



Correlating digital footprints for discovering social connections in crowds

Javier A Espinosa-Oviedo, Genoveva Vargas-Solar, Vassil A Alexandrov, José L Zechinelli-Martini

► To cite this version:

Javier A Espinosa-Oviedo, Genoveva Vargas-Solar, Vassil A Alexandrov, José L Zechinelli-Martini. Correlating digital footprints for discovering social connections in crowds. Proceedings of the 3rd IEEE International Conference on Smart Data , Jun 2017, Exeter, United Kingdom. hal-01583808

HAL Id: hal-01583808

<https://hal.science/hal-01583808>

Submitted on 7 Sep 2017

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.

Correlating digital footprints for discovering social connections in crowds

Javier A. Espinosa-Oviedo¹⁵ Genoveva Vargas-Solar³⁴⁵ Vassil Alexandrov¹²⁶ José L. Zechinelli-Martini⁵⁷
javier.espinosa@bsc.es genoveva.vargas@imag.fr vassil.alexandrov@bsc.es joseluis.zechinelli@udlap.mx

Barcelona Supercomputing Centre, BSC-CNS, Spain¹

Catalan Institution for Research and Advanced Studies, ICREA, Spain²

French Council of Scientific Research, CNRS, France³

Laboratory of Informatics of Grenoble, LIG, France⁴

French-Mexican Laboratory of Informatics and Automatic Control, LAFMIA⁵

Tecnológico de Monterrey, ITESM, Mexico⁶

Fundación Universidad de las Américas, Puebla, UDLAP, Mexico⁷

Abstract—This paper presents an approach for guessing the degree of “social” connection among individuals and groups moving and evolving within real environments that uses information gleaned from omnipresent surveillance and individuals digital foot prints. This task is supported by observations of individuals’ behaviour within urban spaces when they are alone, and when they are part of a group (i.e., crowd). This knowledge is used as reference to predict their possibility of forming communities and how they can establish relationships with other communities in the presence of specific events (e.g., alarm, disaster). We address the challenge of combining an individual’s location with a real-time graphic vision of the urban environment she is moving within, using data produced by GPS, mobile and telephone networks, and security cameras. Discovering communities within crowds leads to social graphs. Computing and processing these graphs requires computing and memory resources we therefore on our HPC infrastructure for performing tests.

Graph storage; graph analytics; social connections in crowds; data correlation.

I. INTRODUCTION

We are living in the era of massive exodus and delicate security issues, where millions of children are lost in Europe, adults migrating from the LATAM region to North America are missing, women disappear without trace as they flee from war, hunger, disease and very bad local economic and political situations. These people certainly act as individuals but also as part of crowds that are moving across different regions. These exoduses create security problems that countries must deal with, to provide aid as they cross their territories, and particularly to be sure that they are not contacted by mafias that take advantage from their situation to act against society. We believe that technology, and particularly data management and visualization, can provide tools that can help to control this complex situation. For instance, by providing tools for monitoring exoduses’ behaviour (i.e., on-line and post-mortem analytics), one could identify behavioural patterns and adapt decision making in real-time.

This paper presents an approach for guessing the degree of “social connection” among individuals and groups evolving within real environments that uses information gleaned from omnipresent surveillance and individuals digital foot prints. This task is supported by observations of individuals’ behaviour within urban spaces when they are alone, and when they

are part of a group (i.e., crowd). This knowledge is used as reference to predict their possibility of forming communities and how they can establish relationships with other communities in the presence of specific events (e.g., alarm, disaster). We address the challenge of combining an individual’s location (like the one performed by the application Friends by Google) with a real-time graphic vision of the urban environment she is moving within (like the one provided by Google earth) using data produced by GPS, mobile and telephone networks and security cameras.

The level of detail depends on the size of individuals’ digital footprints, on the degree of access control associated to these data, on her access rights to personal information and on the privacy laws of the geographic space in which observed people is moving. For example, a consumer with few access rights would see people in a video as points or avatars (like video game characters) moving in the street. Yet a consumer with enough access rights can see the actual person walking in a given street. Since there are no cameras everywhere and videos do not provide enough information for deciding the degree of “social cohesion” among people, in our approach the footprint consists of the digital data related to an individual available on Internet (on social networks, files, seen in cameras, or explicitly provided by people). Discovering communities within crowds leads to social graphs. Computing and processing these graphs requires computing and memory resources and therefore our HPC infrastructure for performing tests.

The remainder of the paper is organized as follows. Section II synthesizes related work concerning the data mining, social graphs exploitation techniques and computational science methods that address the generation and exploitation of social graphs. Section III introduces our approach for collecting footprints data for computing social graphs. Section IV describes the experiments we designed to qualitatively and quantitatively asses our proposal. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORK

We are interested in techniques that use crowdsourcing (explicit and implicit) for collecting data that contain information about the way people evolve in public and private places. These data collections can be used as input for learning

crowd behaviour and simulating it in a more accurate and realistic manner. The advance of location-acquisition technologies like GPS and Wi-Fi has enabled people to record their location history with a sequence of time-stamped locations, called *trajectories*. Some work has been carried out using cellular networks for user tracking, profiting from call delivery that uses transitions between wireless cells as input to a Markov model [3]. Wolf and others [12] used stopping time to mark the starting and ending points of trips. The comMotion system [10] used loss of GPS signals to detect buildings. When the GPS signal was lost and then later re-acquired within a certain radius, comMotion considered this to be indicative of a building. This approach avoided false detection of buildings when passing through urban canyons or suffering from hardware issues such as battery loss. [12] introduces a social networking service, called GeoLife, which aims to understand trajectories, locations and users, and mine the correlation between users and locations in terms of user-generated GPS trajectories. GeoLife offers three key applications scenarios: (1) sharing life experiences based on GPS trajectories; (2) generic travel recommendations, e.g., the top interesting locations, travel sequences among locations and travel experts in a given region; and (3) personalized friend and location recommendation.

Existing work in robotics and autonomous vehicles for example applies automatic learning techniques for making them autonomous while they evolve in open spaces. They often use collected data for example, for training classifiers, to reproduce the behaviour of the crowd in synthetic environment. One of the challenges in data analytics is to do prediction by deducing behaviour models of the observed subject. A model is a collection of data on some aspect of a subject's behaviour that, when associated with a limited set of contextual clues, yields predictions on what behaviour the subject will engage in next. Based on this notion, there is work like the one described in [1] that use location as context to infer other data such as the presence of other people. Predestination [8] is an approach that leverages an open-world modelling methodology that considers the likelihood of users visiting previously unobserved locations based on trends in the data and on the background properties of locations. Multiple components of the analysis are fused via Bayesian inference to produce a probabilistic map of destinations. The proposed algorithm was trained and tested using a database of GPS driving data gathered from 169 different subjects who drove 7,335 different trips.

The challenge with respect to social density within crowds is that the prediction and understanding of the crowd through observed data must be combined with personal data to determine whether individuals can be related among each other and form a community. Once communities are identified within crowds, it can be possible to determine the influence that such communities have among each other depending on the situation. For example, whether a community puts pressure on other groups or individuals in the crowd, or whether they create fear and instability. Social networks analytics, data mining, network science, and more recently social digital science research have addressed this issue and proposed different methodologies. Viral marketing, for instance is a vast domain

willing to understand the degree of influence of individuals within social to study how this influence is spread across these networks. These methods work on existing social networks making abstractions on the process in which social connections are established. Other academic and industrial works (e.g. social network providers) focus on the discovery of social connections for individuals according to different criteria. The easiest strategy for example is to connect people that share some social connection with others if they can have some social link ("I am related with the friends of my friends"). Works like [6] apply data mining and network science techniques [4], or bioinspired methods to discover some kind or "social" connection among people. The possibility of applying one technique or another depends on the amount of information available from the people that can have potential social links. The more information about their profile, the easier and more accurate are the discovered social links. In some scenarios, the information about the individuals that can potentially establish some social link is sparse or it is not integrated in one single collection. Information about individuals must be first harvested, correlated and filtered to generate useful data that can be then analysed. Beyond the volume, the variety and the degree of veracity of data, the task is difficult because data is guided by different access control, privacy level, different legal constraints, and quality. Besides, discovering correlations among such data collections can be challenging. Works concerning correlations [2] have addressed a simplified version of the problem where raw data stem from the same context. Leveraging such an assumption is still an open issue, that we are willing to address.

III. DISCOVERING SOCIAL COHESION IN CROWDS

We apply data analytics techniques (temporal and spatial reasoning) for computing trajectories and for identifying crowds, that is people grouped in a sufficient close spatial region that adopt a specific "behaviour" referring to four well known naïve crowd patterns: (i) casual crowd which is loosely organized and emerges spontaneously (people forming it have very little interaction at first and usually are not familiar with each other); (ii) conventional crowd resulting from more deliberate planning with norms that are defined and acted upon according to the situation; (iii) expressive crowd forms around an event that has an emotional appeal; and (iv) acting crowd where members are actively and enthusiastically involved in doing something that is directly related to their goal.

The objective of our analytics study is to identify possible social connections among people involved in a crowd (cf. fig. 1). For example, people sharing similar political and social ideals, people living geographically close, with the similar occupation, leisure interests, job, and with common contacts. In our approach and since for the time being we do not use images recognition we address crowd profile by measuring people social cohesion when they are in specific spatial regions during a time interval. This requires a continuous analysis of the evolution of the status of the areas of an urban space.

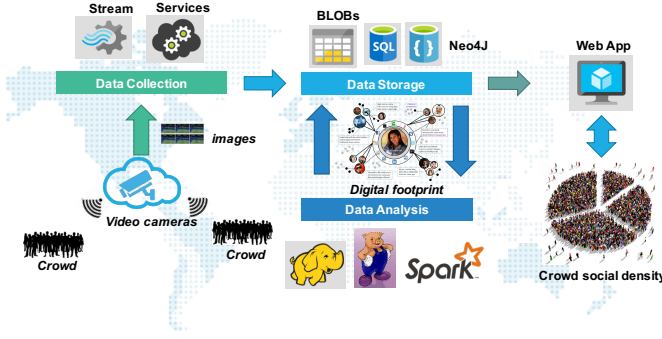


Figure 1. General approach for discovering social cohesion in crowds

A. Harvesting data from social networks

Data harvesting is done using different data collections: (i) the continuously harvested observations of the geographical position of individuals (that accept sharing their position) along time; (ii) the images stemming from cameras observing specific “critical” urban areas, like terminals, airports, public places and government offices; (iii) data produced by social networks and applications like Twitter, Facebook, Waze and similar.

Given the geographical location of the potential crowd, that we name *target location*, we access social networks Twitter, Foursquare, Swarm to “track people” located close to this *target location*. For collecting data, we assume that data providers are services implemented by a REST architecture or a SOAP API. These services are public and they can have specific SLA constraints, like the number of requests per hour (e.g., Twitter) or some authentication ones (e.g., Facebook). Other providers like pages or sites do not have explicit constraints but we assume that they are governed by privacy and authorship rights determined by their country of origin.

- 1) *On demand* data providers must be queried through a specific interface or crawled to harvest data. The frequency in which data is collected is specified in the program interacting with the provider. The consumer invokes the batch method through the network with its input data. The speed of the network introduces the transfer time cost that is determined by the data size and the network's conditions (i.e., latency and throughput). Depending on the type of network, it can have a monetary price (e.g., 3G) also determined by data size. Once the method invocation arrives to the hosting device, the service provider receives the request and associates a predefined method invocation price. Afterwards, the method instance processes the request during an execution time (i.e., method response time) which is determined by the method throughput given by the number of processed requests in an interval of time (e.g., each minute) and the state of the device such as memory or CPU usage. The request implies the usage of the network interface, service provider and method execution. Those processes spend the battery of the device (i.e., battery consumption) entailing a battery cost. Finally, the output (i.e., method response) is sent back to consumer through the same network and, as input data aforesaid, output data contributes to data transfer time

and to monetary cost. Both, input and output data define the data size measure.

- 2) *Stream* providers work under a subscription strategy. A *continuous data provider* exports a method *subscribe()* used by a consumer to start receiving streams at some rate and for a given period of time (e.g., by executing an *unsubscribe()* method, for a predefined period of time, until something happens). The general process implemented to interact with this type of providers is done as follows: the consumer invokes the continuous method through the network with its input data. Then, the method instance starts processing results and it sends the results every time interval (the so-called production rate). The production rate can be determined by consumer needs. For instance, “give my current position every five minutes” where ‘five minutes’ is the expected production rate. Produced data is then sent to the consumer who processes it immediately or after a threshold defined by the number of tuples received, or the elapsed time, or a buffer capacity. This threshold is named processing rate. Both production rate and processing rate impact the execution time cost, execution price cost, and battery consumption cost.

We assumed that providers are autonomous in the sense that they can modify their interfaces, authentication protocols and privacy and authorship rules whenever they want and our data collection services must deal with these changes. We do not have information about the production rate of the streams and changes in Web pages and sites. In a first approach, we tuned the collection manually but we also collected information about services behaviour to automatize the tuning process and ensure the collection of fresh unique data. We used JSON as data model and we then implemented document processing tasks to characterize the content of collected data.

B. Measuring crowd density

The occupation density of specific urban regions is measured separately according to the political organization of the space (quarters, areas) in every database. We use sliding windows to partition continuous data flows with respect to time intervals and we use the political division of the urban space for filtering, grouping the data and computing density per urban region. This is a straightforward yet somehow costly computation not because of the amount of data but because it should be continuously computed, and both data and density results are stored for performing other analytics processes.

Having different visions of the density of urban spaces given different data collections enables to perform other types of data analytics on the n-tuples region-density and to cluster regions both taking into consideration their density and their geographic position. Accordingly, we generate a “crowd social cohesion model” showing a view of the dynamic distribution of communities within the crowds that occupy the urban space and how they evolve along time.

C. Measuring social density

Not all crowds need to be managed within urban spaces, there are some that happen every day in public transport and others need attention and must be better profiled. The first challenge is to be able to discriminate. Using data collections

harvested from social networks, some correlations are computed to identify connected people that were in the same urban place at some time interval, and that might have been part of a crowd event. Not all individuals sharing the same spatial region necessarily participate in a crowd. Thus, the first operation to solve is given a set of individuals located in the same geographic region at the same time interval, whether an individual located at the same space-time belongs to the crowd.

A naïve way of evaluating the predicate *belongs-to*, is determining whether there is a “social” connection between an individual and at least one of the potential crowd members. Using this analytics operation, it is possible to draw an urban occupation map and propose some connections among individuals occupying the same urban region at the same time interval. The objective of this classification is to identify possible outsiders (i.e., individuals that do not really belong in the group). Those are the ones in which we are particularly interested because our guess is that within those outsiders we might find human traffic dealers, for example.

We adopt two complementary strategies for computing what we call social density of crowd (see fig. 2). First, we compute a *digital footprint of individuals* retrieving the social graphs of individuals from their social networks accounts. These individuals are located in a geographical location where a possible crowd has been observed by exploring videos from surveillance cameras. The analysis of videos can be done manually or automatically applying tracking methods [7]. This process is out of the scope of this paper. We assume that when we start the process for computing social density, the potential crowd has been already identified and located.

As shown in fig. 2, we compute each individual *digital footprint graph* by correlating and integrating data from the social graphs from their different social networks. Filtering strong related contacts, we verify whether they are themselves present in the crowd. Given a set of *digital footprints graphs* we discover possible relations with other individuals using two relations discovery techniques.

- 1) The one proposed by [5] for computing *social density*. It starts from a portion of the *digital footprints graphs* consisting of all individuals present in the target location. This process “completes” the graph with new discovered relations tagged with precision probabilities.
- 2) We use the technique described in [9] for identifying communities within the *digital footprints graphs* tagged with new relations. This results in a *social density graph* where nodes represent groups of individuals evolving in the same geographical situation for a given time interval, which are related among each and that can be grouped according to different criteria (e.g., those with similar job, interests, sharing contacts, living in the same region, working for the same company, married, single).

The process is computationally expensive because we are looking for graphs intersections of possibly thousands of individuals. Indexing and graph processing heuristics are used to reduce the size of graphs and graph collections that must be explored.

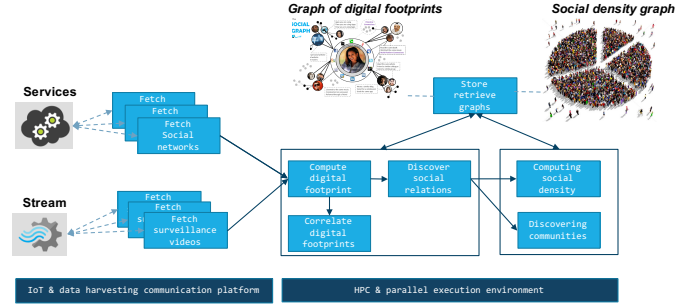


Figure 2. Data & graph processing workflow

D. Profiling and predicting the crowd behaviour

Once the crowd has been modelled in terms of its participants, it is possible to observe its behaviour as if it was an individual: (i) characterize its elements (find the leaders, the followers and eventually the outsiders); and (ii) predict the evolution of its behaviour for example, probability of conflict, space utilisation and risk maps.

Once we have created a first connection graph describing how people possibly participating in the crowd are connected among each other, we apply methods to determine which is the degree of influence of each participant of the crowd. This is done by computing the influence of the elements of the crowd towards other elements according to their contacts network and that can eventually be participants in the crowd. The analysis ends up with groups of outsiders that must be inspected to understand their presence in the crowd, and their possible role in the event. Again, since the inclusion of participants in the crowds evolves along time, the contacts’ network represented by graphs and the influence of participants varies too. The computational cost is important considering semi-post-mortem computations. Of course, the ultimate objective is to be able to observe this evolution in real time which introduces scalability problems that must be addressed with GPU architectures.

The final task is to predict the behaviour of the crowd. Therefore, we use the notions of space occupation and probability of conflict, by searching behaviour patterns in the evolution of the crowd status: it emerges with some individuals, increases its size, in achieves the maximum of participants and then it fades. This “life cycle” happens within a space occupation process that can be controlled by its inherent behaviour but that can also be determined by external factors e.g., police, troublemakers. We define the status of the crowds with a set of attributes including, the approximate number of participants, the main trajectories of the group, the spatio-temporal region occupied by the group, the possible outgoing directions in which participants can move within the urban space, triggering and termination event.

Instead of delivering textual or graphical results of these analytics operations we provide 2D and 3D visualizations that can reproduce the observations and simulate the behaviour of the crowd according to real data. See our current work in [6] for details about visualization issues. The following section focusses on the tasks related to data processing and link discovery.

IV. EXPERIMENTATION

We exploit data collections about the way people move in public places for learning about transit and crowd behaviour. In a first experiment, the characteristics of the characters are either inferred from the movement, or explicitly provided by people. We use the *GeoLife GPS trajectory dataset* [12] with data of 182 users, 17,621 trajectories of *ca.* 1.2 million Km. and 48,000+ hours. These data are used to compute spatio-temporal people flows in real crowds.

A. Computing trajectories

Geolife GPS trajectory dataset is used for computing the trajectories of the people observations it contains. We used Big Data cleaning and processing tools particularly the language PigLatin for computing such trajectories. The process is done by defining four declarative expressions as shown in fig. 3:

- Q₁: Load the observations clean and prepare the data collection with respect to the initial meta-data of the observations. For computing trajectories three attributes were necessary, the initial and termination times and the transportation modes. The idea is that a trajectory is defined as a set of locations observed at a given moment identified by a time stamp. The PigLatin program presented in the figure implements this process.
- Q₂: Load the GPS logs filtering the latitude, longitude and time stamp according to a predefined schema defined for this purpose (see fig. 3).
- Q₃: Finally, the third query estimates the trajectories as sequences of locations where the sequence is determined by the time stamps. With these computed trajectories, it is possible to perform other analytics operations, and reconstitute the movement of people in the corresponding urban space, in this case Beijing.

Even if they seem simple due to the expression power of PigLatin, the execution of these queries can be computationally costly given the volume of data we processed. The PigLatin execution environment was installed in a cluster of 8 machines and executed in parallel. We thus obtained results in reasonable execution time. The computed trajectories were

```

-- Step 1: Load Labels
Labels = LOAD '$INPUT_PATH/Labels.txt' using PigStorage('\t') as (
  start_time: datetime,
  end_time: datetime,
  transportation_mode: chararray);

-- Step 2: Load GPS logs
GPS_logs = LOAD '$INPUT_PATH/Trajectory' using PigStorage(',') as (
  latitude: float,
  longitude: float,
  altitude: int,
  timestamp: datetime);

-- Step 3: Find trajectories
CR = CROSS Labels, GPS_logs;

-- Remove points not belonging to a trajectory
CR = FILTER @ BY
  ToMilliSeconds(GPS_logs::timestamp) >= ToMilliSeconds(Labels::start_time) AND
  ToMilliSeconds(GPS_logs::timestamp) <= ToMilliSeconds(Labels::end_time);

-- Group points belonging to the same trajectory
Trajectories = GROUP @ BY (Labels::start_time, Labels::end_time, Labels::transportation_mode);

```

Figure 3. Computing trajectories

stored in an integrated data collection on top of which we performed other operations as described in the following lines.

B. 2D individuals tracking

Since our objective is to track individuals and crowds in urban spaces. We computed heat maps to aggregate the trajectories of users during specific time intervals and we used a map service to visualize them. The result, as illustrated in fig. **Erreur ! Nous n'avons pas trouvé la source du renvoi.**, shows those itineraries that are popular in red or thick lines. These heat maps concern the Beijing trajectories computed using the *Geolife GPS trajectory dataset*.

Heat maps enable tracking individuals within a specific urban space and see how they move during a specific time interval. The figure shows the simple code for computing heat maps that again serve as new data on which it is possible to perform more analytics. These analytics concern for example identifying the most visited regions, observe rush hours in certain regions, and eventually identify crowds. For the time being our experimentation is done post-mortem.

C. Computing social cohesion and communities in crowds

Fig. 5 shows an intuitive view of the data structures that we process to compute social cohesion and discover communities in crowds. Conceptually it consists of a complex two level graph called *digital footprint graph*.

In the first level a node represents an individual's profile integrating personal data integrated from all her social network accounts. Each data entry is qualified with information about its provenance and privacy degree obtained from the privacy contract associated to the associated account. The attributes associated to the node is given by the union of the attributes of the ones describing the personal data of a person in each social network that is integrated. For every attribute the node uses the following structure:

```

<attribute, type,
  provenance: string, %social network
  privacy level: {public, friends, friends-
    of-friends}>

```

A node maintains a set of special connections with the identity nodes of the social graph of the user stemming from every social network. Finally, a node is connected to other nodes with a relationship tagged with a value depending

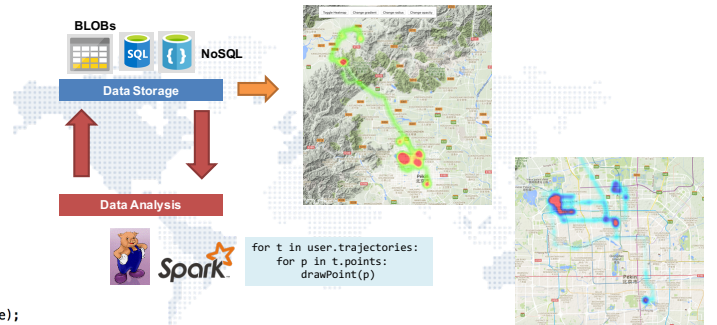


Figure 4. Heat maps aggregating the tra-

whether both nodes are connected in one or several social networks. In our experiments the integration is a process with few technical challenge. We start from an initial set of social graphs from 182 accounts that accepted to participate in the experiment. We used the API's provided by three social networks: Facebook, Twitter, Foursquare, and we accessed to the Whatsapp agenda of the users for building the first level of the footprint graph.

With the footprint graph we apply two algorithms, one for discovering possible connections. Some of these connections can be computed directly by querying the footprint graph. Particularly, retrieving the mutual connections which is already computed by some social networks API. Using this list of mutual connections, it is possible to compute mutual multi-social network connections. Other links are more difficult to infer and require applying specific strategies. In our current experiments, we do not apply them.

The second type of analytics that we apply is that of identifying communities within the digital footprint graph (see leftmost side of fig. 5). To simplify but still have a complete version of the approach we applied a k-means strategy that enabled us to identify communities with similar (i) social behaviour (e.g., visit and like the same type of places on Foursquare, visit similar places using geographically tagged posts on Twitter and Facebook); (ii) similar online behaviour (e.g., the average time spent connected to a given social network per day, the number of posts published in a week).

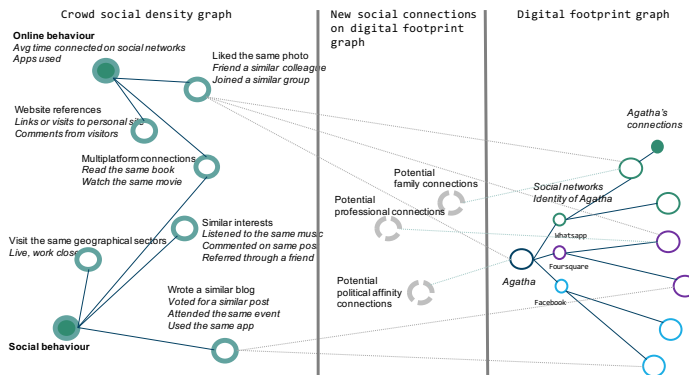


Figure 5. Crowd social density based on individuals digital footprint

V. CONCLUSIONS AND FUTURE WORK

This paper introduced our approach and first experiment results concerning the correlation of digital footprints to discover social connections of individuals potentially participating in crowds. The approach applies existing data mining and computational science techniques for processing data, but contributes on the management of series of graphs that are produced and exploited because of the different data processing tasks.

Given the volume of data to be processed and the complexity of the computed graphs, it is not judicious to proceed using the brutal force for processing them, so in every step indexes are associated to computed graphs and the input collections of graphs are pruned to reduce the execution time despite the fact the we rely on parallel settings for executing

them. For the time being the graphs we process only contain textual information, yet in our context information about the crowds can come in videos and images. We want to include this type of data in our analytics processes and correlated video and textual information to build our digital footprint graphs.

ACKNOWLEDGMENT

This work is part of the project CONCERNs granted by Microsoft which fosters collaborations between the Barcelona Super Computing Centre (BSC), the French Council of Scientific Research (CNRS) and the Fundación Universidad de las Américas Puebla (UDLAP), and it has been partially funded by the CNRS UMI 3175 LAFMIA and the CONACyT of the Mexican government.

REFERENCES

- [1] Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*. 7, (2003).
- [2] Aszalós, László, and Tamás Mihálydeák. "Correlation Clustering by Contraction, a More Effective Method." *Recent Advances in Computational Optimization*. Springer International Publishing, 2016. 81-95.
- [3] Bhattacharya, A., Das, S.K.: LeZi-Update: An Information-Theoretic Framework for Personal Mobility Tracking in PCS Networks. *Wireless Networks*. 8, (2002).
- [4] Barabási, Albert-László. *Network science*. Cambridge University Press, 2016
- [5] Garcia-Gasulla, D., Alvarez-Napagao, S., Tejeda-Gómez, A., Oliva-Felipe, L., Gómez-Sebastià, I., Bejar, J., Vázquez-Salceda, J.: Social network data analysis for event detection. In: *Proc. of the 21st European Conference on Artificial Intelligence (ECAI'14)*. , Prague, Czech Republic (2014).
- [6] Howei, Mohammad, and David Nguyen. *VIRAL MARKETING*. Diss. 2016.
- [7] Hu, Ninghang, Henri Bouma, and Marcel Worring. "Tracking individuals in surveillance video of a high-density crowd." *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012.
- [8] Krumm, J., Horvitz, E.: Predestination: Inferring Destinations from Partial Trajectories. In: *Proc. of the 8th Int. Conference on Ubiquitous Computing (UBICOMP'06)*. , Orange County, CA, USA (2006).
- [9] Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2014.
- [10] Marmasse, N., Schmandt, C.: Location-Aware Information Delivery with ComMotion. In: *Proc. of the HUC 2nd Int. Symposium of Handheld and Ubiquitous Computing*. , Bristol, UK (2000).
- [11] Genoveva Vargas-Solar, Javier A. Espinosa-Oviedo, Hugo Perez, Isaac Rudomin, José Luis Zechinelli-Martini. Modelling crowds in urban spaces: from Big data to smart secure regions. In *Proceedings of the 7th International Supercomputing Conference in Mexico (ISUM'106)*. 11-13 de Abril, Puebla, Mexico (in press).
- [12] Wolf, J., Guensler, R., Bachman, W.: Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data. In: *Notes from Transportation Research Board, 80th annual meeting*. , Washington, D.C, USA (2001).
- [13] Zheng, Y., Xie, X., Ma, W.-Y.: GeoLife: A Collaborative Social Networking Service among User, location and trajectory. *IEEE Data Engineering Bulletin*. 33, (2010).