Modeling Variation in Mobile Download Speed in presence of Missing Samples

Mah-Rukh Fida*‡, Marie Roald*†, Evrim Acar* and Ahmed Elmokashfi*
*SimulaMet, Center for Digital Engineering, Oslo, Norway
†Oslo Metropolitan University, Oslo, Norway
‡ School of Computing and Engineering, University of Gloucestershire, United Kingdom

Abstract—A stably fast mobile broadband connectivity is key to customer retention. Mobile networks, however, suffer unpredictability in performance. Analyzing variability in network speed is, therefore, challenging since it tends to exhibit patterns at several time scales. Additionally, frequently monitoring it over time, is costly. In this paper, we analyze speed measurements from 78 stationary probes, spread across Norway. Monitoring was performed thrice per day across the year, to assess performance of the two largest network operators. Despite being unique, the dataset involves a non-trivial extent of missing data. This study investigates the effect of missing data on the extracted performance patterns. We capture patterns with tensor factorizations, that show that missing data at random has a minimal effect on the identified patterns, and that depending upon the determinism of an operator's performance, the acceptable size and structure of missing data varies. Our analysis shows that, for a probe, the difference in speed variation between real and imputed speed values can be around 7% for up to 40% missing data. We also identify that congestion, routine maintenance and sub-optimal network configuration cause high speed variability. These findings can help operators improving their offerings and deciding on optimal performance monitoring frequency.

Index Terms—Download speed, speed variation, tensor factorizations, imputation, missing data.

1 Introduction

Every new generation of mobile cellular technology has chiefly focused on enhancing the data transfer speed [1]. So far, speed and coverage have been the most important factors that both consumers and policy makers consider when deciding between service providers and when assessing the state of the service in general. Consumers are not only interested in the achievable maximum speed but also in how stable the speed is. Networks with fluctuating performance can quickly lose customers. Unfortunately, the broadband speed of a mobile network is influenced by a multitude of factors which makes ensuring speed stability an uneasy endeavour. These include coverage, number of users, and network configuration, among others [10], [36], [39], [43], [48]. Hence, understanding and characterizing variability in mobile broadband speed is paramount for customers, network operators and regulators [25], [42]. In fact, this has recently become highly relevant as mobile operators in different countries have started offering fixed wireless access as a key technology to bridge the broadband gap [3]. In this paper, we analyze a year-long download speed time series collected by 78 stationary probe nodes. These probe nodes are spread in urban settings across Norway and use commercial subscriptions to connect to the two biggest mobile operators in the country. To capture variation in network download speed, we conducted frequent speed tests from each probe node at both off-peak (i.e., early morning) and peak hours (i.e., noon and evening). This should constitute three measurement samples per day per probe node, but our data set suffered from missing samples due to either measurement artefacts or network problems. The goal of the paper is, therefore, two-fold: (i) to retrieve patterns of variability in download speed across probes at the representative off-peak and peak day hours, weekly and monthly time scales, and (ii) to investigate how much data can be missing and how the missing data can be structured without the extracted patterns being distorted. This latter assessment can guide future monitoring strategies as high measurement frequency is costly for data caps, storage and processing resources.

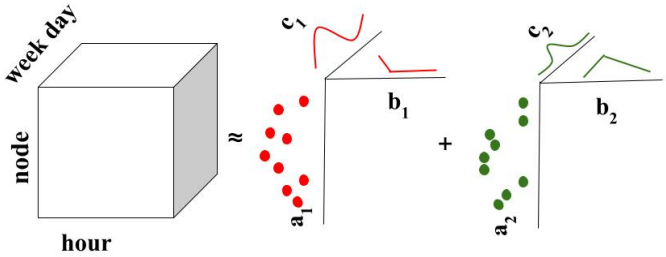


Figure 1: Two sets of patterns extracted from a 3-way tensor using a tensor factorization method.

To extract patterns from a multiway data set, we leverage tensor factorizations [8], [33], [41], as is depicted in the Figure 1 example. Over the past years, tensor factorizations have been successfully used to reveal patterns in various application domains (e.g., data mining [6], [13], [52], neuroscience [5], [50], and chemometrics [21]). In terms of network performance analysis, unlike previous work which mainly focused on either measuring network availability, reliability, speed or interactions between transport protocols and the mobile network [12], [16]–[18], [20], [45], we conduct a lon-

gitudinal study of speed variations and assess the accuracy of the derived tensor factorization models in presence of missing data. The models generated by tensor factorizations are easily explainable and offer important insights into speed variations, like the role of maintenance activity and differences between operators in terms of infrastructure and network management routines. We have already communicated these findings to one of the operators, which is factoring them into its network management plans.

The size and structure of missing samples in our datasets varies across probe nodes and operators. For example, 14% of the nodes in the first operator and 21.7% of the nodes in the second operator have 50% or more missing data samples. To impute missing data, we use a traditional method, namely, Kalman smoothing [26] with structural time series and with a *level* model [30]. We observe that our tensor model is robust to random missing data, and, in general, to all types of missing data if the probe node in question has a stable coverage profile.

Our analysis on the impact of missing data quantifies measurement sampling frequencies that are needed for describing mobile broadband speed at a given location. We find that low frequency measurements such as every second day are sufficient. This is excellent news since all efforts that measure mobile broadband speed always need to balance cost and utility (e.g., to accurately measure full capacity of a connection, the Ookla speed test uses enough data to flood the connection [4]). To the best of our knowledge, this paper is the first to look at the impact of different missing data sizes and structures on the accuracy of mobile broadband speed measurements. Furthermore, it is the first to analyze speed across spatial and several temporal dimensions by leveraging tensor factorizations. We believe that our approach and results can be of interest to network operators as well as the wider Internet measurement community.

The rest of the paper is organized as follows: Section 2 discusses the measurement data set and the motivation of the study. Section 3 briefly introduces tensor factorizations, and outlier removal steps. In Section 4, we discuss the analysis of the data as well as the patterns extracted from *variation tensors* constructed using data set with missing (i.e., non-imputed) data vs. imputed data. Section 5 discusses the reasons behind similarities/differences between the patterns extracted from *variation tensors* constructed using imputed vs. non-imputed data. It also investigates how the amount and structure of missing data affect the reliability of performance-pattern retrieval. After discussing the related work in Section 6, we conclude with lessons learnt from the study and future work in Section 7.

2 Dataset and Motivation

We leverage download speed measurements from 78 stationary probes that are part of the NorNet Edge infrastructure in Norway [35]. Each probe is a single-board computer that connects to two mobile networks, Op_1 (Operator 1) and Op_2 (Operator 2), via miniPCI modems that support up to LTE CAT-6, using commercial subscriptions. To measure speed, we use a command line client for testing Internet speed using Ookla's speedtest.net. The test is based on downloading files of increasing size from Ookla's servers and is used to estimate available bandwidth. The test runs three times a day at 2:00 am, 2:00 pm and 7:00 pm local time to capture different traffic profiles. The download speed measurement data spans a period of 317 days from January 1, 2018 to mid-October 2018. The probes nodes are in active state when speed measurement runs are performed. They are active since the probe nodes, in addition to the speedtest, regularly send one packet per second to a server to track round trip time and packet loss and exchange management traffic with the backend. However, these additional performance statistics are out of the scope of this study.

2.1 Variation in download speed

Figure 2 shows the distributions of the measured download speed for both operators. Op_1 exhibits higher download speeds and larger overall absolute speed variations than Op_2. Not only are there huge differences among the median download speeds across the nodes, but there are also substantial differences in observed speeds at the same node. The differences in median download speeds across nodes are up to 76 Mbps and 53 Mbps for Op_1 and Op_2, respectively. Op_2, however, shows a higher relative speed variability, as is shown in Figure 3 (a). This difference is because the variations in observed speeds were much higher for Op_2 than for Op_1 at the same nodes. In the worst case, the same node experienced a difference of up to 2400% between its highest and lowest observed download speed for Op_2. For Op_1, the worst case difference for a node was 72%. In Figure 3 (a), we display the empirical cumulative distribution functions (ECDFs) of the relative speed variability at the probe nodes, which is estimated by dividing the inter-quartile range of speeds by the median speed value observed by a probe node. Almost 60% of Op_2 nodes observe higher variation in speed compared to Op_1.

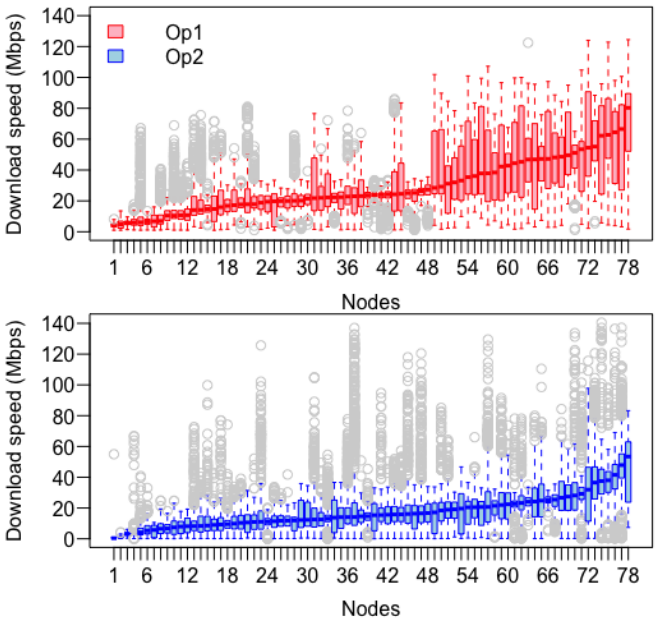


Figure 2: The distribution of download speeds per measurement node.

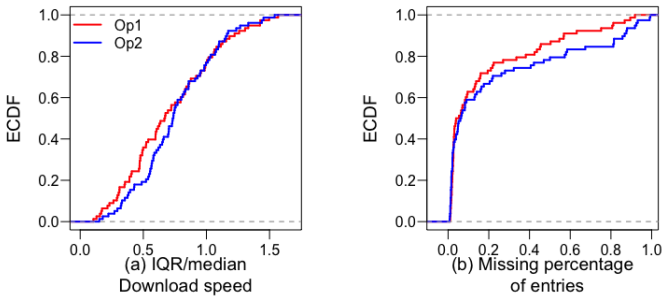


Figure 3: (a) Relative variability in download speed of nodes and (b) the distributions of percentage of missing speed measurements across nodes.

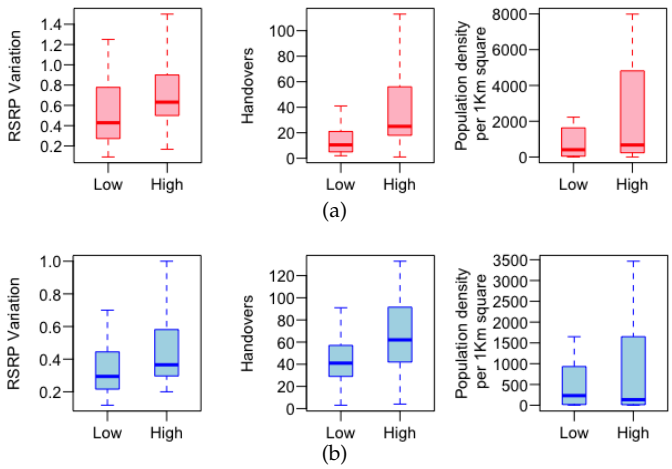


Figure 5: Metadata values for nodes of (a) Op_1 and (b) Op_2, with low and high variation in their observed download speed.

2.2 Features causing speed variability

A high variability in speed translates into performance uncertainty. A key step towards resolving this issue is to identify patterns in speed variability, which can reveal when, where and by how much speed varies. Such patterns can help an operator understand temporal and spatial properties of speed variability, which may guide further network planning and provisioning decisions.

Besides the timings of the speed tests, we have recorded a set of metadata to help contextualize and understand the speed measurements. The metadata includes the Reference Signal Received Power (RSRP) – a measure of the quality of the received signal, the identifier of the serving cell, and the population density at the location of the probe node. The network performance also gets impacted by many other contextual features (including changes in operator configuration or upgrades, construction, and foliage patterns). However, these features are not captured in the metadata.

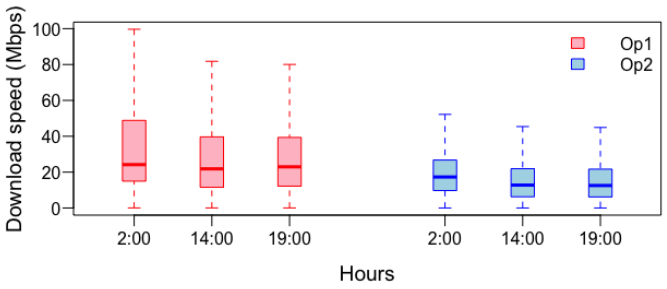


Figure 4: Download speed at off-peak (i.e., 2:00) and peak hours (i.e., 14:00 and 19:00) of day.

The time of the measurements is also a feature, and it is shared across all the nodes. Starting with this feature, Figure 4 shows that the speed varies at different hours of a day. This is expected since traffic profiles vary as a function of the time of day, e.g., peak hours are characterized by larger traffic volumes and thus lower speeds. The opposite is true for off-peak hours. The median speed at 2:00 a.m. is 19% and 43% higher than during the day for Op_1 and Op_2, respectively. For example, with Op_1 the worst connection in terms of variation had a speed drop by 123 Mbps. For Op_2, the same number was 139 Mbps.

Unlike time, RSRP, handover frequency and population density are not shared across nodes. These features can be the cause of differences in speed variation across nodes served at the same time by the same operator. To verify the impact of the non-shared features, we divide our nodes into two groups, one with high and one with low variability in observed download speed. The former includes the top 25% of nodes in terms of speed variability, while the latter includes the bottom 25%. For Op_1 the speed variation for the top 25% of the nodes is 126% more than the bottom 25% of the nodes, whereas for Op_2 this difference is 87%. We then examine the effect of the three features on speed variation. The first is coverage status, which we capture by computing the amount of variation in RSRP experienced by each node. Variation is computed by dividing the inter-quantile range of the observed RSRP values (in dBm) by the median observed RSRP. A higher variation indicates unpredictable radio connectivity. Second is the extent of handovers, in which we capture the number of times a node changes its serving cell. Frequent handovers can be caused by, e.g. poor coverage, sub-optimal configurations [9] or load balancing. In our dataset, however, poor coverage can not be the cause of the frequent handovers since we have controlled for coverage and removed probes with poor coverage (i.e. nodes which were not always on 4G). We hypothesize that these handovers are related to load balancing. We looked into the nodes that switched to different cells and found that a handover can vary median speeds by up to 50 Mbps. Among these nodes, there were only a few nodes that had a permanent switch to a cell. In most situations the nodes switched among the same set of cells randomly.

The last metadata feature is related to radio congestion. Here we use the population density as a proxy for radio congestion, which is in a 1 square km around the node. The population numbers are obtained from the database of *statistics Norway* [2]. Figure 5 compares the distributions of the three metadata features for nodes with high and low speed variability. For Op_1, nodes with a wider RSRP variation as well as those that have more frequent cell handovers exhibit higher variation in speed. Same is the case with Op_2. Interestingly, with Op_2, nodes encounter more handovers than that of Op_1. The correlation between

speed variability and population density seems weak, but we see that few of the nodes with very high population density are facing highly variable speed performance.

To rigorously confirm the correlation between the three features and speed variability, we check whether the difference between the distributions in Figure 5 is statistically significant. For example, for Op_1 we compare the distribution of RSRP variation for nodes with low speed variability to that for nodes with high speed variability. We do the same for all operator and feature combinations. We employ the R package $lmPerm$ [49], which executes a non-parametric permutation test, to compare the distributions. Table 1 presents the p-values for operator and feature combinations. If we use a p-value of 0.05 as a threshold for rejecting the null hypothesis that the two samples are from the same population, we find that intensity of variation in RSRP values and frequency of handovers are statistically significant, both for Op_1 and Op_2.

Table 1: Statistical significance (p-value) of the features towards variation in download speed value based on permutation testing.

Operator	RSRP variation	Handover	Population Density
Op_1	0.008	0.013	0.25
Op_2	0.024	$2.2e^{-16}$	0.27

The results for both operators are consistent since a poor/unstable coverage may trigger handovers. Further, though a comparatively larger population density does not impact performance variation significantly; a very high population density indeed makes the performance unpredictable due to highly variable network load over time.

In the current study, we analyze speed variability patterns observed by probe nodes from the perspective of their shared feature, i.e., measurement timings. Since we have a year-long data, we examine speed variability as an effect of different times of the day, week days and months of the measurement year. Secondly, we evaluate the similarity of the speed variation patterns drawn from measurement samples with missed observations (i.e., raw dataset) to the patterns recovered from the dataset after imputing missing speed values. This is done, to investigate the reliability of patterns derived from a measurement dataset that misses a substantial amount of samples.

2.3 Missing Samples

Our measurement study spanned over 317 consecutive days and was performed thrice a day. Thus, from each probe node, a total of 951 measurement samples were expected, but none of the probe nodes had 951 samples. For Op_1 the minimum number of samples collected by a probe node is 78, and for Op_2 it is 4. Like other real-world datasets, our data has missing observations. These are caused by both measurement artefacts (e.g., a probe was unavailable at the time) and network outages.

Overall, we collected 60763 and 56346 measurement samples from Op_1 and Op_2, respectively. For Op_1, 18% of the total speed data is missing whereas 24% of the samples are missing in Op_2 dataset. Figure 3 (b) shows that 14% and

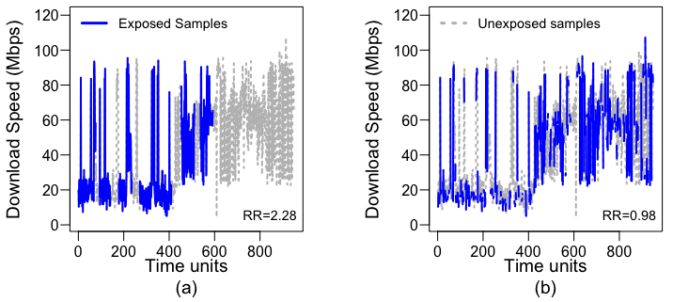


Figure 6: Missing sample structure of node (a) `19` and (b) `58`, superimposed upon the download speed measurements of node `43`.

21.7% of the connections to Op_1 and Op_2, respectively, have 50% or more missing samples.

Missing samples cannot be ignored, as they can introduce biases in the analysis of the measurements, which may be the case with our dataset. To illustrate this challenge, Figure 6 depicts the download speed values at the 941/951 measurement time units for the node with identifier `43` of Op_1. This node is one of the rare nodes with few missing samples. Two other nodes with identifiers `19` and `58` have almost half of their samples missing. The structures of missing samples for both these nodes are not identical. To see any bias, introduced by missing samples, we first superimpose the missing structure of node `19` and `58` respectively, on the node `43`. Each of the two superimpositions divide the samples of node `43` into two parts i.e., 'exposed' and 'unexposed'. For example the 'exposed' part in Figure 6 (a) consists of speed samples of node `43` from time units for which node `19` also has speed samples, while the 'unexposed' part consists of the samples at time units for which node `19` has not captured any speed samples. Figure 6 (b) shows the 'exposed' and 'unexposed' speed samples for node `43`, corresponding to observed and missing samples time units of node `58`. Next, we compute the median of all the 941 speed samples of node `43` i.e., 37.89 Mbps. In the case of unbiased sampling, we expect that both 'exposed' and 'unexposed' parts will have around half of their speed values above the overall median value of 37.89 Mbps. To confirm this, for the two different missing structures, we conduct a test of biased sampling that is called Relative Risk (RR) [11], defined by

$$RR = P_{\text{exp}}/P_{\text{unexp}}, \tag{1}$$

where P_{exp} denotes the fraction of 'exposed' samples with their speed values above the overall median of 37.89 Mbps and P_{unexp} computes the same for the corresponding 'unexposed' part. If the RR is different from 1, then sampling bias is present. By using the structure of missing data of the node with identifier `58`, we obtain an $RR = 0.98$, which indicates that the missing data structure does not induce a bias. However, the missing data structure of the node with identifier `19` yields an $RR = 2.28$, which indicates that the missing data induces a bias. Thus, it is clear that care should be taken when analyzing datasets with missing samples.

In this paper, we do not analyze the absolute download speed values; we are rather interested in deriving

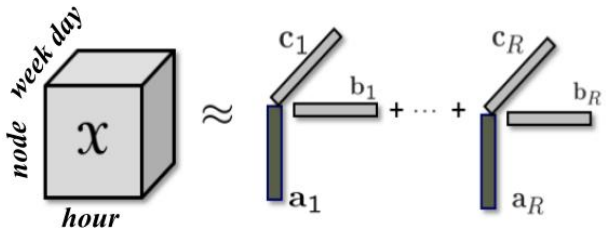


Figure 7: An R-component CP model of a third-order tensor $\mathcal{X}$.

reliable speed variation patterns of Op_1 and Op_2. We, hence, hypothesize that any amount of missing data that comes with a real-world dataset such as ours, does not impact the actual patterns of performance variability. This will be the case when both the patterns from raw (i.e., with missing samples) and imputed datasets are similar. In case the hypothesis fails, it is worthwhile to identify the particular structure, percentage of missing data as well as characteristics of performance data that do not impact the accuracy of patterns derived from the performance data set of a network operator. As a by-product, an analysis like this can help an operator in deciding on the optimal monitoring frequency, both in terms of cost and accurate performance assessment.

3 Tensor Factorizations

To explore variations in download speed across different probe nodes and time scales, we re-arrange our data as higher-order tensors, and refer to them as *variation tensors*. Higher-order tensors can be seen as extensions of matrices to multiway arrays with more than two dimensions. Following this, a matrix is a second-order tensor, and a vector is a first-order tensor. Tensor factorizations provide a natural framework to extract patterns from higher-order datasets [8], [33], [41]. More specifically, we re-arrange our data as two different types of third-order *variation tensors* to capture the variability at different granularity. The first *variation tensor* is in the form of a <nodes, hours, week days> third-order tensor, while the second one is in the form of a <nodes, hours, months> tensor. The first tensor aims to capture hourly and weekly variations across nodes, while the second aims to capture hourly variations in different months of the year. Note that we did not study fourth-order *variation tensors* with modes <nodes, hours, week days, months>. This is because ordering the data in such a way would lead to a maximum of 4 to 5 speed samples, which is insufficient for deriving variation values.

Among various tensor factorization methods, we focus on the CANDECOMP/PARAFAC (CP) model [23], [29]. The CP model is considered as one of the generalizations of singular value decomposition (SVD) to higher-order tensors. An R-component CP model expresses a third-order tensor, $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, as a sum of R rank-one tensors, as follows:

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{2}$$

where $\circ$ denotes the vector outer product. $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R] \in \mathbb{R}^{I \times R}$ is the factor matrix corresponding to the *nodes* mode, $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R] \in \mathbb{R}^{J \times R}$ is the factor matrix for the *hours* mode and $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R] \in \mathbb{R}^{K \times R}$ is the factor matrix corresponding to the *week days* mode for a <nodes, hours, week days> tensor (see Figure 7) and *months* mode for a <nodes, hours, months> tensor. The CP model provides a summary of the data tensor $\mathcal{X}$ by approximating the data as $\hat{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$, also denoted as $\hat{\mathcal{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ in short.

Unlike matrix factorizations which suffer from rotational freedom, the CP model is unique under mild conditions (up to permutation and scaling ambiguities) [34], [44]. Therefore, the CP model often reveals easily interpretable factors/patterns, and is preferred over other tensor methods when the goal is pattern discovery. For instance, if $\mathcal{X}$ is a <nodes, hours, week days> tensor, each column of $\mathbf{A}$ can reveal collections of nodes whose speed variation is related, $\mathbf{B}$ indicates how the speed variation varies for different hours of the day and $\mathbf{C}$ indicates how the speed variation differs for different days of the week. In the case of a <nodes, hours, months> tensor, the $\mathbf{C}$ matrix represents how the network speed varies across different months.

To fit a CP model to an incomplete tensor, we minimize the following weighted objective function [6]:

$$L(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{i,j,k} w_{ijk} \left(x_{ijk} - \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \right)^2, \tag{3}$$

where $w_{ijk} = 1$ for known tensor entries x_{ijk}, and $w_{ijk} = 0$ for missing entries.

3.1 Selecting the number of components

Choosing the right number of components, i.e., R in Eqn. 2, is essential to capture the patterns that underlie the data. In this subsection, we describe the metrics we use to evaluate the performance of the models and determine the appropriate number of components.

3.1.1 Model fit

A standard way of measuring how well the model describes the data is the model fit defined as follows:

$$\text{Fit (\%)} = \left(1 - \frac{\| \mathcal{X} - \hat{\mathcal{X}} \|^2}{\| \mathcal{X} \|^2} \right) \times 100, \tag{4}$$

where $\| . \|$ denotes the Frobenius norm. If the model fully explains the data, then the fit is 100%. How the model fit changes for different numbers of components shows if adding more components results in a substantial improvement in terms of explaining the remaining part in the residuals.

3.1.2 Core consistency

One diagnostic approach for determining the number of components in CP models is the core consistency diagnostic [22]. It determines the "appropriateness" of an R-component CP model to the data, and is defined as follows:

$$\text{Core Consistency} = 100 \times (1 - \frac{\sum_{i=1}^{R} \sum_{j=1}^{R} \sum_{k=1}^{R} (g_{ijk} - t_{ijk})^2}{R}), \tag{5}$$

where $\mathcal{G} \in \mathbb{R}^{R \times R \times R}$ is the estimated Tucker core given the CP factor matrices. The Tucker model [47] is a more flexible

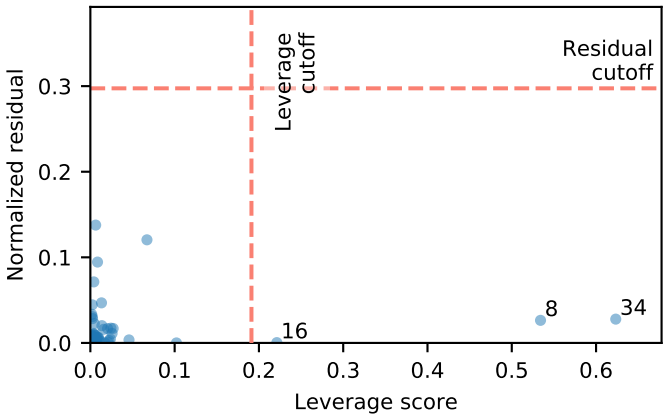


Figure 8: Identification of outlier samples (in the *nodes* mode) using normalized residuals and leverage scores.

tensor model than the CP model and its core, $\mathcal{G}$, can be a full tensor with nonzero entries everywhere. On the other hand, $\mathcal{T}$ denotes CP's core, i.e., an $R \times R \times R$ tensor with nonzero entries only on the super-diagonal. Note that, due to this flexibility, the Tucker model is not unique without additional constraints on the factor matrices and the core. Given the CP factor matrices, if off super-diagonal elements of $\mathcal{G}$ are close to zero, the R-component CP model can be considered an appropriate model. Low core consistency values close to zero or negative are considered to be a sign of an invalid CP model, whereas close to 100% is interpreted as a good model.

3.2 Implementation details

In our experiments[1], for fitting CP models, we used `cp_opt` [7] (for complete *variation tensors* that are constructed using imputed datasets) and `cp_wopt` [6] (for incomplete *variation tensors* constructed using raw datasets) from the Tensor Toolbox (v3.1) [14]. Each CP model was fit using limited memory BFGS algorithm with bound constraints (LBFGS-B)[2] using 100 different random initializations, and the solution with the lowest function value was used for further analysis (after validating the model's uniqueness). To improve the interpretability of the components, we imposed non-negativity constraints in all modes when fitting the CP model. Non-negativity constraints do not allow any element to be negative, which facilitates the interpretation.

3.3 Outlier removal: A pre-processing step

To make sure the extracted patterns (i.e., components) are not distorted by outliers, we filter out outliers as a pre-processing step, from a tensor. After fitting the CP model to a *variation tensor*, outliers can be the nodes that have high leverage values and those that are poorly described by the model.

1. https://github.com/mahfida/SpeedVariation. The repository contains raw and processed speed datasets and metadata information along with R and MATLAB scripts for imputation, factorization, outlier detection, factors' similarity check and other supporting methods.
2. https://github.com/stephenbeckr/L-BFGS-B-C

To detect nodes that have a strong influence on the model, we use the leverage score h_i, of node i, defined as follows:

$$h_i = [\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T]_{ii}. \tag{6}$$

An interpretation of h_i is the inverse number of nodes with a similar model representation as node i [31]. To determine if a node has very strong influence, we use a heuristic based on p-values [19]. A cutoff value equivalent to a Bonferroni corrected p-value of 0.01 (i.e., 0.01 divided by the number of nodes) is selected. The nodes with a leverage score above this threshold are classified as outliers and excluded.

To detect if node i is poorly described by the model, we consider its residual, r_i, given by

$$r_i = \sum_{j,k} \left(x_{ijk} - \hat{x}_{ijk}\right)^2. \tag{7}$$

A high residual indicates that the node is not well described. Similar to the leverage cutoff, we determine the cutoff for the residuals with a heuristic based on p-values [40] (selecting a Bonferroni corrected p-value of 0.01).

The process of fitting CP models and removing outlier nodes is repeated until there are no outlier nodes. This leaves 76 and 70 nodes for the <nodes, hours, week days> tensors, and 73 and 60 nodes for the <nodes, hours, months> tensors of Op_1 and Op_2, respectively. To make the two versions of *variation tensors* comparable i.e., one generated from raw dataset and other from imputed dataset, the outliers nodes in the first one are filtered out as a pre-processing step from both the *variation tensors* before their final decomposition.

3.4 Similarity of components

To measure the similarity between components captured from the two corresponding decomposed *variation tensors*, we use the congruence value. The congruence for two components, i.e., two rank-one tensors $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$, is defined as [46]:

$$\text{cong}(\hat{\mathcal{X}}, \hat{\mathcal{Y}}) = \frac{|\mathbf{a}^\mathsf{T}\mathbf{p}|}{\|\mathbf{a}\|\|\mathbf{p}\|} \times \frac{|\mathbf{b}^\mathsf{T}\mathbf{q}|}{\|\mathbf{b}\|\|\mathbf{q}\|} \times \frac{|\mathbf{c}^\mathsf{T}\mathbf{r}|}{\|\mathbf{c}\|\|\mathbf{r}\|}, \tag{8}$$

where $\hat{\mathcal{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ and $\hat{\mathcal{Y}} = \mathbf{p} \circ \mathbf{q} \circ \mathbf{r}$ are third-order tensors. A congruence value close to 1 indicates highly similar components.

4 Analysis of Variation in Download Speed over Time

To capture speed variation observed by a node at a particular time unit t, we calculate inter-quartile range of the download speed samples collected at t divided by the median of these samples. Using this process, we generate two types of third-order *variation tensors*: (i) The modes in the first one are <nodes, hours, week days> where t represents combinations of an hour of a day and a day of a week. The analysis of this tensor is to reveal hourly and weekly patterns in speed variation observed by different nodes. (ii) The second type is in the form of a <nodes, hours, months> tensor, where t denotes an hour of a day in a month of a year. Through the analysis of this tensor, we aim to reveal hourly and monthly variations.

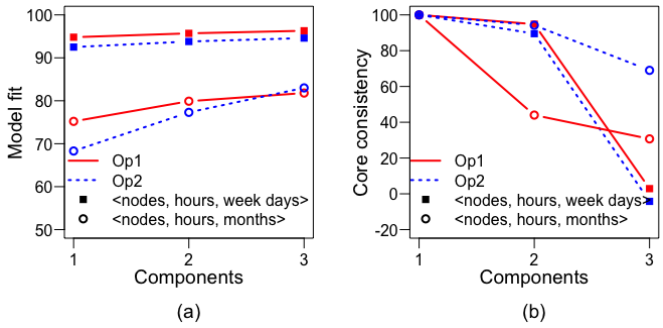


Figure 9: Model fit and core consistency (%) values for CP models of *original* tensors, with different number of components.

4.1 Original Tensors: Variation tensors constructed using data with missing entries

<nodes, hours, week days> tensor. The data from each operator consists of up to 3 speed measurements per day per node for the first 317 days of 2018. We denote this data as 'observed' data in the following text, since the raw data set only consists of observed speed samples without imputing the missing measurements. In the whole measurement duration, for every hour and week day pair, we have up to 46 download speed values per node. We generate a single speed variation value for those speed values by calculating their inter-quartile range (IQR). We then divide each IQR value by its respective median speed value to be able to compare speed variations across nodes. If a node n_i has less than 5 download speed values in an hour h_j and a week day w_k pair, we set the corresponding tensor entry to missing. The reason is, 5 is the least size of a measurement sample set to find the range of its middle 50% of sample values or to derive the first and third quantile for its IQR calculation.

<nodes, hours, months> tensor. To see if the variation in speed is consistent over the year, we also generate a *variation tensor* in the form of a <nodes, hours, months> tensor. Here, we have up to 31 entries to calculate the speed variation (i.e., IQR/median) for each cell of the tensor. Again, we set tensor entries to missing when we have less than 5 download speed observations.

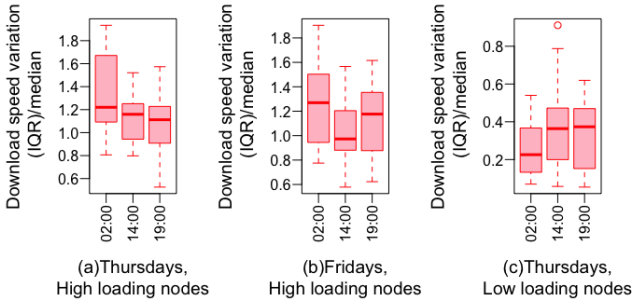


Figure 10: Variation in download speed at off-peak hours compared to peak hours on (a) Thursdays and (b) Fridays observed by top 25% nodes with high loadings, and on (c) Thursdays by bottom 25% nodes, from component 2 of Op_1 <nodes, hours, week days> *original* tensor.

4.2 CP model of original tensors

The two types of tensors described in Section 4.1 use only the observed download speed measurements. In other words, the missing download speed values are not imputed before tensor construction. We refer to the *variation* tensors constructed using observed, non-imputed data as *original* tensors.

4.2.1 *Factors from Op_1 dataset*

We model the <nodes, hours, week days> tensor using a 2-component CP model (see Figure 9 for model fit and core consistency values) as it provides high model fit of 95.7% with core consistency value close to 100%. Increasing the number of components increases the model fit slightly, but results in a sharp drop in core consistency; therefore, we focus on the 2-component model.

The first component of factorized <nodes, hours, week days> tensor (Figure 11 (a)) points out that the nodes with high loadings observed an obviously high variation in download speeds at 2:00 pm and 7:00 pm. The component captures changes in speed during busy hours. This pattern is dominant across the week, except on Thursdays, and especially on Wednesdays and at weekends. Contrary to this, the second component extracts a pattern that was prominent on Thursdays, again with higher intensity for nodes with high loadings (i.e., the nodes with high coefficients). This component captures the network maintenance activities conducted by the operator Op_1, especially on the early morning of Thursdays followed by Friday, for the group of nodes. Note that maintenance may occur on different days for different groups of nodes. Figure 10(a) and (b) show the maintenance effect for top 25% nodes with high loadings, on Thursdays and Fridays. For bottom 25% of nodes (in terms of loadings), no maintenance activity seems to be occurring on Thursdays (Figure 10(c)).

The <nodes, hours, months> tensor is modelled, too, with a 2-component CP model for Op_1. Despite its low core consistency value, of around 44%, we prefer 2-component over 1-component model due to its higher model fit of 80%. Figure 11 (b) shows that both components depict low variation at 2:00 am compared to 2:00 pm and 7:00 pm. The mean daily behaviour seems to dominate in the *hours* mode. In the month mode, each component captures the mean behaviour for different months.

4.2.2 *Factors from Op_2 dataset*

The *original* tensors of Op_2 are also modelled using a 2-component CP model. The models give a high core consistency value of 88.8% and 94.2%, and model fit values of 93.8% and 77.3% for <nodes, hours, week days> and <nodes, hours, months> tensors, respectively (see Figure 9). Increasing the number of components results in a significant drop in core consistency. The first component of the CP model of the *original* tensor of <nodes, hours, week days> in Figure 11 (c) indicates that a set of nodes exhibited a slightly more variation in the download speed at 2:00 am between mostly Monday and Thursday whereas the second component reveals that some nodes showed the lowest variation in speed at 2:00 am on Fridays.

The first component of Op_2 is a mixture of speed variations caused by network maintenance, and a network

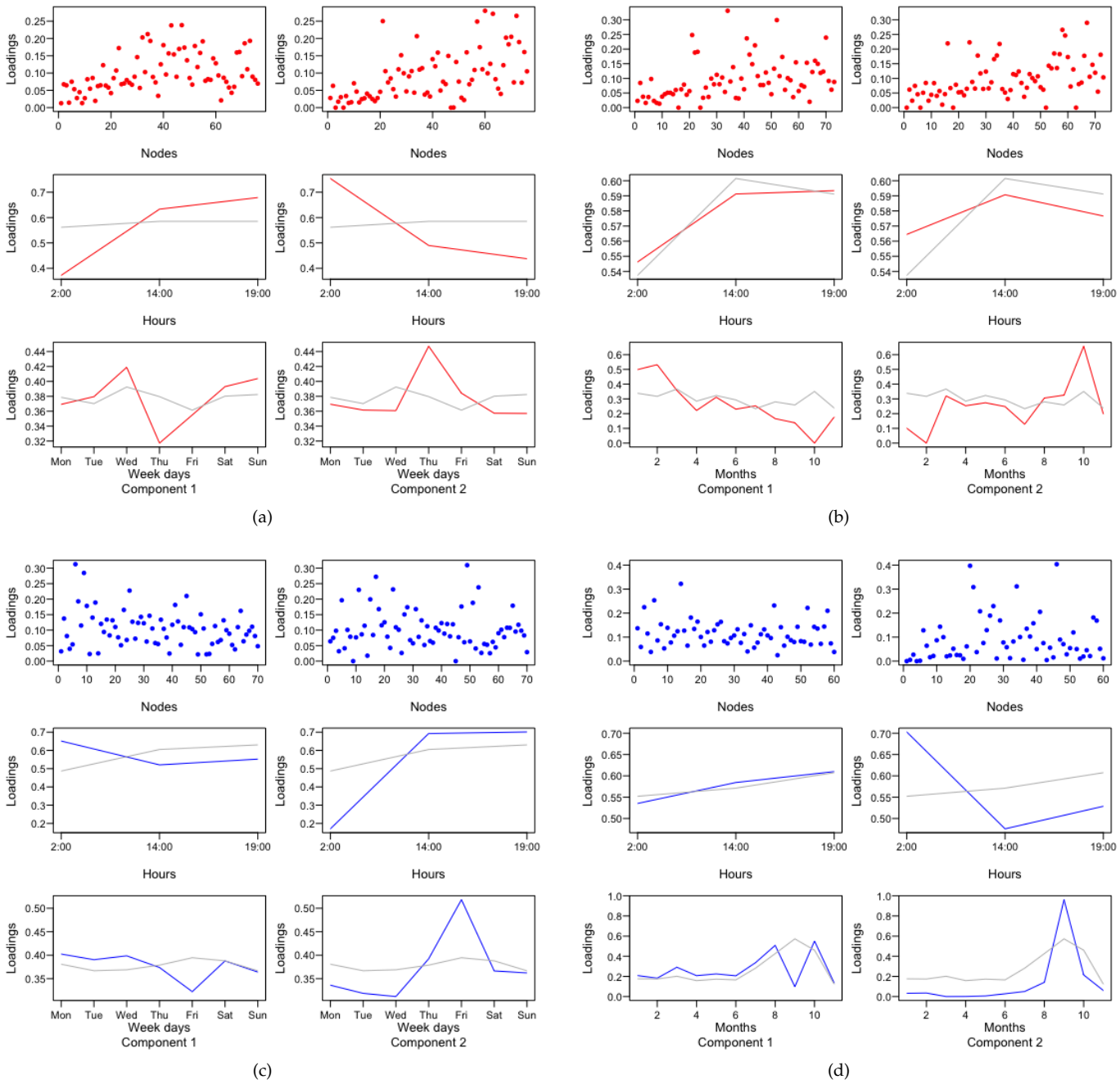


Figure 11: CP Models of *Original* Tensors. 2-component CP models of (a, c) <nodes, hours, week days> and (b, d) <nodes, hours, months> *original* tensors of Op_1 and Op_2, respectively. The red lines indicate loadings of the *original* tensor of Op_1 and the blue lines indicate loadings of the *original* tensor of Op_2. The gray lines indicate mean values of the modes in the corresponding tensor.

upgrade that happened in September 2018 (see Figure 11 (d)) affecting a group of nodes and causing more load at off-peak hours than at peak hours. Figure 12 (a) depicts that high variation in download speed, at 2:00 a.m compared to the peak hours, was observed by 25% of the nodes on Tuesdays, followed by around 16% of the total probe nodes on Wednesdays. In other words, network maintenance is undertaken on different days for different sets of nodes. Overall, the differences between hours loadings are small. The component that shows significant differences, between different hours of the day, is the one described by the second component. It reveals low speed variations at off-peak hours and high variation due to increase in mobile broadband activity during day time. For Op_2, this pattern is very obvious from Thursdays to Sundays.

The first component of the CP model of <nodes, hours, months> tensor, shown in Figure 11 (d), reveals a pattern close to the mean hourly speed and is mostly observed during the months of July, August and October. The pattern shows similar download speed variation over different hours. Contrary to this, component 2 depicts that there are nodes that observed largest variation at 2:00 am but

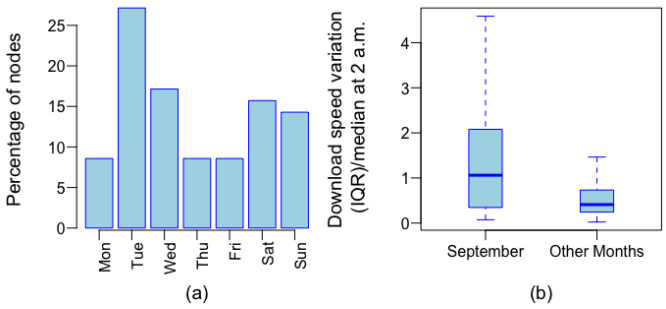


Figure 12: (a) Percentage of nodes, that observed high variation in download speed with Op_2 in off-peak time instead of peak hours on different days of the week and (b) unexpectedly high speed variation with Op_2, at off-peak hours, in September 2018.

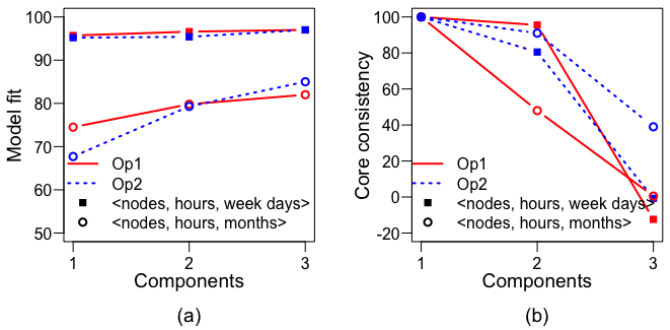


Figure 13: Model fit and core consistency (%) values for CP models of *adjusted* tensors, with different number of components.

only in September 2018. Taking all the nodes, the box-plots in Figure 12(b) depict that in the month of September the variation in speed at 2:00 am was more than double compared to the rest of the months. This hints at network upgrade activities during September, that we confirmed for Op2.

5 Impact of Missing Data

About a third of the analyzed measurement time series per operator involves at least 10% missing data. Understanding the interplay between missing data and pattern identification is important in two ways:

- It is not unusual to miss measurement samples during periodic network performance tracking campaigns. Reasons may include network outage, exceeding data capacity caps, link disruptions or issues with probe nodes. We, therefore, aim to first get insights into the robustness of the identified patterns under different structures and sizes of missing data. Significant changes in the identified patterns after imputations decrease the confidence in them.
- Second, it helps in designing future measurement strategy and picking the right monitoring frequency for a network operator. This is important, since unlike fixed measurements, mobile monitoring has a cost element due to the subscription pricing model.

In this section, we study the effect of imputing missing download speed samples on the patterns extracted from *variation tensors*.

5.1 Missing Data Imputation

To perform imputation, we apply Kalman smoothing separately on the measurement time series of each of the nodes. Kalman smoothing works by assuming a structural time series (Struct TS) model, consisting of a latent state, observations, and a process that transforms the latent state into an observation. We then fit this structural model using maximum likelihood estimation for all time-steps simultaneously [26]. This approach provides the model parameters and, importantly, the latent state for all time steps. Using this latent state, we can impute the missing data.

We used the *random walk plus noise* (or *level*) model [30] to impute the missing values. This method was selected as it provided a smaller root mean squared error (RMSE) compared with other univariate time series imputation methods (see appendix for details). To fit the model, we used the R library imputeTS [38].

5.1.1 *Variation tensors constructed using imputed data and their CP models*

After imputing the missing values in the download speed time series data of each node, we generate the <nodes, hours, week days> and <nodes, hours, months> *variation tensors* for both operators. We then model these tensors using 2-component CP models, as compared to 3-component models since 2-component models provide a better trade-off between model fit and core consistency values (see Figure 13).

To differentiate the tensors generated after imputing missing download speed values, we refer to them as *adjusted* tensors. They are derived with datasets from the same nodes and in the same way as the *original* tensors, i.e., the inter-quartile range divided by the median of the download speed values mapping to the tensor cell. The difference between *original* and *adjusted* tensors is that due to missing value imputation, none of the cells in the *adjusted* tensors is missing.

5.1.2 *Factors of the* adjusted *tensors*

As discussed in the subsection 3.2, we use `cp_opt` [7] to factorize *adjusted* tensors, as these are complete *variation* tensors with none of the cells empty.

Despite multiple nodes having a large number of imputed entries and some with large contiguous imputed blocks (replacing the missing samples and blocks), the 2-component CP model of *adjusted* tensors of Op_1 is very similar to that of the *original* tensors, as shown in Figure 14(a, b). The similarities between the components of CP models of *original* and *adjusted* tensors using Equation 8 are given in Table 2, where the value corresponds to the product of congruence values for two components as we have 2-component CP models. Congruence values of 0.93 and 0.90 between the components from *original* and *adjusted* tensors of Op_1 and similar shapes of their components, in Figure 14 (a, b), indicate that imputation has little impact on the identified patterns.

Table 2: Similarity based on congruence between the CP factors of *original* and *adjusted* tensors.

Tensor	Op_1	Op_2
<nodes, hours, week days>	0.93	0.55
<nodes, hours, months>	0.90	0.92

Unlike Op_1, the *original* and *adjusted* tensor components of <nodes, hours, week days> tensor from Op_2 have a low similarity value of 0.55.

Though visually the patterns revealed by the first and second components of *original* tensor (Figure 11 (c)) seem similar to the second and first components from *adjusted*, there are differences in the ranges of the normalized loadings of the week days in component 1, and hour loadings in component 2, of the *adjusted* tensor (Figure 14 (c)). On the other hand, as Figure 14 (d) shows, we do not see much differences between the components of *original* and *adjusted* <nodes, hours, months> tensors. This is also confirmed by their high congruence value of 0.92.

5.2 Differences and similarities after imputation

All tensors, except for Op_2's <nodes, hours, week days>, are stable under imputation, i.e., the resulting CP models are almost identical. To understand the reasons behind similarities and differences between factors from *original* and *adjusted* tensors, we take a closer look at the structure of the missing data and its interplay with imputation. We examine whether the imputation alters the statistical properties of the download speed time series. Such an alteration may result in a different CP model. To explain any structural differences between the two tensors, we examine the layout of the underlying missing data. Note that the process of imputation is highly influenced by data points that are close in time to missing data. We hypothesize that if a statistical property, of interest, from a time series with missing data is similar to its imputed time series, then the underlying size and structure of missing entries does not adversely impact the measurement campaign. In this study the statistical property of interest is the patterns of download speed variations that are experienced by the users of a mobile broadband network.

5.2.1 Imputation impact on speed variation

Imputing the missing data reduces the download speed variability for both operators. This impact, however, is more pronounced for Op_2 (see Figure 15). Despite this general trend, some nodes see an increase in variation. Across all nodes, the maximum drop in speed variation is 1.07 for Op_1 and 1.16 for Op_2, whereas the maximum increase is by 0.33 and 0.93 for Op_1 and Op_2, respectively.

To gain better insight into the impact of imputation, we divide our nodes into four groups based on the observed speed datasets. The four groups are: 1) 'low': the lowest 25% of nodes and 2) 'high': the highest 25% of nodes in terms of speed variation, 3)'small': the nodes with small missing chunks and 4)'large': the nodes with large missing chunks. Here, a chunk denotes a contiguous set of time units that lacks download speed information for a node. To separate the nodes with 'small' and 'large' missing chunks, we compute the median 'M' of the largest missing chunk sizes of all nodes. We then label any node whose largest missing chunk is smaller than the 'M' as having 'small' and those whose largest missing chunk is greater than the 'M' as having 'large' missing chunks.

Figure 16 shows the distribution of the absolute difference in speed variation between the observed and imputed data for all four categories and both operators. Nodes that are characterized by a 'high' variation in speed or a 'large' missing chunk size to start with exhibit a higher absolute difference between their download speed variation values, derived from their respective time series with missing data and imputed data. The difference, however, is more evident for Op_2. For Op_1, the top 25% nodes with high variation in speed and large missing chunks experience 68% and 57% higher absolute difference than those with low variation and random small missing chunks in download speed measurements. For Op_2, these differences are much higher, i.e., at 2000% and 225%, respectively. It, thus, seems that the difference between the *original* and *adjusted* tensors for Op_2, in Figure 15, can be linked to nodes with high variability as well as nodes with larger contiguous missing samples in their measurement time series.

Takeaways: When performance of a network such as download speed is highly variable over time, the missing observations over large time duration are harmful for the accurate analysis of the network's performance. The study shows that in the above situation the frequency of measurements should be high. To capture both the correct behaviour of a network and reduce the cost of measurement, the measurement campaign can however be turned off at random timings.

5.2.2 Interplay between imputation and tensor organization

The tensor construction involves grouping together measurements in accordance with the target structure (i.e., the modes). After seeing the impact of imputing missing entries from the observed time series, we now proceed to assess their impact on the derived 3-way tensors.

The two panels in Figure 17 confirms the role of tensor construction. As expected, the <nodes, hours, months> tensor involves more aggregation and thus exhibits lower variation than the <nodes, hours, week days> tensor. Also recall that an imputed value is influenced by observed measurements closer in time. Compared to <nodes, hours, week days>, download speed measurements more closer in time map to the same cell in the <nodes, hours, months> tensor thus exhibiting lesser variation especially in its *adjusted* tensor version. We also see that imputation increases the difference between the two tensor organizations for Op_2 (i.e., the median difference increases from 44% to 110%) while slightly decreases it for Op_1.

We next assess the way the extent of missing data impacts the difference between the *original* and *adjusted* tensors. To this end, we identify the missing data percentage that is associated with each cell of the *original* tensors. We then group cells based on similarity in their missing data percentages. For each group we compute the median absolute difference between the corresponding cells of *original* and *adjusted* tensors, as is shown in Figure 18 (a, b). We

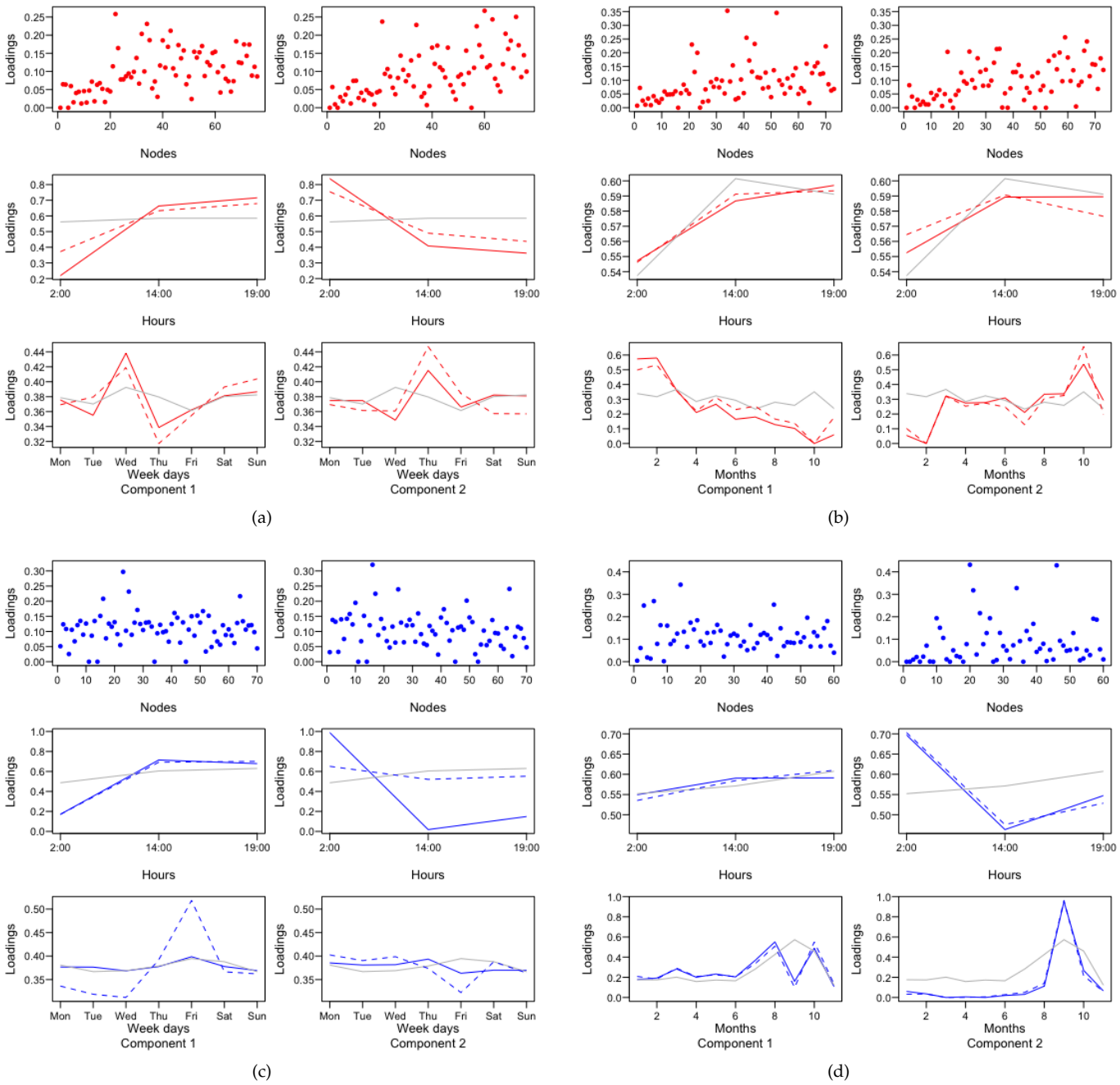


Figure 14: CP Models of *Adjusted* Tensors. 2-component CP models of (a, c) <nodes, hours, week days> and (b, d) <nodes, hours, months> *adjusted* tensors of Op_1 (shown as solid red lines) and Op_2 (shown as solid blue lines), respectively. The dashed lines show the loadings of the corresponding *original* tensors, and solid gray lines show their mean values.

also compute the median percentage change in the speed variation in *adjusted* tensor cells compared to the *original* tensors (Figure 18 (c, d)).

Irrespective of the tensor organization, when size of the missing data mapped to a tensor cell is small— the absolute difference between the speed variation values of *original* and *adjusted* tensor cells is small. It gradually increases and peaks between 40% and 70% missing data depending on the tensor.

As expected with increasing size of missing data there is an increasing trend in the 'percentage' difference between the cells of the two tensors (Figure 18 (c and d)); the 'absolute' difference however falls again as the missing entries increase beyond 70%. This is explainable as for *original* tensor cell when there are too few and similar observed values, the IQR/Median tends to be small. For *adjusted* tensor, in case of large missing blocks in speed data same value is imputed, which leads to a drop in speed variation. Note that in absolute terms, <nodes, hours, months> can tolerate a higher percentage of missing data than the <nodes, hours, week days>, again due to the way imputation is done and the tensors are formed.

Takeaways: The way missing data is imputed and a tensor is organized impacts the variation in download speed that

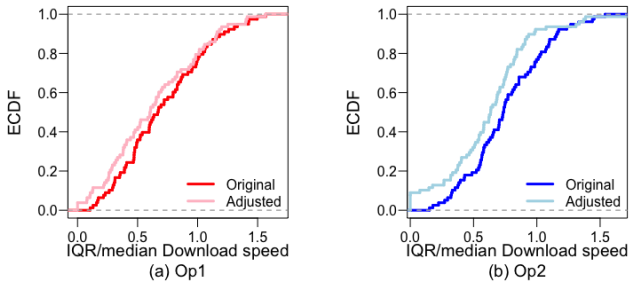


Figure 15: Distribution of speed variation in *original* and *adjusted* tensors.

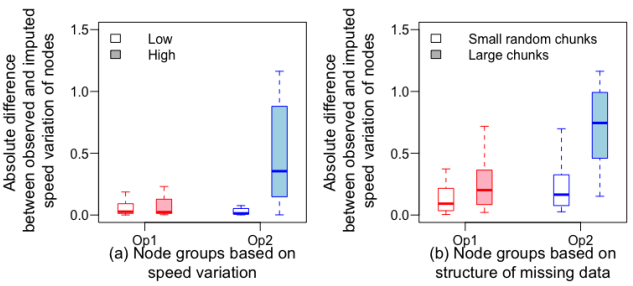


Figure 16: Absolute difference between the observed and imputed speed variation of nodes, under different features of the observed data.

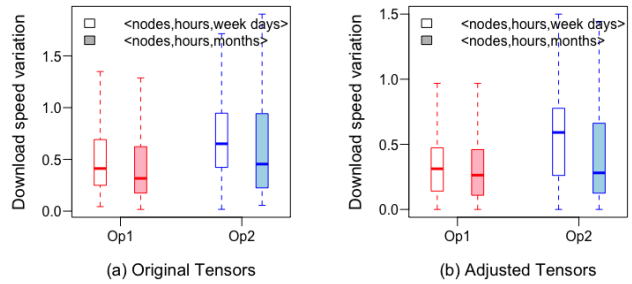


Figure 17: Variation in download speed values per operator with both versions (*original* and *adjusted*) of each tensor type.

is mapped to a tensor cell. Overall, a tensor model is fairly robust to large missing data, up to 40% for one of the operator. This is a positive result, which essentially means that a low frequency measurement process can be adequate for characterizing variability in download speed.

6 Discussion

Speed variability patterns. We have found that mobile broadband download speed exhibits variability at multiple time scales, i.e., hours, weeks and months. Identifying such intricate patterns would have been difficult if it was not for the use of tensor factorizations, which allow for analysis of multiway data. We measured two large mobile broadband operators that are characterized with different patterns of download speed variability. The variability increases at peak hours (i.e., 2:00 pm and 7:00 pm), for both operators. The difference between peak and off-peak is stronger for Op_2, which hints at smaller available capacity. This is inline with our previous study that quantified congestion in these two

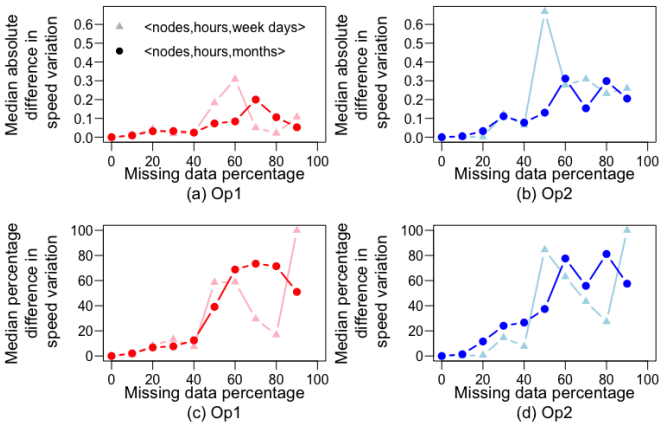


Figure 18: (a, b) Median absolute and (c, d) median percentage difference between speed variation values of cells, missing a certain percentage of download speed data, from *original* and *adjusted* tensors.

operators [28]. We have also observed that Op_1 exhibits non-trivial variability at 2:00 am, which is related to maintenance activity. Notably these activities were concentrated on Thursdays and Fridays. Interestingly, impact of maintenance is less visible for Op_2. We conjuncture that the effect of maintenance is overshadowed by the higher variability during peak hours in comparison with Op_1. Since maintenance is often infrequent, its effect will only be visible if the overall variability is small. Both operators have exhibited monthly patterns, which probably coincide with months where operators were performing network upgrades. We have confirmed this for Op_2. These findings indicate that multiway analysis of network performance data can help gauging the effect of procedures that operators perform as well as help comparing different operators across multiple dimensions.

Missing data. The constructed tensor models are robust to large amount of missing data. For the median node, the difference in speed variation when imputing up to 40% missing data is between 0.02 and 0.07 (as depicted by Figure 18) (a,b), i.e. 7%-26% depending on the operator (see Figure 18) (c,d). This attractive property can be traced to the general low variability in the measured speed. Furthermore, missing data at random seems to have minimal impact on the identified patterns. These findings have implications for efforts that are focused on tracking mobile broadband speed. Any such effort has to balance measurement overhead and utility. Our results show that a lower sampling rate, for example every second day especially in the case of Op_1, can capture the same patterns. We, therefore, believe that both measurement platforms like ours and measurement campaigns by operators can benefit greatly by integrating data analysis methods such as tensor factorizations revealing interpretable patterns into the process that determines measurement configurations. For instance, one could start with a high sampling rate for a period of time, then apply tensor factorizations while sub-sampling the collected data to determine an acceptable rate of missing data. This rate can be used to reconfigure the next rounds of measurements. We leave the details of such a process to our future work.

High variability in speed. Time of the day is a shared feature among probe nodes and a cause of variations in download speed, e.g., at peak hours the speed drops due to large number of concurrent users. The intensity in speed variation, however, varies across nodes and the network operators. Other than the time feature, nodes that experience considerable variations in RSRP and/or frequent handovers are also characterized by comparatively non-deterministic download speed. While variations in RSRP and frequent handovers can be closely related (e.g., a drop in RSRP may trigger a handover), the fact that both RSRP and speed did exhibit high values at times indicates the existence of an acceptable configuration. Furthermore, as mentioned in subsection 2.2, we did not include probes with poor coverage (i.e., those that observed mode switching). Instead, the handovers observed in our dataset may be caused by load balancing.

As far as the measurement sampling-rate is concerned, our results indicate that we need to re-adjust it for nodes with a large RSRP variation or frequent handovers. Conversely, we recommend network operators to start tracking both features and troubleshoot all cases with higher variability. To the best of our knowledge, this is not done today.

7 Related Work

Broadband performance quality and stability help users audit their connectivity costs and regulators to make informed decisions about policies and infrastructure investments. For example studies like [20], [45] conduct monitoring campaigns to probe the network speed experienced by mobile broadband (MBB) customers from various regions.

To assess causes of differences in performance quality of cellular networks from multiple dimensions, Nikravesh et al. [39] conducted a large-scale measurement study spanning 17 months from diverse set of devices and carriers located throughout the world. They found that there were significant differences in performance both within and across carriers. The performance metrics consisted of HTTP GET throughput of file hosted on a Google Server and round trip time latency to $www.google.com$. The study revealed that differences in performance are partially explainable by regional and time-of-day patterns, radio access technology, signal strength and geographic location of a probe. Importantly, it was observed that performance was inherently unstable, with some carriers providing relatively more or less predictable performance for some specific day hours. A similar study is conducted recently by Midoglu et al. [36]. Their work is based on longitudinal speed tests across multiple MBB networks from different European countries. They studied the effect of network context including radio access technology, signal strength, mobility, day hour and week day on speed quality.

Instead of analyzing performance variation of a cellular network under normal condition, Baena et al. [15] performed a long-term measurement study, during crowded events (i.e., football matches) to get an insight on the extent to which performance of different services and operators varies before, during and after a social event. They used virtualized probes in cellular network OAM[3] for monitoring and collected a number of KPIs along with downlink and uplink throughput. They observed that though throughput speed decreases during and after the social event finishes for few operators, there are operators who retain stability in their performance even during high demand. Similarly, change of serving cell ID and better RSRP brought improvement in speed but not for every operator. To get an insight from the perspective of real users, Walelgne et al. [48] used one year-long crowdsourced download speed measurements from commercial cellular networks. They assessed changes in speed from the perspective of mobile user activity, network operator, smartphone models, link stability, time of day, handover frequency, area, signal strength and day of week. They also built a machine learning model to assess stability of a cellular network and to classify reasons for network instability, using 'minimal information' such as a device model, radio technology, signal strength, and battery level. Some studies such as [10], [43] investigated performance differences in MBB networks from the perspective of sub-urban/tribal, urban and rural areas. Baltrunas et al. used a dedicated measurement infrastructure to track mobile broadband reliability and packet loss in mobile networks [16]–[18].

All the works mentioned above examined the behaviour of download speed under different settings, but did not study (i) the extent of variation in performance that a network can experience under a shared state, and (ii) the accuracy of their performance analysis with network monitoring not void of missing observations. Our work is novel in a way that it provides a multiway analysis on stability of MBB networks over different time units, i.e., a shared entity by network nodes. It then assesses the accuracy of the derived patterns of download speed variation, with missing data of different sizes and structures.

We use multiway data analysis tools, i.e., tensor factorizations, to reveal interpretable patterns of variation in download speed at different hours of the day, days of the week and months of measurement year. Tensor factorizations have been used in data mining for extracting the underlying patterns from complex datasets [8], [41]. They have been recently used in applications of data communications and mobile networks [24], [51] to find patterns of resource utilization in data centers, and to recover missing features in a network traffic measurement set. In our previous work [27], we used the CP tensor model to understand the outage behaviour of mobile broadband networks. Unlike these mentioned studies, in this paper we analyze stability over time in the download speed performance via tensor factorizations.

As for the second challenge of missing data, previous studies have employed various strategies to handle this problem such as imputation [26], [32], [38] and interpolation [37]. We, however, do not aim to increase the accuracy of our analysis by imputing the missing data. Our goal is rather to assess the percentage and structure of missing data that can be tolerated in deriving reliable performance patterns of a network.

3. Operations, administration and maintenance (OAM) system

8 Conclusion

In this paper we have employed tensor factorizations to analyze variations in mobile broadband download speed under different time units. We also evaluate the impact of missing observations on the reliability of derived network performance patterns. This is done both to assess the tolerance of the derived performance patterns to missing samples, that are unavoidable in a network monitoring campaign and to guide on future network monitoring frequencies.

Our data set involves one year measurements from 78 stationary probes, that are placed in various urban settings in Norway, and connect to the two largest network providers in the country. The monitoring frequency is thrice per day, throughout the measurement year. The measurement tests include download speed tests for both network providers. The aim is to unveil the extent of variation in download speed that network subscribers observed across the year. To extract the variation at different time scales, we re-arranged the observed variations in two tensors, <nodes, hours, week days> and <nodes, hours, months>.

Fitting a CP model to our tensors, we have found that mobile broadband speed exhibits variability at several time scales. Overall, we have established that high variability is chiefly related to maintenance activity, high network load and network upgrades. Apart from these causes, probes that are subject to frequent handovers, a possible sign of sub-optimal network configuration, have suffered high speed variability. This signature offers network operators a starting point to identify areas with unstable performance and focus on mitigating the underlying causes. This is particularly important as networks are moving to offer fixed wireless access as a replacement to traditional broadband connectivity or as a way to close the gap in broadband coverage.

As for the reliability of the inferred patterns, we find that when the missing data structure is random, then irrespective of the intensity of instability in download speed, the derived speed variation patterns are reliable. In the case where the missing structure has large contiguous time chunks, with high variations in the speed, the accuracy of derived performance patterns are doubtful.

In the future, we plan to quantify optimal monitoring frequency for the network operators on the basis of their tolerance to missing data and determinism in their speed values. We further aim to see if our models can be enhanced by including side information such as handover frequency, signal strength and the spatial locations of network nodes.

References

[1] Mobile broadband. https://en.wikipedia.org/wiki/Mobile_broadband. Accessed: 31-10-2021.

[2] Statistics norway: Population. https://www.ssb.no/en/befolkning/statistikker/folkemengde. Accessed: 1-12-2019.

[3] Broadband for all with fixed wireless access. https://www.cnbc.com/advertorial/broadband-for-all-with-fixed-wireless-access/, 2020. Accessed: 30-10-2021.

[4] Speedtest methodology: The definitive source for network performance metrics. https://resources.ookla.com/hubfs/Ookla%20Speedtest%20Methodology%202020.pdf, 2020. Accessed: 30-10-2021.

[5] Evrim Acar, Canan A. Bingol, Haluk Bingol, Rasmus Bro, and Bulent Yener. Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13):i10–i18, 2007.

[6] Evrim Acar, Daniel Dunlavy, Tamara Kolda, and Morten Mørup. Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems*, 106:41–56, 2011.

[7] Evrim Acar, Daniel M. Dunlavy, and Tamara G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, 25(2):67–86, February 2011.

[8] Evrim Acar and Bulent Yener. Unsupervised multiway data analysis: A literature survey. *IEEE Transactions on Knowledge and Data Engineering*, 21(1):6–20, January 2009.

[9] Rami Ahmad, Elankovan A. Sundararajan, Nor E Othman, and Mahamod Ismail. Handover in lte-advanced wireless networks: state of art and survey of decision algorithm. *Telecommunication Systems*, 66(3):533–558, 2017.

[10] Mark Akselrod, Nico Becker, Markus Fidler, and Ralf Lübben. 4G LTE on the road - what impacts download speeds most? In *86th IEEE Vehicular Technology Conference, VTC Fall 2017, Toronto, ON, Canada, September 24-27, 2017*, pages 1–6. IEEE, 2017.

[11] Chittaranjan Andrade. Understanding relative risk, odds ratio, and related terms: as simple as it can get. *The Journal of clinical psychiatry*, 76(7):21865, 2015.

[12] Eneko Atxutegi, Åke Arvidsson, Fidel Liberal, Karl-Johan Grinnemo, and Anna Brunström. TCP performance over current cellular access: A comprehensive analysis. In *Autonomous Control for a Reliable Internet of Services: Methods, Models, Approaches, Techniques, Algorithms, and Tools*, pages 371–400, Cham, 2018. Springer International Publishing.

[13] Brett W. Bader, Michael W. Berry, and Murray Browne. Discussion tracking in Enron email using PARAFAC. In Michael W. Berry and Malu Castellanos, editors, *Survey of Text Mining: Clustering, Classification, and Retrieval, Second Edition*, pages 147–162. Springer, 2007.

[14] Brett W. Bader, Tamara G. Kolda, et al. MATLAB tensor toolbox version 3.1, June 2019.

[15] Eduardo Baena, Sergio Fortes, Özgü Alay, Min Xie, Håkon Lønsethagen, and Raquel Barco. Cellular network radio monitoring and management through virtual ue probes: A study case based on crowded events. *Sensors*, 21(10):3404, May 2021.

[16] Dziugas Baltrunas, Ahmed Elmokashfi, and Amund Kvalbein. Measuring the reliability of mobile broadband networks. In *Proceedings of the 2014 conference on internet measurement conference*, pages 45–58, 2014.

[17] Dziugas Baltrunas, Ahmed Elmokashfi, and Amund Kvalbein. Dissecting packet loss in mobile broadband networks from the edge. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 388–396. IEEE, 2015.

[18] Dziugas Baltrunas, Ahmed Elmokashfi, Amund Kvalbein, and Özgü Alay. Investigating packet loss in mobile broadband networks under mobility. In *2016 IFIP Networking Conference (IFIP Networking) and Workshops*, pages 225–233. IEEE, 2016.

[19] David A. Belsley, Edvin Kuh, and Roy E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series. Wiley, 1980.

[20] Peter Boyland. The state of mobile network experience. https://www.opensignal.com/sites/opensignal-com/files/data/reports/global/data-2019-05/the_state_of_mobile_experience_may_2019_0.pdf, May 2019. Accessed: 30-10-2021.

[21] Rasmus Bro. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171, 1997.

[22] Rasmus Bro and Henk A. L. Kiers. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17(5):274–286, June 2003.

[23] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35:283–319, 1970.

[24] Waltenegus Dargie. Identification of resource utilisation patterns in data centers using tensor decomposition. In *28th International Conference on Computer Communication and Networks, ICCCN 2019, Valencia, Spain*, pages 1–6. IEEE, August 2019.

[25] Bidit Lal Dey, Wafi Al-Karaghouli, Stanimir Minov, Mujahid Mohiuddin Babu, Angela Ayios, Syed Sardar Mahammad, and Ben Binsardi. The role of speed on customer satisfaction and switching intention: A study of the UK mobile telecom market. *Inf. Syst. Manag.*, 37(1):2–15, 2020.

[26] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. OUP Oxford, 2012.

[27] Mah-Rukh Fida, Evrim Acar, and Ahmed Elmokashfi. Multiway reliability analysis of mobile broadband networks. In *Proceedings of the Internet Measurement Conference, IMC 2019, Amsterdam, The Netherlands, October 21-23, 2019*, pages 358–364. ACM, 2019.

[28] Mah-Rukh Fida, Andres Ocampo, and Ahmed Elmokashfi. Measuring and localising congestion in mobile broadband networks. *IEEE Transactions on Network and Service Management*, pages 1–15, 2021.

[29] Richard A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, 16:1–84, 1970.

[30] Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press, 1990.

[31] Peter J Huber. *Robust statistics*, volume 523. John Wiley & Sons, 2004.

[32] Washington Junger and Antonio Ponce de Leon. Imputation of missing data in time series for air pollutants. *Atmospheric Environment*, 102:96–104, 2015.

[33] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

[34] Joseph B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.

[35] Amund Kvalbein, Džiugas Baltrūnas, Kristian Evensen, Jie Xiang, Ahmed Elmokashfi, and Simone Ferlin-Oliveira. The Nornet Edge platform for mobile broadband measurements. *Computer Networks*, 61:88–101, 2014.

[36] Cise Midoglu, Konstantinos Kousias, Özgü Alay, Andra Lutu, Antonios Argyriou, Michael Riegler, and Carsten Griwodz. Large scale "speedtest" experimentation in mobile broadband networks. *Computer Networks*, 184:107629, 2021.

[37] Massimiliano Molinari, Mah-Rukh Fida, Mahesh K. Marina, and Antonio Pescape. Spatial interpolation based cellular coverage prediction with crowdsourced measurements. In *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, pages 33–38, 2015.

[38] Steffen Moritz and Thomas Bartz-Beielstein. imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1):207–218, 2017.

[39] Ashkan Nikravesh, David R. Choffnes, Ethan Katz-Bassett, Z. Morley Mao, and Matt Welsh. Mobile network performance from user devices: A longitudinal, multidimensional analysis. In Michalis Faloutsos and Aleksandar Kuzmanovic, editors, *Passive and Active Measurement*, pages 12–22, Cham, 2014. Springer International Publishing.

[40] Paul Nomikos and John F MacGregor. Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37(1):41–59, 1995.

[41] Evangelos E. Papalexakis, Christos Faloutsos, and Nicholas D. Sidiropoulos. Tensors for data mining and data fusion: Models, applications, and scalable algorithms. *ACM Transactions on Intelligent Systems and Technology*, 8(2):Article 16, 2016.

[42] Lucia Schiavoni. Regulators eye mobile data speeds. https://www.fiercewireless.com/europe/regulators-eye-mobile-data-speeds, 2014. Accessed: 30-10-2021.

[43] Anika Schwind, Florian Wamser, Stefan Wunderer, Christian Gassner, and Tobias Hoßfeld. Mobile internet experience: Urban vs. rural - saturation vs. starving? *CoRR*, abs/1909.07617, 2019.

[44] Nicholas D. Sidiropoulos and Rasmus Bro. On the uniqueness of multilinear decomposition of n-way arrays. *Journal of Chemometrics*, 14(3):229–239, 2000.

[45] Varadharajan Sridhar, Girish Kumar, and M. Badrinarayan. Analysis of crowdsourced data for estimating data speeds across service areas of India. *Telecommunication Systems*, 76(4):579–594, 2021.

[46] Giorgio Tomasi and Rasmus Bro. A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis*, 50(7):1700–1734, 2006.

[47] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[48] Ermias Andargie Walelgne, Jukka Manner, Vaibhav Bajpai, and Jörg Ott. Analyzing throughput and stability in cellular networks. In *2018 IEEE/IFIP Network Operations and Management Symposium, NOMS 2018, Taipei, Taiwan*, pages 1–9. IEEE, April 2018.

[49] Bob Wheeler and Marco Torchiano. *lmPerm: Permutation Tests for Linear Models*, 2016. R package version 2.1.0.

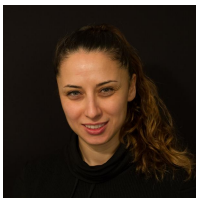
[50] Alex H. Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I. Ryu, Krishna V. Shenoy, Mark Schnitzer, Tamara G. Kolda, and Surya Ganguli. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 98(6):1099–1115, 2018.

[51] Kun Xie, Can Peng, Xin Wang, Gaogang Xie, Jigang Wen, Jiannong Cao, Dafang Zhang, and Zheng Qin. Accurate recovery of internet traffic data under variable rate measurements. *IEEE/ACM Trans. Netw.*, 26(3):1137–1150, 2018.

[52] Kejing Yin, Ardavan Afshar, Joyce C. Ho, William K. Cheung, Chao Zhang, and Jimeng Sun. LogPar: Logistic PARAFAC2 factorization for temporal binary data with missing values. In *KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1625–1635, 2020.

Mah-Rukh Fida is a senior lecturer at School of Computing and Engineering, University of Gloucestershire, UK. She was post-doctoral research fellow at Simula Metropolitan Centre for Digital Engineering in Norway during Nov.2018-March 2020. Sponsored by the prestigious UK Commonwealth Award, she completed her PhD from The University of Edinburgh, UK in 2018. She received degree of MS (IT) with distinction from IMSciences, KP, Pakistan in 2010. Before PhD, Mah-Rukh served as a lecturer in Computer Science, at Shaheed Benazir Bhutto Women University, Pakistan. Dr. Mah-Rukh's existing work focuses on measurement and analysis of performance within mobile broadband networks. Her research interest includes impact of context, content and interplay of different network components on end-to-end quality of experience.

Marie Roald is a PhD student at Simula Metropolitan Center for Digital Engineering and Oslo Metropolitan University (Oslo, Norway). She received her BSc and MSc in applied mathematics from the University of Oslo in 2015 and 2018, respectively. Her research interests include data mining and tensor factorizations as well as their applications for analyzing time-evolving, multiway datasets.



Evrim Acar is a Chief Research Scientist at Simula Metropolitan Center for Digital Engineering (Oslo, Norway). She holds an M.S. and a Ph.D. in Computer Science from Rensselaer Polytechnic Institute (Troy, NY) and a B.S. in Computer Engineering from Bogazici University (Istanbul, Turkey). Her research focuses on data mining methods, in particular, tensor factorizations, and multi-modal data mining using coupled factorizations of higher-order tensors and matrices, as well as their applications in diverse disciplines. Prior to joining Simula, Evrim was a faculty member at the University of Copenhagen (Denmark), and a postdoctoral researcher at Sandia National Labs (Livermore, CA). She is the holder of the Danish Council for Independent Research Sapere Aude Young Elite Researcher Award.

Ahmed Elmokashfi is a Research Professor at Simula Metropolitan Centre for Digital Engineering in Norway. He is currently heading the Centre for Resilient Networks and Applications (CRNA), which is part of the Simula Metropolitan Centre that is funded by the Norwegian Ministry of Transport and Communication. Dr. Elmokashfi's research interest lies in network Measurements and Performance. In particular, he has been focusing on studying the resilience, scalability, and evolution of the Internet infrastructure; the measurement and quantification of robustness in mobile broadband networks; and the understanding of dynamical complex systems. Over the past few years,he has been leading and contributing to the development, operation and management of the NorNet testbed infrastructure, which is a countrywide measurement setup for monitoring the performance of mobile broadband networks in Norway. Dr. Elmokashfi received his PhD degree from the University of Oslo in 2011.