Complex-Network Modelling and Inference

Lecture 14: Random Graphs: HOT and COLD

Matthew Roughan
<matthew.roughan@adelaide.edu.au>
https://roughan.info/notes/Network_Modelling/

School of Mathematical Sciences,
University of Adelaide

January 14, 2025

Section 1

HOT graphs

Problem with "random" graph

- Random graphs are formed by taking lots of small, random operations, and building a larger graph
  - most dependencies are quite local
  - bottom-up construction
  - hope for "emergent" behaviour
  - lots of small, local behaviour produces global structure
- This is NOT how real, technological, physical networks arise
  - the are designed
  - often from the top down
  - often optimised against a set of constraints
- Even if a random network has all the metrics right, does that mean it is actually a good model?
  - for instance, most random graph models don't include any redundancy component
  - consequently, they can be vulnerable to failures
  - real networks have designed redundancy

Example: Internet

- It was noticed early on that the "Internet" has a power-law degree [FFF99]
  - this was one of the motivators for preferential attachment
  - ignore some of the holes in the paper for the moment
- The preferential attachment model has "central" high-degree node
  - if one of these fails, a preferential-attachment network might become partitioned
  - is this a worry for the REAL Internet

HOT

$$\begin{aligned} HOT &= \text{Highly Optimised Tolerances} \\ &= \text{Highly Organised Tradeoffs} \\ &= \text{Highly Optimised Topology} \\ &= \text{Heuristically Optimised Topology} \end{aligned}$$

- Its a generic theory related to emergence or power laws
- The idea is that power-laws emerge from system that have been highly optimised
- This is a pretty superficial take on it – there is a lot more, but we only need to see how it applies to networks

HOT graphs

- Assume that networks are design through an optimisation process
- For the Internet
  - nodes are *routers* and edges are links between them
    - objective is to minimise the cost of these
  - routers have a maximum number of *ports*
    - effectively a maximum node degree
    - *backbone* routers, have a few, high-speed ports, but are very fast
    - *edge* routers have many low-speed ports
  - capacity constraint
    - must have enough capacity to carry given traffic

Is HOT a random graph model?

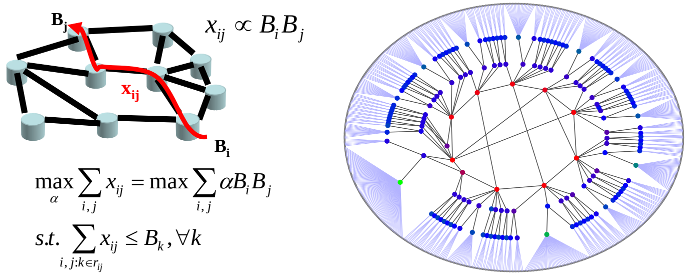
- HOT is different from other random graph models
  - the operations aren't random (mostly)
  - the optimisation is nearly deterministic
- The randomness comes from the environment
  - in this case the (random) traffic that must be carried
- Random *traffic matrix*

$$T_{ij} = \alpha B_i B_j,$$

where the B_i are non-negative random variables.

HOT mark I [?]

Its hard Fig5 to model costs so instead flip the problem around, so instead



(a) (b)

Node degree distribution [LAWD04]

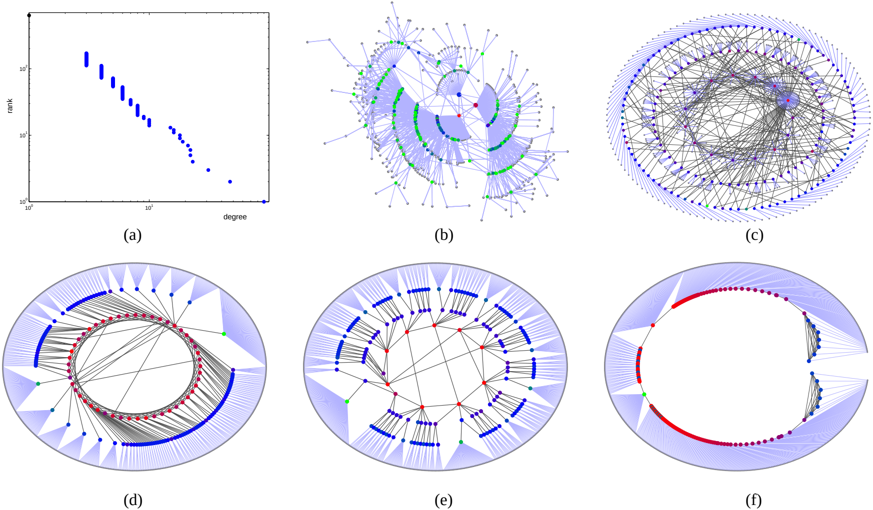


Figure 6: Five networks having the same node degree distribution. (a) Common node degree distribution (degree versus rank on log-log scale); (b) Network resulting from preferential attachment; (c) Network resulting from the GRG method; (d) Heuristically optimal topology; (e) Abilene-inspired topology; (f) Sub-optimally designed topology.

Abilene ecosystem of networks [LAWD04]

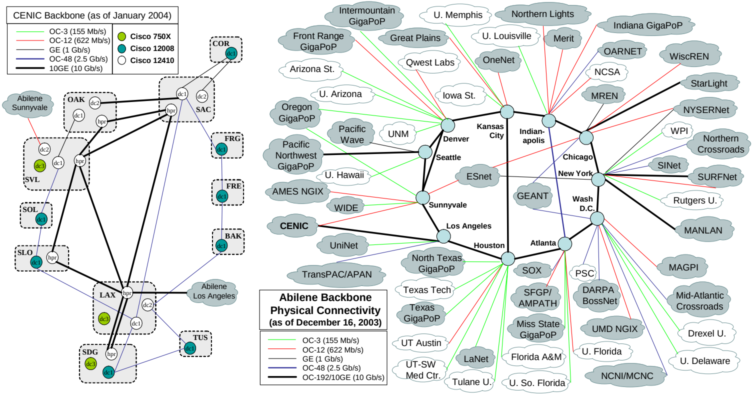
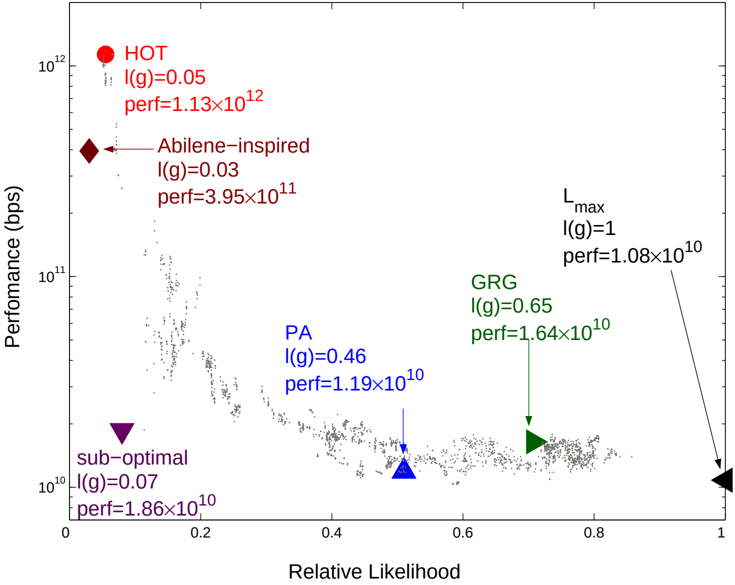


Figure 4: CENIC and Abilene networks. (Left): CENIC backbone. The CENIC backbone is comprised of two backbone networks in parallel—a high performance (HPR) network supporting the University of California system and other universities, and the digital California (DC) network supporting K-12 educational initiatives and local governments. Connectivity within each POP is provided by Layer-2 technologies, and connectivity to the network edge is not shown. (Right): Abilene network. Each node represents a router, and each link represents a physical connection between Abilene and another network. End user networks are represented in white, while peer networks (other backbones and exchange points) are represented in gray. Each router has only a few high bandwidth connections, however each physical connection can support many virtual connections that give the appearance of greater connectivity to higher levels of the Internet protocol stack. ESnet and GEANT are other backbone networks.

Performance



For every complex problem there is an answer that is clear, simple, and wrong.

M.L. Mencken

- Multiple models produce the same node-degree distribution
- They are VERY different

Section 2

COLD graphs

COLD [BRB14]

COLD = Combined Optimized Layered Design

- HOT pointed the way
  - but the optimisation model is not quite right
  - we really want to optimise actual costs
  - but costs in networks are complex
- Also, constraints in real networks are complex
  - traffic must be carried
  - ports limits must be respected
  - but real designs also include redundancy
- The resulting optimisation is too complex (for me)
- Tackle the problems in layers
  - the top-layer is *inter-PoP*
    - ⋆ PoP = Point of Presence
    - ⋆ here we optimise simplified costs
  - second layer is between routers
    - ⋆ here we build redundancy

COLD top layer optimisation

- Inter-PoP
  - recognise that long links cost more
  - so for the moment ignore links inside one city/PoP
  - avoids some (router) constraints
  - reduces the size of the problem to be tractable
    - optimisation is NP-hard
- Randomness
  - PoP locations (as in SERNs)
  - traffic matrix (as in HOT)

COLD costs

Link costs

$$c_e = k_0 + k_1 \ell_e + k_2 \ell_e w_e,$$

where

$$\begin{aligned} \ell_e &= \text{the length of link } e, \\ w_e &= \text{the capacity of link } e, \\ k_i &= \text{a set of constants.} \end{aligned}$$

Node cost = a "complexity" cost

$$c_i = k_3 I(i \in N_H)$$

when node $i \in N_H$ means it is a "hub" or a "core" node

Internet Topology Zoo

- We have a collection of (Internet) network topologies
  http://www.topology-zoo.org/
- One interesting thing is the variety of networks
  - some look like Abilene
  - others are hub-spoke networks
  - others are more meshy
- The costs are flexible to allow all of these combinations

COLD

- Optimisation is solved with a GA (Genetic Algorithm)
  - we need to use a heuristic because the problem is NP-hard
  - GA is still slowish, *i.e.*, $O(n^3)$
- Results
  - are nicely tunable
  - we can't dispose of any of k_i (no simpler model)
  - model parameters have operational meaning (costs)
  - can match real variations of statistics
- Matlab code
  https://github.com/rhysbowden/COLD

Layering, hierarchy, redundancy and structure

- Now we need to build the router layer
  - incorporate redundancy
- But, often, lower layer is build with structure
  - hierarchy
  - graph operators

We'll start talking about graph operators next

Estimation

- We haven't worked out how to do estimation yet!

Estimation

- We haven't worked out how to do estimation yet!
- But I have a good idea how

Further reading I

- Rhys Bowden, Matthew Roughan, and Nigel Bean, *COLD: PoP-level topology synthesis*, CoNext (Sydney, Australia), December 2014.

- M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the Internet topology*, ACM SIGCOMM, 1999.

- Lun Li, David Alderson, Walter Willinger, and John Doyle, *A first-principles approach to understanding the Internet's router-level topology*, Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications (New York, NY, USA), SIGCOMM '04, ACM, 2004, pp. 3–14.

- W. Willinger, D. Alderson, and J.C. Doyle, *Mathematics and the internet: A source of enormous confusion and great potential.*, Notices of the AMS **56** (2009), no. 5, 586–599, `http://www.ams.org/notices/200905/rtx090500586p.pdf`.