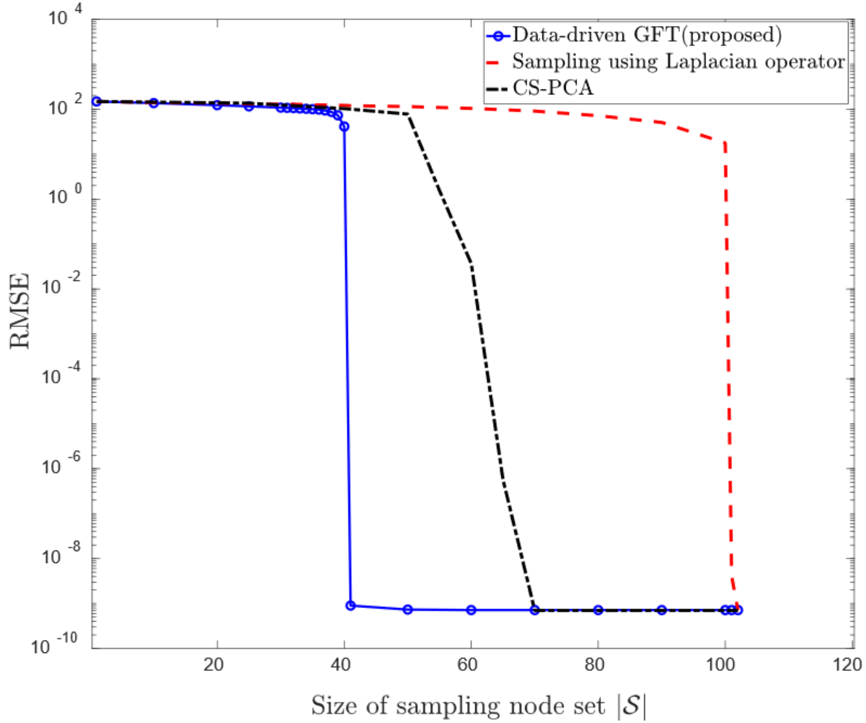
Recovery of Full Network Dynamics from Sampling a Subset of WDN Nodes



(a) (b)

Performance Comparisons

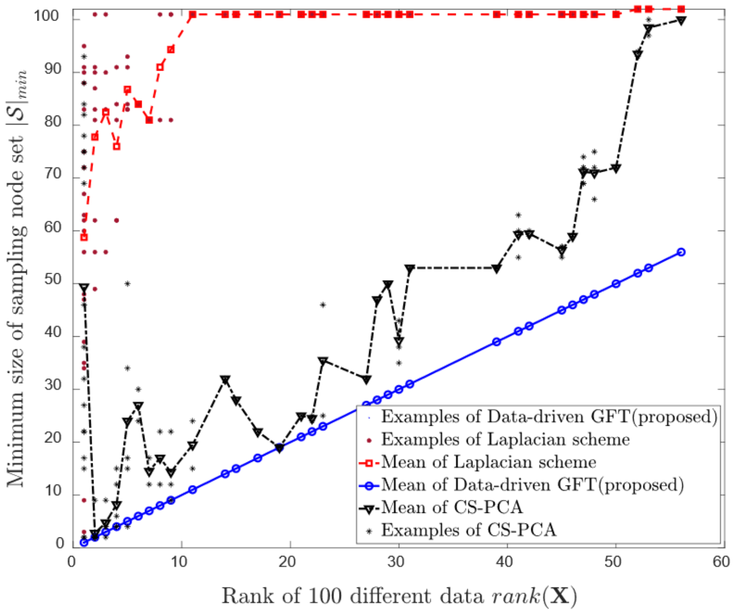




Performance Comparisons

Comparison of Size of sampling node set such that RMSE< 10^{-8} among different sampling methods.

	Methods	Sampling node set size, s.t. RMSE<10^{-8} Data with $rank(\mathbf{X}) = r \leq N$
Graph sampling	Data-driven	r
	Laplacian	$\geq r, \quad \leq N$
Compressed sensing	DCT basis	$\geq (N+K-r)r/K \geq r, \quad \leq N$
	PCA basis	$\geq (N+K-r)r/K \geq r, \quad \leq N$

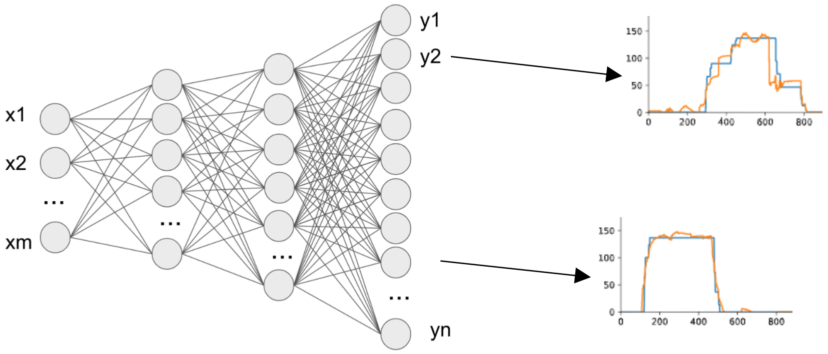


Deep Sampling using Neural Networks

Further reduce sample size and recover features in high dimensions.

1. Training of artificial neural networks to learn high-dimensional features.
2. Rank of critical sampling nodes informs Deep Sampling input data size.

Nodes with sensors

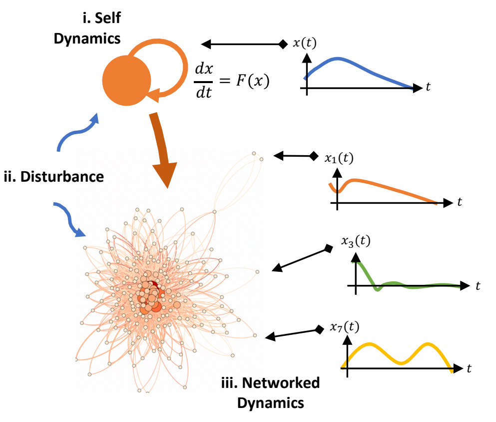


Reconstructed dynamics in nodes with sensors

With 5x less sensor data than state-of-the-art, we can achieve a highly accurate reconstruction of the dynamic response.

2019 Bell Labs Prize Entry: Deep Sampling on Dynamic Networks

a) Explicit Dynamics

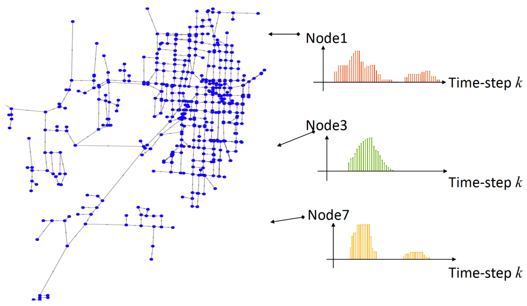


$$\frac{dx_n}{dt} = F_n(x_n) + \sum_{m=1}^{N} a_{nm} W_{nm}(x_n, x_m) + \eta_n$$

b) Implicit (Data-Driven) Dynamics

i. Water Distribution Network with Reynolds-Averaged Navier-Stokes (RANS) Dynamics

ii. Sensors monitor water pressure and contamination

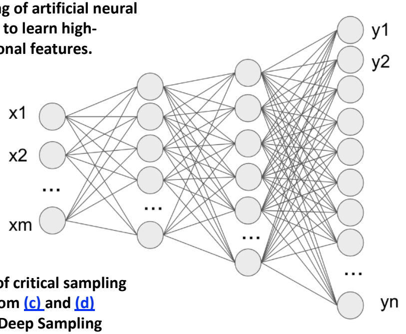


Heterogeneous functions governed by higher order PDE with dynamic parameters.

$$c\bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = c\bar{f}_i + \frac{\partial}{\partial x_j}\left[-\bar{p}\delta_{ij} + \mu\left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i}\right) - c\overline{u_i' u_j'}\right]$$

e) Deep Sampling with Minimum Data

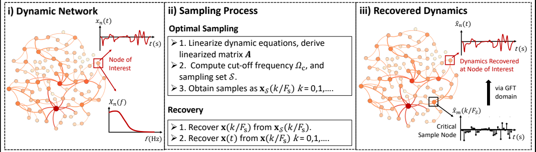
i. Training of artificial neural network to learn high-dimensional features.



ii. Rank of critical sampling nodes from (c) and (d) informs Deep Sampling input data size.

iii. Insight: Theory-driven input data can dramatically reduce sample size and recover features in high dimensions.

c) Graph Fourier Transform for Explicit Dynamics



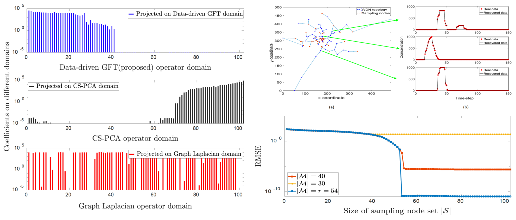
i. Native network: each node's dynamic has different spectral properties.

ii. Sampling theory to identify critical sampling nodes and their sampling rate.

iii. Sparse samples from critical nodes in (ii) enable recovery of all network dynamics.

iv. Insight: Yields knowledge of how graph-time domain sampling relates to network structure and dynamics.

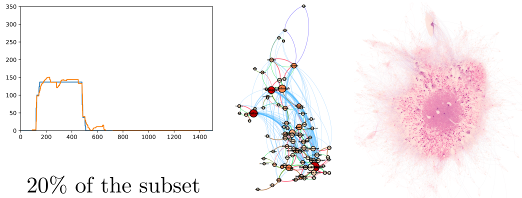
d) Graph Fourier Transform for Implicit Dynamics



i. Proposed data-driven GFT yields potential for band-limited sampling, which improves over CS and Laplacian approaches.

ii. Recovery with arbitrarily low errors, reducing sampling size by 10x compared to Laplacian and 2x compared CS approaches.

f) Deep Sampling Recovery of Dynamics



i) Recovery with low errors for discrete classification. Pathway to reducing sensors by 10x compared to approaches in (c) and (d).

ii) High impact in infrastructure sensing for Digital Twins, and reducing cost of data collection in social and biological sciences.

Main Contributions

Theoretical research

- Optimal sampling in dynamic dynamic networks
- Dynamics reconstruction using GFT and NNs
- Research in innovative neuroevolution techniques
- Green AI

Engineering

- Design of innovative infrastructure digital twins
- *10x* reduction required sensors
- *10x* improvement in dynamics reconstruction
- *10x* reduction in data collection
- Task-specific applications: resources management, predictive maintenance, etc.

Summary & Looking at Current & Future Work

1. Currently: Working with Industrial & Gov. Stakeholders to Deliver Impact
2. Next Steps: Data to Inform Posterior Risk Estimates
3. Developing EPSRC Fellowship & EPSRC P. Grant (Co-I)

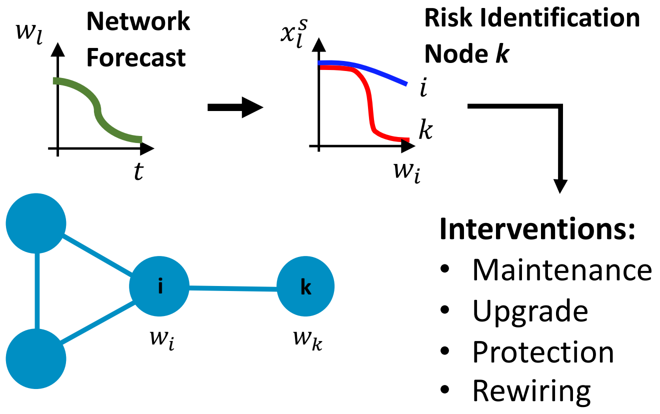
CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Summary, Next Steps & Impact

We have only been working on these important questions for 12 **months and have a long way to go.**

Better understand the relationship between:

- **Resilience & Robustness: Built & Natural Ecosystems**
- **Local Dynamics and Global Topology in Networked Ecosystems**



We can identify vulnerable nodes at the risk of losing resilience and this may go on to inform infrastructure operators.

We still need to couple ecosystems together and model higher dimensional dynamics.

We want the work here to inform real-time sensing and control systems as well as the design of new systems (topology and components).

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

References

[1] “Universal Resilience Patterns in Complex Networks,” J. Gao et al., Nature, 2016

[2] “Node Level Resilience Loss in Dynamic Complex Networks,” G. Moutsinas et al., NPG Sci. Rep., to appear, 2018

[3] ”Trophic Coherence Determines Food-Web Stability,” S. Johnson et al., PNAS, 2014

[4] “Drought Rewires the Core of Food Webs,” X. Lu et al., Nature Climate Change, 2016

[5] “Resilience or Robustness: Identifying Topological Vulnerabilities in Urban Rail Networks”, A. Pagani, et al., Royal Society Open Science, 2019

[6] Water Distribution Network Data Source: University of Kentucky - http://www.uky.edu/WDST/

[7] “Stability of Traffic Load Balancing in Complex Wireless Networks,” G. Moutsinas et al., IEEE Access, submitted, 2018

[8] “Optimal Sampling in Joint Time- and Graph-Domains for Dynamic Complex Networks with Graph Bandlimited Initialization,” Z. Wei, W. Guo et al., IEEE Access, submitted, 2019

[9] "Quantifying Networked Resilience via Multi-Scale Feedback Loops in Water Distribution Networks," A. Pagani, F. Meng, G. Fu, M. Musolesi, W. Guo, preprint, 2019 [arXiv]

[10] "Optimal Sampling of Water Distribution Network Dynamics using Graph Fourier Transform," Z. Wei, A. Pagani, G. Fu, I. Guymer, W. Chen, J. McCann, W. Guo, IEEE Trans. Network Science and Engineering, submitted, 2019 [arXiv]

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).

Thank you for listening

weisi.guo@warwick.ac.uk
wguo@turing.ac.uk

CHANCE funded by Lloyds Register Foundation. CoTRE funded by EPSRC (EP/R041725/1).