

Network Reliability Metrics

Finding the balance between topology, flow and statistical approaches

Juan Pablo Bertucci

I

Outline

- Introduction
- Possible formulations
- Our formulation
- Case study
- Final thoughts

Problem Description (1): Infrastructure and hazards

Infrastructure: the basic physical and organizational structures and facilities needed for the operation of a society or enterprise.

Hazard: event or agent with the potential to cause harm to a vulnerable target.

In the case of networks usually we are interested in the *flow* capabilities of something in the network post-hazard.

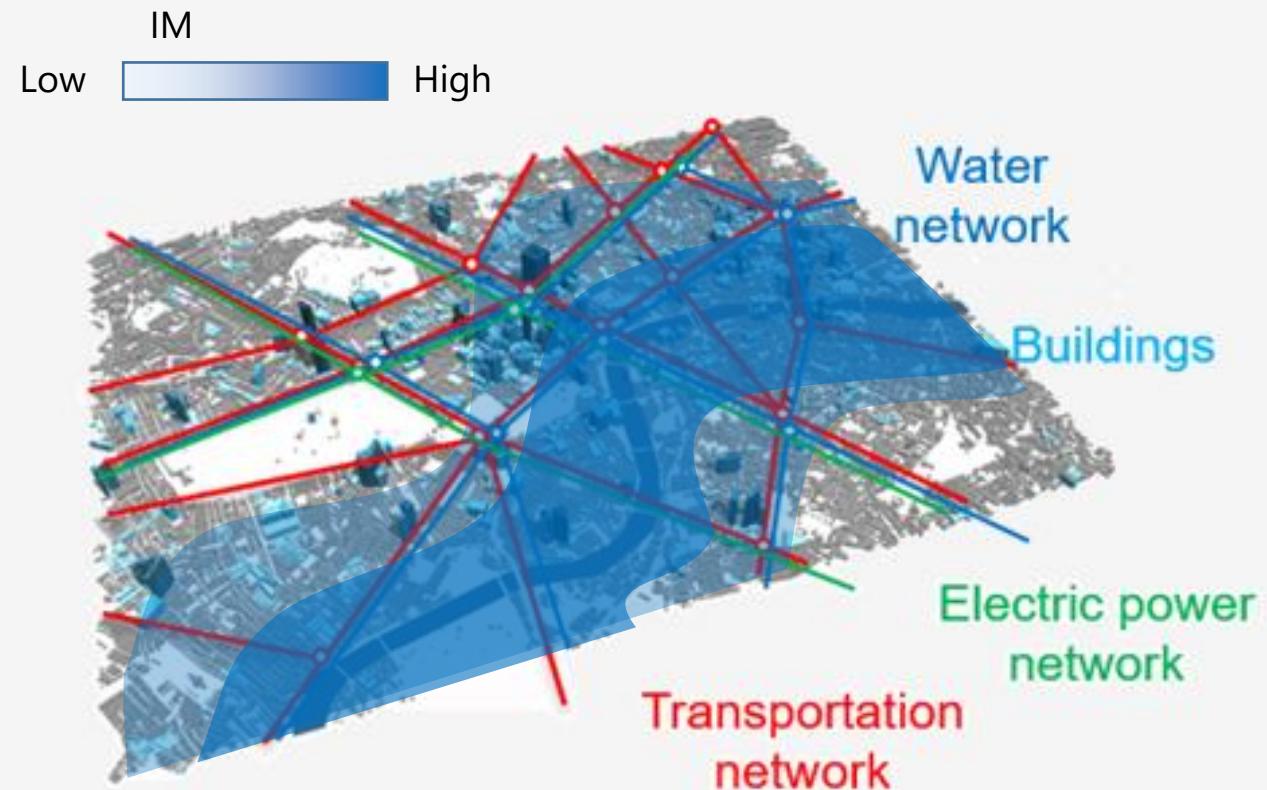
At a more basic level, we are interested in having all places of interest *at least connected*.



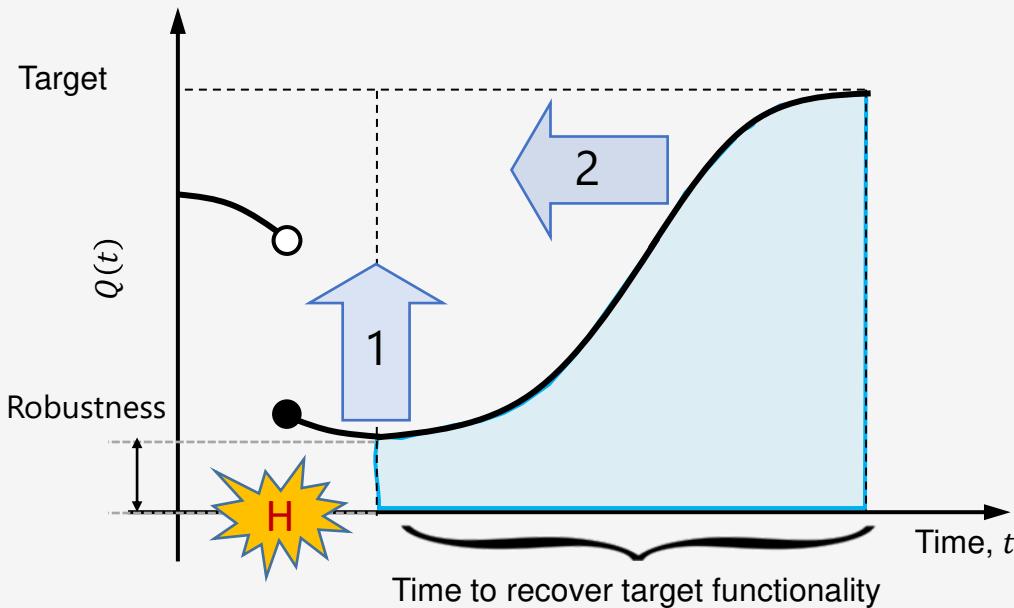
Problem Description (2): How to represent my network and event?

I am the administrator of a network and want to prepare it the best way possible to affront a hazard:

- Conceptualize the infrastructure network as a graph object. (Collection of interrelated spatial vertices)
 - At each vertex a certain level of service is provided
- Conceptualize the hazardous event as a spatial event with an intensity measure at each spatial point
 - This interferes in some way with the service provided
- How to gauge performance before, during, and after the event?

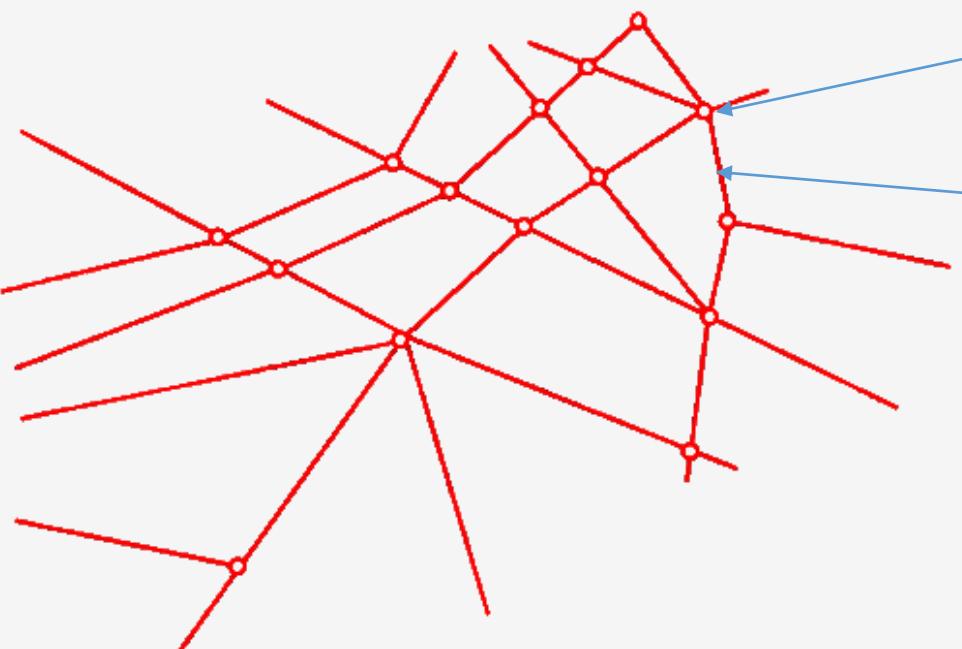


Problem Description (3): What do I want to improve? Resilience



- There is a response (negative) of the system to the event.
- What *system state* is of interest?
 - Something descriptive of level of service
 - Something descriptive of the cost
 - Something related to my network design
 - (layout, capacity, component reliability)
- Defining the metric $Q(t)$ is critical!
 - Capturing the loss of functionality
 - Descriptive of what could shift the curve in the directions (1) and (2)

Representing systems: graph theory



- Nodes, vertices, points, centres (i), (j)
 - Properties: Location, population, demand, $\# = n$
- Arcs, edges, links (i, j)
 - Properties: Start point (origin) End point (Destination), costs, lengths, capacity, time
- Graph: Collection of nodes joined by arcs. $G(N, A)$
 - Directed or undirected.
- Path: Sequence of connected arcs joining a specific origin (s) and destination (t)

Bollobás, B. (1998). *Modern Graph Theory*. New York: Springer Science+Business Media.

Dieter, J. (2008). *Graphs, Networks and Algorithms*. Berlin: Springer-Verlag

Newman, M. (2010). *Networks, An Introduction*. Oxford: Oxford University Press.

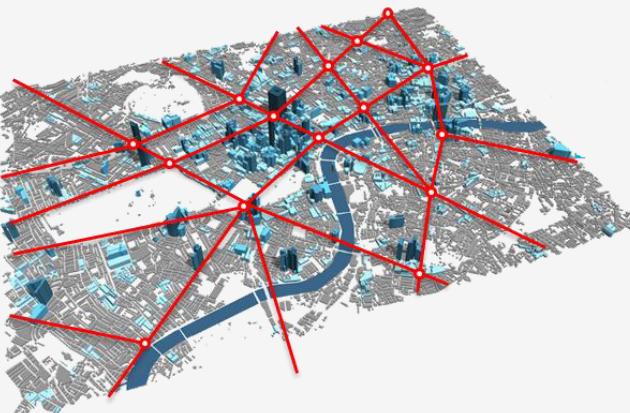
Outline

- Introduction
- Possible formulations
- Our formulation
- Case study
- Final thoughts

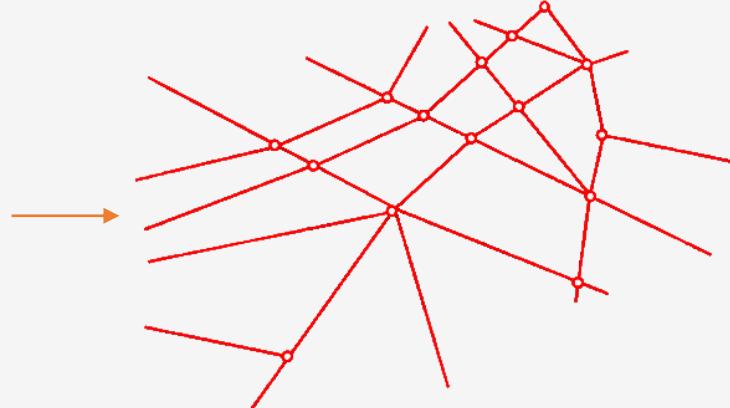
Network representation (1): adjacency matrix

There are many ways to write down the network as a graph object, the most common are:

- Edge list: $e_i = (s_i, t_i)$
- Adjacency matrix: $\mathbf{A} = [a_{ij}]$



Infrastructure Model



Network

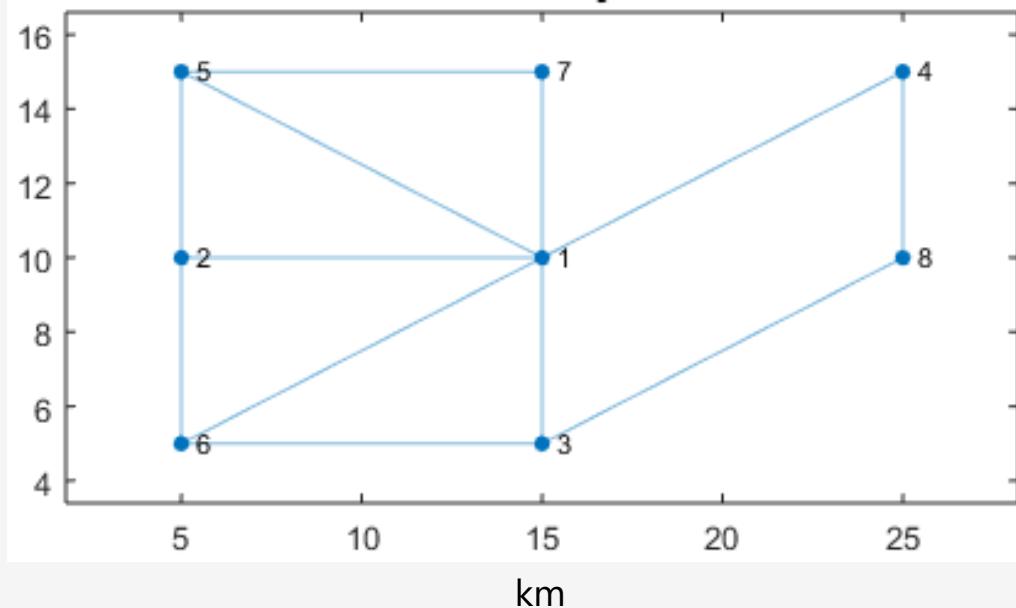
$$\mathbf{A} = \begin{array}{c|cccccc|c} & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ & 1 & 0 & 0 & 1 & 1 & 1 & 3 \\ & 0 & 1 & 1 & 0 & 0 & 0 & 4 \\ & 0 & 0 & 1 & 0 & 0 & 0 & 5 \\ & 0 & 0 & 1 & 0 & 0 & 0 & 6 \\ \hline \text{From} & & & & & & & \\ \text{To} & 1 & 2 & 3 & 4 & 5 & 6 & \\ \end{array}$$

Adjacency Matrix

Benchmark network

Data:

- 8 Nodes (x, y positions)
- 12 links



Original network considered:

- All links as bridges
- Nodes with different population values

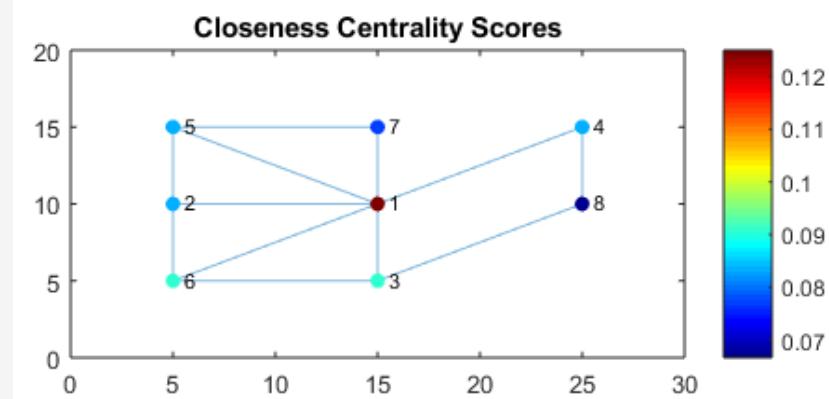
	To								
	1	2	3	4	5	6	7	8	
From	1	0	1	1	1	1	1	0	0
2	1	0	0	0	1	1	1	0	0
3	1	0	0	0	0	0	1	0	1
4	1	0	0	0	0	0	0	0	1
5	1	1	0	0	0	0	0	1	0
6	1	1	1	0	0	0	0	0	0
7	1	0	0	0	1	0	0	0	0
8	0	0	1	1	0	0	0	0	0

Kang, W.-H., Song, J., & Gardoni, P. (2008). Matrix-based system reliability method and applications to bridge networks. *Reliability Engineering and System Safety*, 1584-1593

Guidotti, R., Gardoni, P., & Chen, Y. (2017). Network reliability analysis with link and nodal weights and auxiliary nodes. *Structural Safety*, 12-26

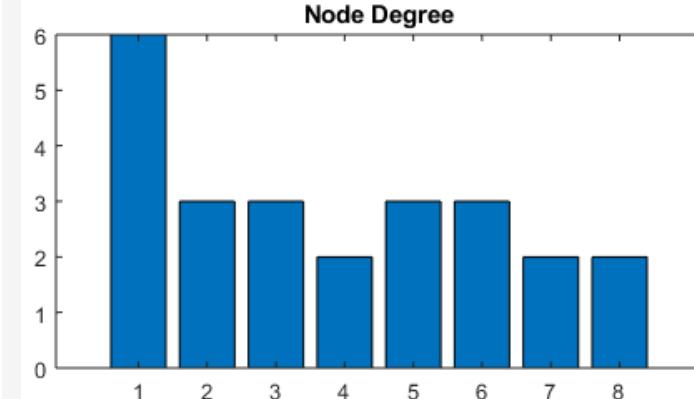
Some metrics that can be obtained from the graph structure

Characteristic path length : how many jumps do I have to make to reach different nodes? Average?



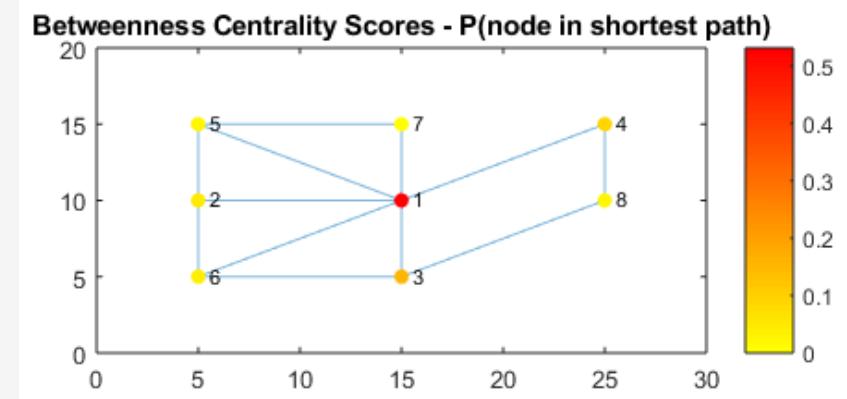
$$l_c(s) = \frac{1}{N-1} \sum_{s \neq j} l(s, j)$$

Degree: How many connections does each node have?



$$k_s = \sum_j a_{sj}$$

Centrality: How many shortest paths use a given node/edge?



$$g(e) = \sum_{e \in E} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

$$g(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'Small World' Networks. *Nature*, 440-442

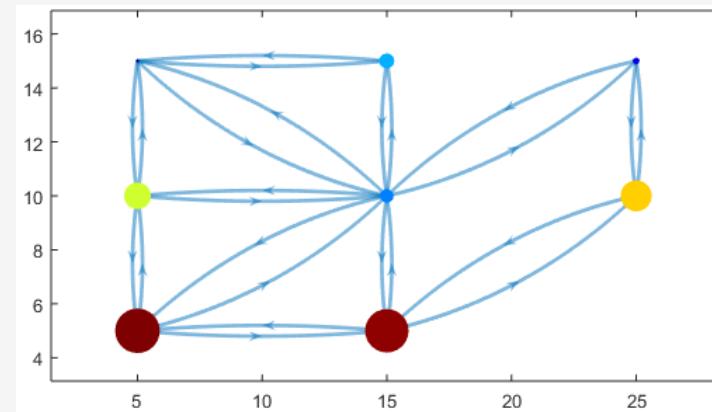
Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 47-97

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 378-382.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 1-101

Shortcomings of core topology metrics

- How to translate to demand and quality of service? Operational costs?
- Definition of network failure? Partial component failures are not allowed.
- Natural indicator of importance under homogeneous conditions. But how to add information about origin-destinations, populations, or other requirements consistently?

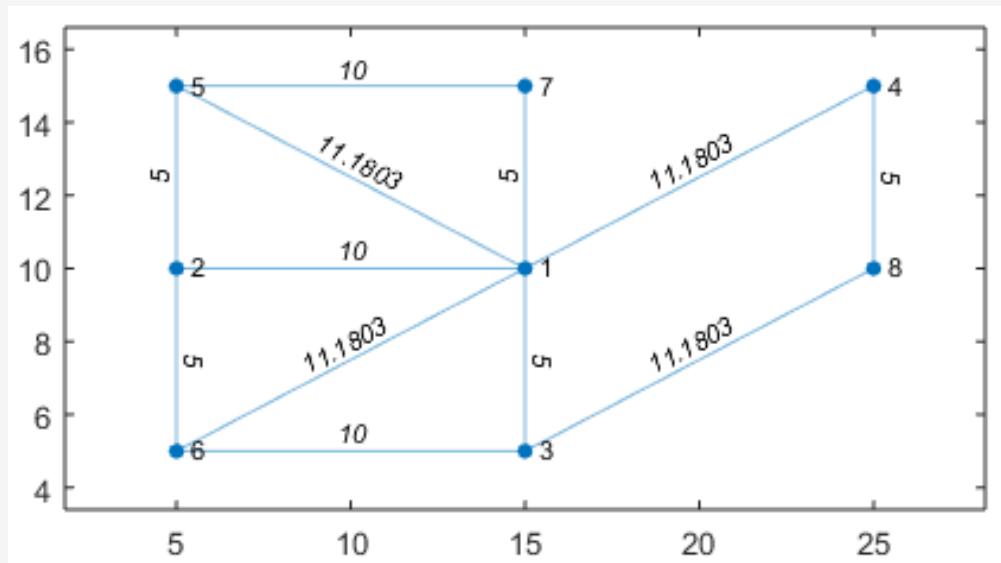


- They represent the topology at hand, we are generally interested in some related physical aspects.

Network representation (2): Cost weighted adjacency matrix

Consider there is a cost function (weight) for every edge. $C: e \rightarrow \mathbb{R}$

Graphs can be fully characterized by the weighted adjacency matrix $C = [c_{ij}]$



From	To								
	1	2	3	4	5	6	7	8	0
1	0	10	5	11.18	11.18	11.18	5	0	0
2	10	0	0	0	5	5	5	0	0
3	5	0	0	0	0	10	0	11.18	0
4	11.18	0	0	0	0	0	0	0	5
5	11.18	5	0	0	0	0	0	10	0
6	11.18	5	10	0	0	0	0	0	0
7	5	0	0	0	10	0	0	0	0
8	0	0	11.18	5	0	0	0	0	0

Weighted Adjacency Matrix (km)

Weighted connectivity metrics (1): diameter and eccentricity

1) Diameter (Average of shortest paths from s to j)

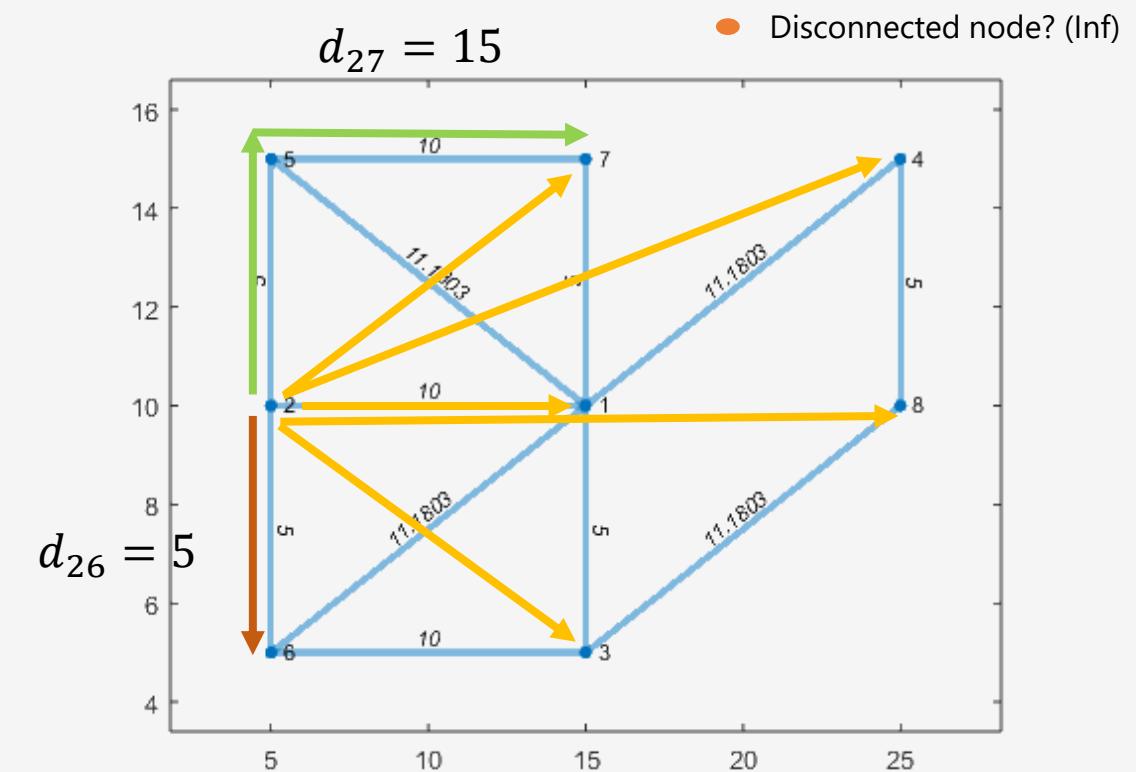
$$\text{Local} \quad \delta_s = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq s}}^n d_{sj}$$

$$\text{Global} \quad \delta = \frac{1}{n} \sum_{i=1}^n \delta_i = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n d_{ij}$$

2) Eccentricity: ("standard deviation" of δ_i)

$$\text{Local} \quad \zeta_s = \sqrt{\frac{1}{(n-1)-1} \left[\sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{d_{sj}}{d_{s,opt}} - \bar{\delta}_s \right)^2 \right]}$$

$$\text{Global} \quad \zeta = \sqrt{\frac{1}{n(n-1)-1} \left[\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{d_{ij}}{d_{i,opt}} - \bar{\delta} \right)^2 \right]}$$



$$\delta_2 = \sum_{j=1}^8 \delta_{2j} = \frac{(\delta_{21} + \delta_{23} + \dots + \delta_{27} + \delta_{28})}{7}$$

Weighted connectivity metrics (2): Efficiency and heterogeneity

3) Efficiency: (Reciprocal of diameter)

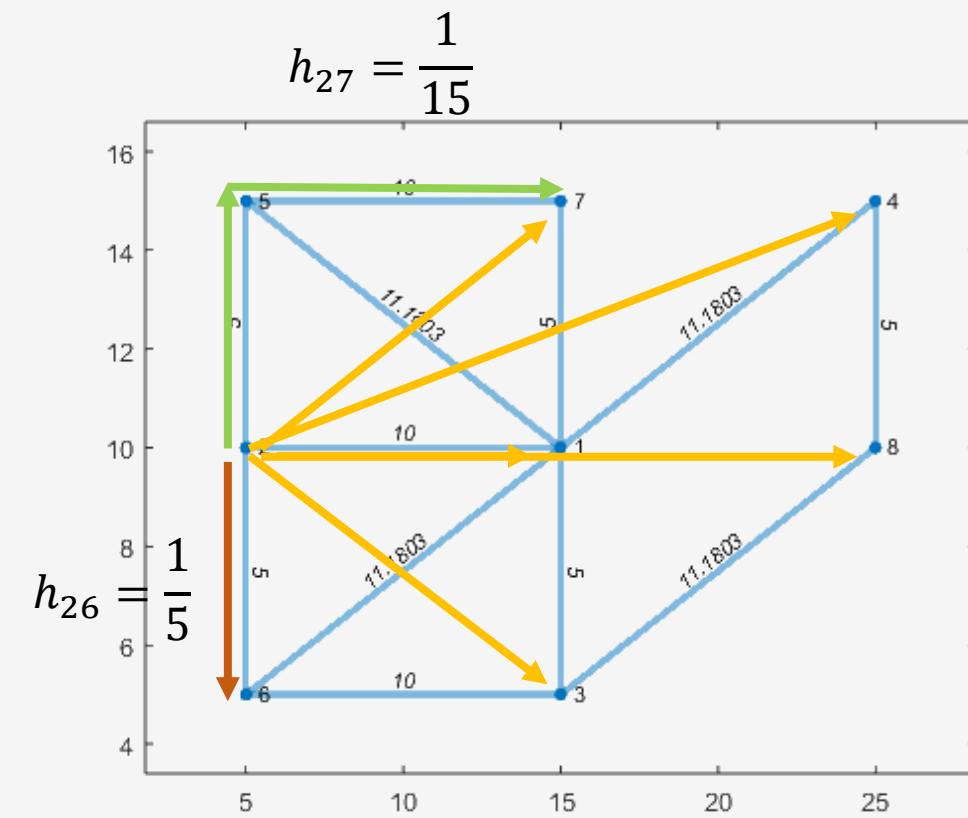
Local $\eta_s = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq s}}^n \frac{1}{d_{sj}} = \frac{1}{n-1} \sum_{j=1}^n h_{sj}$

Global $\eta = \frac{1}{n} \sum_{i=1}^n \eta_i = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n h_{ij}$

4) Heterogeneity ("standard deviation" of η_i)

Local $\psi_i = \sqrt{\frac{1}{(n-1)-1} \left[\sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{h_{ij}}{\eta_{i,opt}} - \bar{\eta}_i \right)^2 \right]}$

Global $\psi = \sqrt{\frac{1}{n(n-1)-1} \left[\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{d_{ij}}{\eta_{i,opt}} - \bar{\eta} \right)^2 \right]}$

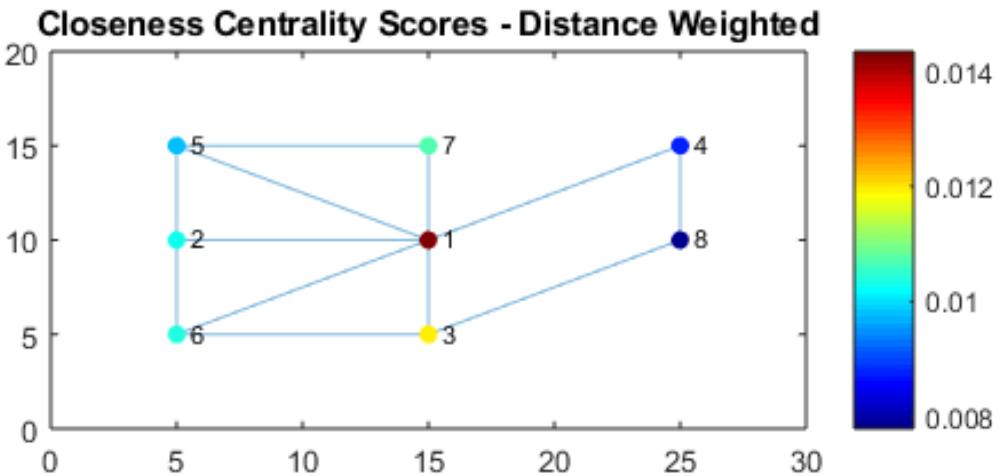


$$\eta_2 = \sum_{j=1}^8 h_{2j} = \frac{(h_{24} + h_{23} + \dots + h_{28})}{7}$$

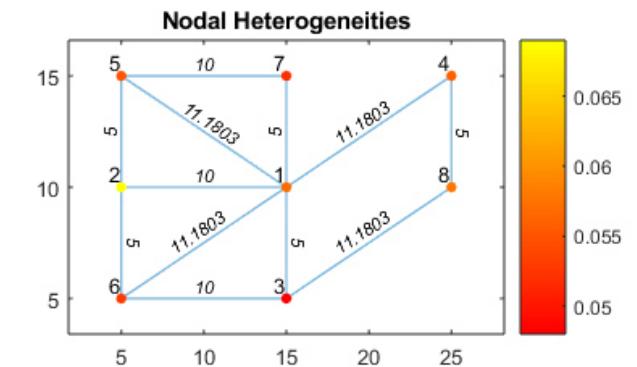
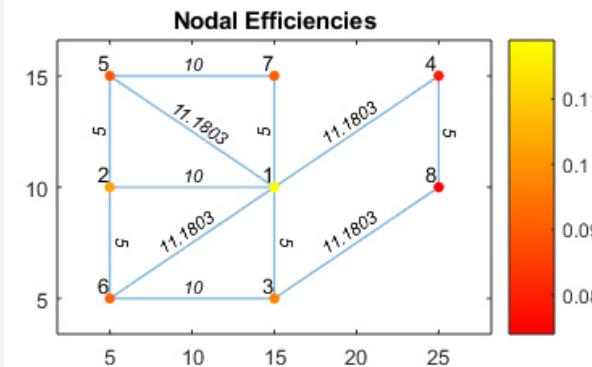
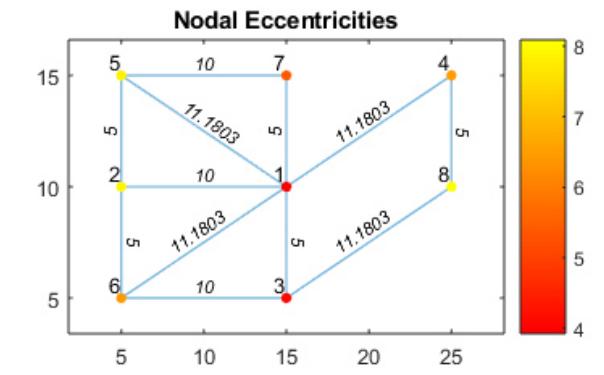
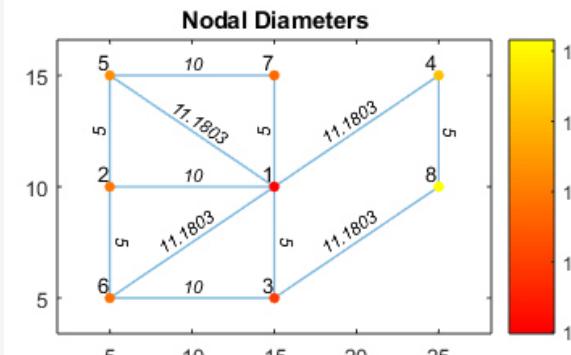
Some interesting metrics that can be obtained from weighted representation

Weighted betweenness

$$g(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$



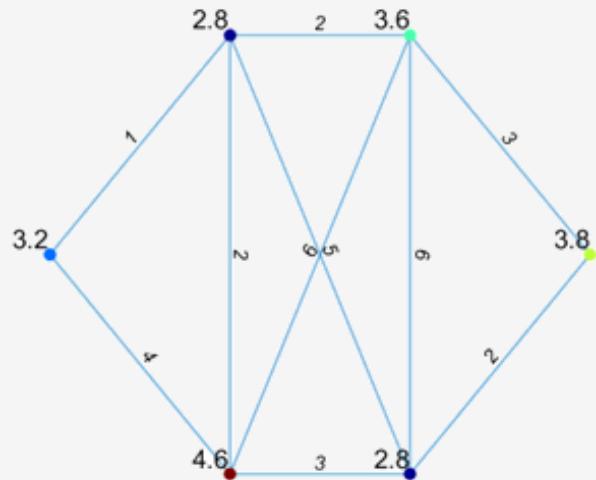
Diameter, eccentricity, efficiency, heterogeneity



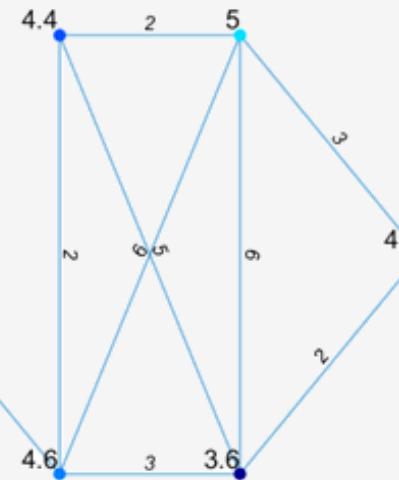
Sample simulation



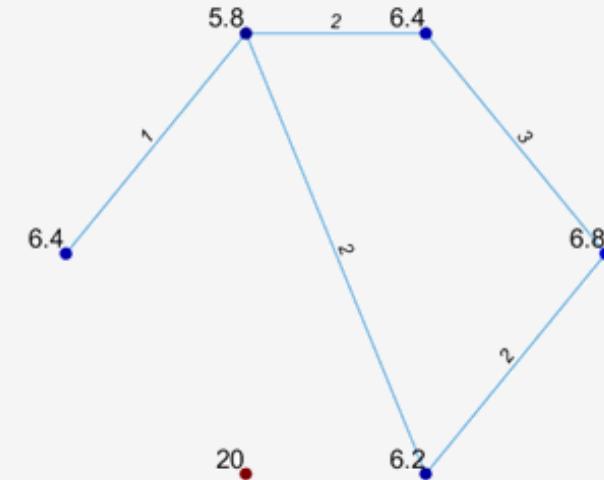
IM_1



$IM_2 > IM_1$



$IM_3 > IM_2$



$$\delta = 3.47$$

$$\delta = 5.00$$

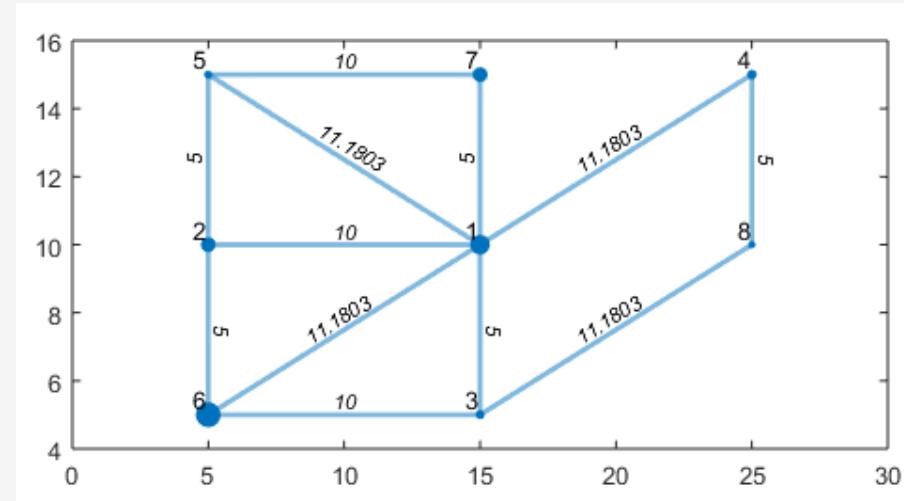
$$\delta = 8.60$$

As the hazard intensity measure increases the diameter increases (locally and globally), i.e: the weighted connectivity drops

Problems with cost weighted approach

Picking the right link **cost** reflecting service level and operational costs.

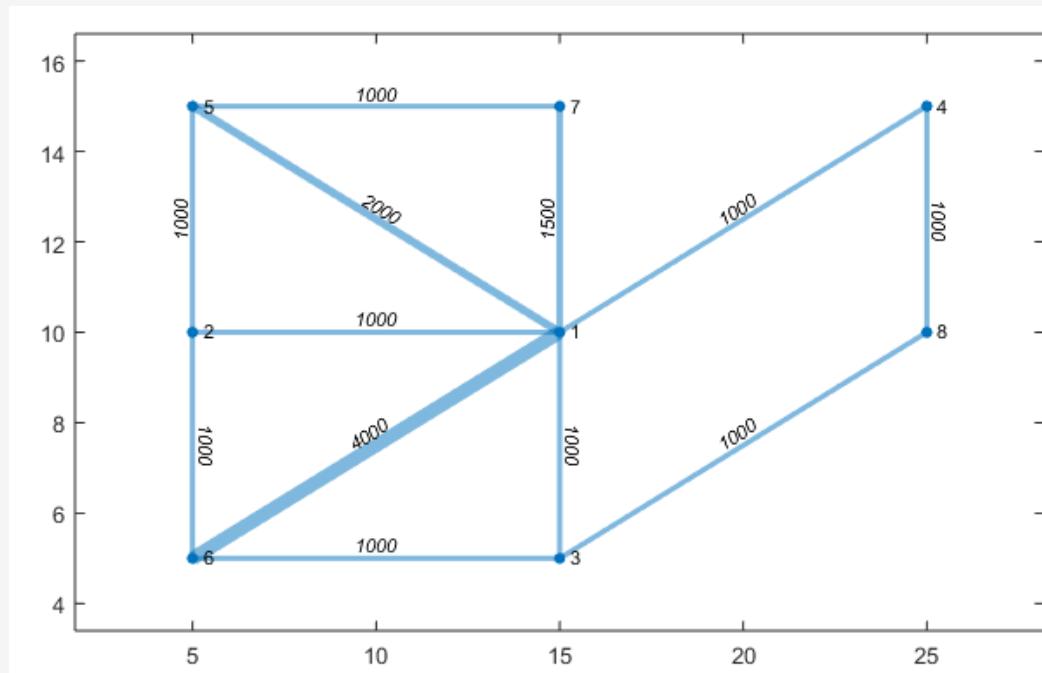
What if some nodes are more critical in the network than others? (example: relative population size)



If the cost chosen and the node relevancy is accounted for, there is still the question of *how much flow* can the network actually provide to each node?

Network representation (3): Capacity weighted adjacency matrix

There is a capacity function for each link, yielding the maximum amount of flow units that it can conduct
 $U: e \rightarrow \mathbb{R}$

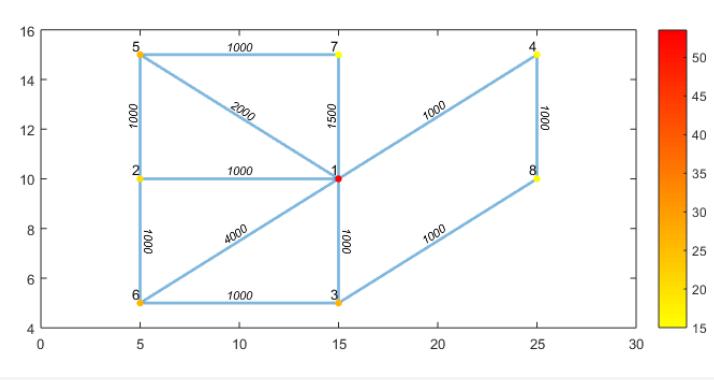


	To								
From	1	2	3	4	5	6	7	8	
1	0	1000	1000	1000	2000	2000	1500	0	
2	1000	0	0	0	0	1000	1000	0	0
3	1000	0	0	0	0	0	1000	0	1000
4	1000	0	0	0	0	0	0	0	1000
5	2000	1000	0	0	0	0	0	1000	0
6	2000	1000	1000	0	0	0	0	0	0
7	1500	0	0	0	1000	0	0	0	0
8	0	0	1000	1000	0	0	0	0	0

Capacity-Weighted Adjacency Matrix

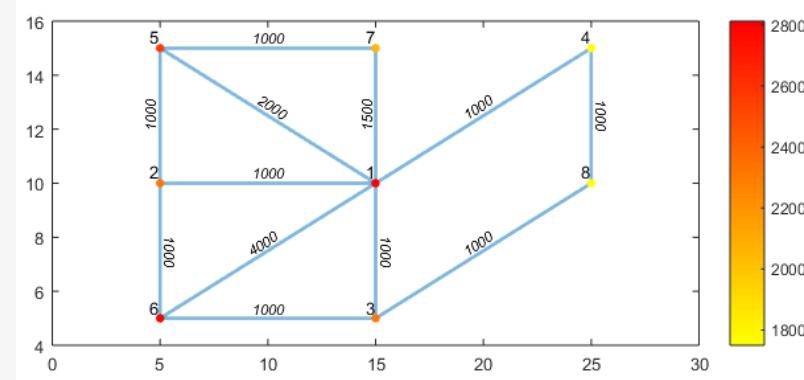
Some interesting metrics we can get from capacitated representation

Weighted degree: number of connections weighted by strength

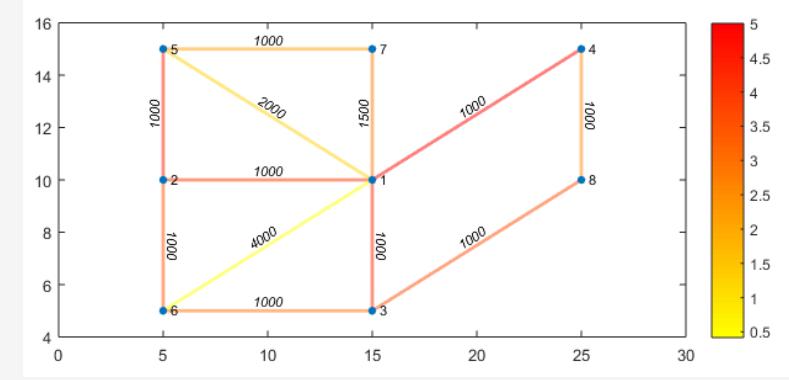


$$s_i = \sum_{j=1}^N a_{ij} u_{ij}$$

Mean maximum flow: How much flow can get there on average



Min cut edges: which edges are critical for flow more often



- Dispersion measures between different nodes indicates relative performance
- Analysis under disruption situation: get $Q(t)$

Problems with flow metrics

- What are the flow dynamics?
- What happens with flow interactions? Interdependency between different OD pairs.
- I have a max flow, but what is the demand? (V/C)
- I can fulfil a certain demand, but how effective is the distribution? (water, i want good pressure.
Transportation, I want to get there fast.)
- I can fulfil a certain amount, but how costly is the operation? What are the final recommendations?
- Some composite indices exist, but they involve the weighing of different properties. (NRI, WIPW)

Are there intrinsic metrics that...

- Are simple enough
- Are general enough
- Are informative enough
- Capture the events of interest and network reliability

global metrics

consider flows

applicable to
extreme events

practical for
design

Outline

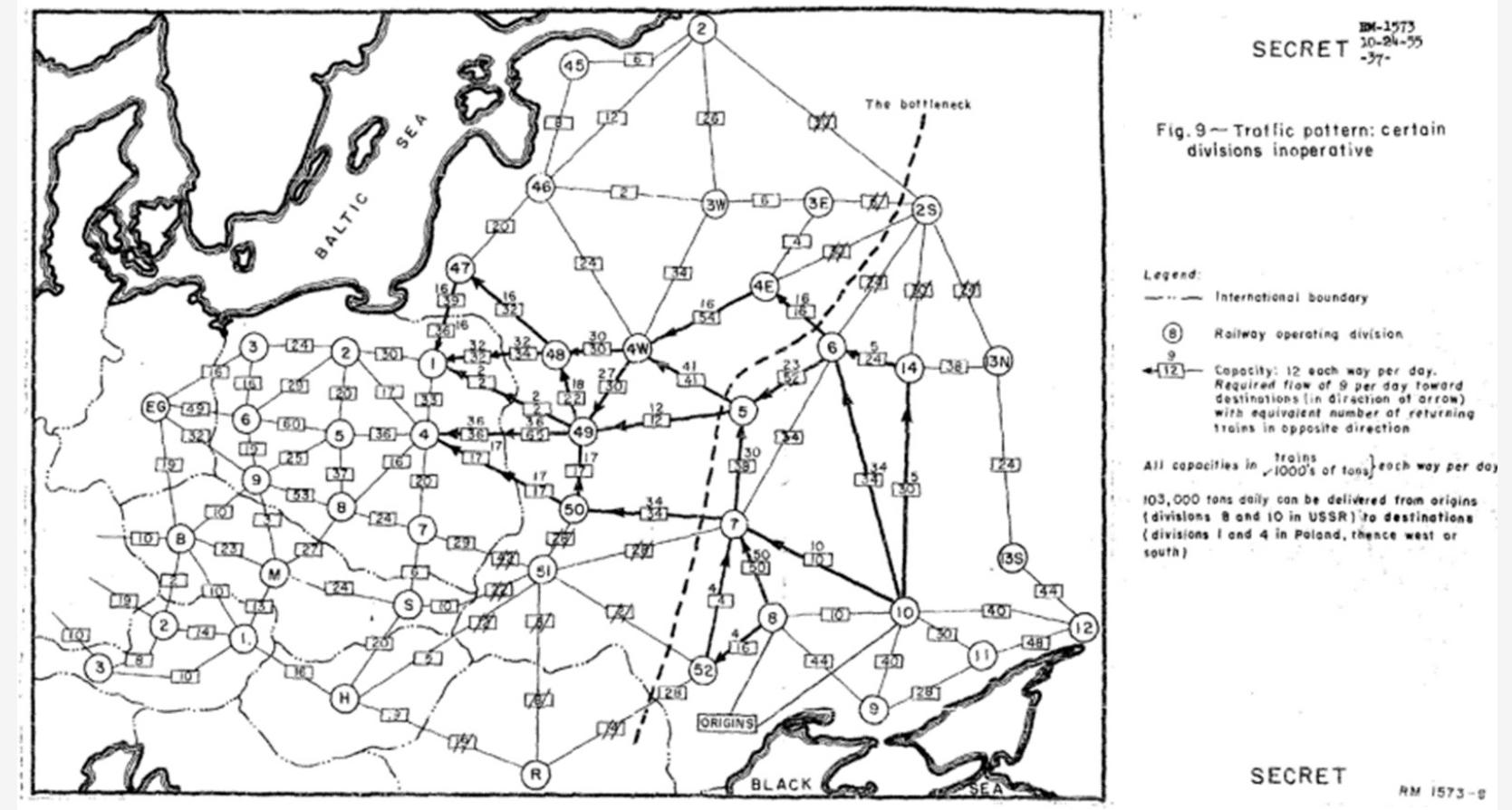
- Introduction
- Possible formulations
- Our formulation
- Case study
- Final thoughts

Graph theory → Optimization / Operations Research

There is an intimate relation between topology and reliability

But we need to account for the flow and a real equilibrium that is taking place in our network.

Understanding the fundamental metrics of topology, can we extend them in a general way that accounts for our problems?



Capacitated minimum cost flow formulation

$$z_{st}^* := \min \sum_{(i,j) \in A} c_{ij} x_{ij}$$

S.t:

$$\sum_{j:(s,j) \in A} x_{sj} = T \quad \sum_{i:(i,t) \in A} x_{it} = -T$$

$$\sum_{j:(i,j) \in A} x_{ij} - \sum_{j:(j,i) \in A} x_{ji} = 0 \quad \forall i \in N \setminus \{s, t\}$$

$$x_{ij} \leq u_{ij} \quad \forall (i, j) \in A \quad \xrightarrow{\hspace{1cm}} \quad \omega_{ij}$$

x_{ij} = flow on arc (i, j)

u_{ij} = capacity of arc (i, j)

c_{ij} = unit cost of transiting arc (i, j)

Objective: Minimize traversal cost

Origin and destination constraint

Flow conservation along path

Capacity Constraint (and dual)

This linear program solves the feasible minimum cost routing of T from s to t

Extended diameter formulation: minimum cost flow

$$z_{ij}^* := \min \sum_{(i,j) \in A} c_{ij} x_{ij}$$

$$\sum_{j:(s,j) \in A} x_{sj} = T \quad \sum_{i:(i,t) \in A} x_{it} = T$$

$$\sum_{j:(i,j) \in A} x_{ij} - \sum_{j:(j,i) \in A} x_{ji} = 0 \quad \forall i \in N \setminus \{s, t\}$$

$$x_{ij} \leq u_{ij} \quad \forall (i,j) \in A$$

x_{ij} = flow on arc (i,j)

u_{ij} = capacity of arc (i,j)

c_{ij} = unit cost of transiting arc (i,j)

Objective: Minimize traversal cost

Origin and destination constraint

Flow conservation along path

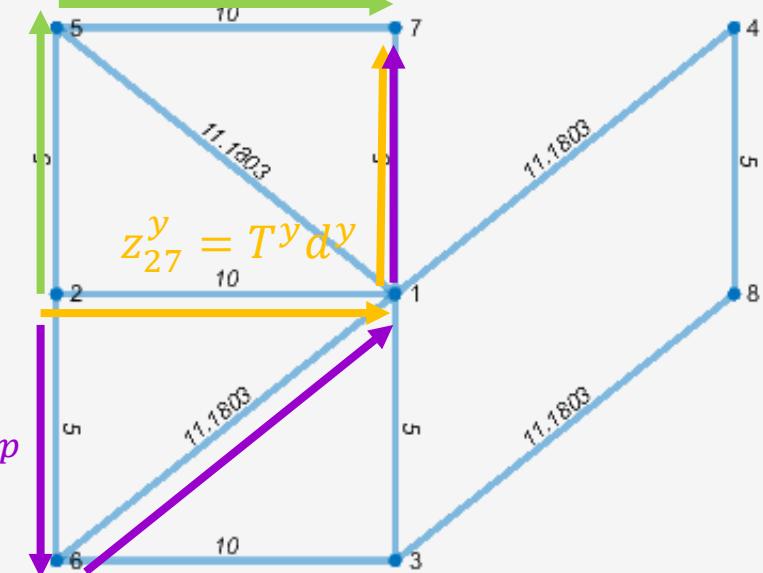
Capacity Constraint

$z_{st,T}^* = \text{minimum total traversal cost from } s \text{ to } t \text{ for flow } T \text{ (ex: veh. mi. travelled)}$

$$z_{27}^g = T^g d^g$$

$$z_{27}^y = T^y d^y$$

$$z_{27}^p = T^p d^p$$



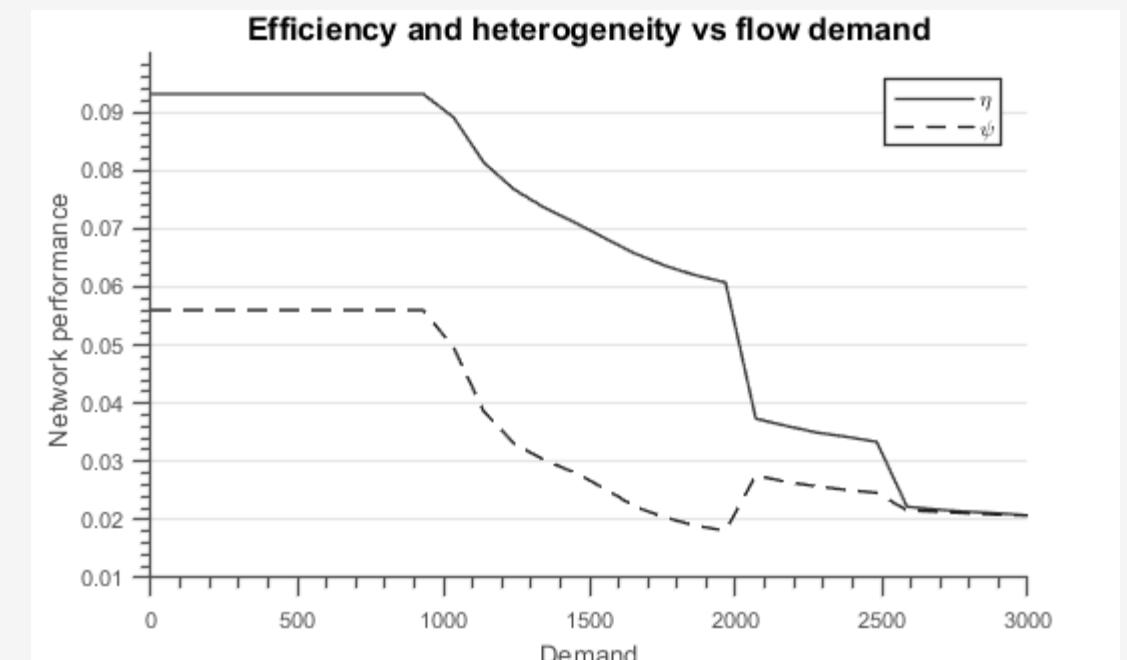
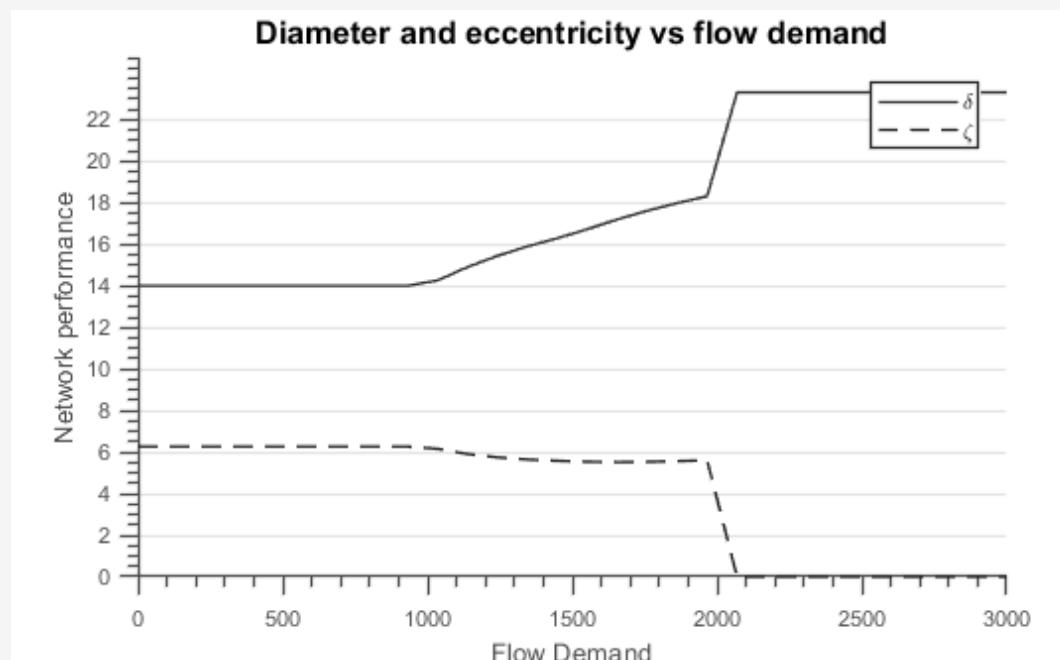
Define new "shortest path" between s and t as:

$$\frac{z_{st,T}^*}{\sum_{i \neq j} x_{ij}} = \frac{z_{st,T}^*}{T} = \delta_{st}^{mc} = \text{generalized shortest path from } s \text{ to } t \text{ (shortest hyperpath)}$$

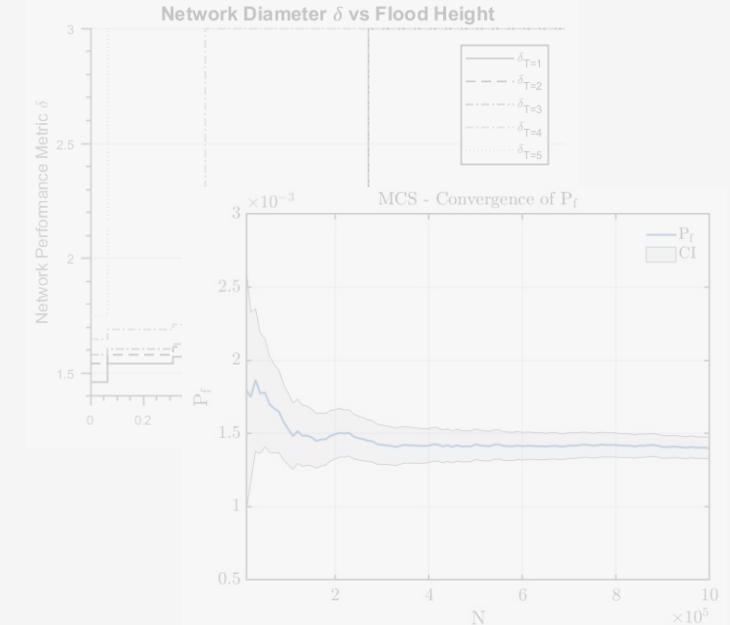
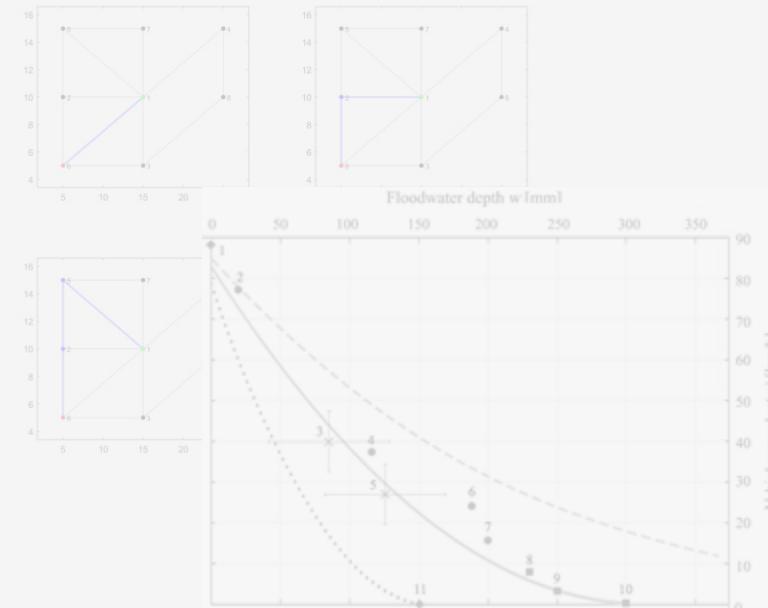
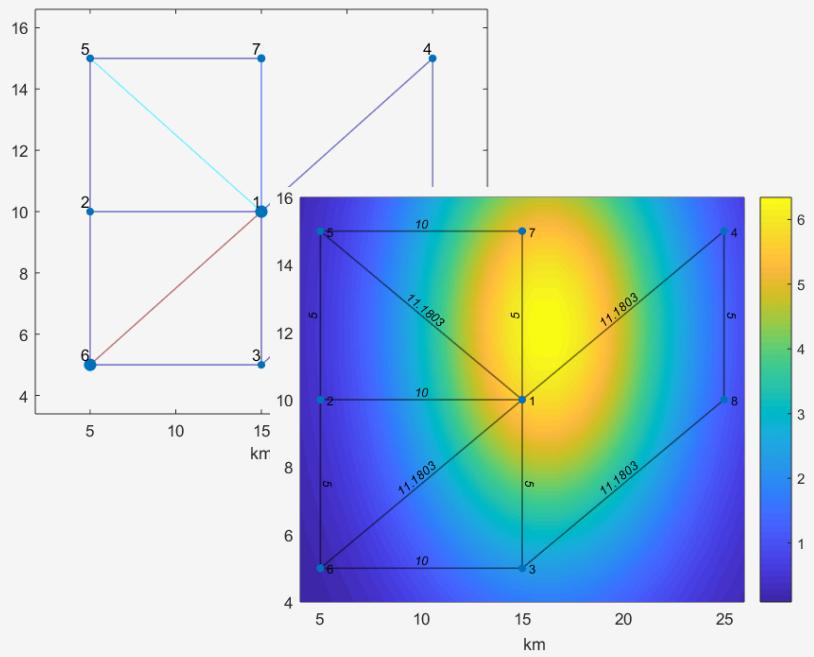
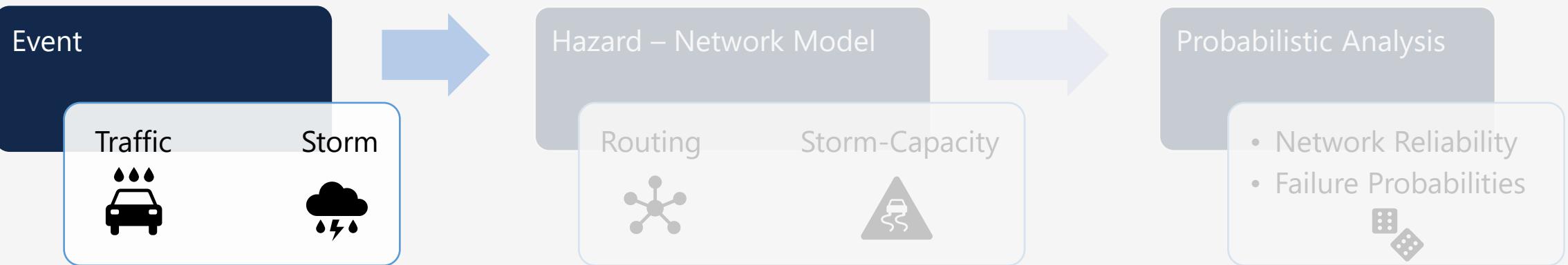
Part 1: stress test

Generate a set of starting paths, successively find paths as they get saturated with increasing flow requirements.
(Augmenting paths)

Characteristic curve: Performance of the network with increasing demands, where are the critical performance drops.

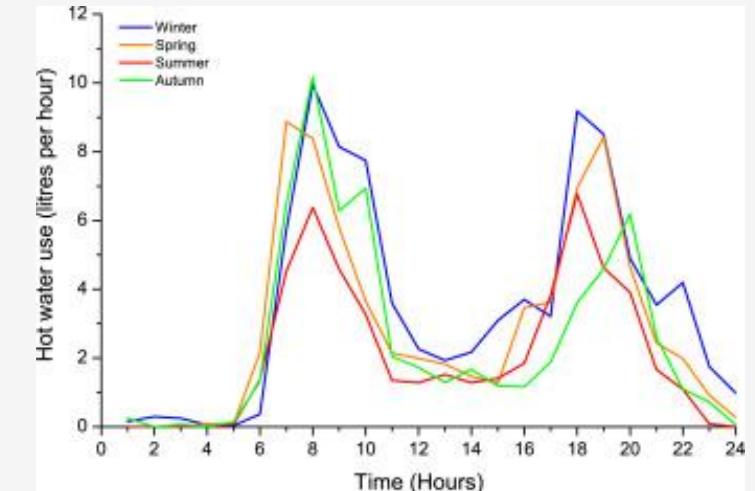
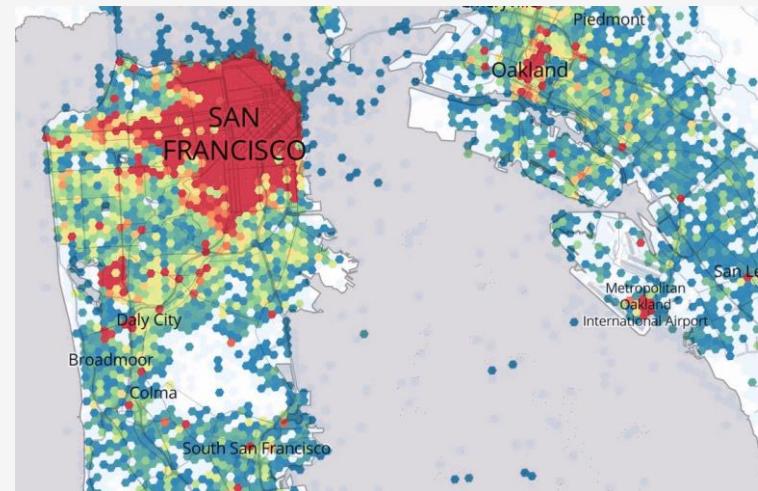
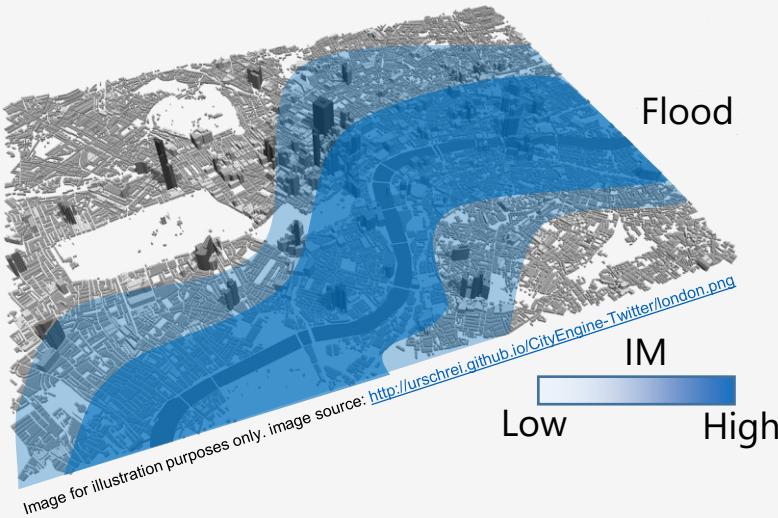


Part 2: probabilistic analysis (traffic example)



Hazard and demand modelling simplification: Shape vs Intensity

Frequently there are specific hazards that we prepare for in certain locations. We assume that studying a specific hazard has shape parameters given Θ_s , and vary its intensity through IM

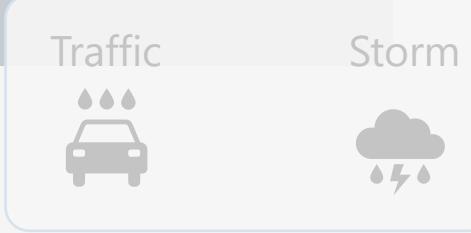


Although a strong assumption and limiting the analysis, it is very relevant for two reasons:

- makes the *problem* convex
- It is applicable for most practical purposes

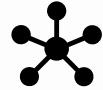
Part 2: probabilistic analysis (traffic example)

Event



Hazard – Network Model

Routing

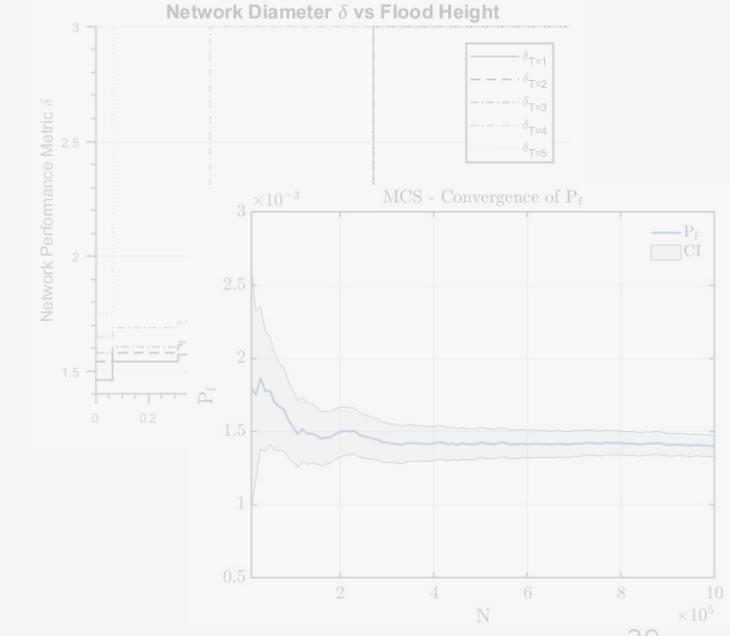
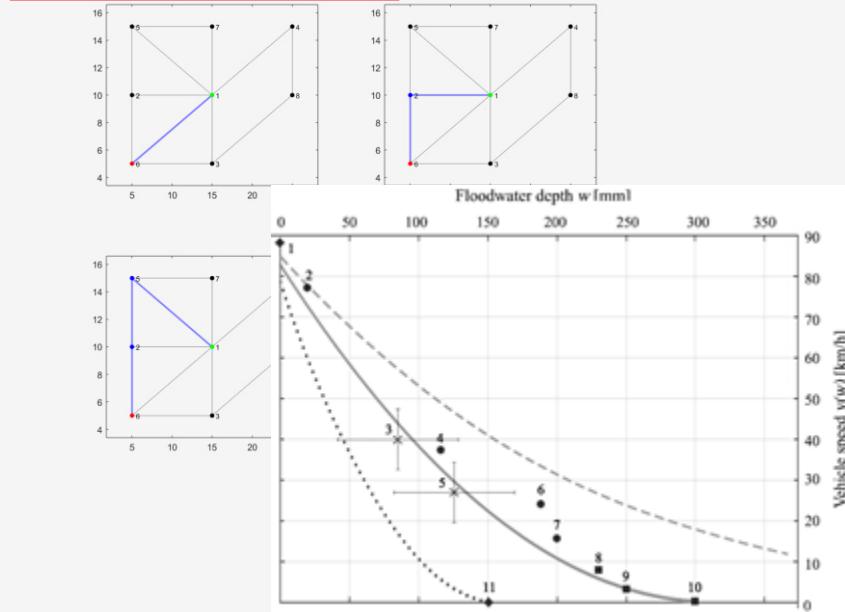
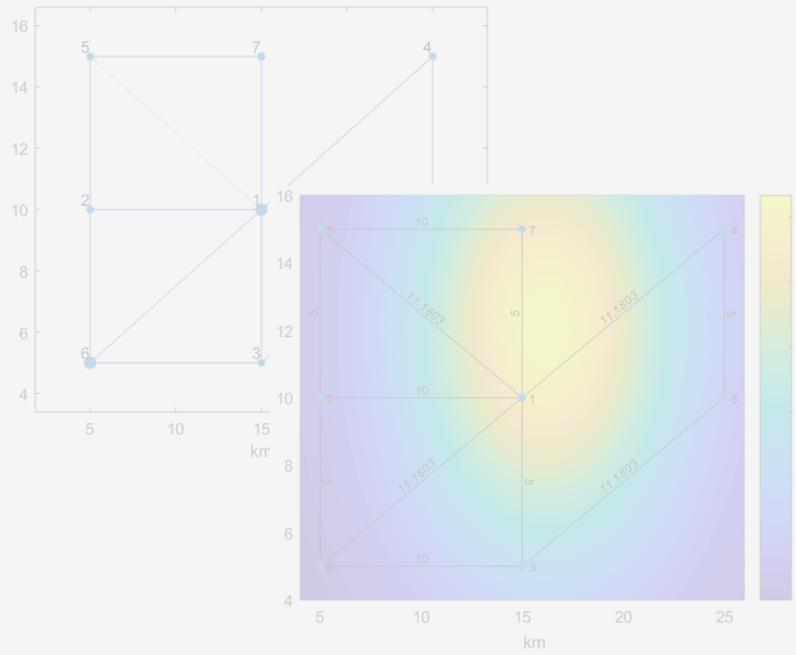


Storm-Capacity



Probabilistic Analysis

- Network Reliability
- Failure Probabilities



Overcoming computational cost: Path based formulation

$$1. \quad z_{st}^* := \min \sum_{p \in P} \left(\sum_{(i,j) \in A} c_{ij} x_{ij}^p \right) f(p)$$

Such that:

$$2. \quad \sum_{p \in P} d_{ij}(p) f(p) \leq u_{ij}$$

Keys:

c_{ij} = cost on arc (i, j)

$\sum_{(i,j) \in A} c_{ij} x_{ij}^p$ = cost of a path p

$$3. \quad \sum_{p \in P} f_p = T$$

$f(P)$ = flow on a path p

$d_{ij}(P_{st}) = \mathbf{1}[(i, j) \in P_{st}]$

$$4. \quad f_p \geq 0 \quad \forall p \in P$$

The variables we now explore are paths instead of links. Is this actually helpful?

Overcoming computational cost (2): Restricted path subproblem

1) Given a path made of links $(i, j) \in P$, we can quickly evaluate:

$$C(P) = \sum_{(i,j) \in A} c_{ij} x_{ij}^p \quad - \text{cost of path } P$$

2) We solve:

$$\min \sum_{k \in K} \sum_{p \in P} C(P) f(P)$$

s.t.:

$$\sum_{k \in K} \sum_{p \in P} \delta_{ij}(P) f(P) \leq u_{ij} \quad \forall (i, j) \in A \quad \xrightarrow{\text{orange arrow}} \omega_{ij}$$

$$\sum_{p \in P} f(P) = d^k \quad \forall k \in K \quad \xrightarrow{\text{orange arrow}} \sigma_k$$

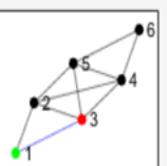
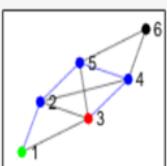
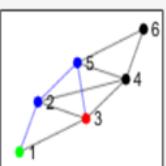
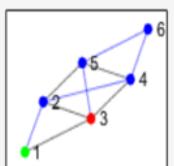
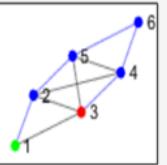
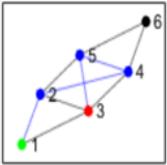
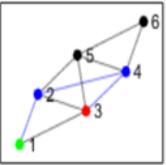
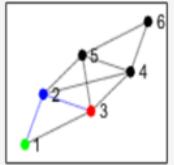
Dual Variables

$$f(P) \geq 0 \quad \forall k \in K \text{ & } \forall p \in P$$

3) Obtain dual variables, define a new network over “modified path costs”:

$$c_P^{\sigma, w} = c(P) + \sum_{(i,j) \in P} \omega_{i,j} - \sigma$$

Overcoming computational cost (3): Column Generation Procedure



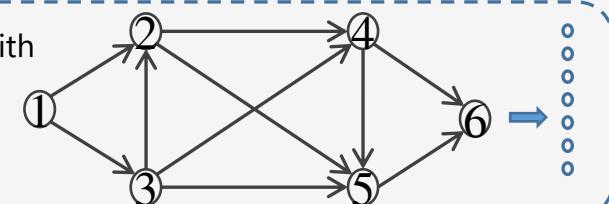
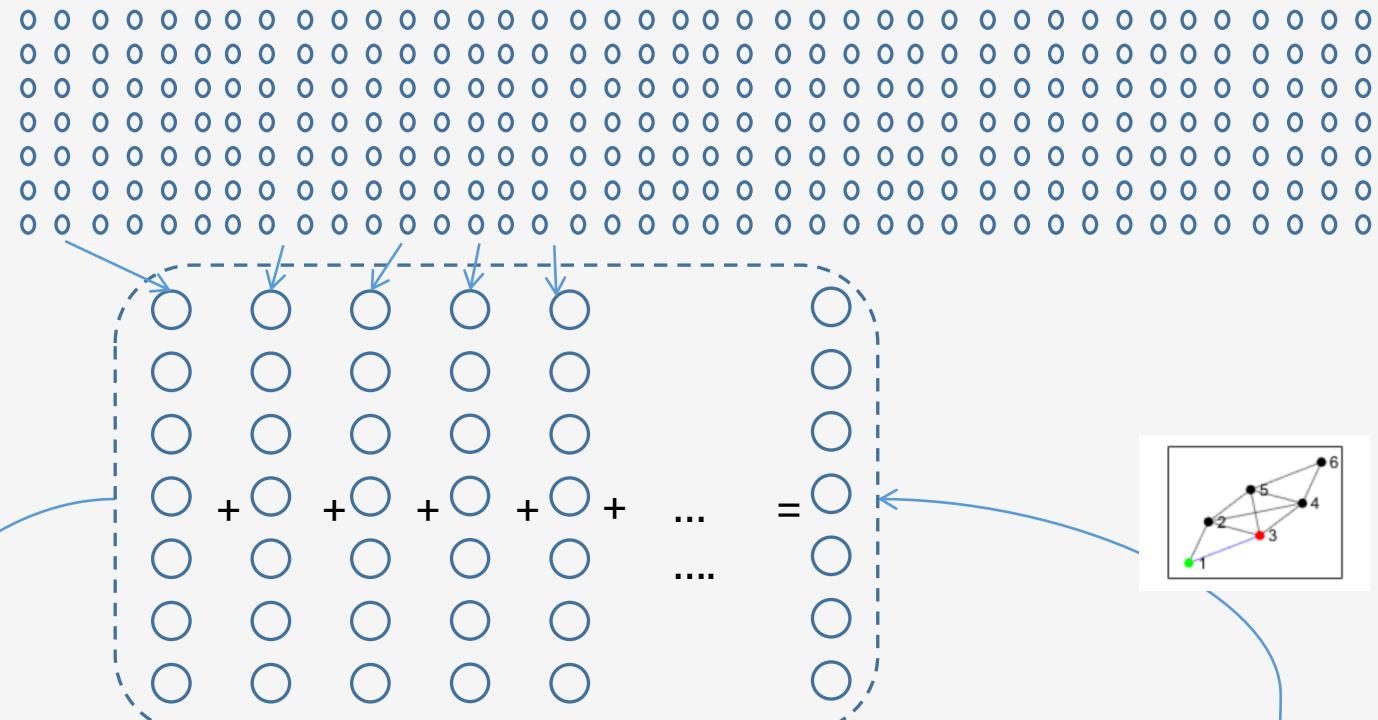
The initial search space is exponentially large (NP-Hard)

We use column generation to add relevant paths

Store the structure obtained for faster successive evaluations!!

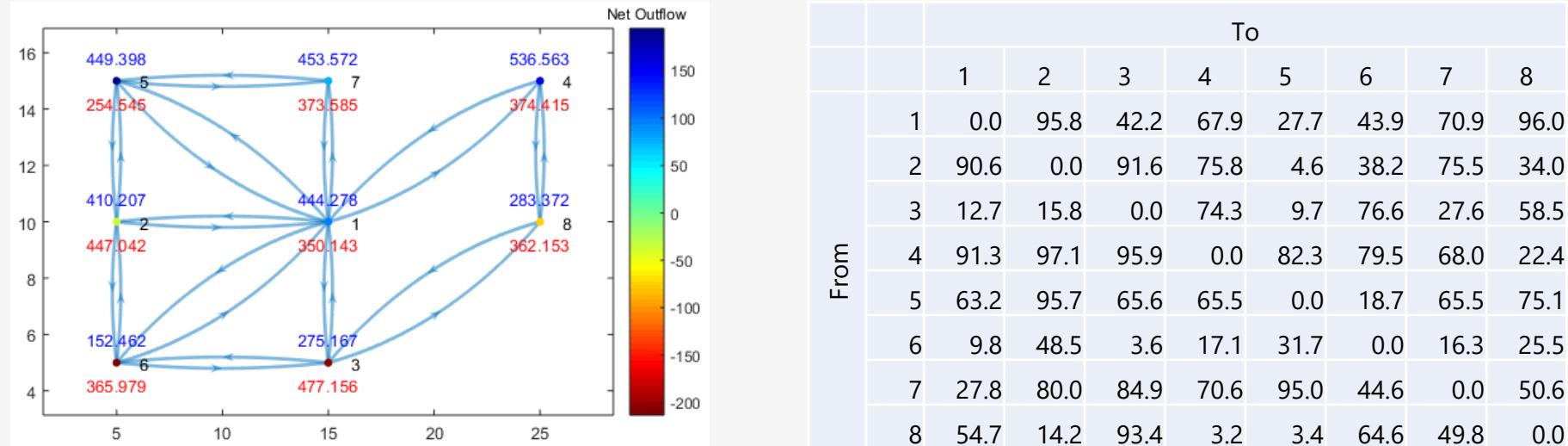
Restricted
master
Problem:

"Pricing"
Problem:



Extra problem constraint: Interaction between flows -> MCF

The problem addressed so far considers how to route one source to one destination in the absence of other flows. We have to switch into a multi commodity flow analysis.

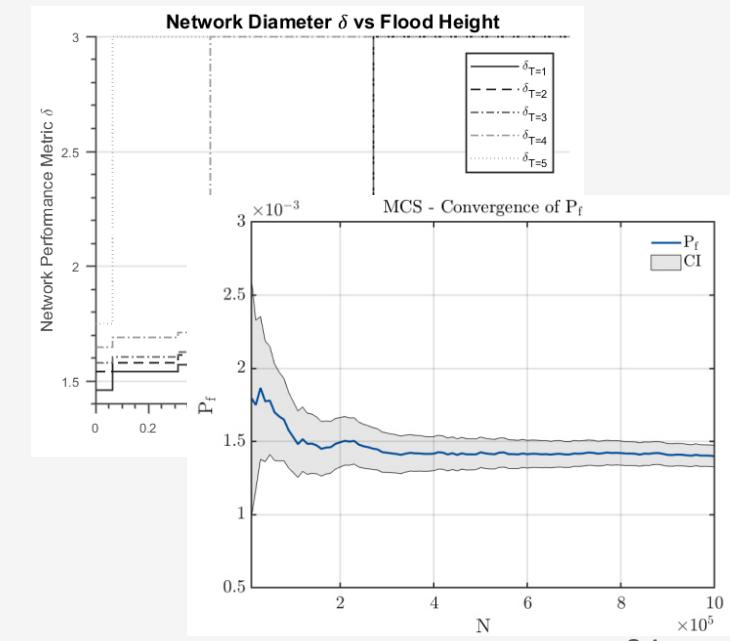
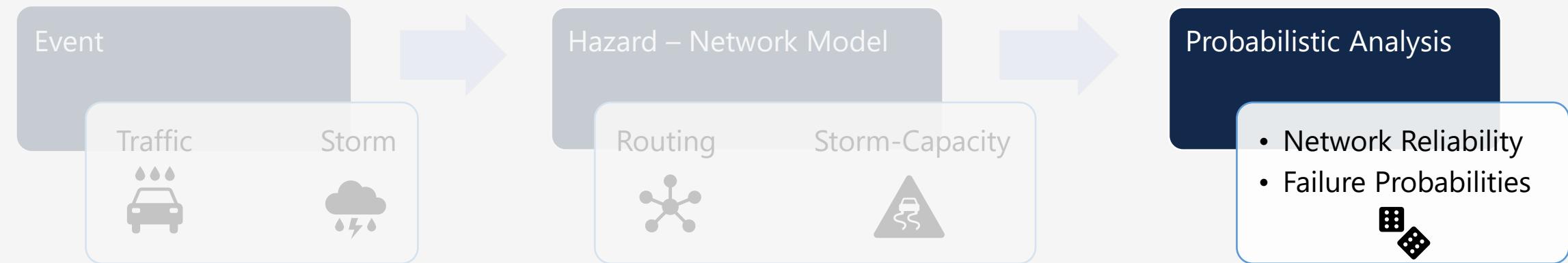


If we are interested in having a specific set of origins, that go to a specific set of destinations, we need to specify each origin-destination pair as a new "commodity" $k = (s,t)$

The indicator function $d_{ij}(p)$ becomes $d_{ij}^k(p)$

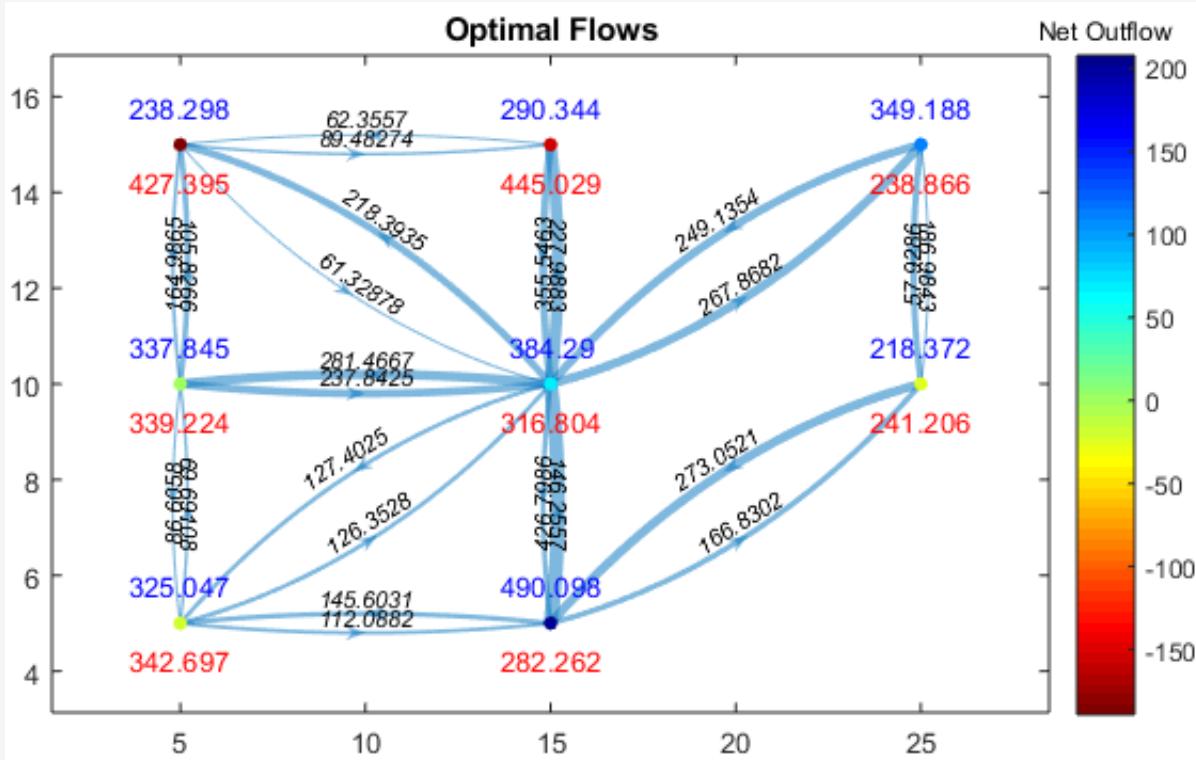
$$\sum_{p \in P^k} d_{ij}^k(p) f(p) \leq u_{ij}$$

Part 2: probabilistic analysis (traffic example)

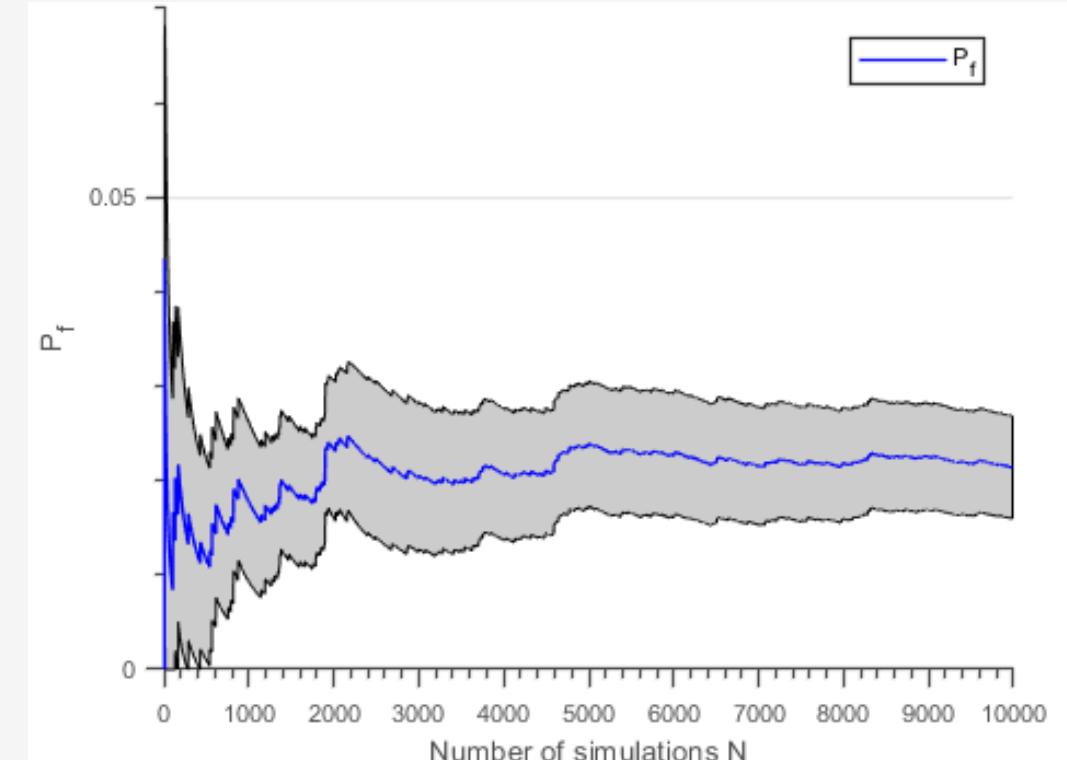


Example: one iteration and markov chain monte carlo

Sample MCF flows for a specific demand and IM



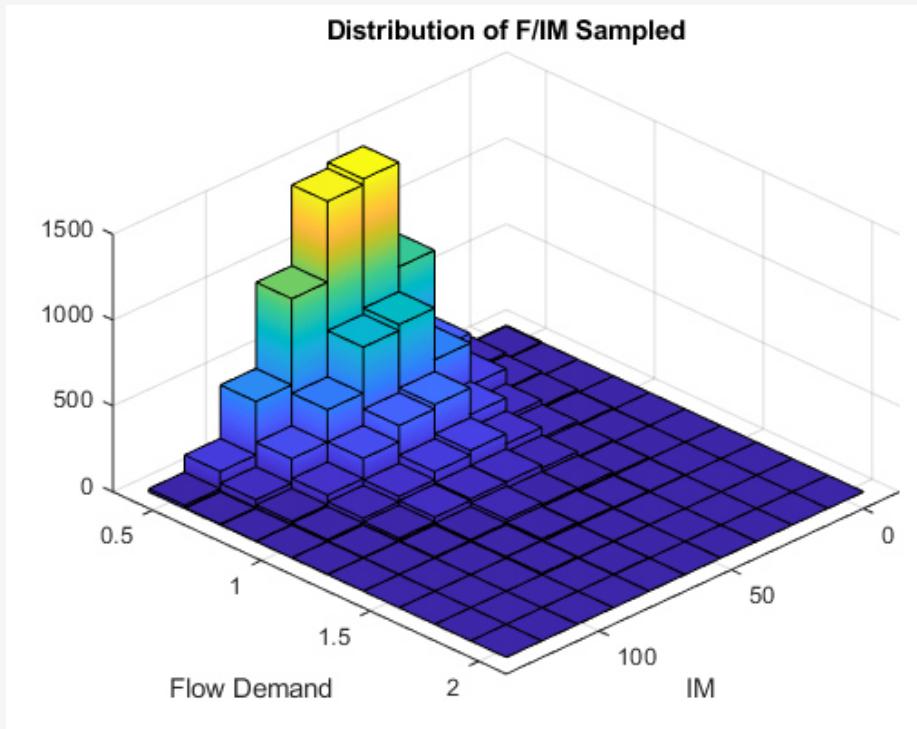
P_f convergence plot over 10,000 simulations



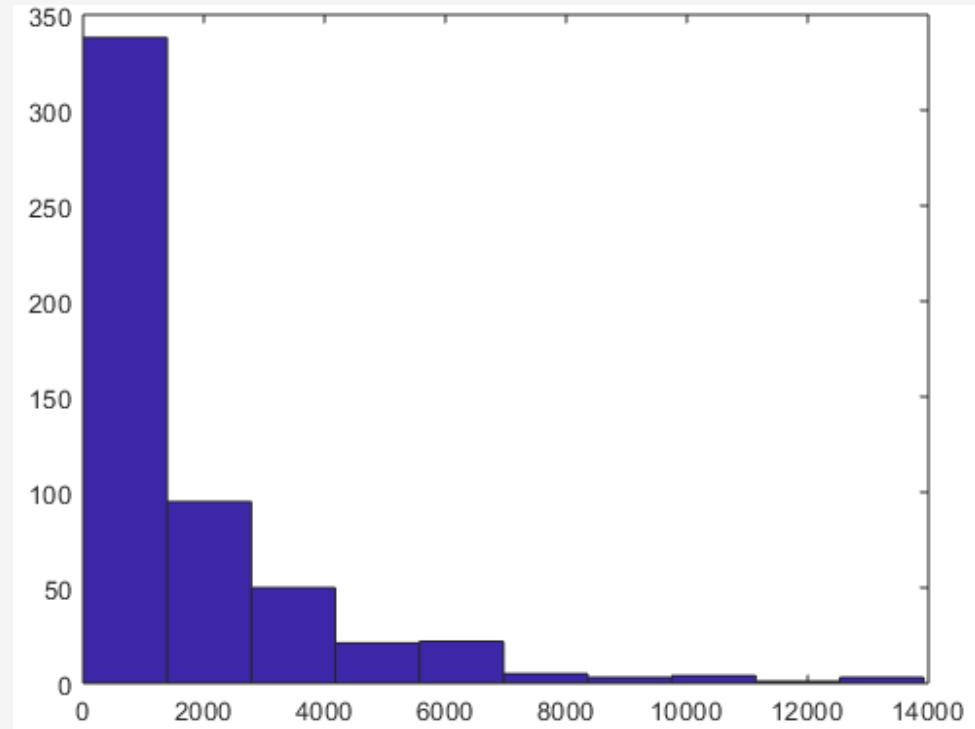
Each solution of the MCF problem yields a “hyperdistance” for the pair (s,t) . We call a “failure” each time the network could not fulfil a demanded O-D trip at required traffic level T .

Markov chain monte carlo parameter space exploration

Distribution of IM and flow intensities sampled



Distribution of extra operational costs

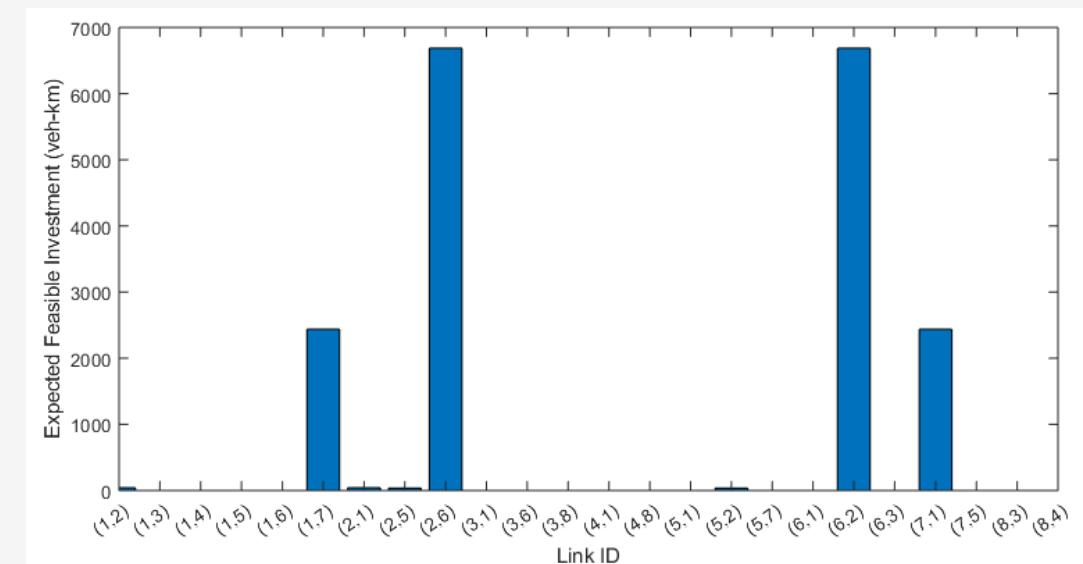
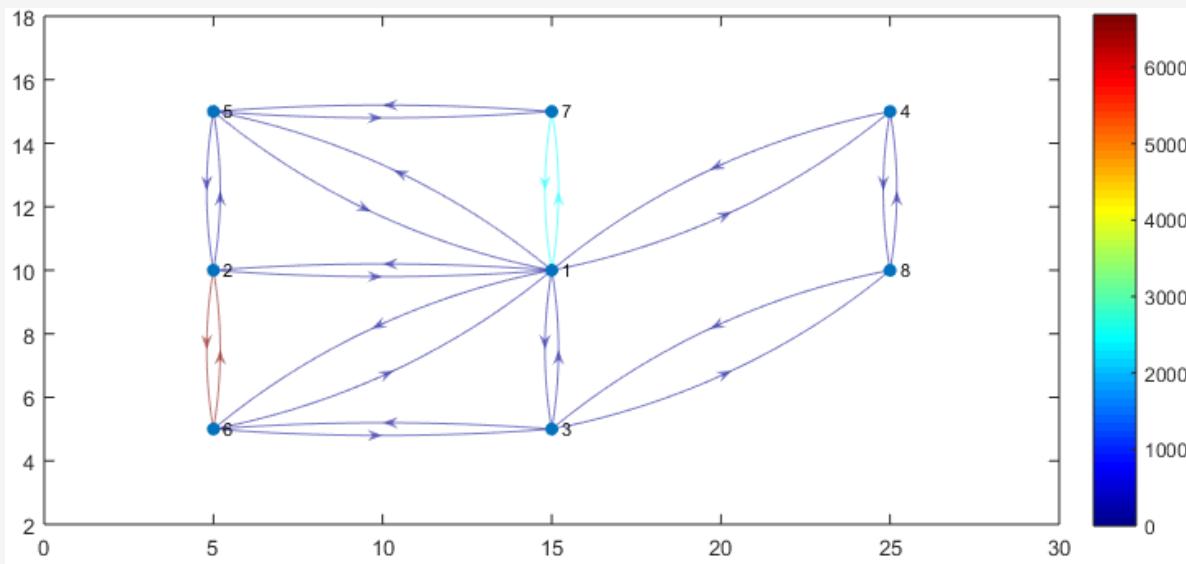


Efficient sampling of the parameter space -> probabilistic distribution of network performance “extra-costs”!

Component level reliability

As with previous approaches, one can count how many times components failed and obtain $P(E_i \mid \delta_{loss} \geq L)$

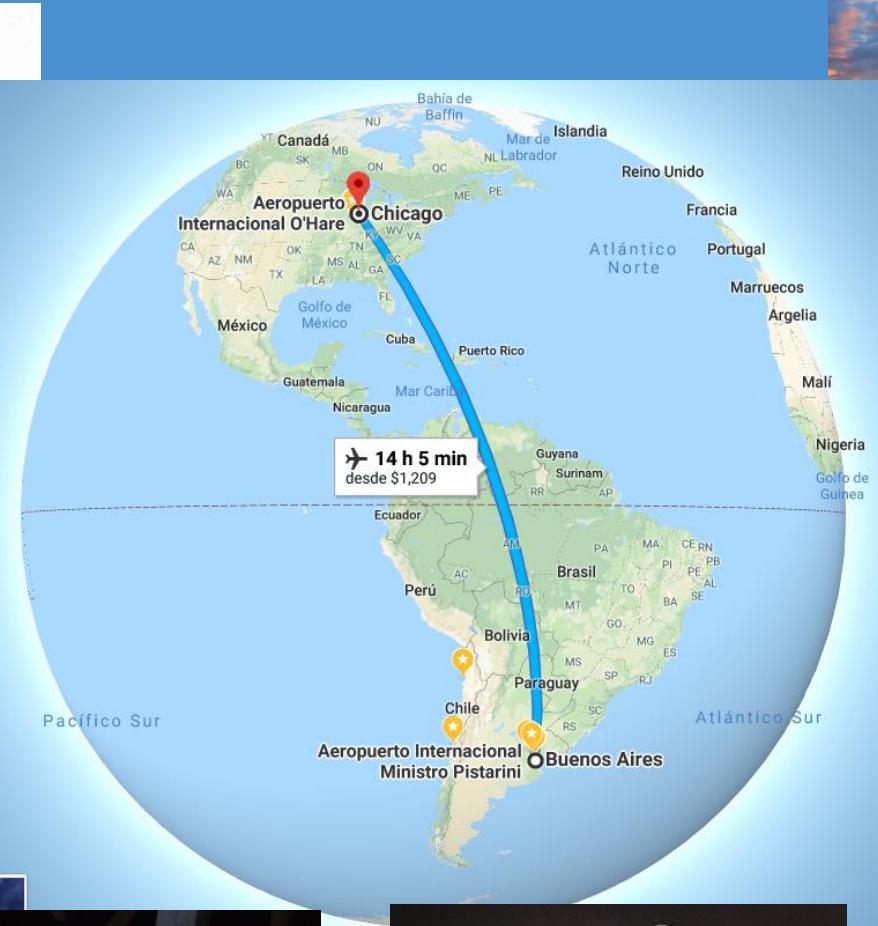
With the dual values now it can be obtained when the probability that capacity becomes an issue, and furthermore the expected losses that could be averted by increasing capacity!



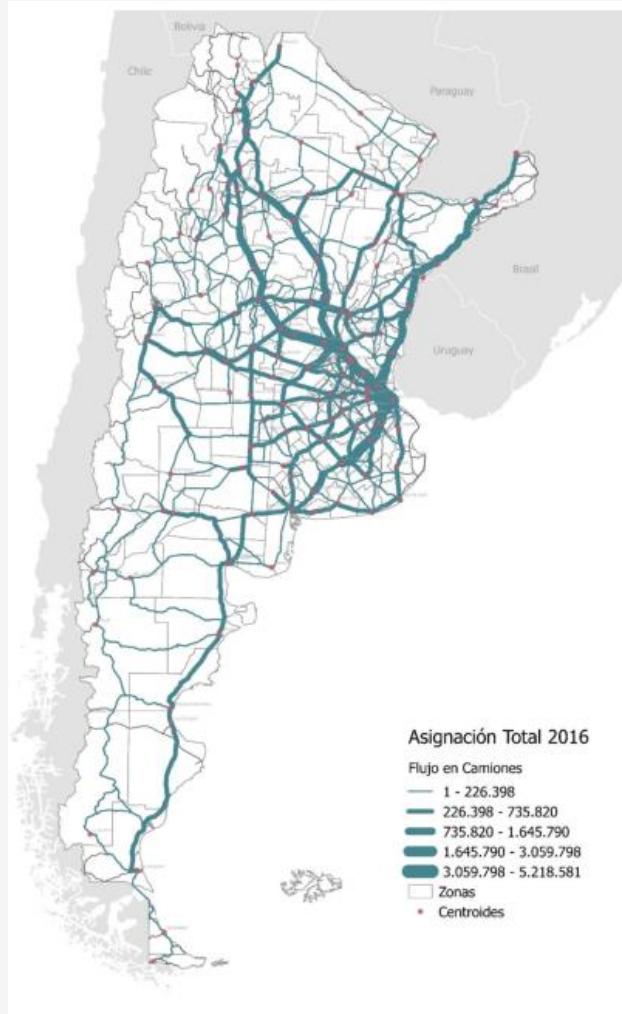
This formulation allows the quantification of losses due to capacity bottlenecks! *Prices capacity and reliability*

Outline

- Introduction
- Possible formulations
- Our formulation
- Case study : Argentina
- Final thoughts

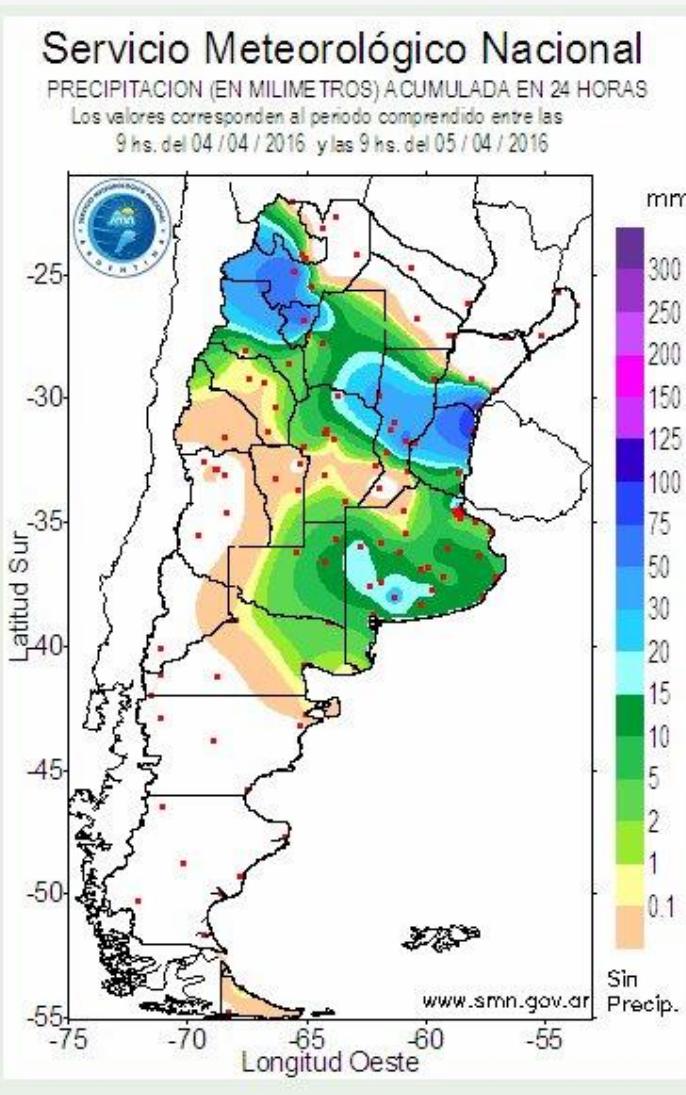


Argentina: freight road network dataset



- 3462 Edges (directed)
- 1426 Nodes
- Origin-Destination Matrix for annual traffic
 - Discriminated by activity type
- Capacities and velocities given as per Highway Capacity Manual (2016)

Hazards : floodmaps

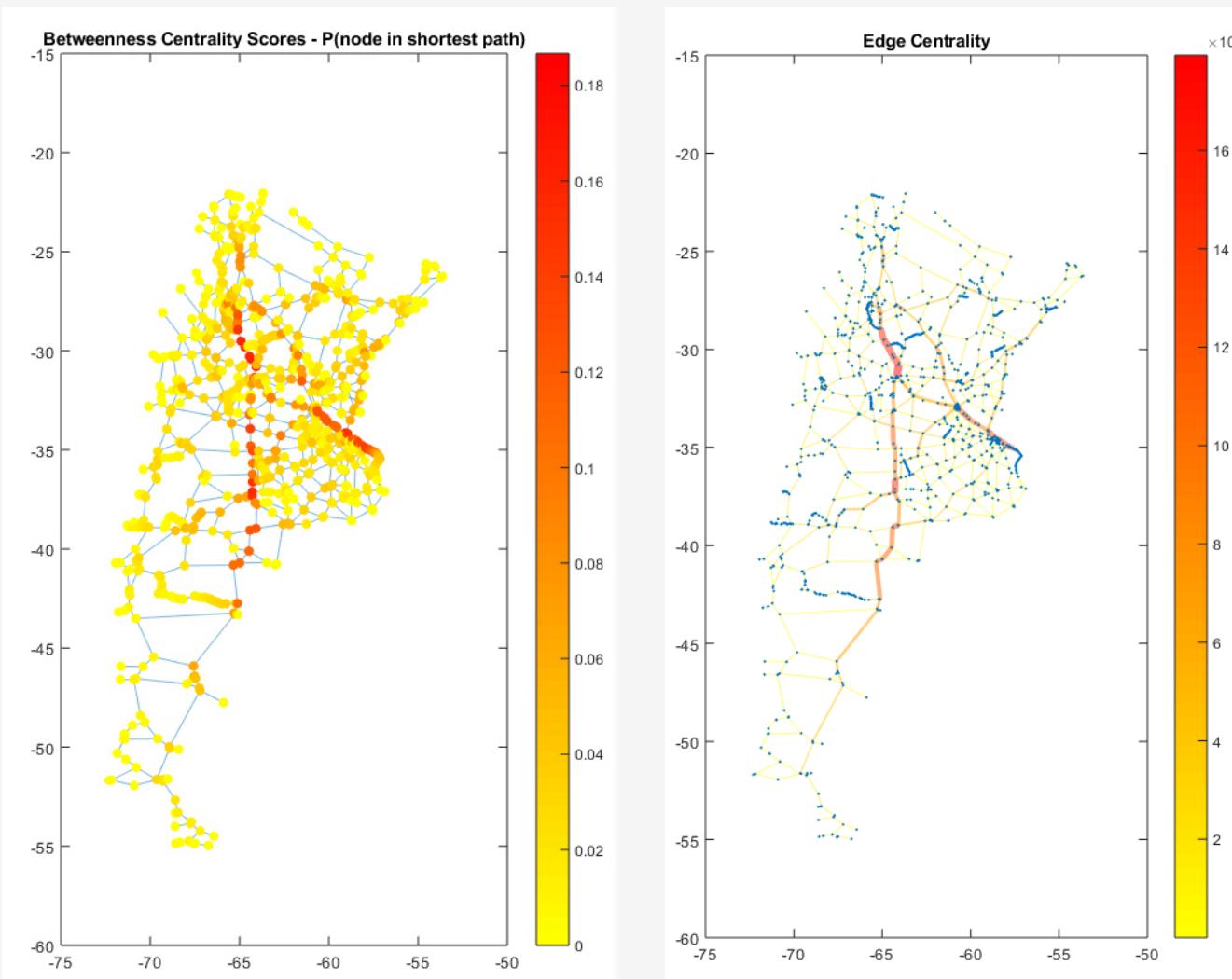


<http://floodlist.com/america/argentina-uruguay-floods-april-2016>

<https://blogs.worldbank.org/transport/transport-and-climate-change-putting-argentina-s-resilience-test>



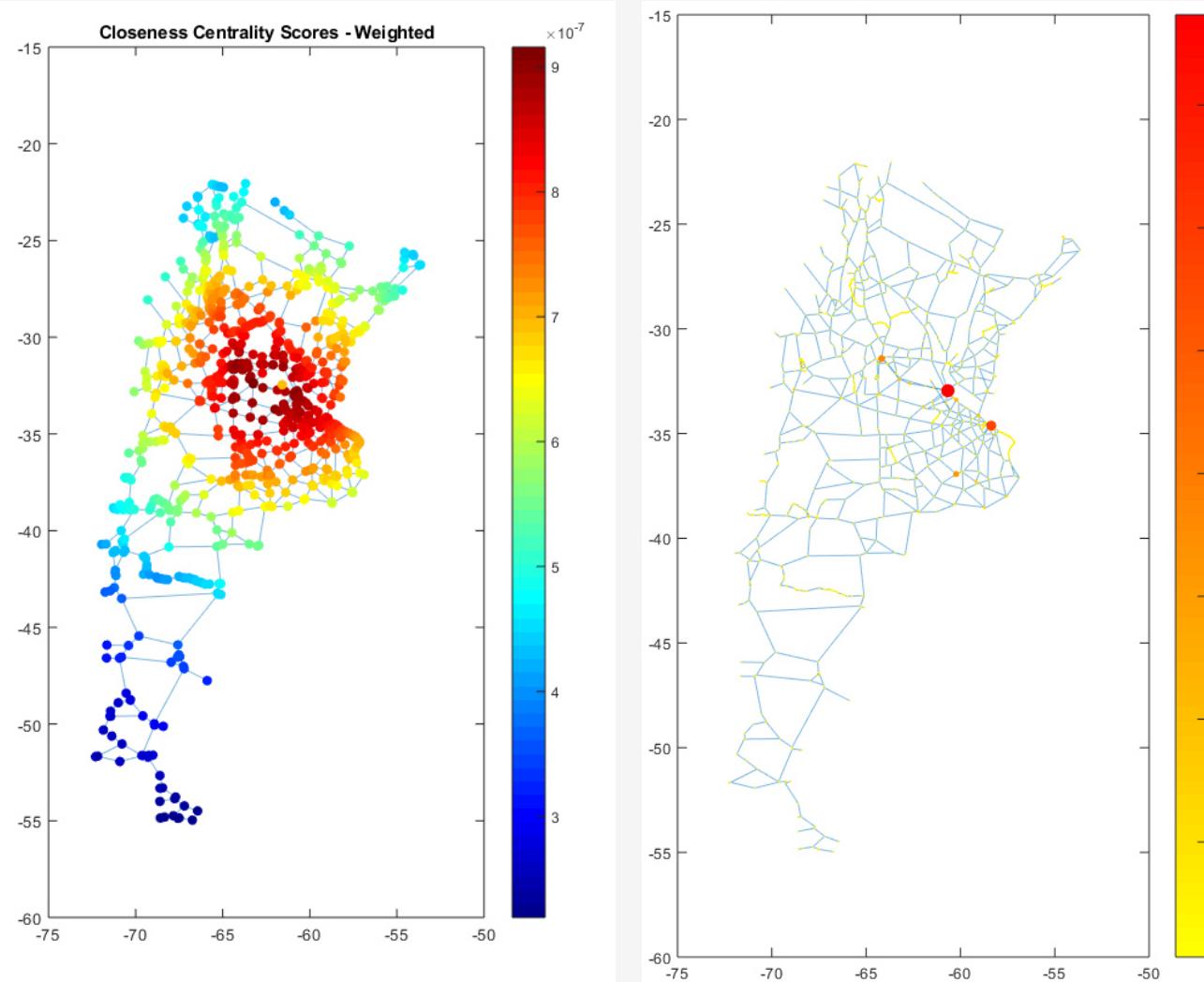
Topology indices



The topology indicators already reveal:

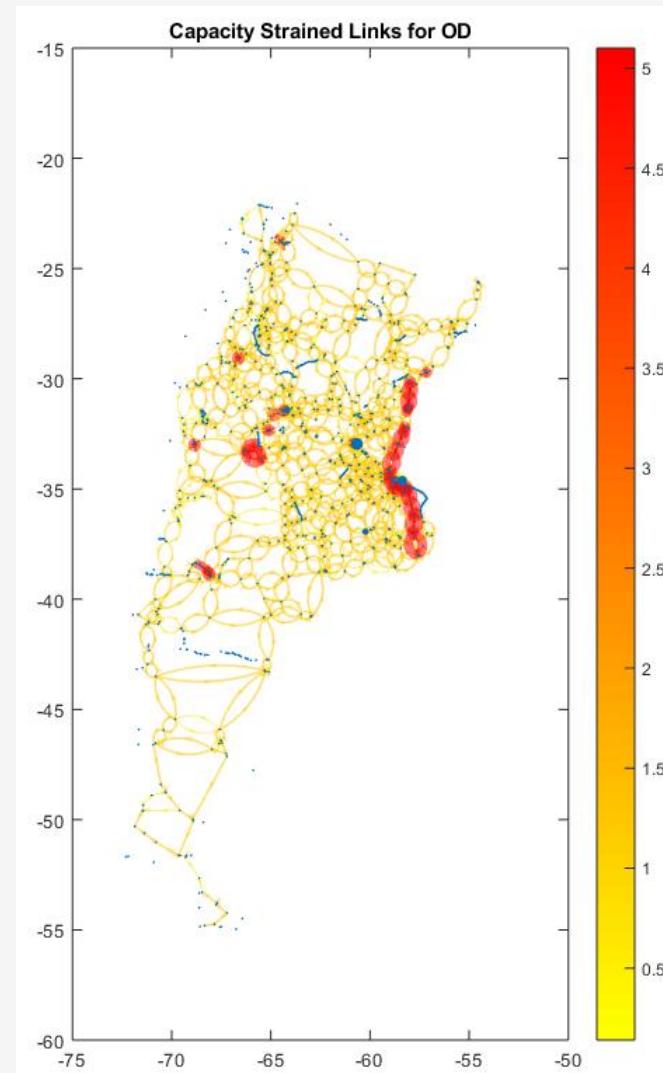
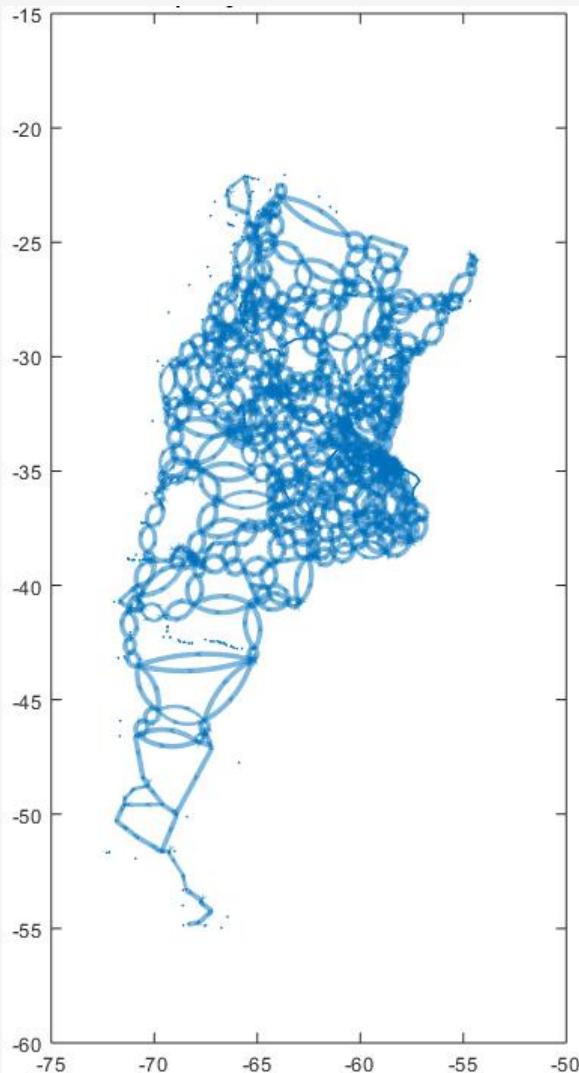
- Disparity between the connectivity levels of the central areas of the country (understandable and reinforced by economic activity)
- Main corridors already arise from strategic central locations and joining of high productivity areas.

Central nodes and Origin-Destinations



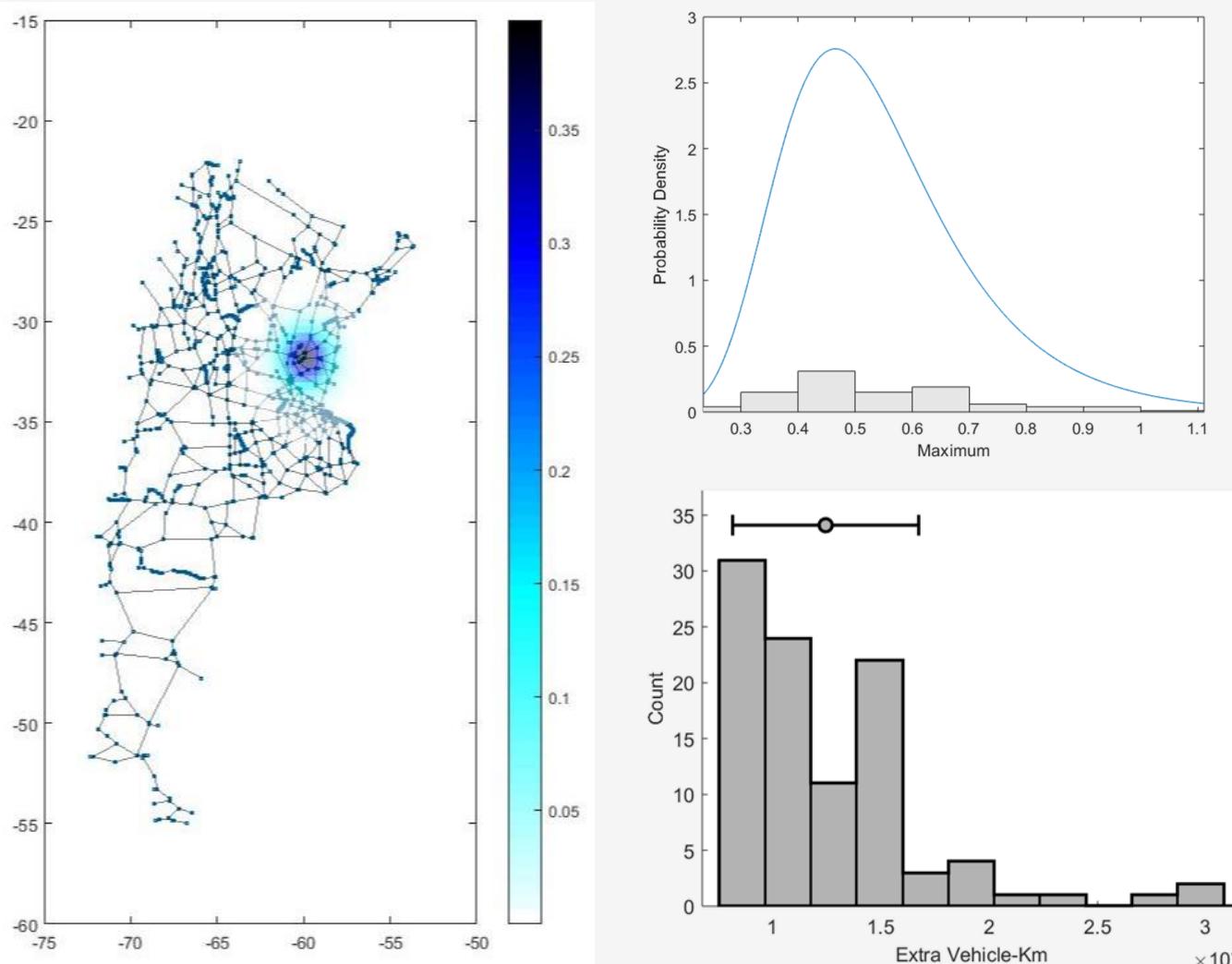
- It can be seen from the weighted centrality scores the central locations of the country
- However these do not necessarily reflect the actual centres of production and export; which also vary for different products.

a preliminary analysis



- Solve the multicommodity flow for one instance, without disruptions
- We observe the usage of the network for standard operational conditions is particularly strong in
- Origin destination for different products, obtain the flows expected (or possible) and the links that act as bottlenecks in common traffic event.
- Can see *which capacities are constraining* for the network as a whole (or for a specific industry/sector)

modelling the hazard



- Simulate instances of “low” frequency events of interest: littoral region (1982, 1983, 1992, 1998, 2003, 2015, 2016, 2019)
- MLE fit for extreme value events: GEV function with parameters μ, k, σ .
 - Include watershed model for future work
- Simulate and obtain *distribution* of expected extra costs of the system routing due to a flooding event
- Extra vehicle costs due to re routing and unmet demand (to be modelled)

Outline

- Introduction
- What has been done
- Our formulation
- Case study: Argentina
- Final thoughts

Summary

- Graph theory : Choosing the best representation for the problem at hand
- Network analysis: transitioning from graph theory to optimization
 - Connectivity may be enough for social capability indicators, but not for capacity planning at a multimodal level
- New methodology consistent with previous approaches and with *more relevant information for network level decision making*. "Simple but not simpler".
- Computational methods: efficient successive search and storage of problem solution structure are key to have a size-manageable problem

Future Work



Streamline code further, sort out Argentina.



Compare with NN structure (a different paper)



Unserved demand: super sink/source



Define failure thresholds (monetary, service, population)

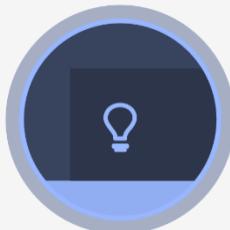


Probabilistic capacity curves for flooding (a different paper)



Finish paper and graduate!

Thanks for listening and happy Friday!



Questions? Comments? Advice? Tickets to Argentina?

