Group Detection in Mobility Traces

Yung-Chih Chen, Elisha Rosensweig, Jim Kurose, and Don Towsley
University of Massachusetts Amherst
{yungchih, elisha, kurose, towsley}@cs.umass.edu

***Abstract*—Analysis and modeling of wireless networks greatly depend on understanding the structure of underlying mobile nodes. In this paper we present two clustering algorithms to determine the number of groups and their identities: k-means chain and spectral clustering. Different from traditional k-means clustering, k-means chain can identify the number of groups in dynamic graphs, and the chaining process can also keep track of group trajectories over the entire trace. The second approach uses spectral clustering, which measures the similarities between each node pair to group nodes of similar behaviors. We show that critical information of a mobility trace, such as the number of groups and group members, can be precisely extracted with little or no prior knowledge of the properties of a trace.**

I. Introduction

Analysis and modeling of wireless networks greatly depends on understanding the underlying models that mobile nodes follow. One type of mobility model is that of group-oriented or leader-oriented movement, in which the set of moving nodes, or units, can be divided into several disjoint sub-sets where the movement of units within each sub-set is highly correlated. By separating out the groups from each other, more precise analysis of intra-group behavior can be achieved, and inter-group mobility correlation can be observed and analyzed. In this work we present several tools for analyzing mobility traces and determining the number and identity of these groups. We describe these tools in detail and compare their performance over both real and synthetic traces.

II. Clustering Algorithms

k-means clustering is a method used to partitioning n observations into k different clusters. With a given number k and a set of objects $(x_1, x_2, ..., x_n)$, k-means clustering targets to partition the n objects into k different sets: $S = (S_1, S_2, \ldots, S_k)$ so as to minimize the sum of squares of objects to the mean μ_i of each cluster S_i s.t.

$$\arg\min_{s} \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

The mean values μ_i can be thought of as points in the sample space, and are reffered to commonly as the cluster *centroids*. In this section we focus on determining the first parameter (k), which is more elusive and fundamental, while the others (μ_i) are discussed in the implementation and results section. We took two major approaches to determine the most likely candidate for the number of clusters k:

A. K-means Chains

1) Unit Loyalty: Here we introduce the concept of loyalty in group mobility patterns. Presumably, when a soldier is recruited to a troop or group, the soldier will stick to that group regardless of changes in the environment (i.e., the soldier should remain *loyal* to the leader over the entire maneuver). Thus, consider an algorithm that would apply k-means to each snapshot in the trace. By enumerating the number of times each unit changes groups between snapshots for a set of possible values of k, we can imagine that when $k = 1$, each identity doesn't change group over the simulation. When k increases, the frequency of changing group label for a specific unit will increase until k equals to the total number of nodes in the network.

2) Chaining Process and Group Trajectory: We refer to such sequences of k-means as a "k-means chain", as the behavior at every snapshot is determined by the previous one, creating a dependency chain between snapshots [1]. Using this approach, we use the following measure for selecting the best k: count the number of times each unit changes its *label*, and select the k that locally minimizes this value [2]. We refer to the number of label changes a unit undergoes as the *loyalty score* of that unit. While we are chaining group behavior between two consecutive snapshots, we can also keep track of the movement of group centroids over the entire period. By doing this, k-means chain not only determines the number of the groups of the most likelihood, but also shows the trajectories of each identified group.

B. Spectral Clustering

Spectral clustering methods [2] begin with a set of (possibly arbitrary) objects $x_1, \ldots, x_n$ for which there exists a pairwise similarity (or weight) matrix $S = (s_{ij})$, $i, j = 1 \ldots n$, which represents the similarity between every pair of nodes i and j.

Taking a Laplacian L of S (normalized or unnormalized), the representation of the abstract data points x_i are then changed to points $y_i \in \mathbb{R}^k$, where k is the given number

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

[1] This method can be thought of as expanding the classic k-means from dealing solely with static data to dynamic, sequential data as well.

[2] For this k, the main cause for group changes will be when groups are close together and units on the edge of the group can be mislabeled.

of groups to be detected. The algorithm then performs k-means over the rows of the k eigenvectors with the smallest eigenvalue over L. The index of each row indicates the index of the corresponding node, and clustered rows here indicate clustered nodes in the original data-set.

III. Trace Data

We collect traces following the *Reference Group Mobility Model (RPGM)* [3] from a mobility trace generator, implemented in the *IMPORTANT* tool [4]. In this model, each group has a logical center (i.e., the leader) whose motion defines the behavior and the trajectory of the whole group.

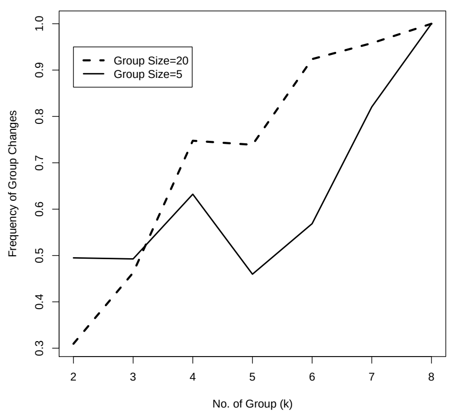

Fig. 1. 5 groups (of size 5 and 20) v.s. different k values. Loyalty function suggests local minimal at $k = 5$.

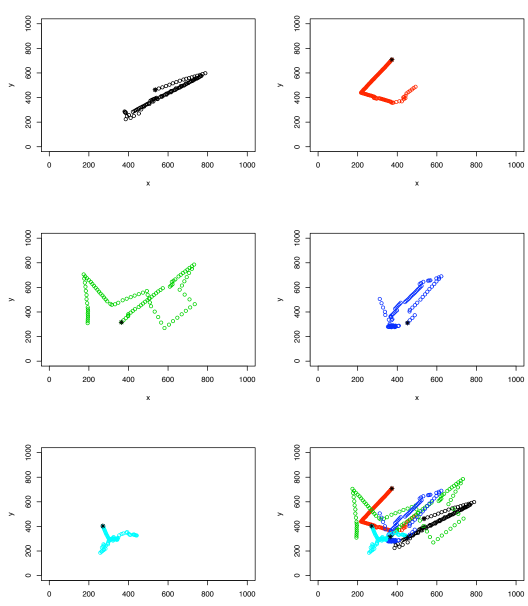

Fig. 2. 5 group trajectories (and the superposition at the bottom right)

A. K-means Chains and Group Trajectory

Our experiments, performed over synthetic traces, indicate that units are most "loyal" to their previous label when the correct value of k is used. Fig.1 shows this loyalty function over different values of k where the trace has 5 distinct groups, and as we can see, minimizes for $k = 5$.

After applying the loyalty function to find the correct k, we then follow the chaining process described in section II-A to separate all the five groups apart. Through out the chaining process, each group centroid's trajectory can be identified as well. The other five plots in Fig.2 are the trajectories of the five groups.

B. Spectral Clustering

In order to select the best k using spectral clustering, we ran for each trace a series of 50 stability checks, and determined that the correct k is the one for which the highest percentage of these tests were passed. Fig.3 shows the *support* for a specific value k is therefore the fraction of successful tests.

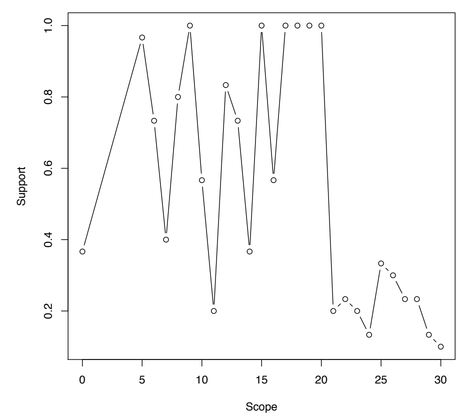

Fig. 3. For the correct number of groups (5), we plot here the support given by the stability tests for different values of h in the h-nearest neighbors version of the similarity matrix.

IV. Summary and Conclusions

In this paper we set out to determine the groups as they emerge from a mobility trace. Our goals were two-fold: 1) to correctly develop a measurable method for group detection in the trace and the group membership as well, and 2) to identify the group trajectory followed by the group. Once groups are discerned, one can proceed to determine the group leader, or the relationship between different groups.

References

[1] J.B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *The 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967*

[2] U. von Luxburg, "A Tutorial on Spectral Clustering," *Tech. Report No. TR-149, Max Planck Institute for Biological Cybernetics, 2006*

[3] X. Hong, M. Gerla, G. Pei, C-C. Chiang, "A Group Mobility Model for Ad Hoc Wireless Networks," *IEEE MSWiM, August 1999*

[4] F. Bai, N. Sadagopan, A. Helmy, "The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks," *Ad Hoc Networks, Elsevier 2003*