Eigenbehaviors: Identifying Structure in Routine

Citation: N. Eagle and A. Pentland, "Eigenbehaviors: identifying structure in routine," Behavioral Ecology and Sociobiology, vol. 63, May. 2009, pp. 1057-1066.

As Published: http://dx.doi.org/10.1007/s00265-009-0739-0

Publisher: Springer Berlin Heidelberg

Persistent URL: http://hdl.handle.net/1721.1/49446

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Eigenbehaviors: Identifying Structure in Routine*

Nathan Eagle[1,2] and Alex (Sandy) Pentland[1]

[1]MIT Media Laboratory, Massachusetts Institute of Technology,

E15-383, 20 Ames St., Cambridge, MA 02139, USA

[2]The Santa Fe Institute

1399 Hyde Park Rd, Santa Fe, NM 87501, USA

Current word count (exc. Refs etc): 4875

Corresponding Author: Nathan Eagle, The Santa Fe Institute,

1399 Hyde Park Rd, Santa Fe, NM 87501. Email: nathan@mit.edu

23

* This contribution is part of the special issue "Social Networks: new perspectives" (guest editors: J Krause, D Lusseau and R James)

Abstract

Longitudinal behavioral data generally contains a significant amount of structure. In this work we identify the structure inherent in daily behavior with models that can accurately analyze, predict and cluster multimodal data from individuals and communities within the social network of a population. We represent this behavioral structure by the principal components of the complete behavioral dataset, a set of characteristic vectors we have termed eigenbehaviors. In our model, an individual's behavior over a specific day can be approximated by a weighted sum of his or her primary eigenbehaviors. When these weights are calculated halfway through a day, they can be used to predict the day's remaining behaviors with 79% accuracy for our test subjects. Additionally, we demonstrate the potential for this dimensionality reduction technique to infer community affiliations within the subjects' social network by clustering individuals into a "behavior space" spanned by a set of their aggregate eigenbehaviors. These behavior spaces make it possible to determine the behavioral similarity between both individuals and groups, enabling 96% classification accuracy of community affiliations within the population-level social network. Additionally, the distance between individuals in the behavior space can be used as an estimate for relational ties such as friendship, suggesting strong behavioral homophily amongst the subjects. This approach capitalizes on the large amount of rich data previously captured during the Reality Mining study from mobile phones continuously logging location, proximate phones, and communication of 100 subjects at MIT over the course of nine months. As wearable sensors continue to generate these types of rich, longitudinal datasets, dimensionality reduction techniques such as eigenbehaviors will play an increasingly important role in behavioral research.

Introduction

While discrete observations of an individual's idiosyncratic behavior can appear almost random, typically there are repeating and easily identifiable routines in every person's life. These patterns become more apparent when the behavior is temporally, spatially, and socially contextualized. However, building models of long-term behavior has been hampered due to the lack of contextualized behavioral data. Additionally, traditional Markov models work well for specific set of behaviors, but have difficulty incorporating temporal patterns across different timescales (Clarkson 2002). We present a new methodology for identifying the repeating structures underlying behavior. These structures are represented by *eigenbehaviors*, the principal components of an individual's behavioral dataset.

To capture these characteristic behaviors, we compute the principal components of an individual's behavioral data. The principal components are a set of vectors that span a 'behavior space' and characterize the behavioral variation between each day. These eigenbehaviors are the eigenvectors of the covariance matrix of behavior data; the heavily weighted vectors generally represent a type of repeated behavior, such as sleeping in late and going out on the town. A linear combination of an individual's eigenbehaviors can accurately reconstruct the behavior from each day in the data. However, we show that our subjects' behavior can be approximated with 90% accuracy using only the six primary eigenbehaviors – the ones that have the largest eigenvalues and account for the most variance. By providing this type of behavioral caricature, it is possible for the primary eigenbehaviors to be used to accurately predict an individual's subsequent behavior. We subsequently show how eigenbehaviors can be applied not only to individual behavior, but also be used to characterize the behavior of communities within the population's social network. Particular groups of friends can have their own collective 'behavior space' which corresponds to the common behaviors of the community. How well these behavior spaces approximate an individual's behavior depends on how the individual is similar to others in her social network. Measuring the Euclidean distance between an individual's behavior and the behavior space of a specific community within the social network can be used to identify affiliations, relationships, and similarity between individuals.

There has been an extensive number of research efforts focused on modeling individual and group behaviors. Due to the breadth of these efforts, we will be limited here to providing only a sample of related research projects. Some closely related work in the Computer Supported Collaborative Work (CSCW) community comes from Begole et al’s techniques for “rhythm modeling” within the workplace. Through analysis of the computer usage of workgroup members, Begole et al demonstrated the potential to extract patterns in behavior of both individuals and teams (Begole et. al 2003). Although primarily used for location-based applications, electronic badges can also generate rich data on individual behavior within a workplace. The exposed manner in which they are worn allows line-of-sight sensors, such as infrared (IR), to detect face-to-face interactions. Some of the earlier badge work to sense human behavior was done in the 80's and early 90's at Olivetti Labs (Want et. al 1992). Developments in ultrasound tracking have greatly improved the ability to localize the badge, enabling a wide range of just-in-time information applications (Schilit et. al 1993; Addlesee et. al 2001). Fogarty et al. expands this work by using low level sensor data to establish extremely accurate estimates of human interruptibility (Fogarty et. al 2005).

Outside the office, GPS has been used for location detection and classification (Asbrook and Starner 2003; Liao et. al 2004; Wolf et. al 2001), but the line-of-sight requirements generally prohibit it from working indoors. As an alternate approach, there has been a significant amount of literature regarding correlating cell tower ID with a user's location (Bar-Noy and Kessler 1993; Bhattacharya and Das 1999; Kim and Lee 1996). Laasonen et al. describe a method of inferring the significant locations from the cell towers by calculating graph metrics from the adjacency matrix formed by proximate towers. They were able to show reasonable route recognition rates, and most importantly succeeded in running their algorithms directly on the mobile phone (Laasonen et al 2004). In the activity and pattern recognition communities, there has been a variety of work using techniques to estimate an individual's location and projected trajectory given a variety of sensor data such as GPS, wifi base-station positioning, and accelerometer data. Hightower and Borriello along with Patterson et al., among others, have demonstrated the potential of particle filters for route recognition (Hightower and Borriello 2004; Liao et al 2004; Patterson et al 2003).

In machine vision and computer graphics, eigenrepresentations have become one of the standard techniques for many tasks. While behavior is perhaps not as characteristic of an individual as a face, many analogies hold between the analysis of an individual's behavior and his facial features. Just as digital imaging created a wealth of data to train and test facial analysis tools, the explosive growth of mobile phones is beginning to enable much more comprehensive computational models of complex human behavior. Eigendecomposition is used in face and object recognition (Turk and Pentland 1991), shape and motion description (Pentland and Sclaroff 1991), and data interpolation (Pentland 1992), and computer animation (Pentland and Williams 1989). More recently it has been used in a wide variety of robotic and control applications.

Methods

To apply eigendecomposition for behavior and social network analysis, a large repository of behavioral data is necessary. In this paper we make use of the publically available Reality Mining dataset representing the behavior of 100 subjects at MIT during the 2004-2005 academic year (Eagle and Pentlad 2006). Seventy-five of the subjects were either students or faculty in the same laboratory, while the remaining twenty-five were incoming students at the business school adjacent to the laboratory. Of the seventy-five students and staff at the lab, twenty were incoming masters students and five were incoming freshman. The data were collected using one hundred Nokia 6600 smart phones pre-installed with a version of the Context application from the University of Helsinki (Raento et. al 2005). The information collected included call logs, Bluetooth devices in proximity, cell tower IDs, application usage, and phone status (such as charging and idle). The study generated approximately 400,000 hours of data on subjects' location, proximity, communication and device usage behavior.

The collection of deeply personal human behavioral data raises justifiable concerns over privacy. While these concerns are legitimate and should be explored, the dataset we are using was collected during a social science experiment, conducted with human subject approval and consent of the subjects. Additionally, these techniques for extracting the underlying structure inherent within behavioral data are not only applicable to human populations. Eigenbehaviors are suitable for analysis of any regularly sampled behavioral data, making it also a potential analysis tool for longitudinal studies of animal behavior, where concerns about privacy are greatly reduced (Krause et. al 2009).

Finally, this paper will not make the claim that the subjects in the Reality Mining study are a representative sample of society. However, regularity in behavior is not an exclusive trait of people at MIT. For many people, weekdays consist of leaving home in the morning, traveling to work, breaking for lunch, and returning home in the evening. People's daily routines are typically coupled with routines across other temporal scales, such as going out on the town with friends on Saturday nights, or spending time with family during the December holidays. Animals exhibit similar behavior patterns, both on a daily and seasonal cycle. The remainder of this paper will be focusing on a particular technique to quantify these universal patterns in the behavior of individuals and communities within a social network.

While we have successfully applied our eigenbehavior technique to a wide range of multimodal data, for purposes of clarity in this section we will only focus on temporal location data. For this example, we characterize person *I* by data shown in Figure 1 as *B(x,y)*, a two-dimensional *D* by 24 array of location information, where *D* is the total number of days that person *I* has been in the study. *B* contains *n* labels corresponding to behavior, where in our case these labels are *{Home, Elsewhere, Work, No Signal, Off}*. It has been previously shown that these labels were generated with a conditioned Hidden Markov Model with over 95% accuracy (Eagle and Pentland 2006), and while there still is noise in the signal, for our purposes we'll take them as ground truth. To perform the analysis, we convert *B* into *B′,* a *D* by *H (*where *H* is *24*n)* array of binary values, shown in Figure 1. Γ_i is row *i* of *B′* and represents an individual's behavior over day *i*; Γ_i can be represented by a single point in an *H*-dimensional space. A set of D days can then be described as a collection of points in this large space.

Due to the significant amount of similar structure in most people's lives, days are not distributed randomly though this large space. Rather, they are clustered, allowing the individual to be described by a relatively low dimensional 'behavior space'. This space is defined by a subset of vectors of dimension *H* that can best characterize the distribution of behaviors and are referred to as the primary eigenbehaviors. The top three eigenbehaviors that characterize the individual shown in Figure 1 are plotted in Figure 2. The first eigenbehavior corresponds to either a normal day or a day spent traveling (depending on whether the associated weight is positive or negative). The second eigenbehavior has a corresponding weight that is positive on weekends and negative on weekdays, analogous to the characteristic behavior of sleeping in and spending that night out in a location besides home or work. The third eigenbehavior is emphasized when the subject is in locations with poor phone reception.

Results

Eigenbehaviors for Individuals

For each subject, the Reality Mining data set provides us with a set of days' behaviors, $\Gamma_1, \Gamma_2, \Gamma_3 \ldots \Gamma_D$, for a total of D days, where an individual day's behavior vector, Γ_i, has H dimensions. Following the same notation as Turk and Pentland, the average behavior of the individual is $\Psi = \frac{1}{D}\sum_{n=1}^{D}\Gamma_n$. And $\Phi_i = \Gamma_i - \Psi$ is the deviation of an individual day from the mean. Principal components analysis is subsequently performed on these vectors generating a set of H orthonormal vectors, u, which best describes the distribution of the set of behavior data when linearly combined with their respective scalar values, λ. These vectors and their corresponding scalars are the eigenvectors and eigenvalues of the covariance matrix of Φ, the set's deviation from the mean.

$$\begin{aligned} C &= \frac{1}{H}\sum_{n=1}^{H}\Phi_n\Phi_n^T \\ &= AA^T \end{aligned}$$

where the matrix $A = [\Phi_1, \Phi_2, \Phi_3, \ldots \Phi_M]$. Each eigenbehavior can be ranked by the total amount of variance it accounts for in the data, which is essentially the associated eigenvalue. The vectors with the highest eigenvalues are considered an individual's primary eigenbehaviors. The next section will discuss how these primary eigenbehaviors can be used for behavioral data reconstruction and prediction.

An individual's primary eigenbehaviors represent a space upon which all of his days can be projected with differing levels of accuracy. Figure 3 shows the projection of each day onto spaces created using an increasing number of these primary eigenbehaviors. It can be seen that while the reconstruction of each day using 40 eigenbehaviors for this particular subject nearly perfectly matches the original data, six eigenbehaviors captures a significant portion of the variance in the individual's behavior. Figure 4 shows the accuracy of representing behavior using a varying number of eigenbehaviors for the three different groups of subjects in the Reality Mining study. It is interesting to note that the space formed by the six primary eigenbehaviors describes individuals within the business school community of the social network with 90% reconstruction accuracy, but the senior lab students with 96% accuracy. This leads us to the conclusion that senior lab students exhibit more behavioral regularity than their business school counterparts.

While there are many techniques for creating predictive models that can generate a sequence of future data given training, eigendecomposition differentiates itself in an important way. Although many of life's patterns can be modeled as a Markov process, whereby the future state depends on the current state and observational data, these types of models have difficulty capturing correlations that span beyond several time slices. For many subjects, sleeping late in the morning is coupled in the same eigenbehavior with going out that evening – a hard pattern to recognize when using traditional models, but one that is highlighted when generating an individual's characteristic behavior spaces.

Figure 4 shows that the top six primary eigenbehaviors provide a characteristic behavior space from which an individual deviates less than 10% of the time. When these six eigenbehaviors are calculated for an individual, it becomes possible to infer the projection of an entire day using only information from a portion of that day. We use these approximations to develop predictions of an individual's subsequent behavior. To test this concept, for each subject we calculated a behavior space using the individual's six primary eigenbehaviors and weights generated from the first twelve hours of a subject's day. Through the linear combination of these weights and the subject's primary eigenbehaviors, a 12-element vector is created containing one of three location states (home, work, elsewhere). Each element in the vector corresponds to the predicted location of the subject for the subsequent hours from noon to midnight. Figure 5 shows the distribution of accuracy scores for the subjects when the sequence of 12 hours is compared with the subject's actual location over the same 12 hours.

Eigenbehaviors for Social Networks

In the previous section we have demonstrated that we can use data from Bluetooth-enabled mobile phones to discover a great deal about an individual's patterns of activities by reducing these complex behaviors to a set of principal components, or eigenbehaviors, characteristic of the individual. In this section we will demonstrate the possibility of inferring the relationships and community affiliations within the social network of the population based on a comparison of these eigenbehaviors.

The social network of the subjects in the Reality Mining study has a high amount of clustering based on affiliation, as shown in Figure 6. It is reasonable to assume that each of these different groups of subjects (Sloan business school students, Media Lab incoming students, Media Lab senior students, and MIT staff) have characteristic behaviors associated with the community affiliation. It is possible now to identify the eigenbehaviors of these particular communities within the social network and project individuals onto this behavior space. How well the community's behavior space explains an individual's behavior, as measured by the Eucleadean distance between the individual and the principal components of the community's behavior space can then be used to infer the individual's affiliation. Additionally, we demonstrate that the distance between a pair of subjects within the community is proportional to the probability the two individuals are connected within the friendship network.

The mathematics behind applying the eigenbehavior technique to a community of M actors is identical to that described in Section 2, with the exception that several of the variables have different interpretations. We now use a matrix B with each row corresponding to the average behavior of a particular individual in the community. After a similar transformation to B', a matrix of M by H, it becomes possible to generate eigenbehaviors of the community as a whole. The primary eigenbehaviors correspond to the community's characteristic behaviors.

While we later will show results that incorporate a variety of data including location, phone usage and people in proximity into the community behavior space, for explanative purposes, we will show data related to solely Bluetooth proximity events for the three main groups of subjects: incoming business school students, incoming lab students, and senior lab students. Figure 7 shows the mean behaviors for each group, Ψ_j, while Figure 8 depicts the top three eigenbehaviors $[u_1^j, u_2^j, u_3^j]$ of each group.

As expected, the top eigenvector in each of the groups closely corresponds to the mean. For individuals within the business school community, there is particular emphasis during the school's coffee breaks at 10:30. Besides this emphasis, the other pattern is simply reflective of the standard course times (nine until noon, a lunch break, and the subsequently afternoon courses). The lab students have less of an enforced structure on their day. While the entire group of incoming lab students is taking courses, along with approximately half of the senior students, these courses can be selected by the students from anywhere in the institution and typically are not attended by many other subjects. However, each of the lab students has an office within the lab and typically works from there when not in class. While the two groups of lab students share virtually identical principal eigenbehavior, the secondary eigenbehaviors are more telling about the differences. It is common knowledge around the lab that incoming students tend to get overwhelmed by over-commitments to coursework and research leading to late nights at the workplace. This characteristic is emphasized from the group's second and third eigenbehaviors with an emphasis from 20:00 to 2:00.

When a community's behavior space is created from the aggregate behavior of its individual members, it becomes possible to determine the similarity of the members by identifying how accurately their behavior can be approximated by the community's primary eigenbehaviors. Because the Reality Mining dataset contains data for both incoming and senior students, it is possible to verify the onset of concordance between the incoming lab students and the rest of the laboratory. Likewise it is possible to distinguish communities by their aggregate behavior, such as business school students and engineering students. An individual's behavior (Γ) can be projected onto the j community's behavior space through the following transformation.

$$\omega_k^j = u_k^j \left(\Gamma - \Psi_j\right)$$

for k=1,..., H and Ψ_j is the mean behavior of the community. Ψ_j for Bluetooth encounters of senior lab students, incoming lab students, and business school students is shown in Figure 7.

These weights form a vector $\Omega_j^T = \left[\omega_1^j, \omega_2^j, \omega_3^j, \ldots \omega_{M'}^j\right]$ which is the optimal weighting scheme to get the new behavior as close as possible to the behavior space. Each element in the vector gives a scalar value corresponding to the amount of emphasis to place on its respective eigenbehavior when reconstructing the original behavior Γ. By treating the eigenbehaviors as a set of basis behaviors, the vector Ω^T, can be used to determine which person k the individual is most similar to in a particular community, j. We follow the method of Turk and Pentland by using Euclidean distance as our metric for describing similarity.

$$\varepsilon_{j_k}^2 = \left\|\Omega^j - \Omega_k^j\right\|^2$$

where Ω_k^j are the reconstruction weights for the *k*th person in community *j*. Figure 9 shows values for ε_j, the distance between one business school student and other subjects. This method can also be applied to data from a single individual to determine which days are most like the ongoing one.

Instead of comparing one individual to another, it is also possible to determine how much an individual 'fits in' with the community as a whole by determining the distance ε as the difference between the individual's projection onto the behavior space of a community and the individual's original behavior. We again use Euclidian distance to calculate the difference between the mean-adjusted behavior, $\Phi^j = \Gamma - \Psi^j$, and its projection onto the community's behavior space $\Phi_b^j = \sum_{i=1}^{M_j'} \omega_i^j u_i^j$.

$$\varepsilon_j^2 = \left\| \Phi^j - \Phi_b^j \right\|^2$$

When determining the affiliation of an individual, there can be four possible outcomes, as shown on Figure 10. The dark gray plane represents the community behavior space, containing any set of behaviors that would constitute being part of the community. The first option has the input behavior on the behavior space as well as proximate to other individuals, Ω_{j_3}, within the behavior space. The second example can be approximated accurately by the behavior space, but there are no other individuals in the same area of the space. Input three appears to have something in common with some members in the community's behavior space, however contains behavioral elements that cannot be reconciled within the behavior space. Lastly, four is a disparate input neither near the behavior space nor any individual in the space.

Until now, we have been focusing on analysis of Bluetooth or location data independently, but this technique enables us to aggregate multimodal datasets. Instead of limiting a community to only one behavior space, for our affiliation classification we generate a set of primary eigenbehaviors for each type of data captured. This enables us to determine every group's Bluetooth, location and phone usage behavior space. When these spaces are computed, it is subsequently possible to calculate each individual's Euclidian distance from each space. Figure 11 shows the distances for each subject from the three business school behavior spaces. We used cross validation to prevent the test subject's data from contributing to the generated behavior space, and were able to classify whether each subject was a member of the business school community with 96% accuracy.

Lastly, the projected clustering of individual subjects onto the behavior space shown in Figure 11 has an additional interesting characteristic beyond affiliation inference. By simply measuring the distance between two individuals within this behavior space, it becomes possible to estimate the probability the pair is connected within the social network of the population. Figure 12 shows that the probability of friendship tails off dramatically as distance increases, until it converges on a steady-state probability of friendship that appears to be irrespective of the behavioral differences between the pair. This relationship follows a distribution qualitatively similar to that discovered within an online friendship network and the physical, geographic distance between each pair of users (Liben-Nowell et al 2005).

Discussion

We have shown that eigenbehaviors can be used effectively to extract the underlying structure in the daily patterns of human behavior, predict subsequent behavior, infer community affiliations, and estimate the probability of a tie within the friendship network of the population. We are currently building applications that leverage this new technique in two main realms, behavior-based segmentation and data interpolation.

We have found that communities within a population's social network tend to be clustered within the same behavior space. It seems reasonable that this type of behavioral homophily is present in a variety of social networks. It should be possible for practitioners, using virtually any type of longitudinal behavior data, to similarly quantify the behavior space of a particular group or individual of interest using the eigenbehaviors technique described above. If strong behavioral homophily is present in the data, it should equally be possible to infer an individual's affiliations by quantifying the individual's distance from a community's behavior space.

When collecting large amounts of data from many subjects of an extended period of time, data loss is unavoidable. The Reality Mining logs account for approximately 85.3% of the time since the phones have been deployed. Approximately 5% of this is due to data corruption, while the majority of the missing 14.7% is due to the phones being turned off. However, with a set of these characteristic eigenbehaviors defined for each individual, it becomes possible to generate a rich synthetic dataset from the approximations of the individual's eigenvalues over a particular time window of interest. Using the behavior space generated from an individual's six primary eigenbehaviors, we have shown we can generate a 12-hour chunk of data with 79% accuracy. If we incorporated the individual's future behavioral data as well as the past, this accuracy should continue to increase.

It is inevitable that the next generation of wearable sensors will be appropriate for the long-term passive monitoring of an increasing set of living creatures. The behavioral data generated from these new devices will require fundamentally new techniques for analysis. To analyze data of such magnitude, eigendecompositions are useful because they provide a low-dimensional characterization of complex phenomena. This is because the first few eigenvectors of the decomposition typically account for a very large percentage of the overall variance in the signal. Because only few parameters are required, it becomes easier to analyze the individual and community behavior, and thus possible to predict the behavior of the individual elements as well as the behavior of the system as a whole.

These unique properties make eigenbehaviors ideal as a representation of daily movements, interactions, and communication behaviors. The low dimensional representation provided by the eigendecomposition will allow us to characterize an individual quickly, match him to similar individuals, and predict his behavior in the near future. The technique also provides us with a representation of the behavior characteristic of a community as a whole and enables us to estimate the probability of a tie within the larger social network of the population. As rich, longitudinal behavioral data becomes increasingly available, it is our hope that these techniques will prove useful to researchers studying a wide range of human and animal behavior.

References

Addlesee M, Curwen R, Hodges S, Newman J, Steggles P, Ward A, Hopper A (2001). Implementing a Sentient Computing System. IEEE Computer Magazine, Vol. 34, No. 8, August 2001, pp. 50-56.

Ashbrook D, Starner T (2003) "Using GPS to learn significant locations and predict movement across multiple users", Personal and Ubiquitous Computing 7: 275-286.

Bar-Noy A, Kessler I, (1993) "Tracking mobile users in wireless communication networks," IEEE Transactions on Information Theory, 39(6): 1877-1886.

Begole, JB, Tang JC, Hill R (2003) Rhythm Modeling, Visualizations, and Applications. Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2003).

Bhattacharya A, Das SK (1999) "LeZi-update: an information-theoretic approach to track mobile users in PCS networks. In: Proceedings of the International Conference on Mobile Computing and Networking, Seattle, WA.

Clarkson B (2002) "Life Patterns: structure from wearable sensors", Massachusetts Institute of Technology.

Eagle N, Pentland A (2006) "Reality Mining: Sensing Complex Social Systems", Journal of Personal and Ubiquitous Computing, 4, 255-268.

Fogarty J, Hudson SE, Atkeson CG, Avrahami D, Forlizzi J, Kiesler S, Lee JC, Yang J (2005). Predicting Human Interruptibility with

Sensors. ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 12, No.1, March 2005, pp. 119-146.

Hightower J, Borriello G (2004) "Particle Filters for Location Estimation in Ubiquitous Computing: A Case Study," in Proceedings of the Sixth International Conference on Ubiquitous Computing (Ubicomp 2004), (Nigel Davies, Elizabeth Mynatt, and Itiro Siio, eds.), pp. 88-106.

Kim SJ, Lee CY (1996) "Modeling and analysis of the dynamic location registration and paging in microcellular systems," IEEE Transactions on Vehicular Technology, 45(1):82-90.

Krause J, Lusseau D, James R (2009) Animal social networks: an introduction. Behav Ecol Sociobiol.

Laasonen K, Raento M, Toivonen H (2004) "Adaptive On-Device Location Recognition", In: Proceedings for Pervasive, pp 287-304.

Liao L, Fox D, Kautz H (2004) "Learning and Inferring Transportation Routines" In: Proceedings for the National Conference on Artificial Intelligence (AAAI-04), San Jose, CA.

Liben-Nowell D, Novak J, Kumar R, Raghavan P (2005) "Geographic routing in social networks", Proceedings of the National Academy of Sciences, 102 (33), 11623-11628.

Patterson D, Liao L, Fox D, Kautz H (2003) "Inferring High-Level Behavior from Low-Level Sensors," Ubicomp 2003: 73-89.

Pentland A, Williams J (1989) "Good Vibrations: Modal Dynamics for Graphics and Animation", ACM Computer Graphics, Vol. 23, No. 4., pp 215-222.

Pentland A, Sclaroff S (1991) "Closed-Form Solutions for Physically Based Shape Modeling and Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 13, No. 7., pp 715-730.

Pentland A (1992) "Fast solutions to physical equilibrium and interpolation problems", The Visual Computer. Vol. 8, No. 5-6., pp. 303-314.

Raento M, Oulasvirta A, Petit R, Toivonen H (2005) "ContextPhone – A prototyping platform for context-aware mobile applications". IEEE Pervasive Computing. IEEE Pervasive Computing, 4 (2).

Schilit B, Adams N, Gold R, Tso M, Want R (1993) "The ParcTab mobile computing system." In Proceedings of the Fourth Workshop on Workstation Operating Systems, pp. 34-39.

Turk, M., and Pentland, A., "Eigenfaces for Recognition", J. of Cognitive Neuroscience. Vol 3, Number 1., (1991) 71-86.

Want R, Hopper A, Falcao V, Gibbons J (1992) "The active badge location system," ACM Transactions on Information Systems, vol. 10, pp. 91-102.

Wolf J, Guensler R, Bachman W (2001) "Elimination of the travel diary: an experiment to derive trip purpose from GPS travel data". In: Proceedings from the Transportation Research Board 80th annual meeting, Washington, DC.

List of Figures

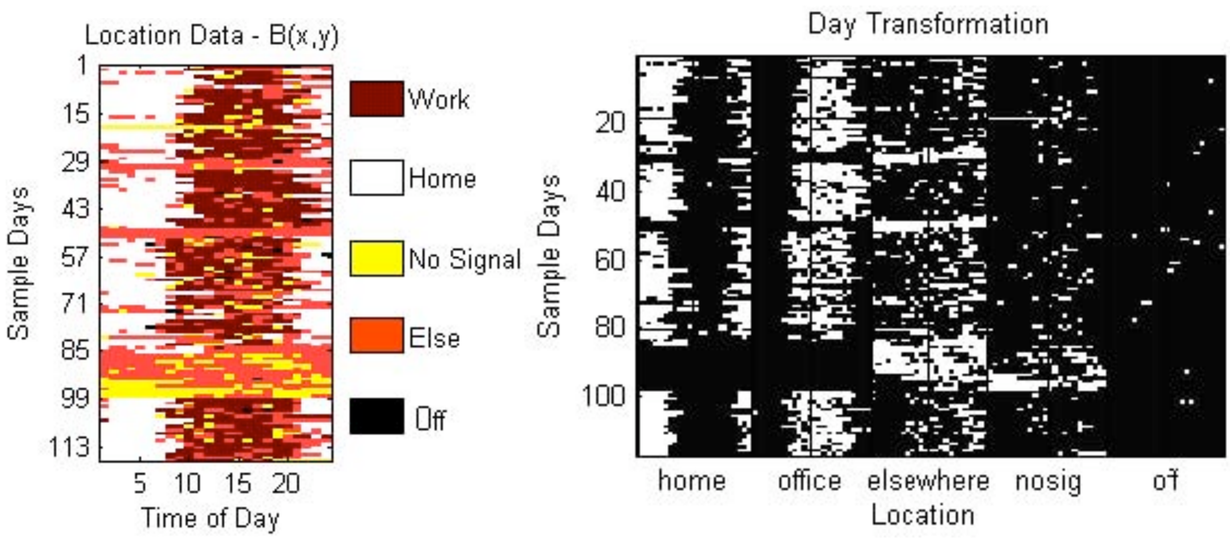


Fig 1. Transformation from B to B' . The plot on the left corresponds to the subject's behavior over the course of 113 days for 5 situations. The same data can be represented as a binary matrix of 113 days (D) by 120 (H, which is 24 multiplied by the 5 possible situations).

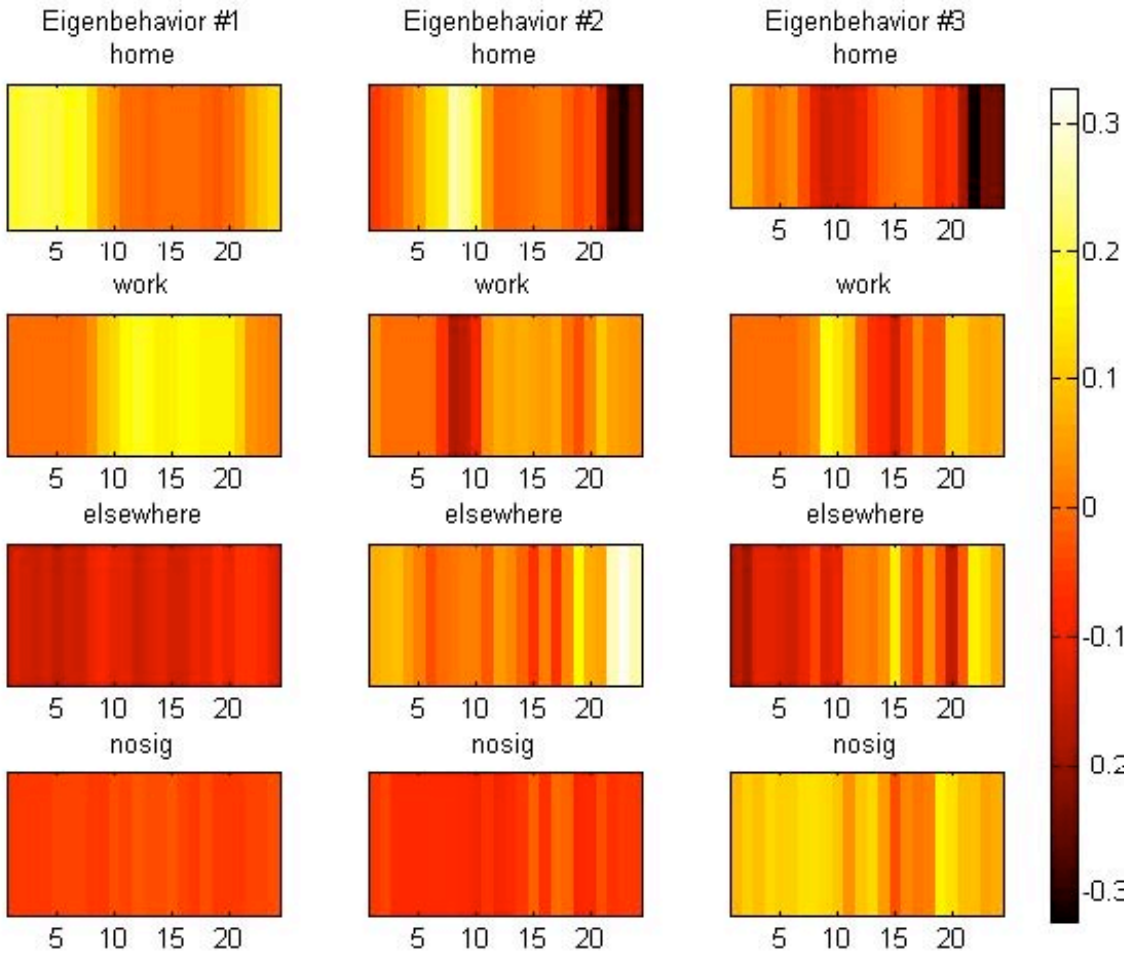


Fig 2. The top three eigenbehaviors, $[u_1, u_2, u_3]$, for Subject 4. The first eigenbehavior (represented with the first column of three figures) corresponds to whether it is a normal day, or whether the individual is traveling. If the first weight is positive, then this eigenbehavior shows that the subject's typical pattern of behavior consists of midnight to 9:00 at home, 10:00 to 20:00 at work, and then the subject returns home at approximately 21:00. The second eigenbehavior (and similarly the middle column of three figures) corresponds to typical weekend behavior. It is highly likely the subject will remain at home past 10:00 in the morning and will be out on the town ('elsewhere') later that evening. The third eigenbehavior is most active when the individual is in locations where the phone has no signal.

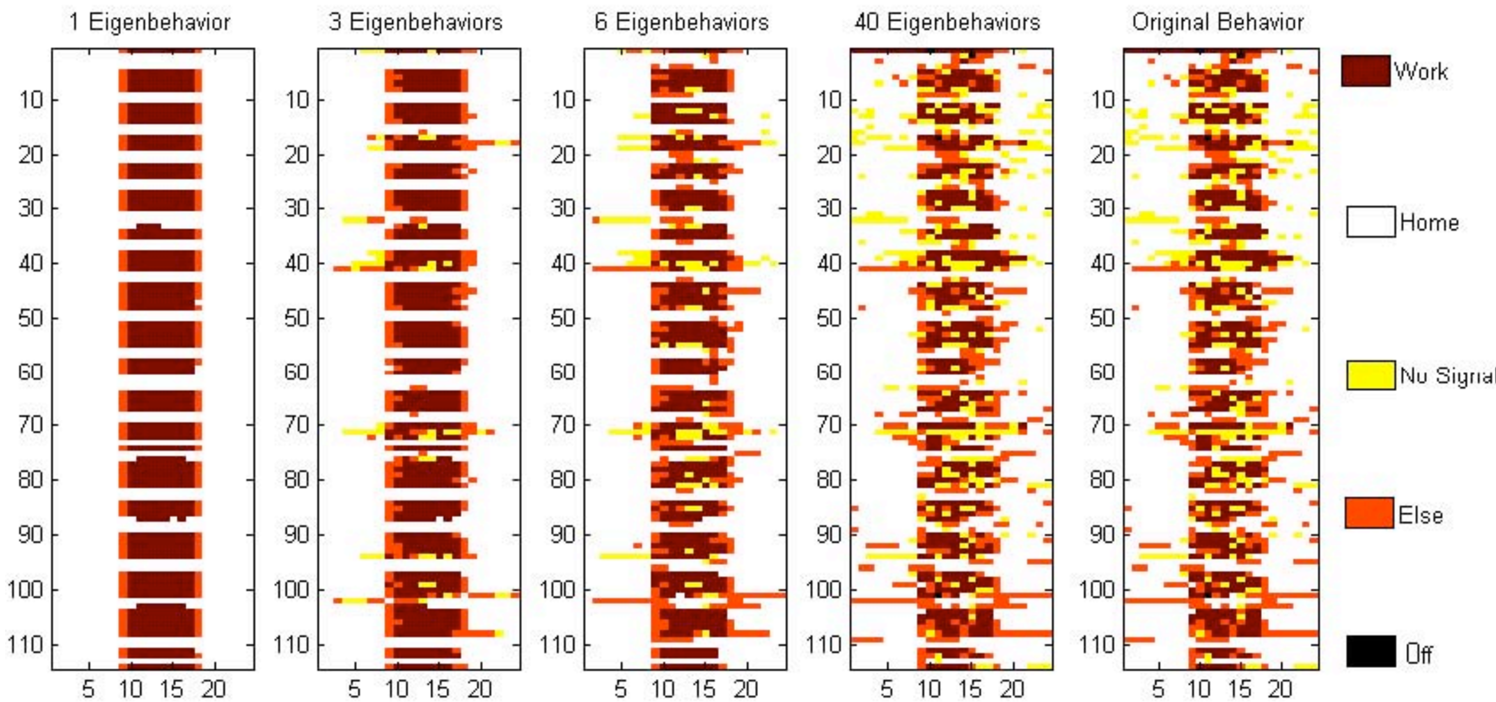


Fig 3. Behavior approximation of 115 days using a varying number of eigenbehaviors. The left-most figure corresponds to behavioral approximation using only one eigenbehavior. The approximation accuracy increases with the number of eigenbehaviors.

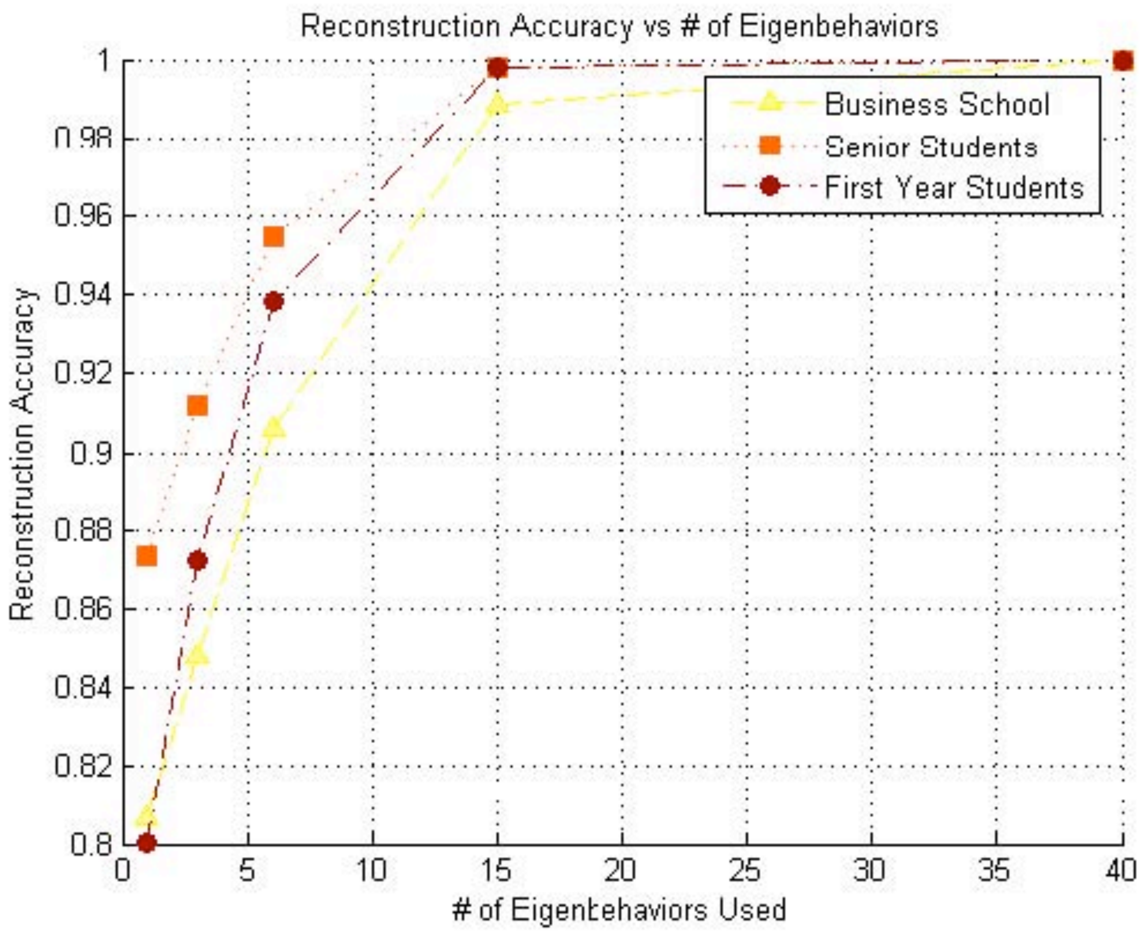


Fig 4. Approximation error (y-axis) for the different subject groups as a function of the number of eigenbehaviors used (x-axis) with the states off and no signal removed.

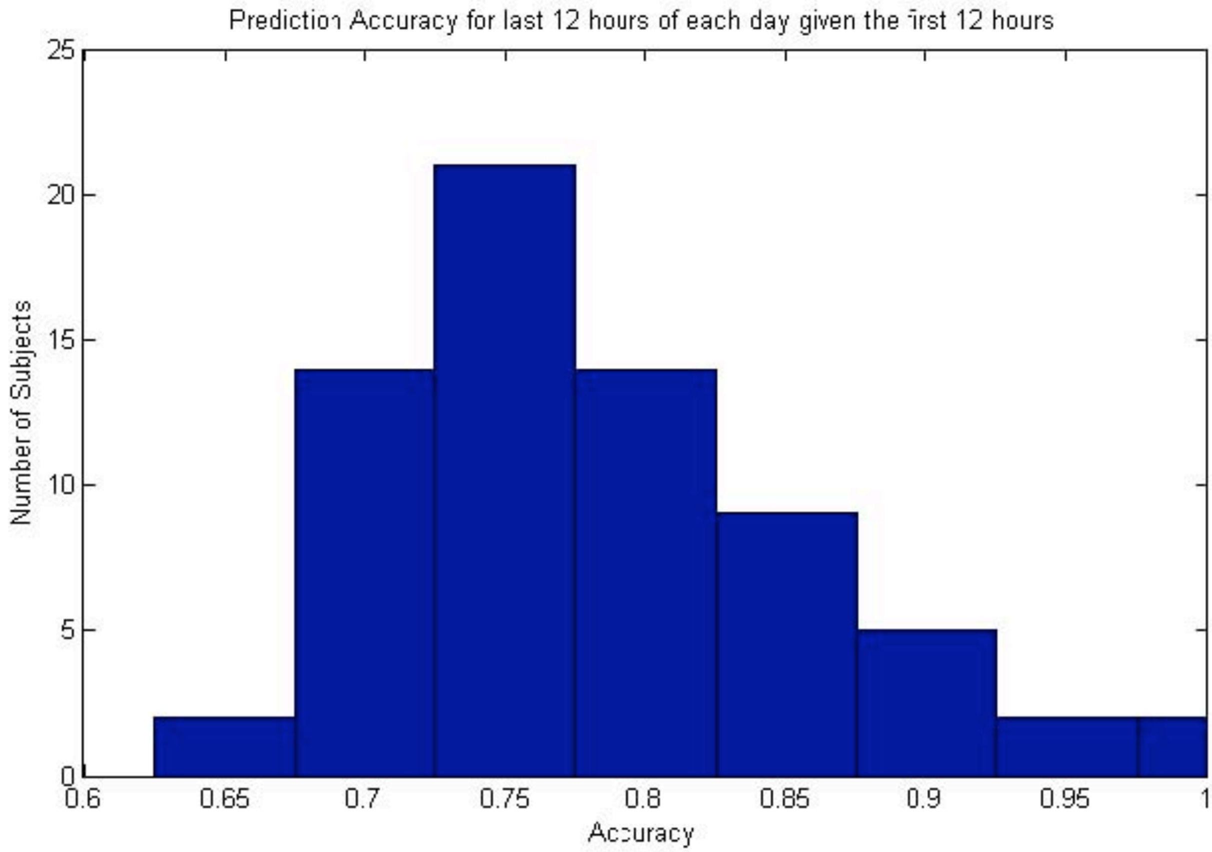


Fig 5. Behavior prediction accuracy for behaviors from noon to midnight given the previous 12 hours of behavioral data and the six primary eigenbehaviors for each subject, an average of 79% accuracy is obtained.

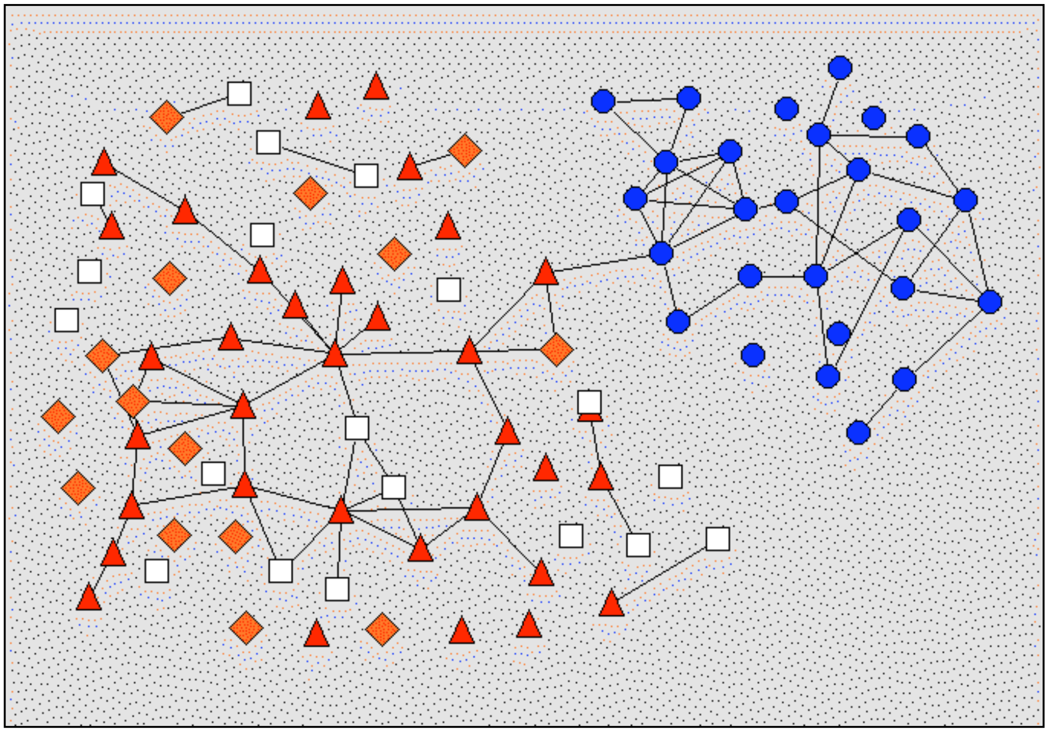


Fig 6. The social network of the population. The blue circles represent the community of business school students. The red triangles are senior lab students, the orange diamonds represent the incoming students, and the white squares represent the laboratory staff and faculty.

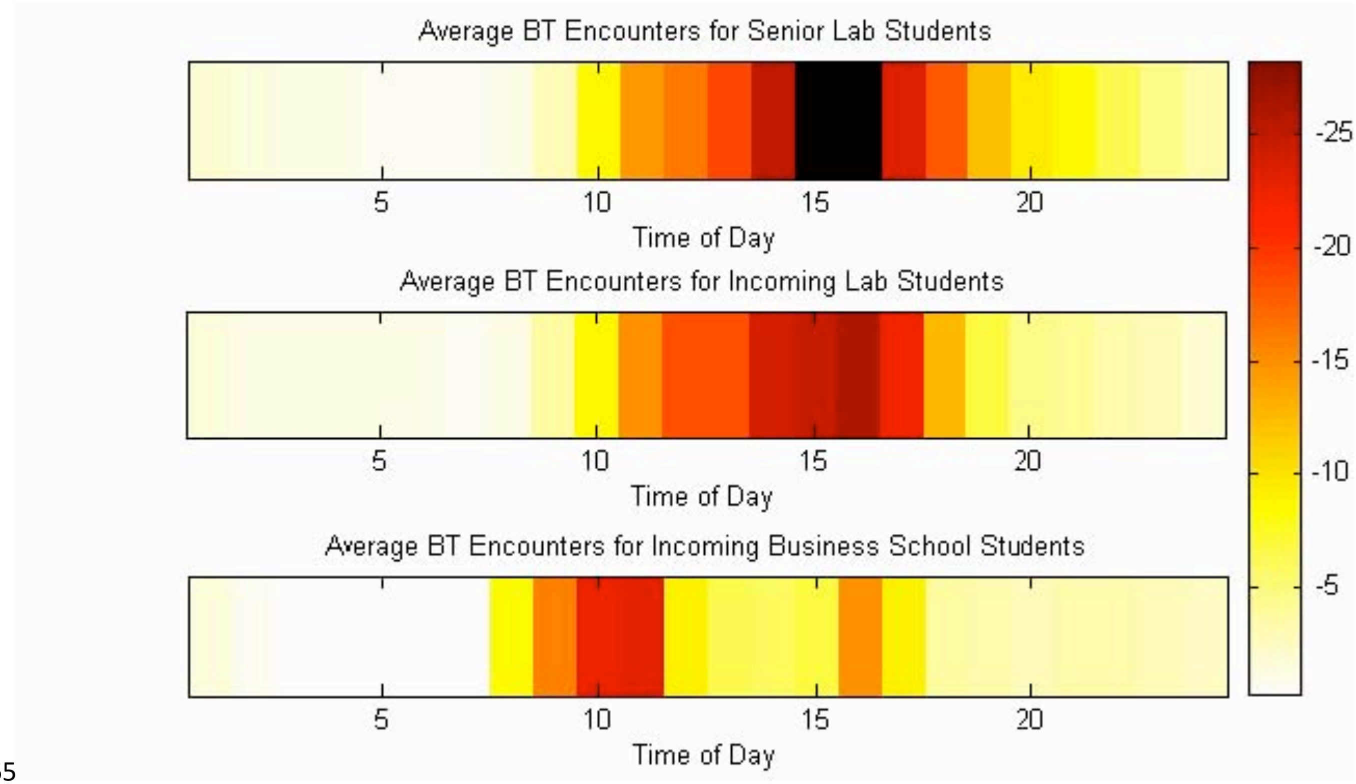


Fig 7. The average number of Bluetooth devices seen, Ψ_j, for the senior lab students, incoming lab students, and incoming business school students. The values in these plots correspond to the total number of devices discovered in each hour of scanning over the course of a day (with time of day on the x-axis).

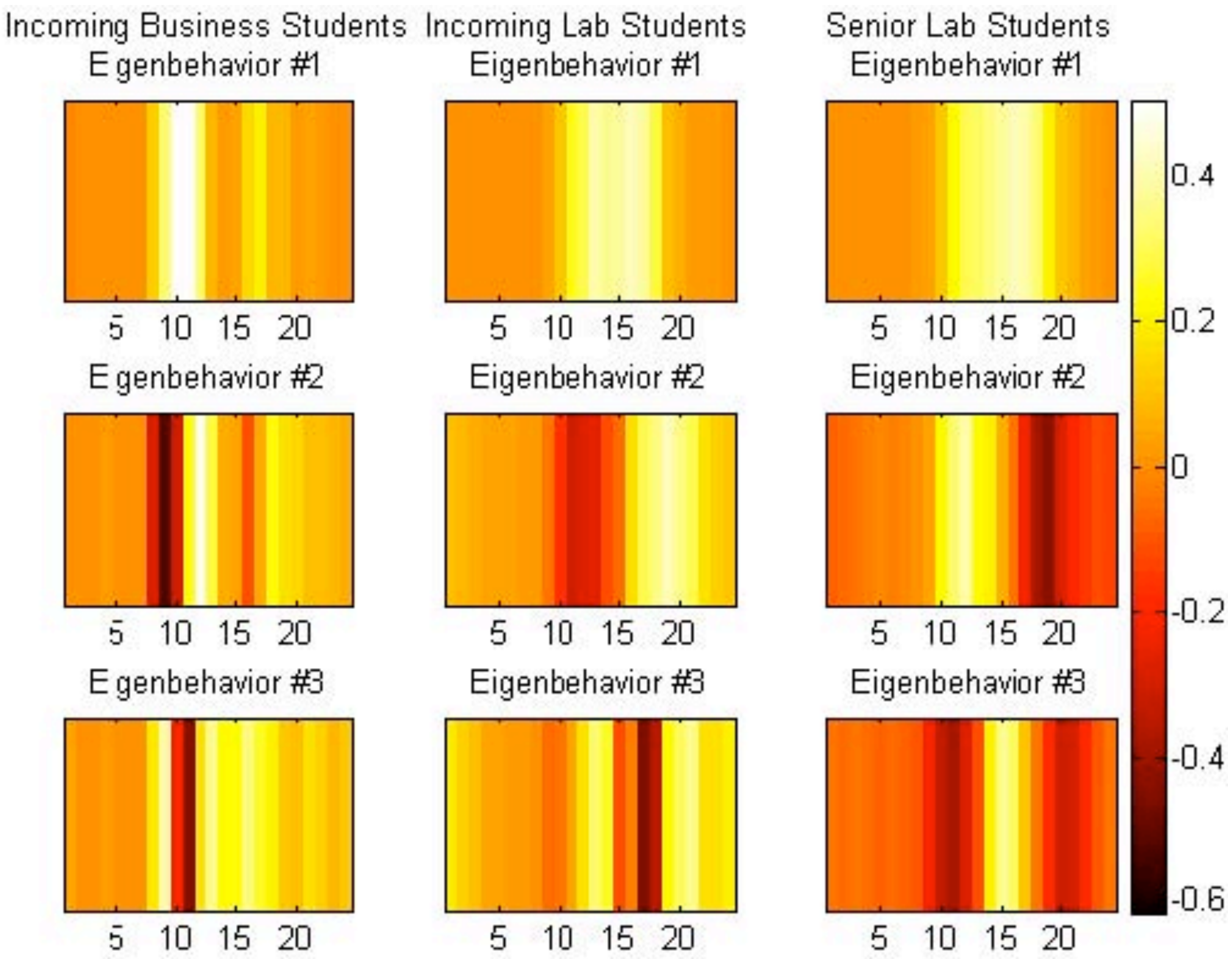


Fig 8. The top three eigenbehaviors $[u_1^j, u_2^j, u_3^j]$ for each group, *j,* comprised of the incoming business school students, incoming lab students and senior lab students. The business school coffee break at 10:30 is highlighted in their first eigenbehavior. Comparing the second eigenbehaviors for the Media Lab students, it can be seen that the incoming students have developed a routine of staying later in lab than the more senior students.

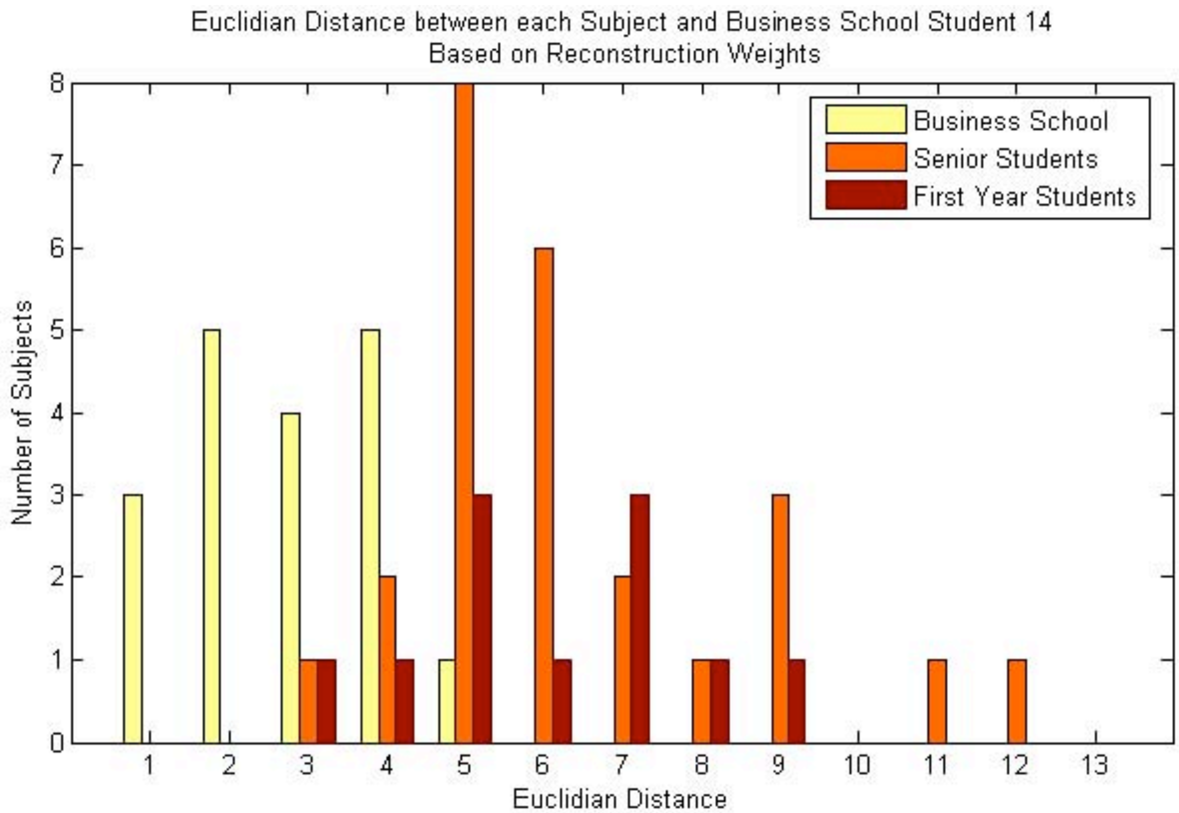


Fig 9. Values corresponding to ε_j , the Euclidian distance between each subject and a single business school student. The distance between two individuals reflects the similarity of their behavior.

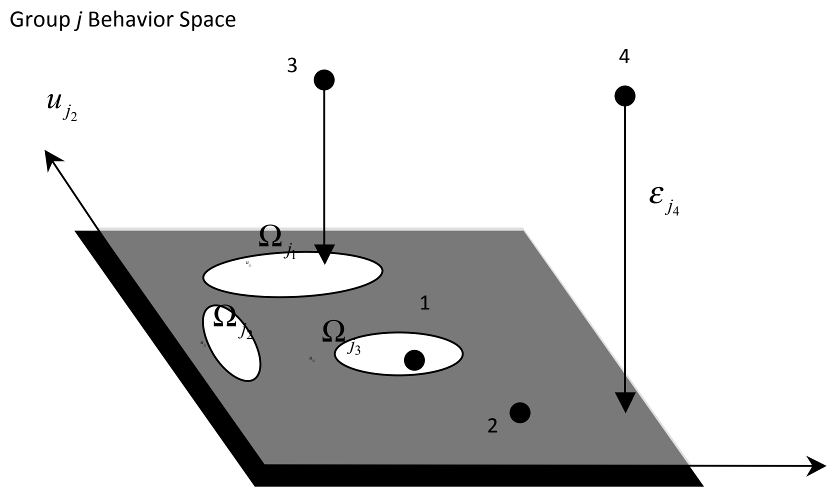
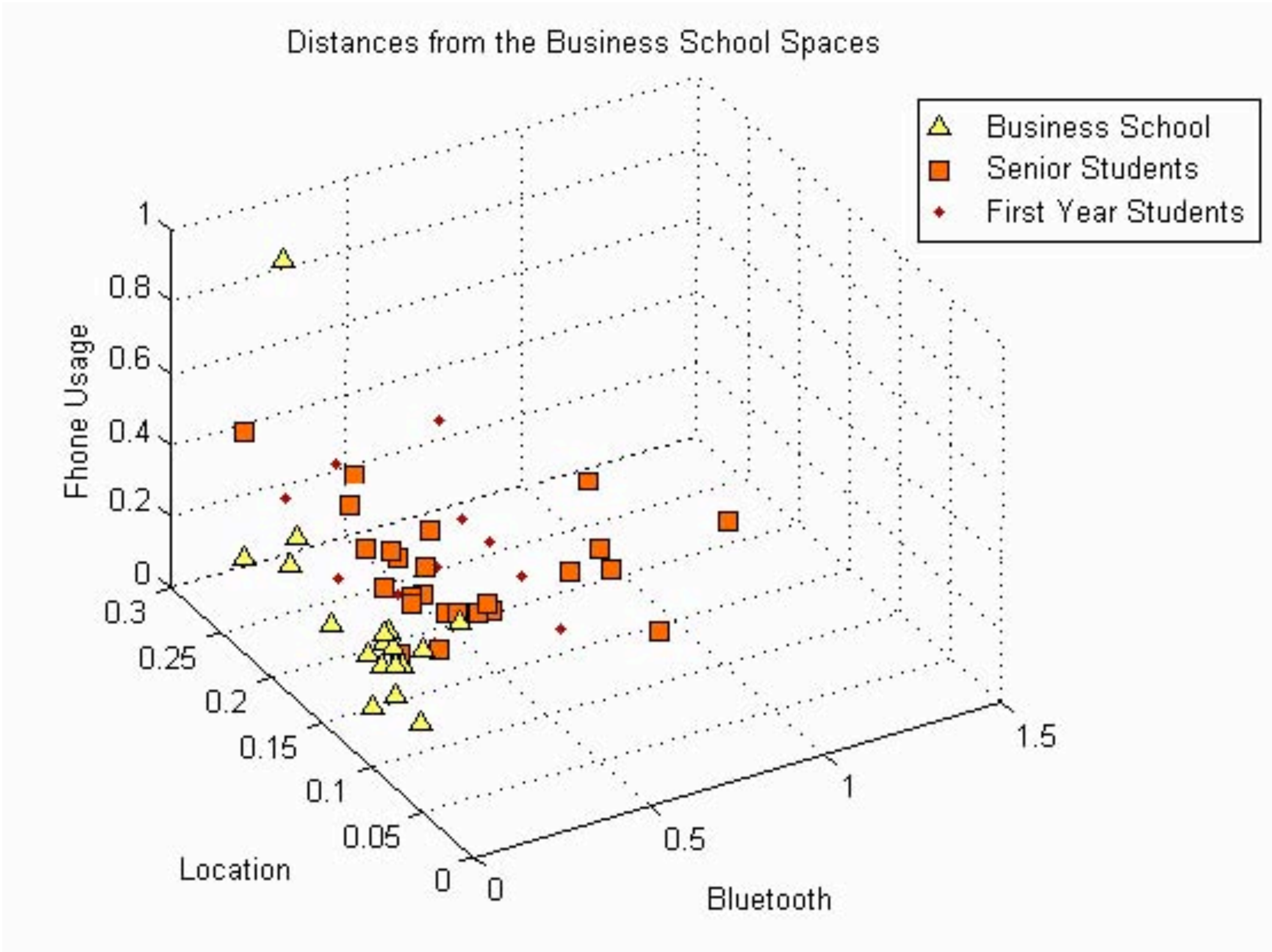


Fig 10. A toy example of community behavior space. Individuals 1 and 2 are on the behavior space and can be affiliated with the community. Individual 1 can also be affiliated with the particular clique, Ω_3^j. There is much more distance between 3 and 4 and the behavior space, and there-fore their projections onto the behavior space do not yield an accurate representation of the two people.

520
521



522 **Fig 11.** The cross-validated distance ε_j between the three groups of students and the Bluetooth, Location and Phone Usage business
523 school behavior spaces.

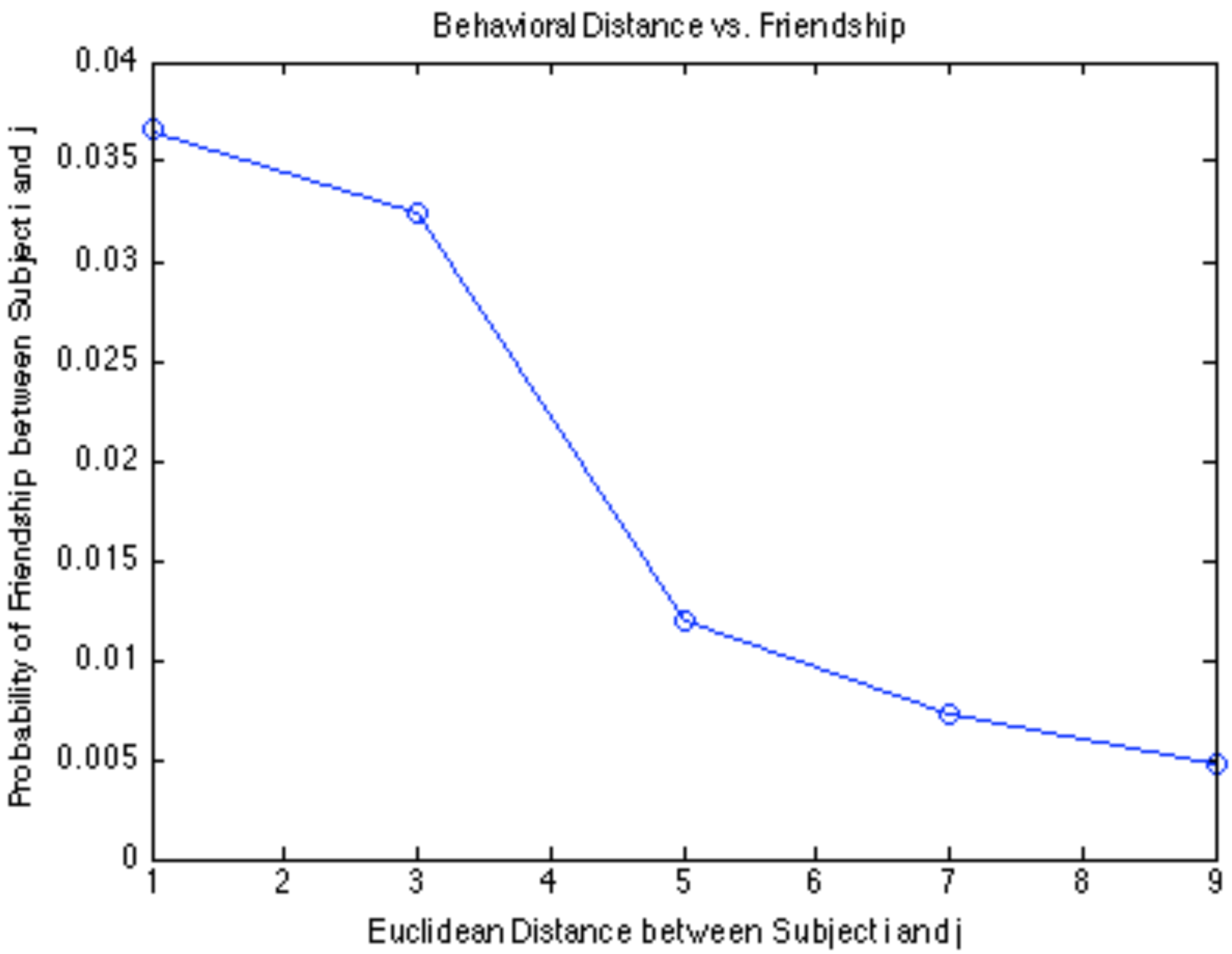


Fig 12. Behavioral Distance vs. Probability of Friendship. The Euclidean distance between every subject's projection onto the behavior space is calculated and compared with whether a friendship was reported between the two individuals. The figure suggests strong behavioral homoplipy, that is, subjects with similar behavior are more likely to be friends.