

Efficient Monitoring of End-to-End Network Properties

David B. Chua, Eric D. Kolaczyk, Mark Crovella

Abstract

It is often desirable to monitor end-to-end properties, such as loss rates or packet delays, across an entire network. However, active end-to-end measurement in such settings does not scale well, and so complete network-wide measurement quickly becomes infeasible. More efficient measurement strategies are therefore needed.

Previous work, examining this problem from a linear algebraic perspective, has shown that for exact recovery of complete end-to-end network properties, the number of paths that need to be monitored can be reduced to approximately the number of links in the network. In this paper we ask whether measurement strategies of even greater efficiency are possible. We recast the problem as one of statistical prediction and show that end-to-end network properties may be accurately predicted in many cases using a significantly smaller set of carefully chosen paths than needed for exact recovery. We formulate a general framework for the prediction problem, propose a simple class of predictors for standard quantities of interest (e.g., averages, totals, differences), and show that linear algebraic methods of subset selection may be used to make effective choice of which paths to measure. We explore the accuracy of the resulting methods both analytically and numerically, in the context of real network topologies of varying size. The feasibility of our methods derives from the low effective rank of routing matrices as encountered in practice, which appears to be a new observation of interest in its own right. The resulting framework, which is quite general, appears to hold promise for studying and improving the efficiency of monitoring of end-to-end-network properties.

Index Terms

Network measurements; monitoring; routing matrices; sampling; statistics.

I. INTRODUCTION

IN many settings it can be useful to have a global view of aggregate conditions across an entire network. For example, network operators often monitor latency in their networks via ping measurements between routers. Larger scale measurement projects, such as the Internet End-to-end Performance Monitoring project (IEPM) attempt to estimate regional or network-wide averages of metrics including latency and packet loss [1]. Such network-wide metrics can be useful for gauging typical user experience over time (as in the IEPM project), monitoring network performance during network failures, choosing alternate paths in an overlay, or gauging the network's response to exceptional events such as Y2K or 9-11.

Such systems are generally organized around all-pairs network monitoring; each endpoint monitors the paths to all other endpoints. However this architecture cannot scale up to the level required for large-scale network or overlay monitoring, since the number of measurements made scales with the square of the number of network endpoints. Furthermore, this approach involves highly redundant measurements; many links in the network are repeatedly measured.

Recent work [2] has shown that it is possible to reduce the number of end-to-end measurements to approximately the number of "virtual links" (identifiable link subsets) – which typically grows much slower than the square of the number of nodes – and yet still recover the complete set of end-to-end path properties exactly. This result is based on a linear algebraic analysis of routing matrices. A routing matrix is a binary matrix that specifies which links appear in which end-to-end paths. Such a matrix G has size (# paths) \times (# links), and $G_{i,j} = 1$ if and only if link j is found along the route taken by path i . The results in [2] follow from the fact that the rank of G , which is generally equal to the number of independent paths in the network, tends to be

David Chua (dchua@math.bu.edu) and Eric D. Kolaczyk (kolaczyk@math.bu.edu) are with the Dept. of Mathematics and Statistics of Boston University. Mark Crovella (crovella@cs.bu.edu) is with the Computer Science Dept. of Boston University. This work was performed while M. Crovella was at Laboratoire d'Informatique de Paris 6 (LIP6), with support from Centre National de la Recherche Scientifique (CNRS), France and Sprint Labs. This work was supported in part by a grant from Sprint Labs, by NSF grants ANI-9986397 and CCR-0325701, and by ONR award N000140310043.

much smaller than the total number of paths. Since a maximal set of such independent paths can be used to reconstruct any other path in the network, it is sufficient to monitor only this set. A method for choosing such a set, based on a linear algebraic method of subset selection is given in [2].

In this paper we ask whether network path properties may be monitored *approximately*, as opposed to exactly, in an accurate fashion using a further-reduced set of end-to-end measurements. We are concerned in particular with approximate summaries of delays or packet losses along all paths in a network. These approximate network summaries may be used in a variety of ways. They can be used to form per-path averages so as to give an “average user experience” measure. Furthermore, they may be used as a lightweight monitoring system for detecting large changes in network conditions, which can then trigger more expensive, detailed network monitoring. Finally, such methods can be used to efficiently decide between sets of alternative paths to remote hosts, as is required for BGP-based route optimization [3], [4].

We show that it is indeed feasible to produce accurate summaries of network-wide path properties based on a highly reduced subset of end-to-end measurements – typically much smaller than the number of independent paths. Typical results show that it is possible to estimate metrics like network-wide averages with relative error of less than 10%, while only measuring about 20% - 30% of the paths needed by previous methods.

To accomplish this, we recast the problem as one of statistical prediction, based on sampling from a finite population. We formulate a general framework for this prediction problem and present a general class of linear predictors for common network quantities of interest, such as averages, totals, and differences.

Besides placing our work on a firm statistical footing, our use of linear model theory provides insights into algorithms that can assist in answering a key question: if we are only to measure a subset of paths in a network, which paths should we measure? We use the theory to motivate an algorithm for path selection and show that it performs quite well in practice.

The success of our approach hinges on the observation that, in practice, routing matrices have low *effective* rank. This means that all the paths in the network can be *approximately* constructed from a set of paths even smaller than the rank of G . Using routing information from a variety of real networks, we show that such routing matrices generally show effective rank much smaller than their actual rank. We explore the reasons for low effective rank, and argue that it is a remarkably common property in routing matrices. The implication is that certain links in the network are more important to measure than others when forming an approximate, network-wide view. This phenomenon is interesting in its own right, and has not been remarked upon before in the networking literature.

The remainder of the paper is organized as follows. In Section II, we provide background on the problem and explain the importance of a routing matrix’s rank. In Section III, we illustrate the phenomenon of low effective rank in routing matrices, and explore reasons why this is the case. In Section IV, we present our methodology for statistical prediction of aggregate network path properties. A detailed numerical evaluation of the methodology is given in Section V. Finally, in Section VI we conclude and discuss future work.

II. BACKGROUND

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a strongly connected network graph, where the nodes in \mathcal{V} represent network devices, and the edges in \mathcal{E} , links between those devices. Additionally, let \mathcal{P} be the set of all paths on the network. Let $n_v = |\mathcal{V}|$, $n_e = |\mathcal{E}|$, and $n_p = |\mathcal{P}|$ denote the number of devices, links, and paths, respectively.

Now consider a metric measured on the paths $i \in \mathcal{P}$ of the network whose value $y \in \mathbf{R}^{n_p}$ is linearly related to the value $x \in \mathbf{R}^{n_e}$ of that same metric over the links $j \in \mathcal{E}$. In particular, we are interested in the case where $n_p \gg n_e$ and the linear relation between y and x is given by $y = Gx$, where $G \in \{0, 1\}^{n_p \times n_e}$ is a routing matrix whose entries simply indicate the traversal of a given link by a given path i.e.,

$$G_{i,j} = \begin{cases} 1 & \text{if path } i \text{ traverses link } j \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

For example, letting α denote the packet loss probabilities on paths, and β , the corresponding probabilities on links, and assuming independence among loss-events on links, the relation between the path-wise and link-wise loss probabilities becomes $y = Gx$ on a logarithmic scale, where $y_i = \log(1 - \alpha_i)$ and $x_j = \log(1 - \beta_j)$.

We take the monitoring of aggregate conditions on a network to mean the knowledge of some summary statistic(s) of values in y , such as averages, medians, standard deviations, or quantiles. In this paper we will restrict our attention to summaries that take the form of linear combinations $l^T y$ of the elements of y , for some

$l \in \mathbf{R}^{n_p}$. Three cases are of particular interest. If $l_i \equiv 1/n_p$, then $l^T y$ is simply the network-wide average. Conversely, if $l_i = 1$ for some fixed i and zero otherwise, then $l^T y$ is the value of the metric on path i alone. Finally, if $l_i = 1/n_A$ for paths $i \in A$, and $l_i = -1/n_B$ for paths $i \in B$, for some subsets of paths A and B of size n_A and n_B , then $l^T y$ is just the difference of the averages on these subsets.

Of course, if one knows that one is interested in only a specific path or handful of paths, then the optimal strategy would be to simply monitor that path(s). Here we are interested in strategies that allow for the capacity for whole-network monitoring, such as in the form of a single network-wide average or averages over various and possibly changing subsets of the network. Our work is motivated by the simple observations that (i) exact knowledge of $l^T y$ in such a setting would seem to require measurement of the y_i on all n_p paths in \mathcal{P} , and (ii) such a requirement quickly can become prohibitive, since $n_p = O(n_e^2)$ and thus scales poorly with the size of the network. Therefore, we might hope instead to obtain sufficiently accurate approximate knowledge of $l^T y$ through measurements on only some comparatively small but representative subset of paths.

An initial solution to this problem is provided in [2], where it is noted that in order to obtain the values of all elements y_i of y , the measurement of all n_p paths is unnecessary: one need only measure an appropriate subset of $q = \text{Rank}(G) \leq n_e$ paths $i \in \mathcal{P}$. To see this, let \tilde{G} denote any matrix formed from rows of G that span the row space $\text{Row}(G)$, and let \tilde{y} denote the corresponding elements of y . Since $y = Gx$ we certainly have $\tilde{y} = \tilde{G}x$, and so for any \tilde{x} satisfying $\tilde{y} = \tilde{G}\tilde{x}$ we have $\tilde{G}(\tilde{x} - x) = 0$. Furthermore, since every row of G is a linear combination of rows of \tilde{G} we have that $G(\tilde{x} - x) = 0$. Thus any solution \tilde{x} of the reduced system will be a solution of the full system, which means that y may be obtained as $G\tilde{x} = y$. Finally, recall that the dimension of the row space of a matrix G is equal to its rank q , and hence q linearly independent rows of G are sufficient to span $\text{Row}(G)$. A more extensive development of this argument may be found in [2], and an algorithm based on the QR-decomposition is provided for selecting the rows that form \tilde{G} . An analogous presentation may be found in [5], in the context of Boolean algebras, for the problem of detecting link failures.

Through the above argument, the rank q of the routing matrix G is found to be an important quantity in regards to the sampling of paths for network monitoring. However, often it is the case that while a matrix technically has some rank q , it *effectively* has some lower rank q' , sometimes significantly lower. This means that some $q - q'$ rows (columns) of the matrix are nearly (i.e., within some range of numerical tolerance) linear combinations of the remaining q' . In such situations, numerical procedures typically can be developed around the use of just those q' rows (columns) that are essentially as accurate as those that use the full set of size q . With respect to the present paper, the relevance of this observation is the implication that networks with routing matrices G of effective rank $q' \ll q$ may potentially allow for efficient monitoring using a very small number of paths. We show in the remainder of this paper that this is indeed the case.

Note that in using less than q paths we are giving up the ability to recover y (and functions thereof) exactly. Viewed abstractly, without the surrounding networking context, the problem we are facing can be viewed profitably as one of statistical prediction. Specifically, we wish to predict a linear function $l^T y$ of the values y in a finite population, based on a sample, say y_k , of some subset of k of those values. From this perspective, a number of fundamental issues, regarding which paths to measure and how best to use the information in the measurements, can be given natural statistical interpretations in terms of statistical risk. We exploit this in developing our methodology, as will be seen in Section IV. Of course, in any statistical prediction problem, the degree to which useful predictions may be made relies heavily on the amount of predictive information there is in the measurements taken. In the context of our present problem, this issue essentially boils down to a question of how much sharing there is of links among paths in the network. And the latter can be determined through study of the routing matrix G .

III. ROUTING MATRICES AND EFFECTIVE RANK

As a motivating example, consider the Abilene high-performance network which serves Internet2 (the U.S. national research and education backbone). This network, depicted in Figure 1, consists of 11 nodes located across the continental United States. Note that there are only $15 \cdot 2 = 30$ directed links. With so few links from which to choose, many paths are forced to share. Intuitively, this suggests that many of the rows in G should be similar. That is, for many paths $i, i' \in \mathcal{P}$, the entries $G_{i,j}$ and $G_{i',j}$ should differ only for a relatively few links $j \in \mathcal{E}$. Therefore, the vectors $G_{i,\cdot}$ and $G_{i',\cdot}$ should be nearly dependent in many cases, which suggests that the effective rank of G for the Abilene network should be noticeably less than 30.

The singular value decomposition (SVD) is a standard device by which the discussion of effective rank is made more precise [6]. The SVD of a matrix $M \in \mathbf{R}^{m \times n}$ is a decomposition of the form $M = U\Delta V^T$



Fig. 1. Map of Abilene

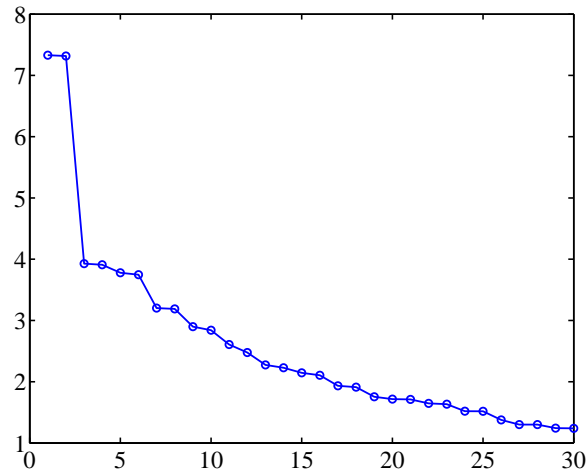


Fig. 2. Spectrum of the routing matrix G for Abilene.

where U and V are $m \times m$ and $n \times n$ orthogonal matrices, respectively, and Δ is an $m \times n$ diagonal matrix, whose diagonal entries $\delta_i \geq \delta_{i+1}$ are known as the singular values. A key property of this decomposition is the relation

$$Mv_i = \delta_i u_i \quad (2)$$

between the columns u_i and v_i of U and V , respectively. As a result, the columns of U and V provide a natural and useful basis when working with M . The singular values themselves give us insight into the structure of M . For example, the largest i for which $\delta_i > 0$ determines the rank of M , and the manner in which the δ_i decay is useful in gauging the effective rank. We note that the SVD of a matrix M is intimately related to the eigen-decomposition of MM^T . In particular, the square root of the eigenvalues of MM^T are in fact the singular values of M , and the collection of either is often called the spectrum of the corresponding matrix.

Consider Figure 2, in which the singular values for a typical routing matrix G from the Abilene network are shown. The sharp decay is evidence of a non-trivial amount of linear dependence (and hence similarity) among the rows of G . Note that there is a sharp drop between the second and third largest singular values that forms a ‘knee’ in the spectrum. The effective rank of a matrix is determined by looking for a large gap that separates the singular values into subsets of large and small singular values [6]. Therefore, the drop in Figure 2 suggests that as little as two paths may be enough to recover meaningful information about the rest of y .

The nature of the dependence among the paths in G is captured by the singular vectors v_i . Note that by (2), each path in G is a weighted sum of the v_i s. Since the v_i s are orthogonal, they capture independent patterns that arise among the paths in G , and the magnitude of contribution of each v_i to the set of all paths in G is measured by δ_i . Thus the v_i s corresponding to the largest δ_i s represent the ‘‘highways’’ that are most commonly used among the paths in G .

Figure 3 show the first four distinct v_i s for an Abilene routing matrix. These four paths capture roughly

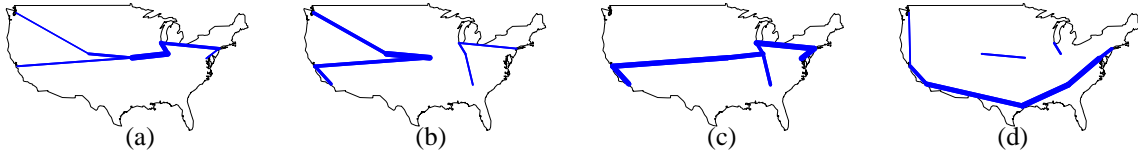


Fig. 3. First four distinct singular vectors of Abilene routing matrix

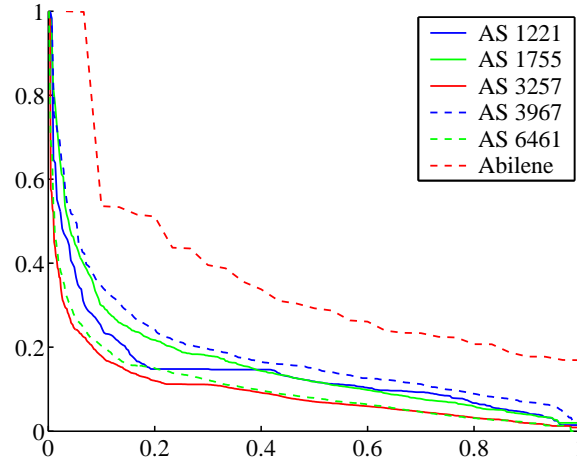


Fig. 4. Spectra of G for five networks measured by Rocketfuel. The Abilene spectrum from Figure 2 has been plotted for comparison. (Note: The spectrum for each network has been rescaled by the largest singular value δ_1 and the indices re-scaled to the unit interval.)

70% of the total energy in the routing matrix. In this figure we have plotted the link corresponding to each vector component whose squared value is greater than 0.01. The components plotted typically constitute 95% to 99% of the energy in the singular vector. Each link is drawn with a thickness that is roughly proportional to the magnitude of the component.

Figure 3(a) shows v_1 , which is the strongest pattern present in Abilene routes. The Indianapolis-Kansas City link has greatest contribution, indicating that this link is heavily shared among the paths in G , which befits its position as a central link in the network. It is also notable that this “highway” consists of a connected east-west path through the network. The other singular vectors are similarly interpretable: the next two (Figure 3(b) and (c)) capture modifications of this path, in conjunction with addition of the most immediately connected north-south links; and the fourth (Figure 3(d)) captures the southern east-west path through the network.

We have found that the dominance of a small number of heavily-used paths as exhibited by Abilene appears to be a common phenomenon in ISP networks. In fact, this effect can be even more pronounced in larger networks. To assess this effect in larger networks, we used data gathered by the Rocketfuel topology mapping engine [7], [8]. In these networks, nodes correspond to routers and links to inter-router connections. We looked at five different ISPs, ranging in size from 80 nodes and 147 links to 164 nodes and 328 links.

As with Abilene, we assess the effective dimension of these networks from their spectra. Figure 4 presents the spectra for these networks. For each network, a significant knee occurs by the 100th singular value, indicating an effective dimension much smaller than the rank of the routing matrix.

The reason for the remarkable consistency in these results across different networks is not obvious. It is clear that in general, networks will be engineered such that paths tend to share links, since link sharing amortizes investment and increases traffic aggregation (improving manageability). However, it appears that even in networks that are not explicitly engineered in this way, a high degree of path sharing can occur. We illustrate this point using synthetic routing matrices.

To show that low effective rank in routing matrices may be a remarkably common phenomenon, we illustrate its presence in an extreme form of topology bearing little resemblance to a network graph. We take the complete graph on 100 nodes i.e., K_{100} , equip the $\binom{100}{2}$ edges with weights that are drawn independently and identically from an exponential distribution with mean one, and route between all nodes using shortest paths.

This experiment was repeated 10 times and the resulting spectra are shown in Figure 5, along with the same ISP spectra shown in Figure 4. It is evident that the spectra of the random graphs are strikingly similar to those

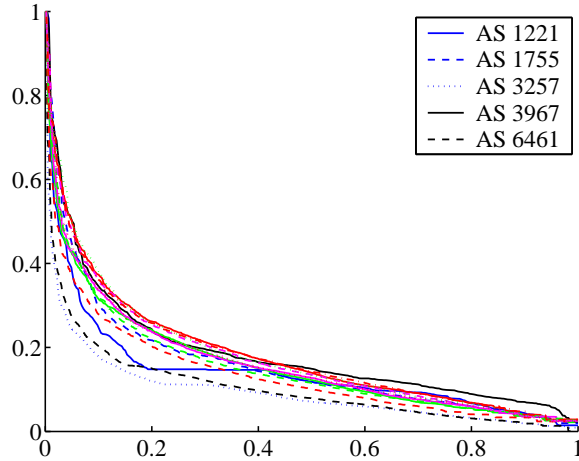


Fig. 5. Spectra for the routing matrices of ten K_{100} networks, based on shortest-path routing with independent and identically distributed exponential(1) edge-weights. Overlaid are the scaled spectra for the Rocketfuel datasets.

of the ISP networks, despite the lack of any overt engineering involved in the construction of the K_{100} graphs.

In summary, it is this decay in the spectra of routing matrices that makes the methods we develop in this paper efficient. It should be noted that, as far as the authors can tell, this general phenomenon does not seem to have been remarked upon in previous literature and in and of itself seems to warrant further study.

IV. PREDICTION OF END-TO-END NETWORK PROPERTIES

In this section we describe a statistical framework for prediction of end-to-end network properties. Our choice of predictors and their properties, given a set of measured paths, are described in Section IV-A. The problem of selecting which paths to measure is addressed in Section IV-B.

A. Statistical prediction from sampled paths

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a network graph, x a metric (such as a function of delay or packet loss) on the set \mathcal{E} of n_e links, G an $n_p \times n_e$ routing matrix, and $y = Gx$ the corresponding aggregation of the metric to the set \mathcal{P} of n_p paths, as defined in Section II. If we define μ and Σ to be the mean and covariance of x , respectively, we note that the corresponding values of y are simply $\nu = G\mu$ and $V = G\Sigma G^T$, respectively. The majority of the work that follows will be in terms of just the first two moments of x and y , which allows us to avoid having to make full distributional assumptions.

Now fix $k \leq \text{Rank}(G)$ and let y_k denote the values y_{i_1}, \dots, y_{i_k} of the metric for k paths $i_1, \dots, i_k \in \mathcal{P}$ that are to be sampled (i.e., measured). Similarly, let y_r denote the values for those $n_p - k$ paths that remain. If we let G_s be those rows of G corresponding to the k paths i_1, \dots, i_k , and G_r , the remaining rows, then we may rewrite G in the form

$$G = \begin{bmatrix} G_s \\ G_r \end{bmatrix}, \quad (3)$$

and we may write y as

$$y = \begin{bmatrix} y_s \\ y_r \end{bmatrix} = \begin{bmatrix} G_s x \\ G_r x \end{bmatrix}. \quad (4)$$

Furthermore, the mean and covariance of y may be written as

$$\nu = \begin{bmatrix} \nu_s \\ \nu_r \end{bmatrix} = \begin{bmatrix} G_s \mu \\ G_r \mu \end{bmatrix} \quad (5)$$

and

$$V = \begin{bmatrix} V_{ss} & V_{sr} \\ V_{rs} & V_{rr} \end{bmatrix} = \begin{bmatrix} G_s \Sigma G_s^T & G_s \Sigma G_r^T \\ G_r \Sigma G_s^T & G_r \Sigma G_r^T \end{bmatrix}. \quad (6)$$

Of special interest to us in what follows will be when $\Sigma = \sigma^2 I$, wherein the link measurements are uncorrelated and share a common variance $\sigma^2 > 0$. In this case, the expression for V in (6) simplifies to terms proportional to products of G_s and G_r .

Recall that our goal is to extract accurate, approximate knowledge of a given linear summary of global network path conditions i.e., $l^T y$, from the k sampled path values in y_s . This goal can be viewed as a particular instance of the classical problem of prediction in the statistical literature on sampling [9]. If the quality of a predictor, say $p(y_s)$ is measured by its mean squared prediction error (MSPE) i.e., $E[(l^T y - p(y_s))^2]$, then it is known that the best predictor is given by the posterior mean

$$E[l^T y | y_s] = l_s^T y_s + E[l_r^T y_r | y_s] , \quad (7)$$

where l_s and l_r are defined analogous to the other quantities above.

Note that the optimal predictor in (7) is simply the sum of the known component $l_s^T y_s$ and the optimal predictor of the remaining component $l_r^T y_r$. Unfortunately, the specific form of the latter will depend on the joint distribution of the elements of y , the determination of which often may not be evident in practice. In such cases, it is common to restrict attention to some smaller, simpler sub-class of possible predictors, with the class of linear predictors of the form $b^T y_s$ being a canonical choice. In that case, the best *linear* predictor (BLP) of $l^T y$ can be shown to be of the form

$$a^T y_s = l_s^T y_s + l_r^T G_r \mu + l_r^T c^* (y_s - G_s \mu) , \quad (8)$$

where c^* is any solution of $V_{sr} = V_{ss} c^{*T}$. The derivation of this result follows similarly to the analogous result for simple linear statistical models in [10]. If V_{ss} is invertible, it is natural to set $c^* = V_{rs} V_{ss}^{-1}$; otherwise, c^* may be defined through a generalized inverse. We note that whether or not V_{ss} is invertible, which without loss of generality we will assume throughout most of what follows, depends on both G and Σ .

Unfortunately, the BLP in (8) depends on μ , which is generally unknown in practice; therefore, the BLP is an ideal that cannot be computed. However, since all of the elements of y involve at least some fraction of the elements in x , it may be possible to obtain a useful estimate of μ from the measured values y_s . Statistical theory suggests solving the corresponding generalized least squares problem

$$\min_{\mu} (y_s - G_s \mu)^T V_{ss}^{-1} (y_s - G_s \mu) \quad (9)$$

to obtain the estimate

$$\hat{\mu} = [G_s^T V_{ss}^{-1} G_s]^{-1} G_s^T V_{ss}^{-1} y_s . \quad (10)$$

Here M^{-} denotes a generalized inverse of a matrix M . The choice of which generalized inverse to use is essentially a choice of what values to assign to the unidentifiable values in x . In later sections we will use the Moore-Penrose inverse which has the effect of assuming links not part of a sampled path to have value zero, although this is not the only choice possible.

The substitution of $\hat{\mu}$ for μ in (8) produces an estimate of the BLP, say $\hat{a}^T y_s$, that is a function of only the measurements y_s , the routing matrix G , and the link covariance matrix Σ . Under certain conditions, estimated BLPs (E-BLPs) produced in this manner will be unbiased i.e., their expected values will match that of $E[l^T y]$, and they will in fact be the best among all linear, unbiased predictors, in the sense of having minimum MSPE within that class [10]. Of course, it is easy to see that in the current setting these properties will not hold: the prediction of y_i for a path i running entirely across links not traversed by any of the measured paths i_1, \dots, i_k can clearly be biased. Nevertheless, it is not unreasonable to expect that $\hat{a}^T y_s$ may still be a useful predictor, and the numerical results in Section V bear out this expectation.

For our numerical results, and indeed for most of the rest of this paper, we focus on the case where $\Sigma \propto I$. It is instructive to first consider the BLP under this condition. Note that the covariance among measured paths has the form $V_{ss} \propto G_s G_s^T$, and therefore is invertible if and only if these paths correspond to a set of independent rows in G . Assuming this to be the case, the MSPE for the BLP is given by¹

$$\text{MSPE}(a^T y_s) = \sigma^2 l_r^T G_r (I - B_s) G_r^T l_r , \quad (11)$$

¹Formulas for MSPEs for the case of general link covariance Σ are straightforward generalizations of those that we provide here and below, but are less interpretable and instructive, and therefore are omitted for reasons of space.

where $B_s \equiv G_s^T(G_s G_s^T)^{-1}G_s$. We note that B_s is just an orthogonal projection matrix, representing projection onto the rows of G_s , and is thus symmetric and idempotent. Therefore, the MSPE in (11) is just proportional to the square of the (Euclidean) norm of the residual of $G_r^T l_r$ after projection onto the rows of G_s . For example, if l_r consists of only a single one, for some path i not in the sample, and zeros otherwise, then this MSPE simply measures the degree to which the row in G for path i lies in the span of the rows in G_s .

Now consider the construction of the E-BLP. With $\Sigma \propto I$, the generalized least squares estimator $\hat{\mu}$ in (10) reduces to

$$\hat{\mu} = [G_s^T(G_s G_s^T)^{-1}G_s]^{-1}G_s^T(G_s G_s^T)^{-1}y_s. \quad (12)$$

This expression may be further reduced in an informative manner. In particular, since $y_s = G_s x$, we find that $\hat{\mu} = B_s^- B_s x$. Furthermore, if we choose to define the generalized inverse B_s^- through the Moore-Penrose inverse, which we denote by M^+ for an arbitrary matrix M , then properties of the latter lead to the further reduction $\hat{\mu} = B_s x$. Finally, noting that the orthogonal projection of x onto $\text{Row}(G_s)$ may also be expressed as $G_s^+ G_s x$, we have that $\hat{\mu} = G_s^+ y_s$.

Therefore, the E-BLP when $\Sigma \propto I$ can be expressed as

$$\hat{a}^T y_s = l_s^T y_s + l_r^T G_r G_s^+ y_s \quad (13)$$

or, more compactly, as

$$\hat{a}^T y_s = l^T G G_s^+ y_s. \quad (14)$$

In other words, the E-BLP in this setting can be seen to be the result of three steps: (i) estimating x by its Moore-Penrose generalized inverse $\hat{x} = G_s^+ y_s$, (ii) mapping that estimate to a vector of predicted path values $\hat{y} = G \hat{x}$, and (iii) predicting $l^T y$ by $l^T \hat{y}$. It should be noted, of course, that in the case of more general Σ such a simplification does not continue to hold, although the relevant formulas are straightforward to produce.

Seen in this light, the first step in our method involves inference of network-internal conditions from end-to-end measurements. This problem, known as network tomography [11] [12], has been well studied and a range of methods have been proposed for its solution. Our work differs in that we motivate the particular solution we adopt starting from the theory of sampling with linear models, and show how the tomography problem fits into a sampling setting. Furthermore, we show that using linear model theory drives us to make use of the structure of the routing matrix to make informed choices about which paths to sample (as described in the next section). Thus, although the use of the generalized inverse is not as sophisticated as some methods used for network tomography, it focuses attention on the information needed for good path selection. We also note that more sophisticated methods for solving the network tomography step can certainly be used in conjunction with our path-selection algorithms.

The MSPE for the E-BLP in (14) can be shown to have the form

$$\begin{aligned} \text{MSPE}(\hat{a}^T y_s) &= E[(\hat{a}^T y_s - l^T y)^2] \\ &= \text{MSPE}(a^T y_s) + \text{bias}^2(\hat{a}^T y_s), \end{aligned} \quad (15)$$

where

$$\text{bias}(\hat{a}^T y_s) = -[(I - B_s)G_r^T l_r]^T \mu. \quad (16)$$

In other words, the MSPE for the E-BLP can be expressed as a sum of the MSPE for the BLP plus an additional bias term. Note that the bias term consists of two parts: the link mean μ and a multiplying matrix $(I - B_s)G_r^T l_r$. The latter consists of the components of unmeasured paths relevant to l_r that lie orthogonal to $\text{Row}(G_s)$. So we can see from examination of (11) and (16), precisely how the choice of which rows are included in G_s , and their relationship to the un-included rows in G_r , plays a fundamental role in determining the accuracy of our predictor $\hat{a}^T y_s$. Therefore, we next turn our attention to the matter of selecting these rows.

B. Path selection

Suppose we have resources to measure a total of k paths in a network. Which k paths should be measured? In light of the statistical framework we have just derived, a natural response is that measurements should be taken on those k paths in \mathcal{P} that collectively minimize the expression for $\text{MSPE}(\hat{a}^T y_s)$ in (15), among all subsets

of k paths. However, this expression depends on the unknown values μ and σ^2 , and so this minimization is not generally feasible in practice. Therefore, we choose instead to minimize the expression for $\text{MSPE}(a^T y_s)$ in (11), which allows us to avoid this problem. This approach is intuitively appealing, given that the two predictors are equivalent up to knowledge of μ . Furthermore, the results of the numerical experiment described in Section V-C confirm its robustness across various changes in parameterization.

Now consider the relevance of the number of paths k we wish to measure. For $k = \text{Rank}(G)$, any set of paths corresponding to k linearly independent rows of G will suffice, since these rows will span $\text{Row}(G)$ and the entire vector of path values y may therefore be recovered. This observation forms the basis of the work in [2], as explained in Section II. Note that (11) will be zero in this case a priori. Motivated by the results in Section III, our interest in this paper is in the case where fewer than $\text{Rank}(G)$ paths are to be measured—possibly much fewer—and therefore not all of $\text{Row}(G)$ will be spanned. In that case (11) will be nonzero, and it may be seen that in seeking to minimize this quantity we are faced with the task of finding a subset of k paths such that the corresponding rows span as much of $\text{Row}(G)$ as possible.

From an algorithmic perspective, this task is equivalent to the problem of subset selection in the field of computational linear algebra. See [6, Ch 12], for example. This problem is NP-complete and so exact, efficient algorithms for its solution do not exist. However, a variety of methods have been proposed for finding an approximate solution. Given a matrix M and choice of integer k , these methods generally seek to find a subset of k linearly independent rows (or columns) of M that accurately approximate its first k singular dimensions and form a well-conditioned submatrix. For calculating the numerical results in Section V, we adapted the method described in Algorithm 12.2.1 of [6, p. 574], which rests on the use of a particular QR factorization² with column pivoting. In particular, to choose an efficient set of k paths to measure, we first compute $U(:, 1:k)^T P_{(k)} = QR$, where U derives from the SVD $G = U\Delta V^T$ and $P_{(k)}$ is an $n_p \times n_p$ permutation matrix defined by the column pivoting. We then take G_s to be the submatrix formed by the first k rows of $P_{(k)}^T G$. We refer the reader to [6] for a more detailed motivation for and description of the general algorithm.

Note that $P_{(k)}$ will depend on k , as indicated by our notation, and therefore must be computed separately for each choice of k . It therefore follows that the subsets of rows selected from G for successively larger choices of k will not necessarily be nested. In some settings it might be desirable to have a nested sequence of subsets across k , such as when adaptively adding or removing paths in an online manner. In that case, one solution to this problem is to first compute $P_{(k_{max})}$, where k_{max} is the largest number of paths of interest, and then to successively form G_s from the first k rows of $P_{(k_{max})}^T G$ for all relevant $k \leq k_{max}$. The comparative performance of the nested and non-nested approaches is among the various issues we explore in Section V.

In summary, for the particular predictor in (14), the algebraic method of subset selection should provide an effective means for selecting a set of k paths, for given choice of k , that approximately minimizes the MSPE in (11) among all such sets. Overall, the complexity of the calculations behind this E-BLP under this approach is $O(n_p^2 n_e)$. This figure is dominated by the calculation of the SVD of G , since the QR factorization with column pivoting underlying our selection of paths is of $O(k^2 n_p)$ complexity, and the computation of G_s^+ is only $O(n_e^2 k)$.

V. NUMERICAL RESULTS

We conducted a series of numerical studies in order to obtain a preliminary assessment of the merits of our proposed framework for efficient monitoring of end-to-end network properties. The results of these studies are presented in this section, where we summarize a combination of analytic and simulated quantities calculated in the context of the Abilene network and the Rocketfuel topologies discussed in Section III. We explore both prediction of the network-wide average, as a case of canonical interest, and prediction of individual path-wise metrics, as an arguably extreme case.

A. Efficient monitoring of a network average

The results in Section III suggest that measurements on a relatively small number of paths should be sufficient for summarizing network-wide end-to-end properties with some accuracy. The linear prediction framework described in Section IV-A yields a natural class of predictors for general network summaries of the form

²We note that [2] also propose the use of an algorithm based on a QR factorization. However, the details of their algorithm differ from that used here, essentially because they need not be concerned with the condition of G_s , since their G_s will be of full rank i.e., $k = \text{Rank}(G)$.

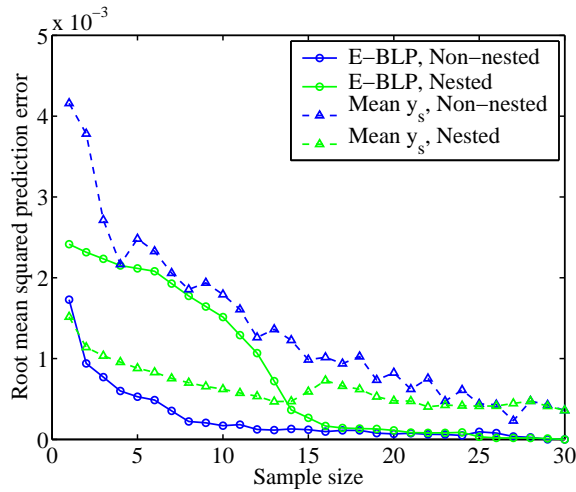


Fig. 6. Root mean squared prediction error for the E-BLP $\hat{a}^T y_s$ and the sample mean \bar{y}_s using non-nested and nested row selection methods.

$l^T y$, and the quality of these predictors, in the special case of $\Sigma \propto I$ upon which we focus, is found to be directly related to the choice of paths measured. The method of subset selection we propose to use for choosing which paths to measure, as sketched in Section IV-B, incorporates elements of the same SVD of the routing matrix G that underlies our results in Section III. Putting these three observations together, it is reasonable to expect that, for predictions based on paths chosen in this manner, the MSPE should decay fairly quickly as a function of the number of paths chosen. In order to test this expectation, we study the quality of the predictor $\hat{a}^T y_s$ through its statistical risk $\text{MSPE}(\hat{a}^T y_s)$, as a function of k , in the case where $l^T y$ is the network-wide average i.e., $l^T = (1/n_p, \dots, 1/n_p)$.

We begin with the Abilene network, described in Section III, for which calculations of the MSPE in (15) are shown in Figure 6. Here and throughout the rest of this section, the MSPE is shown on a square-root scale (i.e., RMSPE), so that it may be interpreted in the same units as the measurements themselves. As mentioned in Section II, the basic equations $y = Gx$ describe the relationship between link-based and path-based measurements under a variety of metrics, including packet loss. In the Abilene network typical link loss rates have been observed to be around $\beta_j \approx 0.1\%$ [1], which corresponds to values of $x_j = \log(1 - \beta_j)$ around $\log(1 - 0.001) \approx -0.001$. Therefore, we set the expected link values μ_j identically equal to -0.001 for all links j . In addition, we chose to set the common variance σ^2 at $(0.001)^2$, so that the standard deviation σ is proportional to the mean.

Results in Figure 6 show the RMSPE of the E-BLP for both the nested and non-nested methods of subset selection mentioned in Section IV-B. The former method was implemented with $k_{max} = 30$, which is the rank of the Abilene routing matrix used. Not surprisingly, results for the nested method are worse than those for the non-nested method, although interestingly, the methods become quite comparable by about $k = 15$. The RMSPE for the E-BLP in the non-nested case is seen to decay quickly until $k = 10$, at which point it is roughly of the same order of magnitude as the individual link means μ_j .

By way of comparison, we also show the performance of the sample mean \bar{y}_s , a simple alternative linear estimator of the network-wide average. Two points are worth noting. First, under both methods of subset selection, the sample mean never comes close to the level of accuracy achieved by the E-BLP under either method by $k = 15$. That this should be the case is not surprising, since these methods are geared towards minimizing the MSPE of the E-BLP, and not that of the sample mean. Second, by definition the sample mean cannot be expected to have zero RMSPE by the time $k = \text{Rank}(G)$, as in the case of the E-BLP, which explains its elevated behavior for large k . Based on the results of Figure 6, we use only the E-BLP with non-nested path selection for the remainder of this section.

In Figure 7 we show the same RMSPE calculations, with the same settings for μ and σ^2 , for the three smallest³ of the Rocketfuel topologies encountered in Section III. To facilitate comparison across networks,

³Calculations for the other two topologies presented memory issues for the machines used for this paper.

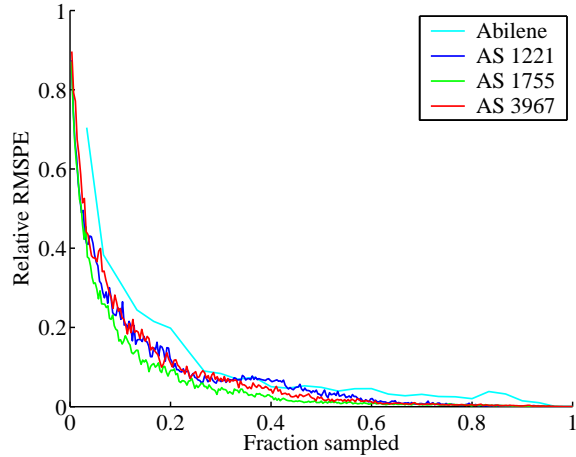


Fig. 7. Relative RMSPE = $\text{RMSPE}/E[l^T y]$ as a function of Fraction sampled = Sample size/Rank(G), where Rank(G) is the minimum number of paths needed for $\hat{a}^T y_s$ to predict $l^T y$ without error.

TABLE I

COMPARISON OF NETWORKS IN FIGURE 7, SHOWING NUMBER OF PATHS BY WHICH RMSPE IS LESS THAN 10% OF $E[l^T y]$.

Network	n_p	Rank(G)	RMSPE < 10%
Abilene	110	30	8
AS 1221	10716	306	65
AS 1755	7482	318	61
AS 3967	6162	280	66

the RMSPE was scaled by $E[l^T y]$ on the y-axis and the x-axis was rescaled to values on the unit interval, thereby adjusting both axes for differences in network size. The results for Abilene are shown as well. The similarity across networks is somewhat remarkable, with the relative RMSPE for all but the smaller Abilene network at roughly 10% using only roughly 20% of the paths (Abilene is seen to achieve this level just beyond 25%). Table I provides an alternative representation of this point, in terms of actual numbers of paths. Together, these results provide strong evidence for our central point, namely that network-wide averages of end-to-end properties may be predicted with high accuracy from measurements on a relatively small fraction of carefully chosen paths.

B. Observed prediction error: a simulation study

The RMSPE is a measure of statistical risk and, as such, is simply a single analytical summary of expected behavior of our predictor over all possible network conditions specified in connection with a given choice of μ and σ^2 . To gain insight into the variation in this behavior over individual instances of network conditions, we conducted a simulation study. A vector of link values x was simulated 1000 times under each of two models, the path values $y = Gx$ were computed, and the E-BLP (with non-nested subset selection) was calculated each time. Each x_j was drawn independently and identically distributed as $N(-0.001, 0.001)$ in the first model and as $-\text{Exp}(0.001)$ in the second model. These two distributions share the same mean and variance, so the MSPE is the same in either case. But their higher order moments (and hence quantities like skewness and kurtosis) are markedly different.

The results for the normal distribution are shown in Figure 8, while those for the exponential distribution are shown in Figure 9. In each case boxplots are used to summarize the prediction error $\hat{a}^T y_s - l^T y$ over all 1000 trials of the simulation, for each k , where again $l^T = (1/n_p, \dots, 1/n_p)$ corresponding to the network-wide average. The results shown are for the Abilene network; similar results were observed for the Rocketfuel topologies but have been left out due to considerations of space. The boxplots for the two simulations are quite similar, with the prediction errors for the exponential distribution showing a slightly stronger skewness and heavier tails. Their means mimic the values of the RMSPE in Figure 6, and their variation decreases

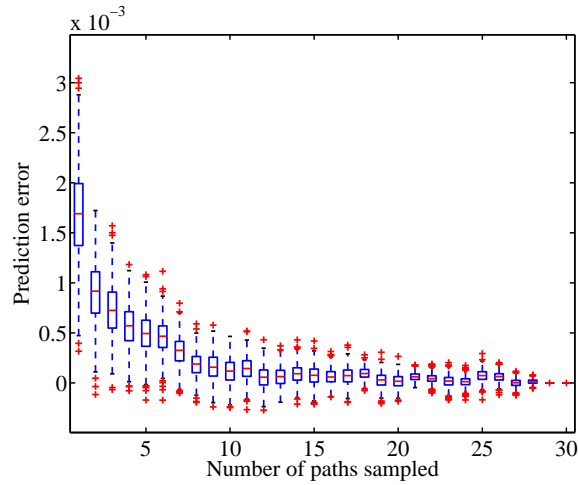


Fig. 8. E-BLP for simulated $x_i \sim N(-0.001, 0.001)$. Total of 1000 runs using the Abilene routing matrix. $E[l^T y] = -2.45 \times 10^{-3}$.

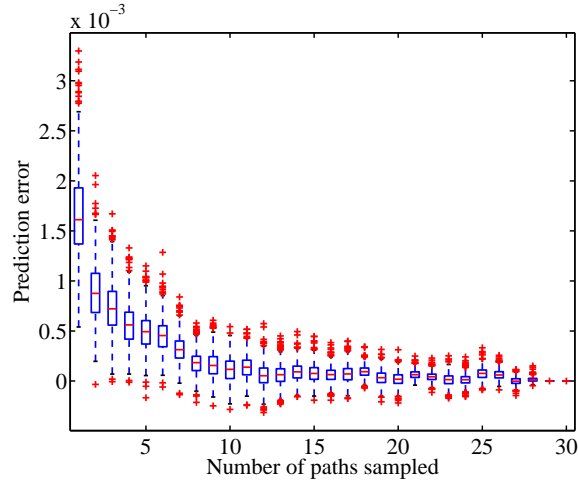


Fig. 9. E-BLP for simulated $x_i \sim -\text{Exp}(0.001)$. Total of 1000 runs using the Abilene routing matrix. $E[l^T y] = -2.45 \times 10^{-3}$.

with increasing k . Overall, they suggest that, for monitoring the network-wide average, (i) the RMSPE is a reasonably accurate summary of the sample-to-sample behavior of our predictor, and (ii) that behavior is fairly robust to the underlying distribution.

C. Robustness of prediction to variation across links

We have seen in the previous two numerical studies that it is possible to obtain quite accurate predictions of network-wide behavior based on only a modest number of measured paths. Yet clearly any predictor using only the information from measured paths i.e., y_s , can be made to do arbitrarily poorly on unmeasured paths i.e., y_r , if the latter are composed of sufficiently many links unseen by any path in the former and the values x_j on those unseen links are sufficiently extreme. This can be seen clearly in the bias term (16) of the MSPE for the E-BLP (14). Nevertheless, perhaps surprisingly, we have found that prediction of the network-wide average appears to be fairly robust to such imbalances.

Specifically, we conducted the following study with the Abilene network. For each of $k = 5, 10, 15, 20$ and 25, the set of paths selected by our algorithm was used to partition the links in the network into two sets: those corresponding to at least one measured path and those corresponding to none. For the links included in the measured paths, we set $\mu_j = -0.001$, as in our previous studies. We then varied the value of μ_j for the other links across four settings: $-0.1, -0.01, -0.001$, and -0.0001 . These values cover a fairly substantial range of loss rates observed in real networks. For computational convenience, we kept the variance σ^2 fixed at

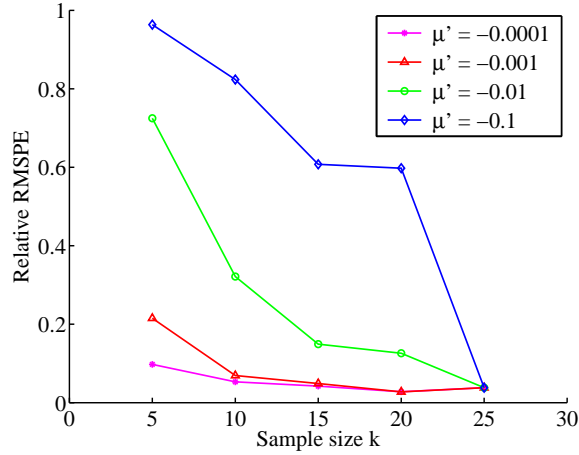


Fig. 10. Relative RMSPE = $\text{RMSPE}/E[l^T y]$ for the E-BLP with nonuniform link means μ_j (i.e., $\mu_j = -0.001$ on measured paths and $\mu_j \equiv \mu'$ on unmeasured paths).

$(0.001)^2$ across all links. The RMSPE derived from (15) was then computed for each of these four scenarios, across the five values of k .

The results are shown in Figure 10, where the RMSPE has been standardized by $E[l^T y]$ in each case. The relative RMSPE is seen to never rise above 1 in any of the cases considered. The pattern observed is as would be expected, where larger imbalances and smaller values of k lead to larger relative error. For all but the case where the mean for links on unmeasured paths is -0.1 , the relative error has decayed to less than 20% by $k = 15$. Closer examination reveals that the vast majority of the links in the Abilene network are in fact contained in at least one measured path from $k = 15$ onwards, which would seem to offer some confirmation of the sensibility of the choice of paths made by our algorithm, despite the fact that it is aimed at minimizing the MSPE in (11) – not (15) – and therefore operates without knowledge of μ .

D. Prediction on individual paths

All of our studies so far have assumed that it is the network-wide average that is being monitored, as a canonical quantity of interest. In reality, interest might instead focus on averages or totals (which are simply rescaled averages) on sub-networks (e.g., overlays), or even differences between sub-network averages or totals, any of which can be represented in the form $l^T y$ and hence dealt with in our framework. As an extreme situation, we consider the problem of predicting the measurement y_i for a single path $i \in \mathcal{P}$, in which case $l^T = (0, \dots, 0, 1, 0, \dots, 0)$ where the 1 is in the i -th component of l .

Of course, if one was only interested in the measurement for a single path, then the most direct and accurate approach would be to measure that path directly. However, measurement might not be feasible for a given path, or even if so, it will not be feasible for all paths simultaneously in networks of realistic size. Furthermore, the single path values y_i are the building blocks of any other linear function $l^T y$, and as such their study helps to further our understanding of the extent to which prediction may be done accurately.

In Figure 11 is shown the relative RMSPE of the E-BLP in (14), using non-nested selection of paths, for predicting the path values y_i individually for each path. Each RMSPE has been standardized by the quantity $E[l^T y]$ again, which in this case is just the sum of the means μ_j over the links j in path i . The calculations were done for the Abilene network, using the same parameter settings as in our first study (i.e., $\mu_i \equiv -0.001$ and $\sigma^2 = (0.001)^2$). Boxplots have been used to summarize the results across all unmeasured paths, for each k ; measured paths are ‘predicted’ with zero error and are not shown. Super-imposed on this plot, for the purpose of comparison, is the relative RMSPE for predicting the network-wide average (i.e., the lowest curve in Figure 6). That the boxplots sit above this curve is to be expected, given that the MSPE for the network-wide average is itself an average of errors that include the zero prediction error for measured paths. While the performance clearly varies among paths and improves with increasing k , the overall plot communicates a strong message in showing that the relative error – even when predicting metrics for individual paths – is generally well below 1.

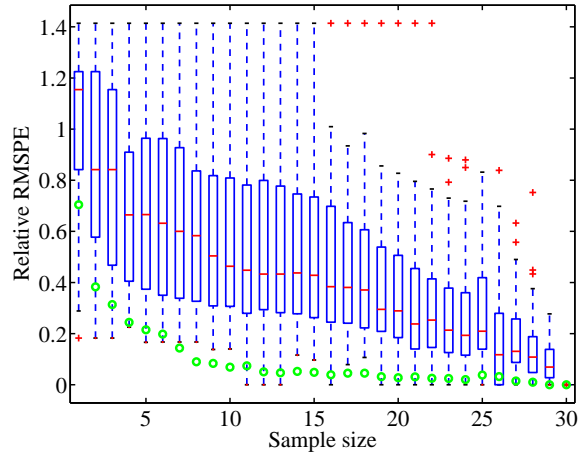


Fig. 11. Boxplots of the relative RMSPE for single path prediction using $\mu_i = -0.001$ and $\sigma^2 = (0.001)^2$. Each box excludes paths that have been sampled, and thus have zero prediction error. The relative RMSPE for the network-wide average has been overlaid for comparison (purple circles).

VI. CONCLUSION

In this paper we have argued that it should be possible to monitor end-to-end network properties with reasonable accuracy, based on sampling a relatively small number of paths. Our experimental results confirm this assertion, showing that we can often obtain estimates with relative error of less than 10% when sampling only 20-30% of the minimum number of paths required for zero error. Our approach holds promise for more efficient network monitoring with a number of specific applications, including forming estimates of network-wide conditions; lightweight monitoring; and comparison of alternate paths to a set of destinations.

Our methods are based on the statistical theory of linear prediction, from which we are able to formulate a general framework for the problem of monitoring a certain class of end-to-end network properties. Having done so, we are able to motivate the use of algorithms from computational linear algebra for the proper choice of paths to be monitored. More generally, however, we believe that a fundamental contribution of this paper is the observation and illustration of the fact that end-to-end network monitoring can be profitably posed as a problem of statistical prediction (i.e., of end-to-end properties) in a finite population (i.e., of paths). As there exists a rather sizable literature on this problem, in various degrees of generality, it is clear that the basic statistical paradigm adopted here can be extended in many ways.

For example, as mentioned in Section IV, although we present results primarily for the case in which the link metrics are uncorrelated and share a common variance, all of the relevant analytical expressions extend without difficulty to the case of arbitrary link covariance. However, the connections with subset selection algorithms will be less obvious. More generally, a fuller treatment using full distributional properties of the link or path metrics (as opposed to just the first two moments through μ and Σ), if it is felt that these may be specified with some confidence, would in principle allow for prediction using the optimal posterior solution in (7), although likely at the cost of an increased computational burden. Lastly, for network summaries that do not take a linear form i.e., such as standard deviations and quantiles, extensions beyond the linear prediction framework are possible.

Critical to the feasibility of our framework is the additional observation that routing matrices generally exhibit the phenomenon of low effective rank. This observation does not appear to have been made in the literature to date. We have explored this phenomenon in both real and simulated networks, and shown that it is remarkably prevalent. In the particular case of the Abilene network, we show how low effective rank translates into well traveled “highways” that are shared by many end-to-end paths.

The results in this paper are suggestive of a number of valuable next steps. From a standpoint of basic investigation, the reasons behind low effective rank of routing matrices are worth exploring; an understanding of why this effect occurs and what affects the degree to which it holds would help inform future work in this area. We are currently exploring this issue. We also intend to study the scaling properties of our methods on larger ISP networks. Although our method performs quite well on the moderate-sized Abilene network, our experimental results suggest that it is possible that relative savings in measurement cost may be even greater

on medium to large scale networks. Actual online implementation of our proposed framework will undoubtedly raise additional interesting issues, such as how best to deal with changes in routing, the failure of links, and networks with load-balanced paths.

REFERENCES

- [1] Les Cottrell, "ICFA SCIC network monitoring report," 2004.
- [2] Yan Chen, David Bindel, and Randy H. Katz, "Tomography-based overlay network monitoring," in *Proceedings of the 2003 ACM SIGCOMM conference on Internet measurement*. 2003, pp. 216–231, ACM Press.
- [3] Geoffrey Brown, "Internet address clustering for intelligent route control," Submitted for publication. Available at <http://www.cs.indiana.edu/~geobrown/>.
- [4] Internap, Inc., "http://www.internap.com."
- [5] Hung X. Nguyen and Patric Thiran, "Active measurement for multiple link failures: Diagnosis in IP networks," in *Passive and Active Measurements Workshop*, 2004.
- [6] G. H. Golub and C. van Loan, *Matrix Computations*, The Johns Hopkins University Press, London, 2 edition, 1989.
- [7] Neil Spring, Ratul Mahajan, and David Wetherall, "Measuring ISP topologies with Rocketfuel," in *Proceedings of ACM SIGCOMM*, 2002.
- [8] Ratul Mahajan, Neil Spring, David Wetherall, and Tom Anderson, "Inferring link weights using end-to-end measurements," in *Proceedings the ACM SIGCOMM Internet Measurement Workshop*, 2002, pp. 231–236.
- [9] Richard Valian, Alan H. Dorfman, and Ricard M. Royall, *Finite Population Sampling and Inference: A prediction approach*, Wiley Interscience, 2000.
- [10] Ronald Christensen, *Plane answers to complex questions*, Springer Texts in Statistics. Springer-Verlag, New York, 1987.
- [11] R. Caceres, N.G. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Transactions on Information Theory*, pp. 2462–2480, 1999.
- [12] M. Coates, A. Hero, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Processing Magazine*, May 2002.