Describing MANETs: Principal Component Analysis for Sparse Mobility Traces

Hector D. Flores, †
Stephan Eidenbenz,††
Rudolf H. Riedi,†
Nicholas Hengartner ††

Technical Report, TR2006-08,
Department of Statistics, Rice University,
August 2006

Accepted and presented at
Intern. Workshop on Performance Evaluation of Wireless Ubiquitous Networks,
Malaga, Spain, October 2006

Department of Statistics†
Rice University, 6100 Main Street, Houston, TX 77005, USA

Discrete Simulation Sciences (CCS-5)††
Los Alamos National Laboratory
Los Alamos, NM 87545, USA

Email: {hflores, riedi}@rice.edu; {eidenben, nickh}@lanl.gov;

Abstract
Data collected in realistic mobility traces for mobile ad hoc networks (MANETS) is intrinsically high dimensional. Principal Component Analysis (PCA) is a good tool for reducing the data dimemsion by extracting important features of the data. We propose a method for computing principal components using iterative regression for high dimensional matricies with missing values with an application to node degree time series. We expand this method to handle an additional dimension of information for a defined neighborhood ancestry of node degree, exposing patterns when they exist. We test our methodology on node degree data from a simulated university campus model (Pedsims) and real campus data. Results indicate that in both cases, the student's major field of study along with class schedule are strong factors to differentiate mobile node degree time series. The ability to detect differences is a powerful tool for application specific network management, allowing for: optimal placement of routers, design of specialized protocols for various user populations and lending insight to gauging the energy/bandwidth needs of mobile devices.

Supported in part from NSF, grant number ANI-0338856, Texas ATP, project number 003604-0036-2003 and Los Alamos National Laboratory, (CCS-5)

1. INTRODUCTION

The most salient characteristic of mobile ad hoc networks is their dynamic topology manifested as devices move around or even enter and exit the network. Understanding the network changes due to mobility on different time horizons is pertinent to an efficient if not proper functioning of the network on various layers of the protocol stack. Despite the importance of this issue, studying mobility and its impact on ad hoc networking still relies mainly on simulation and has only few tools of analysis to draw from.

To allow for analytical tractability and ease of simulation, the models of choice build on the principles of independent node decisions and a simple Markovian decision structure such as Random Walk, Random Waypoint, Random Direction, Gauss-Markov, Boundless Random, and others (see for example [3, 4]). Other mobility models [6] and similarly [14] incorporate adjustable levels coordination or correlation of decisions and realism of decision rules as to reflect and model complex tasks. With increasing complexity these models become naturally more application specific. Understanding how these models affect network performance is vital.

The aim of this paper is to introduce a tool which allows for careful and detailed analysis of mobility traces from any mobility model. Indeed, existing methodologies are typically ad hoc estimation of simple parameters such as time averages which shine little to no light on the dynamic aspects which are so prominent in these traces. For example, two mobility models that have the same average speed, node degree distribution, etc. can still have different time dynamics (spatial-temporal correlation) that are undetectable by averages. A major issue that hinders most COTS (commercial off the shelf) analysis tools (or at least biases their outcomes) for time series observations of mobility traces is the fact that nodes enter and exit the network, thereby creating missing data observations in a large observation window. This paper presents examples to show how poor handling of missing data is detrimental to pattern finding among mobile nodes, and we extend PCA so that it can handle missing data in a larger observed dimension. Though the idea has been applied in various application domains [11, 5, 12, 13], we believe that the introduction of this tool for analyzing mobility traces and our novel extension is new.

The degree (number of devices within range) for each individual provides a reasonable first description of the network at a given time. As the network changes, we track the time series of node degree. We demonstrate the usefulness of PCA by comparing node degree time series observed from various simulations of mobility models, showing that pattern-finding is possible when patterns exist and neglecting to account for missing observations properly can lead to erroneous conclusions.

In Section 2, we propose to use Principal Component Analysis (PCA) as a tool for summarizing and understanding the ensemble of time series of degree distribution (one for each device). We apply the methods to two examples: random waypoint (RWP) and Pedsims, a more realistic mobility model. Our analysis shows how the more realistic mobility differs from the synthetic RWP. In particular, it identifies clear clusters of individuals that have similar connectivity patterns.

Similar data can be collected in the real world. Su *et. al.* conducted experiments at the University of Toronto involving distributing PDAs to students for the purpose of collecting the Bluetooth periodic inquiry requests to describe proximity relationships[10]. A feature of this data (and more realistic simulations) is the presence of missing data that makes conventional application of PCA impossible. An alternative method to compute PCA for handling missing data is presented in Section 3.

Moving beyond description of individuals with similar node degree patterns, an important feature of the network topology is the persistence of the network. Devices need to be in range for some minimum amount of time to exchange handshakes/data. Given a mobility trace, one can compute a node's neighborhood ancestry. We define neighborhood ancestry in the following way: Each mobile node at some time t has n identifiable neighbors within range, then the k-th neighborhood ancestry of a node is the number of identifiable ancestors that are present at time $t - k * \Delta t$ that are also present at time t and all intermediate timesteps. For example, if at time t a node has neighbors (A, B, C) and at time $t - k * \Delta t$ we find that the same node had neighbors (A, B), then the k-th neighborhood ancestry of this node is 2. Note that the sequence of neighborhood ancestry numbers is a monotonically decreasing sequence in k for a fixed t, with the rate of decrease dependent on parameters of the mobility model and associated devices (speed, size of the simulation space, transmission range, and the size of Δt).

Analysis of the neighborhood ancestry requires a novel extension of the iterative methods for PCA outlined in Section 3, which we present in Section 4. We conclude this paper with a discussion on implications for ad hoc routing protocols and MANET experiment design in Section 6.

2. ANALYZING MOBILITY FEATURES WITH PCA

There appear to be two main trends in current research for collecting network statistics: global and local. Global statistics typically involve averages over time, e.g. average link duration, average number of hops, average speed, etc., and in general involve a single value to summarize an entire simulation. Rather than examining such end-values, we consider metrics that can be computed on a discrete time scale, local for each mobile node.

Of interest to current networking researchers is the node degree time series (local), or node degree distribution (global). For the remainder of the paper, we define *node degree* of node n_i as the total number of nodes within 100 meters of n_i. In this paper we consider local node degree time series, and the data structure can be identified as a two-way table. We can utilize PCA to examine such a table.

The following examples illustrate our findings with conventional PCA on two simulated mobility models: Random Waypoint (RWP) and Pedsims. In Random Waypoint, node initial positions are uniformly distributed in the space. At the start of the simulation, each node chooses a random destination and random speed and proceeds to that location. When the node arrives, the node "rests" for a fixed amount of time, and then again chooses a random destination and speed. This activity continues until a predetermined halt to the simulation. Random Waypoint has been studied in great detail [3] and has been shown to have an initial period of instability (mobility is not in a steady-state) which we have taken care to remove. Since all nodes behave according to the same algorithm, we do not expect to see any clustering

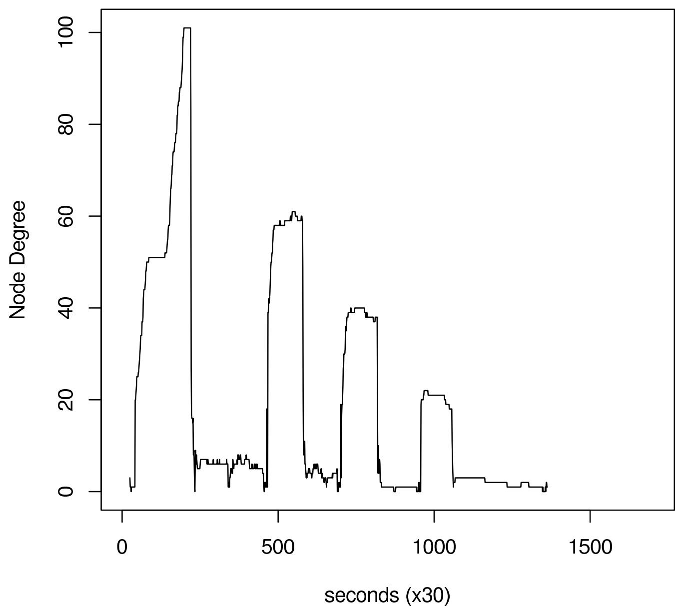

Figure 1: Sample Pedsims Node Degree Timeseries

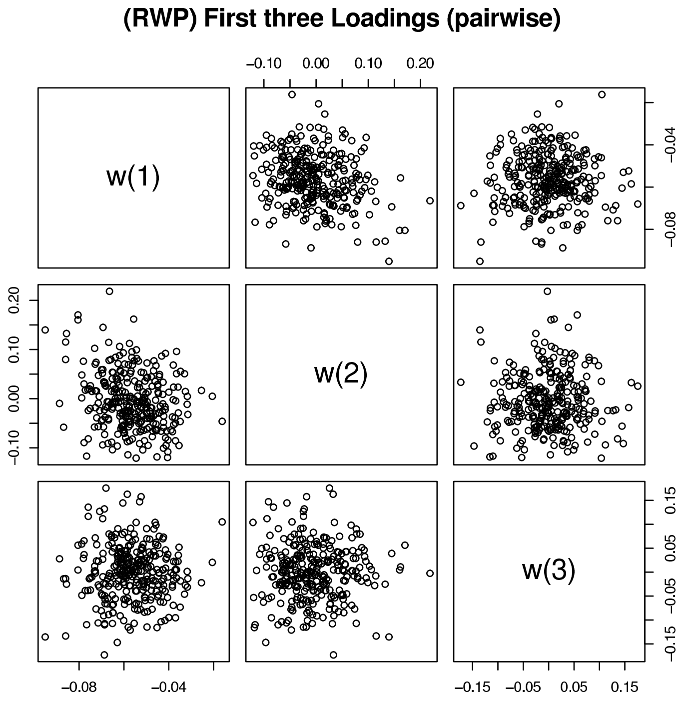

Figure 2: (RWP) Pairs plot of first three loadings

of nodes with similar mobility patterns, or more generally, with similar node degree time series.

Pedsims [6] was built as a geographically accurate, schedule-based, pedestrian-speed mobility model. The chosen geography was a university campus, where the mobile nodes (pedestrians) move about the campus along the sidewalk structures only, according to generated schedules. The authors claim that Pedsims is a more realistic mobility model than RWP that does not exhibit steady-state behavior (see sample node degree time series in Figure (1)). There exists spatial-temporal correlation built into the mobility model, and we expect similar relationships in node degree. With relationships embedded in the model, we expect PCA to effectively pattern-find.

2.1 Principal Component Analysis

When all timeseries data are complete, PCA is performed in the following manner: Let $\boldsymbol{Y}$ be a data matrix of dimension $t \times n$, where the n columns are time series of node degrees for our example. Compute the eigenvalue/eigenvector decomposition of $Y^t Y$, using any desired method (can also be accomplished generally using Singular Value Decomposition). The result is a vector of eigenvalues λ_i and corresponding eigenvectors v_i for $i = 1 \ldots n$. Note that it is customary to order $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. Note also that the vectors v_i form an orthonormal basis for $Y^t Y$. The ith principal component is defined as v_i such that i is the index of the eigenvector with corresponding eigenvalue such that $\lambda_{i-1} > \lambda_i > \lambda_{i+1}$. In our example, we consider the first three principal components which have the largest eigenvalues. Intuitively, most of the important patterns are in the first few principal components.

In the following examples we plot the first three eigenvectors as time series, which represent the main trends of node degree as a function of time. We define $w(i) = Y^t * v_i$ as the ith "loading". We compute $w(1)$, $w(2)$, and $w(3)$ and plot these vectors pairwise in the indicated scatterplots. The axis on these plots vary according to the scale of the weights v_i. Informally, any clustering seen in the scatterplots is indicative of nodes having similar degree time series.

2.2 Example 1: Random Waypoint

We simulated a standard Random Waypoint mobility model with 300 nodes in a large area (1400 x 800 meters) with pedestrian speed (uniformly sampled between 2 and 4 mph), collecting node degree every 30 seconds for 2500 seconds, removing the first 500 seconds as the 'burn-in' period. The resulting plots demonstrate what we have come to expect from Random Waypoint (after burn-in is removed). Relative stationarity of node degree (Figure 3) is demonstrated by a relatively constant first component and oscillating second and third component. No particular differences between the mobile nodes with respect to node degree (Figure 2) are seen which is as expected. Consider then a more realistic mobility model called Pedsims where a clustering of nodes with respect to their degree time series is expected [6].

2.3 Example 2: Pedsims

In the Pedsims mobility design, cohorts of students with similar schedules are expected to have similar node degree time series, so a PCA on the data should reveal clusters in the loadings. Overall, the schedules of the students are themselves correlated, so a regular periodic pattern is expected to be found across all time series. Figure(5) shows how PCA finds the underlying periodic pattern of the first three principle components, and the clusters represented in the pairs plot should be indicative of individuals with similar node degree time series (possibly individuals moving in range together) (Figure 4). The clustering that is produced, however, is not an expected characteristic of the model. Also note the uncharacteristic decrease and increase of the principal time series near the beginning and end of the series,

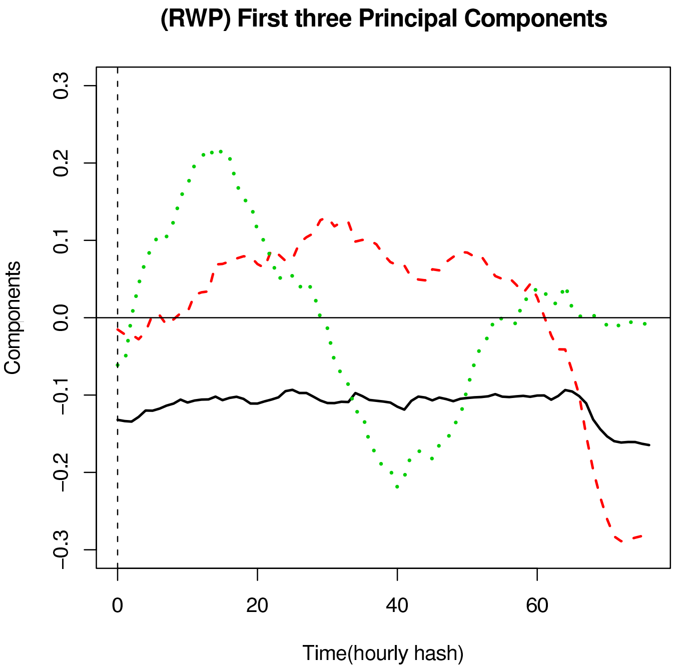

Figure 3: (RWP) Time series plot of first three eigenvectors

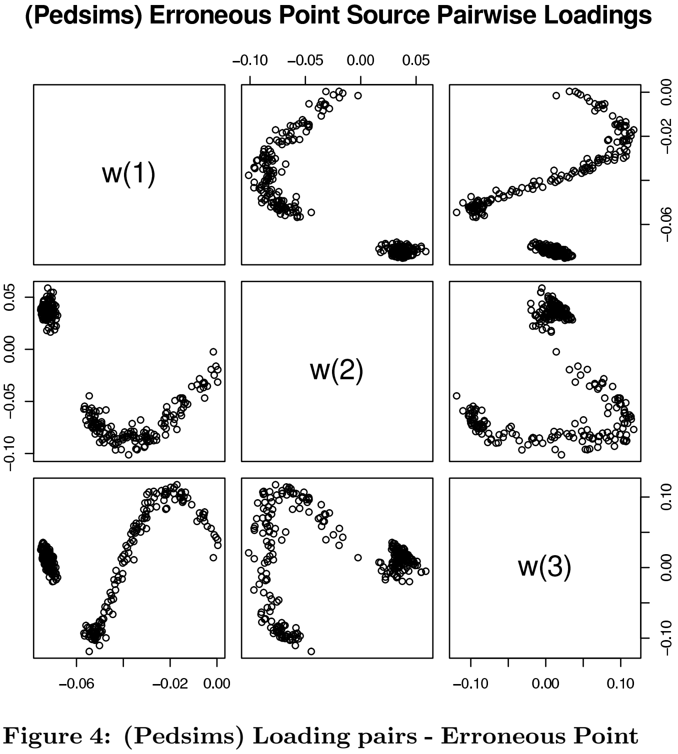

Figure 4: (Pedsims) Loading pairs - Erroneous Point Source

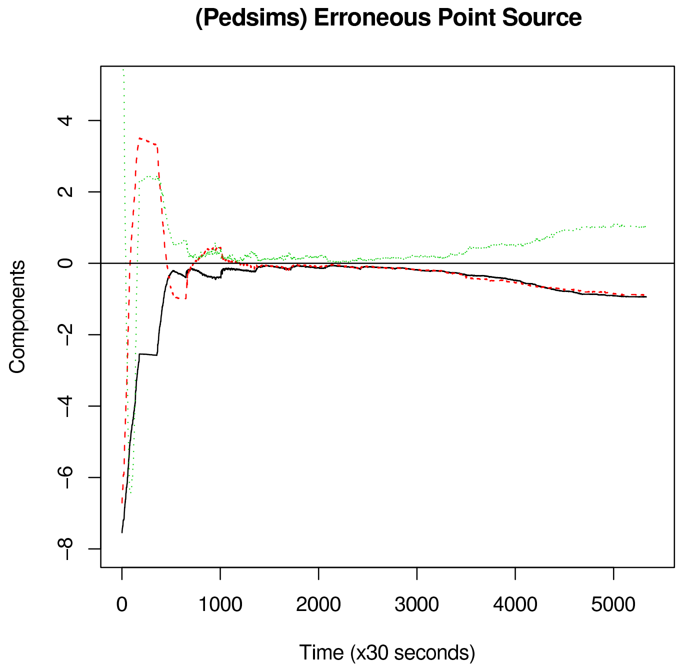

Figure 5: (Pedsims) Time series of the first three Components - Erroneous Point Source

respectively. It turns out that this trend dominates PCA's ability to pattern-find.

We can, of course, explain this behavior. A feature of Pedsims is that mobile nodes enter and leave the model somewhat freely (as they would on a university campus), so computing node degree is not always possible for the life of the simulation. The characteristic jump is not an underlying characteristic of the model, but is a result of including node degree where none exists. In this case, the trace was an input for the *Network Simulator 2* (ns2), which requires that all node positions be defined at the start of the simulation (and exist throughout the entire simulation). Node degree was computed erroneously large near both the beginning and end of the simulation, because Pedsims has only a small number of start and endpoints where the nodes collect. In short, there exists an "erroneous point source" for the nodes at the beginning and end of the simulation. Imputing the values to 0 appears to help capture more of the structure, in the time series (Figure 6), but the extra values produce clustering (Figure 7) which not useful. These plots are titled "0 != NA", indicating imputation of values to 0 is not equivalent to missing data. This is a simple example to show how flawed improper data imputation can be.

3. ITERATIVE METHOD FOR PCA WITH MISSING DATA

The problem of computing PCA for a sparse matrix has been considered by many authors[12, 2, 11, 1] and the regression approach in particular is clearly defined by [13]. We remind the reader that a sparse matrix is a matrix with more 0 values than non 0 values. We are using the term "sparse", however, to describe a matrix with mostly missing data.

3.1 Computing PCA with Missing Data

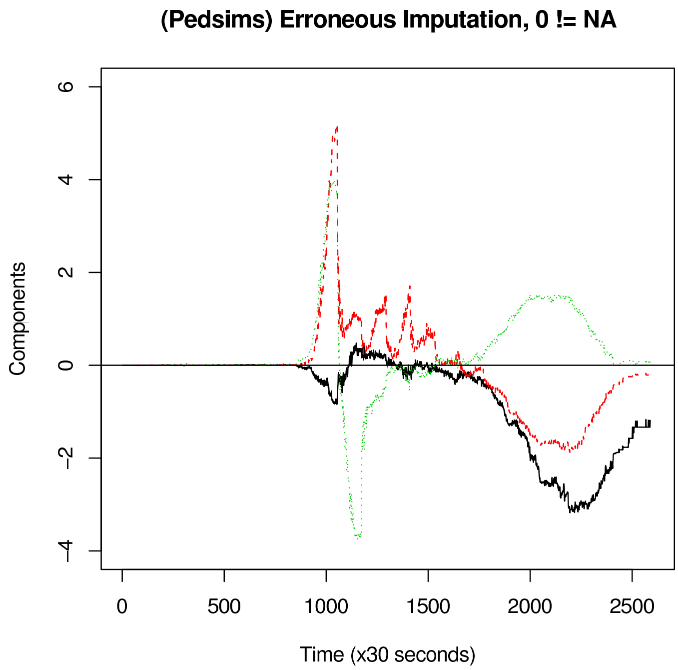

Figure 6: (Pedsims) Time series plot of first 3 eigenvectors: Erroneous 0 values

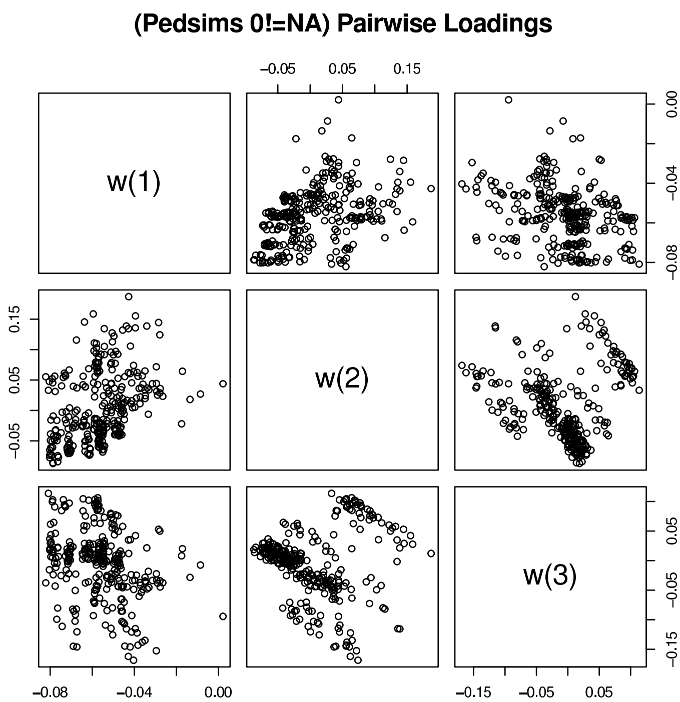

Figure 7: (Pedsims) First three loadings : Erroneous 0 values

We take the approach of [13], making 25 iterations for each component. We note that convergence occurs somewhat quickly, so 25 iterations appears to be sufficient. Understanding the relationship between the sparseness of the data and the rate of convergence of the components was not the focus of this study, and remains an open research question. The reader should continue to interpret the pairwise scatterplots and time series plots as before.

Our implementation of the algorithm is as follows:

3.1.1 *PCA - Missing Data Algorithm*

Let $(i, t) \in \mathbb{I}$, where $\mathbb{I}$ is the set of paired indicies of non-missing data elements in the i-th timeseries. We model $y_{it} = \theta_i \phi_t + \epsilon_{it}$ where ϵ_{it} is $N(0, \sigma^2)$. Note that y_{it} represents the node degree node i at time t, non-missing. One can imagine ϕ_t as the archetypical node degree time series of node i and θ_i as a scaling factor of phi_t. Using a similar methodology as [13], we formulate least squares estimates for θ, ϕ:

$$\hat{\theta}_i = \frac{\sum_t \phi_t y_{it}}{\sum_t (\phi_t)^2} \tag{1}$$

$$\hat{\phi}_t = \frac{\sum_i \theta_i y_{it}}{\sum_i (\theta_i)^2} \tag{2}$$

Since the parameters depend on each other, we use the following iterative algorithm for estimation of any number of eigenvectors:

Given $\epsilon_1 > 0$ and $\epsilon_2 > 0$.
for all i desired "principal vectors" **do**
  Initialize $\theta[N]$, $\phi[T]$ as random vectors of sizes N, T respectively.
  $\boldsymbol{\theta} \leftarrow \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$;
  $\boldsymbol{\phi} \leftarrow \frac{\boldsymbol{\phi}}{\|\boldsymbol{\phi}\|}$;
  repeat
    Compute (1) for all i.
    $\boldsymbol{\theta} \leftarrow \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$;
    Compute (2) for all t.
  until $\|\boldsymbol{\theta}_{current} - \boldsymbol{\theta}_{last}\|_2 < \epsilon_1$ AND $\|\boldsymbol{\phi}_{current} - \boldsymbol{\phi}_{last}\|_2 < \epsilon_2$
  Reassemble $\boldsymbol{y}$ with estimate $\hat{y_{it}} = \theta_i * \phi_t$
  $\boldsymbol{y} \leftarrow \boldsymbol{y} - \hat{\boldsymbol{y}}$
  Report θ, ϕ as the ith desired "principal vector".
end for

Intuitively, θ is the first few loadings that would be computed from a regular PCA. We display θ_i graphically using pairwise scatterplots as usual. Then ϕ_t is the first few principal components, and it is displayed in the usual manner as time series plots. Note that since our linear model is multiplicative, we absorb any proportionality constant by renormalizing all but one variable in the algorithm. Failing to do this will force an "identifiability" statistical issue with the parameter estimation.

3.2 Example 3: Pedsims PCA with missing data

We continue Section 2.3 to demonstrate the usefulness of the method and explore the difference of the results.

We note that the border effects from the erroneously imputed data is now corrected for in the principal time series (Figure 9), and a more accurate clustering of the first few

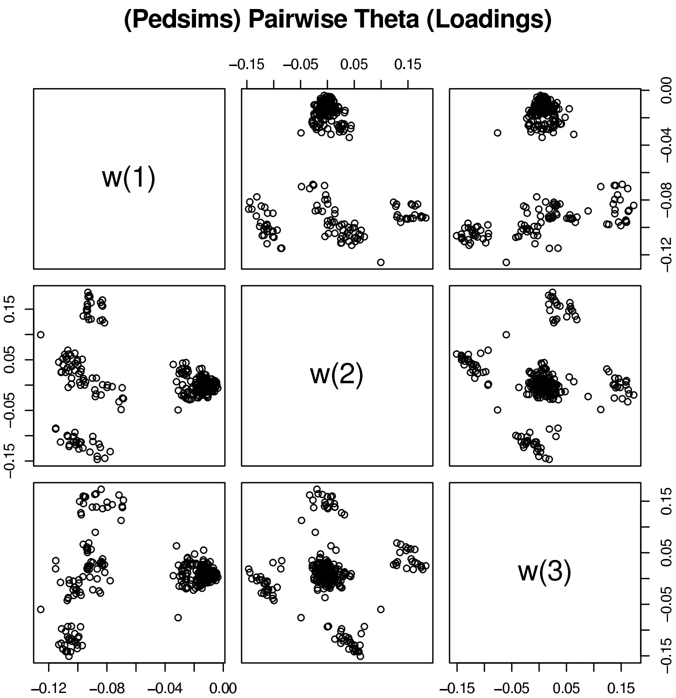

Figure 8: Pedsims: Theta pairs accounting for missing data

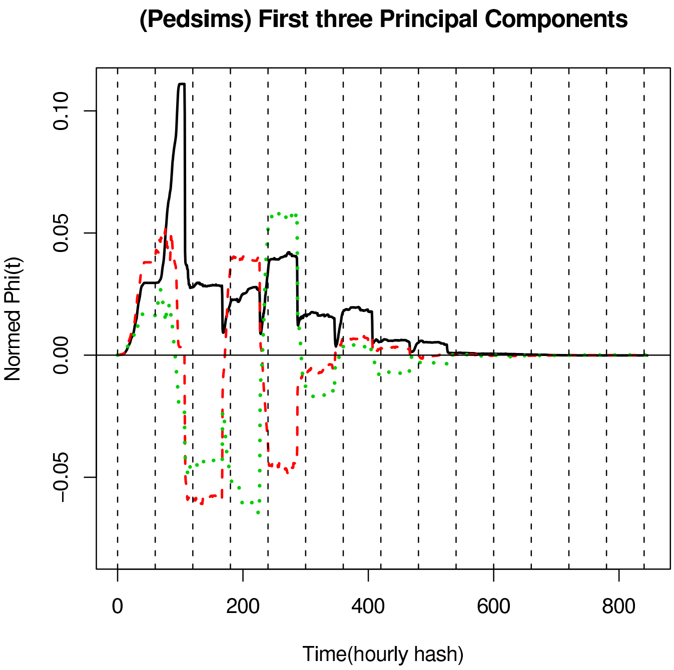

Figure 9: Pedsims: Phi(t) accounting for missing data

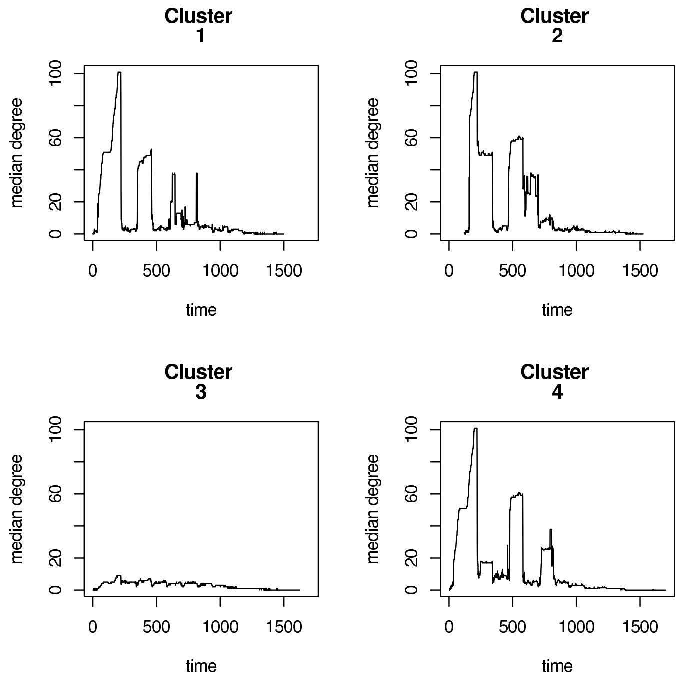

Figure 10: Time series for each cluster

loadings occurs, even in the 2nd and 3rd loading (Figure 8). This clustering clearly seperates individuals with similar node degree time series (as expected), giving us a clear idea of three main patterns (the fourth cluster at $(0, 0)$ is made up of individuals that don't fit into the other three groups) of mobility, effectively decreasing the dimensionality of the problem.

In Figure 10, the median time series plot for each cluster of nodes is presented. The third plot is represents the median time series plot for science and humanities majors, while the other three plots correspond to business majors with three different schedules. This pattern is explainable. In Figure (11) it is clear that there is only one business related activity location (B) which is somewhat isolated from the science/humanities locations [(S,H) respectively], in terms of our defined range (100 meters). In Pedsims, students of a particular major are more likely to move to activity locations associated with their major than other locations. Note that all individuals at an activity location are considered within range. As such, the business majors are very likely to have a high node degree (though perhaps not a high level of connectivity with the rest of the campus) and differ only by the number of classes assigned. A similar pattern is found in the clustering of the second and third component.

4. MODELING NEIGHBORHOOD ANCESTRY DATA

Our main contribution is then to solve the problem of PCA modeling for data in a 3-way table with missing data. Consider that we collect node degree as described above and additionally, for every node and timestep, we collect the neighborhood ancestry for K timesteps in the past (see Section 1 for neighborhood ancestry definition).

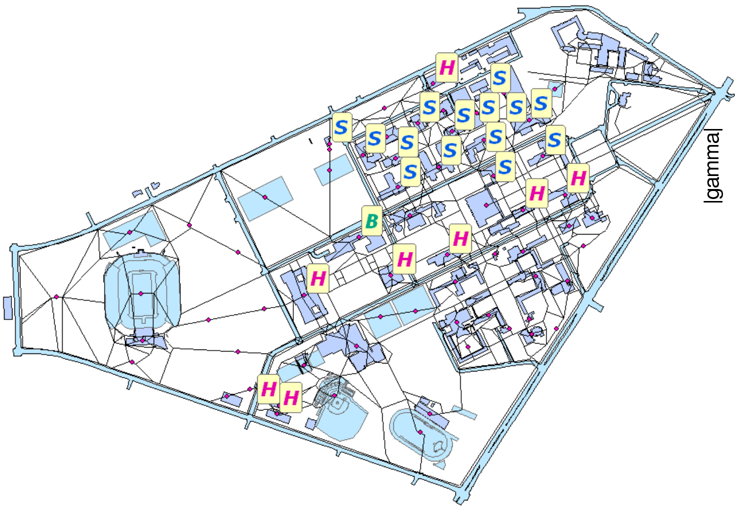

Figure 11: Rice University Map with assigned Activity Types

4.1 Model and algorithm

Let $i = 1, \ldots, N$; $t = 1, \ldots, T$ and $k = 0, \ldots, K$. Then let $y_{itk} = \theta_i \phi_t \gamma_k + \epsilon_{itk}$ where ϵ_{itk} is $N(0, \sigma^2)$. Note that y_{itk} represents the node degree ancestry of node i at time t for lag k, and that y_{it0} is simply the node degree for node i at time t. We formulate least squares estimates for θ, ϕ, and γ as the following:

$$\hat{\theta}_i = \frac{\sum_{t,k} \phi_t \gamma_k y_{itk}}{\sum_{t,k} (\phi_t \gamma_k)^2} \tag{3}$$

$$\hat{\phi}_t = \frac{\sum_{i,k} \theta_i \gamma_k y_{itk}}{\sum_{i,k} (\theta_i \gamma_k)^2} \tag{4}$$

$$\hat{\gamma}_k = \frac{\sum_{i,t} \theta_i \phi_t y_{itk}}{\sum_{i,t} (\theta_i \phi_t)^2} \tag{5}$$

The algorithm used to estimate is nearly identical to that used in Section *3.1.1*, save that we now renormalize both θ_i and ϕ_t. We do not renormalize γ_k for the same identifiability reasons as given in Section *3.1.1*. Intuitively, since node degree ancestry is a decreasing sequence we expect γ_k to be decreasing in k. However, this procedure allows us to discover the main dynamics of how γ_k decreases for each model.

For purposes of illustration and comparison, we first give 2 examples of computed γ_k for Random Waypoint, and Pedsims, where $k = 0, 1, \ldots, 5$. The θ_i and ϕ_t are produced nearly identical to those plots shown above, and are not reproduced here. In Section 5 we apply our method to real data from [10].

4.2 Random Waypoint Example for γ

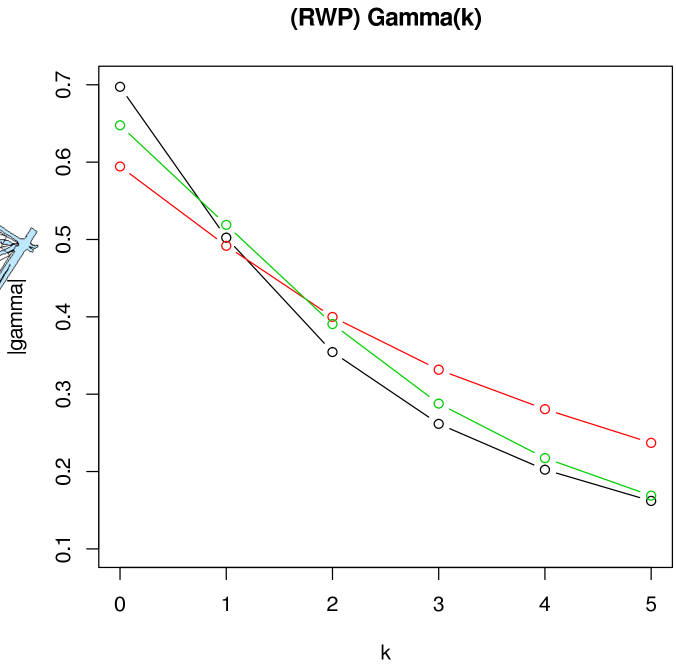

Figure 12: Neighborhood Ancestry trends for RWP

As expected, the first three principal "ancestories" are similar, indicating (once again) no large differences in behavior by the nodes (see Figure 12). The shape is somewhat linear, with a negative slope, implying that the ancestry of each node decays at a constant rate. This is an interesting result, since we can now compare mobility models against Random Waypoint with respect to γ_k. The slope of this line is an inidicator for this, which should change depending on many initial mobility parameters, namely the speed at which the nodes move. Classifying which parameters have an impact on the decay rate remains an open question to our knowledge.

4.3 Pedsims Example for γ

Once again, it's not surprising that structure is be found. However, γ_k here indicates 2 main patterns, a steady decrease in ancestry and a constant ancestry. This is due again to the nature of Pedsims, which our method picks up quite nicely. In the current version of Pedsims (a univeristy model), students move between activity locations according to a particular schedule. The decreasing neighborhood ancestry then is explaining the regular moving about of the students, while the constant ancestry is capturing the ancestry patterns of the times these individuals remain at a location for some time (class or otherwise). One can conceivably model this as a two-state stochastic process and perhaps design routing protocols that are specific to this application.

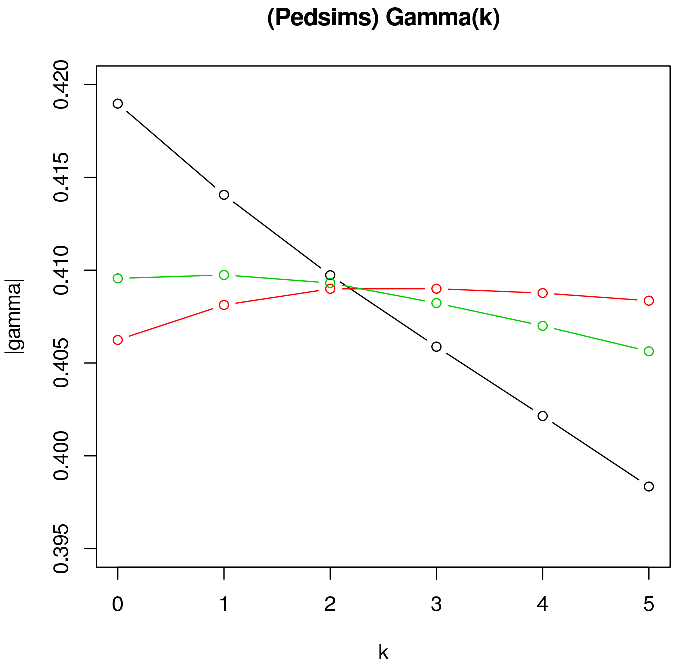

Figure 13: Neighborhood Ancestry trends for Pedsims

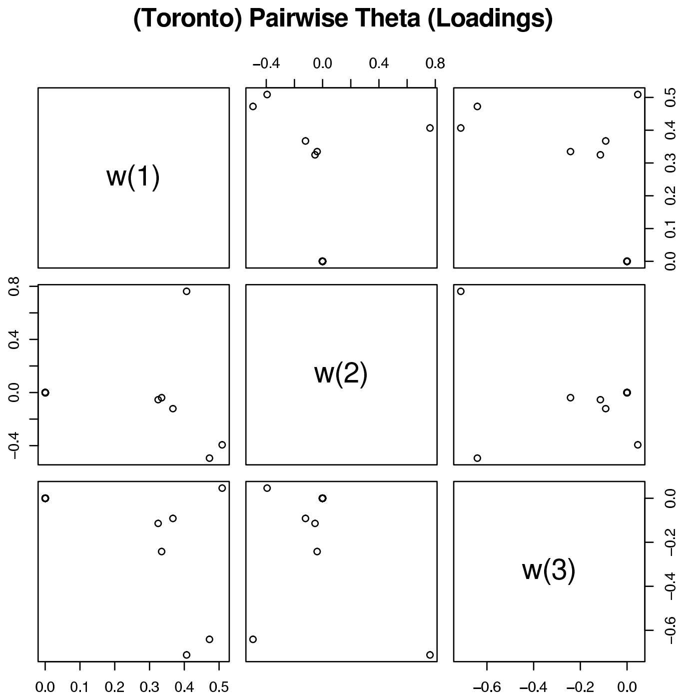

Figure 14: (Toronto) θ pairs

5. REAL DATA EXAMPLE

The data provided by [10] is more sparse (in terms of missing data) than the simulated Pedsims data. It was collected over the course of several days with many hours (and days) of observations missing, for 21 devices handed out to Computer Science (CS) and Electrical and Computer Engineering (ECE) students.

Figure (14) shows shows a modest clustering of the mobile devices with all CS students clustered together and the ECE students scattered about (labeling not shown). This is quite similar to the Pedsims data, as once again major field of study appears to be the chief cause of the clustering. The principal time series are shown for the first 1200 seconds only, as the remainder of the series is flat. Differences between nodes based on activity is not clear. We hypothesize that high activity near the beginning of the series between disjoint time periods is indicative of "awareness bias", where the student may have subconsciously charged and activated the device before class (perhaps as required by the project research). Keeping the device on afterwards appeared to show only small incidental contact with other mobile nodes.

In general we would hope the patterns in the degree ancestry would be similar to Pedsims in Section (4.3), and there are similarities. The first 2 components have a sharply decreasing trend, indicative of the device hearing other devices in the first timestep and not in subsequent steps. The 3rd component however is similar to Pedsims, in that it appears to have an element of constant ancestry behavior.

One problem with this dataset is the number of devices, as this strictly pushes the first few components of γ_k to have a sharp decreasing pattern. Another is the length of recorded time where nothing was observed. Clearly the method does well with missing data in Pedsims, but the missing data in this dataset outscales most of the key information at hand. How to sample of the timeseries to mine the important parts where information was recorded is an issue that needs resolving.

6. DISCUSSION

The problem of computing PCA on sparse matricies is not a new, but the need for it in MANET research appears to be. Much research is being poured into the construction of realistic mobility models to assist with design and testing of ad hoc networks, and invariably connected with realistic mobility is missing data that occurs when observing users/objects/mobile nodes move in and out of the observation range. We agree that realistic mobility models are needed (and should be used) and it is our hope that we have motivated the need for analysis of such data structures, and shed light on a tool that can be flexible and useful for any metric desired.

We took an additional step, however, and showed that such a tool can be expanded a single dimension with good results. The number of iterations necessary for convergence was relatively small (25–50) and our estimates appear to make sense with knowledge of the model. We also note that too many iterations can lead to oversmoothing of the components, In one case, analysis of the Pedsims data with over 150 iterations smoothed out most of the internal patterns identified. The ability to correctly gauge the number of iterations is both a feature and a detriment to the tool. The results of the method remain sensitive to mobility parameters, namely speed of the nodes and size of the space.

Some implications of our work are the following: Consider Pedsims. Using PCA for missing data one can feasibly collect mobility data and attempt to identify node degree patterns in time. Knowledge of the general time series can assist engineers to build networks with capacity that is adaptable to the temporal patterns. With knowledge of individual

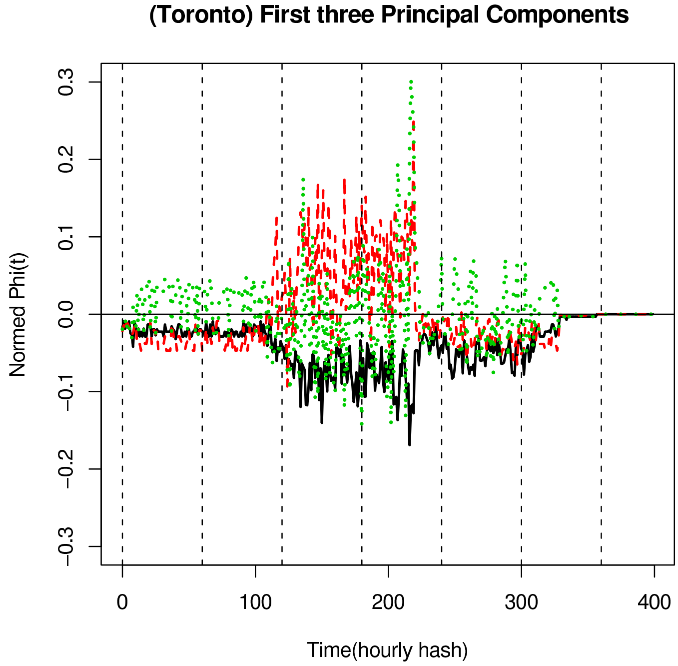

Figure 15: (Toronto) $\phi(t)$

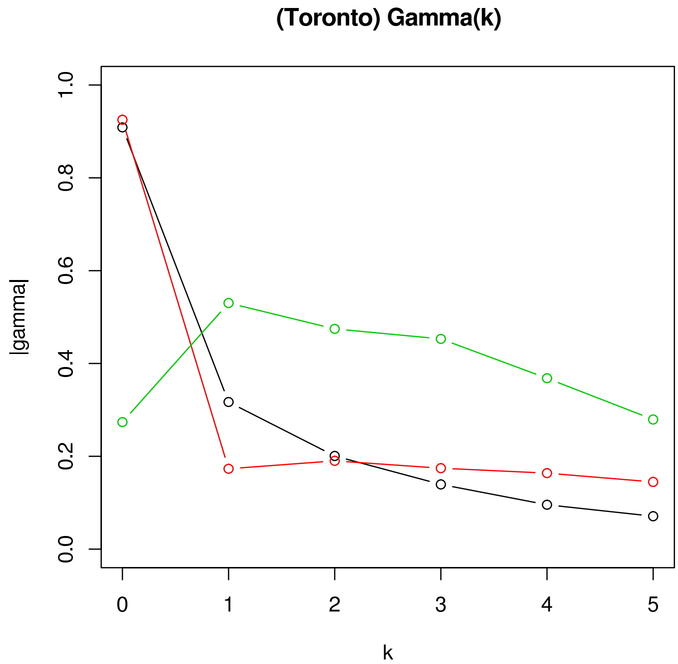

Figure 16: Neighborhood Ancestry trends for Toronto data

clustering patterns, one could generate a more intelligent routing protocol which would perhaps need to flood/explore less often or only under certain circumstances. Finally, we suggest that knowledge of the neighborhood ancestry is a good indicator of "routing-table-decay", being the rate at which new routes need to be discovered. If this is known and is believed to be stationary one can conceivably determine how much power a device requires to satisfy the needs of the group in question, since power is directly related to the amount of flooding the mobile device must do. Let the reader note that we only used node degree as an example metric, as other metrics may be applied for this methodology with differing implications. We leave it to the community to continue the exploration.

The obvious question remains: can we continue to increase the number of dimensions (in the sense of the multiplicative regression model we offered) and still obtain reasonable estimates of overall trends in each dimension? What are the numerical stability issues? More pragmatically, can we even imagine data structures in networking that make sense in higher dimensions? Being mindful of the so-called "curse of dimensionality", we seek to answer these questions in future research.

Acknowledgements: Financial support comes in part from NSF, grant number ANI-0338856, Texas ATP, project number 003604- 0036-2003, and Los Alamos National Laboratory (CCS-5).

We would like to thank Jing Su from the University of Toronto for generously providing the Toronto dataset.

7. REFERENCES

[1] A. d'Aspremont, L. Ghaoui, M. Jordan, G. Lanckriet. A direct formulation for sparse PCA using semidefinite programming. June 2004.

[2] M. W. Berry. Large Scale Sparse Singular Value Computations. International Journal of Supercomputing Applications. 1992, 6:13–49

[3] J. Boudec, M. Vojnovic. Perfect Simulation and Stationarity of a Class of Mobility Models. Infocom 2005.

[4] T. Camp, J. Boleng, V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, 2(5):483 502, 2002.

[5] C. Chennubhotla, A. Jepson. Sparse PCA: Extracting Multi-scale Structure from Data, IEEE International Conference on Computer Vision, 2001.

[6] H. Flores, S. Eidenbenz, N. Hengartner, R. Riedi and J. Smith. PedSims: Building towards Realism of Mobility Models for Wireless Networks. *Technical Report* Los Alamos National Laboratory, 2006.

[7] W. Hsu, A. Helmy. Analyzing Principal Characteristics of User Association Patterns and Eigen-behavior in Wireless LAN Traces. 2006.

[8] A. Jardosh, E. Belding-Royer, K. Almeroth, S. Suri. Real-world Environment Models for Mobile Network Evaluation. Journal on Selected Areas in Communications special issue on Wireless Ad hoc Networks, March 2005.

[9] J. E. Jackson. A User's Guide to Principal Component Analysis. Wiley Series in Probability and Statistics, 1991.

[10] J. Su, A. Chin, A. Popivanova, A. Goel, E. de Lara. User Mobility for Opportunistic Ad-Hoc Networking. Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04) - Volume 00. 2004.

[11] H.Y.Shum, K. Ikeuchi, R. Reddy. Principal Component Analysis with Missing Data and Its Application to Polyhedral Object Modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence. September 1995 (Vol. 17, No. 9) pp. 854-867.

[12] T.Wiberg. Computation of Principle Component Analysis when data are missing. Proceedings of the Second Symposium on Computational Statistics, pp. 229–236, 1976.

[13] H. Zou, T. Hastie, R. Tibshirani. Sparse Principal Component Analysis. 2004.

[14] Q. Zheng, X. Hong, J. Liu, "An Agenda Based Mobility Model21" *anss* , pp. 188-195, 2006.