Hierarchical Traffic Matrices: Axiomatic Foundations to Practical Traffic Matrix Synthesis

Paul Tune* and Matthew Roughan* and Chris Wiren†
* ARC Centre of Excellence for Mathematical and Statistical Frontiers at the University of Adelaide, Australia,
E-mail: {paul.tune,matthew.roughan}@adelaide.edu.au
† Defence Science Technology Group, Australia, E-mail: Chris.Wiren@dst.defence.gov.au

***Abstract*—The traffic matrix of a network is useful in a variety of applications: network planning and forecasting, traffic engineering and anomaly detection. Much work has focused on estimating traffic matrices, but methods are often tested on limited data. There is then the possibility of unrepresentativeness of the datasets, and the lack of generalizability of the subsequent results. Synthesis can help alleviate this problem.**

In this paper, we examine a fundamental question: what constitutes a good class of statistical models for traffic matrix synthesis? The results of our study is the definition of a set of *axioms* specifying structure on traffic matrix models, including the incorporation of organizational structure (hierarchies) in network traffic. We introduce the Hierarchical Traffic Matrix (HTM) which satisfies these requirements. We then study the hierarchical structure of the GEANT network, a research network based in Europe, to validate our ideas. Finally, we illustrate how structure in traffic matrices can affect network topology design.

I. Introduction

The traffic matrix (TM), which describes the volume of traffic between all pairs of nodes in a network, is an important aspect of the network. TMs are used in several applications such as forecasting future traffic for network planning [27] and topology generation [6], and as a baseline for anomaly detection [17,39].

Most of the literature has focused on estimations of TMs, but data is limited. Most TM data from commercial networks is proprietary: network operators are not keen to release such data as they fear they may lose their competitive edge. Publicly available data, on the other hand, generally derives from research and education networks (RENs) such as Abilene [21] or GÉANT [34], and may not be representative of commercial networks. From a statistical viewpoint, however, even the release of a single commercial operator's TM data would not quell the question of representativeness, and so the current deficit is unlikely to be remedied in the foreseeable future.

Also, researchers require a large ensemble of TMs to test new routing protocols or traffic engineering techniques. TM synthesis models can generate large ensembles of TMs to help with both problems.

In our previous work [33], we proposed that synthesis models incorporate prior knowledge about TMs systematically (*e.g.,* the non-negativity of network traffic). Prior knowledge can be viewed as *constraints* (assumptions), which can be naturally incorporated into models using the the principle of maximum entropy [13] (MaxEnt). These models have a very important advantage: assumptions are laid out explicitly, and therefore the model has no hidden assumptions.

There are some drawbacks, however. First, specifying constraints on TMs of a large network is onerous, and potentially error-prone. For instance, the GÉANT network and its associated national European REN partners have of the order of thousands of routers, so specifying constraints for each end-point is tedious. Second, publicly available GÉANT data provide TMs at the core-level Points-of-Presences (PoPs). Fine-grained data at the end-points is not available, yet researchers may need to generate TMs at this finer granularity.

Moreover, in general, there are several other unresolved issues in TM modeling:

- A network is constantly evolving through the inclusion of new nodes and links. It stands to reason that TM models of the network should change as well, but how should this be done consistently?
- If we only have highly aggregated views of the TM, what can we do if we want to generate TMs at a finer level, such as at the end-points? How can we "zoom" into a part, or out to find a general trend?
- Content Distribution Networks (CDNs) distribute content locally, resulting in traffic concentrated in local regions. How can we include such locality properties?
- How can we incorporate a particular traffic structure of data centers, as these networks have very different traffic properties from traditional carrier networks?

In this paper, we tackle the problem at a *foundational* level. Previous approaches, including our prior MaxEnt models, relied on detailed knowledge gained, perhaps, from data. Here we introduce new types of *consistency* constraints, which may not necessarily come from data. To that end, we define a set of *axioms*, properties large sets of TMs should obey without reference to data. The goal is not to fully describe the TMs. Instead, given there is potentially a large space of possibilities, we want to define a few "rules" that encompass a large enough set of important models, and are useful for generating practical and controllable ensembles of synthetic TMs.

For instance, real networks define zones for security and management purposes. Members within a zone are often statistically homogenous, so the traffic they generate are statistically similar. Furthermore, for instance, in the GÉANT [34] network, PoPs are organized in levels, *i.e.,* comprising a hierarchy. Yet, there has been, as far as we are aware of, no attempt to capture hierarchical structure in TM models, though TMs has been analyzed in multiple resolutions before [36].

We introduce the *Hierarchical Traffic Matrix (HTM)* model, which arises naturally under the set of TM axioms we propose. The HTM also incorporates knowledge about network zones and hierarchy. This is an improvement over previous models and definitions of the TM, including our previous work. Allowing constraints to be specified in zones simplifies the specification of the model, yet captures general trends about the TM, and allows us to "zoom" into a part of the network's traffic to capture locality properties, thus resolving the aforementioned issues.

Our contributions in this paper are the following:

- we lay the foundation of an axiomatic framework for TM synthesis models,
- the HTM, which arises naturally from a set of axioms and captures the interaction between the hierarchical structure of a network and the TM, and
- an example of the model application on real-world TM data from GÉANT.

We then apply the HTM to artificially generate network topology and show how hierarchical structure in the TM is reflected in the generated network topology. The generated topologies form *zones* based on the HTM's structure, in contrast to previous works [6,33] that showed network topologies are largely insensitive to the TM.

II. Background and Related Work

A. Modeling traffic matrices

A general introduction to TM models can be found in [32], but we summarise some of this literature here. Many TM models have been proposed over the years [7,30,35]. Successive models [11,20,28,37,38] based on better assumptions of network traffic, resulted in better inference techniques. However, most of the modelling has been aimed at inference rather than synthesis.

Roughan [26] proposed using a random gravity model as a spatial model for TM synthesis, and Oikonomou [23] noted that it is a MaxEnt model for traffic under a certain set of assumptions. In fact, if one examines the transportation literature, that insight is even older [25]. This is a special case of the framework presented here.

Our work differs from prior art in two ways. First, to the best of our knowledge, there have been no attempts in past works in formulating a set of axioms for TM statistical models. We introduce novel *consistency constraints* that are useful in developing synthesis models. We believe that an axiomatic approach will further advance the state-of-art in modeling TMs (and network traffic) by encouraging clear thinking about TMs.

Moreover, past models do not account for organizational structure in the network as we do in the Hierarchical Traffic Matrix in §IV. Hierarchical structures exist in real networks [24] and we show this in an example of a real European network (see §IV-B). Aggregating traffic from IP addresses to subnets and to PoPs to form TMs at various levels of coarseness have been discussed before and a taxonomy was proposed [19], but this is different from incorporating hierarchical structure of the network topology itself into TMs.

Hierarchical models themselves have been studied in geographical modeling. In this context, multiscale spatial models have proven useful [8,16]. Such models also appear in image analysis [10]. However, spatial structure in these fields is different from what we consider here, because geographic locality has direct implications for the data. For instance, in image analysis, we can consider an image as a matrix. Each row has a strong spatial relationship with other rows, so permuting an image row-wise will distort the original image. In contrast, the labels of the rows of a TM are arbitrary so permutation of rows is no less arbitrary.

The closest work to ours is a proposed framework for multi-resolution analysis of TMs [36], but their work focuses on analysis rather than modeling. To our knowledge, there have been no attempts to apply hierarchical models to TMs, as we do so here.

To develop useful statistical models from these ideas we use the principle of maximum entropy, which we describe next.

B. The principle of maximum entropy

The application of parsimony to describe collected data dates back to the late 18th century [29]. The principle of insufficient reason often credited to Pierre Simon Laplace, states that given an observation, one should remain undecided about all the potential events explaining the observation so we assign equal probability to potential events. Laplace and Jacob Bernoulli considered the concept to be intuitively obvious, seeing little need to formalize it.

It wasn't until the concept of *information entropy* [9] was introduced that significant progress was made. Shannon's information entropy is defined as $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$, where X is a discrete random variable taking values in set $\mathcal{X}$ with probability mass function $p(x)$, with the convention $0\log 0 = 0$. Entropy measures the average uncertainty of a random source.

The uniform distribution over $\mathcal{X}$ represents the *least committed distribution* in the absence of further data. The entropy is maximized by this distribution, *i.e.,* $p(x) = 1/|\mathcal{X}|, \forall x$. Equal weights are assigned for all outcomes of X, consistent with the intuitive notion of being least committed.

If new information, formulated as constraints, is available, the principle of maximum entropy states that the best strategy is to choose a distribution that maximizes entropy subject to these constraints. Therefore, the solution is one that conforms to known observations about the data and no more than that.

Jaynes [13] used the Shannon entropy to construct the maximum entropy (MaxEnt) framework. Since we consider continuous random variables as an approximation of the traffic volumes, we use the *differential entropy* [9], denoted by

$$h(f) = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\mathbf{X})\log f(\mathbf{X})\, d\mathbf{X} \tag{1}$$

for a random TM $\boldsymbol{X}$ distributed with density $f(\mathbf{X})$. Let $\mathcal{C} = \left\{\mathbf{X}\,|\,\{\phi_\ell(\mathbf{X}) = a_\ell\}_{\ell=0}^{L}\right\}$ be a convex set of $L+1$ constraints on $\mathbf{X}$, with $\phi_0(\mathbf{X}) = 1$ since $f(\mathbf{X})$ is a density function. The MaxEnt optimization problem is

$$\max_{f(\mathbf{X})} h(f) \text{ s.t. } \mathbf{X}\succeq \mathbf{0},\ \mathbf{X}\in\mathcal{C}, \tag{2}$$

where $\mathbf{X} \succeq \mathbf{0}$ denotes non-negativity of $\boldsymbol{X}$. The problem is convex[1] so any local maximum is the global maximum, the solution being the *maximum entropy model*.

The optimal solution comes from the Boltzmann distribution [13], generically given by

$$f(\mathbf{X}) = \frac{1}{Z} \exp\left(- \sum_{\ell=1}^{L} \lambda_\ell \phi_\ell(\mathbf{X}) \right), \ \mathbf{X} \succeq \mathbf{0}, \tag{3}$$

where $\{\lambda_\ell\}_{\ell=1}^{L}$ are the Lagrange multipliers obtained via the Calculus of Variations applied on (2), Z is the normalization factor (note: the constraint $\phi_0(\mathbf{X})$ is included via Z) and $\mathcal{X}$ is the support of the distribution. There is only one Lagrange multiplier per constraint. These distributions appear in our models, with their form determined by the constraints, by (3).

MaxEnt was applied to TM synthesis modeling in [33]. In that work, various statistical constraints on traffic (spatial, temporal and spatio-temporal) were considered, which resulted in variants of the random gravity model [26], and several other new models. However, no axioms for TM models were proposed, nor was hierarchical structure included. In this paper, we do both, and also incorporate MaxEnt with hierarchy to develop new TM models.

III. Axiomatic Foundation

A. The case for axioms

An *axiom* is often stated to be a premise so evident as to be accepted as true without controversy. Axioms are the starting point – the bedrock – on which foundations of a subject are built. In mathematics, however, there is a subtle implication that an axiom is considered "unprovable". It is an assumption made to start a chain of logic, but without incontrovertible truth in itself.

The classical case is Euclidean geometry. For centuries, mathematicians wrestled with the axioms of geometry, eventually realizing that they could be altered, leading to new and interesting non-Euclidean geometries. So an axiom fulfils this dual role of stating the properties that we believe true (without proof), and partitioning an unruly space of possibilities into meaningful classes for which we can prove useful properties.

With modeling, such mathematical niceties are often discarded in favor of empirical data. In the era of "Big Data", statisticians and computer scientists have presumed that axiomatic approaches to modeling might be ignored in favor of "realistic" models, or even that modeling might be discarded altogether in favor of machine learning and related techniques.

There are, however, many cases where data **cannot** ever replace clear thinking, notably in the generation of synthetic complex-structured data such as large structured matrices, or graphs. The reason data cannot just fix our problem in these cases is that when dealing with such problems we are aiming to create an N^2 dimensional dataset, with large N, and inhomogeneous structure. Real measurements of the world contain at most a few samples of such. For instance, we have only one Internet, *i.e.,* one sample of such a network. We could view multiple snapshots of this network, but they are not independent samples.

[1]Strongly convex under linear constraints.

We therefore want to describe some objects in N^2 space, but have $K \ll N^2$ samples. Statistically, we cannot hope to derive a sensible model. The solution most commonly applied is to assume that the object in question actually has homogeneous structure (*i.e.,* repeats in some way). We then look at statistical properties of this structure: for instance, in graphs, we look at properties of node degree, centrality, or clustering, and we effectively project this very high-dimensional object into a few dimensions. Generally, at most two dozen such graph characteristics are considered. By assuming independence between sub-components, and homogeneity across them, we can reuse the sub-components of the graph to obtain statistical characterizations. For example, if all nodes are assumed alike and independent instances of a generic "node", we can use each node to provide one sample of the node-degree distribution.

However, the data we observe isn't homogeneous. Real networks can have subnetworks that are both dependent and different! Statistical summaries based on the homogeneity assumption may appear to support the assumption of homogeneity though, making it appear that simple random graphs are a viable model. But this is hidden in complex matrix and graph analysis – our goal here is to make such assumptions completely explicit.

When using such data we are often trying to answer some questions about the effect of some set of properties as they vary, but we only have a few datasets, and these don't exhibit the range of properties we care about. So we need to account for cases outside of the range of the available data, but do so in a constrained manner, otherwise the space to explore is too large. Moreover, we should generate multiple (non-isomorphic) instances in order to provide statistical analysis of results. Sometimes we need to generate a large number.

The result is that data, while helpful, cannot solve the problems of modeling of large, complicated matrices or graphs.

Instead, we go back to basics. We consider an axiomatic framework for TMs. Remember though, that while we will choose axioms based on their "self-evident" truth, we consider these as features that we can incorporate, or not, to constrain the exploration of the modeling space, not as "facts."

B. TM definition

We first define an abstraction of TMs. Let Ω be a non-empty set of nodes of a network (graph), with $|\Omega| = N$. Nodes may be *physical* or *logical*, where the former is tied to geographic location, while the latter is related to the Internet's address structure, such as the IP address space. However, here we arbitrarily label each node by a unique integer $k = 1, 2, \cdots, N$ to emphasize that spatial relationships play less part here. For instance, typically, there is little similarity between traffic from adjacent IP addresses.

Traffic matrices evolve over time, but here we shall examine *purely spatial TMs* (for a discussion of spatiotemporal TMs, see [33], from which it should be clear that many of the ideas presented here generalise).

Let $\boldsymbol{X}$ be the TM. We adopt the convention where each row i of $\boldsymbol{X}$ denotes the outgoing traffic from the source, i, to each j, while each column j denotes the incoming traffic from all i into a node, the destination, j, so the TM entry $X_{i,j}$ describes the volume of traffic from i to j (measured in bytes or packets per time interval) for all pairs $(i, j) \in \Omega \times \Omega$.

C. The class of models and axioms

A class $\mathcal{C}$ of statistical TM models is defined as an ensemble of matrices $\boldsymbol{X}$ satisfying at least one constraint and which is equipped with a probability measure $\mu(\cdot)$. A probability measure, informally, assigns a value between 0 and 1 to the set of events in a probability space. We could be more formal in our definitions by bringing in the machinery of measure theory [5], however, we favor a lightweight approach to motivate this work, as deeper theory is not fundamental to our arguments and would detract from the main points.

Information about the ensemble to be simulated will be incorporated in the form of constraints. There are three types of constraints: *deterministic*, or hard constraints on allowable matrices, *probabilistic*, soft constraints on the measure that specifies the probability of matrices, and *consistency* constraints that incorporate properties of the ensemble under transformations. The former two were included in past work, the latter is new here, but we describe each in detail below.

Deterministic constraints. Deterministic constraints are hard constraints because they restrict the space of TMs. Deterministic constraints can be *equality* or *inequality* constraints. For example, we usually require TM entries to be non-negative, since they represent traffic. An example of an equality constraint is fixing the total traffic of a TM via $\sum_{i,j} X_{i,j} = T$, so all TMs in an ensemble have their total traffic equal to T, as used in oblivious routing design [3,40].

A more subtle example is that some existing definitions of TMs do not allow self-traffic, *i.e.,* there are zeroes on the diagonal. This would, however, depend on the definition of the node. For instance, with IP address-to-IP address TMs, self-traffic should be zero, but a PoP-to-PoP TM can have self-traffic that describes the internal traffic within a PoP. If self-traffic is not allowed, then we require $X_{i,i} = 0$ for all i.

Note that deterministic constraints may not just be linear constraints, though typical TM models include only such.

Let the region of admissible TMs constrained by the deterministic constraints be denoted as $\mathcal{X}$. We assume that the constraints result in a set of matrices that is non-empty, is not a singleton and whose elements are not all isomorphic under permutation[2]. We still require a way to synthesize TMs from $\mathcal{X}$, and this requires that we impose a probability measure, which we describe next.

Probabilistic constraints. We could specify the probabilistic measure on our ensemble explicitly, but we rarely have enough information to do so. Instead, we have some constraints on the measure. To avoid technical issues, we assume that the constraints result in a non-empty set of positive measure.

Let $\mathbb{E}_\mu$ be the expectation operator with respect to measure $\mu(\cdot)$. Constraints, for instance, could be

$$\mathbb{E}_\mu[\mathbf{A}\boldsymbol{X}] = \mathbf{B}, \tag{4}$$

[2]The arbitrary nature of the indices means that matrix and its permutations are equivalent for our purposes.

i.e., the set of linear constraints on the expected values of the components of the members of the ensemble of TMs. As an example, the random gravity model class [26] is defined as the class of matrices

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}/T, \tag{5}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are both random vectors such that

$$\mathbb{E}_\mu[U_i] = r_i, \ \mathbb{E}_\mu[V_j] = c_j,$$

and r_i, c_j are the mean row i and column j sum constraints, *i.e.,* the average sum of the entries of the rows and columns, respectively, and T is the mean total traffic constraint on the TMs. The measure on these spaces can be specified indirectly, for instance in [26], the class above was given a measure implicitly via the random gravity model, or the MaxEnt measure which we will apply in the next section.

Consistency. There are certain facts: for instance, if we take a IP-address-level TM, and add the traffic from IP addresses within given prefixes and form a matrix of these, the result will still be a traffic matrix, albeit at a different level of aggregation. We might therefore expect that the new aggregated ensemble preserve some properties of its more fine-grained parent.

Consistency conditions generalise this idea — after some transformation, we expect that the matrix will have some of the same properties. We express this by requiring that some set of deterministic or probabilistic constraints retain their functional form after a transform on the TM.

Suppose we start with an ensemble $\mathcal{X}$ of TMs with a set of constraints, such as (4), on the measure $\mu(\cdot)$. Let there be a mapping $\phi : \mathcal{X} \to \mathcal{Y}$, *i.e.,* from the class $\mathcal{C} = (\mathcal{X}, \mu)$ to $\mathcal{C}' = (\mathcal{Y}, \mu')$ such that a measure exists and satisfies

$$\mu'(\mathbf{y}) = \int_{\mathcal{X}} \mu(\mathbf{x}) \mathbf{1}[\phi(\mathbf{x}) = \mathbf{y}], \tag{6}$$

where $\mathbf{1}[\cdot]$ is an indicator function. In non-measure theoretic terms, (6) is simply a statement about the marginal distribution

$$\Pr(\boldsymbol{Y} = \mathbf{y}) = \int_{\phi^{-1}(\mathbf{y})} \Pr(\phi(\boldsymbol{X}) = \mathbf{y} \,|\, \boldsymbol{X} = \mathbf{x}) \, d\mathbf{x}, \tag{7}$$

where ϕ^{-1} is the inverse map of the transform ϕ.

Now, we want to ensure that constraints on any TM statistical model are still consistent after transformation of the TM via a mapping ϕ. As an example, suppose we have a constraint, for all i, j,

$$\mathbb{E}_\mu\Big[\sum_{i,j} X_{i,j}\Big] = T, \tag{8}$$

that is, the expected total traffic in the TM.

Take a partition $S_1, S_2, \cdots, S_K$, *i.e.,* disjoint subsets such that their union is Ω. An important transform is the *aggregation transform*: $\phi(\mathbf{X})$ takes $\mathbf{X}$ and results in $\mathbf{Y}$ with

$$Y_{i,j} = \sum_{k \in S_i} \sum_{\ell \in S_j} X_{k,\ell}. \tag{9}$$

Under the transform ϕ, (8) remains consistent since

$$\mathbb{E}_{\mu'}\Big[\sum_{i,j} Y_{i,j}\Big] = T, \tag{10}$$

remains true, and therefore the measure $\mu'(\cdot)$ is consistent under aggregation, with respect to constraint (8).

Similar to the aggregation transform we can define a *sub-sampling transform*: partition Ω into two subsets, S and $\Omega \backslash S$, and restrict our attention to the entries of members in $\Omega \backslash S$, *i.e.,* form $Y_{i,j} = X_{i,j}$ for $i, j \in S$. This allows us to "zoom in" on a submatrix of the TM, but the constraint (10) is only consistent under the subsampling transform if we restrict the probability of certain matrices.

On the other hand, a measure cannot be consistent under aggregation with respect to the constraint

$$\mathbb{E}_{\mu}[X_{i,j}] = 1, \tag{11}$$

because under transform (9), using the same functional form as (11) would form the new constraint

$$\mathbb{E}_{\mu'}[Y_{i,j}] = 1, \tag{12}$$

but, due to the linearity of expectation, $\mathbb{E}_{\mu'}[Y_{i,j}] = |S_i||S_j|$, *i.e.,* a different functional form from (11).

The random gravity model (5) is an example of a generative model that is consistent under the aggregation transform, since the row, column and total traffic sum constraints still hold after a change of measure. It is, however, not consistent with a no self-traffic constraint under this transform, *i.e.,* the diagonals $X_{i,i} = 0$. For instance, after aggregating entries in the Origin-Destination TM, where the labels are IP addresses, the resulting Ingress-Egress TM does satisfy the no self-traffic constraint, because there may be traffic between IP addresses in a PoP. An illustrative example of this is found in [32].

So, *consistency under constraints* means that the model specified in class $\mathcal{C}$ can be transformed without changing these constraints in class $\mathcal{C}'$, though the actual values could change.

TMs are transformed in different ways for analysis or for storage. We have seen that one such operation is the aggregation operation, where sub-matrices of the TM are summed together to form a lower-dimensional TM with $N' < N$ rows and columns. Often, there is an interest in different levels of aggregation. For instance, planning a network requires coarser partitions of traffic, say at the PoP-level, while finer levels at the IP level are needed for anomaly detection. When we specify constraints at a finer level, we want the constraints to carry over after aggregation, and we expect constraints defined on the TM model to be consistent in this sense.

Clearly, not all constraints will be consistent under a transform. Likewise, given a constraint, there may be a limited set of consistent transforms. The key point, however, is that we want consistency of the constraints to apply to transforms that we deem important, such as the aggregation transform.

We require another condition. A TM model has *consistency of measure*, if after a transform $\psi : \mathcal{X} \to \mathcal{Y}$, the measure remains unchanged *i.e.,* $\mu'(\mathbf{Y}) = \mu(\mathbf{X})$. What this means is that changes to the TM should not affect the underlying probability density.

One example is the measure consistency under the *permutation transform*, where, when the rows or columns of a TM ensemble are permuted, the model's probability density remains unchanged. However, one cannot simply permute any row or any column in a TM. For example, permuting rows in a PoP-level TM is valid, but swapping a member in PoP A with another in PoP B fundamentally changes the structure of the PoP-level TM. In the latter case, the TM model has changed. Thus, we need a constrained definition of permutation.

Suppose there are L levels of aggregation, 1 being the coarsest, say at the PoP-level, and L the finest, say the IP address level. Let $S_{1,\ell}, S_{2,\ell}, \cdots, S_{K,\ell}$ be disjoint subsets of Ω such that their union is Ω at level $1 \le \ell \le L$. Let $X_{S_{i,\ell},S_{i,\ell}}$ be the submatrix of a TM X of nodes in $S_{i,\ell}$. Also, let $\mathbf{P}$ denote a permutation matrix (note: the identity matrix is also a permutation matrix). Then, the class $\mathcal{C}$ has consistency of measure under permutation if

$$\mu'(\mathbf{P}\mathbf{X}_{S_{i,\ell},S_{i,\ell}}\mathbf{P}^{\mathrm{T}}) = \mu(\mathbf{X}),$$
$$\mu'(\mathbf{P}\mathbf{X}_{S_{i,k},S_{j,\ell}}\mathbf{P}^{\mathrm{T}}) \neq \mu(\mathbf{X}), \text{ for } i \neq j, \text{ or } k \neq \ell.$$

That is, we require measure consistency when permuting within a group at some level of aggregation, but not when permuting between groupings. For example, routers are assumed to be exchangeable within a PoP, but not between PoPs (within the model). So measure consistency applies only to members of the subset at a level, but not members in different subsets or different levels.

In summary, we require models to obey deterministic and probabilistic constraints, as explained in [33] do too. What's new is our requirement of model consistency:

To reiterate, given a transform $\phi : \mathcal{X} \to \mathcal{Y}$, which implies a measure $\mu'(\mathbf{Y})$, we can define:

Axiom 1 (Consistency under constraints): A model is consistent under transform ϕ with respect to its constraints, if its set of deterministic and probabilistic constraints in the new class $(\mathcal{Y}, \mu')$ take the same form.

Axiom 2 (Consistency of measure): A model has consistency of measure under transform ϕ if $\mu'(\mathbf{Y}) = \mu(\mathbf{X})$.

D. Discussion

There is a limit to any axioms and constraints. They may apply in one setting but not another.

For instance, we could propose a "scaling" transform, where a TM is multiplied by a scalar. In many settings this should result in consistent constraints, however, that presumes that there are no capacity constraints imposed by the network. If not, congestion could alter the carried load (carried load is the usual observable when dealing with traffic matrices).

It is important to choose appropriate axioms/constraints for the setting of interest.

IV. Hierarchy and Models

A. Hierarchical Traffic Matrices

Network nodes, for instance, hosts and routers, are often grouped by location (physical PoPs), or logical groupings (by function, purpose or access type) [24]. We refer to such groupings in general as *zones*. Network zoning, or network segmentation and segregation, is considered a best practice in network management because it simplifies network management and improves security [2].

For larger networks, there might be multiple levels of grouping. For instance GÉANT connects many smaller European national RENs, each independently managed, and each with its own internal structure, *e.g.,* PoPs, OSPF areas and so on. In this section we consider natural models that arise in the presence of such hierarchy.

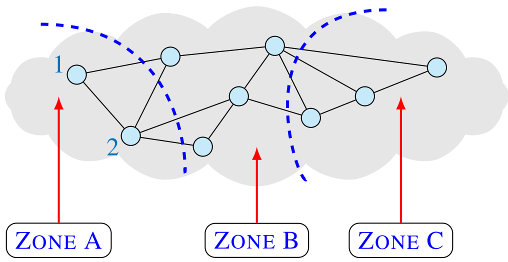


Fig. 1: *Example of zoning in a network. Here, the network is divided into three zones: A, B and C. Zone A has two nodes labeled 1 and 2.*

It is natural to model members within a zone as possessing similar characteristics. This suggests a TM model defined in terms of a hierarchy: the coarse level covers the network zones, with members within the zones forming the finer level.

The HTM seeks to incorporate the information we may have about the traffic properties of each zone. Conceptually the idea is simple: traffic parameters describe the traffic per zone, with the idea that the nodes within a zone are statistically homogenous. If there is a set of nodes which are not, then these nodes can be separately grouped into their own zone.

Figure 1 presents an example of a network divided into three zones. The HTM is

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{A,A} & \mathbf{X}_{A,B} & \mathbf{X}_{A,C} \\ \mathbf{X}_{B,A} & \mathbf{X}_{B,B} & \mathbf{X}_{B,C} \\ \mathbf{X}_{C,A} & \mathbf{X}_{C,B} & \mathbf{X}_{C,C} \end{bmatrix}.$$

Each sub-matrix $\mathbf{X}_{i,j}$ describes the traffic traversing from Zone i to Zone j. Here, the diagonals describe the traffic within zones, while the off-diagonal elements describes traffic between zones. Within each zone, we can specify the sub-matrix, for instance, in Zone A,

$$\mathbf{X}_{A,A} = \begin{bmatrix} X_{1,1}^{A,A} & X_{1,2}^{A,A} \\ X_{2,1}^{A,A} & X_{2,2}^{A,A} \end{bmatrix}.$$

The hierarchy naturally lends itself to aggregation and permutation consistency requirements. However, here, we still need to develop statistical models that (i) have this property, and (ii) come with a practical synthesis algorithm.

B. Example hierarchy in real networks

Before moving on to the models, we study an example of hierarchical structure on actual networks. We look at the GÉANT network, a REN, based in Europe. We chose GÉANT because it is a well-documented network [12,22,34].

Figure 2 presents the geographic and logical structure of GÉANT (2010). Notably, in the logical view, we see groups of geographically related networks that link both through GÉANT, and directly. For instance SANET, ACOnet, CESNET and PSNC (Austria, Czech Republic, Poland and Slovakia), connect directly, and NORDUnet connects a sub-group of national RENs in Scandinavian countries (denoted IS, NO, SE and FI in the northern region of the geographical layout).

In previous work [1], it was shown that a direct application of a gravity model to GÉANT TM data [34] without accounting for the structure of NORDUnet resulted in a poor fit. GÉANT TM data only provides the traffic between directly connected RENs, so the traffic between members of NORDUnet is not available. The gravity model was modified to account for NORDUnet members by assuming that each member is statistically homogenous. Once the group information from NORDUnet's members was included, the modified gravity model fits the data extremely well [1, Figure 4, p. 13].

Thus a TM synthesis model in this context should (i) take into account the nature and structure of the hierarchy, (ii) be able to exploit the available information at the level given.

C. MaxEnt Hierarchical models

We now apply MaxEnt to the problem. The beauty of MaxEnt is that, without additional knowledge on the members within a zone, they will automatically be statistically homogeneous (satisfying our requirements) but if more detailed information is available it can be incorporated. We use here constraints based on the type of data that has typically been available (in much of the literature), and we present two models: the hierarchical gravity model and the intra- and inter-zone (IIZ) model.

Hierarchical gravity model. Suppose a network has M zones. There are three basic probabilistic constraints for the model:

- *fixed mean incoming (ingress) traffic per zone,* $\mathbf{c}$;
- *fixed mean outgoing (egress) traffic per zone,* $\mathbf{r}$; and
- *fixed mean overall total traffic,* T (which is implicitly determined by the previous two constraints, but which we add here for clarity).

We can increase control by specifying the variance parameters of above quantities, essentially adding another three probabilistic constraints.

Suppose Zone A and Zone B have N members each. The traffic generated by a member within the group then is a simple scaling of the traffic of the zone. If traffic constraints between zone A and B are given by r_A and c_B, and there are N nodes within the zone, then entries within the zone have the form

$$\boldsymbol{X}_{A,B} = T\bar{\boldsymbol{U}}\bar{\boldsymbol{V}}^{\mathrm{T}} \tag{13}$$

where now $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are IID random vectors where entries come from the exponential distribution with mean $r_A/(NT)$ and $c_B/(NT)$. Members within a zone are statistically homogeneous by MaxEnt. This form is similar to the random gravity model [26,33], but with a scaling factor $1/N^2$ for each entry.

Intra- and inter-zone (IIZ) model. Additional knowledge can be obtained by measuring the traffic between zones. For instance, consider the GÉANT TM data which provides a TM over the top-level of GÉANT but no visibility into distribution of traffic within NORDUnet. The constraints are specified in terms of the average total traffic within and between the zones, and represented in a matrix $\mathbf{T}$, *e.g.,*

$$\mathbf{T} = \begin{bmatrix} T_{A,A} & T_{A,B} & T_{A,C} \\ T_{B,A} & T_{B,B} & T_{B,C} \\ T_{C,A} & T_{C,B} & T_{C,C} \end{bmatrix}. \tag{14}$$

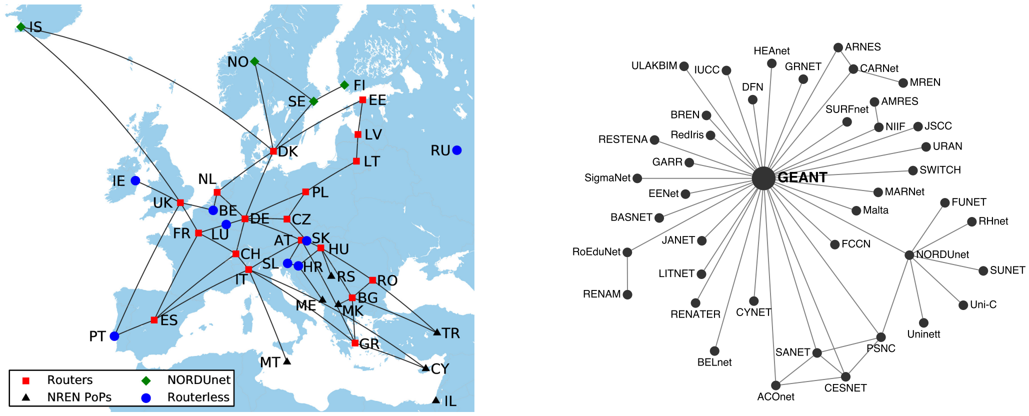


Fig. 2: *The GÉANT network topology (as of 2010).* ***Left****: Geographical layout,* ***Right****: Logical topology.*

MaxEnt again treats each node within a zone as independent and homogenous, but with constraint on the sums over the groups of nodes given by the elements of $\mathbf{T}$. Hence (again assuming constraints on means but not variances) the elements $X_{i,j}$ will be independent exponential random variables with mean given by $T_{S_i,S_j}/|S_i||S_j|$ where $i \in S_i$ and $j \in S_j$.

The IIZ model is more refined than the hierarchical gravity model, since its constraints are more detailed than the latter model.

In either model, adding a variance constraint changes the exponential distributions used into truncated normal distributions [33].

A major advantage of HTMs in synthesis as compared to other TM models is that, rather than specify parameters for each node in the network, parameters for the zones need only be specified. This could allow the model to be matched to a setting such as GÉANT where only limited data is available, or aid in model specification. For instance, in data center networks, hosts and routers number in the hundreds of thousands. Trying to specify parameters for all these nodes would be an ardous and error-prone task. Thankfully, data center TMs are known to have strong locality effects [4,14].

D. Relation to the axioms

Here, we show that the HTM naturally arises under the simple set of axioms and constraints outlined in §III.

Formally we defined the hierarchy by a set of nested *partitions* of the set of nodes Ω [16]. Take a collection of subsets $S_{i,j} \subset \Omega$, $i = 1, 2, \cdots, K$, and level $1 \leq j \leq \ell$ that satisfy

- for each level j, the collection $\{S_{i,j}\}_{i=1}^{N_j}$ must form a proper partition of Ω, and
- each subset $S_{i,j}$ can be expressed as the union of a unique set of elements in the partition at level j, *i.e.,* $S_{i,\ell} = \cup_{i' \in child(i)} S_{i',\ell+1}$, where $child(i)$ is the collection of indices of the children of $S_{i,\ell}$.

Here $j = L$ is the finest, and $j = 1$ the coarsest level of aggregation.

HTMs are *consistent under aggregation with respect to non-negativity* as a direct result of summation.

HTMs are *consistent under aggregation with respect to row- and column-sum constraints* in the aggregation transform (9). Linearity of expectations requires this.

However, constraints such as the no self-traffic constraint *i.e.,* the diagonal entries are zero, might hold at the finest level, but not at the coarser levels, and so the model is not consistent under aggregation with respect to this constraint – but we did not require it to be.

HTMs have *consistency of measure* under permutation between zones, or members within a zone, but this consistency does not apply between members of different zones. For instance, permutation is not admissible for a scenario involving swapping the labels between router 1 in zone A and router 1 in zone B, since these two routers are homogeneous entities only within their zone.

By imposing the MaxEnt measure on our models, the MaxEnt HTM models satisfy another axiom:

Axiom 3 (Subset independence): Let $S_1, \cdots, S_K$ be disjoint subsets of Ω whose union is Ω. Constraints on S_i are independent of other subsets S_j, $j \neq i$.

While more restrictive than the other axioms, there are advantages to subset independence. First, independent subsets can be generated separately, allowing parallelization for fast generation. Second, the MaxEnt HTM models allow a researcher to focus on one submatrix of the TM if need be, so there is greater control over the generation of the ensembles, such as adding locality constraints for CDN TMs. It is also very useful for testing and tuning of applications that require TMs.

V. Example

In this section, we study how hierarchy in TMs affects the design of a PoP-level network topology by applying the HTM to generate such topologies.

Real-world PoP-level topologies are determined by factors such as demographic and geographical constraints, traffic demand, the designer's experience, and financial costs [18]. The Internet Topology Zoo [15] demonstrates that real designs are variable, with some as simple as a hub-and-spoke networks to more complex designs.

If we are to create synthetic topologies then we must have input TMs to the generation process as in [6,33]. Surprisingly, [6,33] found that the resulting designs were largely insensitive to the statistical properties of these TMs. In these works, however, hierarchy in the TM was not included and the topology of the network was primarily determined by the peak TMs, *i.e.,* the TMs with the largest amount of total traffic. We here show that introducing hierarchy can change the network topology; in particular, there is an observable effect of clustering of nodes into zones (see below). Through in this example we show how TMs impact the topology design.

We note that the relationship between TMs and topology is not strictly one-way. Network topology also affects traffic, so it is important to understand the topology of the network to understand the TMs that arise. In fact, it is likely that both the network topology and the TM co-evolve over time. As the network gets larger it can cope with more demand, and the TM properties change, for instance, due to hierarchical structure changes. Conversely, the resulting changes in TM properties, driven by new demand, necessitate upgrades in the network topology.

We discuss how to generate the topologies and how a hierarchical structure in TMs is reflected in the topological structure of a network next.

A. Simulation setup

The topologies we generate have $N = 20$ nodes. Nodes are placed on a 20 by 20 unit square. The plane is divided into 4 quadrants, to simulate the zones, with the zones containing 5, 6, 5 and 4 nodes. Nodes in each quadrant are placed randomly around the center of the quadrant. We will vary the distance of the nodes from the centers of each zone by controlling a parameter λ, where by increasing λ, the average distance from the center increases. We performed similar experiments with other setups, and similar results, so we report the simplest case here for clarity.

Topology design involves tradeoffs between different costs. Here we use COLD [6] to model and solve the resulting combinatorial optimisation problem. COLD applies a genetic algorithm as a heuristic to solve for the topology with the minimum cost, based on a function with four cost parameters:

- k_0: the cost for the existence of a link,
- k_1: the cost for the physical length of a link,
- k_2: the the bandwidth cost over the length of the link, factoring in operating expenses, initial expenditures on equipment *etc.,* and
- k_3: a complexity cost of a PoP with more than one link, called a non-leaf or *core* PoP.

The two costs of primary interest here are costs k_2 and k_3. The idea here is to set the cost of a hub high *i.e.,* k_3 is large, so the nodes are less likely to connect if there is little traffic between them. In this way, a hierarchical structure in the traffic will be expressed in the topology. In our experiment, we set the costs to be $(k_0, k_1, k_2, k_3) = (1, 20, 10, 1000)$. Note that the scales here are different, so k_3 is not 50 times larger than k_2; each cost is different and not trivially comparable.

For hierarchy, we use the IIZ model from § IV-C. We presume we do not possess refined knowledge of traffic of the PoPs in each zone, and so each PoP within a zone is treated as statistically homogenous. We will compare this model against a random gravity model (5) [33], where hierarchy is absent. COLD's parameters are the same for both models.

For fair comparison, both models are scaled to have the same row, column sum and total traffic constraints, so that the random gravity model acting as a control to compare the effect of hierarchy. We also include a parameter σ^2 to control the traffic variation, so the entries of the TMs of both models are distributed according to the truncated normal distribution [33].

We need a metric to measure the differences in the topology between the two models. We use a simple clustering metric: the ratio of the number of outgoing links from a zone (inter-zone) to the number of links contained within a zone (intra-zone). We expect topologies with hierarchical structure to rate low on the metric, since there should be more intra-zone connections than inter-zone connections.

B. Results

As a concrete example, Figure 3 presents two topologies generated by COLD: one generated with the HTM and another with a random gravity model. Figure 3(b) shows that the HTM induces 4 distinct clusters around the corners of the plane, while Figure 3(a) has a meshier topology. We see that the clusters are formed due to higher intra-zone traffic compared to inter-zone traffic in the IIZ model. In contrast, since all nodes are statistically homogeneous in Figure 3(a), there is no such concentration of traffic in any one cluster.

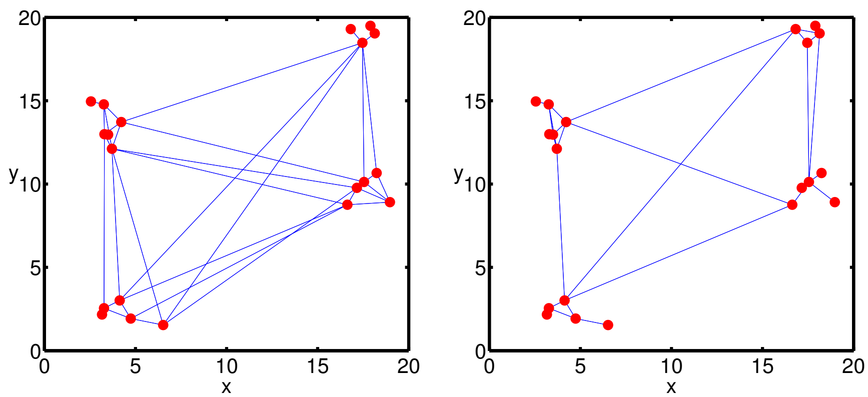
We then investigate what happens when one changes the ratio of traffic in- and between zones in the IIZ model. To that end, we defined a new parameter $0 \leq \gamma \leq 1$ such that the total traffic for each zone $i = 1, 2, \cdots, M$ (*i.e.,* the rows of (14)) obeys

$$r_i = \gamma T_{i,i} + (1 - \gamma) \sum_{j \neq i} T_{i,j}. \tag{15}$$

The extremes, $\gamma = 0$ and 1, imply that the TM has zero intra- and inter-zone traffic respectively. Intuitively, we expect the clustering ratio of the HTM to be smaller than the random gravity model as $\gamma \to 1$, when inter-zone traffic is low. Indeed, this intuition is correct.

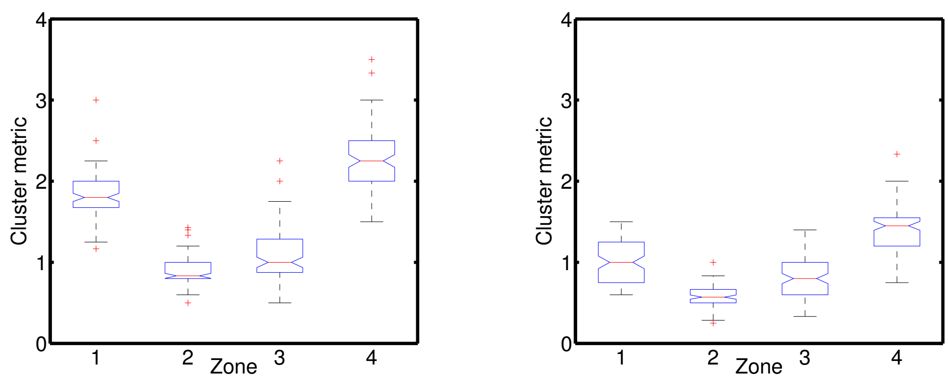
We generated 100 topologies (trials) with the parameter sets $\lambda = 1$ and $\sigma^2 = \{10, 100\}$. As a preliminary result, Figure 4 presents the boxplots of clustering coefficients for $\sigma_s^2 = 1$ and $\sigma^2 = 10$ with $\gamma = 0.6$. For each trial, the node locations remain the same but the input TMs differ between plots (a) and (b). The clustering coefficients are measured per zone, and in plots (b), though technically there are no zones here, for comparison the exact same nodes are grouped into zones as per the case in (a). It is clear that even with different parameters, the clustering coefficients for topologies generated with the HTM is smaller than that of the random gravity model for every zone.

We next vary γ. Figure 5 presents the average clustering ratio for the 4 zones with $\lambda = 1$ and $\sigma^2 = 10$ respectively. The vertical lines represent the 95% confidence interval. As γ increases, we see that the clustering metric for HTM decreases, especially once $\gamma \geq 0.5$. In comparison, the average clustering



(a) No hierarchy (b) With hierarchy

Fig. 3: *Examples of generated topologies with the random gravity model and the IIZ model on the same set of PoPs. The PoP locations were randomly generated on a 20 by 20 unit square in quadrants. In (a), when no hierarchical structure is present, the topology becomes mesh-like with no clear groups. In contrast, in (b), we can see 4 groups of PoPs at the 4 corners of the plane, due to the strong hierarchical structure present in the HTM.*



(a) No hierarchy (b) With hierarchy

Fig. 4: *Boxplots of the clustering metric for each zone for the 100 topologies generated with (a) no hierarchy, (b) hierarchy. Here,* $\lambda = 1$ *and* $\sigma^2 = 10$*. The notch represents the mean, the edges of the box are the 25th and 75th percentile, while the whiskers extend to the most extreme outliers. We can see that there is tighter clustering with the IIZ model as input into COLD.*

ratio for the random gravity model is $\{1.62, 1.28, 1.68, 2.04\}$, so there is a significant decrease after $\gamma \geq 0.5$. Tests with $\sigma^2 = 100$, *i.e.,* larger TM entry variation, showed similar results.

Though the results are not presented here because of space constraints, we tested COLD with other zone parameters, for instance, with 8, 2, 3, and 7 nodes in the 4 zones to test the effect of smaller zones, the results are largely the same, with slightly larger confidence intervals due to the fact that some zones are now larger, and therefore has larger variations in traffic. We performed experiments where we increased both the number of nodes and zones, and the results are similar to the results discussed here.

The key point of our example is that TMs themselves, contrary to results in previous work [6,33], do affect the topology design, especially when designing a network from the ground up. The results suggest that the two conditions influence the appearance of such topologies are (1) a relatively high cost of building a PoP with several links, and (2) hierarchical structure of traffic in the TM, *i.e.,* higher intra-zone traffic than inter-zone traffic. However, though here we see a direct effect from HTMs on the topology, we stress that this relationship goes both ways: traffic is affected by topology, and in turn, when traffic demand grows, the network topology has to change to cope with a changing traffic load.

Our work can be applied to GÉANT, where its topology is found in the Internet Topology Zoo [15]. Code for the HTM models can be obtained from [31].

VI. Conclusion

In this paper, we proposed a set of axioms for TM synthesis models, with the goal of defining a useful class of TM models. Our contributions are two new model consistency axioms. We also propose the HTM, which incorporates the hierarchical structure of the network into the TM. We show that the HTM is the model that arises naturally from our simple set of axioms, and when combined with the principle of maximum entropy,

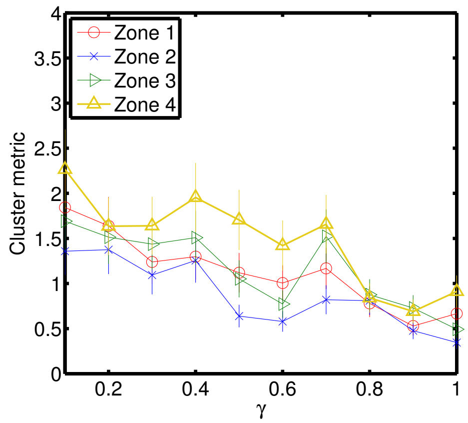


Fig. 5: *The clustering metric for each zone for the 100 topologies generated with varying γ from 0.1 to 1.0, for $\lambda = 10$. The vertical lines represent the 95% confidence interval. As γ increases, we see that the clustering metric for the HTMs decreases. In comparison, the average clustering ratio for the random gravity model is $\{1.62, 1.28, 1.68, 2.04\}$ for the 4 zones, showing a marked decrease in the clustering metric once $\gamma \geq 0.5$.*

results in a class of models that are useful in practice. The HTM was used as input into an application on PoP-level topology design to demonstrate that hierarchical structure in the TM can affect the network design by generating clusters or zones. Future work will focus on advancing the axiomatic framework of TM models, with the eventual goal of laying a strong foundation for future TM models.

References

[1] D. L. Alderson, H. Chang, M. Roughan, S. Uhlig, and W. Willinger. The many facets of Internet topology and traffic. *Networks and Heterogeneous Media*, 1(4):569–600, December 2006.

[2] ANSI/ISA-62443-1-1. Security for industrial automation and control systems part 1-1: Terminology, concepts, and models, 2007.

[3] D. Applegate and E. Cohen. Making intra-domain routing robust to changing and uncertain traffic demands: Understanding fundamental tradeoffs. In *ACM SIGCOMM*, pages 313–324, August 2003.

[4] T. Benson, A. Akella, and D. A. Maltz. Network traffic characteristics of data centers in the wild. In *IMC '10*, pages 267–280, November 2010.

[5] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, 2nd edition, 1986.

[6] R. Bowden, M. Roughan, and N. Bean. COLD: PoP-level topology synthesis. In *CoNEXT '14*, pages 173–184, December 2014.

[7] J. Cao, D. Davis, S. V. Wiel, and B. Yu. Time-varying network tomography: Router link data. *J. Am. Statist. Assoc.*, 95(452):1063–1075, December 2000.

[8] J. B. Collins and C. E. Woodcock. Combining geostatistical methods and hierarchical scene models for analysis of multiscale variation in spatial data. *Geographical Analysis*, 32:50–63, January 2000.

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 2nd edition, 2006.

[10] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, April 1998.

[11] V. Erramilli, M. Crovella, and N. Taft. An independent-connection model for traffic matrices. In *ACM IMC 2006*, pages 251–256, October 2006.

[12] GÉANT. GÉANT partners, July 2011.

[13] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, May 1957.

[14] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chalken. The nature of datacenter traffic: Measurements and analysis. In *IMC '09*, pages 202–208, November 2009.

[15] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan. The Internet Topology Zoo. *IEEE J. Select. Areas Commun.*, 29(9):1765–1775, October 2011.

[16] E. D. Kolaczyk and H. Huang. Multiscale statistical models for hierarchical spatial aggregation. *Geographical Analysis*, 33(2), April 2001.

[17] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. In *SIGCOMM 2004*, pages 219–230, September 2004.

[18] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the Internet's router-level topology. In *SIGCOMM '04*, pages 3–14, August 2004.

[19] A. Medina, C. Fraleigh, N. Taft, S. Bhattacharyya, and C. Diot. A taxonomy of IP traffic matrices. *Proc. SPIE*, 4868(200), July 2002.

[20] A. Medina, N. Taft, K. Salmatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. In *SIGCOMM 2002*, 2002.

[21] NLANR. Abilene Trace Data. http://pma.nlanr.net/Special/ipls3.html.

[22] U. of Adelaide. European NREN model, 2010.

[23] K. N. Oikonomou. Analytic forms for most likely matrices derived from incomplete information. *Int. J. Systems Science*, 43(3):443–458, September 2010.

[24] E. Parsonage, H. X. Nguyen, R. Bowden, S. Knight, N. Falkner, and M. Roughan. Generalized graph products for network design and analysis. In *Proceedings of the 19th IEEE International Conference on Network Protocols*, pages 79–88, October 2011.

[25] R. B. Potts and R. M. Oliver. *Flows in Transportation Networks*. Academic Press, 1972.

[26] M. Roughan. Simplifying the synthesis of Internet traffic matrices. *SIGCOMM Comput. Commun. Rev.*, 35(5):93–96, 2005.

[27] M. Roughan. Robust network planning. In C. R. Kalmanek, S. Misra, and R. Yang, editors, *The Guide to Reliable Internet Services and Applications*, chapter 5, pages 137–177. Springer, 2010.

[28] A. Soule, A. Nucci, R. Cruz, E. Leonardi, and N. Taft. How to identify and estimate the largest traffic matrix elements in a dynamic environment. *SIGMETRICS Perform. Eval. Rev.*, 32(1):73–84, June 2004.

[29] S. M. Stigler. *The history of statistics: The measurement of uncertainty before 1900*. Belknap Press of Harvard University Press, 1986.

[30] C. Tebaldi and M. West. Bayesian inference on network traffic using link count data. *J. Am. Statist. Assoc.*, 93(442), June 1998.

[31] P. Tune. Hierarchical traffic matrix synthesis code, 2016. https://github.com/ptuls/HTM.

[32] P. Tune and M. Roughan. Internet traffic matrices: A primer. In H. Haddadi and O. Bonaventure, editors, *Recent Advances in Networking, Vol. 1*. ACM SIGCOMM, August 2013.

[33] P. Tune and M. Roughan. Spatiotemporal traffic matrix synthesis. In *ACM SIGCOMM*, pages 579–592, August 2015.

[34] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon. Providing public intradomain traffic matrices to the research community. *Computer Communication Review*, 36(1):83–86, January 2006.

[35] Y. Vardi. Network Tomography: Estimating source-destination traffic intensities from link data. *J. Am. Statist. Assoc.*, 91:365–377, 1996.

[36] Y. Zhang, Z. Ge, S. Diggavi, Z. M. Mao, M. Roughan, V. Vaishampayan, W. Willinger, and Y. Zhang. Internet traffic and multiresolution analysis. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*, volume 4, pages 215–234. Institute of Mathematical Statistics, 2008.

[37] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale IP traffic matrices from link loads. In *ACM SIGMETRICS 2003*, pages 206–217, 2003.

[38] Y. Zhang, M. Roughan, C. Lund, and D. Donoho. An information-theoretic approach to traffic matrix estimation. In *ACM SIGCOMM 2003*, pages 301–312, 2003.

[39] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu. Spatio-Temporal compressive sensing and Internet traffic matrices. In *ACM SIGCOMM 2009*, pages 267–278, August 2009.

[40] R. Zhang-Shen and M. McKeown. Designing a predictable Internet backbone with Valiant load-balancing. In *Proceedings of the 13th International Conference on Quality of Service (IWQoS)*, pages 178–192, 2005.