



**HAL**  
open science

## Reality mining: digging the impact of friendship and location on crowd behavior

Yuanfang Chen, Antonio Ortiz Torres, Noel Crespi, Lei Shu, Lin Lv

► **To cite this version:**

Yuanfang Chen, Antonio Ortiz Torres, Noel Crespi, Lei Shu, Lin Lv. Reality mining: digging the impact of friendship and location on crowd behavior. MOBIQUITOUS 2013: 10th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Dec 2013, Tokyo, Japan. pp.142 - 154, 10.1007/978-3-319-11569-6\_12 . hal-01285971

**HAL Id: hal-01285971**

**<https://hal.science/hal-01285971v1>**

Submitted on 10 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Reality Mining: Digging the Impact of Friendship and Location on Crowd Behavior

Yuanfang Chen<sup>1</sup>, Antonio M. Ortiz<sup>1</sup>, Noel Crespi<sup>1</sup>, Lei Shu<sup>2</sup>, and Lin Lv<sup>3</sup>

<sup>1</sup> Institut Mines-Télécom, Télécom SudParis, France

<sup>2</sup> Guangdong University of Petrochemical Technology, China

<sup>3</sup> School of Software, Dalian University of Technology, China

{yuanfang.chen, antonio.ortiz\_torres, noel.crespi}@telecom-sudparis.eu, lei.shu@lab.gdupt.edu.cn, lvlin1023@gmail.com

**Abstract.** Crowd behavior is a natural instinct of human, which directly impacts how we form opinions and make decisions. It is a subject that deserves to be studied since it is common that people change their behavior when being in a group. In pervasive computing research, plenty of work has been directed towards discovering human movement patterns based on wireless networks, mainly focusing on movements of individuals. It is surprising that social interactions among individuals in a crowd is largely neglected. Mobile phones offer on-body tracking and they are already deployed on a large scale, allowing the characterization of user behavior through large amounts of wireless information collected by mobile phones. In this paper, we observe and analyze the impact of friendship and location attributes on crowd behavior, using location-based wireless mobility information. This is a cornerstone for predicting crowd behavior, which can be used in a large number of applications such as traffic management, crowd safety, and infrastructure deployment.

**Key words:** Crowd behavior, Mobile phones, Wearable computing, Complex social networks

## 1 Introduction

With the increasing size and frequency of mass events, such as traffic congestion on a highway, swarming at a tourist attraction, and clogging at weekend shopping sale, the study of crowd dynamics has become an important research area [28, 7]. However, even successful modeling approaches such as those inspired by Newtonian force models are still not fully consistent with empirical observations and are sometimes difficult to be adapted for crowd prediction. With the prevalence of smart devices (such as Smart Cellphone, Tablet PC, etc.), on-body sensing, computing and communication have become widespread [16] (carry-on smart devices can be called “social sensors”). These developments have made it possible to obtain real-time and comprehensive empirical data required by crowd behavior analyses [24]. This “reality mining” [18] is deemed adequate to provide objective measurements of human interactions, and it can be called “honest signals” [25]. These options open up new pathways in Computational Social Science [17], where large amounts of information obtained from wireless mo-

mobile devices can lead to new perspectives for the analysis of crowd behavior and social dynamics.

In order to develop reliable prediction models for traffic management, urban infrastructure deployment, or crowd safety, it is necessary to understand what laws determine the formation of a crowd. While a lot of studies know the “physics” of crowd behavior, it is surprising that social interactions among individuals in a crowd has been largely neglected. Indeed, the great majority of existing studies investigate a crowd as a collection of isolated individuals, and each individual has its own motion speed and direction [20, 1]. However, it turns out that in practice, the majority of individuals do not take actions alone, but in groups with social relationships [26, 10].

In this paper, we first focus on recognizing human crowd behavior by analyzing the data measured by internet-accessible mobile phones from a location-aware online social network. By crowd behavior recognition, we understand that the movement of a large number of individuals has a pattern and can be attributed, depending on relevant parameters such as the friendship between individuals and check-in locations (with time) of these individuals. Based on the recognition, we can realize that it is possible to predict the formation of a crowd from wireless information related to individuals. For instance, the formation of downtown pedestrian flows is related to the interactions between individuals, and these flows can be distinguished by collecting the wireless information of pedestrians.

Second, we investigate how “friendship” and “location” impact human crowd behavior. Nathan Eagle *et al.* have obtained an important result: “Data collected from mobile phones has the potential to provide insight into the relational dynamics of individuals. Furthermore, it is possible to accurately infer 95% of friendship relations only considering the observational wireless data” [9]. This implies that there is a relationship between the friendship and the behavioral patterns of humans. Based on this result, we observe the patterns of friendship in different crowds.

The contributions of this paper are listed as follows.

- We design a crowd recognition model. One of the main challenges in crowd behavior recognition is to infer the most likely crowd behavior using the data collected from a set of persons. We use check-in time and location (*Time* and *Location\_id*). We convert each latitude/longitude coordinate of the earth into a unique *Location\_id* to quantify the track of each individual. Then, a clustering algorithm<sup>1</sup> is used to find the likely crowds.
- We investigate how friendship influences crowd behavior. In order to measure this influence, we use the friendship degree of each user and the probability distribution of various degrees.
- We investigate how individuals’ locations influence crowd behavior. For measuring the influence, we investigate these relationships for different clusters: (i) users and their locations; (ii) check-in time and users’ locations.

---

<sup>1</sup> An Expectation-Maximization (EM) clustering algorithm is used in this paper. The EM assigns a probability distribution for each track record (instance), which indicates the probability of each instance belonging to each of the clusters. The EM can automatically decide how many clusters to create.

The paper is structured as follows. Section 2 introduces the related work. Section 3 presents the data used in our study and statistic analysis. Section 4 shows how to recognize human crowd behavior from datasets. Section 5 reveals the impact of attributes (check-in time, check-in location and friendship) on crowd behavior. Finally, some conclusions are given in Section 6.

## 2 Related Work

Recently, a number of scientific communities, from computer science to physics, have been working in human dynamics. Pedestrian movement patterns have been studied using wireless-based personal location data [23]. In physics, for human crowd behavior analysis, many approaches have been proposed inspired by using fluid dynamics [13], swarms [3] and cellular automata [2]. Car-based human movement patterns have also been studied by utilizing the data from GPS-equipped vehicles [27]. Thereinto, the approaches which are based on wireless mobility information are more reliable, objective and environment-independent compared with model-based approaches, e.g., habit-based model [15]. Model-based approaches are environment-sensitive. Moreover, because the factors of environment influence each other, a simple model is not sufficient to reflect the interactions between these factors. Furthermore, the performance of a model is related to the experience of modeler.

As an important aspect of human dynamics, “crowd dynamics” is worthy to be deeply analyzed, since it helps to extract some useful conclusions about how humans behave when they are in large groups. Further, predicting the formation of a crowd is helpful in some emergency situations, e.g., evacuation route control; and even the prediction is also beneficial for studying and improving the performance of public infrastructures, e.g., network usage during a mass event. However, the prediction of crowd events is still a challenge, even if a lot of new technologies can be used, e.g., GPS-based human trace tracking technology. Moreover, GSM, bluetooth or WiFi localization technologies have been explored to be used to collect sufficient data for analyzing crowd behavior [6].

The prediction of human behavior is the main topic of a number of publications. The method proposed in [14] estimates an object’s future locations, by considering the patterns of recent movements. They present the concept of a Trajectory pattern (T-pattern), a special association rule with a timestamp that is able to define a sequence of locations with a certain probability. They also propose the Trajectory Pattern Tree (TP-T), a data access method that indexes trajectory patterns to efficiently answer predictive queries, and finally, they detail a Hybrid Prediction Algorithm (HPA) that provides accurate prediction for both near and distant time queries. T-patterns are also used in [19], where *WhereNext* is presented, which is a technique to predict the next location of a moving object. It uses an evaluation function that efficiently creates TPTs by considering the previous movements of all moving objects in a certain area. The model presented in [21] is based on behavioral heuristics that predict individual trajectories and collective motion patterns such as the spontaneous formation of unidirectional lanes or stop-and-go waves. These heuristics consider visual information to describe the motion of pedestrians.

Most of the previous research in crowd dynamics has ignored the internal connection between social relationships and crowd formation. While discrete observations of an individual’s idiosyncratic behavior seem to be merely random, and most studies of crowd behavior consider only interactions among isolated individuals, the results presented in [22] show that up to 70% of people in a crowd are actually moving in groups such as friends, couples, or families, concluding that the social relationships affect crowd forming. In addition, group sizes are commonly distributed according to a poisson distribution [12]. Thus, social ties between individuals impact the forming of crowd. In this work, we consider friendship and location relationships between individuals as an important parameter for crowd recognition.

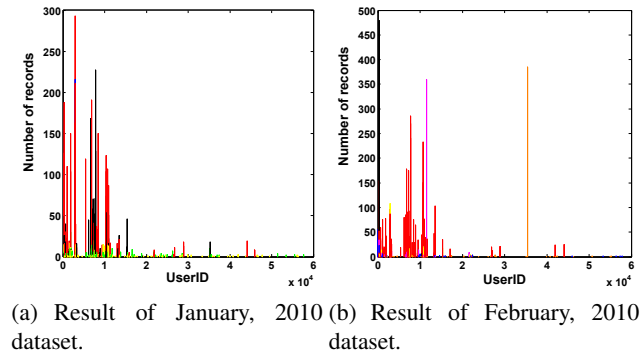
### 3 Data Description

The dataset used for analyzing crowd behavior consists of anonymous check-in data from mobile devices collected by a location-based social networking service provider where users share their locations by check-in, and the friendship between these mobile users is collected using their public API [5]. This aggregated and anonymous mobile device information is used to correlate, model, evaluate and analyze the relationships between the check-in time, locations, friendship and crowd behavior of users in 772, 966 distinct places. The dataset consists of 58, 228 nodes (users) and 214, 078 friend edges (friendship is directed between any two nodes).

**Check-in behavior of users.** Based on users’ check-in behavior, we can obtain users’ locations, and our recognition and observation are check-in-location-aware. Moreover, through check-in frequency, some special places can be inferred. For example, the home location of a user can be defined as the location which has maximum average number of check-ins for a period of time. Manual inspection shows that this infers home locations with 85% accuracy [5]. We can deduce that there is a relationship between users’ social relations (e.g., the kinship of users at home) and check-in locations with check-in frequency.

**Friendship and mobility.** Our study aims at understanding how the location of user  $A$ ’s friend  $B$  affects the movement of  $A$ . Intuitively we are more likely to move to a place in which we have friends (crowd behavior is a natural instinct of human). To quantify this effect we proceed as follows. User  $A$  “visits” the location of friend  $B$ , if  $A$  checks in within radius  $r$  of  $B$ ’s location, and we aim at computing  $p(d)$ , which measures the probability that  $A$  “visits” a friend (near the friend) within the range,  $r = d$  (we set  $r = 10m$ ). In the dataset, using the “location\_id”, we can draw the “ranges” corresponding to different friends and obtain relevant probabilities to which range  $A$  belongs. Moreover, In Fig. 1, we show, for each user, the number of records and to which cluster belongs based on January, 2010 dataset (Fig. 1(a)) and February, 2010 dataset (Fig. 1(b)).

The original dataset provided by the service provider can not be directly used to mine the basic laws that govern human crowd behavior, or even the impact of these laws, so we apply a process to perform an association of the mobility check-in data with the friendship information. The process involves two steps:



**Fig. 1.** Number of records and which cluster belongs for each user in the January and February, 2010 datasets (different colours represent different clusters).

1. Perform a spatial-temporal analysis of the records to detect which users form a crowd (e.g., places in which some people have stopped for a sufficiently long time);
2. Infer in every crowd which users are friends and what kind of friendship they have (1, 2, 3, ...,  $N$ -hop friends, e.g., “1-hop” is for direct friends).

In order to infer “where and when crowds are forming?” from a large number of records, we first characterize the individual activity by mathematical modeling. Each location measurement  $m_i$ , collected for every mobile device, is characterized by a position  $p_i$  expressed in latitude and longitude, and a timestamp  $t_i$ . We also measure the interval time between different check-in activities [11]. The average interval time measured for the whole population is 30 minutes. So within the interval we can detect a gathering of humans from the dataset. Moreover, this time interval is comparable to the average length of real social events.

To confirm the friendship in a crowd, we merge the friendship dataset with the check-in record dataset, e.g., for a crowd, if there is a record of any two users in the friendship dataset, the two users are 1-hop friends, and then we add this information into the check-in record dataset as a value of a new attribute column (friendship). Multihop friendship will be also recognized, and we record the identity number of minimum-hop-count friend for every user as the value of the friendship attribute column.

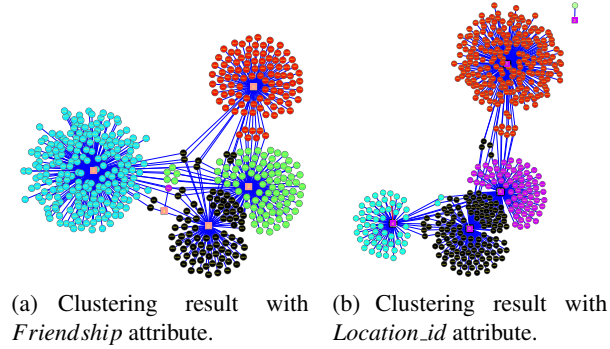
## 4 Crowd Behavior Recognition

We formalize a series of processing steps which can be used to infer crowd behavior from location-based wireless information. In this section we build a mathematical model for recognizing the crowd behavior of population.

First, based on the processed dataset, we confirm whether the crowd behavior of individuals can be identified. Fig. 2 shows the results of data clustering with *Friendship* and *Location\_id* attributes, respectively (these two attributes are evaluation classes for clustering), using Expectation-Maximization (EM) algorithm [8]. From Fig. 2, the

crowd behavior has been recognized using the processed dataset (in our study a crowd is defined in Definition 1). Moreover, we can find that different evaluation classes have different clustering accuracy levels, so the impacts of different attributes on the crowd behavior are different. Fig. 3 shows the number of users for each cluster corresponding to Fig. 2 (Fig. 2 only shows the users who are “1-hop” friends, but in Fig. 3, the users who have multihop friendship are also counted for each cluster).

**Definition 1 (A Crowd).** A group of individuals at the same physical location (the range radius  $r = 10m$ ) at the same time.



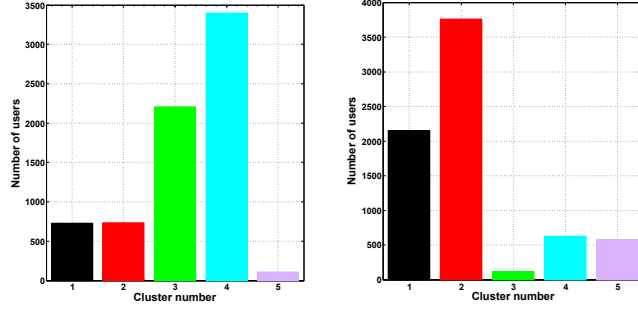
**Fig. 2.** For clearness, we only show one-month clustering result using *Friendship* and *Location\_id* attributes as the classes of clustering evaluation, respectively. 5 clusters can be found for two clustering processes, and we use different colours to distinguish different clusters. Note that the shape for the center node of any cluster is square.

The characteristics of the crowd behavior of each single person can thus be inferred from his/her check-in records. We refer to this as their “individual behavior”. This shows which individuals participate in a specific crowd. From the EM algorithm, a given record belongs to each cluster with certain probabilities. Moreover, the likelihood is a measurement of “how good” a clustering process is and it is increased at each iteration of the EM algorithm. It is worth mentioning that the higher the likelihood, the better the model fits the data.

Second, the clustering process (crowd behavior recognition model) is as follows. We define two parameters: (i) the user  $u$ ’s check-in data  $S^u$  which is a sequence of activity observations about  $u$ ; (ii) a set of unknown values  $\theta$  (i.e., the serial numbers of clusters). These two parameters are used along with a Maximum Likelihood Estimation (MLE):  $L(\theta; S^u) = p(S^u|\theta)$ . Our purpose is to seek the MLE of marginal likelihood. In other words, we need to find the most probable  $\theta$  which the user  $u$  belongs. The EM algorithm iteratively applies the following two steps to achieve our purpose:

1. Expectation step (E step): calculate the expected value of the log-likelihood function under the current established clusters ( $\theta^{(t)}$ ):

$$Q(\theta|\theta^{(t)}) = E_{S^u, \theta^{(t)}}[\log L(\theta; S^u)];$$



(a) Number of users for each cluster using *Friendship* attribute as the class of clustering evaluation.

(b) Number of users for each cluster using *Location\_id* attribute as the class of clustering evaluation.

**Fig. 3.** Number of users for each cluster.

2. Maximization step (M step): find the appropriate value of parameter  $\theta$ , which maximizes this quantity:

$$\theta^{mle} = \arg \max_{\theta} Q(\theta|\theta^{(t)}).$$

MLE estimates  $\theta$  by finding a value of  $\theta$  that maximizes  $Q(\theta|\theta^{(t)})$ , and the estimation result can be flagged as:  $\theta^{mle}$ .

Further, based on our recognition model, we add the spatio-temporal pattern into the crowd behavior (clustering). We use a triple  $q = (\theta, p_i, t_i)$  to replace  $\theta$ . Then the expectation step becomes:

1. Expectation step:

$$Q(q|q^{(t_i)}) = E_{S^u, q^{(t_i)}}[\log L(q; S^u)],$$

where  $p_i$  is the position characteristic of location measurement  $m_i$ , and  $t_i$  is timestamp of  $m_i$ . Moreover,  $q^{(t_i)}$  is a set of current established clusters with their locations and timestamps;

2. Maximization step: choose  $q$  to maximize  $Q(\cdot)$ ,

$$q^{mle} = \arg \max_q Q(q|q^{(t_i)}).$$

Finally, in order to preserve the model integrity, the recognition accuracy must be measured. In our crowd behavior recognition model, the value of the log-likelihood can be used to measure the accuracy. For instance, using the *Friendship* attribute as the evaluation class, based on one-month data, the log likelihood of crowd identification is:  $-16.42186$ , and using the *Location\_id* attribute as the evaluation class, the log likelihood is:  $-16.87742$ . Their accuracy is different, and the *Friendship* attribute is more effective for improving the recognition ability of the model.

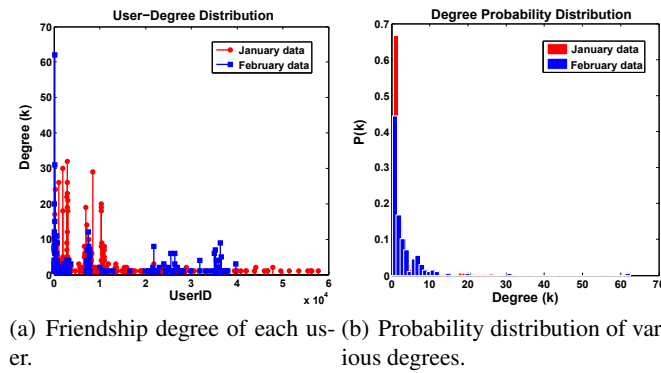


## 5 Impact of Attributes on Crowd Behavior

Most previous studies of crowd behavior only consider interactions among isolated individuals, and assume that the motion of individuals is random. The work in [22] affirms that the walking behavior of pedestrians is affected by social relationships, such as friends, couples, or families walking together. The results presented in Fig. 2, show that some population attributes impact the crowd motion of humans, and these impacts are different for diverse attributes. In this section, we go deeper into the impact of these effects by analyzing how friendship and location affect crowd behavior.

### 5.1 Impact of Friendship on Crowd Behavior

In this section, we analyze the impact of social relationship (friendship) on the complex dynamics of crowd behavior. For this, we use the empirical data of the motion of individuals by means of check-in recordings of public areas. Observations are made under varying-density collections of population.

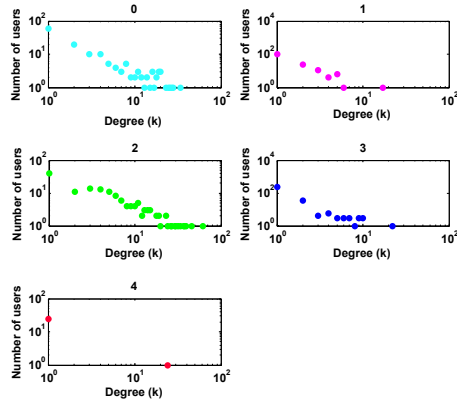


**Fig. 4.** Friendship degree of each user and probability distribution of various degrees (January and February, 2010 datasets).

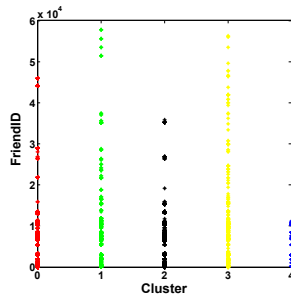
First of all, we analyze the friendship of a group of people. Fig. 4 shows the friendship degree of each user and the probability distribution of various degrees (based on the data of January and February, 2010). We can find that friendship exists in almost all observed users. It means that friendship is an important influencing factor for human behavior. It is worth noting that a Poisson distribution is met for the friendship degree probability distribution of all observed users. Moreover, from Fig. 4(b), friendship degree is less than 10 for almost 90% users. So the crowding propensity of an individual is not primarily oriented by friendship.

Secondly, several crowds exist in the observed group of people and we investigate the friendship degree probability distribution of each crowd. Fig. 5 shows the friendship degree distributions of 5 clusters which are based on January, 2010 dataset. For some

crowds, friendship degrees show approximative scale-free power-law distributions, e.g., crowds 0, 1, 2, and 3.



**Fig. 5.** Relationship between friendship degree and number of users for each crowd (January, 2010 dataset).



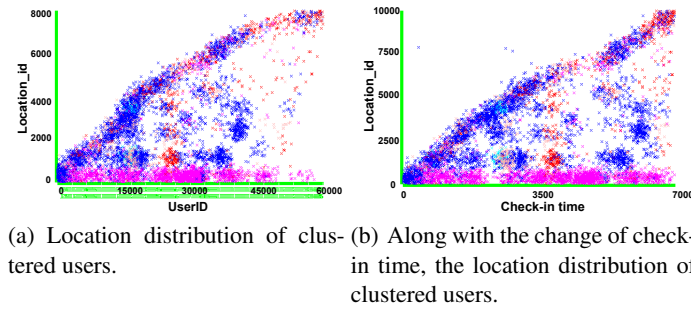
**Fig. 6.** IDs of users (there is friendship among these users) for each crowd.

Finally, how many users have friendship between each other for each cluster? Fig. 6 exposes an interesting phenomenon: even though the average friendship degree of cluster 2 is larger than that of cluster 3 (shown in Fig. 5), the number of users who have friendship in cluster 2 is less compared to cluster 3. This means that there is no relation between the number of users (who have friendship) and friendship degree in a crowd.

### 5.2 Impact of Location on Crowd Behavior

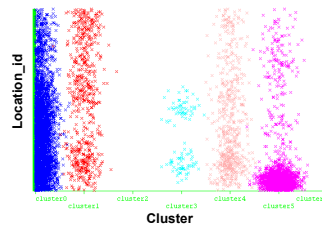
With the increasing ubiquity of location sensing included in mobile devices, we realize the arising opportunity for human crowd analysis through mobile wireless information from the real world context. Thereby, the spatial locations of individuals become a

kind of important and available information for extending the analysis and modeling of human crowd behavior to the physical world. Most prediction models of crowd events use some kind of location information. In [4], Calabrese *et al.* believe that the attendees of crowd events are related to areas: “Sport events such as baseball games attract about double the number of people which normally live in the Fenway Park area. Moreover, those events seem to be predominantly attended by people living in the surrounding of the baseball stadium, as well as the south Boston area”. From our investigation, we can find the impact of location on crowd behavior. Fig. 7 shows *location\_id* distribution of users.



**Fig. 7.** Relationship among users, check-in time, clusters and locations (different colours denote different clusters).

Analyzing Fig. 7, we can see that users are obviously clustered, and an approximate diagonal line divides the User-Location (see Fig. 7(a)) and Check-in time-Location (see Fig. 7(b)) space. Moreover, we can observe: the range of activity for any user is limited; and duration time is different for different crowds in different locations. We can assert that human crowd behavior is centralized; namely “hot spot” exists. We use this character of crowd behavior to do centralized population monitoring or urban public facility deployment.



**Fig. 8.** Location distribution of users in each cluster.

Figure 8 shows the location distribution of users for each crowd. We can find that the distances between users are different (the location distribution of users is uneven) in

any crowd. For instance, in cluster 3, there are two small subclusters. We believe that the *small-world phenomenon* exists in crowd behavior.

## 6 Conclusion

The analysis of crowd behavior can give us an idea of how we behave when we are part of a group. Different actions can be taken by individuals when being surrounded by others. In this paper, we have focused on analyzing crowd behavior, considering check-in data from location-aware mobile social networks.

There are several parameters that affect the forming way of a crowd. Social relationships, e.g., friendship, as well as current locations, can impact the way that a crowd is formed. Understanding how these parameters determine crowd formation and evolution is the first step prior to the creation of predictive approaches. After grasping the relationship between these parameters and crowd behavior, our ongoing research focuses on the prediction model of crowd behavior. This model will be useful for preventing disasters from the pushing of crowds, facilitating efficient massive event planning, or even for traffic management.

## Acknowledgment

This work was supported by the EU ITEA 2 Project 11020, “Social Internet of Things-Apps by and for the Crowd” (SITAC).

## References

1. Antonini, G., Bierlaire, M., Weber, M.: Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological* **40**(8), 667–687 (2006)
2. Bandini, S., Manzoni, S., Vizzari, G.: Crowd behaviour modeling: from cellular automata to multi-agent systems. *Multi-Agent Systems: Simulation and Applications* pp. 204–230 (2009)
3. Bellomo, N.: Modeling crowds and swarms: congested and panic flows. *Modeling Complex Living Systems* pp. 169–188 (2008)
4. Calabrese, F., Pereira, F., Di Lorenzo, G., Liu, L., Ratti, C.: The geography of taste: analyzing cell-phone mobility and social events. *Pervasive Computing* **6030**, 22–37 (2010)
5. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: *Proceedings of the 17th SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1082–1090. ACM (2011)
6. Cook, D.J., Das, S.K.: Pervasive computing at scale: transforming the state of the art. *Pervasive and Mobile Computing* **8**(1), 22–35 (2012)
7. Davis, L.: Predicting travel time to limit congestion at a highway bottleneck. *Physica A: Statistical Mechanics and its Applications* **389**(17), 3588–3599 (2010)
8. Do, C.B., Batzoglu, S.: What is the expectation maximization algorithm? *Nature biotechnology* **26**(8), 897–899 (2008)

9. Eagle, N., Pentland, A.S., Lazer, D.: Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* **106**(36), 15,274–15,278 (2009)
10. Ge, W., Collins, R.T., Ruback, R.B.: Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(5), 1003–1016 (2012)
11. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. *Nature* **453**(7196), 779–782 (2008)
12. Griesser, M., Ma, Q., Webber, S., Bowgen, K., Sumpter, D.J.: Understanding animal group-size distributions. *PloS one* **6**(8), e23,438:1–9 (2011)
13. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. *Physical review E* **51**(5), 4282–4286 (1995)
14. Jeung, H., Liu, Q., Shen, H.T., Zhou, X.: A hybrid prediction model for moving objects. In: *Proceedings of the 24th International Conference on Data Engineering*, pp. 70–79. IEEE (2008)
15. Jiao, Y., Liu, Y., Wang, J., Wang, J.: Model for human dynamics based on habit. *Chinese Science Bulletin* **55**(24), 2744–2749 (2010)
16. Ko, M.H., West, G., Venkatesh, S., Kumar, M.: Online context recognition in multisensor systems using dynamic time warping. In: *Intelligent Sensors, Sensor Networks and Information Processing Conference*, pp. 283–288. IEEE (2005)
17. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. *Science* **323**(5915), 721–723 (2009)
18. Mitchell, T.M.: Mining our reality. *Science* **326**(5960), 1644–1645 (2009)
19. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: Wherenext: a location predictor on trajectory pattern mining. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 637–646. ACM (2009)
20. Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., Theraulaz, G.: Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences* **276**(1668), 2755–2762 (2009)
21. Moussaïd, M., Helbing, D., Theraulaz, G.: How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences (PNAS)* **108**(17), 6884–6888 (2011)
22. Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G.: The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS one* **5**(4), e10,047:1–7 (2010)
23. Paul, U., Subramanian, A.P., Buddhikot, M.M., Das, S.R.: Understanding traffic dynamics in cellular data networks. In: *Proceedings of the 30th International Conference on Computer Communications*, pp. 882–890. IEEE (2011)
24. Pentland, A., Choudhury, T., Eagle, N., Singh, P.: Human dynamics: computation for organizations. *Pattern Recognition Letters* **26**(4), 503–511 (2005)
25. Pentland, A.S., Pentland, S.: *Honest signals: how they shape our world*. MIT press (2008)
26. Tang, L., Liu, H.: Toward predicting collective behavior via social dimension extraction. *Intelligent Systems* **25**(4), 19–25 (2010)
27. Trasarti, R., Pinelli, F., Nanni, M., Giannotti, F.: Mining mobility user profiles for car pooling. In: *Proceedings of the 17th SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1190–1198. ACM (2011)
28. Zhang, X., Weng, W., Yuan, H., Chen, J.: Empirical study on unidirectional dense crowd during a real mass event. *Physica A: Statistical Mechanics and its Applications* **392**(12), 2781–2791 (2013)