



# **The SIGSPATIAL Special**

**Newsletter of the Association for Computing Machinery  
Special Interest Group on Spatial Information**

---

**Volume 11      Number 3      November 2019**

# The SIGSPATIAL Special

The SIGSPATIAL Special is the newsletter of the Association for Computing Machinery (ACM) Special Interest Group on Spatial Information (SIGSPATIAL).

ACM SIGSPATIAL addresses issues related to the acquisition, management, and processing of spatially-related information with a focus on algorithmic, geometric, and visual considerations. The scope includes, but is not limited to, geographic information systems.

## **Current Elected ACM SIGSPATIAL officers are:**

- Chair, Cyrus Shahabi, University of Southern California
- Past Chair, Mohamed Mokbel, University of Minnesota
- Vice-Chair, Goce Trajcevski, Iowa State University
- Secretary, Egemen Tanin, University of Melbourne
- Treasurer, John Krumm, Microsoft Research

## **Current Appointed ACM SIGSPATIAL officers are:**

- Newsletter Editor, Andreas Züfle, George Mason University
- Webmaster, Chrysovalantis (Chrys) Anastasiou, University of Southern California

For more details and membership information for ACM SIGSPATIAL as well as for accessing the newsletters please visit <http://www.sigspatial.org>.

The SIGSPATIAL Special serves the community by publishing short contributions such as SIGSPATIAL conferences' highlights, calls and announcements for conferences and journals that are of interest to the community, as well as short technical notes on current topics. The newsletter has three issues every year, i.e., March, July, and November. For more detailed information regarding the newsletter or suggestions please contact the editor via email at [azufle@gmu.edu](mailto:azufle@gmu.edu).

Notice to contributing authors to The SIGSPATIAL Special: By submitting your article for distribution in this publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights:

- to publish in print on condition of acceptance by the editor,
- to digitize and post your article in the electronic version of this publication,
- to include the article in the ACM Digital Library,
- to allow users to copy and distribute the article for noncommercial, educational or research purposes.

However, as a contributing author, you retain copyright to your article and ACM will make every effort to refer requests for commercial use directly to you.

Notice to the readers: Opinions expressed in articles and letters are those of the author(s) and do not necessarily express the opinions of the ACM, SIGSPATIAL or the newsletter.

**The SIGSPATIAL Special (ISSN 1946-7729) Volume 11, Number 3, November 2019.**

# Table of Contents

	Page
<b>Message from the Editor</b> ..... <i>Andreas Züfle</i>	1
<b><u>Section 1: ACM SIGSPATIAL 2019 Students Research Competition: Grand Finalists</u></b>	
<b>Flash: Scalable Spatial Probabilistic Graphical Modeling</b> ..... <i>Ibrahim Sabek</i>	3
<b>Discovering Human Activity Community in a City</b> ..... <i>Tianyuan Huang</i>	7
<b><u>Section 2: SIGPSPATIAL 2019 Workshop Reports</u></b>	
<b>1<sup>st</sup> ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)</b> ..... <i>Marc van Kreveld, Bettina Speckmann, Matei Stroila, Goce Trajcevski</i>	9
<b>1<sup>st</sup> ACM SIGSPATIAL International Workshop on Geo-computational Thinking in Education (GeoEd 2019)</b> ..... <i>Amr Magdy, Coline Dony</i>	12
<b>1<sup>st</sup> ACM SIGSPATIAL International Workshop on Geospatial Data Access and Processing APIs (SpatialAPI 2019)</b> ..... <i>Ahmed Eldawy</i>	14
<b>1<sup>st</sup> ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)</b> ..... <i>John Krumm, Cyrus Shahabi, Andreas Züfle</i>	15
<b>2<sup>nd</sup> ACM SIGSPATIAL International Workshop on Advances on Resilient and Intelligent Cities (ARIC 2019)</b> ..... <i>Bandana Kar, Olufemi A. Omitaomu, Xinyue Ye, Shima Mohebbi, Guangtao Fu</i>	18
<b>2<sup>nd</sup> ACM SIGSPATIAL International Workshop on GeoSpatial Simulation (GeoSim 2019)</b> ..... <i>Joon-Seok Kim, Hamdi Kavak, Dieter Pfoser, Andrew Crooks, Carola Wenk, Sarah Wise</i>	20
<b>3<sup>rd</sup> ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI 2019)</b> ..... <i>Song Gao, Shawn Newsam, Liang Zhao, Dalton Lunga, Yingjie Hu, Bruno Martins, Xun Zhou, Feng Chen</i>	23

<b>3<sup>rd</sup> ACM SIGSPATIAL International Workshop on Analytics for Local Events and News (LENS 2019)</b> .....	25
<i>Haiquan Chen, Federico Iuricich, Amr Magdy</i>	
<b>3<sup>rd</sup> ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities 2019)</b> .....	27
<i>Bruno Martins, Ludovic Moncla, Patricia Murrieta-Flores</i>	
<b>3<sup>rd</sup> ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec 2019)</b> .....	30
<i>Panagiotis Bouros, Tamraparni Dasu, Yaron Kanza, Matthias Renz, Dimitris Sacharidis</i>	
<b>3<sup>rd</sup> ACM SIGSPATIAL International Workshop on Prediction of Human Mobility (PredictGIS 2019)</b> .....	34
<i>Satish V. Ukkusuri, Kaoru Sezaki, Takahiro Yabe, Kota Tsubouchi</i>	
<b>8<sup>th</sup> ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (BigSpatial 2019)</b> .....	38
<i>Ashwin Shashidharan, Varun Chandola, Ranga Raju Vatsavai</i>	

### **Section 3: Event Announcements**

<b>Call for Submissions: The Sixth International ACM SIGMOD Workshop on Enriched Geo-Spatial Data (GeoRich 2020)</b> .....	40
<i>Panagiotis Bouros, Amr Magdy, Matthias Renz, Mohamed Sarwat, Andreas Züfle</i>	

# Message from the Editor

Andreas Züfle  
Department of Geography and GeoInformation Science,  
George Mason University  
Email: azufle@gmu.edu

Dear SIGSPATIAL Community,

The newsletter serves the community by publishing short contributions such as SIGSPATIAL conferences' highlights, calls and announcements for conferences and journals that are of interest to the community, as well as short technical notes on current topics.

In this November 2019 issue, the first section highlights the winners of our 4th ACM SIGSPATIAL Student Research Competition (SRC) held at the ACM SIGSPATIAL Conference 2019. Since the SRC papers are included in the proceedings, I have broken the tradition of including *all* SRC papers in the Newsletter. Instead, only the two winners of the SIGSPATIAL SRC (undergraduate and graduate), which will represent our community in the ACM Student Research Competition Grand Finals, are included. The winners are Tianyuan Huang, South China University of Technology (undergraduate) and Ibrahim Sabek, University of Minnesota (graduate), whose SRC papers are found in the first section.

**Good luck Tianyuan Huang and Ibrahim Sabek in the 2020 ACM SRC Grand Finals!**

Furthermore, please join me in congratulating Fandel Lin, National Cheng Kung University, for winning the *second place* in the 2019 SRC Grand Finals in the undergraduate category for his 2018 SIGSPATIAL SRC paper "*An Intelligent and Interactive Route Planning Maker for Deploying New Transportation Services*". More details on the 2019 SRC Grand Finals can be found at <https://src.acm.org/>. Special thanks to the chairs of the student research competition 2019: Theodoros Chondrogiannis and Kyriakos Mouratidis.

The second section consists of 12 workshop reports of workshops that were co-located with SIGSPATIAL 2019 and held on November 5th 2019:

- 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)
- 1st ACM SIGSPATIAL International Workshop on Geo-computational Thinking in Education (GeoEd 2019)
- 1st ACM SIGSPATIAL International Workshop on Geospatial Data Access and Processing APIs (SpatialAPI 2019)
- 1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)
- 2nd ACM SIGSPATIAL International Workshop on Advances on Resilient and Intelligent Cities (ARIC 2019)
- 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation (GeoSim 2019)
- 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI 2019)
- 3rd ACM SIGSPATIAL International Workshop on Analytics for Local Events and News (LENS 2019)

- 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities (GeoHumanities 2019)
- 3rd ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec 2019)
- 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility (PredictGIS 2019)
- 8th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (BigSpatial 2019)

I would like to sincerely thank all workshop organizers for their generous contributions of time and effort. These reports are paramount, as all SIGSPATIAL workshops happen in parallel and thus, it is impossible to concurrently attend all the interesting keynotes, paper presentations and other exciting events.

The third section consists of one workshop announcement for the Sixth International ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data, which is highly relevant to our community.

I hope that you will find the newsletters interesting and informative and that you will enjoy this issue. You can download all Special issues from:

<http://www.sigspatial.org/sigspatial-special>

Yours sincerely,

Andreas Züfle

SIGSPATIAL Newsletter Editor



# **The SIGSPATIAL Special**

## **Section 1: The 4<sup>th</sup> ACM SIGSPATIAL Student Research Competition (SRC) Grand Finalists**

---

**ACM SIGSPATIAL**  
**<http://www.sigspatial.org>**

# Flash: Scalable Spatial Probabilistic Graphical Modeling

Ibrahim Sabek

Department of Computer Science and Engineering, University of Minnesota, USA

Email: sabek001@umn.edu

## ABSTRACT

The current explosion in spatial data raises the need for efficient spatial analysis tools to extract useful information from such data. Spatial probabilistic graphical modeling (SPGM) is an important class of spatial data analysis that provides efficient probabilistic graphical models for spatial data. Unfortunately, existing SPGM tools are neither generic nor scalable when dealing with big spatial data. In this work, we present *Flash*; a framework for *generic* and *scalable* spatial probabilistic graphical modeling (SPGM). *Flash* exploits Markov Logic Networks (MLN) to express SPGM as a set of declarative logical rules. In addition, it provides spatial variations of the scalable RDBMS-based learning and inference techniques of MLN to efficiently perform SPGM predictions. We have evaluated *Flash*, based on three real spatial analysis applications, and achieved at least two orders of magnitude speed up in learning the modeling parameters over state-of-the-art computational methods.

## KEYWORDS

Spatial Probabilistic Graphical Models, Spatial Analysis, Markov Logic Networks, Scalability

## 1 INTRODUCTION

There is a plethora of spatial data being generated at the moment. For example, space telescopes generate up to 150 gigabytes weekly spatial data, medical devices produce spatial images (X-rays) at a rate of 50 petabytes per year, and a NASA archive of satellite earth images has more than 500 terabytes. This raises the need for efficient spatial analysis solutions to extract insights and useful patterns from such data. *Spatial probabilistic graphical modeling* (SPGM) represents an essential class of spatial analysis techniques, which exploits probability distributions and graphical representations (e.g., spatial hidden Markov models [12]) to describe spatial phenomena and make predictions about them [28]. SPGM has revolutionized many scientific and engineering fields in the past two decades including health care, risk analysis, and environmental science (e.g., [3, 12]). However, existing SPGM techniques have a scalability issue. In particular, they were originally designed for running on a single machine and hence suffer from the limited computation resources (e.g., see [7, 14, 30]). Such techniques can not scale beyond implementing prototypes over small spatial datasets.

Meanwhile, Markov Logic Networks (MLN) [22] was introduced to efficiently build complex learning and inference models over big data in a declarative manner. Basically, MLN combines first-order logic rules with probabilistic graphical models to represent statistical learning and inference problems with few logical rules (e.g., rules with imply and bit-wise AND predicates) instead of thousands of lines of code. With MLN, data scientists and developers can focus

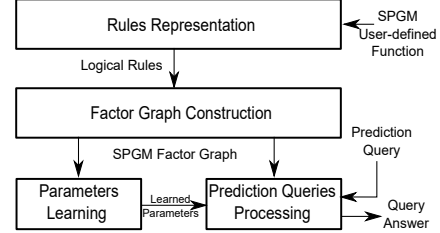


Figure 1: *Flash* System Architecture.

their efforts only on developing the rules that represent their applications (e.g., knowledge base construction, data cleaning, genetic analysis). Although the recent advances in MLN frameworks [21] helped to scale up the performance of typical spatial analysis applications (e.g., spatial regression [25, 27], and spatial-aware knowledge base construction [24, 26]), MLN was never exploited to scale up the performance of SPGM techniques.

In this paper, we propose *Flash*; a framework for scalable spatial probabilistic graphical modeling (SPGM) using Markov Logic Networks (MLN). *Flash* has the following three main features: (1) *Declarativity*: *Flash* expresses any SPGM application with logical semantics, and allows developers to implement it using a set of logical rules. (2) *Efficiency*: *Flash* translates the equivalent MLN rules of any SPGM application into SQL queries using an efficient grounding technique [29], and then executes these queries inside scalable database engines. In addition, *Flash* provides spatial variations of the RDBMS-based learning and inference algorithms of MLN [21] to perform scalable SPGM predictions (e.g., predictions over models with millions of nodes). (3) *Abstraction*: *Flash* allows developers to build a myriad of spatial analysis applications as a set of user-defined functions (UDF) without the need to worry about the underlying SPGM computation. As a case study, we equipped *Flash* with the implementation of three fundamental SPGMs; spatial Markov random fields (SMRF) [6], spatial hidden Markov models (SHMM) [12], and spatial Bayesian networks (SBN) [7]. The following sections explain the architecture of *Flash*, the implementation details of these three supported SPGMs, and the preliminary evaluation results.

## 2 FRAMEWORK OVERVIEW

*Flash* adopts a modular system architecture as shown in Figure 1. It consists of four main modules, described briefly as follows:

**Rules Representation.** This module is responsible for generating an equivalent representation of logical MLN rules to any user-defined SPGM input. These rules have two main properties: (1) they contain first-order logical predicates (e.g., bitwise-AND, and imply) that capture the SPGM semantics; (2) they are associated with *weights* that represent the original SPGM parameters (Examples

<sup>1</sup>This work is partially supported by the National Science Foundation, USA, under Grants IIS-1907855, IIS-1525953 and CNS-1512877.



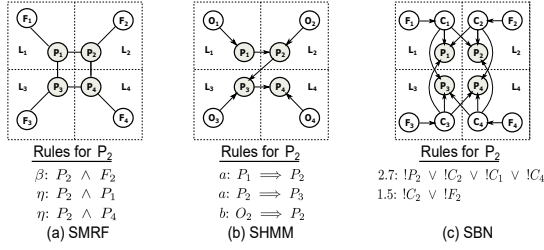


Figure 2: SMRF, SHMM, and SBN Representations in *Flash*.

are in Section 3). The generated rules follow the syntax of a DBMS-friendly Datalog-like language, called DDlog [29], which can be efficiently processed with any relational DBMS (e.g., PostgreSQL) during the factor graph construction module.

**Factor Graph Construction.** This module takes the generated rules as input and uses them to build a *factor graph* [33] in a scalable way. The factor graph is the main data structure used to represent any MLN model, where the weights of the graph nodes correspond to the weights of rules (i.e., SPGM model parameters). To efficiently populate this factor graph, *Flash* adapts a scalable grounding technique from [29] that translates the generated rules into SQL queries, and then applies such queries on the input application data to obtain the final factor graph that is equivalent to the SPGM input.

**Parameters Learning.** This module learns the unknown weights of the constructed factor graph (i.e., weights of rules), which in turn specify the final SPGM parameters (e.g., spatial hidden Markov model [12] parameters). *Flash* proposes a pseudo-likelihood learning algorithm that adapts an efficient variation of a sampling-based gradient descent optimization technique to compute the gradient of the SPGM pseudo-likelihood and then determine the weights.

**Prediction Queries Processing.** This module is responsible for answering prediction queries over the SPGM model (e.g., what is the probability of a specific event to happen?). Basically, it takes the prediction query along with the factor graph and its learned weights as inputs, and produces a prediction output associated with its confidence probability. Prediction queries can be answered using traditional Gibbs sampling-based inference algorithms over factor graphs [21]. However, such algorithms perform sequential sampling over the factor graph nodes which results in slow convergence to the inference answer in case these nodes have spatial dependencies as in SPGM applications [18]. Instead, *Flash* employs a variation of Gibbs Sampling that exploits a *concliques*-based traversal pattern [18] to efficiently sample spatially-dependent nodes in parallel while guaranteeing the rapid convergence.

### 3 CASE STUDIES IN FLASH

*Flash* supports the implementation of three common spatial graphical models; spatial Markov random fields (SMRF) [6], spatial hidden Markov models (SHMM) [12], and spatial Bayesian networks (SBN) [7], as case studies. Figure 2 gives toy examples on the logical representation of these three models in *Flash*, where each model is defined over 4-cells grid, and the neighborhood of any cell  $l$  is assumed to be the cells that share edges with  $l$  only.

**Spatial Markov Random Field (SMRF).** SMRFs are powerful and important tools for modeling spatial data and building analysis

applications. They have been widely used in different areas of spatial statistics [13, 23, 34]. As with many other areas of statistics, a major challenge for spatial analysts is dealing with massive data sets. This is particularly problematic for SMRFs due to the need for matrix operations that involve very large matrices can be computationally prohibitive, specially in the case of Gaussian processes. Existing approaches tried to solve the scalability issues in two categories. The first category focused on developing fast matrix computations that exploit the sparsity of matrices in SMRF models (e.g., [10]). However, utilizing sparsity does not seem to be among the more promising strategies as it does not fit the dense data cases. The second category introduced fast likelihood approximations for the Gaussian-based SMRF models (e.g., [4]). However, this category is not generic enough to capture other arbitrary SMRF models (i.e., the interactions between random variables in the SMRF model are not captured with multivariate Gaussian distributions).

In contrast to existing approaches, *Flash* provides a scalable approach for SMRF models by introducing a first-order logic representation, where there is an equivalent weighted *bitwise-AND* predicate for each pair of connected variables. In this case, the predicates' weights correspond to the SMRF parameters that need to be learned. Figure 2(a) shows a small SMRF model with a prediction  $P_l$  and feature  $F_l$  at each cell  $l$ . Each prediction  $P_l$  has undirected edges with feature  $F_l$  at this cell and each neighboring prediction variable. For example, predication  $P_2$  is connected with feature  $F_2$  and neighboring predictions  $P_1$  and  $P_4$ .

**Spatial Hidden Markov Models (SHMM).** The hidden Markov model (HMM) is a doubly embedded stochastic method based on probability theory, which can be used in a sequence labeling problem. It describes the process of randomly generating non-observable state sequences from a hidden Markov chain and generating an observation from each state to produce an observable sequence. In spatial hidden Markov model (SHMM), the sequences are generated on spatially-correlated random variables.

While all existing SHMM solutions are innovative, they face severe scalability issues when dealing with big spatial data. The scalability challenge is mainly because these solutions were not originally designed for the big data era or to exploit new high performance computing environments [2]. In contrast, *Flash* scales up the performance of SHMM by providing an equivalent MLN representation, where any state/state or observation/state pair is mapped to a weighted *imply* predicate, and the resulting weights correspond to the SHMM parameters. Figure 2(b) shows a small SHMM model with a hidden state  $P_l$  and observation  $O_l$  variables at each cell  $l$ . Each observation  $O_l$  has a directed edge to state  $P_l$  at this cell. In addition, SHMM imposes an ordered spatial dependence among neighboring locations, where it uses z-curve ordering technique to build a sequence that preserves the spatial dependence between prediction variables (e.g.,  $P_1$  has a directed edge to  $P_2$ , and  $P_2$  has another one to  $P_3$ , etc).

**Spatial Bayesian Networks (SBN).** Numerous applications model the probability of an input event to occur based on other *causal* events that have spatial dependencies with the input event. These applications include meteorology [8], risk analysis [3, 17], and environmental science [11, 19]. For example, business analysts forecast the budget and likely costs of water infrastructure networks based on failure events in water mains at neighboring sites [17]. A

typical solution to model the *causal dependencies* between events in all these applications is to employ spatial Bayesian networks (a.k.a spatial Bayesian belief networks) [16, 20]. These networks are directed probabilistic graphs whose nodes represent variables corresponding to events over neighboring locations, and the edges represent the casual relationships between these variables. For example, two events "rain" and "flood" at neighboring locations  $x$  and  $y$ , respectively, can be represented as two random variables, where the *rain* variable is a cause (i.e., parent node in the graph) for the *flood* variable. Existing solutions of spatial Bayesian networks can not scale beyond implementing prototypes over small spatial and spatio-temporal datasets [7, 20]. Meanwhile, Markov Logic Networks (MLN) are recently used to scale up the performance of classical Bayesian networks that do not consider spatial dependencies between random variables (e.g., Bayesian Logic Networks [15]).

In *Flash*, we exploit Markov Logic Networks (MLN) to represent the SBN models. *Flash* provides an equivalent weighted combination of *bitwise-OR* and *negation* predicates for each causality relation (i.e., directed edge). The weights of these predicates are calculated from the input prior probabilities of SBN. Figure 2(c) shows a small SBN model with a prediction variable  $P_l$  at each cell  $l$  which is affected directly by a status variable  $C_l$  and indirectly by a feature variable  $F_l$  (i.e.,  $F_l$  has a direct edge to  $C_l$ ). In addition, each prediction  $P_l$  is affected by the status variables at the neighboring cells.

## 4 EXPERIMENTS

In this section, we experimentally evaluate the accuracy and scalability of *Flash* in building SPGM models for three spatial analysis applications. In these applications, we compare the performance of *Flash* with *ngspatial* [14], *shmm* [30], and *bnspatial* [7] tools when building SMRF [6], SHMM [12], and SBN [7] models, respectively.

### 4.1 Experimental Setup

**Applications.** The details of the three applications, along with their datasets, used in our experiments are described as follows:

**Bird Monitoring.** This application predicts the existence of a bird species across a certain area. Ornithologists model this problem using SMRF [14] as shown in [1], where the area is divided by a two-dimensional grid. Each grid cell holds a binary prediction variable indicating the presence or absence of the bird at this cell, and a set of feature variables that help predicting the value of this prediction variable. Then, the prediction at any cell is determined based on the values of feature variables at this cell along with a set of predicted or observed values at neighbouring cells. As a case study, we use the daily distribution of a certain bird species, namely Barn Swallow, from Ebird dataset [31], which contains more than 360 Million observations collected over North America. We define a grid of 84K cells, and map each observation to one cell. Then, we build the SMRF-based prediction model of the bird existence at cells with no observations.

**Safety Analysis.** The objective of this application is to infer the safety level (e.g., low, medium and high) at a bunch of neighbouring locations simultaneously based on reported incidents at these locations. This application has been usually represented with SHMM [12] as shown in [5], where the safety level at each location is considered a hidden state to be predicted and the reported incident at this location is an observation that affects the prediction

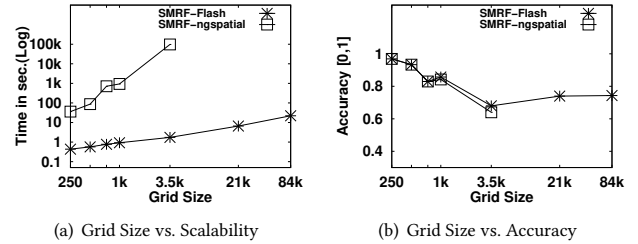


Figure 3: Study of SMRF Model Scalability and Accuracy.

value. As a case study, we use the official Chicago crime dataset repository [9], which contains around 7 Million reported incidents (i.e., observations) over 500K grid locations.

**Land Use Change Tracking.** The objective of this application is to determine whether there will be a change in the land use or not. For example, the land in a location  $l$  could be suitable for agriculture, however, given certain factors (e.g., crowded neighbourhoods), it is expected to be for human use soon. We model this application as SBN problem. As a case study, we use a grid dataset containing one Million cells of land cover distribution over Minnesota state, and is compiled from national land cover data repository [32].

In each application, we randomly select 15% of its grid data for testing, and use the rest data for training any SPGM model.

**Environment.** We run all experiments on a single machine with Ubuntu Linux 14.04. Each machine has 8 quad-core 3.00 GHz processors, 64GB RAM, and 4TB hard disk.

**Metrics.** We use the total running time of learning the parameters of any SPGM model as a scalability evaluation metric, and the ratio of correctly predicted cells using the learned model to the total number of test cells as an accuracy evaluation metric.

### 4.2 Experimental Results

**4.2.1 Study of SMRF Scalability and Accuracy.** In this section, we compare the performance, both scalability and accuracy, of *Flash* with *ngspatial* [14], when learning and using the SMRF models that are built for five different sizes of Ebird grid data.

Figure 3(a) shows the running time for each algorithm to learn the SMRF parameters while scaling the grid size from 250 to 84k cells. For all sizes, *Flash* was able to significantly reduce the running time compared to *ngspatial*. Specifically, *Flash* and *ngspatial* have an average running time of 4.7 seconds and 5.5 hours, respectively. This means that *Flash* has at least three orders of magnitude reduction in the running time over *ngspatial*. Note that the *ngspatial* curve in Figure 3(a) is incomplete after a grid size of 3.5k cells because of a failure in satisfying the memory requirements needed for its internal computations. In contrast, the running times for *Flash* are complete. This shows the *Flash* efficiency when scaling up the grid size regardless of the model specified.

Figure 3(b) shows the accuracy for each algorithm while using the same grid sizes in Figure 3(a). As can be seen in the figure, *Flash* has almost the same accuracy achieved by *ngspatial* at small grid sizes, while it is slightly more accurate (4% more) than *ngspatial* at the grid size of 3.5k cells. Note that the *ngspatial* curve is incomplete for grids with sizes more than 3.5k cells as in Figure 3(a).

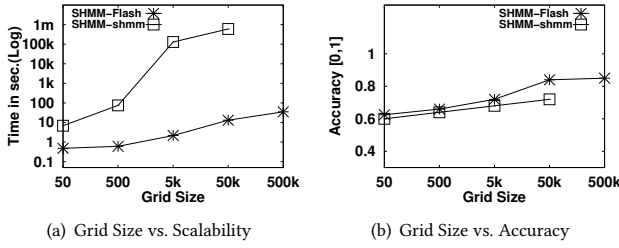


Figure 4: Study of SHMM Model Scalability and Accuracy.

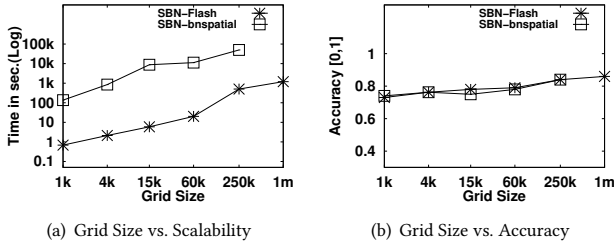


Figure 5: Study of SBN Model Scalability and Accuracy.

**4.2.2 Study of SHMM Scalability and Accuracy.** In this section, we compare the performance, both scalability and accuracy, of *Flash* with *shmm* [30], when learning and using the SHMM models that are built for five different sizes of Chicago crime grid data.

Figure 4(a) shows the running time for each algorithm while scaling the grid size from 50 to 500k cells. We can observe from the results that *Flash* has an average three orders of magnitude less running time than *shmm*. In contrast to *ngspatial* in Figure 3(a), *shmm* is more scalable to relatively large grid sizes (e.g., 50k cells), but, still can not complete the running for huge sizes like 500k cells.

Figure 4(b) shows the prediction accuracy for each algorithm while using the same grid sizes in Figure 4(a). We observe that *Flash* is consistently more accurate than *shmm* at all sizes, yet, *Flash* has a larger improvement ratio when the grid size becomes larger (the improvement ratio can reach to 18%). Note that the *shmm* curve is also incomplete for the grid with 500k cells as in Figure 4(a).

**4.2.3 Study of SBN Scalability and Accuracy.** In this section, we compare the performance, both scalability and accuracy, of *Flash* with *bnsptial* [7], when learning and using the SBN models that are built for six different sizes of Minnesota land use data.

Figure 5(a) shows the running time for each algorithm while scaling the grid size from 1k to 1 million cells. In general, *Flash* is much faster than *bnsptial* in all cases, however the ratio of improvement in case of SBN is less than its counterpart in SHMM. We observe from the results that *Flash* has at least two orders of magnitude less running times than *bnsptial*. The running times of *Flash* range from 0.6 sec (min. value) to 20 hours (max. value).

Figure 5(b) shows the accuracy for each algorithm while using the same grid sizes in Figure 5(a). As shown, *Flash* does not improve so much over the accuracy already obtained by *bnsptial*. In general,

the main objective of *Flash* is to speed up the computation steps of SPGM models, while keeping the same accuracy obtained by the best state-of-the-art techniques or increasing it, if possible.

## REFERENCES

- [1] R. Ambrosini et al. Modelling the Progression of Bird Migration with Conditional Autoregressive Models Applied to Ringing Data. *PLOS ONE*, 9(7):1–10, 2014.
- [2] L. Anselin et al. Spatial Econometrics in an Age of CyberGIScience. *IJGIS*, 2012.
- [3] S. Balbi et al. A Spatial Bayesian Network Model to Assess the Benefits of Early Warning for Urban Flood Risk to People. *Natural Hazards and Earth System Sciences*, 2016.
- [4] S. Banerjee, A. Gelfand, A. Finley, and H. Sang. Gaussian Predictive Process Models for Large Spatial Data Sets. *Royal Statistical Society Journal*, 70(4), 2008.
- [5] F. Bartolucci et al. A Latent Markov Model for Detecting Patterns of Criminal Activity. *Royal Statistical Society Journal*, 170(1):115–132, 2007.
- [6] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Royal Statistical Society Journal*, 1974.
- [7] bnsptial: Spatial Implementation of Bayesian Networks. [cran.r-project.org/web/packages/bnsptial/](http://cran.r-project.org/web/packages/bnsptial/), 2019.
- [8] R. Cano, C. Sordo, and J. M. Gutiérrez. *Applications of Bayesian Networks in Meteorology*, pages 309–328. Springer, 2004.
- [9] Chicago Crime Data, 2019. [data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzpq8t2/](http://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzpq8t2/).
- [10] D. Cornford, L. Csató, and M. Opper. Sequential, Bayesian Geostatistics: A Principled Method for Large Data Sets. *Geographical Analysis*, 37(2), 2005.
- [11] J. Gonzalez-Redin et al. Spatial Bayesian Belief Networks as a Planning Decision Tool for Mapping Ecosystem Services Trade-offs on Forested Landscapes. *Environmental Research*, pages 15–26, 2016.
- [12] P. J. Green and S. Richardson. Hidden Markov Models and Disease Mapping. *JASA*, pages 1055–1070, 2002.
- [13] M. Haran. *Gaussian Random Field Models for Spatial Data*, pages 449–478. Chapman and Hall/CRC, 2011.
- [14] J. Hughes. ngspatial: A Package for Fitting the Centered Autologistic and Sparse Spatial Generalized Linear Mixed Models for Areal Data. *The R Journal*, 2014.
- [15] D. Jain, K. von Gleissenthall, and M. Beetz. Bayesian Logic Networks and the Search for Samples with Backward Simulation and Abstract Constraint Learning. In *KI 2011: Advances in Artificial Intelligence*, 2011.
- [16] F. V. Jensen. *Introduction to Bayesian Networks*. Springer-Verlag, 1996.
- [17] G. Kabir, S. Tesfamariam, A. Francisque, and R. Sadiq. Evaluating Risk of Water Mains Failure using a Bayesian Belief Network Model. *European Journal of Operational Research*, 2015.
- [18] M. Kaiser et al. Goodness of Fit Tests for a Class of Markov Random Field Models. *The Annals of Statistics*, pages 104–130, 2012.
- [19] D. Landuyt, S. Broekx, R. D’hondt, G. Engelen, J. Aertsens, and P. L. M. Goethals. A Review of Bayesian Belief Networks in Ecosystem Service Modelling. *Environmental Modelling and Software*, 2013.
- [20] M. Mello, B. Rudorff, M. Adami, and D. Aguiar. An R Implementation for Bayesian Networks Applied to Spatial Data. *Procedia Environmental Sciences*, 2011.
- [21] F. Niu et al. Tuffy: Scaling Up Statistical Inference in Markov Logic Networks Using an RDBMS. *PVLDB*, 4(6):373–384, 2011.
- [22] M. Richardson and P. M. Domingos. Markov Logic Networks. *Machine Learning*, pages 107–136, 2006.
- [23] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005.
- [24] I. Sabek and M. F. Mokbel. Sya: Enabling Spatial Awareness inside Probabilistic Knowledge Base Construction. In *ICDE*, 2020.
- [25] I. Sabek, M. Musleh, and M. Mokbel. TurboReg: A Framework for Scaling Up Spatial Logistic Regression Models. In *SIGSPATIAL*, pages 129–138, 2018.
- [26] I. Sabek, M. Musleh, and M. F. Mokbel. A Demonstration of Sya: A Spatial Probabilistic Knowledge Base Construction System. In *SIGMOD*, 2018.
- [27] I. Sabek, M. Musleh, and M. F. Mokbel. RegRocket: Scalable Multinomial Autologistic Regression with Unordered Categorical Variables Using Markov Logic Networks. *ACM TSAS*, 5(4):27:1–27:27, 2019.
- [28] S. Shekhar et al. Identifying Patterns in Spatial Information: A Survey of Methods. *WIREs: Data Mining and Knowledge Discovery*, pages 193–214, 2011.
- [29] J. Shin et al. Incremental Knowledge Base Construction Using DeepDive. *PVLDB*, 8(11):1310–1321, 2015.
- [30] shmm: An R Implementation of Spatial Hidden Markov Models. [github.com/mawp/shmm](https://github.com/mawp/shmm), 2019.
- [31] B. Sullivan et al. eBird: A Citizen-based Bird Observation Network in the Biological Sciences. *Biological Conservation*, pages 2282–2292, 2009.
- [32] USGS National Land Cover. [archive.usgs.gov/archive/sites/landcover.usgs.gov/landcoverdata.html](http://archive.usgs.gov/archive/sites/landcover.usgs.gov/landcoverdata.html), 2019.
- [33] M. Wick et al. Scalable Probabilistic Databases with Factor Graphs and MCMC. *PVLDB*, 3(1-2):794–804, 2010.
- [34] J. Zhu, H.-C. Huang, and J. Wu. Modeling Spatial-temporal Binary Data using Markov Random Fields. *JABES*, 2005.

# SRC: Discovering Human Activity Community in A City

Tianyuan Huang  
South China University of Technology  
Guangzhou, China  
tianyuanhuang@outlook.com

## ABSTRACT

This study investigates human activity community in a city by conceptualizing it as a network embedding problem. In order to learn the latent representations of activity-travel patterns from individual daily trajectories, network embedding learns a vector space representation for each type of activity place as a node connected by movement links to preserve the structure of individual activities. The proposed approach is applied to mobile positioning data at the individual level obtained for a weekday from volunteers at Guangzhou City. Assessments are conducted to validate individual decision making for several types of activities by a field survey. This study contributes to a general framework for discovering individual activity-travel patterns from human movement trajectories.

## CCS CONCEPTS

• Information systems → Geographic information systems

## KEYWORDS

Network embedding, Graph representation, Geographic knowledge discovery, Urban planning

## 1 INTRODUCTION

Discovering communities of places by human activities can enable a variety of valuable applications. First, it can provide a quick understanding of a complex city. Trajectory data can identify people's activity-travel patterns and have a better interpretation of complex activities in urban areas. Second, given that one user's trajectory data reflects that user's behavioral preferences, the massive amount of trajectory data of a city can reflect the general behavioral characteristics of its urban residents and the general demand for urban space. In this way, extracting behavioral characteristics from user trajectory data can effectively evaluate the rationality of urban functional layout, in order to assess urban planning strategies.

There has been recent research [3] that used human trajectory and remote-sensor data to discover regions of different functions in a city. There is also research [1] that characterize the relationship among the POIs based on how people access them. In this paper, we introduce a network embedding (Node2Vec) [2] technique, which has been proven successful on multi-label classification in several real-world networks from diverse domains, into individual trajectories analysis. By constructing a network of places for human activities as nodes connected by movement links, activity-travel patterns can be extracted as latent features of the nodes. The latent

features will capture the neighborhood similarity between different places and encode people's daily interactions with different places in a continuous vector space with a relatively small number of dimensions.

## 2 METHODOLOGY

The input data are individual movement trajectories represented as sequences of locations that are attached to Point of Interests (POIs). We seek to construct a network from the trajectory data and then apply the Node2Vec Algorithm to learn an embedding representation for activity places, based on which we analyze communities of human activity places.

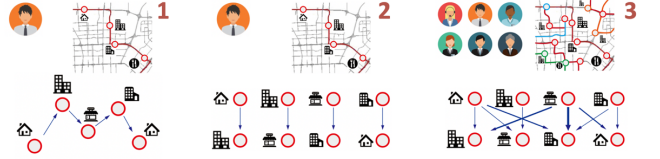


Figure 1: Overview of Network Construction Process

### 2.1 Construction of Activity Networks

A network is formally defined as  $G = \langle V, E \rangle$  in which entities (the nodes in  $V$ ) are linked by ties (the edges in  $E$ ), representing any sort of connection, similarity or interaction [1]. We propose to construct an activity network with a set of nodes which correspond to the set of activity places, represented by POIs, where users stay to perform some activities. The overall process is shown in Figure 1. Given a universal set  $V$  of  $n$  places  $x_k$  in a city, one visits a subset of  $m$  places within  $V$  and it forms one's trajectory that can be represented as a series of place, for example,

$$traj_i = [x_1, x_2, \dots, x_k, \dots, x_m], x_k \in V \quad (1)$$

which is demonstrated in Step 1 of Figure 1. In step 2, a user's trajectory can be decomposed into a set of Origin-Destination links; each link connects two consecutive places in the series. We denote this set as one's activity set (2).

$$act_i = \{[traj_i[j], traj_i[j + 1]] \mid j \in [0, m - 1]; j \in \mathbb{Z}\} \quad (2)$$

As (2) is a multiset, it can be written as:

$$act_i = \{[x_k, x_l]^{f([x_k, x_l])} \mid [x_k, x_l] \in A(act_i); x_k, x_l \in traj_i; x_k, x_l \in V\}$$

where  $A(\cdot)$  gives the underlying set of  $act_i$ , which only contains unique links.  $f(\cdot)$  is a function mapping a link to its number of occurrences in the multiset. In Step 3, we construct an activity network  $G$  by aggregating all individuals' (the set  $I$ ) activity sets to form a directed weighted graph.

$$G = \langle V, E \rangle, E = \sum_{i \in I} (act_i) \quad (3)$$

where  $\sum(\cdot)$  is the sum operation of multiset for all  $act_i$ . In other words,  $E$  contains all the unique links from a combination of all

individuals' activity sets; the weight for a unique link is obtained from the total occurrences of this link in the combination.

## 2.2 Network Embedding

We apply Node2Vec [2] to the networks of activity places and learn a mapping of places to a low-dimensional space of features that maximizes the likelihood of preserving the network structure of nodes. We then discover the communities of human activity places based on the learned embedding presentations.

## 2.3 Community Discovery

In this study, communities are groups of highly interactive and densely connected places that are frequently visited by individual trajectories. These places will be mapped as points close to each other in the embedding space. In order to determine the similarity between places, we calculate cosine similarity between the embeddings. Formally, *similarity* between embedding  $A$  and  $B$  could be defined as:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

For a given place, a rank of other places can be found based on the calculated cosine similarity (4). A community can be delineated by choosing the top-k ranked places surrounding the target place. Alternatively, community detection can be done with clustering techniques [1].

## 3 EXPERIMENTAL RESULTS

The set of collected observations in a weekday in Guangzhou is composed of 48,178 points for 2,575 volunteers in the Guangzhou city. We compute the network of 1,852 nodes and 21,742 edges and learn continuous feature representations for the nodes. Figure 2 shows the visualization of our network.



Figure 2: Visualization of the Trajectory Network

Ranking of Similarity	Xinfeng Residential District	Hualin International Plaza	Shisanhang Apparel Market	Qingping Medicine Market
1	Liwanhu Park	Hualin Temple	Hengbao Shopping Center	Sun-Yat-San Memorial Hospital
2	Xinglong District	7 Days Inn	Xinya Commercial Center	Guangzhou Hospital of Chinese Medicine
3	Duobao Market	Minghui Plaza	Yutian Center	Xinfeng District

Table 1: Selected Training Results

Each point is associated with a POI where user stays to perform some activities. In order to discover places having high interactions, we choose four target POIs with a relatively high frequency of visits in the dataset and find the top 3 most similar POIs of each target POI based on the embeddings and the cosine similarity measure. Table 1 shows the above selected results. To validate our network embedding approach, we complete a field investigation. Specifically, we interviewed 50 people in those four different places and requested them to label two common destinations in their daily activities. We then aggregated the results and found the top 3 most visited POIs for each place. It shows that over 50 percent of their choices are assigned into the same POIs with high similarity in Table 1.

## 4 DISCUSSION AND FUTURE WORKS

We proposed an exploratory study on human activity community based on POI data and a network embedding approach. The novelty of this work is two-fold. First, we presented an algorithm to build a complex network that synthesizes trajectories of places that people visit to conduct daily activities. Second, we employ a network embedding technique to learn a representation for activity places as nodes connected by movement links to preserve the structural human activities. Some potential future works include discovering socioeconomic groups from individual trajectories and simulation of human travel patterns to assist urban-transport planning.

## ACKNOWLEDGMENTS

The author would like to thank Dr. Zhaoya Gong from the University of Birmingham and Dr. Qiwei Ma from Tsinghua University, for their extensive guidance.

## REFERENCES

- [1] I. R. Brilhante, M. Berlingerio, R. Trasarti, C. Renso, J. A. F. d. Macedo and M. A. Casanova, "Come Together: Discovering Communities of Places in Mobility Data," 2012 IEEE 13th International Conference on Mobile Data Management, Bengaluru, Karnataka, 2012, pp. 268-273.
- [2] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 855-864.
- [3] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, 186-194.



# **The SIGSPATIAL Special**

## **Section 2: Workshop Reports**

---

**ACM SIGSPATIAL**  
**<http://www.sigspatial.org>**



# The 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)

Marc van Kreveld<sup>1</sup>, Bettina Speckmann<sup>2</sup>, Matei Stroila<sup>3</sup>, Goce Trajcevski<sup>4</sup>

<sup>1</sup>Department of Information and Computing Sciences, Utrecht University, the Netherlands

<sup>2</sup>Department of Mathematics and Computer Science, TU Eindhoven, the Netherlands

<sup>3</sup> HERE Technologies, Chicago, USA

<sup>4</sup>Software Engineering Program, Iowa State University, Ames, USA

## 1 Introduction

Modern technology allows us to track essentially anything that moves, be it animals, people, vehicles, or hurricanes. As a result, many efficient computational methods have been developed to analyze movement data, including methods for similarity analysis, clustering, segmentation, classification, and pattern detection. However, movement rarely occurs in isolation and to truly understand movement data it is of paramount importance to understand the intrinsic and extrinsic factors that influence movement, such as or health conditions or motivation (intrinsic) or the (natural) environment, weather, and other surrounding entities (extrinsic). Often the data that describes these factors is available together with the tracked object data for analysis, but comparatively few computational techniques fully utilize the potential of such multifaceted data. This workshop brought together researchers who are interested in developing computational techniques to analyze movement data in conjunction with other data sources that capture (some of) the factors which influence movement.

## 2 Preparing the program

To select contributed talks for the workshop, a program committee was formed, consisting of a representative selection of active researchers in the workshop topic. The call for papers was distributed several months before the workshop, and eight submissions were received. One submission clearly did not follow the submission guidelines and was out of scope, so it did not enter the reviewing process. The other seven submissions were reviewed by three PC members each. The PC provided high-quality feedback in their reports, which helped to improve the quality of the submitted papers. In case of disagreements between PC members, we followed a discussion cycle to assess how severe shortcomings of the papers were. Then we cleaned up the reviews to ensure positive or constructive feedback only. The PC members are:

- Maike Buchin, Ruhr-University Bochum, Germany
- Somayeh Dodge, University of California, Santa Barbara, USA
- Christian S. Jensen, Aalborg University, Denmark
- Marc van Kreveld, Utrecht University, the Netherlands (chair)
- Wouter Meulemans, TU Eindhoven, the Netherlands
- Bettina Speckmann, TU Eindhoven, the Netherlands

- Matei Stroila, HERE, Chicago, USA
- Goce Trajcevski, Iowa State University, Ames, USA
- Rob Weibel, University of Zurich, Switzerland

The process led to the acceptance of five submissions and the rejection of two. The accepted submissions were revised for publication in the ACM Digital Library, as was an abstract from the invited speaker.

### 3 The workshop

The workshop was held as a half-day event just before the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2019) in Chicago on Tuesday, November 5 in the afternoon. 24 people attended the workshop, in nearly all cases for the full duration. The five contributed talks were 20 minutes long; the invited talk was 55 minutes long. This included time for questions. As the last scheduled part of the workshop, time was reserved for around-the-table discussions with the presenters, who each sat at different tables and discussed with several of the participants. This was a nonstandard item of a typical workshop, and we were happy to see that it was appreciated by the workshop attendees.

The first paper, entitled “Understanding Movement in Context with Heterogeneous Data”, addressed nine main challenges in movement, focusing on mobility and heterogeneity [2]. The second paper, entitled “Shared Micro-mobility Patterns as Measures of City Similarity”, explored the possibilities of using e-bike and e-scooter data to understand and compare cities [5]. The third paper, entitled “Inferring Semantically Enriched Representative Trajectories”, assumed that a cluster of similar trajectories is enriched with semantic information and considered how this can help in computing a median trajectory for that cluster [6]. The fourth paper, entitled “Latent Terrain Representations for Trajectory Prediction”, addressed the interplay between terrain and chosen trajectories by vehicles or on foot, in order to predict traversed paths [3]. The fifth paper, entitled “A Repository of Network-Constrained Trajectory Data”, discussed the challenges involved in setting up a repository of trajectories constrained to a network (often road network) while supporting the most important query types [4]. Finally, the invited presentation, entitled “Location Graphs for Movement Data Modeling, Analytics, and Visualization”, addressed the representation of context data that may influence movement, including the spatial, temporal, and semantic component [1].

All papers together revealed that there are many facets of multifaceted movement data computation and analysis that give rise to interesting and innovative research.

The workshop organizers thank all submitters, presenters, PC members, and participants for their contributions to this successful workshop.

### References

- [1] Craig Barnes. Location graphs for movement data modeling, analytics, and visualization. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*, 2019.
- [2] Onur Derin, Aniket Mitra, Matei Stroila, Bram Custers, Wouter Meulemans, Marcel Roeloffzen, and Kevin Verbeek. Understanding movement in context with heterogeneous data. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*.
- [3] Andrew Feng and Andrew S. Gordon. Latent terrain representations for trajectory prediction. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*, 2019.



- [4] Stefan Funke and Sabine Storandt. A repository of network-constrained trajectory data. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*, 2019.
- [5] Grant McKenzie. Shared micro-mobility patterns as measures of city similarity. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*, 2019.
- [6] Jana Seep and Jan Vahrenhold. Inferring semantically enriched representative trajectories. In *Proc. 1st ACM SIGSPATIAL International Workshop on Computing with Multifaceted Movement Data (MOVE++ 2019)*, 2019.

# 1st ACM SIGSPATIAL Workshop on Geo-computational Thinking in Education (GeoEd 2019)

## Chicago, Illinois, USA - November 5, 2019

Amr Magdy  
University of California, Riverside  
amr@cs.ucr.edu

Coline Dony  
American Association of Geographers  
cdony@aag.org

The 1st International Workshop on Geo-computational Thinking in Education (GeoEd 2019) was held in conjunction with the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2019). The workshop is intended to bring together experts from both geography (or related) and computer science disciplines who have primary interest in geospatial data and technologies, either from academia or industry, to discuss the grand challenges towards improving existing learning pathways through integration of geo-computational thinking in higher education. This could impact a variety of disciplines that increasingly deal with geospatial data beyond geography, such as social sciences, environmental sciences, public policy, climatology, and other geo-related disciplines. The workshop speakers and attendants have discussed their vision on challenges and opportunities of various topics within the workshop scope.

The workshop is part of a bigger vision by the two organizers to integrate their research expertise with education to impact future of learning in various geo-related disciplines. In 2018, the American Association of Geographers has announced the Encoding Geography initiative, which is a long-term commitment to build capacity and broaden participation for computational thinking within the geography discipline. The initiative has several goals towards strengthening the future of geography, including training new generations of undergraduate and graduate students for the geospatial technology industry which is having a growing and significant impact. A 2017 global impact study on geospatial services estimates that this industry creates approximately 4 million direct jobs and generates 400 billion U.S. dollars globally in revenue per year. The growth of this industry is increasing the demand for graduates with training in both geography and computational thinking (geo-computational thinking), but they are hard to find. The limited availability of learning pathways towards geo-computationally intensive jobs requires employers across the public and private sectors to choose between hiring a geographer with limited or no computational skills, or a computer science graduate with limited or no expertise in spatial thinking and geographic information.

This first version of the workshop was a series of invited position papers and talks from teaching and research experts in geospatial technologies from both computer science and geo-related disciplines, from both academia and industry. The workshop has attracted five position papers, one keynote speech, and one panel discussion. Total seven attendants registered for the workshop. However, the actual number of attendants were thirteen along the day, with seven to nine attendants maintained at all the times.

The workshop has achieved its goal with bringing up together experts and conducted a set of productive discussions and talks. A detailed summary of the day discussion is being prepared collaboratively by the attendants and organizers to serve as an input for the research community to follow up on the identified research gaps and opportunities, and for the following anticipated versions of the workshop.

The workshop papers and discussions have identified several grand challenges and educational research opportunities on the topics of interests. The workshop has featured an excellent keynote talk by Prof. Shashi

Shekhar from the Department of Computer Science and Engineering at the University of Minnesota - Twin Cities, where he directs the Spatial Computing Research Group. The talk was titled “*Spatial Computing Education: A Perspective*” . The talk has quickly went through the journey of Prof. Shekhar over a quarter century of spatial computing educational activities, including in-class courses, online MOOCs, co-authoring and co-editing impactful books, articles, and encyclopedias, in addition to several members of his group who joined academia as educators in the field. Then, the talk has discussed several challenges that face spatial computing education. This included a rich set of examples for misusing language terms between different disciplines and stereotypes that lead to underestimating the impact of each discipline on other disciplines with detailed examples on each. The talk finally discussed several educational models and efforts to incorporate them in spatial computing courses, discussing several behavioral, cognitive, and socio-cultural factors that affect them. A main message was the significance of learning computational thinking skills for students to be considered beyond simply programming skills to prepare for changes during their careers. The talk has triggered a very interactive discussion with several workshop attendants along the following sessions of the day.

# The 1st ACM SIGSPATIAL International Workshop on Geospatial Data Access and Processing APIs (SpatialAPI 2019) Chicago, IL, USA - November 5, 2019

Ahmed Eldawy  
eldawy@ucr.edu  
University of California, Riverside, USA  
(Workshop Chair)

With the increasing amounts of geospatial data, there is a growing demand on by developers and researchers to analyze geospatial data efficiently. The SIGSPATIAL community is both a provider of new systems with cutting edge technology on maintaining and processing geospatial data, and a user for all these systems. The SpatialAPI workshop is designed to help the SIGSPATIAL community by growing the knowledge of the existing well-established systems that are available for accessing and processing geospatial data. This includes, but is not limited to, web APIs, programming libraries, database systems, and geospatial extensions to existing systems.

This year, we had a public call for two-page proposals for tutorials. Each proposal included a summary of the tutorials and an outline of the topics covered by the tutorial. We also identified top leaders to serve as advisory board to help steering the workshop. We received a total of six tutorials which were reviewed by the Chair and the advisory board to select five of them. In the selection process, we considered the quality of the tutorial proposal, the diversity of the proposals, and the goals of the workshop. We had a total of 11 registrations for the workshop which includes the Chair and the speakers.

During the workshop, we had a morning plenary session where each speaker gives a five-minute presentation for each tutorial. During that session, there were 17 attendees in the room. After that, we split the room into two parts where two tutorials are running in parallel, two in the morning and three in the afternoon. The total number of attendees in both sessions was around 15-18 during the entire day. To avoid lengthy session, we instructed the speakers to split each tutorial into two 45-minute parts, instruction and development. The instruction part (45 minutes) gives an overview of the topic, the functions covered in the tutorial, and prerequisites to run the development part. The development part (45 minutes) gives hands on experience with actual code development and system interaction. Most of the audience were ready with their laptops to engage into the experience.

At the end of the day, we had a discussion about the workshop with the presenters and attendees to improve the workshop in the future. In general, the feedback was positive and everyone thought they had a good experience in the workshop. They also liked the mix of instruction and development parts in the tutorials and some of them mentioned it would be better to give more flexibility with the length of these two parts rather than the 45-45 minutes structure. The attendees also mentioned it would be better to have a single track rather than two parallel tracks as many of them wanted to attend all the tutorials but were not able to. From the organization perspective, we think that the workshop was well-executed and the presenters did a good job in preparing and presenting the tutorials. One issue we would like to address in the future is to work more with the presenters to make sure there is a well-structured development part that the audience can follow. Some of the tutorials required a lengthy setup process and if the audience missed it or could not do it, they missed most of the development experience.

# Spatial Gems 2019 Workshop Report

## The 1st ACM SIGSPATIAL International Workshop on Spatial Gems

John Krumm<sup>1</sup>, Cyrus Shahabi<sup>2</sup>, Andreas Züfle<sup>3</sup>

<sup>1</sup>Microsoft Research AI, Microsoft Corporation, USA

<sup>2</sup>Computer Science Department, University of Southern California, USA

<sup>3</sup>Department of Geography and Geoinformation Science, George Mason University, USA  
jckrumm@microsoft.com, shahabi@usc.edu, azufle@gmu.edu

### 1 Introduction to Spatial Gems

Researchers and practitioners working with spatial data often develop fundamental new techniques they would like to share with their community. These are not necessarily new research results, not yet in any textbook, but they are interesting, self-contained techniques for doing something useful in the domain of spatial data. We call these techniques “spatial gems”.

The goal of this workshop is to publish several spatial gems contributed by the participants. While a gem may have already been published as a small part of a paper, extracting it into a gem makes it much more likely to be found and used by others. Good gems will stay relevant for a long time. Each gem will be two to six pages long. Where appropriate, a good gem will include numerical examples so programmers can verify their implementations, but it should not be a research paper with results on multiple test cases. Spatial gems should be reproducible and usable. Thus, we encourage authors to provide implementation details and code whenever possible. Code can be included in short blocks of code in the paper, or longer code can be shared in an open source repository with a pointer in the paper. At the workshop, participants work together to edit all the accepted submissions for clarity and utility, with the goal of creating a reference archive of spatial techniques.

### 2 Spatial Gems 2019 Submissions

The 1st ACM SIGSPATIAL International Workshop on Spatial Gems (Spatial Gems 2019, <https://www.spatialgems.net>) was held in conjunction with 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems in Chicago, Illinois, USA. The first Spatial Gems Workshop had nine paper submissions which were reviewed by eight program committee members. Each paper was assigned to three reviewers and all papers received at least two reviews. While the number of submissions was low, the quality of the submissions was high. Among the seven accepted papers, reviewers recommended 13 accepts, two neutrals, and one reject and the workshop had 10 registrations.

### 3 Spatial Gems 2019 Workshop Program

Spatial Gems puts the “work” back in “workshop”. Therefore, the goal of this workshop is not only to present research, but also for workshop attendees to team up and to hands-on work on each others papers in  $\LaTeX$  during the workshop. For this purpose, the workshop was split into three parts:

1. Paper Presentations
2. Editing Session I
3. Editing Session II

#### 3.1 Paper Presentations

After a short introduction by the chairs, the workshop kicked off by having ten-minute presentation for each accepted spatial gem:

- First, Gil Wolff introduced a heat map segmentation algorithm to automatically detect high density areas among a background of low density areas [7] (“Heat map segmentation”),
- then, ABM Musa presented an algorithm for online trajectory compression that allows to specify error and delay bounds [5] (“Online location trajectory compression”),
- Tin Vu demonstrated a standard method for generating synthetic spatial data that can be used in benchmarking and scalability tests [6] (“Spatial data generators”),
- then, Andreas Züfle showed an efficient, complete, and sufficient techniques to determine spatial domination of multidimensional rectangles [1] (“Complete and sufficient spatial domination of multidimensional rectangles”),
- Next, Randolph Franklin presented simple representations of polygon and polyhedra for efficient computation of area and volume [2] (“Minimal representations of polygons and polyhedra”),
- Joon-Seok Kim described a simplification for polygons specialised for building footprints [3] (“Simplification of indoor space footprints”),
- Finally, given two normally distributed location measurements of a moving object, John Krumm presented an approach to compute the speed distribution of this object [4] (“Speed distribution from normally distributed location measurements”).

During the presentations, we had about 15 people in the room. After these short presentations, workshop attendees were split into small teams of two to three people to work hands-on the  $\LaTeX$  source code of each others papers in two editing sessions.

#### 3.2 Editing Session I

To ensure that attendees were assigned to teams that match their interests, each workshop attendee anonymously provided a list of papers that they preferred working on. These preferences were fed to a matching algorithm during the coffee break at 10:30am and workshop attendees were assigned to the resulting groups after the break.

After the coffee break, at 11:00am, the assigned groups got together and received printed hard-copies of each others papers. Each team had 30 minutes to read each others paper to provide constructive feedback. Authors were instructed to focus on improving readability and ease of understanding, to improve the impact and usability

of each others spatial gem. After this reading phase, teams had 30 minutes to discuss and implement changes in the  $\text{\LaTeX}$ source code. For this purpose, all workshop papers were required to share their code in Overleaf to work concurrently on their papers. Changes of the papers included minor edits related to grammar and typos, but also major changes to clarity such as adding examples, adding motivation, or removing unnecessary sections for brevity.

During the nature of hands-on working on each others papers, the editing sessions mainly had authors and workshop organizers working on the papers, as well as a few students looking to improve their  $\text{\LaTeX}$ writing skills. The room had an average of 10 people in the room during these sessions.

### 3.3 Editing Session II

The matching algorithm used for Editing Session I was re-run during the lunch break from 12:30-2:00pm, subject to the constraint of not assigning any groups that had previously been assigned in Editing Session I. All workshop attendees reassembled at 2:00pm, were assigned to their new groups, and received updated print out of their assigned paper(s) to reflect that changes made in Editing Session II.

Again, each team was given 30 minutes to read each others paper and another 30 minutes to discuss and implement changes directly in the  $\text{\LaTeX}$ code. Editing Session II took much longer than the planned 2:00-3:30pm, as workshop members had now read many of each others papers, such that discussions about styles and best practises of paper writing were discussed across teams. Also, since we had an odd number of workshop attendees, each editing session had one team of three, which required each team members to read, discuss and edit two other paper. Discussions and editing continued well beyond the final break at 3:30pm.

After the extended Edition Session II, the workshop decided not to implement a third editing session, as most workshop attendees had already read most other workshop papers, and especially the papers that they were most interested in. Thus, we decided to extended Session II, and afterwards, concluded the workshop at around 4:00pm.

## References

- [1] T. Emrich, H.-P. Kriegel, A. Züfle, P. Kröger, and M. Renz. Complete and sufficient spatial domination of multidimensional rectangles. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [2] W. R. Franklin and S. V. G. de Magalhães. Minimal representations of polygons and polyhedra. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [3] J.-S. Kim and C. Wenk. Simplification of indoor space footprints. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [4] J. Krumm. Speed distribution from normally distributed location measurements. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [5] A. Musa, J. Biagioni, and J. Eriksson. Online location trajectory compression. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [6] T. Vu, S. Migliorini, A. Eldawy, and A. Bulussi. Spatial data generators. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.
- [7] G. Wolf. Heat map segmentation. In *1st ACM SIGSPATIAL International Workshop on Spatial Gems (SpatialGems 2019)*. ACM, 2019.

# ARIC 2019 WORKSHOP REPORT

## The 2nd ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities Chicago, IL, USA November 5, 2019

Bandana Kar<sup>1</sup>, Olufemi A. Omitaomu<sup>1</sup>, Xinyue Ye<sup>2</sup>, Shima Mohebbi<sup>3</sup>, Guangtao Fu<sup>4</sup>

<sup>1</sup>Oak Ridge National Lab, TN, USA

<sup>2</sup>New Jersey Institute of Technology, NJ, USA

<sup>3</sup>University of Oklahoma, OK, USA

<sup>4</sup>University of Exeter, Exeter, England

The advancements in sensor technology and ubiquity of connected devices has enabled the generation of large volume of disparate, dynamic and geographically distributed data both by scientific communities and citizens. With astonishing technological innovations and convergence, there have been major changes in peoples daily activities and social interaction. The socio-technological innovations motivate the concept of smart and connected cities. A smart city, however, is subjected to the same challenges as a conventional city, such as environmental damages, hazard impacts, access to services and resources, due to continuous population and economic growth. Therefore, it is imperative to improve our understanding of Resilient and Intelligent Cities in order to leverage technologies and artificial intelligence to tackle the challenges cities face, which range from climate change, public health, traffic congestion, economic growth, to digital divide, social equity, political movements, and cultural conflicts, among others.

Currently, the discussion about making a city intelligent and resilient are occurring on two parallel planes. The challenge is to plan and design intelligent cities under the framework of resilience so that real-time knowledge discovery from both dynamic data streams and static data sets can be accomplished to help practitioners and researchers with their policy decisions. While current developments in data science and artificial intelligence has enabled real-time analytics of online and static data sets, the issue of modeling urban plans to ensure an intelligent city that is also resilient need to be well understood to maximize the benefits of connected technologies, which is the focus of this workshop.

Following our successful ARIC 2018, the 2nd International Workshop on Advances on Resilient and Intelligent Cities (ARIC 2019) was timely. The second workshop brought together researchers and practitioners to address the challenges of integrating large-scale computing, geospatial analytics, and urban sciences in building intelligent and resilient cities. The workshop provided a platform to discuss research areas and issues in modeling urban design by considering sensor technology, edge computing, interactive visualization, modeling and simulation, and advanced data analytics.

ARIC 2019 (<https://aric2019.com/>) was held in conjunction with the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems on Nov. 5, 2017 in Chicago, IL, USA. The workshop received 17 technical submissions, and after a rigorous review process, 10 papers were accepted for full presentation, reaching an acceptance rate of 59%. Each accepted short paper was allocated 20 minutes for presentation, while accepted long papers were assigned 30 minutes. The workshop attracted 40 participants from different background and had 15 participants who registered for the workshop.



The technical program included a keynote from Dr. Rajesh Sankaran from Argonne National Laboratory, who is the R & D lead for the Chicago's Array of Things project in the morning session. Dr. Sankaran's keynote titled *Array of Things: A Fitbit for a City* discussed the computational and technological advantages and limitations of using urban sensors and edge computing to address smart and resilient city requirements. Following the keynote, a vision paper titled *Vision for a Holistic Smart City (HSC)-Integrating Resiliency Framework via Crowdsourced Community Resiliency Information System (CRIS)* was presented by Barnali Dixon and Rebecca Johns from University of South Florida. The paper presented a vision to integrate smart city initiatives in a resiliency framework to increase the integration of socio-technological innovations to meet the demands of smart and resilient cities. Following this presentation, Christa Brelsford (Oak Ridge National Laboratory (ORNL)), Rudy Arthur (Univ. of Exeter), Gautam Thakur (ORNL) and Hywel Williams (Univ. of Exeter) presented the paper titled *Using Digital Trace Data to Identify Regions and Cities*. The paper presented the idea of determining regions and cities by using digital trace data that capture human-dynamics such that a deeper understanding of social networks can be obtained for resilient city initiatives. The third paper in the session titled *Mobility Pattern Analysis for Power Restoration Activities Using Geo-Tagged Tweets* was presented by Bandana Kar (ORN) and Jacob Ethridge (Univ. of Tennessee, Chattanooga). The paper presented an analytical framework to capture mobility patterns using digital trace data to aid with real-time emergency management and response efforts. Following this paper, Ashlynn Daughton, Chrysm Watson Ross, Geoffrey Fairchild and Sara Y. Del Valle from Los Alamos National Laboratory presented their paper on *Topic Modeling to Contextualize Event-Based Datasets: The Colombian Peace Process* that discussed the role of digital trace data in understanding and capturing public sentiments during peace process to influence decision-makers. The last paper of the session titled *In-Database Geospatial Analytics using Python* was presented by Avipsa Roy (Arizona State Univ.), Edouard Fouche (Karlsruhe Institute of Technology), Rafael Rodriguez Morales (TU Dresden) and Gregor Moehler (IBM Deutschland Research and Development GmbH). In this paper, the authors presented a new method to perform fast and seamless spatial analysis without in-memory data storage, which can be used for near real-time analytics.

The afternoon session started with the paper titled *ADMSv2: A Modern Architecture for Transportation Data Management and Analysis* by Chrysovalantis Anastasiou, Jianfa Lin, Chaoyang He, Yao-Yi Chiang and Cyrus Shahabi from Univ. of Southern California. This paper presented an end-to-end data-driven system for near real-time and historical analytics of streaming and static transportation big data using machine learning. Following this paper, Alina Klerings (Ruprecht-Karls-Universität Heidelberg), Shimin Tang and Zhiqiang Chen (Univ. of Missouri-Kansas City) presented the paper titled *Structuralizing Disaster-scene Data through Auto-captioning* about an end-to-end deep learning framework with a linked CNN-LSTM architecture for auto-captioning of disaster scene data to help with damage assessment using crowdsourced images. Jerry Mount, Yazeed Alabbad and Ibrahim Demir of The Univ. of Iowa presented the third paper titled *Towards an Integrated and Realtime Wayfinding Framework for Flood Events*. The paper presented a graph-theory based framework to identify flood impacted road networks in real-time to help with navigation. Zonglin Meng, Bo Pen and Qunying Huang from Univ. of Wisconsin-Madison presented the fourth paper titled *Flood Depth Estimation from Web Images*. In the paper, the authors presented a mask R-CNN framework to determine flood depth using crowdsourced images. The last paper of the session titled *Semantics-enabled Spatio-Temporal Modeling of Earth Observation Data: An application to Flood Monitoring* was presented by Kuldeep Kurte (ORNL), Abhishek Potnis and Surya Durbha from Indian Institute of Technology, Bombay. This paper presented a semantic model named Dynamic Flood Ontology (DFO) to extract spatial and temporal information about flood impact areas from remote sensing imagery. A discussion of the participants about future directions concluded the workshop.

We sincerely thank the keynote speaker and the authors for presenting and discussing their papers as they contribute to the purpose of ARIC 2019. We also thank the program committee members for their time and effort in reviewing and evaluating the submitted papers. We hope that the proceedings of ARIC2019 will contribute to the field and stimulate new research. We also hope that the workshop series will continue to provide a leading international forum for researchers, developers and practitioners in the field of computing, urban and geospatial sciences, and data analytics to identify current and future areas of research and promote applications in practice.

# GeoSim 2019 Workshop Report

## The 2nd ACM SIGSPATIAL International Workshop on Geospatial Simulation

Joon-Seok Kim<sup>1</sup>, Hamdi Kavak<sup>1</sup>, Dieter Pfoser<sup>1</sup>, Andrew Crooks<sup>1,2</sup>, Carola Wenk<sup>3</sup>, Sarah Wise<sup>4</sup>

<sup>1</sup>Department of Geography and Geoinformation Science, George Mason University, USA

<sup>2</sup>Department of Computational and Data Sciences, George Mason University, USA

<sup>3</sup>Department of Computer Science, Tulane University, USA

<sup>4</sup>Centre for Advanced Spatial Analysis, University College London, UK

{jkim258,hkavak,dpfoser,acrooks2}@gmu.edu, cwenk@tulane.edu, s.wise@ucl.ac.uk

Space has long been acknowledged by researchers as a fundamental constraint which shapes our world. As technological changes have transformed the very concept of distance, the relative location and connectivity of geospatial phenomena have remained stubbornly significant in how systems function. At the same time, however, technology has allowed us to begin to bring tools like simulation to bear on our understanding of how such systems work. While previous generations of scientists and practitioners were unable to gather spatial data or to incorporate it into models at any meaningful scale, new methodologies and data sources are becoming increasingly available to researchers, developers, users, and practitioners. This flowering of different approaches is occurring simultaneously across many fields, and at every point in the research process.

The 2nd ACM SIGSPATIAL International Workshop on Geospatial Simulation was held in conjunction with 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems in Chicago, Illinois, USA. GeoSim 2019 brought together researchers and practitioners from a range of disciplines to disseminate their cutting-edge research in geospatial simulation. Similar to GeoSim'18[4], GeoSim'19 had 11 registered participants which was in the middle range of all workshops at ACM SIGSPATIAL 2019. In addition to these participants, at times the workshop had upwards of 25 participants ranging from academia (e.g. The University of Utah, The University of Tokyo, University College London, University of Illinois at Urbana-Champaign, Leibniz-University Hanover) to industry (e.g. Here, Amazon, AT&T).

This half-day workshop comprised two dense sessions. The workshop featured one invited talk with six oral presentations. Dr. Andrew Crooks kicked off the first session with welcome and opening remarks, having brief time for participants to introduce each others. In the first session, as a proxy of the authors, Xiqi Fei from George Mason University presented work of AT&T Labs-Research entitled “*SimCT: Spatial Simulation of Urban Evolution to Test Resilience of 5G Cellular Networks*” [2] which proposed a plausible scenario in cellular network industry. Terence Lines from University College London presented the paper entitled “*Simulating and Modeling the Signal Attenuation of Wireless Local Area Network for Indoor Positioning*” [5] which demonstrated how to apply simulation to WLAN for indoor positioning. Dr. Alexander Hohl presented his work “*Spatiotemporal Simulation: Local Ripley's K Function Parameterizes Adaptive Kernel Density Estimation*” [1].

In the second session, Dr. Sarah Wise gave an invited talk entitled “*Geospatial Simulation, Three Ways*” which is on applied geospatial simulation help to advance this mission and orient the program. Shaofeng Yang from the University of Tokyo presented his work under the title “*Firm-level Behavior Control after Large-scale Urban Flooding Using Multi-agent Deep Reinforcement Learning*” [7]. Two papers leveraging CyberGIS-Jupyter were presented by Rebecca Vandewalle and Dr. Jeon-Young Kang from the University of Illinois at Urbana-Champaign under the titles “*Integrating CyberGIS-Jupyter and Spatial Agent-based Modelling to Eval-*



Figure 1: Group Photo of GeoSim 2019 Attendees

uate Emergency Evacuation Time” [6] and “CyberGIS-Jupyter for Spatially Explicit Agent-based Modeling: A Case Study on Influenza Transmission” [3], respectively.

Based on the submitted papers received, the best paper award was given to Shaofeng Yang, Yoshiki Ogawa, Koji Ikeuchi, Yuki Akiyama, Ryosuke Shibasaki from The University of Tokyo with their paper entitled: “Firm-level Behavior Control after Large-scale Urban Flooding Using Multi-agent Deep Reinforcement Learning.”

## References

- [1] A. Hohl and P. Chen. Spatiotemporal simulation: Local ripley’s k function parameterizes adaptive kernel density estimation. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 16–23, 2019.
- [2] T. Hsiung and Y. Kanza. Simct: Spatial simulation of urban evolution to test resilience of 5g cellular networks. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 1–8, 2019.
- [3] J.-Y. Kang, J. Aldstadt, A. Michels, R. Vandewalle, and S. Wang. Cybergis-jupyter for spatially explicit agent-based modeling: A case study on influenza transmission. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 32–35, 2019.
- [4] J.-S. Kim, H. Kavak, A. Züfle, A. Crooks, U. Manzoor, and P. M. Torrens. Geosim 2018 workshop report the 1st acm sigspatial international workshop on geospatial simulation. *SIGSPATIAL Special*, 10(3):28–29, 2019.
- [5] T. Lines and A. Basiri. Simulating and modeling the signal attenuation of wireless local area network for indoor positioning. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 9–15, 2019.

- [6] R. Vandewalle, J.-Y. Kang, D. Yin, and S. Wang. Integrating cybergis-jupyter and spatial agent-based modelling to evaluate emergency evacuation time. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 28–31, 2019.
- [7] S. Yang, Y. Ogawa, K. Ikeuchi, Y. Akiyama, and R. Shibasaki. Firm-level behavior control after large-scale urban flooding using multi-agent deep reinforcement learning. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, pages 24–27, 2019.

# GeoAI 2019 Workshop Report

## The 3rd ACM SIGSPATIAL International Workshop on GeoAI: AI for Geographic Knowledge Discovery Seattle, WA, USA - November 5, 2019

Song Gao<sup>1</sup>, Shawn Newsam<sup>2</sup>, Liang Zhao<sup>3</sup>, Dalton Lunga<sup>4</sup>, Yingjie Hu<sup>5</sup>, Bruno Martins<sup>6</sup>,  
Xun Zhou<sup>7</sup>, Feng Chen<sup>8</sup>

<sup>1</sup>University of Wisconsin, Madison; <sup>2</sup>University of California, Merced;

<sup>3</sup>George Mason University; <sup>4</sup>Oak Ridge National Lab;

<sup>5</sup>University at Buffalo; <sup>6</sup>University of Lisbon;

<sup>7</sup>University of Iowa; <sup>8</sup>The University of Texas at Dallas.

Nowadays artificial intelligence (AI) is bringing tremendous opportunities and challenges to geospatial research. Big data enable computers to observe and learn the world from many different perspectives, while high performance machines support the development, training, and deployment of AI models within reasonable amount of time. Recent years have witnessed significant advances in the integration of geospatial study and AI in both academia and industry. There have already been many successful studies for both physical environment and human society. Focusing on modeling the physical nature, research has shown that deep learning can improve the representation of clouds that are smaller than the grid resolutions of climate models. Examining the human society, AI and natural language processing methods, such as word embeddings, help quantify changes in stereotypes and attitudes toward women and ethnic minorities over 100 years in the United States. There are also many other applications that effectively integrate AI with problems in geospatial studies, such as vehicle trajectory prediction, high-definition mapping and navigation, historical map digitizing, gazetteer conflation, geographic feature extraction, and place understanding. The 3rd International Workshop on AI for Geographic Knowledge Discovery (GeoAI 2019) builds on the success of the previous workshops in 2017 and 2018. GeoAI is bringing together geoscientists, computer scientists, engineers, entrepreneurs, and decision makers from academia, industry, and government to discuss the latest trends, successes, challenges, and opportunities in the field of artificial intelligence for data mining and geographic knowledge discovery.

GeoAI 2019 received 25 paper submissions in total. After a rigorous peer-review process by the program committee, in total 17 papers (5 full papers, 10 short papers, and 2 vision papers) were accepted by the workshop and selected for presentations. Dr. Xin Chen, the Director of Engineering at HERE Technologies, gave a keynote on “HD Live Map for Automated Driving: An AI Approach” (industry keynote), and Dr. Raju Vatsavai, Professor at North Carolina State University, gave a keynote on “Geospatial AI for Monitoring Crops to Nuclear Proliferation Using Global Earth Observations” (academic keynote). This year’s workshop also featured one wrap-up discussion on “How GeoAI will progress?” Many participants contributed constructive ideas, such as encouraging future submissions with share of datasets for common benchmark testing, engagement between different domains of people, enhance data privacy and data security issues, transparency and reproducibility, and explicit spatial constraints on AI models. Participants also discussed the design of core skills for students who want to build expertise in GeoAI. There are 43 participants who officially registered to GeoAI’19. The workshop attracted in average 50 participants, with a maximum of about 80 participants in the room.

We sincerely thank our program committee for their time and efforts in reviewing the submitted papers. We



hope that the proceedings of GeoAI'19 can stimulate new ideas and make a contribution to this growing field.



Figure 1: GeoAI'19 Workshop in Chicago

# 3rd ACM SIGSPATIAL Workshop on Analytics for Local Events and News (LENS 2019)

## Chicago, Illinois, USA - November 5, 2018

Haiquan Chen  
California State University, Sacramento  
haiquan.chen@csus.edu

Federico Iuricich  
Clemson University  
fiurici@clemson.edu

Amr Magdy  
University of California, Riverside  
amr@cs.ucr.edu

The 3rd ACM SIGSPATIAL Workshop on Analytics for Local Events and News (LENS 2019) was held in conjunction with the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2019). The workshop is intended to bring together experts in the workshop scope to exchange ideas on opportunities, challenges and cutting-edge techniques for local events and news analytics. The workshop has attracted ten submissions, all full-length research papers. Six submissions were accepted for publications; with 60% acceptance rate. The papers are reviewed by 7 program committee members, where each paper is assigned to two reviewers. Total 8 attendants registered for the workshop. The actual number of attendants along the day fluctuated from 8 to 12, with an average of 9 attendants maintained almost all the time. The paper titled *DeepSpot: Understanding Online Opinion Spam by Text Augmentation using Sentiment Encoder-Decoder Networks*, by Avik Nayak, Haiquan Chen, Xiaojun Ruan, and Jinsong Ouyang was awarded the best paper award. The paper titled *Scalable Community Detection over Geo-Social Network*, by Xiuwen Zheng, Qiyu Liu, and Amarnath Gupta was awarded the best presentation award. The workshop has achieved its goal in bringing up together experts and conducted a set of productive discussions and talks.

The workshop has featured two excellent keynote speakers. The first keynote was by Dr, Grant McKenzie from the Department of Geography at McGill University in Montréal, Canada. He leads the Platial Analysis Lab, an interdisciplinary research group that works at the intersection of data science and behavioral geography. His talk title was “*Local Dimensions: Towards Platial data analytics*”. The talk focuses on the multifaceted nature of places. Our ability to experience space can change radically based on the time of the day or even our social or economic condition. The idea that the same space, i.e., the same physical location, can be characterized by way more dimensions than just a point in a 3D space is the foundation of place-based modeling. During the keynote, an example of applications deriving from place-based modeling has been discussed such as temporally enhanced reverse geocoding and, city similarity based on user-generated tourist attraction reviews. The talk has provided a very effective overview of platial data analytics as a field built upon the processing of heterogeneous data and recent advances in computation and statistics.

The second keynote was by Alexander Visheratin from ITMO University, Russia. He is a senior researcher at ITMO University where he has been working on a wide variety of problems related to data storage and processing. His talk title was “*Multiscale event detection and forecasting in complex urban environments*”. The talk focuses on the challenges of collecting and analyzing data for detecting events happening in a big

city. These span from data crawling, to data filtering for removing undesired noisy data, to the challenge of plugging in reliable learning tools for prediction. A very important aspect addressed during the talk is the new opportunities provided by temporary content analysis (like Facebook stories). The talk provided a very complete description of the work done by the research team at ITMO University and it has demonstrated the effectiveness and efficiency of the proposed approach using several real-world datasets.



# GeoHumanities 2019 Workshop Report

Bruno Martins<sup>1</sup>, Ludovic Moncla<sup>2</sup>, Patricia Murrieta-Flores<sup>3</sup>

<sup>1</sup>Instituto Superior Técnico - INESC-ID, University of Lisbon, Portugal

<sup>2</sup>LIRIS UMR CNRS 5205, INSA Lyon, France

<sup>3</sup>Digital Humanities Hub, Lancaster University, United Kingdom  
(Workshop Co-Chairs)

## Abstract

*This article reports on the 3rd ACM SIGSPATIAL Workshop on Geospatial Humanities, held in conjunction with the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. The article outlines the objectives of the workshop, and briefly describes the technical program.*

## 1 Introduction to the ACM SIGSPATIAL Workshop on Geospatial Humanities

The 3rd ACM SIGSPATIAL Workshop on Geospatial Humanities (GeoHumanities'19) was held together with the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, on November the 5th at Chicago, Illinois, USA. The workshop addressed the use of geographic information systems and other spatial technologies in humanities research.

Scholars in the humanities have long paid attention to spatial theory and cartographic outputs. Moreover, in recent years, new technologies and methods have lead to the emergence of a field that is now commonly known as the Spatial Humanities. Despite recent developments, many challenges persist in the application of state-of-the-art techniques (e.g., text geoparsing), which have been showcased in venues such as the ACM SIGSPATIAL conference. The workshop is specifically concerned with the use of geographic information systems and other spatial technologies in humanities, including history, archaeology, cultural heritage, and digital humanities research, placing a strong emphasis on new methodologies that leverage recent technical and scientific developments.

## 2 The Workshop Program

The call for papers resulted in 8 submissions describing high quality research. A program committee of 30 members, which are listed on the workshop website<sup>1</sup>, reviewed the submissions, and 6 papers were accepted for presentation. The workshop program featured two technical sessions, followed by a discussion at the end.

After a brief opening address, the workshop started with the presentation of work by Kamakura et al., concerning a digitisation project focused on the Preah Vihear Temple, i.e. a Hindu temple located on the border between Cambodia and Thailand, built in the Angkor Dynasty. The project involved specialised hardware and software for acquiring 3D point-cloud data regarding this particular structure (i.e., the authors used mobile laser range sensors developed by the University of Tokyo, suitable for measuring wide areas, as well as commercially available tripod-based sensors, together with specialised software that combines the results from multiple iterations of partial scanning). The authors also discussed the use of the 3D point-cloud data to determine the direction of the central axis of the temple, afterwards interpreting this direction in the context of Khmer legends.

The second presentation covered research by Huang et al. focused on spatial down-scaling of population data into high-resolution grids, motivated by the fact that access to this type of information is very useful in a broad

---

<sup>1</sup><http://bgmartins.github.io/sigspatial-geohumanities/program-committee.html>

range of applications. Information on building footprints, made available by Microsoft and refined through land use data from the OpenStreetMap, was used to disaggregate the census tract population data (i.e., the latest ACS 5-year estimates (2013-2017) for the conterminous United States) into a high-resolution grid of 100 meters per cell. The actual down-scaling procedure is based on dasymetric mapping, and the authors compared different weighting scenarios leveraging quantities derived from the building footprints (e.g., counts and relative sizes for the footprints within each cell) and from other remote sensing data sources (e.g., nighttime light intensity). In the discussion period that followed the presentation, Xiao Huang highlighted the fact that accurate information on building footprints does indeed provide very useful information for population down-scaling, significantly outperforming the use of other types of ancillary variables.

In the third presentation, Palumbo et al. discussed approaches for standardising category (e.g. restaurant, temple, hotel, etc.) information in point-of-interest (POI) databases, specifically in the context of data integration from multiple sources (e.g., when fusing multiple alternative data sources of POI data with the OpenStreetMap). The authors argue that having consistent category labels is an important concern when using these data for producing land use maps or population distribution estimates, discussing these two applications in detail. The process of matching the source categories, in some source, into a standardised set of categories (e.g., the OpenStreetMap tags) should involve human assessment, but systems like the SONET graph database can provide a programmatic way to manage the category mapping process, by supporting experts in the organisation, storage, and retrieval of POI categories at multiple levels of abstraction.

After the coffee break, the second session started with Brynne Godfrey presenting a paper by Andris et al., in which the authors described the use of a large call detail record (CDR) dataset, from the Trk Telekom mobile operator, to examine how refugees from the Syrian Refugee Crisis are integrating into Turkish society. Individuals were classified according to their call patterns (i.e., refugees who often call Turkish nationals, refugees who do not call Turks, Turks who often call refugees, and Turks who do not call refugees), and they were also geo-referenced with basis on the cell towers that they use. With this information, the authors analysed spatial patterns in the locations that are more commonly associated to the *bridging* members of Turkish society, and they looked for correlations between (a) the location of the different classes of individuals, and (b) different types of infrastructural and socioeconomic indicators (e.g., distance from the Syrian border, intersection with urban amenities, etc.). Bridging Turks were found to be located near particular infrastructural elements (e.g., places of worship or community centers) more often than their non-bridging counterparts, highlighting the significance of social amenities and meeting places for refugee integration. In the discussion session that followed the presentation, Brynne Godfrey and some of the participants discussed challenges involved in using CDR data for research purposes, while complying with privacy safeguards.

Afterwards, Ludovic Moncla described work developed with his collaborators towards the identification of qualitative *neighbours* for toponyms referenced in an eighteenth-century French encyclopedia (i.e., for place mentions in Diderot and d'Alembert's *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers, par une Société de Gens de lettres* (1751-1772, EDDA)). The authors proposed a graph-analysis method that draws on the co-occurrence relations between toponyms that are present in the corpus of encyclopedia articles, arguing that the co-occurrence network can act as an alternative to geospatial representations based on latitude and longitude coordinates, and as a useful proxy to geospatial proximity when no historical gazetteer exists for the source material's period. In his presentation, Moncla also discussed the use of the qualitative neighbours for toponym disambiguation applications.

Finally, Wang and Hu presented their recently proposed EUPEG benchmarking platform for evaluating text geoparsing systems (i.e., systems for recognising place mentions in textual documents, and for assigning each place name to the corresponding location coordinates). In its present version, the EUPEG platform integrates eight annotated datasets, nine baseline geoparsers, and it reports on eight different performance metrics. The authors used the platform to evaluate the different systems that participated in the SemEval 2019 shared task that focused on Toponym Resolution in Scientific Papers, against the baselines included in the platform. Although some challenges remain, the results show that some of the new geoparsers introduced in the SemEval

competition, e.g. based on state-of-the-art neural models, indeed improve the performances on multiple datasets.

A discussion period followed the last presentation of the second session, covering ideas for future developments and common aspects between the different contributions presented at the workshop (e.g., discussing the possibility of integrating other datasets and systems into the EUPEG platform from Wang and Hu, or discussing the possibility of using large POI datasets for population down-scaling).

In total, the workshop had 9 officially registered participants and, on average, 12 attendees were present at each of the session presentations. We believe GeoHumanities'19 was a successful event that, although small, allowed the participants to explore the contributions that modern GIS and geo-spatial analysis technologies can enable within and beyond the digital humanities.

### **3 Acknowledgments and Final Remarks**

The organizers would like to thank the authors for submitting and presenting their contributions, and also the program committee members for their commitment to the paper review process. We hope that the proceedings<sup>2</sup> of GeoHumanities'19 will inspire new research ideas, and that you will enjoy reading them. The organizers would also like to acknowledge the support provided by Fundação para a Ciência e a Tecnologia (FCT), through the project grant with reference UID/CEC/50021/2019, and also by the Trans-Atlantic Platform for the Social Sciences and Humanities, specifically through the 'Digging into Early Colonial Mexico: A large-scale computational analysis of 16th century historical sources' project with reference HJ-253525-ES/R003890/1.

---

<sup>2</sup><http://dl.acm.org/citation.cfm?id=3356991>

LocalRec 2019 Workshop Report

# The Third ACM SIGSPATIAL Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising

Chicago, Illinois, USA — November 5, 2019

Panagiotis Bouros  
Johannes Gutenberg University Mainz  
Germany  
bouros@uni-mainz.de

Tamraparni Dasu  
AT&T Labs–Research  
USA  
tamr@research.att.com

Yaron Kanza  
AT&T Labs–Research  
USA  
kanza@research.att.com

Matthias Renz  
Christian-Albrechts-Universität zu Kiel  
Germany  
mr@informatik.uni-kiel.de

Dimitris Sacharidis  
Technische Universität Wien  
Austria  
dimitris@ec.tuwien.ac.at

The amount of publicly available geo-referenced data has seen a dramatic explosion over the past few years. Many user activities generate data that are annotated with location and contextual information. Furthermore, it has become easier to collect and combine rich and diverse location information. In the context of geoadvertising, the use of geosocial data for targeted marketing is receiving significant attention from a wide spectrum of companies and organizations. With the advent of smartphones and online social networks, a multi-billion dollar industry that utilizes geosocial data for advertising and marketing has emerged. Geotagged social-media posts, GPS traces, data from cellular antennas and WiFi access points are used widely to directly access people for advertising, recommendations, marketing, and group purchases. Exploiting this torrent of geo-referenced data provides a tremendous potential to materially improve existing recommendation services and offer novel ones, with numerous applications in many domains, including social networks, marketing, and tourism.

Realizing the full potential of geo-referenced data requires new technologies to collect, store, analyze and use the data. It also raises issues in the area of responsibility, accountability, transparency, fairness, adequacy (e.g., avoiding ads in improper places) and preventing misconduct. This in turn means addressing many core challenges and combining ideas and techniques from various research communities, such as recommender systems, data management, geographic information systems, social network analytics, and text mining. By bringing together researchers and practitioners from these communities, the LocalRec workshop provides a unique forum for in-depth discussion about challenges, opportunities, novel techniques and applications related to location-based recommendation, geosocial networks and geoadvertising.

LocalRec 2019 was a full-day workshop (<http://www.ec.tuwien.ac.at/localrec2019/>). The program committee received and evaluated 17 submissions (12 full papers and 5 short papers), out of which 6 papers were accepted as full (acceptance rate of 50%) and 6 papers were accepted as short (overall acceptance rate of 71%). Among the main conference attendees, 22 registered for LocalRec. In most of the sessions there were about 15–20 participants, with about 25 attendees at peak time; see also Figure 1.



Figure 1: Photo from the LocalRec 2019 sessions.

There were two keynotes in the workshop. Siva Ravada from Oracle gave a keynote about *Spatial Analytics for Storing and Analyzing Geosocial data for Geoadvertising*. He provided an overview of challenges and solutions for supporting geoadvertising and other location-related business decisions. For example, when selecting a store location, it is important to estimate the number of potential customers who might use the services of that store. This requires providing information about other stores in the area, transportation to the store location, population size in the area, etc. Another example is presenting distribution of market share of competing companies in different geospatial regions. Solving these problems using ordinary database management systems does not scale, because many expensive geospatial computations are involved. Hence, indexes and query processing methods were developed for these tasks. For recommendations, there is a need to maintain location-based relationships between customers. The talk elaborated on the applicability of Property Graphs for this task, and their use in recommender systems. The talk also surveyed concepts like Augmented Location (location within the context of the user data), Beacon Technology (location targeting, customer mapping, visit tracking, geofencing), Hyper-Local Marketing and Mobile App Localization.

Syagnik Banerjee, from the University of Michigan-Flint, gave a keynote on *Geo-marketing and Situated Consumers: Opportunities and Challenges* [1]. This talk presented potential uses of location and spatial analytics in geoadvertising and geomarketing. The first part of the talk focused on the different ways in which geolocation can be used in marketing, distinguishing between conversion, retention, and discovery objectives. For example, it described uses of augmented reality in supermarkets to provide information to customers while collecting information about the items that customers looked at. It illustrated the effect of location on decisions regarding price, e.g., offering customers a low price on a Web site of a store before they get into the physical store, and avoiding offering the low price after the actual visit at the store. The second part of the talk focused on measuring the effectiveness of a marketing campaign and how it is perceived by customers with respect to their privacy concerns. Both keynote talks gained a lot of attention and led to fruitful discussions.

The research papers were organized into four sessions. The first session focused on *Spatial Analytics* and included two research papers. In The authors of [5] presented an approach for classifying Points of Interest (POIs) into accurate category recommendations using minimum amount of available POI metadata (name, coordinates). This is in contrast to existing techniques that require a wealth of metadata (e.g. reviews, ratings, working hours, price ranges). A methodology for generating sketch maps, i.e., simplified maps that improve readability by visualising only necessary or requested facilities, is discussed in [8]. The paper shows how to automatically select the necessary roads for sketching a map, for given facilities.

The second session focused on *Location Recommendations*, and included four papers. The authors of [9]

presented automatic creation of trip itineraries, while exploiting multiple social media sources. The source-to-target trips are derived using an ant colony optimization algorithm where the total score of a trip depends on its length, the social media popularity of the contained POIs, and the relevance of the POIs with respect to the city's history. In [3], inferring the rating for an unseen hiking trail is studied. The proposed approach examines similar historical trails and appropriately weighs their ratings, to predict the desired rating. Given a sparse geo-textual dataset, the goal of [11] is to predict emotions associated with certain topics at a target location. In the absence of specific location-topic combinations data, the paper applies interpolation based on combination of SVD and Kriging. The last paper of the session presented a study of various aspects of fairness in location recommender systems [13]. Fairness could relate to receivers of recommendations or to recommended locations. The desired notion of fairness is quantified based on outputs of recommenders over a period of time.

The third session included three talks related to *Privacy*. In [10], the authors considered the impact of privacy-by-design requirements on the accuracy of POI recommendations on mobile devices, based on the assumption that the recommender system can only use locally stored user information, i.e., the user trajectory. The focus is on detecting the stay area based on generated recommendations. The authors of [4] discussed the trade-off between utility and privacy in mobility datasets and the evaluation of geo-indistinguishability techniques for two distinct applications over the Geolife dataset. In [6], the authors discussed the problem of location privacy for physical infrastructure maps. Various types of obfuscation techniques, which differ in how they perturb the locations of the infrastructure nodes and/or their connections, were proposed.

The last session included three talks in the areas of *Geoadvertising and Location-Based Services*. In [7], the authors studied ways to measure the effect of location-based advertising based on uplift modeling. They identified significant geo-features for location-based targeting and predicted user segments with high net effect of advertisement. The second paper [12] described the association of semantic labels to entities in an indoor map, based on crowd-sourced images and motion traces. The third paper [14] introduced a novel augmented reality (AR) application for location-based social networks that (1) integrates mobile augmented reality with user generated content, (2) allows users to publish their own content through an augmented reality form on a 3D model aligned to the real-world scenario (with real-world coordinates) displayed by the device, and (3) allows users to interact with user-generated content published by others.

The research papers and the keynote talks emphasized the many facets of location recommendations and geoadvertising. They illustrated the potential of recommendation systems to help users find the geospatial information they need, whether it is points of interest, routes or recommended venues. They described methods for advertisers and marketing people to reach out to potential customers with high precision based on location information, and they also emphasized the need for awareness of privacy and fairness issues.

We thank the authors for publishing and presenting their papers in LocalRec 2019, and the program committee for their professional evaluation and help in the peer-review process. We hope that the proceedings [2] will inspire new research ideas and will help promoting the area of location-recommendations and geoadvertising.

## References

- [1] S. Banerjee. Geomarketing and situated consumers: Opportunities and challenges. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 13:1–13:4. ACM, 2019.
- [2] P. Bouros, T. Dasu, Y. Kanza, M. Renz, and D. Sacharidis, editors. *LocalRec '19: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*. ACM, 2019.

- [3] T. Chondrogiannis and M. Ge. Inferring ratings for custom trips from rich gps traces. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 4:1–4:4. ACM, 2019.
- [4] A. Di Luzio, A. C. Viana, K. Chatzikokolakis, G. Dikov, C. Palamidessi, and J. Stefa. Catch me if you can: How geo-indistinguishability affects utility in mobility-based geographic datasets. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 8:1–8:10. ACM, 2019.
- [5] G. Giannopoulos, K. Alexis, N. Kostagiolas, and D. Skoutas. Classifying points of interest with minimum metadata. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 1:1–1:4. ACM, 2019.
- [6] Y. Gullapalli, J. Koritzinsky, M. Syamkumar, P. Barford, R. Durairajan, and J. Sommers. Bokeh: Obfuscating physical infrastructure maps. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 9:1–9:10. ACM, 2019.
- [7] S. Kawanaka and D. Moriwaki. Uplift modeling for location-based online advertising. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 10:1–10:4. ACM, 2019.
- [8] Y. Kim, H. Fukuyasu, D. Yamamoto, and N. Takahashi. A road generalization method using layered stroke networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 2:1–2:10. ACM, 2019.
- [9] K. D. Mukhina, A. A. Visheratin, D. Nasonov, and L. Manovich. Intelligent sightseeing in immensely manifold cities: Case of 2018 fifa world cup host cities. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 3:1–3:10. ACM, 2019.
- [10] D. Proux and F. Roulland. Mobile recommendation challenges within a strong privacy oriented paradigm. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 7:1–7:9. ACM, 2019.
- [11] A. Sikder and A. Züfle. Emotion predictions in geo-textual data using spatial statistics and recommendation systems. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 5:1–5:4. ACM, 2019.
- [12] X. Teng, D. Guo, Y. Guo, P. Xu, Y. Meng, R. Hu, H. Chai, and Z. Liu. Semap: Automatic semantic mapping for large-scale indoor spaces. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 11:1–11:4. ACM, 2019.
- [13] L. Weydemann, D. Sacharidis, and H. Werthner. Defining and measuring fairness in location recommendations. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 6:1–6:8. ACM, 2019.
- [14] Y. Yue, J. Ding, Y. Kang, Y. Wang, K. Wu, and T. Fei. A location-based social network system integrating mobile augmented reality and user generated content. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*, LocalRec '19, pages 12:1–12:4. ACM, 2019.

# PredictGIS 2019 Workshop Report

## Held in conjunction with ACM SIGSPATIAL 2019

Satish V. Ukkusuri<sup>1</sup>, Kaoru Sezaki<sup>2</sup>, Takahiro Yabe<sup>1</sup>, Kota Tsubouchi<sup>3</sup>

<sup>1</sup>Lyles School of Civil Engineering, Purdue University, USA

<sup>2</sup>Center for Spatial Information Science, University of Tokyo, Japan

<sup>3</sup>Yahoo Japan Corporation, Japan

### Abstract

*This report provides a summary of the 2019 edition of the International Workshop on Prediction of Human Mobility (PredictGIS 2019), which was held in conjunction with ACM SIGSPATIAL 2019, in Chicago, IL on November 5th, 2019.*

## 1 Aim of the workshop

The prediction of human mobility is becoming an attracting field. This topic attracts researchers from a broad range of disciplines from the behavioral sciences, where understanding the complexity of the human behavior is one of the hot topics, to the industrial fields, where such results are applied for many beneficial applications. Recent progress for sensing human mobility via smartphones is boosting this trend. However, due to the complexity and context-dependence of human behavior and the incompleteness and noise of geospatial data collected from various sensors, the prediction of human mobility is still far from solved. This workshop aimed at collecting contributions on the cutting-edge studies in human mobility analysis and modeling which can advance the human mobility prediction research. Potential topics included, but were not limited to:

- Next place prediction of individual mobility
- Modeling crowd or population dynamics
- Predicting human mobility patterns during emergencies and rare events
- Modeling the dynamics of commute flow and migration flow
- Traffic congestion, road usage forecast and optimal vehicle routing
- Social event forecast using geospatial data
- Novel agent based mobility simulators
- Case studies of mobility prediction in academia and industry

Both full papers (10 pages) and short papers (4 pages) that describe timely research contributions in PredictGIS's areas of interest were solicited. This included papers reporting on initial results as well as papers discussing mature research projects or case studies of deployed systems are sought out. Submissions describing big ideas that may have significant impact and could lead to interesting discussions at the workshop were also encouraged.



## 2 Review Process and Attendance

The 3rd PredictGIS workshop was held on 5th November 2019 at the ACM SIGSPATIAL conference in Chicago, USA<sup>1</sup>. PredictGIS 2019, included 12 presentations, of which 7 were full papers, 3 were short papers, and 2 were keynote presentations. The reviewing process was competitive at an unprecedented level this year, with an acceptance rate of 55%, which was lowest among all 3 editions of PredictGIS. A total of 18 papers were submitted to the workshop, and only 10 were accepted for publication and presentation. Every paper was reviewed by at least 2 members from the technical committee, which was composed of over 15 members with expertise in relevant fields such as spatio-temporal data analysis, machine learning, and artificial intelligence. The acceptance/reject decision was made by the program chairs, by selecting the top 10 papers based on the average review scores weighted by the reviewers' level of confidence.

The average number of attendees were around 25 throughout the day, with a maximum of around 30. Attendees gathered from various countries around the world, for example from Brazil, Germany, India, Singapore, South Korea, Japan, China, and the United States (Figure 1). Overall, the workshop attracted papers with various topics, methods, and datasets. The variety of papers increased the number of topics covered in the workshop, and triggered an intense discussion between attendees on the current trends, issues, and also future research opportunities related to the prediction of human mobility.



Figure 1: Attendants of PredictGIS workshop

## 3 Highlights of the workshop

The first keynote talk was given by Professor Stanislav Sobolevsky of New York University titled “Big data for predictive modeling of urban mobility and forecasting impacts of urban solutions”. He discussed the applications of novel large scale data for understanding urban mobility and review some common approaches for its descriptive and predictive modeling. Further, several specific case studies, including the prediction and anomaly detection of the for-hire-vehicle ridership originating at major transportation hubs in New York City and predicting the mode-shift resulting from transportation innovations and policy change in order to assess their economic, social and environmental impacts were introduced. The second keynote talk was given by

---

<sup>1</sup><http://predictgis.umnilab.com/>

Professor Satish Ukkusuri of Purdue University titled “Mobility Analytics in an Era of Accelerated Technological Change”. His talk covered his recent research in the areas of: (1) smart mobility in information rich transportation environments; (2) innovations in connected/autonomous vehicles and (3) resilience of coupled socio-technical networks. Open questions and insights from these research areas were discussed, and ignited an active discussion following the talk.

The contributed talks all focused on the analysis and prediction using spatio-temporal data, but were diverse in the application domains ranging from crime detection [2], traffic accident prediction [8], bike reallocation prediction [5], to next place prediction of human mobility [9, 6]. We saw an increase in the number of the studies that utilized deep neural network architectures, including the paper on clustering spatio-temporal data, which was selected as the best paper of the workshop [1], as well as papers that used network science insights for the analysis of spatio-temporal data [4]. We were able to see a large variety in the data used for the research, including mobile phone location data [7, 10, 3], crime log data [2], bike trajectory data [5], and social network data [9, 4].

As a whole, this workshop had very fruitful discussions along with very interesting and cutting-edge talks from the presenters and the keynote speakers. We would like to thank the presenters and attendees of the workshop for making it a huge success, and also the organizing members of ACM SIGSPATIAL 2019 for giving us an opportunity to hold this workshop.

## 4 Awards

### 4.1 Best Paper Award

“Spatio-Temporal Clustering of Traffic Data with Deep Embedded Clustering” by Reza Asadi (University of California, Irvine) and Amelia Regan (University of California, Irvine)

### 4.2 Runner-up Award

“Grab-Posisi: An Extensive Real-Life GPS Trajectory Dataset in Southeast Asia” by Xiaocheng Huang (Grab-taxi Holdings), Yifang Yin (National University of Singapore), Simon Lim (Grabtaxi Holdings), Guanfeng Wang (Grabtaxi Research and Development Centre), Bo Hu (Grabtaxi Holdings), Jagannadan Varadarajan (Grabtaxi Holdings), Shaolin Zheng (Grabtaxi Holdings), Roger Zimmermann (National University of Singapore)

## References

- [1] R. Asadi and A. Regan. Spatio-temporal clustering of traffic data with deep embedded clustering. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 45–52. ACM, 2019.
- [2] F. C. Freire Nunes Junior, T. L. Coelho da Silva, J. F. de Queiroz Neto, J. Macedo, and W. Silva. A novel approach to approximate crime hotspots to the road network. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 53–61. ACM, 2019.
- [3] X. Huang, Y. Yin, S. Lim, G. Wang, B. Hu, J. Varadarajan, S. Zheng, and R. Zimmermann. Grab-posisi: An extensive real-life gps trajectory dataset in southeast asia. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 1–10. ACM, 2019.
- [4] H. Lim, S. Kim, and J. Heo. Graph analyses of phone-based origin-destination data for understanding urban human mobility in seoul, korea. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 62–65. ACM, 2019.

- [5] T. Mimura, S. Ishiguro, S. Kawasaki, and Y. Fukazawa. Bike-share demand prediction using attention based sequence to sequence and conditional variational autoencoder. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 41–44. ACM, 2019.
- [6] C. Schrekenberger, C. Bartelt, and H. Stuckenschmidt. Enhancing a crowd-based delivery network with mobility predictions. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 66–75. ACM, 2019.
- [7] C. Song, M. Ito, Y. Nishiyama, and K. Sezaki. Using mobile sensing technology for capturing people mobility information. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 33–40. ACM, 2019.
- [8] Y. Takimoto, Y. Tanaka, T. Kurashima, S. Yamamoto, M. Okawa, and H. Toda. Predicting traffic accidents with event recorder data. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 11–14. ACM, 2019.
- [9] D. To, D. Si, and Y. Chen. Predicting traveler next choice of activity with location-based social network data. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 15–23. ACM, 2019.
- [10] K. Tsubouchi, T. Saito, and M. Shimosaka. Context-based markov model toward spatio-temporal prediction with realistic dataset. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Prediction of Human Mobility*, pages 24–32. ACM, 2019.

# The Eighth ACM SIGSPATIAL International Workshop on Analysis for Big Spatial Data Chicago, IL, USA - November 5, 2019

Ashwin Shashidharan<sup>1</sup>, Varun Chandola<sup>2</sup>, Ranga Raju Vatsavai<sup>3</sup>

<sup>1</sup>Environmental Systems Research Institute, USA

<sup>2</sup>Department of Computer Science and Engineering, SUNY Buffalo, USA

<sup>3</sup>Department of Computer Science, North Carolina State University, USA

Since the “Big Data Research and Development Initiative” launched by the White House in 2012, big data has received great attention from industry and federal agencies alike emerging as an important area of research for scientists worldwide. Within the realm of big data, spatial and spatiotemporal data continues to be among the fastest-growing types of data. With advances in remote sensors, sensor networks, and the proliferation of location sensing devices in daily life activities and common business practices, the generation of disparate, dynamic, and geographically distributed spatiotemporal data has exploded in recent years. In addition, significant progress in the ground, air- and space-borne sensor technologies have led to unprecedented access to earth science data for scientists from different disciplines, interested in studying the complementary nature of different parameters. Today, analyzing this data poses a massive challenge to researchers.

The workshop series on Analytics for Big Geospatial Data (BIGSPATIAL), has become one of the key meeting points for researchers in the area of big geospatial data analytics, since 2012. Held every year, in conjunction with the annual ACM SIGSPATIAL conference, this meeting has found strong support from researchers in government, academia, and industry. The workshop provides a platform for researchers and practitioners engaged in addressing the big data aspect of spatial and spatiotemporal data analytics to present and discuss their ideas.

Building on the success of the previous editions to bring together researchers from academia, government and industry, who have been working in the area of spatial analytics with an eye towards massive data sizes, the 8<sup>th</sup> workshop on Analytics for Big Geospatial Data (BIGSPATIAL 2019) was held in conjunction with the 27<sup>th</sup> ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2019) on November 5<sup>th</sup>, 2019. The main motivation of the workshop as has been in previous years was to serve as a forum to exchange ideas, present recent research results and to facilitate collaboration and dialog between academia, government, and industrial stakeholders.

This year we received 8 technical submissions out of which 4 were selected for full presentations at the workshop. The technical program also consisted of a keynote talk by Dr. Dalton D. Lunga from Oak Ridge National Laboratory (ORNL), who is a lead scientist in machine learning-driven geospatial image analytics and a member of the ORNL AI Initiative. His talk titled, “Creating Global Scale Data Layers from Trillion Pixels using Machine Learning: A Journey of Challenges and Opportunities” provided an excellent start to the workshop by laying out the challenges and opportunities for big data researchers from the consideration of trillion pixel capable machine learning systems. The workshop was well-attended with 15 registered participants. Besides the technical and invited talks, two awards were presented at this year’s workshop. This included the best paper award and the best presentation award which was given to Alberto Belussi, Damiano Carra, Sara Migliorini and Mauro Negri for their paper on “Efficient MapReduce Computation of Topological Relations for Big Geometries”.

We would like to thank all the speakers, authors and attendees who participated in the workshop. We express our sincere gratitude to Dr. Dalton D. Lunga for his insightful keynote talk. Also, a special note of thanks to the program committee members, whose reviewing efforts ensured in selecting a competitive and strong technical program. We hope the BIGSPATIAL workshop series will continue to provide a leading international forum for researchers, developers, and practitioners in the field of big geospatial data analytics to identify current and future areas of research.



# **The SIGSPATIAL Special**

## **Section 3: Event Announcement**

---

**ACM SIGSPATIAL**  
**<http://www.sigspatial.org>**

# GeoRich'2020: Call for Submissions

## The Sixth International ACM Workshop on Enriched Geo-Spatial Data

Panagiotis Bouros<sup>1</sup>, Amr Magdy<sup>2</sup>, Matthias Renz<sup>3</sup>, Mohamed Sarwat<sup>4</sup>, Andreas Züfle<sup>5</sup>

<sup>1</sup>Johannes Gutenberg University Mainz, Germany, <sup>2</sup>University of California, Riverside, USA

<sup>3</sup>University of Kiel, Germany, <sup>4</sup> Arizona State University, USA, <sup>5</sup> George Mason University

bouros@uni-mainz.de, amr@cs.ucr.edu, mr@informatik.uni-kiel.de,

msarwat@asu.edu, azufle@gmu.edu

### 1 The Vision of GeoRich: Reaching the Database Community

Spatial and spatio-temporal data management has traditionally been a core topic of the database community. Until 2018, database conferences such as the ACM International Conference on Management of Data (SIGMOD), as well as the International Conference on Very Large Data Bases (VLDB) always had at least one research session on spatial data:

- SIGMOD 2016 (One Sessions): Session 11 (Spatio/Temporal Databases),
- SIGMOD 2017 (Two Sessions): Session 24 & 29 (Spatial and Multimedia Data (1 & 2)),
- SIGMOD 2018 (One Session): Research Session 8 (Spatial Data & Streams)
- VLDB 2016 (Two Sessions): Research 11 & 20 (Spatial Data and Queries 1 & 2),
- VLDB 2017 (Three Sessions): Spatial Data Management 1-3 (No Session Numbers)
- VLDB 2018 (One Session): Research Session: Spatial Data

Yet, in the last year, we have seen a rapid decline in interest of the database community in topics related to spatial data. This is evident by the number of research sessions at these conferences in 2019:

- SIGMOD 2019(No Session):  $\emptyset$ ,
- VLDB 2019 (One Sessions): Research Session 25: Spatial Data

We see that the number of spatial data related research sessions across SIGMOD and VLDB has declined from five sessions in 2017 to only one session in 2019. While this decline is due to the proliferation of SIGSPATIAL as the flagship spatial data conference, this trend isolates us from the database community. This trend reduces the visibility of our research to the SIGMOD community and lowers exposure of our work. It also disrupts valuable cross-disciplinary research between the SIGMOD and the SIGSPATIAL community, by no longer bringing these communities together. To stop this trend, we decided to reboot the SIGMOD@GeoRich workshop on Enriched Geo-Spatial Data, which has been organized as a SIGMOD satellite workshop from 2014-2018, but which was not organized last year in 2019. The goal of this workshop is to bring together and provide a forum for the SIGSPATIAL and the SIGMOD community to discuss topics relevant to spatial data management.

**This is a call for the SIGSPATIAL Community to continue outreach and exposure to the database community. Please consider submitting your work to GeoRich'2020!**

## 2 GeoRich'2020: Call for Papers

The aim of this workshop is to provide a unique forum for discussing in depth the challenges, opportunities, novel techniques and applications on modeling, managing, searching and mining rich geo-spatial data, in order to fuel scientific research on big spatial data applications beyond the current research frontiers. The workshop is intended to bring together researchers from different fields of databases, data-science and geo-science that deal with the management of spatial and spatio-temporal data, social network data, textual data, multimedia data, semantic data and ontologies, uncertain data and other common types of geo-referenced data. In particular, submissions covering topics from the following non-exclusive list in the context of enriched geo-spatial data are encouraged: The topics of interest are related to semantically enriched geo data, such as spatio-textual, spatio-temporal, spatio-social, geo-social network, mobile and wireless data and uncertain data are not limited to:

- Big Spatial Data
- Data-mining and machine learning on enriched geo data
- Data stream management and mobile computing
- Exploratory geospatial analysis
- Geoinformation Systems (GIS)
- Geospatial data fusion
- Leveraging spatial data for smart cities
- Privacy and confidentiality
- Machine learning and artificial intelligence using spatial data
- Managing uncertainty in spatial data
- Managing crowd-sourced spatial data
- Spatial and spatio-temporal data management
- Spatial crowdsourcing
- Spatial econometrics
- Spatial data models and representation
- Spatial query processing and optimization
- Spatial reasoning and analysis
- Spatial recommendation
- Spatial statistical analysis
- Urban analytics and smart transportation

The workshop will be organized in a manner that fosters interaction and exchange of ideas among the participants. Besides paper presentations, time will be allocated to open discussion forums, poster presentations, informal discussions or panels. In addition to regular papers, vision or work-in-progress papers that have the potential to stimulate debate on existing solutions or open challenges are especially encouraged. Proposals for panels on newly-emerging or controversial topics are also especially welcome.

## 3 Important Dates & Submission Instructions

Abstract Registration:	Sunday, March 15th, 2020 (midnight AoE)
Paper Submission:	Sunday, March 22th, 2020 (midnight AoE)
Notification of acceptance:	Sunday, April 26th, 2020 (midnight AoE)
Camera-ready version due:	Sunday, May 9th, 2020 (midnight AoE)
Workshop date:	Sunday, June 14th, 2020 (midnight AoE)

Authors are invited to submit original, unpublished research papers that are not being considered for publication in any other forum. Manuscripts should be submitted electronically as PDF files and be formatted using the ACM templates. Papers cannot exceed six pages in length including all references and appendices. All papers must be submitted through the submission website hosted on EasyChair.

Accepted papers will be included in the informal proceedings to be distributed at the workshop, and will be available on the workshop's Web site. Additionally, the accepted papers will be published online in the ACM digital library. The papers must include the standard ACM copyright notice on the first page. Additional details and instructions can be found on the conference website at:

<https://georich2020.github.io/>



# join today!

# SIGSPATIAL & ACM

[www.sigspatial.org](http://www.sigspatial.org)

[www.acm.org](http://www.acm.org)

The **ACM Special Interest Group on Spatial Information (SIGSPATIAL)** addresses issues related to the acquisition, management, and processing of spatially-related information with a focus on algorithmic, geometric, and visual considerations. The scope includes, but is not limited to, geographic information systems (GIS).

The **Association for Computing Machinery (ACM)** is an educational and scientific computing society which works to advance computing as a science and a profession. Benefits include subscriptions to *Communications of the ACM*, *MemberNet*, *TechNews* and *CareerNews*, full and unlimited access to online courses and books, discounts on conferences and the option to subscribe to the ACM Digital Library.

- ☐ SIGSPATIAL (ACM Member) ..... \$ 15
- ☐ SIGSPATIAL (ACM Student Member & Non-ACM Student Member) ..... \$ 6
- ☐ SIGSPATIAL (Non-ACM Member) ..... \$ 15
- ☐ ACM Professional Membership (\$99) & SIGSPATIAL (\$15) ..... \$114
- ☐ ACM Professional Membership (\$99) & SIGSPATIAL (\$15) & ACM Digital Library (\$99) ..... \$213
- ☐ ACM Student Membership (\$19) & SIGSPATIAL (\$6) ..... \$ 25

## payment information

Name \_\_\_\_\_

ACM Member # \_\_\_\_\_

Mailing Address \_\_\_\_\_

\_\_\_\_\_

City/State/Province \_\_\_\_\_

ZIP/Postal Code/Country \_\_\_\_\_

Email \_\_\_\_\_

Mobile Phone \_\_\_\_\_

Fax \_\_\_\_\_

Credit Card Type: ☐ AMEX ☐ VISA ☐ MC

Credit Card # \_\_\_\_\_

Exp. Date \_\_\_\_\_

Signature \_\_\_\_\_

Make check or money order payable to ACM, Inc

ACM accepts U.S. dollars or equivalent in foreign currency. Prices include surface delivery charge. Expedited Air Service, which is a partial air freight delivery service, is available outside North America. Contact ACM for more information.

### Mailing List Restriction

ACM occasionally makes its mailing list available to computer-related organizations, educational institutions and sister societies. All email addresses remain strictly confidential. Check one of the following if you wish to restrict the use of your name:

- ☐ ACM announcements only
- ☐ ACM and other sister society announcements
- ☐ ACM subscription and renewal notices only

### Questions? Contact:

ACM Headquarters  
2 Penn Plaza, Suite 701  
New York, NY 10121-0701  
voice: 212-626-0500  
fax: 212-944-1318  
email: [acmhelp@acm.org](mailto:acmhelp@acm.org)

### Remit to:

ACM  
General Post Office  
P.O. Box 30777  
New York, NY 10087-0777

SIGAPP



Association for  
Computing Machinery

[www.acm.org/joinsigs](http://www.acm.org/joinsigs)

Advancing Computing as a Science & Profession



# **The SIGSPATIAL Special**

---

**ACM SIGSPATIAL**  
**<http://www.sigspatial.org>**