

Understanding Spatial Relationships in Resource Usage in Cellular Data Networks

Utpal Paul*, Anand Prabhu Subramanian[†], Milind Madhav Buddhikot[‡], Samir R. Das*

*Computer Science Department, Stony Brook University, Stony Brook, NY 11794-4400, U.S.A.

[†] Alcatel-Lucent USA Inc., 600 Mountain Avenue, Murray Hill, NJ 07974-0636, U.S.A.

[‡] Alcatel-Lucent Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974-0636, U.S.A.

Abstract—We conduct a detailed measurement analysis to investigate the spatial characteristics of network resource usage using a large-scale data set collected ‘in situ’ in a nationwide 3G cellular data network. The data set spans over thousands of base stations. We first characterize the spatial correlation in radio resource usage using different statistical techniques. The analysis shows existence of significant spatial correlation that varies during the day, peaking during the middle of the day and waning in the middle of the night. We also use the notion of spectral clustering to show how base stations can be clustered based on how correlated they are in terms of radio resource usage. We show that this produces spatially connected clusters. We also show that only a few clusters exist when clustered optimally. Finally, we use the concept of Granger causality to understand the underlying functional connectivity and flow of influence in the network. We show that roughly one-third of neighboring base station pairs exhibit statistically significant Granger causality, and long causal paths exist in the network. Our observations can lead to development of new techniques for network monitoring and resource management in future cellular data networks.

I. INTRODUCTION

Broadband cellular networks are emerging to be the most common means for mobile data access world-wide. 3G networks such as narrow-band EV-DO Rev. A and wide-band CDMA (W-CDMA) based HSDPA/ HSPA are now commonplace. Higher capacity networks such as LTE and WiMax are also emerging. Various predictions from industry analysts indicate that the volume of data through cellular data networks will increase exponentially in near future [3]. While operators are scrambling to add capacity, there is an apparent lack of understanding of the nature of mobile traffic in the large scale, except a few very recent papers [26], [20], [21]. We posit that understanding of mobile data traffic via measurement and analysis is critical for the resource management in the wireless access networks. In this work, our specific focus is on ‘spatially significant behavior’ in terms of the resource usage on the network infrastructure (i.e., cells or base stations (BS)). Our hope is that our analysis will lead to a better understanding of such behaviors prompting new resource planning, spectrum allocation and network design techniques.

In our previous work, we have studied the traffic dynamics from subscriber and network perspectives as well as its impact on spectrum allocation by conducting a measurement-driven analysis of the data traffic collected *at the core* of a nation-wide 3G network [20]. Our focus in this paper is

to continue analysis on the same data set, but focus on the spatial properties and causal relationships in the network. Our goal is to provide answers to important questions regarding i) how, or if at all, radio resource usage at base stations are spatially correlated, ii) how base stations can be clustered based on the similarity of their resource usage patterns, and iii) whether causal influence exists in the network in that a base station influences the load on other neighboring base stations. These questions are important from the network providers’ perspectives specifically in the context of resource (including spectrum) management and planning.

Our data set spans one week in 2007 and consists of all data traffic associated with close to one million mobile subscribers in a nation-wide network consisting close to ten thousand base stations (both are ballpark numbers).¹ All generated data packet headers (but not including user payloads) and various signalling and accounting packets are captured, archived and later post processed using a tool we have developed.²

The remainder of the paper is organized as follows. Section II focuses on the spatial correlations among base stations. We present an optimal clustering of the base stations based on their resource usage in Section III. In Section IV, we investigate the underlying causal structure of the network. Section V describes the related work, and Section VI concludes the paper.

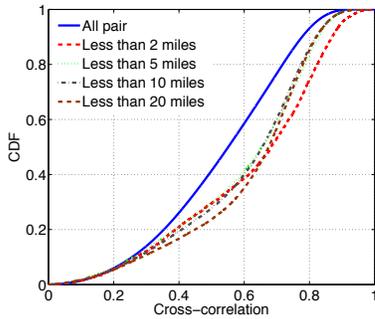
II. SPATIAL CORRELATION

We study different techniques to understand the spatial characteristic in cellular network. The results indicate that there is significant spatial correlation in the network. We also study two metrics representing network resource usage to use in our analysis.

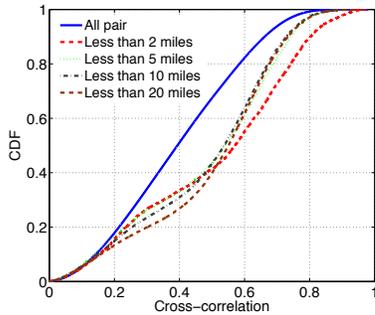
Traffic load in terms of bytes is the most used metric to describe resource usage in a network. In 3G cellular networks, another metric, ‘airtime’, can provide a more realistic indication of spectrum usage. In the commonly used 3G standards (3GPP [5] or 3GPP2 [4]), a subscriber requests and is in turn allocated a radio channel whenever it has data to send. The allocated radio channel is revoked by the

¹Though load volume has increased recently compared to 2007, we believe that the general trend should still hold for more recent data.

²For proprietary reasons, we are unable to provide further details about the nature of the 3G network, network location, data set, packet capture and post-processing techniques. This is not unusual in recent published network-wide studies [26]. In any case, the missing details are not relevant to understanding our analysis for commercially operated networks.



(a) One hour interval



(b) 10 minute interval

Fig. 1. CDF of cross-correlation between pairs of base stations (all pairs as well as pairs within different ranges) based on airtime for different summarization granularity.

network when the subscriber is dormant for a certain period, modeled using the so-called ‘inactivity timer’ (typically about 10 seconds) [11]. The inactivity timer is configurable for different networks [5], [4]. A subscriber can go between active and dormant state multiple times within a single mobile IP session. We refer the amount of time a subscriber holds onto a radio channel (regardless of whether it actually communicates) as the ‘airtime.’ Effectively, the airtime gives us the actual amount of time a subscriber uses radio and spectrum resources.

Given this background, we use airtime for our analysis as it is more directly related to the radio resource usage for the current generation networks. Further exposition of traffic load vs. airtime question is left for a different work. An interested reader can also refer to [21] for existing studies in this regard. We will use the time series of airtime (sec) for two summarization granularities: 1 hour and 10 mins, for each base station for the entire 7 days period. The goal of using two different granularity is to check whether the spatial characteristics depend on the summarization interval. Because of space limitations and also somewhat similar nature of the resulting plots, we will sometimes present the 1 hour data only in the plots.

We will now use the concept of cross-correlation to investigate the nature of spatial correlation of resource usage in the network. Cross-correlation is a standard statistical method of estimating the degree to which two time series are correlated [2], [9]. We compute the cross-correlation of

zero lag between various pairs of base stations using the time series of airtime. Figure 1 shows the CDF of cross-correlation for all pairs of base stations as well as pairs of base stations within different ranges for both 1 hour and 10 mins granularities. Note that cross-correlation, in general, between pairs of base stations is relatively high with the 1 hour interval showing a somewhat higher cross-correlation (median around 0.55 for 1 hour interval and 0.4 for 10 min interval). Also, when categorized into groups of base stations that are within different distances from each other, *closer base stations show significantly higher cross-correlation*. For example, for base stations that are within 2 miles from each other the median cross-correlation for 1 hour interval is around 0.7.

III. CLUSTERING BASE STATIONS

One can think of the pairwise cross-correlation as a similarity measure between base stations. It will be interesting to find out whether we can cluster base stations based on this measure and how such clusters look. The base stations within each cluster then exhibit similar behavior in terms of resource usage. Understanding the nature of such clusters can help the network provider in resource planning, as the provider now can think in terms of clusters or groups instead of individual base stations. The spatial nature of such clusters would be also interesting. For example, if the clusters form large connected components, then it demonstrates spatial significance. This has significant implication in terms of spectrum allocation. This, for example, shows that one can find large geographic regions, as opposed to base stations, that have correlated resource usage behavior.

We use ‘spectral clustering’ for clustering base stations. Spectral clustering is a powerful technique to partition points (in our case, base stations) into disjoint clusters such that points in the same cluster having a high degree of similarity (i.e., correlation) and points in different clusters having low degree of similarity [19], [27], [7]. The clustering algorithm works on the basis of a ‘similarity matrix.’ In our case, the similarity matrix is constructed using the pair-wise cross-correlation values for all pairs of base stations, thus forming a matrix.

We use the spectral clustering technique as presented in [19], [27]. The limitation of this technique is that the number of desired clusters needs to be specified. To compute the ‘optimum’ number of clusters, we follow the technique proposed in [27], where the algorithm self-tunes itself.

We get 4 clusters as the optimum for hourly data and 5 clusters for 10 min interval data. As the results for both the intervals are somewhat similar, we only show the results for 1 hour interval. Figure 2 shows the clustering output for a sample geographic region of size 110 mile \times 110 mile. The map is partitioned into Voronoi cells. Each Voronoi cell approximates the geographic region of one base station’s coverage. The color of each Voronoi cell indicates its cluster. The Voronoi cells widely vary in size – denser in downtowns/city centers and sparser in suburbs. We have also provided zoomed-in versions of two of the denser areas in the map.

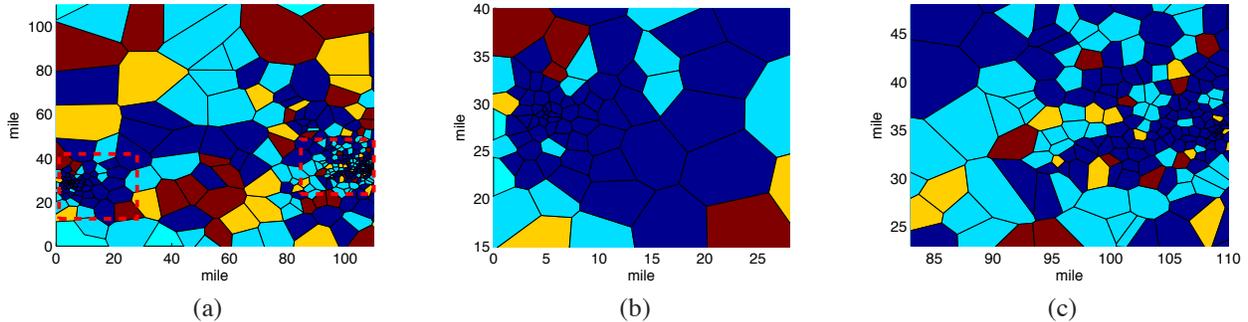


Fig. 2. (a) Spectral clustering of base stations in a 110 mile \times 110 mile region shown by coloring of the corresponding Voronoi cells. Four different colors represent four clusters. (b) and (c) are the zoomed-in versions of the densely deployed base stations in the two areas indicated by red dotted rectangles in (a) on the left and right side, respectively.

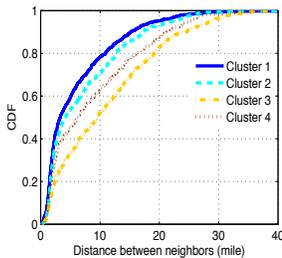


Fig. 3. CDF of the distance between two neighboring base stations in each cluster.

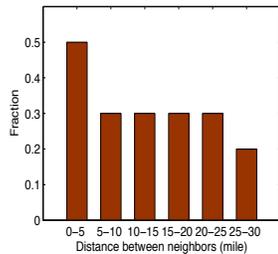


Fig. 4. Fraction of neighbor pairs that are in the same cluster, categorized on distance.

Visual inspection of these colored maps reveals some degree of spatial connectedness of the clusters. Particularly, the ‘blue’ and ‘cyan’ clusters exhibit significant connectedness. Also, note that the ‘blue’ cluster is quite prevalent in the dense (downtown) areas, followed by the ‘cyan’ cluster. The sparser areas, on the other hand, have contributions from all clusters, though ‘cyan’ appears somewhat more prevalent. Figure 3 shows the CDF of the distances between the neighboring base stations in each cluster. The same cluster color is used as before for easy readability. Note that the ‘blue’ cluster is the most densely packed (small cells). The ‘brown’ and ‘yellow’ clusters are mostly concentrated in the suburbs and thus less densely packed (larger cells). This indicates that there is some relationship between cell size and the cluster it belongs to. There is a tendency of cells of similar sizes to cluster together.

Finally, to further analyze the spatial connectedness of the clusters, we investigate the probability of a neighboring pair being in the same cluster. Overall, about 40% of the neighbor pairs are grouped in the same cluster. We also calculate this probability for neighbor pairs at different distances. Figure 4 shows that the neighbors that are geographically closer have a higher tendency to be in the same cluster (upto almost 50%).

IV. CAUSALITY

From correlations, we turn to functional influences in this section. An important metric to understand the underlying functional connectivity in the network is the ‘causal influence’

among the base stations. To keep the computational requirements reasonable, we do this investigation among the neighboring base stations only. The causality relationship among the neighboring base stations can be helpful in predicting the base stations’ loads, and thus allocate the spectrum accordingly in advance. While there are many avenues to pursue this, we use the notion of Granger causality [13], a statistical concept used to measure causality between a pair of time series.

A. Granger Causality

Granger causality (G-causality) determines whether one time series is useful in forecasting another [13]. According to G-causality, one stochastic variable X_2 ‘Granger-causes’ another stochastic variable X_1 if the information in the past of X_2 helps predict the future of X_1 with a better accuracy than is possible when considering only information in the past of X_1 alone [13], [24]. In other words, there is a Granger causality from X_2 to X_1 , if X_2 provides statistically significant information about the future value of X_1 . Such causality relation is not symmetric, meaning that ‘ X_2 Granger-causes X_1 ’ does not necessarily imply ‘ X_1 Granger-causes X_2 ’.

Granger causality was originally used in the field of economics to study the relationship between different economic variables such as GDP, oil price, stock market price, unemployment rate and so on [12], [15]. Recently there has been a growing interest in the field of neuroscience for using G-causality to identify causal interactions in neural data (see, e.g., [23], [25]). Use of Granger causality in communication network measurement and analysis is, however, rare. The lone example we have found is a recent study using Granger causality to understand the relationship between building occupants’ energy usage and their IP traffic [18].

We now present the formal definition of Granger causality. Suppose, we have two time series $X_1(t)$ and $X_2(t)$ both of length T . We can describe the two time series using a bivariate autoregressive model [24]:

$$X_1(t) = \sum_{i=1}^p A_{11,i} X_1(t-i) + \sum_{i=1}^p A_{12,i} X_2(t-i) + \varepsilon_1(t).$$

$$X_2(t) = \sum_{i=1}^p A_{21,i} X_1(t-i) + \sum_{i=1}^p A_{22,i} X_2(t-i) + \varepsilon_2(t).$$

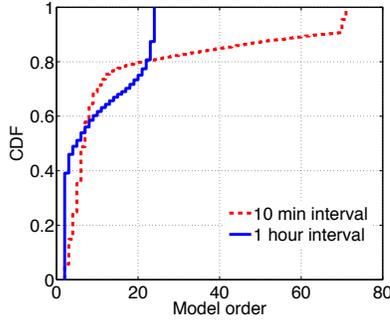


Fig. 5. CDF of model order for each pair of base stations.

Here, $p < T$ is the maximum number of lagged or past observations of X_2 (or X_1) used to predict the current value of X_1 (or X_2) at time t . It is called the model order and is provided as a parameter to the model. There are different criteria to determine the appropriate model order, p , so that the data can be represented correctly. Among them, Bayesian Information Criterion (BIC) [22] or the Akaike Information Criterion (AIC) [6] are mostly used. The matrix, $A = \{A_{mn,i}\}$ contains the model coefficients and ε_1 and ε_2 are the prediction errors or residuals. By definition, $X_2(X_1)$ Granger-causes $X_1(X_2)$, if all the coefficients $A_{12}(A_{21})$ are non-zero (in other words, if the variance of error term $\varepsilon_1(\varepsilon_2)$ is significantly reduced by the inclusion of $X_2(X_1)$ in the first (second) equation). It is important to check whether the causality is statistically significant or not. This can be done using the F-test [24]. To become statistically significant the F-statistic value should be greater than a critical value of the F-distribution for some desired significance threshold, between 0 and 1. A significance threshold closer to zero indicates a stricter test.

In our context, the time series for airtime consumed for a pair of neighboring (in the Voronoi sense) base stations describe the behavior of the two variables X_1 and X_2 . We have used the ‘Granger Causal Connectivity Analysis’ toolbox [24] for MATLAB for our analysis. Akaike Information Criterion (AIC) is used to find the model order p . For the statistical significance test we use 0.05 as the critical value. We test the causality for every neighboring pair of base stations in both directions. In our analysis, 32% of the neighbor pairs show significant causality at least in one direction for 1 hour interval data. This number increases to 40% when airtime is summarized to 10 min interval. *One can conclude that roughly for one third of the neighboring base station pairs there is causality at least in one direction that is statistically significant.* Figure 5 shows the distribution of model orders for each pair of base stations. Note that model order is generally low (median is 5-7) for both the intervals.

To understand the causal properties of the network as a whole, we define a *causality graph* using the pair-wise causal relations [18]. The Granger causality graph is a directed graph

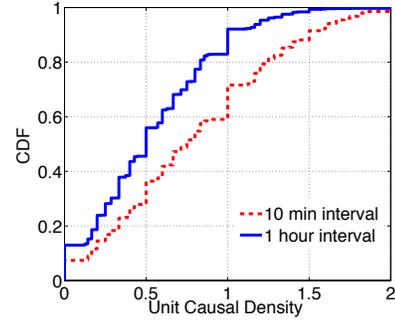


Fig. 6. CDF of unit causal density of each base station.

$G = (V, E)$ where V is the set of vertices, E is the set of edges. Each base station in the network becomes a node in the graph. There is an edge from node a to b in the graph, that is $(a, b) \in E$, if the corresponding base stations are neighbors in the voronoi construction and there is significant Granger causality from a to b . The causal graph allows us to explore an interesting set of causal properties [24] that we will describe now.

B. Causal Density

The *causal density* of the dynamics of a system is a global measure of causal interactivity [24]. It is a single index showing the mean causality over the whole network. High values of causal density indicate that the network elements are globally coordinated in their activity. It is defined as the average G-causality over all the pairs of base stations considered. Causal density bounded in $[0,1]$ gives the average count of significant Granger causality over the whole network. The causal density can be defined using the causality graph as follows:

$$\text{Causal density} = \frac{\sum_{a \in V} \sum_{b \in V - \{a\}} I[(b, a) \in E]}{\sum_{a \in V} |N_a|}$$

where N_a defines the set of neighbors of base station corresponding to node a in the voronoi sense and I is the indicator function. In our analysis we get causal density equal to 0.322 and 0.4 for 1 hour and 10 min interval data, respectively. This indicates existence of statistically significant G-causality on average.

A similar term, *unit causal density* can be defined to find the interaction locally for each base station [24]. This indicates how a base station is causally involved with its surrounding base stations. It is the sum of causal interactions around a base station normalized by its number of neighbors. A node with high unit causal density can be viewed as *causal hub*. Figure 6 shows the distribution of unit causal density. The median of causal density is 0.5 (0.7) and around 20% (40%) of the base stations have causal density greater than 1 for 1 hour (10 min) interval, which indicates *significant G-causality with at least 50% of neighbors in either direction.*

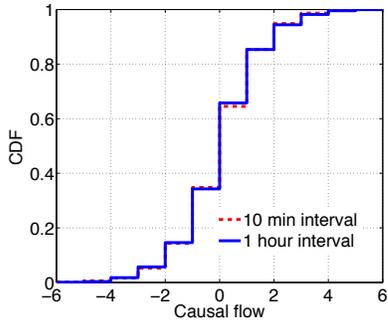


Fig. 7. CDF of causal flow.

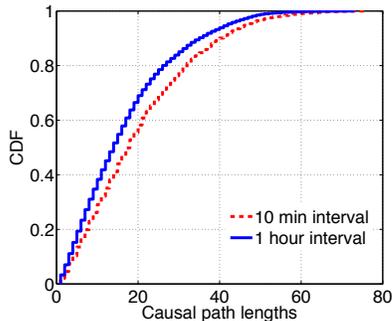


Fig. 8. CDF of path lengths in the causal graph.

C. Causal Flow and Causal Path

It is interesting to explore which base stations produce more causal influence on its neighbors and which are the sink base stations, i.e., are mostly influenced by its neighbors. The concept of *causal flow* is used to discover this. Causal flow of a base station is the difference of causal interaction by it to its neighbors and causal interaction pointed to it by its neighbors [24]. In the context of the causality graph, the causal flow of a node, a is the difference between its out-degree (number of edges from node a) and in-degree (number of edges into node a). A node with highly positive causal flow can be viewed as *causal source* and highly negative causal flow can be viewed as *causal sink*. Figure 7 shows the CDF of causal flows of the base stations in our network. It shows that about 30% of the base stations have large causal flow – either positive (≥ 2) or negative (≤ -2). This observation is similar for both the intervals.

Next we investigate how the influence from one base station propagates or flows in the network. The paths in the causality graph indicate these flows. This somehow indicates the order in which the forecasting and resource allocation should be done for the base stations in the network. We define the *causal path* as a path in the causality graph. We enumerate all such causal paths in the causality graph. Figure 8 shows the CDF of causal path length in the causality graph for both intervals.

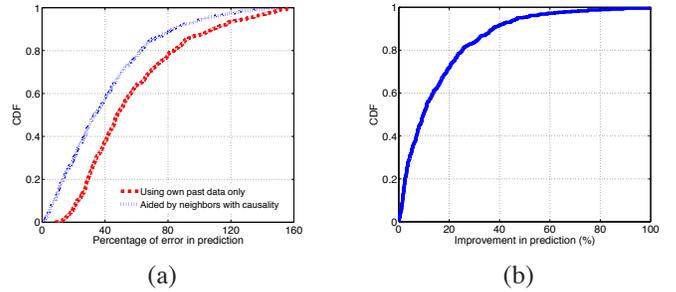


Fig. 9. (a) CDF of average absolute error percentage in prediction and (b) CDF of improvement in prediction.

It shows the presence of very long paths in the graph. The median is around 15 and the 90-percentile path length is 37.

In our future work, we plan to investigate whether the causal flow properties as well as the causal paths have any ‘out-of-band’ property. We are interested in identifying the causal sources and sinks on the map and correlating them with various forms of GIS data. We also like to investigate whether causal paths follow major roadways or neighborhoods with special properties.

D. Load Prediction using Causality Relations

Recall that a base station’s load can be predicted better using the past information of other base stations that ‘Granger-cause’ the first one. Here, we show how knowledge of causality improves prediction. To do the prediction, we use the Auto-regressive Moving Average (ARMA) [1] model. We only use the one-hour interval summarization data. First, for each base station we create the model using its own load of the first 4 weekdays and predict the load of each hour of the fifth weekday of our data set. Next, we do the same using the past load of all the neighboring base stations with causal relationship along with its own history. Figure 9(a) shows the CDF of average absolute error percentage in prediction of each base station using both the techniques, while Figure 9(b) shows the CDF of improvement in prediction (the difference of average absolute error) of each base station. The plots show significant improvement in prediction when causal influences are taken into account.

V. RELATED WORK

Relatively few papers have analyzed cellular network characteristics using a large scale measurement as this. The authors in [14] have shown the distribution of voice call durations analyzing the call logs from a cellular GSM provider. In a recent work [8], the authors have developed a tool to generate synthetic mobile network traffic using different data sets and models providing partial information about mobility and calling patterns. The authors in [10] have investigated some aspects of human dynamics and social interactions using large scale mobile phone records. The authors in [21] have characterized the settings of operational state machine that guides the radio resource allocation policy in a UMTS

network. They have used actual cellular data traces for the investigation. The authors in [16] have analyzed customer tickets collected from a large cellular network to identify potential network problems. The authors in [17] have grouped users and browsing profile simultaneously using real mobile network data collected from a large 3G cellular service provider. The authors in [26] have presented a large scale measurement analysis to characterize the primary usage in cellular voice network. They have investigated the spatial correlation in the network but in a limited scope. In our earlier work [20] we have used the same data set as in this paper and analyzed individual subscriber behaviors, subscriber mobility, and base station traffic dynamics at length. We also investigated some limited amount of spatial correlation properties in [20].

VI. OBSERVATIONS AND CONCLUSION

In our knowledge, our work is the first major study in measurement analysis of a large-scale 3G cellular data network with specific focus on spatial correlation and causality properties. We have made several important observations that have implications in network resource management:

- 1) There is a significant amount of spatial correlation for base stations that are in close proximity. These correlations are time sensitive. They increase during high traffic times (midday) and fall during low traffic times (midnight).
- 2) Spectral clustering based on cross-correlation shows that cells of similar sizes and in the same neighborhoods have a tendency to be clustered together. Also, the number of clusters was found to be small.
- 3) There is a statistically significant causal structure in the network affecting roughly one-third of the base stations. Causal paths tend to be long, indicating long chains of influence in the network.

The above observations can help develop future analysis and forecasting tools to better provision the network, and for better spectrum and radio resource management and planning. For example, spatially significant correlation in the network load indicates that for many monitoring purposes it may be sufficient to sample loads sparsely across time and space. Since there are only few clusters, these techniques can be made quite attractive. This will in effect reduce the network monitoring burden on the part of the operator. Existence of causal influence between neighboring base stations indicates that load forecasting techniques need to use past loads of neighbors for better prediction. Such forecasting may be useful in various resource management decisions, including spectrum management and energy conservation. The model order is typically small, meaning that one does not need to go too much into the past and thus archiving burden is not significant. Further, the existence of long causal paths is interesting and needs to be examined carefully with respect to available out of band information, such as nature of neighborhoods.

ACKNOWLEDGMENT

This work was partially supported by NSF awards 0831791, 0831762, 1117719 and 1117597.

REFERENCES

- [1] Auto-regressive moving average. http://en.wikipedia.org/wiki/Autoregressive-moving-average_model.
- [2] Cross-correlation. <http://en.wikipedia.org/wiki/Cross-correlation>.
- [3] Message from Mobile Wireless Congress 2010. <http://www.analysysmason.com/About-Us/News/Insight/The-message-from-MWC-2010/>.
- [4] Third generation partnership project 2 (3gpp2). <http://www.3gpp2.org/>.
- [5] Third generation partnership project (3gpp). <http://www.3gpp.org/>.
- [6] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19:716–723, 1974.
- [7] F. R. Bach and M. I. Jordan. Learning spectral clustering. Technical report, EECS Department, University of California, Berkeley, Jun 2003.
- [8] R. Beckman, K. Channakeshava, F. Huang, A. Vullikanti, A. Marathe, M. Marathe, and G. Pei. Implications of dynamic spectrum access on the efficiency of primary wireless market. In *Proc. DySPAN*, 2010.
- [9] M. W. Bulach. *Canonical Auto and Cross Correlations of Multivariate Time*. Dissertation.Com, 1999.
- [10] J. Candia, M. Gonzalez, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabasi. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41:1–11, 2008.
- [11] M. Chuah and W. Luo. Impacts of inactivity timer values on UMTS system capacity. In *Proc. of IEEE Wireless Communications and Networking Conference*, pages 897–903, 2002.
- [12] I. Dobre and A. A. Alexandru. The USA shadow economy and the unemployment rate: Granger causality results. *Journal of Applied Quatative Methods*, 5, 2010.
- [13] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969.
- [14] J. Guo, F. Liu, and Z. Zhu. Estimate the call duration distribution parameters in GSM system based on K-L divergence method. In *Proc. WiCom*, 2007.
- [15] C. Hiemstra and J. D. Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. *Journal of Finance*, 49, 1994.
- [16] Y. Jin, N. Duffield, A. Gerber, P. Haffner, W.-L. Hsu, G. Jacobson, S. Sen, S. Venkataraman, and Z.-L. Zhang. Making sense of customer tickets in cellular networks. In *Proc. IEEE INFOCOM*, 2011.
- [17] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao. Profiling users in a 3G network using hourglass co-clustering. In *Proc. ACM MobiCom*, 2010.
- [18] Y. Kim, R. Balani, H. Zhao, and M. B. Srivastava. Granger causality analysis on IP traffic and circuit level energy monitoring. In *Proc. BuildSys*, 2010.
- [19] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 2001.
- [20] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding traffic dynamics in cellular data networks. In *Proc. IEEE INFOCOM*, 2011.
- [21] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Characterizing radio resource allocation for 3G networks. In *Proc. Internet Measurement Conference*, 2010.
- [22] G. Schwartz. Estimating the dimension of a model. *The Annals of Statistics*, 5:461–464, 1978.
- [23] A. K. Seth. Causal connectivity analysis of evolved neural networks during behavior. *Network:Computation in Neural Systems*, 16:35–54, 2005.
- [24] A. K. Seth. A matlab toolbox for granger causality connectivity analysis. *Journal of Neuroscience Methods*, 186:262–273, 2010.
- [25] A. K. Seth and G. Edelman. Distinguishing causal interactions in neural populations. *Neural Computation*, 19:910–933, 2007.
- [26] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary users in cellular networks: A large-scale measurement study. In *Proc. DySPAN*, 2008.
- [27] L. Zelnik-manor and P. Perona. Self-tuning spectral clustering. *Advances in Neural Information Processing Systems*, 17, 2004.