# Connecting Personal-scale Sensing and Networked Community Behavior to Infer Human Activities

**Nicholas D. Lane†, Li Pengyu‡, Lin Zhou*, Feng Zhao†**

†Microsoft Research, ‡Beijing University of Post and Telecommunications, *UC Santa Barbara

**ABSTRACT**

Advances in mobile and wearable devices are making it feasible to deploy sensing systems at a large-scale. However, slower progress is being made in activity recognition which remains often unreliable in everyday environments. In this paper, we investigate how to leverage the increasing capacity to gather data at a population-scale towards improving existing models of human behavior. Specifically, we consider the various social phenomena and environmental factors that cause people to develop correlated behavioral patterns, especially within communities connected by strong social ties. Reasons underpinning correlated behavior include shared externalities (e.g., work schedules, weather, traffic conditions), that shape options and decisions; and cases of adopted behavior, as people learn from each other or assume group norms due to social pressure. Most existing approaches to modeling human behavior ignore all of these phenomena and recognize activities solely on the basis of sensor data captured from a single individual. We propose the *Networked Community Behavior* (NCB) framework for activity recognition, specifically designed to exploit community-scale behavioral patterns. Under NCB, patterns of community behavior are mined to identify social ties that can signal correlated behavior, this information is used to augment sensor-based inferences available from the actions of individuals. Our evaluation of NCB shows it is able to outperform existing approaches to behavior modeling across four mobile sensing datasets that collectively require a diverse set of activities to be recognized.

**Author Keywords**

Mobile Sensing, Activity Recognition, Community Learning

**ACM Classification Keywords**

H.5.2 User/Machine Systems: I.5 Pattern Recognition

## INTRODUCTION

The emergence of smartphones and wearables as viable mobile sensing platforms has dramatically lowered the barrier to building and deploying large-scale sensing systems. However, the same rapid progress is not being made in activity recognition. State-of-the-art classifiers remain unreliable when exposed to the noisy dynamic environments common in the real world [8]. Until current methods for activity recognition advance, behavior and context classification will continue to act as a bottleneck, limiting the extent to which mobile sensing can achieve mainstream adoption.

In this paper, we investigate how an important, but currently neglected, property of human behavior can improve existing models used for activity recognition. Within large populations, communities of socially connected individuals often have correlated behavioral patterns [6][7][16]. For a surprisingly wide variety of behaviors, the actions of individuals often appear connected. For example, two people with a strong social connection will be more likely than a pair of randomly selected individuals to have correlated beliefs and attitudes, which manifest in actions, such as food selection, sleep patterns, and green transportation decisions. Similarly, socially connected individuals will often also share related external factors that shape behavior (e.g., correlated peak work periods, school exams, financial constraints). Various social processes cause these network dependencies to emerge, including homophily [26] (the tendency of similar people to form social bonds), social influence [24] and types of information diffusion [17]. In this paper, we refer to these and related processes as *community behavior*. Such phenomena present a significant opportunity to improve the recognition of behavioral patterns, as the inferences about one individual can be leveraged to improve the inferences made about others with whom they are socially connected.

Conventional activity recognition assumes that people are isolated individuals and models behavior based solely on the sensor data collected at a personal-scale, which is closely related to the physical actions associated with the behavior. We propose to widen this perspective and train models that use readily available forms of sensor data to capture forms of collective behavior between people in user communities. Our *Networked Community Behavior (NCB)* framework models activities by: (1) considering not only personal-scale events but also by making community-scale observations from sensor data; (2) identifying communities with correlated behavior within a user population; and, (3) using the evidence from both the personal- and community-scale to improve the accuracy of recognized behaviors.

Like many activity recognition frameworks, NCB can use, for example, accelerometer or microphone data to inform about the actions and context of individual users (personal-scale information). But NCB also mines from sensor data examples of correlated behavior by analyzing social networks and identifying communities of individuals who demonstrate patterns of related actions (community-scale information). By combining both types of data, NCB constructs a hierarchical net-


*UbiComp '14*, September 13 – 17, 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2632048.2636094

work that ties a lower layer containing soft inferences of personal activities to an upper layer that represents the social network of the user population. Select communities within the network are linked to user inferences based on which activities they impact. NCB adopts techniques from relational learning [11] by incorporating a relational classifier performing collective inference [15, 28], allowing it to make activity inferences using this hierarchical network. These techniques were developed for networked data, such as mining topics from text documents in the presence of citation and authorship graphs, and increase in accuracy as the level of correlation in the data increases. Importantly, the NCB framework remains flexible at the personal-scale level and allows relational learning to be used with already proven approaches (i.e., features and classifiers) for modeling specific activities.

The contributions of this paper are as follows:

- NCB is the first general purpose activity-recognition framework that models human behavior using sensor-based observations at both a personal-scale and community-scale. To the best of our knowledge, NCB is the only framework specifically designed to exploit the correlations that occur in a population due to community behavior.
- We propose new approaches to learning activity recognition models that incorporate: (1) a discovery process for user communities within broader social networks that capture the correlation between the behavior of individuals for certain classes of activity; (2) a training process that builds classifiers that blend personal-scale observations and community-scale relationships; and, (3) an inference step that carefully propagates personal-scale behavior probabilities between historically correlated users.
- Evaluation results using four mobile sensing datasets show that NCB: (1) exploits the opportunities found in correlated collective behavior; and (2) exceeds conventional classifiers that rely only on personal-scale sensor data to model the activities of individuals.

Although NCB is premised on the existence of correlated behavioral patterns within communities, the focus of this paper and its core innovation is the proposed NCB framework itself. As a result, the evaluation examines the efficacy of this framework to leverage such situations, and uses activity recognition datasets that we find contain users with linked behaviors. We do not systemically identify the reason for all incidence of networked user behavior in these datasets. Outside of a few case studies, we rely on the existing social science literature to explain why these situations manifest.

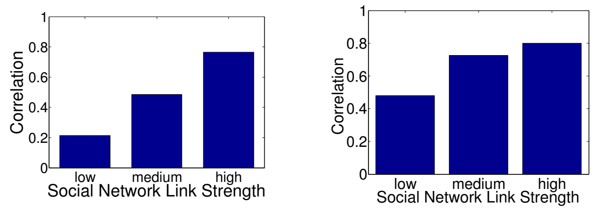
## EVERYONE IS CONNECTED

Existing sensor-based classifiers of everyday actions assume that individuals' behavioral patterns are completely uncorrelated. In this section, we show correlation can occur in real datasets which has implications for activity recognition and survey the various factors that cause such effects.

**Community Behavior.** Correlations among behavioral patterns are strongest between people with strong social ties. Examples of such behavior ranging from sleep patterns, to entertainment preferences, to exercise and transportation habits, can spread through social networks [6][7][16]. These effects suggest that within communities found in social networks, people respond similarly given similar contexts and situations (e.g., how they value sleep hours when they are busy or the likelihood that they will watch a movie or go to a night club). This phenomena is further reinforced by a variety of external factors (i.e., context and situations) shared by members of these communities related to where they live and work; for example, shared weather and traffic conditions, similar destinations or even a specific event like a subway strike will all jointly impact how a group of people commute to work each day. Because alternative transportation modes have similar trade-offs for each person, correlated decisions can naturally result. Additional shared community-wide externalities that can shape individual actions include: proximity of points of interest, relatively similar amounts of disposable income, and working hours of shared occupation types. Controlled experiments show that these types of effect even exist on a small scale, for instance, individuals will select different foods when dining in a group than when they are dinning alone, and will tend toward the norm pattern of a group in terms of making health food choices. The social processes that explain such phenomena (and many others) are varied, as are the terms used to describe them: social influence [24], information cascades [17], information diffusion, group decision making, community dynamics, and dynamic social models. We refer to these effects as community behavior.

Various social phenomena underpin these patterns of correlated behavior; for example, people might consciously or unconsciously adopt the behavior of others, such as individuals they identify as role models, or they observe an outcome they also wish to achieve. More directly, people exchange information with each other that changes their previous behavior and must act within hard environment-based constraints common across the whole community (e.g., bus schedules, store open hours, school holidays). In addition, social pressure often causes a person to conform to existing behavior norms or attitudes. Finally, research indicates that much of this phenomenon can be explained by homophily: people show a strong tendency to develop social ties with those with whom they share correlated behaviors and beliefs. Consequently, many cases of correlated behavior in social networks are the result of self-selection during these networks' formation rather than behavior propagating within them. However, regardless of the process behind the emergence of correlations, the significance to activity recognition is the same.

**Potential Implications for Activity Recognition.** Even though community behavior leads to some level of correlated activity between socially connected people, it may not have a meaningful effect in datasets used for activity recognition. We begin to consider this question by testing two mobile sensing datasets that capture sleep duration (discretized into 4 duration categories) and transportation mode decisions (daily transportation choices – car, bus, subway, walk); each dataset included 27 and 51 participants over 21 and 90 days, respectively. (For further dataset details see Evaluation section).




(a) Transportation Mode (b) Sleep

**Figure 1:** Pairs of people with strong social links have correlated behavior in diverse activities, such as, transportation and sleep.

To test if strongly socially connected people tend to have more correlated behavioral patterns, we first mined social links based on participant co-location and trajectory similarity (adopting the technique of [9] which is also used within NCB). Correlation of behavior was computed over a sliding window of 4 days, using a jaccard index (that ignores strict ordering of events) to compare the sets of sleep duration and transportation decisions between individuals. In Figures 1(a) and 1(b) we report correlation scores (i.e., jaccard index) when grouping our subjects by the strength of social connection, labeling the bottom 20% as "low", the top 20% as "high", and the 20% straddling the 50th percentile as "medium". We find sleep duration categories and transportation mode types are strongly related (approximately 0.79 and 0.77) between those users that also have strong social ties. These figures also show correlation increases as the strength of social connection grows. Although preliminary, such results show a measurable presence of correlation between users and so highlight the potential need for activity recognition frameworks that can leverage such effects.

**Isolated Users Assumption.** Most existing approaches to activity recognition assume that people operate in a social and environmental vacuum, isolated from each other. Excluding a few emerging techniques (e.g., [23, 32, 18]), activity models are based solely on an individual's actions, as observed by sensor data, and ignore the impact of community behavior within the user population.

The key problem is that today's classifiers lack the ability to exploit additional community-scale information when recognizing behaviors, such as those present in the correlated behavior found between social connected groups of people. When discriminative patterns in an individual's sensor data are noisy or ambiguous, the observations available for socially related individuals can provide critical new signals. This effect's usefulness is even more pronounced for longer-term behavior, such as week-long sleep or dietary patterns.

## NETWORKED COMMUNITY BEHAVIOR FRAMEWORK

Unlike conventional approaches to activity recognition, NCB recognizes behavioral patterns of individuals using personal-scale (e.g., accelerometer data related to physical actions) and community-scale (e.g., community behavior) observations simultaneously. Activity inferences can therefore be more robust because NCB leverages not only an individual's sensor data but also the sensor data of other people with correlated behavioral patterns with whom they shares strong social ties.

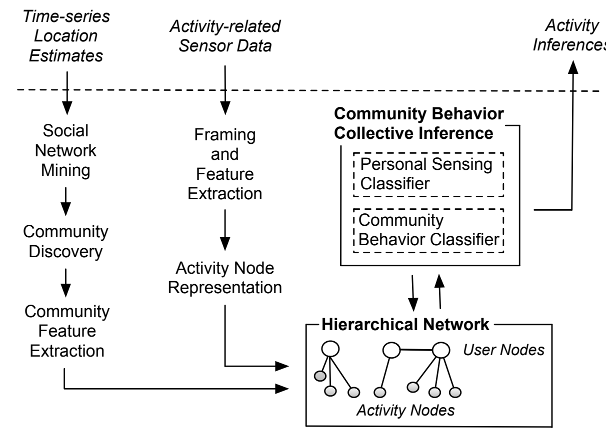



**Figure 2:** Networked Community Behavior Framework

### Overview

Figure 2 illustrates the data-flow and primary components of the NCB framework. This framework is divided into two key phases: the first constructs a hierarchical network linking personal- and community-scale observations; and the second is a learning phase that trains classifiers and performs collective inference to recognize behavioral patterns.

To build the hierarchical network, NCB requires two types of data that can be readily collected from mobile devices: (1) sensor data collected while people perform the activities to be recognized; and, (2) time-series location estimates, from which common trajectories and locations can be mined.

NCB segments the sensor data into chunks appropriate for the activities to be recognized and extracts features for each chunk. This process forms the bottom layer of the network, with each occurrence of an activity corresponding to a network *activity* node and features that act as node attributes. From location traces, NCB extracts a social network graph based on similarities in user trajectories and periods of co-location. This graph forms the top layer of the network with each node at this layer representing a person (and so referred to as a *user* node). NCB then mines communities within the social network, extracting network structural features (e.g., centrality) to further annotate this layer.

NCB adopts techniques from relational learning [11] by incorporating a relational classifier and performing collective inference across the hierarchical network. Two types of classifiers are used in this process – the Personal Sensing Classifier and Community Behavior Classifier; each estimating the probability of potential activity classes. Because NCB assumes the use of conventional supervised learning for this pair of classifiers, labeled sensor data for targeted activities will be necessary to bootstrap the framework. Final inference is decided by performing Relaxation-Labeling [15, 28], which accounts for the discriminative patterns available in personal- and community-scale information.

The NCB framework can be implemented using a combination of smartphones or wearables along with cloud infrastructure. Mobile devices would be responsible for providing the

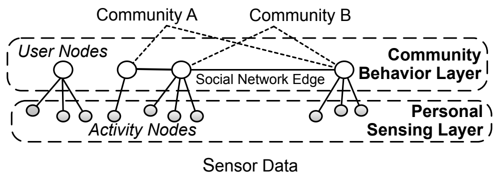



**Figure 3:** Hierarchical NCB Network

sensor data for personal- and community-scale observations. However, activity inference has to occur in the cloud because it is based on observations gathered from potentially all other users. When recognizing user activities, NCB is most effective when the final inference stage (i.e., Relaxation-Labeling) is delayed until the sensor data related to the targeted activities are collected from a significant fraction of all users. This is because such delay increases the opportunity for NCB to exploit any correlated behavior.

### Hierarchical NCB Network

Figure 3 illustrates the hierarchical network NCB uses to represent the relationship between personal-scale and community-scale data. The network's top layer of user nodes contains community-scale relationships. Nodes represent people, edge weights between top-layer nodes indicate the strength of social ties, and node attributes indicate individuals' memberships in communities within the network, along with social network statistics (e.g., centrality) that characterize their relationships with each community. The bottom layer of activity nodes describes personal-scale data. Nodes represent instances of activities occurring with links between the two network layers based on which person is performing the activity. Activity node attributes, in addition to capturing sensor-based features, are also used to indicate the ground truth activity, if known, or a vector of the likelihood for each activity category (i.e., soft-decision vector).

We now detail the function of components shown in Figure 2 that construct each layer of this network.

**Personal Sensing Layer.** We begin by describing the bottom layer of the network, which is constructed using a two-stage process. The information captured in this layer is similar to other conventional activity recognition approaches.

*Framing and Feature Extraction.* The available sensor data (e.g., accelerometer, audio, GPS) is segmented into frames, the length of which depends on the activity being described. Activity specific features are then extracted from each frame.

*Activity Node Representation.* Because each instance of an activity is encoded as a separate node in NCB there are practical limits as to how activities can be represented. For high frequency events, such as basic physical activities (e.g., running, walking), activity nodes can not be created, for example, every minute – thus, alternative representations may be required. However, for many user behaviors this is not a problem – for example, sleep duration, or user mood can be represented using even just a single activity node per day. But in cases of high frequency events, activity nodes can capture a behavior summary; for instance, a daily total duration (e.g., total time spent walking during a day) or daily activity instance count. If a behavior summary is needed additional sensor data processing (e.g., inferring activities and extracting features describing likely summary states) is performed when constructing the Personal Sensing Layer.

**Community Behavior Layer.** Our three-stage process for building the top user node layer begins with extracting a weighted graph that captures the social ties between individuals. In the resulting graph, nodes represent users and edge weights indicate the strength of users' social connection. At this stage, NCB aims to recognize the distinct communities (subgroups of users) within the social network and characterize the relevance each community has on each individual user (i.e., community features). Information related to the membership and the relevance of communities to users is encoded as user node attributes (see Figure 4). Later, during the learning phase, NCB uses this graph to understand which communities impact specific behavior categories.

| Category | Feature |
|---|---|
| Intensity & Duration | $\{NumColoc, NumLoc\} \times \{All, Evening, Weekend\}$<br>NumHours, NumWeekdays, BoundingBoxArea |
| Location Diversity | $\{Avg, Med, Var, Min, Max\} \times \{Entropy, Freq\}$<br>$\{Avg, Med, Var, Min, Max\} \times$ UserCount |
| Mobility Regularity | SchEntropy $\times \{L, LH, LD, LHD\}$<br>SchSize $\times \{LH, LD, LHD\}$ |
| Specificity | $\{Min, Avg, Max\} \times$ TFIDF, PerObvTogether |
| Structural Properties | NumMutalNeighbors<br>$\{Neighborhood, Location\} \times$ Overlap |

**Table 1:** Social Network Mining Features. Adopted from [9].

*Mining Social Network Graph.* We identify social ties between users by mining co-location and user trajectory traces based on GPS or WiFi data. Although co-location might be coarse compared to social networks based on face-to-face interactions (mined, for example, from conversation networks [5]), recent research supports this approach [9][31]. NCB uses a subset of the features developed in [9] to extract the social network from GPS and WiFi data, as listed in Table 1. Essentially, these features estimate the strength of the social link between two individuals based on the similarity of their trajectories or the quantity of co-location. Both factors are normalized to how common the level of co-location or trajectory similarity is within the larger dataset. Each feature is sensitive to how co-location is calculated from location estimates (e.g., GPS); in this work, we assume two individuals are effectively in the same place if they are within 30 meters and these estimates are within 10 minutes of each other – these are the same parameters employed in [9]. To determine trajectories when computing features, we interpolate between the available location estimates (using WiFi or GPS data). Finally, to estimate the intensity of social tie (i.e., edge weights), we simply use a multi-variate regression across all features. Significantly, edge weights below a fixed threshold are ignored. Therefore the social graph is typically not fully connected.

Although in this paper we assume the use of GPS and WiFi data, our NCB framework is agnostic to the source of the social network graph. Without effecting any other part of the framework, NCB could use a graph provided, for example, by user self-report or by mining conversation patterns.

*Discovering Communities.* Certain behaviors of an individual tend to correlate more strongly within specific communities within their social network. For example, a person's lunchtime eating habits might be related to those of work colleagues, whereas connections between sleeping patterns are likely to occur with the individual's partner and family. This observation is supported by findings in the social science literature, such as [24], which demonstrates that political opinions, exercise behavior, and retirement savings choices are impacted by different communities within the study participants' social networks.

NCB copes with this factor by identifying communities within the social network graph. This allows the learning stage to treat links between individuals and certain communities differently from another given that the level of correlation (and the affected activities) are different. Community membership is introduced in the top-layer network by adding additional attributes to each user node based on the communities the individual belongs to (see Figure 4). Algorithms for unsupervised recognition of communities based on social network link structure were previously developed. We adopt the Newman and Girvan [12] community discovery algorithm. This approach iteratively removes edges with high normalized betweenness values[1] while seeking towards optimizing modularity ($Q$), where $Q = \sum A_{i,j} - P_{i,j}$; where $A_{i,j}$ is the adjacency matrix between nodes $i$ and $j$ and $P_{i,j}$ is the expected edges between $i$ and $j$. Essentially, this algorithm groups nodes into communities based on an assumption of what distinguishes a community from the larger social network. Specifically, community members tend to be densely interconnected, and comparatively poorly connected with the rest of the social network.

*Community Feature Extraction.* Studies of collective behavior have shown that the strength and the topology of links within communities affects the behavioral impact on individuals. Feedback loops are one example. The level of density within a community can cause a feedback loop of any effect that increases the strength felt by an individual. For example, people A, B, C might all influence person D positively (e.g., by acting as role models or during discussions), then later persons B, C, D all influence A positively, with A then combining again with B and C to influence D, and so on. Real-world experiments have shown this effect in scenarios such as music tastes and buying decisions [24].

| Category | Feature |
|---|---|
| Centrality | Betweenness, Eigenvector, Closeness, Degree [1] |
| Other | Average Network Clustering Coefficient [10] |

**Table 2:** Community Features.

NCB differentiates the impact strength of various communities the user may belong to by computing a series of previously developed social network features selected based on their ability to summarize an individual's relative position within different communities using only link strength and topology. For each person in the social network, we generate four features *for each* community he or she belongs to. Specifically, we use four variations of centrality and a network clustering coefficient – as detailed in Table 2 and defined in [1] and [10] respectively. These centrality variations measure the relative importance the person has within each community. The network clustering coefficient represents the density of links within the community.

[1]Number of shortest paths between node pairs that include the edge.

### Networked Community Behavior Learning

Training and inference under NCB is based on relational learning and collective inference techniques. The inference of an activity from the hierarchical NCB network occurs in two steps. First, NCB trains two classifiers using personal- and community-scale information. These classifiers – the Personal Sensing Classifier and Community Behavior Classifier – summarize node attributes at each network layer and expose soft decisions as to the likely activity class. Second, NCB performs collective inference using Relaxation-Labeling. This iterative process uses the network weights and classifier outputs to arrive at a final inference result for each activity, and copes with the important issue of limiting error propagation.

**Personal Sensing Classifier.** Our Personal Sensing Classifier uses only data observed from individual users contained within the bottom hierarchy of the NCB network (i.e., the Personal Sensing Layer). The classifier is trained using a separate dataset with sensor data features labeled by hand with the ground-truth activities that occurred. Subsequently when inferences are required, the network is populated with new unlabeled activity nodes based on sensor data gathered from users. This classifier is applied to all new activity nodes and produces a vector of activity class likelihoods (i.e., a soft-decision), as well as an inference confidence value that estimates the certainty by the classifier in this result.

We do not innovate in this part of our framework, and rely on features and classification models recognized as being effective for the specific classes of activity to be inferred. NCB currently assumes the use of static non-temporal classifiers that do not capture sequential information. However, NCB's primary requirement is that the classifiers must be able to make soft classification decisions, although most classifiers already have a soft-decision variant developed.

**Community Behavior Classifier.** Our Community Behavior Classifier, just like the Personal Sensing Classifier, also makes a soft inference of user activity for all activity nodes. However as illustrated in Figure 4, this classifier operates on a combination of features (i.e., node attributes) from user nodes along with soft-decision vectors from the Personal Sensing Classifier that are aggregated from multiple users (i.e., the network neighborhood).

*Network Neighborhood.* The Community Behavior Classifier operates on data from a neighborhood of people. This neighborhood dictates which nodes in the Personal Sensing Layer and Community Behavior Layer will be used to generate features used by the Community Behavior Classifier. Membership in this neighborhood is determined for each ac-

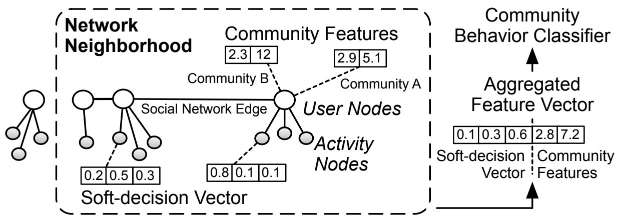



**Figure 4:** Community Behavior Classifier

**Algorithm 1:** Relaxation-Labeling under NCB

**For** $\forall$ unlabeled $a_j \in$ Hierarchical NCB Network
  $a_j.\mathtt{per}\boldsymbol{V} \longleftarrow$ Personal Sensing Classifier $(a_j)$
**End**
**While** $ca_{cv} \geq ca_{th}$ and iterations below threshold $\leq it_{th}$
  **For** $\forall$ unlabeled $a_j \in$ Hierarchical NCB Network
    $\mathtt{temp}\boldsymbol{V} \longleftarrow$ Community Behavior Classifier $(a_j)$
    $a_j.\mathtt{com}\boldsymbol{V} \longleftarrow \beta\,\mathtt{temp}\boldsymbol{V} + (1-\beta)a_j.\mathtt{per}\boldsymbol{V}$
  **End**
**End**
$a_j.\mathtt{inference}\boldsymbol{V} \longleftarrow a_j.\mathtt{com}\boldsymbol{V}$

tivity node by its parent node in the Community Behavior Layer of the NCB network (i.e., user node), which corresponds to the user who generated the data. All adjacent nodes of this user node (i.e., other users with whom the target user is socially connected) will be included in the neighborhood, along with all of their children activities nodes.

*Aggregated Neighborhood Feature Vector.* The feature vector used by the Community Behavior Classifier contains elements generated through an aggregation process, as is standard in relational classification. The purpose of the aggregation process is to determine a set of neighborhood features based on the individual features of each neighborhood member. The features used are: (1) community features (i.e., structural network metrics) from each user node; and, (2) the soft-decision vectors from all activity nodes.

A variety of aggregation processes have been proposed, we find a weighted mean of the soft-decision vectors and community features to be effective for our evaluated datasets. Specifically, for unlabeled activity $j$ performed by user $q$:

$$a\boldsymbol{V}_j = \sum_{i \in \mathcal{N}} \boldsymbol{V}_i\, ic_i \mathrm{e}^{-kt_{i,j}}\, w_{p,q}\, cf_k \tag{1}$$

is used to find an aggregate feature vector ($a\boldsymbol{V}$) based the soft-decision vector ($\boldsymbol{V}_i$) of each activity node ($i$) in the neighborhood ($\mathcal{N}$); precisely the same equation is applied to aggregate community features that are treated as a vector replacing each $\boldsymbol{V}_i$ in the calculation. In Equation 1, $w_{p,q}$ captures the strength of social tie between the user ($p$) tied to the neighborhood activity node ($i$) and the user ($q$) for whom an activity is being inferred. Similarly, inference confidence ($ic_i$) expresses a classifier-specific estimate of the confidence in the inference $\boldsymbol{V}_i$ made by the Personal Sensing Classifier, which is used to cope with error propagation (use of this term is described in the following subsection). Inference confidence is also scaled (using $\mathrm{e}^{-kt_{i,j}}$) based on the time ($t_{i,j}$) elapsed between the two events at the Personal Sensing Layer – diminishing the influence of events that happened in the past; importantly, this allows activity nodes within the neighborhood to be scoped out of consideration entirely beyond a certain threshold of elapsed time. Finally, a community factor ($cf$) regulates the influence of other users based on the communities membership. This factor essentially is the correlation of behavior between all user within the same community for the set of predicted activities, based on available training data.

*Classification.* In contrast to classification at the activity node level, at the user node level the classifier estimates the likelihood of activity classes for a user's activity nodes given the recent and concurrent activities of others with whom they: (1) have strong social ties; and, (2) belong to communities that historically have correlated behavior between members for this activity type. For our experiments we use an SVM classifier because it is effective across our datasets. However, again NCB can use any classifier able to make soft decisions.

**Community Behavior Collective Inference.** Like other methods for collective inference, Relaxation-Labeling [15] makes a Markov assumption and only considers the effects of direct neighbors in the hierarchical network, rather than the effects of neighbors more than one link away.

As detailed in Algorithm 1, Relaxation-Labeling performs inference iteratively. At first only the Personal Sensing Classifier is used on all unlabeled nodes ($a_j$) as an algorithm bootstrapping phase. Subsequently, it applies the Community Behavior Classifier to each node in turn. At every iteration, the inference for an activity node ($a_j$) based on community behavior ($a_j.\mathtt{com}\boldsymbol{V}$) is biased based on the original sensor data ($a_j.\mathtt{per}\boldsymbol{V}$), the extent of this bias is determined by $\beta$ which is set between 0 and 1. Over a number of iterations, the inference of other nodes will impact others as their classification result (i.e., soft-decision vector) changes. Relaxation-Labeling continues until the algorithm converges and the stop condition is reached. We compute this stopping condition based on the fraction of activity nodes ($ca_{cv}$) with changed inference results ($a_j.\mathtt{com}\boldsymbol{V}$) being below a certain threshold ($ca_{th}$) for a specific number of iterations ($it_{th}$). These two parameters (i.e., $ca_{th}$ and $it_{ct}$) are determined experimentally.

Propagation of error is a practical issue to be dealt with when performing any type of collective inference. Node misclassifications can spread because of the networked nature of this inference approach. We cope with this issue by adopting Cautious Collective Classification [25]. As described earlier, a scaling term ($ic_i$) is used when the relational classifier performs aggregation (see Equation 1). The term is calculated during the bootstrapping process of Relaxation-Labeling, when the only the Personal Sensing Classifier is used across all nodes. As a result, low-accuracy nodes (those whose predictions are less reliable) have less influence on other nodes even if they have a high degree of influence. Propagating community behavior is not useful if the node information is unreliable.

| Dataset | Activity Classes |
|---|---|
| *Transportation* | { bus, subway, car, walk, stationary} |
| *Sleep* | { 9+ hours, 9 to 6 hours, 6 to 3 hours, <3 hours } |
| *Diet* | { fast food, cafeteria & food court<br>western restaurant, ethnic restaurant} |
| *Mood* | { `PA` – above median pleasure/activeness<br>`pA` – below median pleasure/activeness<br>`Pa` – above median pleasure/activeness<br>`pa` – below median pleasure/activeness } |

**Table 3:** Dataset Activity Classes.

## EVALUATION

In this section, we study the effectiveness and design of NCB. We find by using collective behavior, NCB classifies a diverse set of activities more accurately than existing approaches.

### Experimental Methodology

We evaluate NCB using four datasets primarily collected using smartphones. All classifier performance metrics are calculated using per-user leave-one-out cross validation.

**Datasets.** Table 3 lists all activity classes contained in the four datasets used in our evaluation. Our first dataset, *Transportation* comprises 51 people who collect GPS traces of their transportation mode behavior for three months. Data is collected not only with smartphones but using other personally worn GPS-enabled devices (e.g., PDAs, personal navigation devices). Users self-report ground-truth transportation modes. Next, *Mood* contains 25 people all of whom provide phone usage data (e.g., SMS, email, phone call, application usage, web browsing, and location). Participants provide usage data for two months; in addition they complete a brief mood survey instrument (based on the Circumplex model [29]) multiple times per day. Under the Circumplex model, user moods are represented as a pair of values corresponding to two dimensions of mood (pleasure and activeness). As noted in Table 3, we use four discrete class of mood adopting the methodology of [21]. Similarly, *Sleep* contains participant provided sleep duration surveys for 27 people over 21 days. Subjects carry Nexus One smartphones which collect microphone, accelerometer and again phone usage data (e.g., time spent recharging). We source *Transportation*, *Mood* and *Sleep* datasets externally, from the authors of and [34],[21] and [19] respectively. The remaining dataset *Diet* we collect ourselves. We provide 20 people with smartphones for 1 month, each phone collects GPS or WiFi data (enabling location estimates). Participants work for the same company and agree to either complete a food diary on their phone or take photos of their meals. Meals that occur at home are ignored. We determine coarse ground-truth of daily meals by coding meals from the food diary or photos into 4 meal categories (see Table 3). Three people perform coding independently, to maintain coding accuracy and consistency.

**Classifier Benchmarks.** We compare NCB to a different benchmark classifier for each dataset. Features and classification models are chosen to provide state-of-the-art performance given the dataset and target activity classes. In fact, the Personal Sensing Classifier used by NCB (see prior section) and the benchmark classifier are identical for each dataset. We refer to this classifier as `single-user`. As a result any differences in performance for NCB and `single-user` are due to the use of additional community-scale information - rather than differences in features or classifiers.

We provide an additional performance reference point by also comparing NCB with CSN [20], a recently proposed activity framework. CSN seeks to personalize classifiers for each user by exploiting inter-user similarity measurements. It requires user meta-data such as their age and sex, along with similar mobility traces to those required by NCB. Because we only have such data for *Transportation* we can provide a comparison for only this dataset. We refer to this classifier as `CSN`.

**Features and Classification Models.** Where possible, we use features and classifiers that have already been shown to be effective in recognizing the activities contained in each dataset, given the sensor data available. Our results assume NCB calculates final user inferences at the end of each day.

For *Transportation* we use a collection of purpose designed features proposed in [34]. Classification is performed using a boosted ensemble of naive Bayes classifiers, with smoothing performed using a simple Markov model [4]. Activity nodes in the hierarchical NCB network represent the most common transportation mode across 30 minute increments.

To classify sleep duration from *Sleep* we adopt a technique proposed in [19]. This approach classifies sleep duration once a day by extracting from phone usage and sensor data the frequency and duration of the following events: (1) the recharging of the phone (recharging events overnight correlate with sleep duration); and (2) prolonged contexts when the phone is both stationary and in a silent setting (this type of context is also correlated with sleep duration). We adopt a similar approach proposed in [21] to classify *Mood*. Under this technique, 60 purpose-designed features from phone usage are used to coarsely mine social interaction (e.g., SMS and phone calls) and daily activities (e.g., applications used and web browsing activity) that are associated with user mood. Final classification for both activities use a SVM model.

Finally, automated tracking of eating habits is not common. Typical solutions rely on image recognition or specialized sensors. Given the sensor data available in *Diet*, we have little existing work to leverage. Instead, we conjecture that spatial and temporal features would be sufficient to recognize a limited selection of meal types – especially if we simplify the problem by ignoring all meals eaten at home (which can be diverse and occur at the same location). Further, we anticipate people exhibit predictable meal selection habits such as selecting the same category of meal at certain restaurants and times of day. For location features we use: (1) tessellated coarse location (using 50 m sq. tiles) and (2) restaurant type (if a restaurant is returned from the FourSquare reverse geo-code API [2], which provides point-of-interest information for a provided GPS co-ordinate). For temporal features we use: (1) an indicator variable for weekend or weekday; (2) a categorical variable depending on which 3 hour window of the day the meal occurs and the day of the week; and, (3) the numbers of hours since the last meal. We provide these features to a SVM to perform meal category prediction.

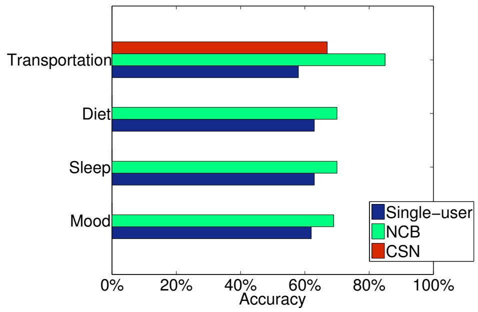



**Figure 5:** Comparison of average classification accuracy for NCB and benchmark classifiers across all datasets.

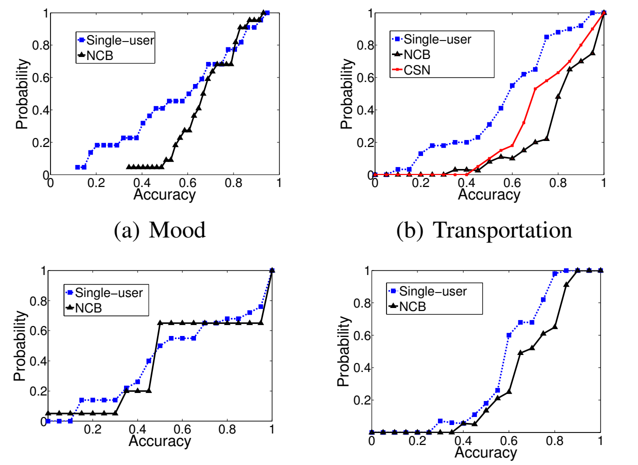
### Classifier Performance Comparisons

We begin with a set of experiments that demonstrate NCB is able to achieve higher levels of accuracy than any included benchmark. Furthermore, we find NCB is able to reduce the volatility in per-user average accuracy.

Figure 5 compares the average classification accuracy for NCB and `single-user` across all datasets. We find NCB outperforms `single-user` by 46.8%, 9.1%, 10.1%, 10.7% for *Transportation*, *Sleep*, *Diet* and *Mood* respectively. NCB and `single-user` use the same underlying features, training data and classifier design. Therefore these gains in accuracy are due to NCB incorporating awareness of community behavior into the modeling process. However, accuracy gains come at the cost of result latency, we perform NCB inference only at the end of each day while `single-user` can perform inferences at anytime.

For *Transportation*, which is the only dataset able to support the use of `CSN`, Figure 5 also shows an accuracy improvement of 31.1% for NCB compared to this benchmark. `CSN` trains separate classifiers that are personalized to the characteristics of each user. In comparison, NCB uses the same generic classifier across the entire population. We find for this dataset, NCB is able to provide improved performance over even personalized classifiers through its collaborative classification stage performed at inference-time. Because `CSN` requires crowdsourced labeled data from users (albeit in small quantities), NCB will be also useful in scenarios where the user is best not involved with the classification process.

Figure 6 presents CDFs of per-person average classifier accuracy across the entire user population, one for each dataset. Stronger classifier performance is signaled in these figures by curves shifted further to the bottom right; this indicates a larger fraction of the population experiences a high accuracy level. From these figures, we learn NCB fairly consistently provides for all users a more uniform level of classification accuracy than `single-user`. For example, Figure 6(d) shows for *Diet* NCB provides 82% accuracy to around 70% of all users. In comparison, under `single-user` the same fraction of users average around 71% accuracy; NCB outperforms `single-user`, in this case by 15%. Consistent with the results of Figure 5, `CSN` is more competitive than `single-user`. As an example, under *Transportation* NCB




(a) Mood (b) Transportation

(c) Sleep (d) Diet

**Figure 6:** CDF of per-person classification accuracy for all datasets. NCB provides more uniform classifier accuracy for the entire user population than the benchmark classifiers.

outperforms `CSN` by 9% at the 57%-mark of the user population. The same reasons discussed earlier for such comparative performance also hold in this case.

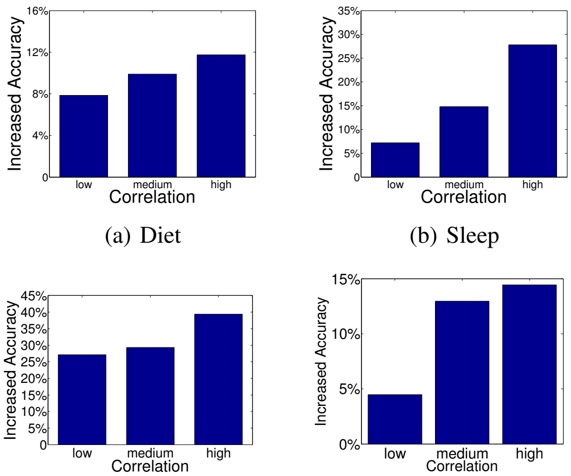
### Understanding Performance Gains

In our next set of experiments, we examine the key reasons underpinning the aggregate performance gains by NCB seen in the earlier results.

Conceptually, NCB is designed to leverage patterns of correlated behavior as a means to increase classifier accuracy. If NCB is operating as expected, we should see datasets with higher levels of correlation (within social networks) experiencing the largest gains under NCB. We calculate correlation using a 4-day moving window containing the activities that occur, with windows compared using the jaccard index (to ignore differences in ordering/timing within the window).We group users based on the percentile of correlated behavior they display, relative to other users, the bottom 20% is labeled as "low", the top 20% as "high", and the 20% straddling the 50th percentile is labeled as "medium". Table 4 presents the correlation rates for each social strength level across all datasets. This table also includes the magnitude of outperformance by NCB compared to `single-user`. When we compare the rank order of NCB outperformance with correlation we find coarse agreement. For example, *Transportation* enjoys both the largest gain due to NCB, in addition to having the largest level of correlation.

| **Dataset** | **Social Network Link Strength** | | | **NCB Gain** |
|---|---|---|---|---|
| | *Low* | *Medium* | *High* | |
| *Transportation* | 0.20 | 0.51 | 0.77 | 46.8% |
| *Mood* | 0.30 | 0.51 | 0.71 | 10.7% |
| *Sleep* | 0.48 | 0.71 | 0.79 | 9.1% |
| *Diet* | 0.08 | 0.22 | 0.58 | 10.1% |

**Table 4:** Population correlation level present in all datasets.




(a) Diet (b) Sleep

(c) Transportation (d) Mood

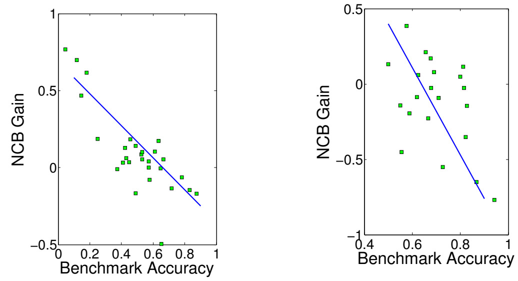
**Figure 7:** NCB increases classification accuracy for individuals with higher levels of correlated behavior within the social network

To further confirm the presence of this effect, we perform an experiment that measures the average per-user accuracy gain under NCB; and then groups users depending on the extent to which they are correlated with others in their social network. For this experiment we reuse the same correlation groups computed for Table 4. Figure 7 shows the results of this experiment across all four datasets. We consistently find those users with higher amounts of correlated activity within the social network benefit the most from NCB. For example, in Figure 7(a) those users in the highest category of social tie strength experience a classification accuracy increase of 12% on average. Similarly, in Figure 7(c) we observe the highest increases compared to any other dataset, with the "high" social tie strength group enjoying nearly a 40% increase.

NCB is also designed to leverage community-scale behavioral patterns to compensate for times when sensor data alone is difficult to interpret. To investigate this expectation we compare the relationship between the per-user classification accuracy for `single-user` and NCB itself. Figure 8 shows this relationship for the *Mood* and *Transportation* datasets. These two figures show when the benchmark classifier struggles (i.e., has low accuracy) the the gains of using NCB (compared to `single-user`) are the highest. For example, Figure 8(a) shows for those users that experience classifier accuracy *increases* of 75% or more under NCB have accuracy values under 20% when using `single-user`; in contrast, those users with `single-user` accuracy of 70% have no gains in accuracy under NCB or actually even perform worse than `single-user`. This relationship holds across all datasets with an average correlation coefficient (r) of -0.66, inline with our expectation.

### Examining Discovered Communities

We conclude our evaluation by examining the communities that NCB discovers and studying their relationship to the recognition process. Due to a lack of meta-data deep analysis is not possible, instead we provide informal observations.




(a) Transportation (b) Mood

**Figure 8:** When `single-user` classifier struggles, NCB provides its largest performance gains. Here we show representative relationships between the performance gain in using NCB compared with the accuracy of the benchmark classifiers.

*Transportation.* Gains in accuracy primarily come from communities related to what appears (from mobility data) to be workplace-based groups. Within these communities shared commute-time behavior is often valuable in correcting both difficult to classify individuals and routes. Common causes of classification errors include "bus" and "subway" that share for some routes very similar features, and periods when missing GPS data significantly skews features values.

*Sleep.* We find a few small-sized communities lead to large accuracy increases for users within these particular communities. Judging from mobility patterns, we believe these communities represent people living in the same household. However, very few of these suspected "household" communities are discovered. Without them users seem to gain very little from NCB. Again, judging from mobility patterns it appears many users either live alone or have no members of their household included in the study. We suspect this observation points to a broader issue – namely, that NCB performance will be sensitive to which sections of a user's social network contribute data; in the case of sleep, for example, it seems to be critical to gather data from the user's household.

*Diet.* Discovered communities roughly correspond to internal work-groups within the company (determined by follow-up interviews as we collect this dataset ourselves). We find NCB improves accuracy during the working week but lowers accuracy slightly during the weekend. This is because not only is the weekend more challenging due to increased diversity but as in the case of *Sleep* the dataset lacks participation from family and friends outside the workplace. However encouragingly, irrelevant communities (work-based sport clubs etc.) are weighted low for this activity class by NCB.

*Mood.* A frequent reason for low-accuracy users (see Figure 6(a)) benefiting from NCB is that incorrect inferences of extreme moods (e.g., `pa`) under `single-user` are often correct under NCB. Most inferences of `pa` under NCB are accompanied by at least some support (i.e., a similar mood change) within the user's nearby community. Importantly, NCB has only a modest negative impact to the detection of the `pa` class overall (a drop of 19%); but further study is required to understand the implications of such detection trade-offs.

## DISCUSSION

In the following section, we briefly survey additional important factors related to the NCB framework.

**Target User Populations.** We expect NCB will prove especially valuable when monitoring populations in relatively controlled environments. For example, when studying the health and wellbeing of students or tracking the productivity of employees in the work-place. Such conditions have many environmental externalities (e.g., common schedules, workloads) in addition to providing circumstances where strong social phenomena that influence behavior often develop. Although NCB is designed to cope with less homogeneous populations, we point out such scenarios as NCB gains will likely be even larger than those reported in this paper.

**Privacy.** Conventionally, activity recognition is performed using data from a single individual. NCB proves that expanding this perspective can lead to improved activity recognition. However, this approach may open new privacy risks – these risks fall into two categories. First, to build the hierarchical network used by NCB mobility or social network data is required; conventional activity models do not need this additional, potentially sensitive, data. Second, to perform the collective inference steps of NCB requires some form of centralized processing and sharing of user data – it is no longer possible for inferences to be done solely on a user's device. Privacy, remains an important open design issue for NCB.

**Latency.** Because NCB leverages the activity patterns of other users it works best when a delay is acceptable before making a final inference decision. For example, before recognizing commute or sleep patterns of any user ideally the behavior of all users is first captured using raw sensor data; as a result, delays of multiple hours might be necessary as people perform actions at different times of the day. For this reason, NCB should be used only in delay tolerant application scenarios – such as most forms of life-logging [3]. Other scenarios that are require near real time inferences are not suitable for NCB.

**Understanding Community Behavior.** In describing NCB, we highlight a variety of circumstances that lead to correlated user behavior (i.e., community behavior). However, we have been unable to systematically verify which of these circumstances underpin the particular correlated behavior NCB exploits in the datasets evaluated. (Though we have made some efforts towards this with specific observations where possible). This is a limitation of the datasets we use which lack the meta-data needed to perform such analysis. We plan in future work to perform such analysis with purpose collected data.

## RELATED WORK

Large-scale mobile sensing presents new opportunities for researchers to address the limitations of existing models of human behavior [33]. NCB is one of a growing number of frameworks (e.g., [22][18][30]) exploring such opportunities.

An emerging body of work considers the recognition of group activities (e.g., [32, 14, 13]). Often these techniques rely on coupled activity models of two or more people, but in doing so assume activities are performed at relatively the same time and place. In contrast, NCB is designed to leverage correlated behavior between socially tied groups of users; importantly, users do not have to perform activities at the same time or place for NCB to improve recognition accuracy. For example, two users may have correlated sleep durations because of work and commute constraints but the time and place they sleep may be completely different; however, NCB is still able to exploit this opportunity to improve model robustness.

Other activity recognition systems are using "groups" in less obvious ways. For example, CSN [20] mines networks of similarity between users to enable the *personalization* of classifiers based on user characteristics, such as age or sex. In fact, CSN uses co-location/proximity measurements, just as NCB does, as one of three metrics it uses to measure similarity. However NCB and CSN use this information towards completely different modeling objectives (and so neither framework shares any machine learning techniques). The objective in CSN is to train personalized classifiers with lower amounts of training data by selectively sharing data among users. In contrast, NCB is most novel at *inference* time. NCB learns how socially connected users manifest collective patterns of behavior, such as reacting similarly towards the actions of each other and their environment. NCB is able to couple this information with sensor data collected directly from the user to infer activities. But unlike the novel training process of CSN, NCB uses standard classifier training.

Aggregate behavior within organizations have been studied from the perspective of understanding, for example, information flows [27]; techniques of this type, that use sensors to measure group interactions automatically, are complementary input signals to the NCB framework. In [23] a similar social process is captured, measuring the spread of political opinions, but for the purpose of understanding the phenomena rather than leveraging it in activity recognition. Similarly, [30] demonstrates how structural information from social networks can assist in the classification of user personality types. However, NCB is the first to examine the use of community behavior to improve general-purpose activity recognition.

Finally, the current design of NCB is grounded in the use of relational classifiers and collective inference. These techniques have been developed to exploit situations of networked data segments requiring classification which are found in, to name just two, online social networks [36] and citation linked articles [35]. To the best of our knowledge, we are the first to use these techniques in the context of activity recognition.

## CONCLUSION

In this paper, we have proposed and evaluated the Networked Community Behavior framework (NCB) for activity recognition. The key contribution of NCB is that it examines an important new direction in the modeling of human behavior. Under NCB, activity recognition operates not only at the conventional personal sensing scale using sensor data gathered solely from individual users; but it also incorporates critical signals observable only at the community-level. With this increased visibility over the whole population, NCB is able to leverage aggregate community-wide behavioral patterns to increase activity inference robustness.

## REFERENCES


1. Sun, J., and Tang, J. A Survey of Models and Algorithms for Social Influence Analysis. *Social Network Data Analytics*, (2011) 177–214.
2. FourSquare Search API. `http://developer.foursquare.com/`.
3. Bell, G., and Gemmell, J. A Digital Life. *Scientific American*, 296:3, (2007), 58–65.
4. Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.
5. Choudhury, T. K. Sensing and Modeling Human Networks. *PhD thesis*, Massachusetts Institute of Technology, 2003.
6. Christakis, N. A., and Fowler, J. H. The Spread of Obesity in a Large Social Network over 32 years. *New England Journal of Medicine*, 357:4, (2007), 370–379.
7. Christakis, N. A., and Fowler, J. H. The Collective Dynamics of Smoking in a Large Social Network. *New England Journal of Medicine*, 358:21 (2008), 2249–2258.
8. Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., and Landay, J. A. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. In *Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI '08 (2008), 1797–1806.
9. Cranshaw, J., Toch, E., Hong, J., Kittur, A., and Sadeh, N. Bridging the Gap between Physical Location and Online Social Networks. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10 (2010), 119–128.
10. daF. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. Characterization of Complex Networks: A Survey of Measurements. *Advances in Physics*, (56):1 (2005), 167–242.
11. Getoor, L., and Taskar, B. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.
12. Girvan, M., and Newman, M. E. J. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*, (99):12 (2002), 7821–7826.
13. Gordon, D., Hanne, J.-H., Berchtold, M., Shirehjini, A. A. N., and Beigl, M. Towards Collaborative Group Activity Recognition using Mobile Devices. *Journal of Mobile Networks and Applications*, (18):3 (2013), 326–340.
14. Gu, T., Wang, L., Chen, H., Tao, X., and Lu, J. Recognizing Multiuser Activities using Wireless Body Sensor Networks. *IEEE Transactions on Mobile Computing*, (10):11 (2011), 1618–1631.
15. Hummel, R., and Zucker, S. On the Foundations of Relaxation Labeling Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3 (1983), 267–287.
16. Kelman, H. Compliance, Identification, and Internalization: Three Processes of Attitude Change. *Journal of Conflict Resolution* 2 (1958), 51–60.
17. Kempe, D., Kleinberg, J. M., and Tardos, É. Influential Nodes in a Diffusion Model for Social Networks. In *Proceedings of the 32nd International Conference on Automata, Languages and Programming*, ICALP '05 (2005), 1127–1138.
18. Lane, N. D., Xu, Y., Lu, H., Eisenman, S., Choudhury, T., and Campbell, A. Cooperative Communities (CoCo): Exploiting Social Networks for Large-scale Modeling of Human Behavior. *IEEE Pervasive Computing*. 10(4):(2011), 45–53.
19. Lane, N. D., Mohammod, M., Lin, M., Yang, X., Lu, H., Ali, S., Doryab, A., Berke, E., Choudhury, T., and Campbell, A. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In *Proceedings of ICST Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '11 (2011).
20. Lane, N. D., Xu, Y., Lu, H., Hu, S., Choudhury, T., Campbell, A. T., and Zhao, F. Enabling Large-scale Human Activity Inference on Smartphones using Community Similarity Networks (CSN). In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11 (2011), 355–364.
21. LiKamWa, R., Liu, Y., Lane, N. D., and Zhong, L. Can Your Smartphone Infer Your Mood? In *Proceedings of the 2nd International Workshop on Sensing Applications on Mobile Phones*, PhoneSense '11 (2011).
22. Longstaff, B., Reddy, S., and Estrin, D. Improving Activity Classification for Health Applications on Mobile Devices using Active and Semi-Supervised Learning. In *Proceedings of ICST Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '10 (2010).
23. Madan, A., Farrahi, K., Gatica-Perez, D., and Pentland, A. Pervasive Sensing to Model Political Opinions in Face-to-Face Networks. In *Proceedings of the 9th International Conference on Pervasive Computing*, Pervasive'11 (2011), 214–231.
24. Mason, W., Conrey, F., and Smith, E. Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence Within Social Networks. *Personality and Social Psychology Review*, 11(3):(2007), 279.
25. McDowell, L., Gupta, K. M., and Aha, D. W. Cautious Collective Classification. *Journal of Machine Learning Research* 10:(2009), 2777–2836.
26. McPherson, M., Smith-Lovin, L., and Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):(2001), 415–444.

27. Olguín, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., and Pentland, A. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man and Cybernetics*, Part B 39(1):(2009), 43–55.

28. Rosenfeld, A., Hummel, R., and Zucker, S. Scene Labeling by Relaxation Operations. *IEEE Transactions on Systems, Man and Cybernetics*, 6:(1976), 420–433.

29. Russell, J. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6):(1980), 1161–1178.

30. Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., and Pentland, A. Friends Don't Lie: Inferring Personality Traits from Social Network Structure. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12 (2012), 321–330.

31. Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. Human Mobility, Social Ties, and Link Prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11 (2011), 1100–1108.

32. Wang, L., Gu, T., Tao, X., Chen, H., and Lu, J. Recognizing Multi-user Activities using Wearable Sensors in a Smart Home. *Journal of Pervasive and Mobile Computing*, 7(3):(2011), 287–298.

33. Lane, N. D. Community-aware Smartphone Sensing Systems. *IEEE Internet Computing*, 16(3):(2012), 60–64.

34. Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09 (2009), 791–800.

35. Shi, X., Li, Y., and Yu, P. Collective Prediction with Latent Graphs In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11 (2011), 1127–1136.

36. Henderson, K., Gallagher, B., Li, L., Akoglu, L., Eliassi-Rad, T., Tong, H., and Faloutsos, C. It's Who You Know: Graph Mining Using Recursive Structural Features. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11 (2011), 663–671.