Complex networks and complex processes

Simon Dobson and Saray Shai
School of Computer Science, University of St Andrews UK

simon.dobson@st-andrews.ac.uk
http://www.simondobson.org

Overview

- Network models of real-world processes
  - Water flow, flooding
  - Epidemiology
  - Urban traffic patterns

Defence Forces Ireland, 2014

- This talk
  - Introduction to the core ideas of network science
  - Examples of what we can find out
  - Future directions

Structure

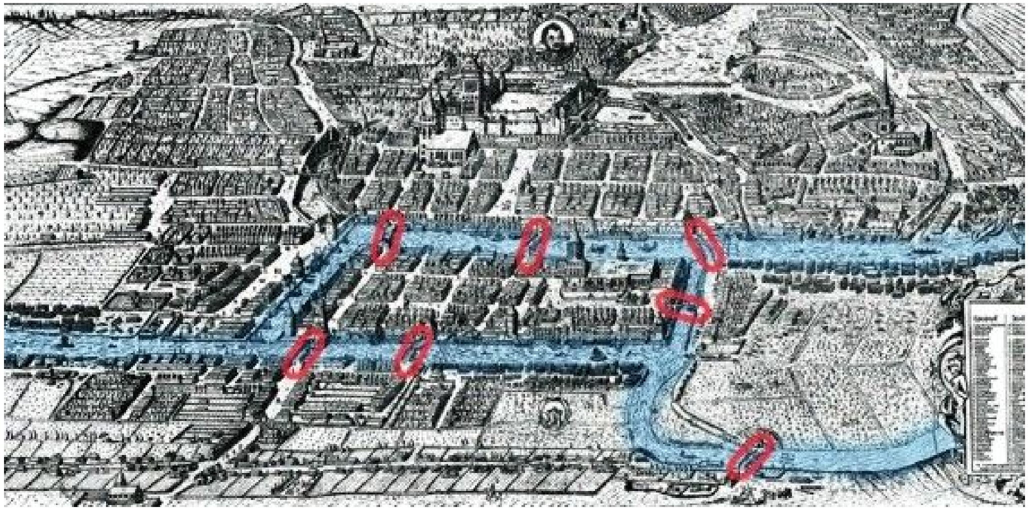
1. Basic ideas
   - Graphs, networks, complex networks
   - Processes
2. Coupled adaptive complex networks
   - Evolution over time
3. Urban networks
   - Working at scale
4. The future

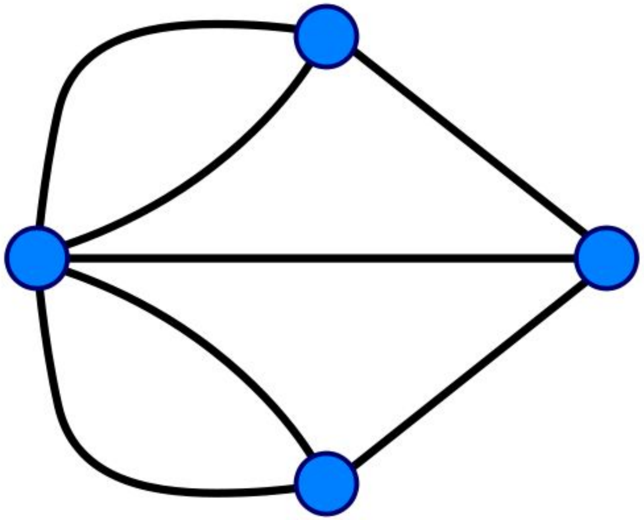
Part 1

A brief yet helpful guide to network science

Graphs: the Königsberg Bridges

- Can one traverse each bridge exactly once?
  - Abstract as a *network* or *graph*
  - Pose the question as a *structural property* of the network





Königsberg is now Kalinigrad, a Russian enclave surrounded by Lithuania and Poland

Random graphs – construction

- Canonical example
  - Start with a graph $G_0 = (N, \varnothing)$ with N nodes and no edges
  - Add edges between nodes chosen at random, with no parallel edges

Erdős and Renyi. *On the evolution of random graphs*. Publications of the Mathematical Institute of the Hungarian Academy of Sciences. 1960.

- Often called the *ER model*
- Questions
  - Probability that G is connected?
  - Average degree $<k>$ of a node?

Random graphs – percolation

- Model how a graph stays connected as it evolves, for example by removing edges
  - Traffic in a city as roads are closed, water flow through soil under silting, ...
- Construction
  - Start with a random graph G of average degree $\langle k \rangle$
  - Remove a fraction $1 - p$ of the edges
- There is a *critical threshold* $p_c = 1/\langle k \rangle$ below which the network becomes fragmented
  - An analytic result: no need to simulate

Random graphs – a précis

- A good model for a surprising number of phenomena
  - Mathematically very tractable
  - $P(deg(v) = k)$ is Poisson as $|N| \to \infty$
  - Normality evens-out variations

- But do *all* networks have this structure?
  - Google's PageRank works because the distribution of links is non-normal, and biases towards links from higher-degree nodes

Power-law networks

- Degree distribution
  - $P(deg(v) = k) = k^{-\lambda}$
  - Large number of low-degree nodes
  - Small number of *hubs* with very high degree
- Small worlds
  - Distances are "unexpectedly small"
  - Social networks, internet page and link distributions, …
  - Offer structure that can be exploited

One tends to find $2 < \lambda < 3$

Bacon numbers and Erdős numbers are surprisingly small relative to the size of the communities they describe

Simon's Erdős number is at most 6. Strangely, his Bacon number may actually be less...

Navigating in a small world – 1

Six degrees of freedom?

- How many handshake degrees of separation?

Navigating in a small world – 2

Six degrees of freedom?

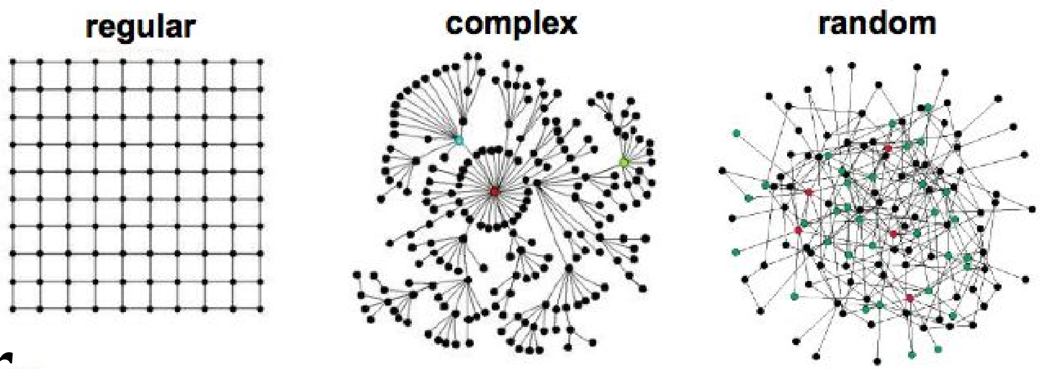
- How many handshake degrees of separation?
  - Simon
    → Prof Muffy Calder OBE
    → The Queen
    → Sir David Attenborough
    → Member of the Biami tribe
  - A hub is on a short path between any pair or nodes, with high probability

Power-law networks are *ultra-small*, with $L \propto \log \log n$

Not only is there a short path, it's relatively easy to find using social features

Complex networks

- Characterised by overview statistics
  - Degree distribution
  - Modularity
- Turn out to be more interesting than either end of the spectrum
  - Processes operate differently
  - Statistically-driven classes of behaviour
  - Hard to understand analytically



Barabasi *et alia*. Science **286**. 1999.

Neither defined by the fine structure nor by assumptions of normal distributions

Construction

- Various construction processes
  - Start with $G = (N, \varnothing)$ where $n = |N| = 1$
  - At each step, add a node n' to N and an edge e' to E connecting n' to some other node in N

This process is due to Barabasi and Albert and gives rise to *BA networks*

- Early nodes have more chances to gain links
  - $n = 1, P(e' = (n', n_1)) = 1$
    $n = 2, P(e' = (n', n_1)) = 0.5$
- A process of *preferential attachment*
  - Base connection on node degree

Generates a power-law network with $\lambda = 3$

Analysis – static

- Basic graph-theoretic properties still of interest
  - Degree distribution, diameter, …
- Often look at *betweenness* metrics
  - Betweenness centrality, the number of shortest paths that intersect a node
  - A measure of how significant a node is to information flowing through the network
  - Disrupting such nodes will disrupt information flowing efficiently

Analysis – dynamic – 1

- Run a *spreading process* over the network
- Susceptible / Infected / Susceptible (SIS)
    - Node start off susceptible, and we seed the network with some proportion ε of infected nodes
    - At each timestep, each infected node infects any susceptible neighbour with probability β
    - ...or recovers back to susceptible with probability α
    - Run until equilibrium
- How does the "illness" spread for different values of α, β, ε? For different topologies?

Analysis – dynamic – 2

- Can sometimes be treated analytically
  - Works well for random networks
  - ...but falls over quickly for complex networks
  - Heterogeneous degree distribution, correlations, ...
- Simulate for varying values of the parameters

$$s = \frac{S}{n} \qquad i = \frac{I}{n}$$

$$\frac{ds}{dt} = \alpha i - \beta s i$$

$$\frac{di}{dt} = \beta s i - \alpha i$$

- Where mathematics meets the halting problem
  - Rules (program) the simplest description

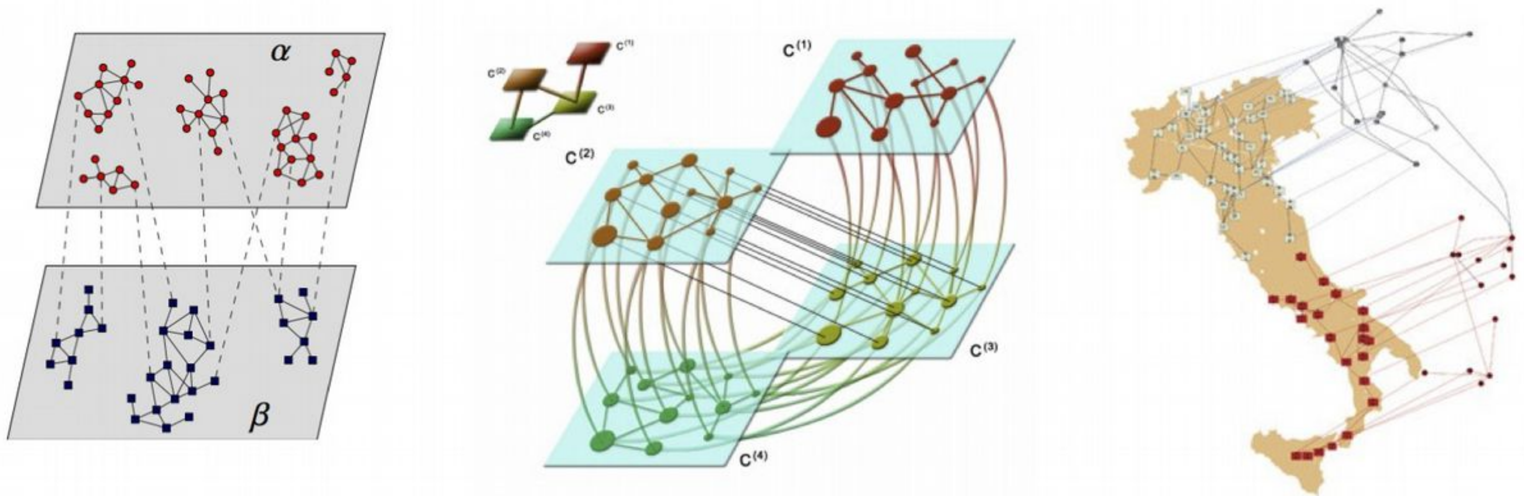
Part 2

Coupled adaptive complex networks

Coupled complex networks

- Phenomena with several different layers
  - "Separate concerns" for abstraction

Also known as *multiplex* networks



Leicht *et alia*; de Domenico *et alia*; Buldyrev *et alia*

  - Processes *within* layers can cross over *between* layers
  - Can't understand a process without understanding the multiplex

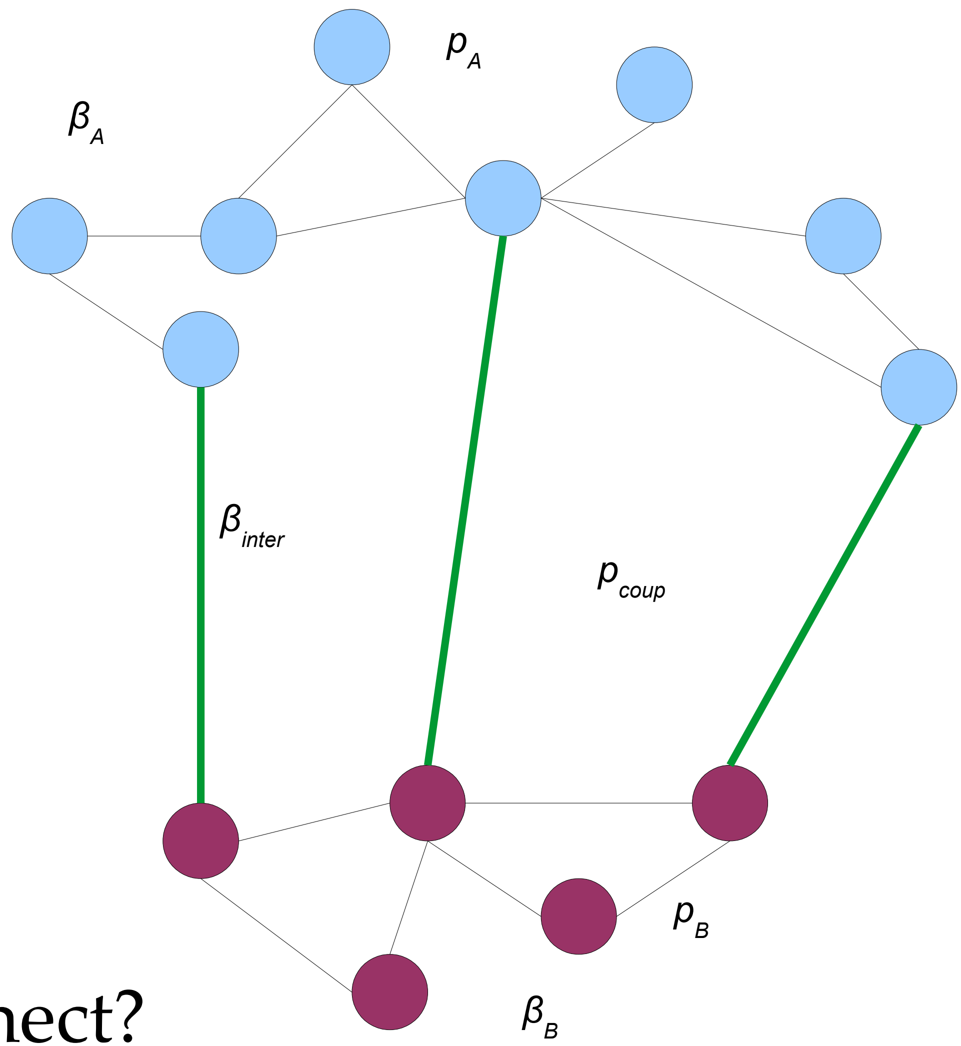
Properties

- Each layer can (and often will) have a different topology
  - ER in layer 1, BA in layer 2, etc
- There may be structure in the coupling
  - High-degree modes in layer 1 couple preferentially to high-degree nodes in layer 2, etc
- The parameters of the spreading process may be different

- So the networks may be fundamentally *different*

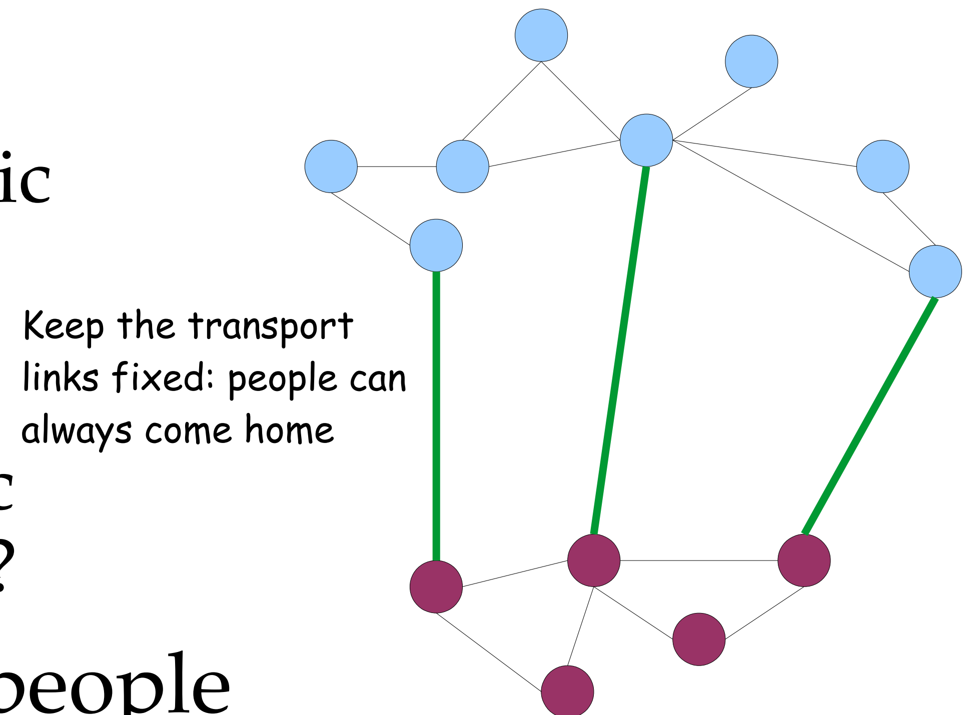
Stereotype construction

- Create the layers
  - *E.g.*, ER network with $p = 0.2$
- Couple nodes
  - Random
  - Biased by degree
  - Inversely biased
- Parameters
  - Different for each layer, and the interconnect?



Adaptive networks and epidemics

- Two countries A and B
  - Seed A with an epidemic
  - A and B connected by transport links
  - How does the epidemic spread in the countries?



- People avoid infected people
  - For any I_A to S_A edge, re-wire to another S_A node with probability γ (similarly for I_B etc)

Shai and Dobson. Effect of resource constraints on intersimilar coupled networks. Physical Review E **86**. 2012.

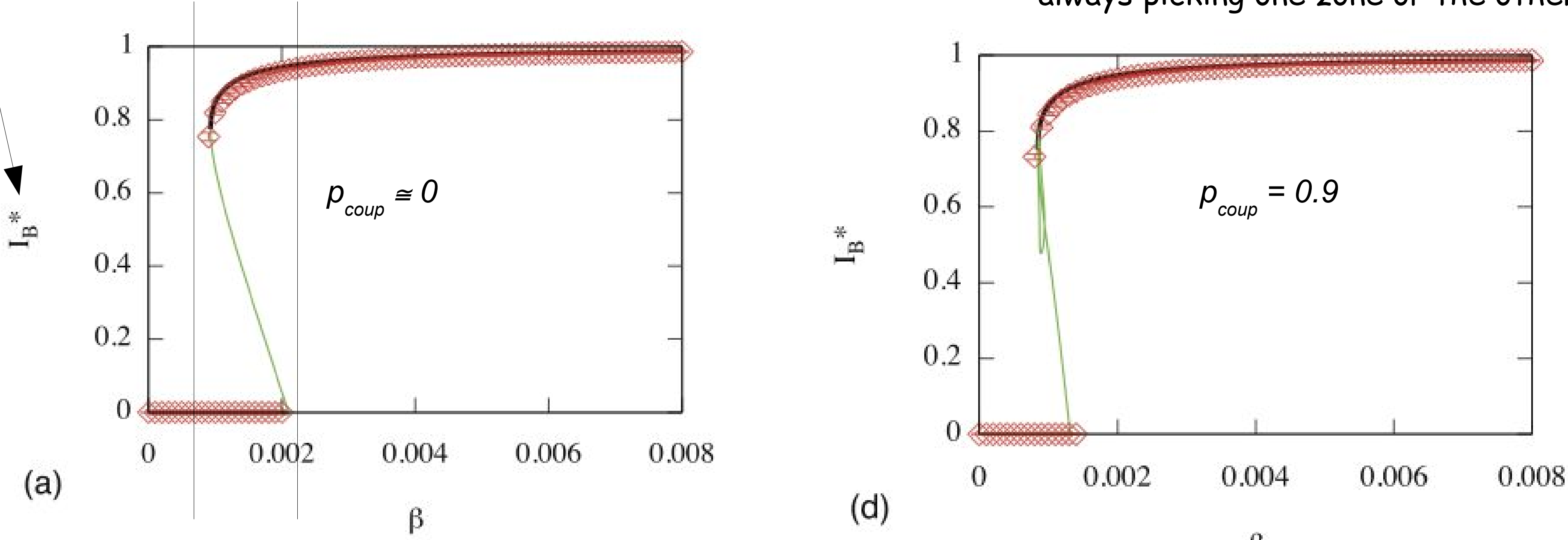
So what?

- Does this structure behave differently to a single network?

Proportion of infected nodes in *B* at equilibrium

Unstable zone where the outcome depends on exact initial conditions (a bifurcation)

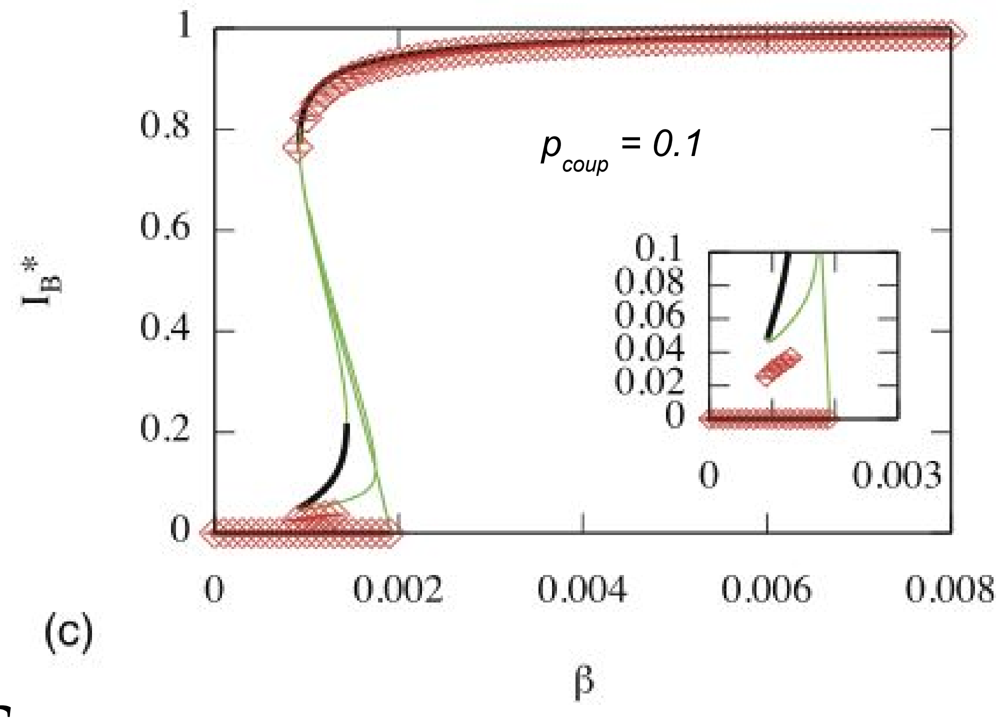
As coupling increases, network becomes more stable in the sense of always picking one zone or the other



Changing rates of infection, common to both networks and the multiplex

Extra option

- There is a critical region in which there is a *third* stable branch
  - Nodes coupled to *A* continue to get infected, but the epidemic doesn't break out
  - Whether cutting off travel is enough to prevent an epidemic crossing between countries depends critically on the infection rate and frequency of travel



Shai and Dobson. Coupled adaptive complex networks. Physical Review E **87**. 2013.

Part 3

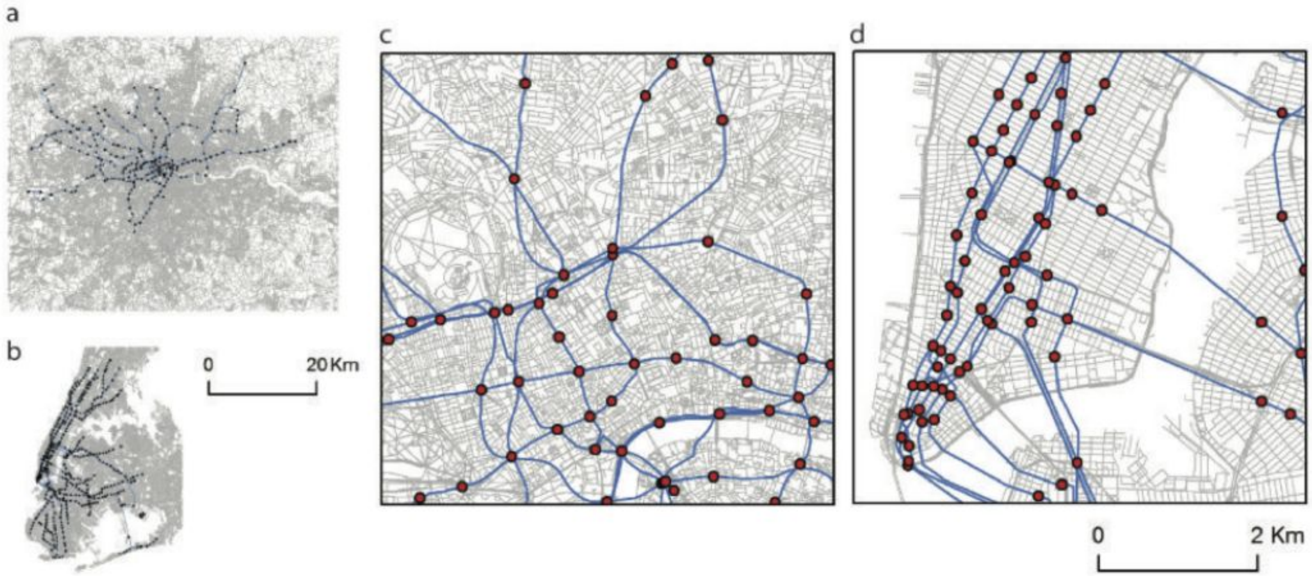
Urban networks

Transport in urban networks

- Large cities only work well if they have good transportation networks
  - How good is "good enough"?
  - How do different transport modalities affect commuting times?
  - How vulnerable are different parts of the transport system?

First get your data...

- Not always easy...



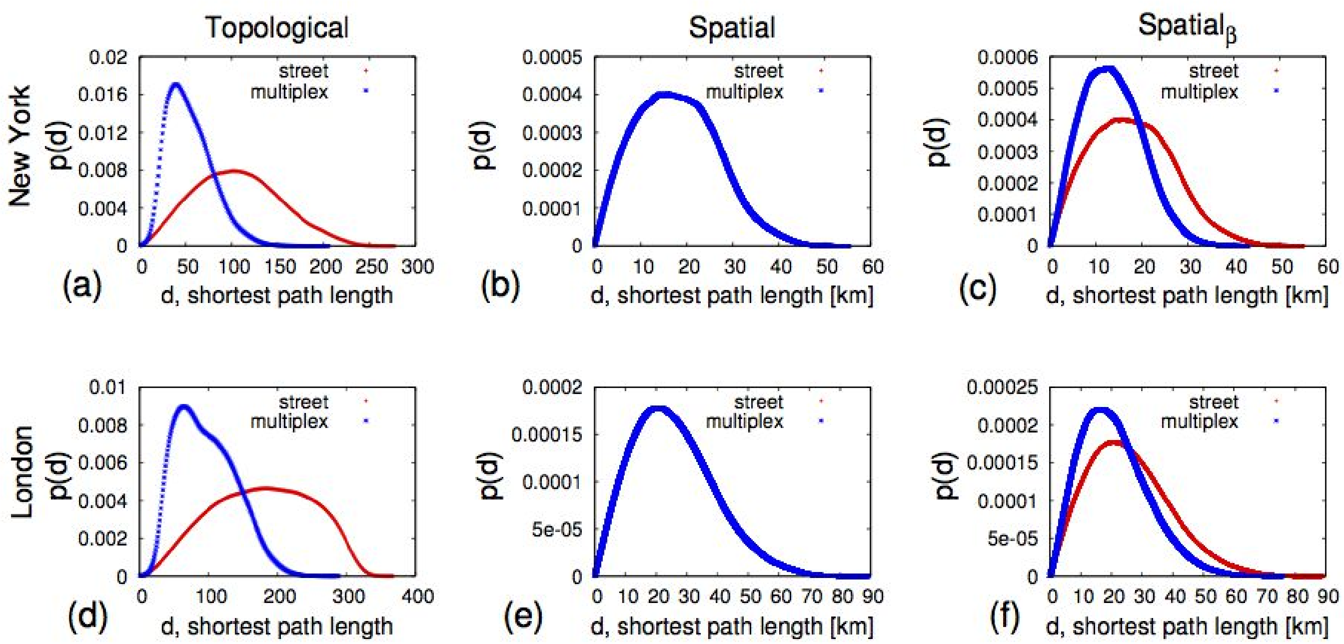
- Government open data is patchy to say the least

Case	N	E	cost (km)	$\bar{l}_{ij}^{geo}$ (km)	$\bar{l}_{ij}^{topo}$	$Diam^{geo}$ (km)	$Diam^{topo}$
London streets	324536	427920	34493.73	25.83	178.16	89.31	368
London subway	263	296	385.98	18.55	14.26	60.3	42
London multi	324799	428479	34886.52	25.78	96.16	89.27	288
New York street	68417	112827	12153.81	17. 94	106.64	55.18	278
New York subway	454	489	416.12	18.87	19.10	57.28	62
New York multi	68871	113770	12579.44	17.91	54.45	55.18	205

Much of this data comes from Open Street Map

Topological properties

- How much the existence of the tube affect journey distances?
  - How much “smaller” does a city get?



Explore different factors for tube journeys being faster than street journeys (*β=0.5*)

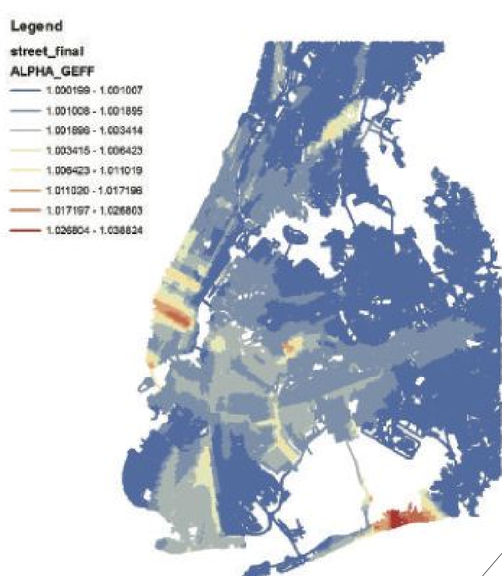
Strano, Shai *et alia*. Multiplex cities: the interplay between coupled transportation networks. In preparation.

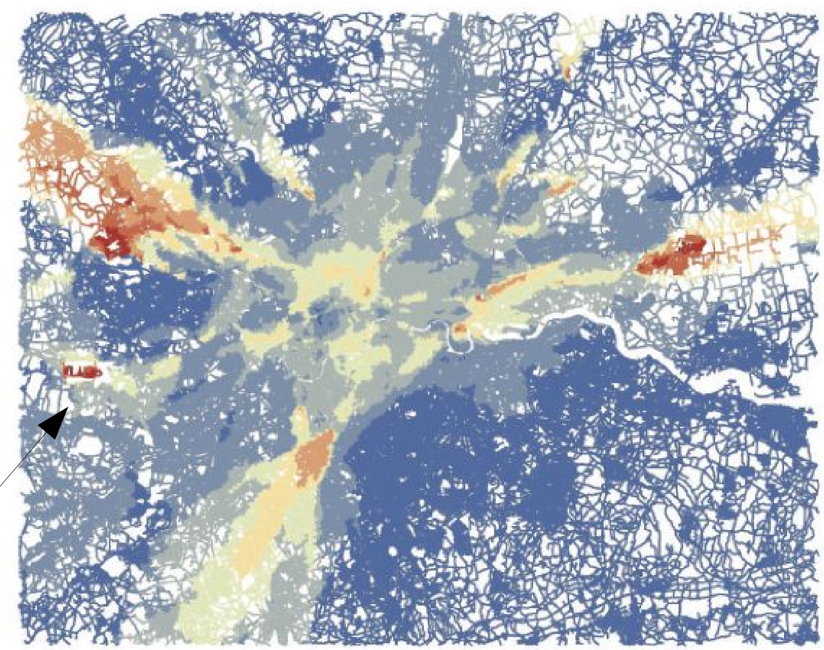
Efficiency

- How does the tube shrink the city?
  - Distance from node v to all other nodes
  - Street *vs* multiplex

$$\alpha_{geff}(v) = \frac{E_{street}(v)}{E_{multi}(v)}$$

Some places become significantly closer when you allow tube rides



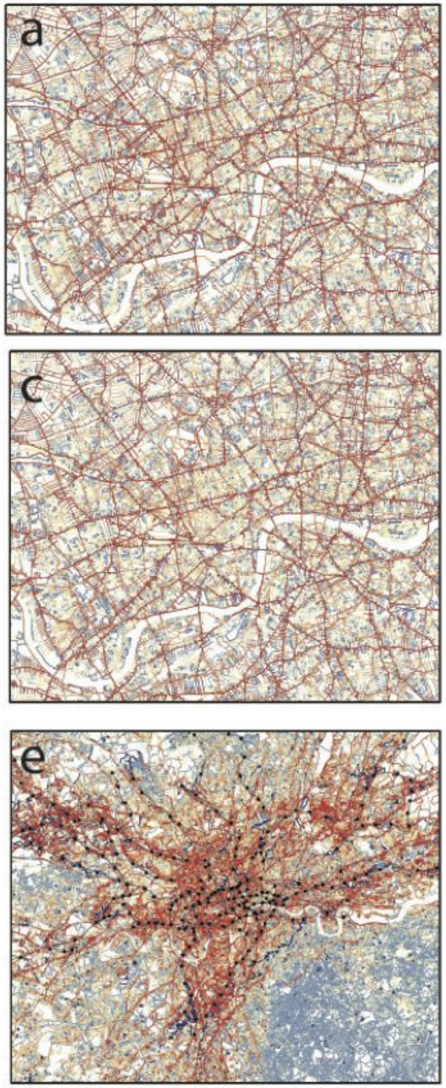


Heathrow

Vulnerabilities

- Which parts of a network are most central to most journeys?

- How does betweenness centrality change when we add the tube network to journeys?



Streets only

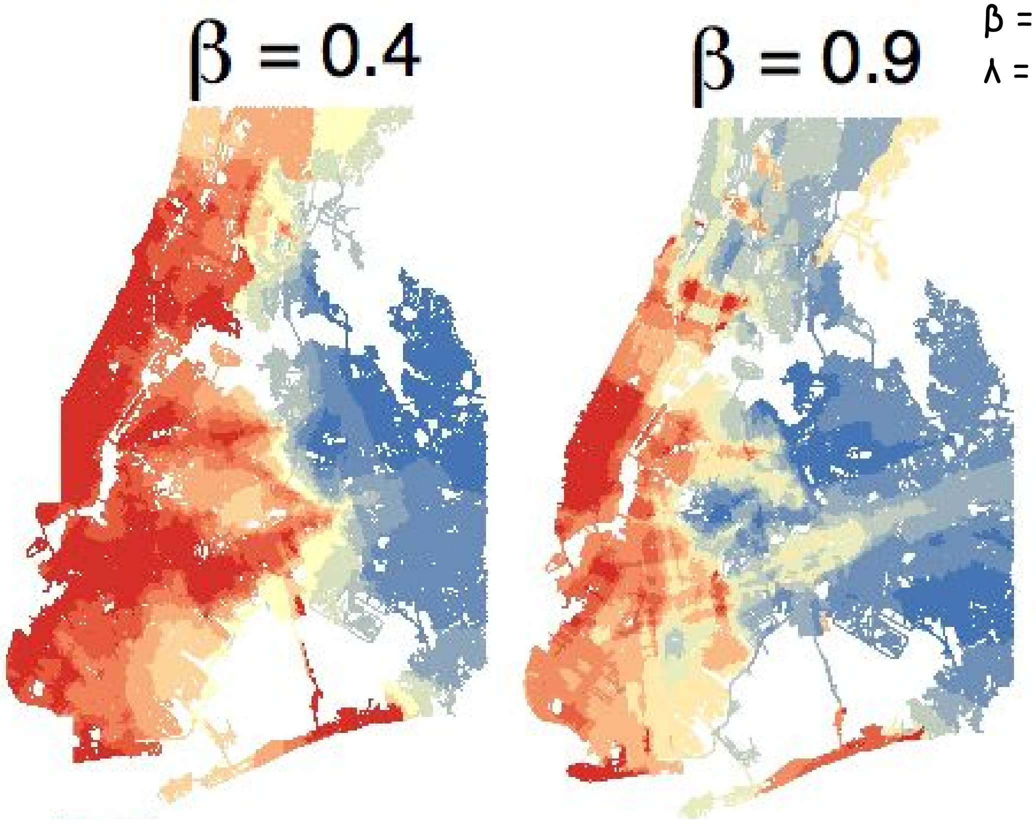
Streets and tube

Impact of tube on shortest paths

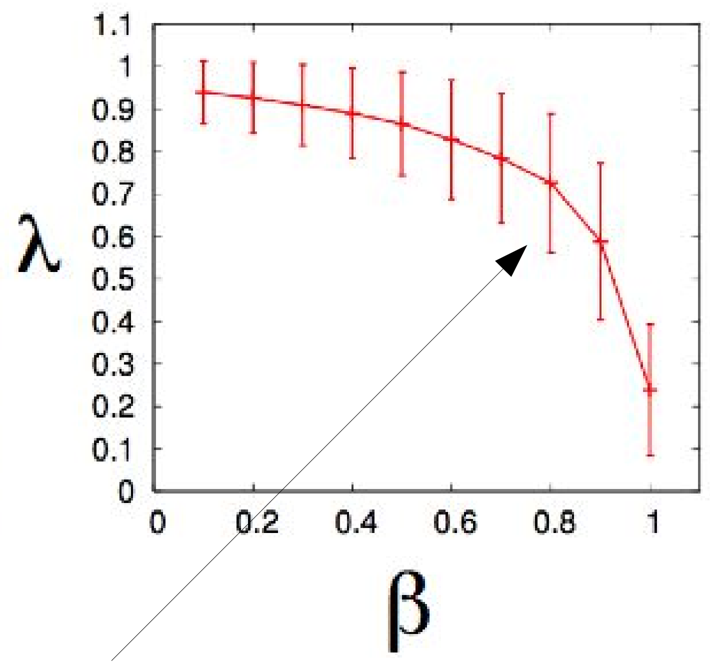
(Spatial model)

Best investments

- What happens if we install a high-speed tube system?



β = Speed factor, tube links 1/β faster than street
λ = Fraction of shortest paths using tube



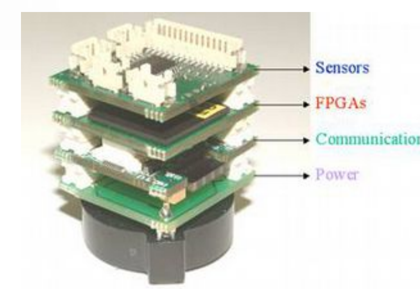
Lots of benefit from
a 25% speed-up

Part 4

Opportunities

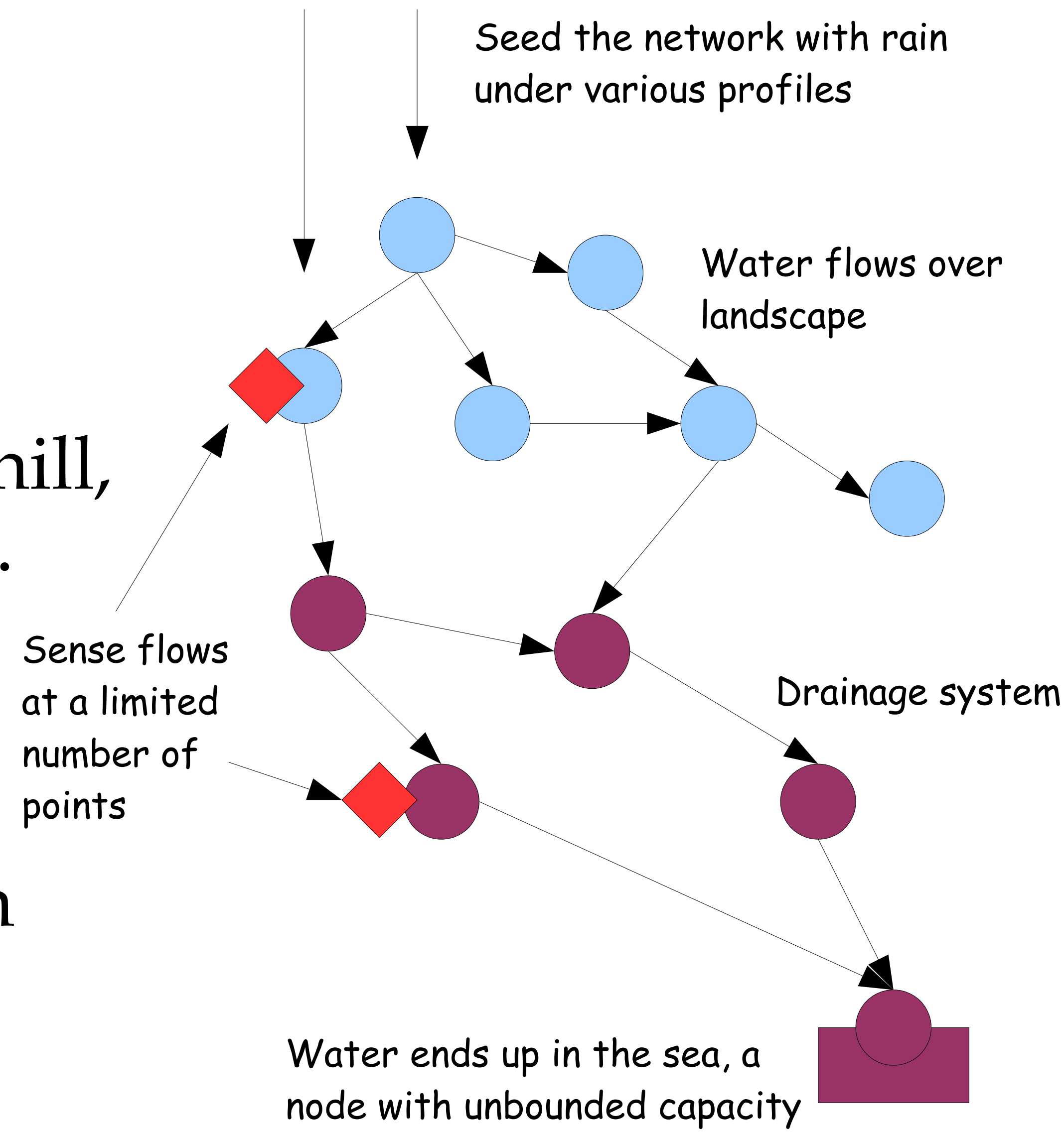
Leveraging processes

- We can now get real network data
  - Synthesise networks for exploration
  - Apply against real networks
  - Verification, what-if analysis
  - Controllability
- Integrate with sensing
  - Grab real behaviour
  - Inject into network to drive or steer processes



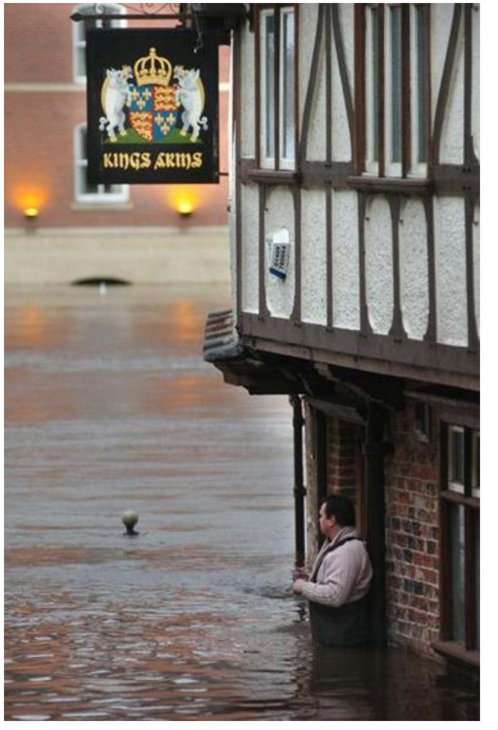
Multiplex model

- Elements
  - Landscape, land use, drainage, ...
- Local rules
  - Water flowing downhill, channel capacities, …
- Process and sensing
  - Confirm predictions
  - Condition simulation



Where we want to go

- How do rivers and towns *really* shed water?
  - How do you control the process?
  - What effects to remedial measures actually have?
  - Does dredging large channels makes this better? – or worse?
  - How does the variance in rainfall affect the network's behaviour?



- Non-trivial dynamics of processes make these questions interesting

Three things to take away

- Network science is a hot topic best not left (just) to physicists
  - Computing is vital
- General results that can be applied to range of phenomena of sufficient complexity
- May combine well with distributed systems and other core areas of interest