Data-driven Human Mobility Modeling: A Survey and Engineering Guidance for Mobile Networking

ANDREA HESS, Trinity College Dublin, Ireland
KARIN ANNA HUMMEL, Johannes Kepler University Linz, Austria
WILFRIED N. GANSTERER, University of Vienna, Austria
GÜNTER HARING, University of Vienna, Austria

Over the last decades, modeling of user mobility has become increasingly important in mobile networking research and development. This has led to the adoption of modeling techniques from other disciplines such as kinetic theory or urban planning. Yet these techniques generate movement behavior that is often perceived as not "realistic" for humans or provides only a macroscopic view on mobility. More recent approaches infer mobility models from real traces provided by positioning technologies or by the marks the mobile users leave in the wireless network. However, there is no common framework for assessing and comparing mobility models.

In an attempt to provide a solid foundation for realistic mobility modeling in mobile networking research, we take an engineering approach and thoroughly discuss the required steps of model creation and validation. In this context, we survey how and to what extent existing mobility modeling approaches implement the proposed steps. This also summarizes helpful information for readers who do not want to develop a new model, but rather intend to choose among existing ones.

Categories and Subject Descriptors: A.1 [**GENERAL LITERATURE**]: Introductory and Survey; C.2.1 [**COMPUTER-COMMUNICATION NETWORKS**]: Wireless Communication; I.6.4 [**SIMULATION AND MODELING**]: Model Validation and Analysis

General Terms: Design, Experimentation, Measurement, Performance

Additional Key Words and Phrases: Mobility Modeling, Representativeness, Realistic Models

1. INTRODUCTION

Mobile networks have changed human life considerably by enabling ubiquitous connectivity and communication for users that are physically mobile. Inherently, wireless network research and development rely on the characterization of the mobility of network users, in other words, on *mobility models*. Mobility characteristics influence the design of mobility management in infrastructure networks such as a cellular network or a Wi-Fi hotspot area (handover or hand-off mechanisms for mobile devices). In turn, handover and association

This work was partially supported by an ERCIM Alain Bensoussan Fellowship, co-funded by the European Union Seventh Framework Programme (EU FP7) under grant 246016, by the Science Foundation Ireland (SFI) under grant 10/IN.1/I1007, and by the Austrian Science Fund (FWF): project S10608. Part of the work was carried out while K.A. Hummel was at ETH Zurich, supported by the EC FP7 Marie Curie program (contract PIEF-GA-2010-276336 MOVE-R).
Authors' address: A. Hess is with the CONNECT Centre for Future Networks, Trinity College Dublin, Dunlop Oriel House, Dublin 2, Ireland; email: hessan@tcd.ie; K.A. Hummel is with the Department of Telecooperation, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria; email: karin_anna.hummel@jku.at; W.N. Gansterer and G. Haring are with the Faculty of Computer Science, University of Vienna, Währinger Straße 29, 1090 Vienna, Austria; email: {wilfried.gansterer, guenter.haring}@univie.ac.at.

DOI:http://dx.doi.org/10.1145/0000000.0000000

events are the marks which moving devices leave in the network, cf. [Janecek et al. 2012]. Infrastructure-less wireless networks, namely mobile ad-hoc and opportunistic networks, even exploit mobility in the sense that network links are established when mobile devices come within transmission range of one another [Conti et al. 2010; Grossglauser and Tse 2002].

Considering the significant impact of the chosen mobility model on the performance of mobile systems under investigation [Camp et al. 2002; Kunz et al. 2001; Newport et al. 2007; Yoon et al. 2003], it becomes clear that a solid foundation is needed for developing valid mobility models.

1.1. Approaches to mobility modeling

Mobility models can be created without the use of observation, based only on assumptions about certain properties of movement, such as velocity or changes in direction. Traditionally such models are referred to as *synthetic models*, which often have only limited agreement with mobility behavior in the real world. A representative of this group is the random waypoint mobility model [Broch et al. 1998; Johnson and Maltz 1996] that describes the movement of individual nodes in straight lines with pauses. Speed, waypoints, and pause time of nodes are chosen independently at random[1]. Shortcomings of the random waypoint model are, e.g., lack of compliance with city topographies or stochastic artifacts, namely, that the stationary spatial distribution of the nodes resulting from random waypoint mobility models is non-uniform [Bettstetter et al. 2003], and that the average node speed does not reach a steady state (but consistently decreases over time) if the smallest possible node speed is chosen as zero [Yoon et al. 2003].

The growing availability of diverse and large-scale mobility traces is an important enabler for *trace-based* (*data-driven*) models. Mobility data are available in infrastructure and ad-hoc mobile networks, such as marks of handovers in the cellular network or Bluetooth-based encounters of mobile devices. Other data originate from position and movement tracking campaigns, e.g., generated by GPS (Global Positioning System), activity tracking, or open government initiatives. Real anonymized data are made available by network operators during initiatives such as the cellular data collected in Ivory Coast [Blondel et al. 2012] and Senegal [de Montjoye et al. 2014] in the context of two Orange D4D challenges, by governmental census data, transport timetables, etc.[2], or by fleet providers who make taxi traces publicly available, such as the Shanghai traces[3] or the San Francisco traces[4]. Traces can be simply replayed or stochastic mobility models may be derived from the observed mobility characteristics. One shortcoming of trace-based models is that traces are collected in specific environments such as a campus, a theme park, or a conference. As a consequence, their general applicability is limited.

Together with the advent of trace-based models, the interest in "realistic" human mobility models, i.e., models *"representing things in a way that is accurate and true to life"* [Soanes and Stevenson 2005], has grown in general. Although no commonly accepted definition of a realistic mobility model exists, this term is widely used intuitively in the mobility modeling literature, cf. [Jardosh et al. 2003; Kim et al. 2009; Munjal et al. 2011; Schwamborn et al. 2010; Treurniet 2014; Vogt et al. 2012; Yoon et al. 2006]. All these articles address the common aim of developing mobility models that capture real-world human mobility more accurately than early synthetic models by including knowledge about the context of movement, such as the topography of an area. Questions arise with respect to the spatial and temporal granularity of movement as provided by the real-world observations, relations

[1]Standard implementations of the random waypoint mobility model use uniform distributions.
[2]An overview of available data is given by the Open Knowledge Foundation: http://census.okfn.org/.
[3]WnSN lab, Shanghai Jiao Tong University: http://wirelesslab.sjtu.edu.cn/.
[4]Cabspotting project of San Francisco Exploratorium: http://cabspotting.org/.

to be captured such as social behavior, etc. Realistic models are typically very sophisticated and there seems to be a need for guidelines on how to develop such models.

While our focus is on human mobility models in the context of mobile networking research, we would like to mention that human mobility modeling is also considered in other fields, such as transport modeling [Hensher and Button 2008] or time geography [Miller 2008]. Our findings may also be of interest in these fields, but a detailed discussion is beyond the scope of this survey.

1.2. Objectives of this survey

Existing related surveys primarily focus on describing and classifying mobility models. Our main objective is to go beyond that by also providing guidelines for researchers on how to develop and assess realistic data-driven human mobility models. For this purpose, we first discuss the foundations of mobility modeling by applying general modeling theory. On this basis, we structure the process of data-driven mobility modeling into several pivotal steps from model creation to model evaluation and validation. This engineering approach allows for a step-by-step quality control of the resulting mobility model.

We connect these guidelines for model engineering to the state-of-the-art in mobility modeling and illustrate how existing modeling approaches realize important modeling steps. Therefore, we survey selected representatives of the most important mobility model classes from the literature. In this part, we put more emphasis on investigating how the proposed model engineering process is reflected in mobility modeling research than on comparing and rating existing mobility models as a whole. Nevertheless, the stepwise discussion may be leveraged to select an appropriate model as well. In particular, for each model engineering step we summarize whether it is discussed explicitly in representative mobility models, which different techniques for implementing it have been investigated, and whether any shortcomings can be identified in existing models. In contrast to earlier surveys, we put a particular emphasis on the *validation* of a mobility model.

In Section 2, we start with summarizing existing related survey literature. We then define basic terminology for human mobility modeling and review the underlying modeling concepts in Section 3. Based on these foundations, we introduce the engineering approach and detail the different steps required for creating a realistic and representative mobility model. We consider general approaches for modeling human mobility irrespective of the specific movement type or application. In particular, pedestrian and vehicular movement are both covered. In Section 4, we detail the considerations to be made during the planning phase, the types of input data used for model building, and the modeling restrictions they impose. In Section 5, we summarize common techniques applied for building a mobility model and context elements commonly covered by state-of-the-art realistic models. Finally, in Section 6 we discuss the validation step and how it is implemented in existing mobility models.

2. RELATED SURVEYS

In the past, several surveys have been presented that give pointers to important aspects of mobility models in the domain of networking. These surveys represent the state-of-the art at the time of their compilation.

In the survey of Camp et al. [2002], which represents the view of more than a decade ago, the authors review and discuss several synthetic mobility models. Major findings are that the (simulated) performance properties of ad-hoc network protocols such as packet delivery ratio, end-to-end delay, average hop count, and protocol overhead strongly depend on the mobility model used in the simulation. Realism of synthetic traces is recognized as an important issue, however, no clear terminology or comprehensive guidelines to follow are provided. The discussion is focused on speed and direction of mobile nodes, on the avoidance of unrealistic movement such as sudden stops or sharp turns, and on restrictions employed by the mobility model such as roads. For entity mobility, the authors recommend the use of the random waypoint mobility model, or alternatives such as the random walk mobility model or the Gauss-Markov mobility model as these models allow for investigating general statistical properties of mobile network protocols. However, concerns are stated with respect to clustering properties of the random waypoint mobility model or the strict straight movement between waypoints of the random walk model. As mobility models exhibit different strengths and weaknesses, this survey further recommends to expose a network protocol either to a model combining the strengths of the discussed models, or to multiple mobility models in series.

As more real data have become available, more recent surveys focus on analyzing trace data and describing their use. Aschenbruck et al. [2011] give an overview of publicly available trace data sets and the dependency of node movement on several aspects such as past movement, other nodes, and geographic restrictions. Moreover, several trace generating tools are described. The main challenges for trace-based analysis are identified as time-based variations, data filtering, limited generalization, the required amount of data, and that the majority of models is not validated by any measure.

A more novel aspect in mobility modeling is the inclusion of social relations as discussed by Musolesi and Mascolo [2009]. Besides providing a survey of mobility models that include hybrid synthetic and trace-based models, a distinctive feature of this survey is the inclusion of models that utilize information from social networks. These models are based on the tendency of humans to form communities and make in particular use of social ties that are extractable, e.g., from online social networks. To capture social ties, a connectivity graph can be constructed starting from an interaction graph that quantifies the number and duration of node interactions. Based on the connectivity graph, a mobility model is derived expressing the likelihood of nodes joining at other nodes' locations. In a related survey including social relationships, Karamshuk et al. [2011] classify the nature of human movements along the spatial, temporal, and social dimension. In order not to be limited by a specific data set in trace-based models, the authors focus on synthetic models, which aim to reproduce driving forces for individual mobility such as social attitude, location preferences, and regular schedules. To cover also perceivable regularity in a person's movement the authors present the concept of human mobility patterns. By including the scale of mobility (building-wide, city-wide, or world-wide), hierarchical mobility modeling is introduced, which is used to classify existing models along the levels. An important but open question remains concerning the understanding of the correlations between the different statistical properties of human movement. For example, the authors consider heavy-tailed inter-contact times, which seem to emerge in the majority of models and could thus be seen as a common related side-effect.

The special case of vehicular mobility is addressed by the survey in [Härri et al. 2009], which summarizes models used in vehicular ad-hoc network (VANET) simulators. Important modeling aspects are the road topology that is created, the intersection policy, multi-lane structures, or speed limitations. The VANET models covered by the survey use either a car following mobility model, defining at least acceleration, deceleration, and reaction time in relation to other vehicles, or a simple model with uniform velocity on the generated paths. This work is one of the rare cases in which model validation is addressed by comparing the generated traces either against observed mobility traces or against traces originating from an already validated model. Concerning evaluation of realism, Fiore et al. [2007] discuss traffic simulators along with minimum requirements for highway or urban scenarios based on observed acceleration patterns of different car following models.

In the most recent survey [Treurniet 2014], a taxonomy is provided to classify human mobility models along important modeling aspects, that are, motion determination (e.g., random speed, collision avoidance, etc.), path and target determination (e.g., based on a random trip), involved group dynamics, and basic characteristics, such as pause time characteristics or spatial constraints. Yet this list should be considered as an open list, and progress in the field might require extending this taxonomy.

In addition to the above mentioned surveys, two books have been published on the topic of mobility models in recent years. In [Roy 2011], a comprehensive collection of major basic mobility models is presented. This book focuses on analytical model characterization for generic use. Differently, in [Santi 2012], mobility models are discussed with respect to their application in mobile network research. Models are presented for WLAN, wireless mesh, cellular, vehicular, and opportunistic networks. This application-oriented book allows for easy identification of candidate mobility models in case the networking context is known.

While the existing surveys and books classify and describe mobility models from several perspectives, they all approach mobility modeling "as a whole". Complementary to these summaries, we introduce a novel engineering approach that zooms into the different steps of mobility modeling and survey how existing approaches implement each step separately. We further find that most of the existing surveys already use a notion of realism which is similar to ours, yet often implicitly includes representativeness. In order to clearly distinguish these two aspects, we differentiate methodologically between building a *representative* baseline mobility model and – on top of this – further improving the *realism* of the model. Finally, we also discuss validation aspects in detail, which so far have not received sufficient attention in the existing literature.

3. THE PROCESS OF HUMAN MOBILITY MODELING

Based on foundations from general modeling theory, we develop a sound concept for mobility modeling. Moreover, we define basic terms and finally propose a generally applicable process for mobility modeling, which has been inspired by the process of workload modeling (cf. [Ferrari et al. 1983]).

3.1. Foundations from general modeling theory

From a theoretical modeling perspective (see, e.g., [Stachowiak 1973]), any scientific model has the following characteristics:

(1) A scientific model is a *representation* of an original (natural or artificial) system. Note, that the original system itself can be a model.
(2) A scientific model is not defined uniquely by the original system. Since the model represents the original system for a specific *purpose* in a specific *context*, the modeling process and its result are influenced by the objective of the investigation and by its context. From a very general perspective, this might even include influence on the modeling process by when the model is created, who creates it, and by whom it is used.
(3) A scientific model is always a *simplified* representation of the original system. This simplification (or reduction) of reality or another model leads to a loss of information. Without any simplification, it would not be a model, but the original system itself. Thus, given the underlying objective and the context of the modeling process, the model does not represent all properties/features of the original system, but only that subset of the properties that is considered relevant.
(4) A scientific model has to be *valid* or *representative*, i. e., despite the reduction of the original system, the model has to agree well (enough) with the original system with respect to a predefined set of relevant criteria. This set of criteria is influenced by the objective and context of the modeling process. If the model turns out to be invalid or not representative in this validation process, then it results from a "wrong" reduction of the original system and may lead to wrong conclusions about the original system. In this case the model has to be changed, i.e., it has to be calibrated or the modeling process has to be reformulated. Once a model has been validated, it is the basis for

further investigations (baseline model). These investigations may include adaptations and modifications in order to simulate modifications of the original system.

In the following, we apply this theory to mobility modeling.

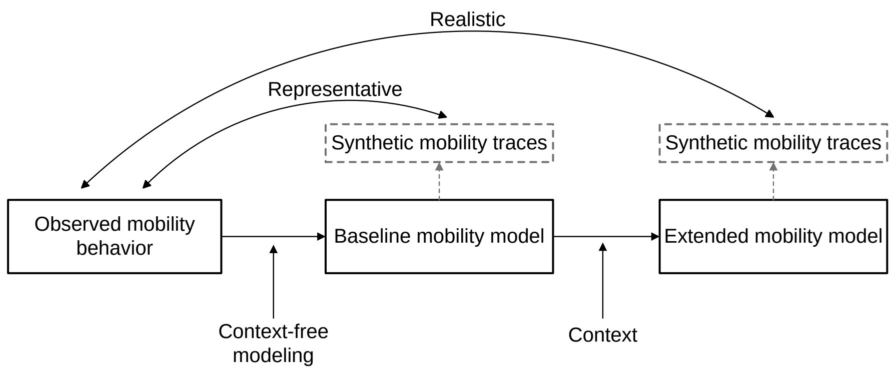

Fig. 1. Schematic view of the mobility modeling process.

3.2. Principles of mobility modeling

A *mobility model* is a simplified representation of the movement of single or groups of mobile entities in a given context, primarily the spatial environment, for a specific purpose during a well defined period of time. The model may consider only one class of mobile entities, meaning that all entities are statistically behaving in the same way, or it may consider several classes of mobile entities (multiclass model). Mobile entities may be people, vehicles, robots, or animals. In this article, we focus on *human* mobility models due to their practical importance in mobile networked systems.

A schematic view of realistic mobility modeling is depicted in Figure 1. The first stage towards a realistic mobility model is to observe human mobility in reality, which results in mobility data such as mobility traces. Mobility traces usually contain at least location information along a time line such as time-stamped geo-location data provided by GPS. From the *observed mobility behavior* a context-free *baseline mobility model* is derived. This model can be used for generating synthetic mobility traces used in a simulation study. For example, changes in direction, velocity, and mobility range of nodes can be described by the statistical properties derived from the observation. The baseline mobility model has to be representative for the observed mobility behavior. This means that the synthetic mobility traces generated by a *representative mobility model* have to agree well enough with a set of observed real-world traces along relevant validation characteristics.

A representative baseline model does not contain explicit information about the context in which the real-world traces were observed. This context may include the purpose of movement (e.g., traveling to workplace or school), geographical structures and limitations (streets, walls, houses, etc.), or e.g., in case of taxi traces, the status of the taxi (occupied, available, on its way to a customer). To introduce such context, a second model is derived, which we call *extended mobility model*. This new model is intended to be more realistic than the baseline model. A *realistic human mobility model* is thus an accurate representation of real-world movement behavior of (groups of) humans with respect to a set of relevant characteristics. Including more context information about, e.g., roads, traffic lights, social relations, etc. allows for creating a more realistic model, yet at the cost of simplicity.

The reasons for introducing this two-stage approach are as follows. Realism may or may not be an issue for a mobility model, while representativeness always has to be assured. Further, introducing context constraints is fundamentally different from the statistical analysis of traces conducted for creation of the baseline mobility model, as it requires semantic knowledge about the context of movement.

3.3. Engineering a mobility model

When creating a *representative* and *realistic* mobility model, several steps have to be taken. To provide comprehensive guidance for those steps, we structure the process of realistic mobility modeling as depicted in Figure 2. The process starts with clarifying the objective and deciding which fundamental characteristics the model should have including generic assumptions and requirements, and determining which observations are needed. In the following modeling steps, mobility is characterized, the baseline model is created, and extended by inclusion of context information. After calibration, the model (with or without extension by context considerations) has to be validated. We detail each step in one of the following sections and survey how existing mobility models implement each step.

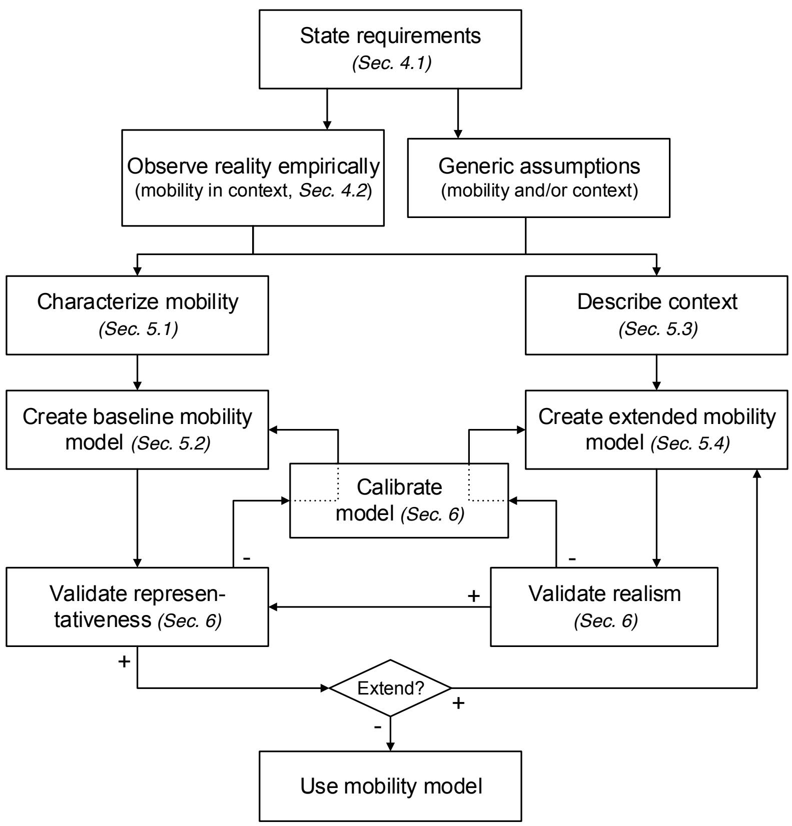

Fig. 2. Engineering a mobility model: steps of the modeling process and dependencies (positive/negative outcomes of a step are indicated by '+' and '–').

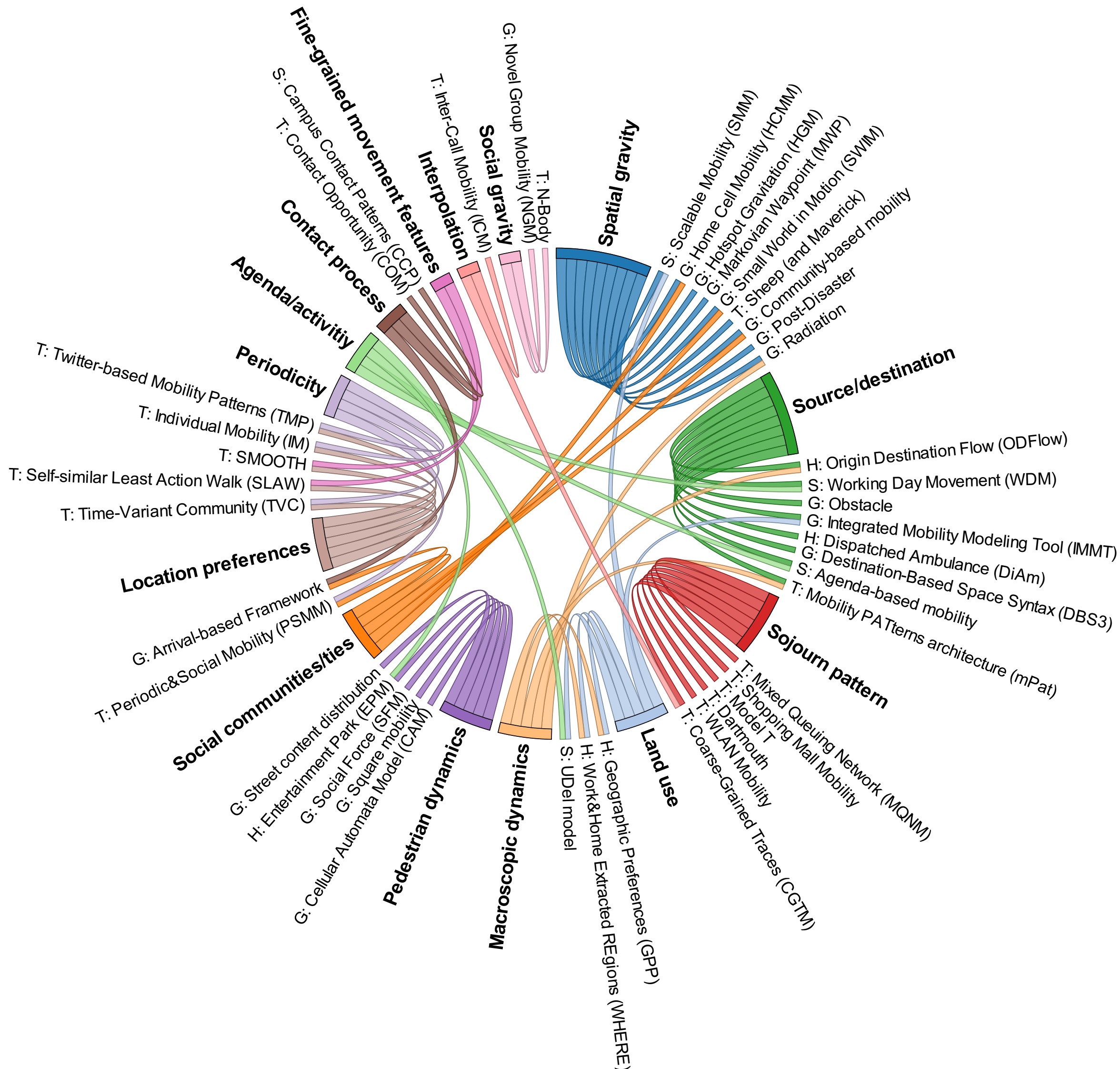

Fig. 3. Overview of existing mobility models reviewed and grouped according to major features. The chords in the diagram connect each mobility model to one or more of its major features (bold). The arc length of each feature corresponds to the number of models to which this feature has been assigned. The letters G, S, T, and H relate to the assigned group, i.e., generic, survey data based, trace data based, and hybrid models, respectively.

3.4. Overview of surveyed mobility models

We include major data-driven mobility models in our survey based on a few criteria, namely, their timeliness and importance for recent mobile networking research (cf. Table IX), as well as the innovative aspects introduced by the models or the model creation process. Here, our focus is set on disclosing relevant methods provided by the models.

There are multiple options to present an overview of the surveyed models by grouping them along distinct features, cf. [Musolesi and Mascolo 2009; Karamshuk et al. 2011;

Treurniet 2014]. A fundamental criterion targets the *foundations* of the model and results in four classes: (*i*) models that are not based on any observations ("generic models"), (*ii*) models based on mobility traces, (*iii*) models based on survey data, and (*iv*) models based on both traces and surveys ("hybrid models"). Traditional synthetic models fall into the category of generic models. As in the literature the term synthetic model also very generally refers to a "non trace-based" model, we introduce the new term "generic model" which allows us to distinguish synthetic from survey-based and hybrid models. We make use of this foundation-driven classification in the remainder of the survey.

In addition, we provide an orthogonal, *feature-driven* classification by identifying major model characteristics originating from the existing mobility modeling literature. This classification provides an introduction to the selected mobility models and allows for finding models based on features. Figure 3 visualizes the outcome of this classification[5] and Table I lists all models together with the corresponding references. The importance of a feature can be visually deduced from the corresponding arc length in Figure 3, e.g., spatial gravity, source/destination, and sojourn pattern are the most frequently implemented features reflecting the spatial and temporal aspects obviously targeted in many of the surveyed mobility models. The features themselves may be grouped into the following classical categories:

Spatial features. **Spatial gravity** stands for models describing attraction (and repulsion) forces of single locations or regions determining the path. An example is the Sheep (and Maverick) model [Morlot et al. 2010] which reproduces the formation of highly populated zones during city-wide events. **Source/destination** models preselect start and end location of a path, based on land use information or similar, and apply standard algorithms, such as shortest path, to route entities between source and destination. Whereas source/destination models actually define the path of each single moving entity from start to end location, spatial gravity models consider multiple attraction and repulsion forces that influence the movement along a path. For example, WHERE (Work and Home Extracted REgions) [Isaacman et al. 2012] models the commuter mobility between residential and commercial areas of a city. Models given the feature **land use** subdivide the analyzed geographical area according to land use surveys reflecting, e.g., spatial constraints. The model Geographic Preferences Prediction (GPP) [Calabrese et al. 2010] derives the movement choices of individuals from the type of geographical areas that are of interest for the collectivity at a give time and are available in a certain travel distance. **Location preferences** are a key modeling feature to determine location targets, e.g., of daily life activities. The Individual-Mobility (IM) model [Song et al. 2010], e.g., realizes location preferences by incorporating mechanisms for exploring new locations and preferentially returning to previously visited locations.

Temporal features. The aspect **sojourn pattern** fits to models focusing on the times people stay in buildings/places (cf. pause times), within reach of network elements, etc., with little or no consideration of the movement between the sojourns. **Periodicity** is a further distinct characteristic addressed by a number of models. The Mixed Queueing Network Model of Mobility (MQNM) [Chen et al. 2012] describes the arrival and sojourns of users at access points of a campus WLAN, each represented by a server queue. The Periodic & Social Mobility Model (PSMM) [Cho et al. 2011] combines the observation that 50% to 70% of movements can be explained by periodic mobility behavior within bounded regions with effects of social relationships causing occasional long distance travels.

[5]Note that for synthetic, generic models known as "random" or "stochastic" models, the used features characterize the add-on (e.g., spatial gravity emanates from a point of interest or pedestrian interaction patterns).

Table I. Overview of reviewed mobility models: mapping of acronyms, names, and references.

Acronym	Full name and reference	Acronym	Full name and reference
Agenda	Agenda-based mobility model [Zheng et al. 2006]	Arrival	Arrival-based Framework [Karamshuk et al. 2012]
CAM	Cellular Automata Model [Blue and Adler 2001]	CCP	Campus Contact Patterns [Srinivasan et al. 2006]
CGTM	Coarse-Grained Traces model [Yoon et al. 2006]	COM	Contact Opportunity Model [Chaintreau et al. 2007]
Community	Community-based mobility model [Musolesi and Mascolo 2006]	Dartmouth	Dartmouth model [Kim et al. 2006]
DBS3	Destination-Based Space Syntax [Vogt et al. 2012]	DiAm	Dispatched Ambulance model [Schwamborn et al. 2010]
Disaster	Disaster mobility model [Nelson et al. 2007]	EPM	Entertainment Park Model [Vukadinović et al. 2014]
GPP	Geographic Preferences Prediction [Calabrese et al. 2010]	HCMM	Home Cell Mobility Model [Boldrini et al. 2009]
HGM	Hotspot Gravitation Model [Du et al. 2012]	ICM	Inter-Call Mobility Model [Ficek and Kencl 2012]
IM	Individual Mobility model [Song et al. 2010]	IMMT	Integrated Mobility Modeling Tool [Markoulidakis et al. 1997]
MA	Movement Activity Model [Hummel and Hess 2013]	Mall	Shopping Mall Mobility Model [Galati et al. 2013]
ModelT	Model T mobility model [Jain et al. 2005]	mPat	Mobility PATterns interference architecture [Zhang et al. 2014]
MQNM	Mixed Queuing Network Model [Chen et al. 2012]	MWP	Markovian Waypoint Model [Hyytiä et al. 2006]
NBody	N-Body mobility model [Zhao and Sichitiu 2010]	NGM	Novel Group Mobility model [Rossi et al. 2005]
Obstacle	Obstacle model [Jardosh et al. 2003]	ODFlow	Origin Destination Flow [Calabrese et al. 2011]
PSMM	Periodic & Social Mobility Model [Cho et al. 2011]	Radiation	Radiation model [Simini et al. 2012]
SFM	Social Force Model [Helbing and Molnár 1995]	Sheep	Sheep (and Maverick) Model [Morlot et al. 2010]
SLAW	Self-similar Least Action Walk [Lee et al. 2009]	SMM	Scalable Mobility Model [Basgeet et al. 2003]
SMOOTH	SMOOTH mobility model [Munjal et al. 2011]	Square	Square mobility model [Desta et al. 2013]
Street	Street content distribution model [Vukadinović et al. 2009]	SWIM	Small World in Motion model [Kosta et al. 2010]
TMP	Twitter-based Mobility Patterns [Jurdak et al. 2015]	TVC	Time-Variant Community Model [Hsu et al. 2009]
UDel	"University of Delaware" simulator [Kim et al. 2009]	WDM	Working Day Movement Model [Ekman et al. 2008]
WHERE	Work & Home Extracted REgions [Isaacman et al. 2012]	WLANM	WLAN Mobility model [Tuduce and Gross 2005]

Group and social features. "Social models" build upon either community structures or ties between single persons. The movement is governed by spatial gravity (location of assigned communities) or social gravity (current location of socially close people). Models associated with the features **social communities/ties** and **social gravity** form the group of social models featuring group dynamics. A more general characteristic that expresses ties between individuals are contacts, thus, the **contact process** is a key characteristic described in some mobility models. Group dynamics are, e.g., modeled by the Novel Group mobility model

(NGM) [Rossi et al. 2005] which considers relationships between the mobility patterns of group leaders and followers. The campus contact pattern model (CPP) [Srinivasan et al. 2006] describes the temporal meeting characteristics of students who attend the same classes on a university campus.

Individual user features. Individual human movement is targeted mainly by models expressing **pedestrian dynamics**, which characterize human (pedestrian) behavior on relatively small areas, e.g., a street or a city square. Similarly, vehicular driver models comprise individual characteristics such as vehicle acceleration or overtaking behavior. In these models the major dynamics result from the dependencies among the vehicles, such as adapting the speed to the speed of the vehicle in front and the overall traffic flow. Other user-centric models are focused on an individual's **agenda/activity**. An example pedestrian dynamics model is the Social Force Model (SFM) [Helbing and Molnár 1995] reproducing typical movement behavior of people in crowded environments, including lane formation with people walking in the same direction or oscillatory changes of the walking direction at narrow passages. In the UDel model [Kim et al. 2009] the movement is governed by an underlying agenda of daily activities (working, at home, exercise, eating out, etc.) which are assigned to certain locations representing start- and endpoint of every trip.

Granularity. **Macroscopic dynamics** express the model's aim not to model individual mobility properties, but coarse-grained mobility, such as flows in and out of a city. For example, the macroscopic model Radiation [Simini et al. 2012] can be applied to model migration flows as well as mobile calling patterns between states of a country. In contrast, **fine-grained movement features** express the focus on detailed mobility characterization – such as extracted from fine-grained location measurements by the SLAW model [Lee et al. 2009]. The aspect **interpolation** represents models that derive fine-grained movement information by enriching coarse-grained data with additional knowledge and assumptions. Example are the models Coarse-Grained Traces (CGTM) [Yoon et al. 2006] and Inter-Call Mobility (ICM) [Ficek and Kencl 2012], which estimate the path traveled in-between network associations of a mobile device.

Making use of the feature-based classification. The classification is intended to guide the search for mobility models. By selecting the respective features of interest, candidate models can be efficiently identified. Then, to assess whether the candidate models are appropriate, our discussion of model properties along the engineering steps (cf. Figure 2) may be leveraged. For instance, in case fine-grained movement and location preferences are of importance in a network study, SMOOTH and SLAW are two representative mobility models addressing these features (cf. Figure 3). The engineering steps detailed in the following sections can be used to further evaluate whether these models are appropriate in terms of the modeled aspects (requirements, assumptions, mobility characterization), the kind of real data sources used (observed reality), level of realism (context), and validation quality. The outcome of such an evaluation can be that a model is sufficient or that it seems to be appropriate but lacks, e.g., the validation step, which has to be added before the model is used in the network study. In addition to this classification, the first steps of the engineering process in which basic design choices are taken also provide guidance for selecting candidate mobility models.

4. PREREQUISITES: STATING REQUIREMENTS AND OBSERVING REALITY

Before creating a mobility model, it is important to clarify the objectives of modeling and to reflect the way physical mobility may be observed. In this section, we thus describe the major decisions to take before creating a mobility model and discuss the choices taken by the different mobility models. For ease of identifying concrete mobility models, we support this discussion by a classification of the models along the choices (Table II and III). Furthermore, we detail different approaches to observe reality and relate them to existing mobility models (Table IV).

4.1. Modeling aim and design choices

The modeling aim and the purpose of an investigation in which a mobility model is used will have a profound influence on the design and the validation of the model. Due to its long tradition in structured modeling, the classical area of workload modeling [Ferrari et al. 1983] provides well-accepted guidelines for model creation. Moreover, workload modeling is partially comparable to mobility modeling since it has to be ensured that synthetically generated test workloads are representative for real workloads. A major concept that we transfer directly to mobility modeling is the definition of fundamental decisions about modeling aim and design choices, which results in these major categories: *modeling view*, *evaluation method*, *granularity level*, *mobility classes*, and the *focus of observation* in terms of concrete aspects of real-world mobility that should be monitored. Tables II and III provide a categorization of the reviewed models along these categories. We find the following:

Table II. Modeling view and evaluation method: The table relates the major categories of model view and model evaluation to mobility models implementing the respective category.

		Generic	Trace-based	Survey-based	Hybrid
Modeling view	*Physical space*	CAM Community DBS3 Disaster HCMM HGM IMMT MWP NGM Obstacle Radiation SFM Square Street	CGTM Dartmouth ICM Mall mPat NBody PSMM SLAW SMOOTH TMP TVC	Agenda SMM UDel WDM	DiAm EPM GPP MA ODFlow WHERE
	System view		IM MQNM ModelT Sheep WLANM		
	Contact process	Arrival SWIM	COM	CPP	
Evaluation method	*Simulation-based*	Arrival CAM Community Dartmouth DBS3 Disaster HGM IMMT NGM Obstacle SFM SWIM	CGTM Dartmouth Mall ModelT mPat MQNM NBody PSMM SLAW SMOOTH TVC WLANM	Agenda CPP SMM UDel WDM	DiAm EPM GPP MA ODFlow WHERE
	Analytical	HCMM MWP Radiation SFM Square Street TMP	COM IM Sheep		

Modeling view. The intention of the model defines the perspective to take during modeling. In case the model is intended to be used to evaluate a networked system, the focus is on the system's perspective and a *system view* model might be used in this case to characterize system-related features, e.g., user density at infrastructure network resources such as access points. On the contrary, in case mobility is viewed in relation to *physical space*, characteristics of the movement itself are to be modeled. The majority of models reviewed considers mobility in the physical space. Other works reproduce the view of a WLAN network (e.g., MQNM [Chen et al. 2012]) or a cellular network (e.g., Sheep [Morlot et al. 2010]). Additionally, a number of models focusing on the node *contact process* reproduce sojourns at

Table III. Modeling design choices: The table relates the categories of major modeling design choices to mobility models implementing the respective category.

		Generic	Trace-based	Survey-based	Hybrid
Level of granularity	*Finest granularity*	CAM Community DBS3 Disaster HGM IMMT MWP Obstacle SFM Street Square	CGTM Dartmouth ICM Mall NBody PSMM SLAW SMOOTH	Agenda UDel WDM	DiAm EPM MA
	Aggregated granularity	Arrival HCMM IMMT NGM Radiation SWIM	COM IM ModelT mPat MQNM NBody Sheep TMP TVC WLANM	CPP SMM UDel	GPP ODFlow WHERE
Mobility classes	*Single-class*	Arrival CAM Community DBS3 HCMM HGM NGM MWP Obstacle Radiation SFM Square Street SWIM	CGTM COM Dartmouth ICM IM Mall ModelT mPat MQNM NBody Sheep SMOOTH SLAW TMP WLANM	Agenda CPP	DiAm EPM GPP ODFlow WHERE
	Multi-class	Disaster IMMT	PSMM	SMM UDel WDM	MA
Focus of observation	*Movement paths*		NBody mPat SLAW SMOOTH TMP		DiAm EPM MA
	Network connections		CGTM Dartmouth ICM IM ModelT mPat MQNM Sheep TVC WLANM		GPP ODFlow WHERE
	Contacts		COM Mall NBody PSMM	CPP	
	Daily routines			Agenda CPP SMM WDM UDel	DiAm EPM GPP ODFlow WHERE

meeting points disregarding movement paths in-between, such as SWIM [Kosta et al. 2010] and CPP [Srinivasan et al. 2006].

Evaluation method. Whereas the evaluation method practically depends on the aspects modeled, the choice of evaluation method might restrict character and complexity of the model. Analytical evaluation usually requires a simpler representation than simulation-based evaluation; while low complexity might be a requirement for the solvability of the analytical model, a simulation model allows more details of reality. Often, model assumptions reflect the simplifications introduced for the sake of analytical tractability, such as constant velocity or uniformly distributed waypoints. Square [Desta et al. 2013] and HCMM [Boldrini et al. 2009] introduce such simplifications. As seen in Table II, analytical models are either generic models or are based on traces, whereas simulation-based evaluation can be found in each model group. Models may be also built for both analytical and simulation-based evaluation such as in the case of Street [Vukadinović et al. 2009], where a street topology is modeled analytically as a queueing network and, at the same time, movement rules for street segments allow also fine-grained simulation.

The fundamental decisions about the aim and evaluation of the model determine basic model properties and design choices. It is worth mentioning that these properties are not all independent of one another.

Level of granularity. Mobility models characterize movement with a particular resolution; the following major levels of model granularity have been identified:

— *Finest granularity – microscopic* mobility modeling level: movement paths of single entities in physical space are modeled. Inherently, most fine-grained models can be categorized as physical space models. Exceptions are, e.g., fine-grained models introducing a cell structure such as CAM [Blue and Adler 2001], in which cells of $0.21\,\mathrm{m}^2$ size are occupied by one node at a time.
— *Aggregated granularity – mesoscopic* and *macroscopic* mobility modeling level: fine-grained physical paths of single entities can be abstracted to a higher, aggregated level. The aggregation of homogeneously moving entities to groups or flows is sometimes referred to as *mesoscopic* modeling, whereas the aggregation of location details is often referred to as *macroscopic* modeling. Aggregating homogeneously moving entities is done in group mobility models (e.g., NGM [Rossi et al. 2005]), which describe people who might travel together in public transport vehicles or who form "flocks" in crowd behavior [Laube et al. 2008], and in flow mobility models (e.g., ODFlow [Calabrese et al. 2011]), which may aggregate large quantities of vehicles moving in the same direction on a road segment. Location details can be aggregated to coarser geographic regions (e.g., WHERE [Isaacman et al. 2012], Radiation [Simini et al. 2012]) or to a cell structure (e.g., SWIM [Kosta et al. 2010]). System view models, such as MQNM [Chen et al. 2012], and contact process view models, describe mobility at an aggregated level as movement details are abstracted.

The UDel model [Kim et al. 2009] includes both levels of granularity since (aggregated) demographic dynamics of the city are modeled in combination with characterizing the physical movement itself. Note that an abstraction from the finest granularity level to an aggregated level is always possible, whereas the reverse direction requires additional information or assumptions. Which level to take is a design decision depending on the purpose of the model. For example, for investigating data traffic at specific points of interest, modeling at the aggregated level of points of interest may suffice. In contrast, fine-grained modeling is needed for studying short-range device-to-device connections.

Mobility classes. Models can incorporate either a single or multiple mobility types. Multiple mobility model classes result from heterogeneous mobility patterns that should be represented by the model. Major examples of different mobility classes included in one model are: (*i*) work/residential/leisure [Basgeet et al. 2003; Ekman et al. 2008], (*ii*) outdoor/indoor [Ekman et al. 2008; Kim et al. 2009], (*iii*) individual/public transport [Ekman et al. 2008; Kim et al. 2009], (*iv*) civilians/first responders [Nelson et al. 2007], and (*v*) periodic/purpose-driven movement such as in PSMM [Cho et al. 2011], where nodes follow basically periodic patterns while occasionally conducting specific purpose travels (e.g., long distance visits of friends/family). Conceptually there are two methods for identifying model classes, they are either defined a priori or inferred from the observation, e.g., by using unsupervised pattern recognition or clustering techniques such as used in [Chen et al. 2011].

Focus of empirical observation. The focus of observation is not only determined by the purpose of the model, but also depends on the model's granularity and mobility classes. In a pedestrian class, it might be the fine-grained walking trajectories of people that are focused on, while for modeling at an aggregated level from a system perspective, data about association events of access points might be collected. We group the foci of the empirical observations appearing in the literature into (*i*) movement paths, (*ii*) network connections, (*iii*) contacts, and (*iv*) daily routines, based on which type of empirical data is selected for model building. Inherently, an observation focus on movement paths results in (or is required for) fine-grained models. Yet in some cases refinement has been done for coarse-grained network traces (which will be discussed in Section 4.3). Most existing survey models and some hybrid ones characterize movement behavior triggered by the sequence of daily activities, determining all trips. Trace-based models can be grouped into models that focus on movement trajectories with preferably high resolution, models that use network association information, and models merely relying on contact data as this allows to study ad-hoc data exchange opportunities.

An additional decision to take is whether and to which extent a model should address realism. The creation of realistic mobility models by including context and their validation are discussed in detail respectively in Sections 5.4 and 6. By introducing boundary conditions, the observation may be temporally or spatially limited and, as a consequence, constrains the mobility model. For example, only working day or rush hour mobility, or the mobility in particular geographic areas may be of interest. Spatial and temporal boundaries concern mainly data collection by empirical observation and data selection and will be discussed in the following subsection.

The planning phase is of particular importance when selecting a suitable existing model for a certain purpose. The categorizations in Tables II and III raise not only awareness for the implications of certain design choices on the nature of a model, but also help to quickly narrow down the pool of candidate models. For example, in the following cases the categorizations along level of granularity and mobility classes are helpful: If a mobility model should be used within a user-level network simulation, mobility models on the macroscopic level do not provide enough detail and can thus be excluded from the selection process. Or someone might be looking for a mobility model capable of covering the entirety of mobile network users in a city while at the same time allowing him to account for different types of mobility impacting network demands, such as public transport travelers browsing the Internet or streaming video versus individuals driving a vehicle.

4.2. Empirical observation of reality

Before a mobility model can be created, movement behavior has to be understood. To reach this aim, recent mobility modeling approaches chose empirical observation of real movement [Aschenbruck et al. 2011; Musolesi and Mascolo 2009]. Whereas basic mobility characteristics are observed to create a representative *baseline mobility model*, the context of movement is also included when creating a realistic *extended mobility model* (cf. Figure 1). Inherently, each observation can capture only certain aspects of real movement and context, which leads to a loss of information about reality.

Observation of mobility results in mobility traces, data from traffic studies, or network usage traces. In the following, we briefly introduce the notion of mobility trace and describe how data are gathered and transformed; Table IV classifies the surveyed models along how data are gathered, how observation is limited, and whether data are transformed before they are used for model creation.

Mobility traces. Observation of moving entities in the real world results in *mobility traces*. When describing the movement of a mobile entity with finest granularity, we define a three-dimensional spatial representation of a mobile entity's movement. Figure 4 shows a 2-D projection of such a continuous movement path of an entity on an area and its potential discrete representation. The continuous path can be approximated by a polygonal path and time-discretization based on either equally-sized intervals or certain events such as a change in velocity or direction. This polygonal path can be described by a sequence of tuples $\vec{x}(t_i) = (x, y, z)_{t_i}$, giving x-, y-, and z-coordinates at time t_i which are connected by straight line segments. It is worth mentioning that an approximation of a time-continuous movement path can never be a complete representation and thus modeling errors already arise here. The length of the interval between t_i and t_{i+1} depends on the modeling context or simply on the available data. Moreover, the length of this time interval determines the minimum length of sojourn times that can be captured.

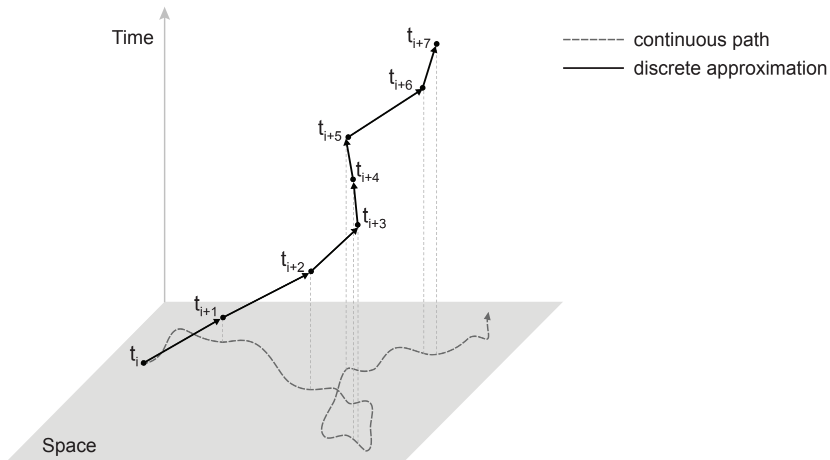

Fig. 4. Schematic view of a continuous path (2-D projection) and its discrete approximation as a space-time polygonal path describing the movement path of a mobile entity.

From the polygonal path representation a vector $\vec{a}(t_i)$ can be derived as equivalent representation, where each $a_i(t_i)$ is composed of characteristics, such as direction of movement, velocity, or even encounters with other mobile entities for the path segment between t_i and t_{i+1} and, in case of event driven modeling, dwell time at t_i. The *mobility trace* of an observation may thus feature the sequence of positions of the polygonal path and a multiple of other derived features.

Gathering data. Data for model building may already be available through public sources or earlier studies, or have to be collected within an empirical observation campaign. In any case data are commonly anonymized to adhere to privacy concerns. Possible problems when using data from other sources are restrictions in granularity or context view, or even completely unknown context. A common problem in particular for network connection data is the lack of network elements' location information. For example, mobile cellular data (e.g., Reality Mining data [Eagle and Pentland 2006]) often provide no more than the cell tower IDs as location information, which precludes deriving mobility information at a fine-grained level, but serves system-view models well. Even when location information is provided, this information is often blurred or coarsened to protect privacy. Notable concepts for degrading the quality of location information by introducing noise include data obfuscation [Duckham and Kulik 2005], which has been proposed in the context of individuals authenticated (and thus not anonymous) for location-based services, and releasing data in differentially private manner such as proposed in [Mir et al. 2013] for mobility traces generated based on the WHERE model [Isaacman et al. 2012]. Coarsening is often done by spatial grouping of similar (sub-)trajectories [Abul et al. 2008; Görnerup et al. 2015] or by aggregating user locations. Example data sets are the D4D challenge data [de Montjoye et al. 2014] blurring cell tower locations by assigning random coordinates within a tower's Voronoi region and the Telecom Italia challenge data [Douglass et al. 2015] aggregating the network activities of all users located in the same cell of a grid structure spanning a city. A popular public online archive for mobility data is the CRAWDAD repository [Yeo et al. 2006]; a survey of trace data sets is presented in [Aschenbruck et al. 2011].

When designers of mobility models carry out their own data collection campaigns, they can better control which data is collected and more relevant features can be measured. However, mobility observation experiments are time intensive, require the consent and willingness of individuals, and provide insights solely into the specific settings of the experiment.

Table IV. Mobility data: The table summarizes the type of data used by each mobility model as well the limits and data transformation applied in the respective modeling approach.

	mPat	MA	SLAW	SMOOTH	DiAm	EPM	NBody	COM	MQNM	TVC	ModelT	Dartmouth	WLANM	CGTM	GPP	ODFlow	PSMM	ICM	WHERE	Sheep	IM	Mall	TMP	Udel	CCP	Agenda	SMM
Data gathering																											
Trace data																											
Geo-location data	•	•	•	•	•	•	•																				
WLAN network data								•	•	•	•	•	•	•													
Cellular network data	•														•	•	•	•	•	•	•						
Bluetooth data								•														•					
Social appl. data																	•				•		•				
Transport check-ins	•																										
Survey data																											
Activity surveys		•			•										•									•	•	•	
Mobility statistics						•										•								•			
Census data																•			•								•
Limitations																											
Spatial limitation																											
City, country	•	•			•		•								•	•	•	•	•	•	•		•	•		•	•
Campus, conference			•	•				•	•	•	•	•	•	•											•		
Theme park, shop			•	•		•																•					
Temporal limitation									•			•	•							•		•			•		
Special data selection							•					•		•	•	•							•				
Re-scaling																											
Refinement												•		•				•									
Location aggregation	•															•			•				•	•			
Entity aggregation							•																				

Mobility trace data can be differentiated into geo-location data, data gathered through WLAN-, cellular-, and Bluetooth networks, and location check-in data from location-based (social) applications:

— In *geo-location traces* (currently, mainly GPS-based traces) an arbitrary position on earth is defined by geographic coordinates for a point in time. For GPS, position accuracy of $2-10\,\mathrm{m}$ can be assumed (tens of centimeters when using, e.g., differential GPS). Major models based on geo-location traces are SLAW [Lee et al. 2009] and EPM [Vukadinović et al. 2014].

— In contrast, *WLAN and mobile cellular data* specify an area given by, e.g., the transmission range of an access point or cell (segment) size featured by a base station, where the moving entity is located at a point in time. The accuracy greatly varies with the range of the base station – e.g., in 2G/3G cellular networks from a few hundred meters in dense urban areas up to tens of kilometers in flat rural areas. To improve the accuracy, some cellular data include position estimation based on triangulation such as in GPP [Calabrese et al. 2010]. Both, WLAN and cellular data, are geographically restricted to a region or country as they are collected by regional network operators. Mobile cellular data are often based on CDR (call detail record) data that, however, capture mobility

only when communication activity is present, such as a call, text messaging, or data transfer. Drawing conclusions for human movement about a population in its entirety may not be possible and mobility models may remain biased. Example models based on CDR data are ODFlow [Calabrese et al. 2011] and WHERE [Isaacman et al. 2012].
— *Bluetooth network traces* can be either connections with stationary Bluetooth devices (position known), or between Bluetooth devices – usually without position measurements. Contact process models based on Bluetooth data are COM [Chaintreau et al. 2007] and Mall [Galati et al. 2013].
— *Social application data* can contain point in time logs for check-ins at locations, as available in location-based social applications like Foursquare [Song et al. 2010; Cho et al. 2011], or location information for geotagged user entries on social media platforms like Twitter [Jurdak et al. 2015]. In addition, these data can provide friendship and followership network graphs such as used in PSMM [Cho et al. 2011] or semantic information extracted from the content of the user entry [Frank et al. 2013].
— *Public transport check-in traces* are a type of data which has been introduced most recently. They are collected by means of contactless smart cards serving as identifier for a traveler entering a transportation means or a subway station (e.g., see mPat [Zhang et al. 2014]).

In addition to trace-based data, *survey data* can be leveraged to extract aspects of mobility. Daily routines are mostly deduced from time use studies, as well as worker meeting studies [Kim et al. 2009; Zheng et al. 2006]. Time tables of students combined with interviews how regularly they attend lessons serve a similar purpose [Srinivasan et al. 2006]. Census data provide resident population [Song et al. 2010] and employment population [Kim et al. 2009] figures for geographic areas as well as statistics about daily commuting distances [Calabrese et al. 2011; Isaacman et al. 2012], etc. These data are usually semantically richer than trace data, but spatially and temporally coarser, and might depend on the accuracy of human memory (cf. the discussion on transport demand data in [Cottrill et al. 2013]). Thus, they usually allow modeling mobility on an aggregated level, while additional data are required to model the movement between waypoints. Additional survey data are pedestrian and traffic statistics [Calabrese et al. 2011; Kim et al. 2009] for modeling velocity, distance-speed relations, vehicle density, etc., or theme park visitor statistics [Vukadinović et al. 2014]. A drawback of survey data in general might be the low update frequency as surveys are often conducted in intervals of several years. Another possible drawback of survey data, their availability, has been improved recently, e.g., due to open data initiatives. An advantage over other data types is that survey data often cover the whole population of an area or a country and are thus unbiased towards special groups, such as mobile phone users currently involved in calls or volunteers participating in a study.

Limitations. In many models, limitations of the observation in space and/or time are present due to practical limitations or study aims.

— *Spatial limitations* refer to restrictions depending on the environments in which data are collected. Not surprisingly, WLAN and Bluetooth traces are collected in closed scenarios such as a campus, conference, shopping mall, whereas the spatial range of geo-location and cellular data varies heavily and can be a whole country (cf. mobile phone data of Senegal [de Montjoye et al. 2014]). Survey data are available for a geographic region, typically a city or a country.
— *Temporal limitations* are existent in several mobility models. Some models focus only on working hours, i.e., night-time or weekend traces are not collected (or eliminated). Trace sets collected during mass gathering events such as the "Fête de la Musique" in Paris for the Sheep model [Morlot et al. 2010] or a concert and football match in Rome [Calabrese et al. 2011], have been used for modeling peak network loads.

In absence of real data, synthetic data may be used. Generic models based on social communities, such as Community [Musolesi and Mascolo 2006], utilize randomly generated social relationship graphs, which can be replaced by real-world social network data as these data become more and more available. Synthetic graphs/matrices showing social connections are often generated by means of the Caveman model [Watts 1999]. While synthetic data can be comparatively easy generated by using existing models, it is usually difficult to achieve representativeness.

4.3. Data cleaning, selection, and transformation

Preparing "raw" empirical data sets for model building often includes data *cleaning*, *selection* along temporal/spatial limits, and transformation such as *re-scaling*. Typical cleaning tasks are eliminating positioning errors or smoothing the ping-pong phenomenon [Yoon et al. 2006], which refers to undesired handovers between two base stations back and forth happening in a short period of time when the mobile device resides, e.g., at the border of two cells. Special selection procedures applied by mobility modeling (cf. Table IV) include selecting cellular network users with a sufficiently high number of connections to reduce phases with unknown position (granularity-based selection) [Calabrese et al. 2010], removing stationary WLAN users [Kim et al. 2006], or selecting taxi cars active at the same time period [Zhao and Sichitiu 2010].

Re-scaling means either down- or up-scaling the size in space or time. When down-scaling temporal granularity, the sampling intervals are enlarged. For example, if positions were measured in irregular intervals of 1–30 s, the intervals could be equalized by maintaining values in gaps of, e.g., 1 min. Increasing the time resolution, on the other hand, can be achieved by interpolating between measured values. Spatial re-scaling can occur when transferring absolute into relative coordinates (physical distances to points of interest). In many of the surveyed papers, these data transformations are not discussed explicitly. However, data transformations have been carried out for publicly available mobility data sets. Examples are the NCSU data [Rhee et al. 2009] with temporal and spatial re-scaling – this data set has been used for the models SLAW [Lee et al. 2009] and SMOOTH [Munjal et al. 2011]. *Refining* coarse-grained traces to finer-grained trajectories is applied in [Ficek and Kencl 2012; Kim et al. 2006; Yoon et al. 2006]: ICM [Ficek and Kencl 2012] estimates trajectories between two consecutive calls (handled by base stations with a distance larger than 3 km) based on known inter-call trajectories of other users. The CGTM model [Yoon et al. 2006] maps transitions between APs installed in different buildings to movements onto known pathways in between. Examples of *coarsening traces* based on *location aggregation* are given in, e.g., [Basgeet et al. 2003; Calabrese et al. 2011; Simini et al. 2012]: Since SMM [Basgeet et al. 2003] describes the mobility between "area zones" of a city, locations are abstracted to four zones spanning the whole city area and its surroundings: city center, urban, suburban, and rural areas. In ODFlow [Calabrese et al. 2011], a grid structure consisting of rectangular cells with 500 m edge length is introduced to model the user density based on mobile phone data. For the evaluation of the Radiation model [Simini et al. 2012], call detail records are aggregated to the total number of phone calls between each pair of cities. *Entity aggregation*, in particular the group forming of nodes, is implemented in the generic (non observation-based) model NGM [Rossi et al. 2005], and NBody [Zhao and Sichitiu 2010].

5. CREATING THE MOBILITY MODEL

After clarifying modeling aim and selecting data sets, the core of mobility modeling, model creation can take place. Based on the collected data sets the *baseline mobility model* describing movement is constructed. By adding context, the *extended mobility model* is created and realism is introduced. We use again a categorization approach to relate the surveyed mobility models to modeled mobility features (Table V) and context characteristics (Table VI). The categories allow us to get a quick overview of important modeling aspects, the discussion emphasizes the most important methods, and the tables provide references to the mobility models implementing methods related to the respective category.

5.1. Characterizing mobility

In general, the movement of a node starts at a location, which might be selected randomly or based on the context of the location – e.g., locations considered as "home" or "meeting point" of a social community to which the node is assigned (cf. Section 5.3). Then, a node commonly traverses several waypoints to eventually reach an end location. The characteristics of this movement are the basis of model creation (and at a later stage they will be used for model validation). They are extracted by statistical analysis and data mining methods.

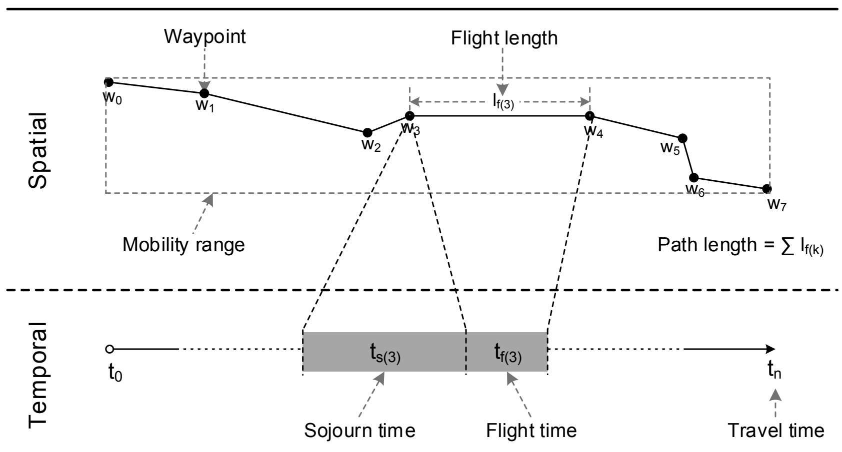

Fig. 5. Common direct mobility characteristics in the space and time domain.

Single mobility characteristics. The properties of movement found in mobility data sets can be quantified by means of distinctive mobility characteristics, which are features describing one aspect of movement. A mobility characteristic can be a direct description, such as velocity, or an indirect description of movement, such as the frequency of location revisits. An overview of direct mobility characteristics is given in Figure 5; models referring to these characteristics are summarized in Table V.

The most common *direct characteristics* are velocity and sojourn time (a.k.a. pause time, dwell time, or waiting time). The location of a sojourn is defined by waypoints or indirectly by the flight length (or jump size), which is the length of a path traveled between two consecutive pause times, and the direction. Change of direction is rarely modeled explicitly as it often results from heading to the next waypoint (which might also be the destination location) or from a topological map. The flight time is the time the entity moves between two waypoints. Characteristics describing an entire trace are travel time, movement range, and path length. With respect to the time dimension, which is often represented by sojourn times and flight times, it has to be mentioned that these time intervals are often not modeled explicitly. In discretized-time models nodes might chose a new location in each time slot, while several (mostly open) mobility models assume continuous movement without sojourn phases. Instead, some of these open models include an acceleration/deceleration pattern, which can be considered to be similar to the move/sojourn pattern.

Aggregated mobility characteristics. Aggregated characteristics describe properties derived from multiple trajectories. *Transition probabilities* determine changeovers between areas, such as city districts in macroscopic models (e.g., in SMM [Basgeet et al. 2003], mPat [Zhang et al. 2014], WHERE [Isaacman et al. 2012]), WLAN access points (ModelT [Jain et al. 2005] and WLANM [Tuduce and Gross 2005]), campus hotspots (Dartmouth [Kim et al. 2006]), or attractive points in the city during mass gatherings (Sheep [Morlot et al. 2010]). In social-based models, transition matrices of movements to other cells express the cell's social attraction (HCMM [Boldrini et al. 2009]) or popularity and distance from home cell (SWIM [Kosta et al. 2010]). Note that the weight for each area is in many cases derived from context information (cf. Section 5.4). Time-variant transition probabilities are modeled in IM [Song et al. 2010] suggesting that the probability for exploring additional locations is decreasing over time while "preferential returns" become more likely. *Arrival* times and rates are commonly modeled for either a point of interest, such as subway stations [Kim et al. 2009] or streets [Vukadinović et al. 2009], or the whole study area is open in case of open mobility models[6], like Square [Desta et al. 2013]. Related to arrival rates, the Sheep model [Morlot et al. 2010] pictures the filling and scattering dynamics of randomly formed concentration areas of human gatherings. When mobility is abstracted to *contacts* the pure contact process might be modeled based on the characteristics inter-contact time and contact duration [Chaintreau et al. 2007].

5.2. Creating a baseline mobility model

When modeling mobility, all relevant aspects of movement have to be described along selected mobility characteristics. Depending on the model type, the respective mobility characteristics may be basic direct movement characteristics (waypoints, flight length, pause time, etc.) but may also express social structures and group dynamics along contact characteristics (contact time, inter-contact time). To capture the variability of the characteristics, stochastic processes and derived distribution functions are employed.

Distribution functions and statistical patterns. For quantifying mobility characteristics, each characteristic may be assigned a distribution function derived from its empirical probability density function (PDF). If the type of distribution function is known (e.g., from the literature), a pre-defined distribution function may be selected and its parameter setting might be adapted to the data set. Otherwise, the data are analyzed and the empirical PDF may be fitted to a set of possible distribution functions following a maximum likelihood estimation method. In this case, it is tested which distribution out of the set is the "best" fitting distribution by means of a *goodness-of-fit test.* If a characteristic has a specific value pattern that does not follow a standard distribution function, the empirical distribution function itself might be applied. However, it is preferable to utilize a known distribution function for the sake of reproducibility of the model.

Sample insights into distribution functions observed by several real data studies are that flight length and pause times follow truncated power law distributions (see IM [Song et al. 2010], SLAW [Lee et al. 2009], SMOOTH [Munjal et al. 2011]). For contact processes, also inter-contact times have been found to follow a truncated power-law distribution [Chaintreau et al. 2007] (this insight can be used for modeling but also for validation). In SLAW, waypoints are modeled as "fractal points" to induce truncated power law flights. This allows to implement attraction to popular places [Lee et al. 2009]; the mobility range of nodes is considered to be heterogeneous.

[6]We use the notion *open mobility model* for models enabling nodes to arrive and depart from the study area at any time. These models differ from the majority of models featuring a constant node number during the entire period modeled.

Table V. Mobility characterization: The table relates the categories of mobility characterization and model creation to mobility models implementing the respective category.

		Generic	Trace-based	Survey-based	Hybrid
Movement characteristics	*Velocity*	CAM DBS3 Disaster HCMM HGM IMMT MWP NGM Obstacle SFM Square Street	CGTM Dartmouth NBody TVC WLANM	Agenda WDM	EPM MA
	Acceleration/ deceleration	CAM IMMT SFM Street	NBody		
	Pause/Sojourn time	DBS3 MWP Obstacle SWIM	CGTM Dartmouth IM Mall MQNM NBody PSMM Sheep SLAW SMOOTH TMP TVC WLANM	Agenda SMM WDM UDel	DiAm EPM MA
	Discretized time	Arrival Community HCMM NGM Radiation	COM mPat	CPP	GPP ODFlow WHERE
	Continuous movement	CAM Disaster HGM IMMT SFM Square Street	ICM ModelT		
	Waypoints	DBS3 HCMM IMMT MWP Obstacle	Dartmouth NBody SLAW SMOOTH TVC	Agenda UDel WDM	DiAm ODFlow
	Flight length/time		IM SLAW SMOOTH TMP TVC		MA
	Direction	Disaster HGM NGM SFM			MA
Aggregated characteristics	*Transition probability*	CAM HCMM Radiation SWIM	Dartmouth ModelT mPat Sheep WLANM	SMM	WHERE
	Arrival rate/Start time	Arrival Square Street	Dartmouth Mall mPat MQNM Sheep	UDel	EPM MA ODFlow
	Contacts	Arrival	COM NBody	CPP	
Common components	*Generic models*	Disaster HCMM HGM Square	Mall Sheep TVC	WDM	EPM
	Shortest path algo.	DBS3 IMMT Obstacle		Agenda UDel WDM	DiAm
	Location probability		ICM IM PSMM TMP		GPP

Common model components. When creating a model, the following components are commonly addressed:

— *Generic models* – a common practice is that mobility models incorporate generic models (such as RWP), for movements within closed spaces (e.g., city square [Desta et al. 2013], cell [Boldrini et al. 2009], shop [Galati et al. 2013]) or for single node groups, like in the Sheep model [Morlot et al. 2010] where a fraction of nodes follows a Random Waypoint-based "maverick" movement, i.e., the nodes are moving independently and not following the group movement. Several generic models are extensions of others, such as HGM [Du et al. 2012], which is a random walk model with entities gravitating towards a hotspot if they arrive in its vicinity.

— *Shortest path algorithms* are applied when assigning source and destination locations to nodes. The locations assigned might be locations of, e.g., home, work, and leisure places in activity-driven models [Ekman et al. 2008; Kim et al. 2009; Zheng et al. 2006]. The assigned locations might also be randomly selected locations, chosen by uniform selection [Jardosh et al. 2003], or chosen such that destinations are more likely to be centrally located in the street grid [Vogt et al. 2012].

— *Location probabilities* – instead of modeling mobility characteristics, predictive models calculate location probabilities for individuals, i.e., the probability that they can be found in a given location at a given time. For example, PSMM [Cho et al. 2011] predicts geographic locations for a given weekday and daytime by means of an expectation-maximization estimation based on the places regularly visited by each person in the data set. In time geography, possible locations of an entity can be modeled probabilistically using a space-time volume representation (cf. [Winter and Yin 2011]). For example, space-time cone representations are based on a known location of an entity, its movement direction and maximum speed. A space-time prism model [Winter and Yin 2010] describes the set of all points, which lie between two known locations of an entity and which can be reached given certain time and space constraints. This modeling method is adopted by ICM [Ficek and Kencl 2012], which predicts the unknown path taken between two known locations inferred from cell tower coordinates by means of kernel density estimation. The estimation of the path relative to the cell towers is based on an aggregation of finer-grained data available for a different set of users.

The newly created baseline mobility model has to be validated against real-world observations – this validation step is discussed in Section 6.2. The implementation of the baseline mobility model outputs synthetic traces used first for validation and finally as input for mobile network simulations. In case a realistic mobility model is aimed for, the model has to be enriched with context information.

5.3. Describing context

Movement takes place in context, such as in a particular place or at a particular time, which influences mobility behavior. Table VI summarizes the context categories found in all surveyed models. The major context categories are geographic, temporal, entity-specific, relational, and demographic context[7].

Geographic context. Geographic context typically includes topological structures derived from map data, and further single elements, such as street types (highway, pedestrian road, etc.) and speed limitations, number of lanes, traffic lights, etc. A street topology may be a simple grid, in indoor environments the corridor topology may be described by a more or less regular three dimensional grid. These topologies impact the mobility model by restricting movement.

On an aggregated level, topological structures might be connected to capture transitions between larger areas such as administrative districts [Basgeet et al. 2003; Simini et al. 2012; Zhang et al. 2014] or cities, or abstracted to a grid structure [Boldrini et al. 2009; Kosta et al. 2010]. A transition of a mobile entity to another area may be triggered by its higher attractiveness due to job opportunities in nearby districts as considered in the migration pattern model Radiation [Simini et al. 2012] (see demographic context) or by being part of a social community, e.g., HCMM [Boldrini et al. 2009] (see social context). Points of interest modeled are mostly locations of work places and homes [Isaacman et al. 2012; Zhang et al. 2014], but also restaurants or general attraction points, such as highly frequented squares for pedestrians or intersections for vehicles. Points of interest found in generic models are hot spot areas – generated through node gravitation to attraction points [Du et al. 2012]

[7]Some trace-based and survey-based models do not consider context.

Table VI. Mobility context: Types of context found in literature related to mobility models featuring the respective context type.

		Generic	Trace-based	Survey-based	Hybrid
Geographic context	*Topology*	DBS3 IMMT Obstacle Street	CGTM	Agenda UDel WDM	DiAm EPM
	Aggregated topology	HCMM Radiation SWIM	mPat	SMM	WHERE
	Indoor map		Mall	UDel WDM	
	Points of interest	Disaster HGM MWP	Dartmouth Sheep TVC	Agenda SMM UDel WDM	DiAm EPM GPP ODFlow WHERE
	Attraction or repulsion	Disaster HCMM HGM IMMT Radiation SFM SWIM	NBody Sheep	SMM	
Temporal context	*Time of day*	Community IMMT	mPat PSMM	Agenda SMM UDel WDM	EPM MA
	Day of week	Community IMMT	PSMM TVC		ODFlow
Entity context	*Roles*	Disaster	Dartmouth Mall MQNM Sheep	Agenda UDel WDM	
	Activities			Agenda UDel WDM	DiAm EPM GPP MA
	Means of transport			UDel WDM	
Relational context	*Interactions*	CAM NGM SFM	NBody	UDel	EPM
	Social relations	Arrival Community HCMM NGM SWIM	NBody PSMM		
Demographic context		IMMT Radiation Square		SMM UDel	ODFlow WHERE
No context			ICM IM ModelT SLAW SMOOTH TMP WLANM	CPP	

or through a discrete time Markov process model that conditions velocity on the location of the two ending waypoints of a transition [Hyytiä et al. 2006] – and locations of disaster events featuring repelling and attracting forces [Nelson et al. 2007].

Temporal context. Typical temporal context characteristics are *time of day* and *day of week*. Many mobility models differentiate mobility behavior along these characteristics. PSMM [Cho et al. 2011] applies both to predict locations of a user for a specific point in time. The generic model Community [Musolesi and Mascolo 2006] allows to define different relationships for daytime (to colleagues at work), evenings, and weekends (to family and friends). The time of day may be used to distinguish between normal, rush hour, and busy hour periods (SMM [Basgeet et al. 2003]), and to schedule activities such as commuting to work (WDM [Ekman et al. 2008], UDel [Kim et al. 2009], Agenda [Zheng et al. 2006]). All in all, the temporal context is used to adapt mobility to usual temporal mobility patterns.

Entity context. This context type addresses the individual moving node. Nodes may take different *roles* within a model, determining their mobility behavior. The concept of roles can be found in [Chen et al. 2012; Ekman et al. 2008; Galati et al. 2013; Kim et al. 2009; Morlot et al. 2010; Nelson et al. 2007]: WDM [Ekman et al. 2008] and UDel [Kim et al. 2009] differentiate between car owners, who drive their vehicles on the street grid, and public transport commuters, who use subways (UDel), buses (WDM), or walk. Roles may also relate to device usage as defined in MQNM [Chen et al. 2012], where laptop users visiting one AP and leaving the network afterwards are differentiated from always-on users wearing, e.g., phones that are constantly roaming between APs. Different social and professional behavior can be expressed by roles: the Sheep model [Morlot et al. 2010] comprises "sheep movers", who follow the crowd of people during mass gathering events, and maverick movers, who move independently; the Post-Disaster model [Nelson et al. 2007] comprises three roles, namely civilians, who flee from the disaster event (except a fraction who is curious about the disaster scene), police, who are attracted by the event, and ambulances oscillating between the disaster area and the hospital.

A likewise prevailing type of entity context are *activities*. Models viewing mobility according to daily routines consider activities such as work, commute, at home, evening, shopping, eating out, exercise, etc. [Ekman et al. 2008; Hummel and Hess 2013; Kim et al. 2009; Zheng et al. 2006]. In more specific settings, such activities can be more detailed such as modeled in the entertainment park model [Vukadinović et al. 2014], which distinguishes between visiting attractions, restaurants, events, or taking rides. Despite the similarities between the role and the activity concepts, they are not the same: activities can be changed whereas a role is held permanently.

Relational context. Relational context refers to interactions between mobile entities, or social relations and ties. Typical *interaction* model components are mechanisms for collision avoidance, such as rules for overtaking slower walking pedestrians [Blue and Adler 2001; Kim et al. 2009] or extrapolating trajectories of surrounding people to adapt direction and speed accordingly [Vukadinović et al. 2014]. A generic model picturing the lane formation behavior of pedestrians walking in the same direction is the Social Force Model (SFM) [Helbing and Molnár 1995]. Coordination-related interaction between nodes belonging to a crowd (also called flock) is modeled in the group mobility model NGM [Rossi et al. 2005]. Information about *social relations* is used to determine either the locations to visit such as community hotspots or points of interest close to a friend's place in SWIM [Kosta et al. 2010] and PSMM [Cho et al. 2011], or to determine which nodes accompany each other during traveling [Rossi et al. 2005]. The influence of joint activities on movement and geographic locations has as well been explored by the accessibility modeling domain, e.g., from an analytical viewpoint in [Neutens et al. 2010]. An illustrative application of the joint accessibility concept in urban geography is provided by Farber et al. [2013], who model the spatio-temporal availability of social interaction opportunities in a metropolitan region considering land use data and commuter flow statistics as input.

Demographic context. Demographic context might be derived from census-based data as well as traffic/pedestrian density statistics. Population characteristics are used to extract mobility characteristics of usual movement, such as commuting to work, etc. In the UDel model [Kim et al. 2009], the population characteristics number of people living within a city, commuting by car, and commuting by public transport, are leveraged. Commuting distances and work/home districts are extracted from census data by the macroscopic model WHERE [Isaacman et al. 2012]. Transitions are further motivated by a region's attractiveness: in the Radiation model [Simini et al. 2012], transitions to other geographic regions are based on a region's job opportunities, which are assumed to be proportional to the region's resident population. In some cases, demographic information is used to create a model complying better to real world observations. Census-based population data are utilized in ODFlow [Calabrese et al. 2011] to approximate the cellular trace-based model to actual population densities. The Square model [Desta et al. 2013] is adapted to pedestrian populations observed by means of public webcams of squares in three cities.

5.4. Creating an extended mobility model

Including context aims at extending the capabilities of the mobility model to capture real world phenomena better. By proceeding this way, mobility characteristics might be directly affected by context. For example, entities must adapt movement to spatial structures such as circumventing obstacles (change of direction) or adapting to other entities, e.g., by adjusting the velocity to the current entity density (CAM [Blue and Adler 2001]). Specific context may only be assigned to some mobility classes in the model, such as sidewalks to pedestrians and multiple-lane roads to cars [Kim et al. 2009]. Moreover, the model might be extended with context-dependent changes between mobility classes, e.g., changeovers from walking to public transport every time a means of transport is available nearby (WDM [Ekman et al. 2008]).

In models defining human activities, context-based adaptation of movement characteristics may also depend on the specific activity. For example, in the model WDM [Ekman et al. 2008] the pause time on the way to work is influenced by the traffic situation, whereas traffic has no influence on sojourn times at dinner or at leisure points of interest.

Finally, a model may be adjusted to its application. For example, if a particular city area is studied (UDel [Kim et al. 2009]), street map, population data such as inhabitants, (un)employed people, etc., and transportation schedules of the given city might be input. To study effects of context changes, context parametrization leaves room for later *model adaption*. An example is the macroscopic model WHERE [Isaacman et al. 2012], which considers shifts in commuting distance triggered by factors such as increasing subway fares.

Typically, there is an open list of optional context parameters to be included in order to transfer a baseline mobility model into an extended mobility model with increased realism. By including more context aspects, the extended mobility model also becomes more complicated leading to a potential loss of generality, increased likelihood of modeling errors, and the need for extensive validation.

6. VALIDATION

Validating a mobility model requires the selection of validation metrics (Section 6.1) for both the representativeness and the realism validation step (Section 6.2), of validation references, and of comparison methods (Section 6.3). The tables in this section again provide pointers to mobility models making use of the respective validation method. The validation reference may be real observations (e.g., an additional data set not used for model building) or another already validated mobility model [Härri et al. 2009]. Comparison methods may range from visualizing deviations between a model and a validation reference to applying statistical tests. Comparing the output of a mobility model, namely the synthetic traces, with a validation reference is termed *direct validation*. *Indirect validation* denotes the strategy of validating the response of the networked system under investigation. Figure 6 depicts the scheme used for direct and indirect validation against observation of the real world.

6.1. Defining validation metrics and adequacy criteria

Mobility and context characteristics may serve as direct validation metrics, whereas network performance metrics can be leveraged as indirect validation metrics. Adequacy criteria[8] determine to which degree these metrics have to comply with expectations in order to achieve a *valid mobility model*.

[8]Definition by Mayes [2009]: *"Adequacy criteria provide the required maximum acceptable difference between the validation experiment and the computational model response features."*

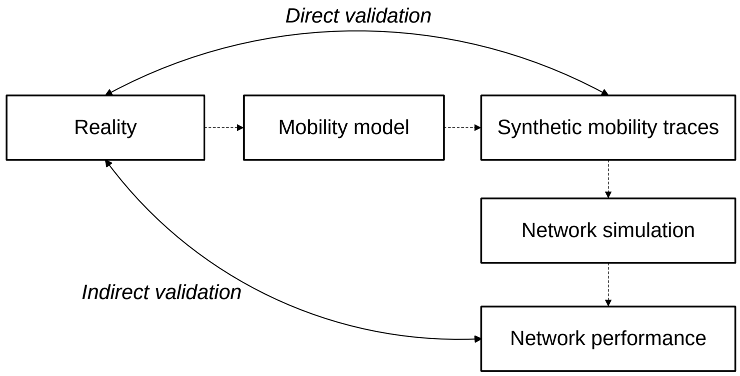

Fig. 6. Direct and indirect validation of a mobility model against real world observation.

To validate representativeness, *context-independent metrics* have to be defined. In principle, any known mobility characteristic may be leveraged (cf. Figure 5). Among those, major mobility characteristics are: mobility range defined by a geometric shape, flight length determining the covered distance from one waypoint to the next, overall path length of a trip, pause and flight time, and start and travel time. Further revisiting location and contact patterns between devices are features that may be used. Among these metrics, the flight length is a commonly used validation metric for fine-grained models [Lee et al. 2009; Munjal et al. 2011]. On coarse-grained level, path length [Chen et al. 2012], daily range [Isaacman et al. 2012], and number of visited distinct locations [Song et al. 2010] are metrics that have been selected for validation in the past.

Context-dependent metrics, on the other hand, are needed to evaluate a model's realism, since the evaluation of movement characteristics under context influence or the effect of movement on context is targeted. Influencing context may relate to space such as the topography of a region, time (daytime, season), etc., whereas, e.g., social relations may be affected by co-location patterns. Investigated effects of movement in context are pedestrian or vehicular density at a crossing (CGTM [Yoon et al. 2006]) or at a street segment (DBS3 [Vogt et al. 2012], UDel [Kim et al. 2009]). Furthermore, the deviation of a predicted location from the actual location has been used as a metric (PSMM [Cho et al. 2011], ICM [Ficek and Kencl 2012]).

6.2. Validation process – representativeness and realism

Both for validating representativeness and realism, statistical methods are used to evaluate whether the respective metrics meet the adequacy criteria. Table VII gives an overview of the validation provided for the surveyed mobility models. A first observation is that many models do not report on validation of representativeness or realism at all. Particularly generic models are often compared to other synthetic/generic models to highlight optimized model properties without connecting the model to empirical observations. To give an example, it is shown for the Hotspot Gravitation Model (HGM) that in the visual comparison the resulting trajectories appear more regular than trajectories from other generic models [Du et al. 2012].

Representativeness. Baseline models are validated by evaluating their *representativeness* with respect to context-independent metrics. Validation either leverages *built-in metrics* that are features already used for model creation or *external metrics*, which have not been used for model creation and allow to take a different perspective when rating the validity of a mobility model. Whereas built-in metrics allow to validate whether the characteristics

Table VII. Validation of representativeness, realism, or no validation as implemented by the respective surveyed mobility models.

		Generic	Trace-based	Survey-based	Hybrid
Representativeness	*Built-in metrics*		ModelT mPat SLAW SMOOTH TMP WLANM		DiAm
	External metrics		IM MQNM TMP		MA WHERE
Realism	*Location-based*	DBS3 Radiation	CGTM Dartmouth ICM mPat MQNM PSMM Sheep TVC	UDel	GPP ODFlow WHERE
	Contact-based	Community SWIM	Mall NBody SLAW SMOOTH TVC	WDM UDel	EPM
No validation		Arrival CAM Disaster HCMM HGM IMMT MWP NGM Obstacle SFM Square Street		Agenda CPP SMM	

are well modeled (e.g., statistics are sound), external metrics are needed to rate the actual compliance of the entire model with the original system. The more independent external metrics are considered, the more confidence in the model's validity can be achieved. In practice, one common context variation originates from the different topologies of pathways or streets used in simulation. A simple topology is a grid, more advanced topologies may be defined by multiple lanes, sidewalks, etc. The validation mobility traces may now be compared to the synthetic traces generated by a mobility model with a specific topology. One major finding is that in particular the flight length is sensitive to the street structure, as discussed in [Mayer and Waldhorst 2011]. In this work, a random walk and social model are exposed to different grid-based graph structures and the resulting inter-contact time distributions show that the different graph structures have a strong impact on both models.

Realism (and representativeness). Extended mobility models have to be validated both in terms of their *realism and representativeness* by comparing context-dependent and context-independent characteristics of synthetic and expected mobility behavior. In hitherto literature, more stress has been laid on realism than on representativeness (cf. Table VII), reflected by *location-based* and *contact-based* validation approaches. Most frequently the effects of the modeled mobility on visited locations or user population is evaluated. Reference data for location-based realism validations might be WLAN traces [Hsu et al. 2009], GPS traces [Hsu et al. 2009; Munjal et al. 2011], or statistics such as traffic volume data [Kim et al. 2009]. The validation of generic mobility models in terms of realism is targeted for social community-based models such as SWIM [Kosta et al. 2010], for which contact patterns are validated with Bluetooth-based trace data. It can be observed that validating contact-based realism is often carried out leveraging Bluetooth contact trace data, cf. also WDM [Ekman et al. 2008], TVC [Hsu et al. 2009], and SMOOTH [Munjal et al. 2011].

Realism versus representativeness. Although realism is often not explicitly differentiated from representativeness, realism does not automatically imply representativeness. As an example, let us assume a fine-grained mobility model which has been created based on GPS traces of individuals, who move within an urban area as pedestrians or using various means of transport. The representativeness of the baseline model is validated using the context-*independent* metrics direction change and flight length. The model is later extended with context information by mapping the trajectories to a street topology. The realism of the extended model is validated by means of context-*dependent* metrics, such as density at a pedestrian crossing. Obviously, the mapping onto a street topology may have a strong impact on the context-*independent* metrics as, e.g., adapting a trajectory to the course of a road will affect direction changes and introducing traffic lights will cause more sojourn times and shorter flights. Thus, it is not clear whether the new trajectories in the extended model are still representative with respect to direction changes or flight lengths due to the impact of the context information. Regarding the implication of a negative result of the validation of representativeness of the extended model, it is an open question whether and under which circumstances it is reasonable (in view of the modeling aim) to remove context information distorting representativeness. Alternatively, one might argue that it is reasonable to discard single trajectories that do not pass both validation runs, while this would distort the node density, which is a metric for realism in the example above.

6.3. Methods for validation and model calibration

Validation methodology. Several validation approaches are mentioned in recent works, each having its benefits and weaknesses. Validation usually begins with a *visual comparison* of the distribution of validation metrics, which is done for the majority of models surveyed (cf. Table VIII). However, visual comparison only gives a very basic idea of how well a model fits a validation reference and what could be a possible cause for a lack-of-fit. Moreover, visual validation does not allow the designer of a mobility model to make general statements about the model's representativeness or, e.g., to quantify and compare the representativeness of several models for the same validation references. The reliability of the validation process can be reinforced by additionally *computing the deviation* between observed values of mobility characteristics and those occurring in the synthetic traces. High reliability is ensured when a standardized statistical test, such as a goodness-of-fit test, is applied. Most of these tests would, e.g., reject a distribution function exhibiting an inadequate fit without the need to define a well-chosen threshold. Only a few works perform a goodness-of-fit test – common tests in use are: the Kolmogorov-Smirnov test [Chen et al. 2012], the mixed-sample method [Ficek and Kencl 2012], or calculating the Kullback-Leibler divergence [Lee et al. 2012; Meyer et al. 2011] or the AIC (Akaike Information Criterion) [Jurdak et al. 2015].

For comparison data sets or a reference mobility model are needed. A good practice followed by the discussed models is validation with *other data sets* that have not been used in the model building process. An example is the generically formulated Radiation model [Simini et al. 2012] that is validated with multiple data types, such as census data and inter-city travels extracted from cellular network data. If one trace data set is applied for both model building and validation, it is reasonable to *split the set into training and test data* such as done for ModelT [Jain et al. 2005] and WLANM [Tuduce and Gross 2005].

As more and more implementations of mobility models are publicly available, it becomes possible to evaluate models also against other already existing models. This has been done for SMOOTH [Munjal et al. 2011], whose validation metrics are compared to SWIM [Kosta et al. 2010], SLAW [Lee et al. 2009], and TLW [Rhee et al. 2008]. mPat [Zhang et al. 2014] is compared to the other coarse-grained models Radiation [Simini et al. 2012] and WHERE [Isaacman et al. 2012] for different time scales as well as daytimes, which led to observations such as that all models perform best for morning commuter mobility confirming the high predictability of this special mobility case. In another work, the presence probability at all possible locations between two cell towers estimated by the interpolation model ICM [Ficek and Kencl 2012] is validated against a space-time prism model [Winter

Table VIII. Validation methodology and data sets, and mobility model calibration as implemented by the respective surveyed mobility models.

		Generic	Trace-based	Survey-based	Hybrid
Methods	*Visual comparison*	Community Radiation SWIM	Mall ModelT mPat Sheep SLAW SMOOTH TVC	UDel	EPM
	Computation of deviation	DBS3 Radiation	Dartmouth ModelT mPat NBody PSMM TVC WLANM		GPP ODFlow WHERE
	Goodness-of-fit test		ICM MQNM SLAW TMP		
Data sets	*Other data sets*	Community Radiation SWIM	ICM mPat SMOOTH TVC	UDel WDM	
	Training/test data		ModelT mPat PSMM WLANM		DiAm
Models	*Other mobility models*		ICM mPat PSMM SLAW SMOOTH TMP		
Calibration		Radiation SWIM	Sheep	SMOOTH TVC	EPM

and Yin 2010] (see Section 5.2). Dedicated to the validation of space-time volume models, [Kobayashi et al. 2011] provide an analytical tool for assessing the spatial error propagation from model parameters to prisms and prism-prism intersections, which indicate potential contacts with other moving entities.

Model calibration. In case of a negative validation result, the model needs to be *calibrated* in an iterative process, i.e., the model is adapted and its representativeness or realism is rated repeatedly until the adequacy criteria are fulfilled. Possible calibration means to improve a non-representative or unrealistic model are (*i*) adapting parameter values and ranges, (*ii*) adding or discarding parameters, and (*iii*) changing parameter relations (dependencies). Most commonly, the term “calibration” is used to refer to the aspect of “parametrization”: Parameter values of a mobility model are adapted to match a set of traces or a particular context.

The studied models address calibration by changing particular model parameter values (or value ranges). Methodologically of interest is the approach presented by Sheep [Morlot et al. 2010], which adjusts the number of city zones, total number of nodes, maverickness rate (the probability that a node is not influenced by other nodes), and pause time to create user densities for three attraction areas in Paris that match the observed total number of sent text messages derived from GSM traces. Another important representative of model calibration is demonstrated in the validation of EPM [Vukadinović et al. 2014]. The model is adapted to properties of the given environment (a theme park) and the mobility observed there. The study area is adapted according to the theme park map specifying the location of points of interest such as attraction areas or restaurants. Other context parameters are derived from visitor statistics (e.g., guest arrival rate) and time schedules (e.g., time of events, ride durations). Transition probabilities between points of interest and sojourn times are derived from GPS traces. After calibration, synthetic traces are generated that are the basis for the contact-based validation of realism. Other works conducting calibration processes aim at recreating patterns observed in different types of trace data, e.g., by fitting model parameters (see Radiation [Simini et al. 2012], SMOOTH [Munjal et al. 2011], SWIM [Kosta et al. 2010], TVC [Hsu et al. 2009]).

Whereas changing parameter values or value ranges can be easily performed, adding or removing parameters or relations is more complicated. In particular in case multiple (partly dependent) forces have to be considered such as social ties and physical environment, more sophisticated model calibration techniques are needed. This is an open challenge in mobility modeling research.

7. SUMMARY OF OBSERVATIONS AND TRENDS IN MOBILITY MODELING

Mobility modeling has a long tradition in wireless networking. Table IX gives an overview of the reviewed mobility models for wireless networking as they appeared in the literature over time[9]. We can observe that the first mobility models have been deployed to analyze cellular networks, e.g., to support capacity planning. Early models could draw on fundamental concepts from traffic engineering, such as land use forecasting (cf. [Herbert and Stevens 1960]), spatial gravity (cf. [Evans 1973]), or source/destination path computation (cf. [van Vliet 1978]). Later, finer-grained models have been introduced. Generally, trace-based models have become predominant in the last few years providing a good foundation for realistic mobility modeling.

Table IX. Timeline of mobility models presented for wireless networking evaluations. Upper part: number of presented mobility models using a particular type of data set (both for model building and validation). Lower part: number of models along distinguishing features of the models. The column headers give the total number of models published (it should be noted that in several instances a single model utilizes more than one type of data set).

	Year	≤'01	'03	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15
	# Mobility models published	3	2	3	6	3	1	4	6	3	8	3	1	1
Data	Public transport check-ins												1	
	Social appl. data								1	1				1
	Cellular network data								3	1	3		1	
	GPS data							1	2	1	2	1	1	
	Bluetooth data				1	1	1		1	1		1		
	Activity/mobility statistics				2		1	1	1		1	1		
	WLAN data			2	2	1	1				1			
	Census data		1							1	2			
	Generic	3	1	1	2	1		2	1		4	1		
Distinguishing features	Pedestrian dynamics	2						1			1	1		
	Source/destination	1	1		1		1		1	1	1		1	
	Land use	1	1					1	1		1			
	Spatial gravity		1	1	2	1		1	1		3			
	Sojourn pattern			2	2						1	1		
	Social gravity			1					1					
	Interpolation				2						1			
	Agenda/activity				1		1	1			1	1		
	Social communities/ties				1			1	1	1	1			
	Contact process				1	1			1					
	Location preferences					1		1	1	1				1
	Periodicity					1			1	1				1
	Fine-grained movement							1		1	1	1		
	Macroscopic dynamics								1	1	2		1	

In the following, we summarize to what extent the main steps of the mobility model engineering process are reflected in the current literature as well as trends observed (details are discussed in Sections 4–6).

[9]We selected the models with respect to their fundamental novelty compared to models that have been respectively introduced earlier.

Empirical observation of reality. First we note that 16 out of all 44 mobility models we surveyed are not based on any empirical observation. In particular early mobility models are either generic models or rely on nation-wide data collection systems (census data), which provide demographic and topographic data. We further observe that the availability of data for mobility and network behavior studies is a driving force for realistic mobility modeling. Data gathering and filtering are steps of increasing importance in mobility modeling. Already in pioneering works in user behavior mining of network data, user mobility analysis was addressed, e.g., 1999 in the work of Tang and Baker [1999] for a city-wide wireless mesh network. In particular the wide availability of WLAN measurements paved the way for the development of *trace-based* mobility models. Bluetooth-based contact data and social/location-based service data enable modeling of social communities and social ties. However, as Bluetooth data capture only contacts of devices without spatial information, these data are less often employed for modeling (cf. Table IV), but very useful for validating contact statistics produced by a mobility model. Models concerned with macroscopic dynamics on larger areas mostly rely on cellular network data and on census data. With the growing availability of GPS data, finer-grained traces and trace-based models became feasible also on larger scale such as a city, state, or a country. Yet a large enough representative user base for generating these traces has to be assured.

Model creation. State-of-the-art mobility modeling processes as reflected in the current literature do in general not differentiate between model building and model validation. They further do not differentiate between building a baseline mobility model first and extending it later by taking into account specific context information to improve realism. By differentiating, however, we find that modeling decisions taken become clearer, models can be better evaluated, and compared with one another. Regarding realism, 8 of the surveyed models do not include any context, which concerns mostly trace-based models.

Implementations of several models are publicly available. A few simple synthetic models are implemented in general network simulators (e.g., ns-3[10] or OMNet++[11]). More realistic models are implemented in simulators for special types of mobile wireless networks (see, e.g., theONE [Keränen et al. 2009] for opportunistic networks). Moreover, specialized simulators with the main aim of generating mobility traces integrate a wide range of models, such as BonnMotion [Aschenbruck et al. 2010], transport models simulators (MATSIM [Horni et al. 2011], SUMO [Behrisch et al. 2011], etc.), and simulators of global mobility dynamics (e.g., GLEAM [Van den Broeck et al. 2011]). For a detailed listing of available implementations we refer the reader to the survey of Aschenbruck et al. [2011].

Model validation. 15 of the mobility models we have surveyed do not discuss any validation. This often concerns generic models enhancing simple synthetic models. If a validation step is discussed, it is more often targeting the realism of a mobility model (22/44) than its representativeness (11/44). Mobility modeling practice uses visual comparison, statistical test for fitness of the model against real data, and evaluation against other existing models. Both representativeness and realism are explicitly validated for only 5 of the mobility models surveyed.

Guidance for mobile networking researchers (searching for a mobility model). Introducing mobility models along a structured engineering process allows for selecting a model by evaluating the distinct features of candidate models. The evaluation includes rating the selected data sources (and their representativeness for the specific application) and scrutinizing the modeled features of mobility. Finally, it should be evaluated whether the model has been properly validated. The selection process may best start with Figure 3, which relates features to mo-

[10]https://www.nsnam.org/
[11]http://www.omnetpp.org/

bility models, resulting in a set of candidate models that may be evaluated step-by-step. Alternatively, as the validation is a fundamental step, the selection of a mobility model may also start with using Table VIII, or with any other "step". It is important to note that the selection of an appropriate mobility model – or the need to create a new model – depends on the application context, which determines the modeling aspects of interest as well as those that can be disregarded for reasons of simplicity and negligible impact on the performance of the studied network. For example, in opportunistic networks, contact characteristics are more important than spatial characteristics, whereas for location-based routing algorithms, realistic spatial mobility modeling is essential.

8. CONCLUSIONS

We surveyed and reviewed the recent literature on mobility modeling dedicated to mobile networking. More than 40 models for human mobility have been proposed in the last decade providing a mature understanding of human mobility. Yet structured creation of mobility models has been missing, which we addressed by a modeling framework for engineering realistic models of human mobility. The modeling framework formalizes a process that is often implicitly carried out in existing data-driven mobility modeling efforts. Intended to provide guidance and not to set up a strict set of rules, the framework raises awareness of the important aspects of a model's representativeness and realism, with the aim to increase the validity and comparability of mobility models.

Our analysis reveals that in particular the validation step in data-driven modeling requires improvement. About one half of the surveyed mobility models are introduced without an explicit discussion of model validation, neither in terms of representativeness nor in terms of realism. The introduction of *benchmark traces* for validation is one option to counteract the lack of explicit model validation and make mobility models comparable.

Yet an upcoming challenge with respect to traces used for model creation and benchmark traces is to keep them up-to-date. Wireless user devices have become increasingly mobile, leading to outdated traces as in the case of WLAN traces of more stationary notebooks that are not representative for smartphone mobility. This trend will continue for wearable devices that exhibit an even higher degree of mobility, which will cause additional demand for adapting existing data sets.

REFERENCES

Osman Abul, Francesco Bonchi, and Mirco Nanni. 2008. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*. IEEE Computer Society, Washington, DC, USA, 376–385. DOI:http://dx.doi.org/10.1109/ICDE.2008.4497446

Nils Aschenbruck, Raphael Ernst, Elmar Gerhards-Padilla, and Matthias Schwamborn. 2010. BonnMotion: a mobility scenario generation and analysis tool. In *Proceedings of the International ICST Conference on Simulation Tools and Techniques (SIMUTools)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, Article 51, 10 pages. DOI:http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8684

Nils Aschenbruck, Aarti Munjal, and Tracy Camp. 2011. Trace-based mobility modeling for multi-hop wireless networks. *Computer Communications* 34, 6 (2011), 704–714. DOI:http://dx.doi.org/10.1016/j.comcom.2010.11.002

Dharmayashdev Rai Basgeet, Pascal Dugenie, Alistair Munro, Dritan Kaleshi, and J. Irvine. 2003. SMM: mathematical framework of a scalable mobility model. In *Proceedings of the 6th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWIM)*. ACM, New York, NY, USA, 74–81. DOI:http://dx.doi.org/10.1145/940991.941005

Michael Behrisch, Laura Bieker, Jakob Erdmann, and Daniel Krajzewicz. 2011. SUMO–Simulation of Urban MObility. In *Proceedings of the Third International Conference on Advances in System Simulation (SIMUL)*.

Christian Bettstetter, Giovanni Resta, and Paolo Santi. 2003. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *IEEE Transactions on Mobile Computing* 2, 3 (2003), 257–269. DOI:http://dx.doi.org/10.1109/TMC.2003.1233531

Vincent D. Blondel, Markus Esch, Connie Chan, Fabrice Clérot, Pierre Deville, Etienne Huens, Frédéric Morlot, Zbigniew Smoreda, and Cezary Ziemlicki. 2012. Data for Development: the D4D Challenge on Mobile Phone Data. *CoRR* abs/1210.0137 (2012), 1–10. http://arxiv.org/abs/1210.0137

Victor J. Blue and Jeffrey L. Adler. 2001. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B: Methodological* 35, 3 (2001), 293 – 312. DOI:http://dx.doi.org/10.1016/S0191-2615(99)00052-1

Chiara Boldrini, Marco Conti, and Andrea Passarella. 2009. The sociable traveller: human travelling patterns in social-based mobility. In *Proceedings of the ACM International Symposium on Mobility Management and Wireless Access (MobiWAC)*. ACM, New York, NY, USA, 34–41. DOI:http://dx.doi.org/10.1145/1641776.1641782

Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. 1998. A Performance Comparison of Multi-hop Wireless Ad-hoc Network Routing Protocols. In *Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*. ACM, New York, NY, USA, 85–97. DOI:http://dx.doi.org/10.1145/288235.288256

Francesco Calabrese, Massimo Colonna, Piero Lovisolo, Dario Parata, and Carlo Ratti. 2011. Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 141–151. DOI:http://dx.doi.org/10.1109/TITS.2010.2074196

Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2011. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *Pervasive Computing* 10, 4 (2011), 36–44. DOI:http://dx.doi.org/10.1109/MPRV.2011.41

Francesco Calabrese, Giusy Di Lorenzo, and Carlo Ratti. 2010. Human mobility prediction based on individual and collective geographical preferences. In *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 312–317. DOI:http://dx.doi.org/10.1109/ITSC.2010.5625119

Tracy Camp, Jeff Boleng, and Vanessa Davies. 2002. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing* 2, 5 (2002), 483–502.

Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. 2007. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing* 6 (2007), 606–620. DOI:http://dx.doi.org/10.1109/TMC.2007.1060

Yung-Chih Chen, J. Kurose, and D. Towsley. 2012. A mixed queueing network model of mobility in a campus wireless network. In *Proceedings of the IEEE INFOCOM*. IEEE, 2656–2660. DOI:http://dx.doi.org/10.1109/INFCOM.2012.6195673

Zhuo Chen, Lu Wang, and Nelson H.C. Yung. 2011. Adaptive Human Motion Analysis and Prediction. *Pattern Recognition* 44, 12 (2011), 2902–2914. DOI:http://dx.doi.org/10.1016/j.patcog.2011.04.022

Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, New York, NY, USA, 1082–1090. DOI:http://dx.doi.org/10.1145/2020408.2020579

Marco Conti, Silvia Giordano, Martin May, and Andrea Passarella. 2010. From opportunistic networks to opportunistic computing. *IEEE Communications Magazine* 48, 9 (2010), 126–139. DOI:http://dx.doi.org/10.1109/MCOM.2010.5560597

Caitlin Cottrill, Francisco Pereira, Fang Zhao, Inas Dias, Hock Lim, Moshe Ben-Akiva, and P. Christopher Zegras. 2013. Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board* 2354 (2013), 59–67. DOI:http://dx.doi.org/10.3141/2354-07

Yves-Alexandre de Montjoye, Zbigniew Smoreda, Romain Trinquart, Cezary Ziemlicki, and Vincent D. Blondel. 2014. D4D-Senegal: The Second Mobile Phone Data for Development Challenge. *CoRR* abs/1407.4885 (2014), 1–11.

Michael Solomon Desta, Esa Hyytiä, Jörg Ott, and Jussi Kangasharju. 2013. Characterizing content sharing properties for mobile users in open city squares. In *Proceedings of the 10th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*. IEEE, Banff, Alberta, Canada, 147–154. DOI:http://dx.doi.org/10.1109/WONS.2013.6578340

Rex W. Douglass, David A. Meyer, Megha Ram, David Rideout, and Dongjin Song. 2015. High resolution population estimates from telecommunications data. 4, 1 (2015), 1–13. DOI:http://dx.doi.org/10.1140/epjds\%252fs13688-015-0040-6

Zhenguo Du, Peilin Hong, and Kaiping Xue. 2012. Hotspot Gravitation Mobility Model. *IEEE Communications Letters* 16, 2 (2012), 193–195. DOI:http://dx.doi.org/10.1109/LCOMM.2011.111611.111272

Matt Duckham and Lars Kulik. 2005. A Formal Model of Obfuscation and Negotiation for Location Privacy. In *Proceedings of the Third International Conference on Pervasive Computing (PERVASIVE)*. Springer, 152–170. DOI:http://dx.doi.org/10.1007/11428572_10

Nathan Eagle and Alex Pentland. 2006. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.* 10, 4 (2006), 255–268. DOI:http://dx.doi.org/10.1007/s00779-005-0046-3

Frans Ekman, Ari Keränen, Jouni Karvo, and Jörg Ott. 2008. Working Day Movement Model. In *Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models*. ACM, New York, NY, USA, 33–40. DOI:http://dx.doi.org/10.1145/1374688.1374695

Suzanne P. Evans. 1973. A relationship between the gravity model for trip distribution and the transportation problem in linear programming. *Transportation Research* 7, 1 (1973), 39–61. DOI:http://dx.doi.org/10.1016/0041-1647(73)90005-1

Steven Farber, Tijs Neutens, Harvey J. Miller, and Xiao Li. 2013. The Social Interaction Potential of Metropolitan Regions: A Time-Geographic Measurement Approach Using Joint Accessibility. *Annals of the Association of American Geographers* 103, 3 (2013), 483–504. DOI:http://dx.doi.org/10.1080/00045608.2012.689238

Domenico Ferrari, Giuseppe Serazzi, and Alessandro Zeigner. 1983. *Measurement and Tuning of Computer Systems*. Prentice Hall, Englewood Cliffs, New Jersey. 523 pages.

Michal Ficek and Lukas Kencl. 2012. Inter-Call Mobility model: A spatio-temporal refinement of Call Data Records using a Gaussian mixture model. In *Proceedings of the IEEE INFOCOM*. IEEE, 469–477. DOI:http://dx.doi.org/10.1109/INFCOM.2012.6195786

Marco Fiore, Jérôme Härri, Fethi Filali, and Christian Bonnet. 2007. Understanding Vehicular Mobility in Network Simulation. In *IEEE 4th International Conference on Mobile Adhoc and Sensor Systems (MASS)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–6. DOI:http://dx.doi.org/10.1109/MOBHOC.2007.4428738

Morgan R. Frank, Lewis Mitchell, Peter Sheridan Dodds, and Christopher M. Danforth. 2013. Happiness and the Patterns of Life: A Study of Geolocated Tweets. *Scientific Reports* 3 (Sept 2013), 2625. http://dx.doi.org/10.1038/srep02625

Adriano Galati, Karim Djemame, and Chris Greenhalgh. 2013. A mobility model for shopping mall environments founded on real traces. *Networking Science* 2, 1-2 (2013), 1–11. DOI:http://dx.doi.org/10.1007/s13119-012-0011-1

Olof Görnerup, Nima Dokoohaki, and Andrea Hess. 2015. Privacy-preserving mining of frequent routes in cellular network data. In *Proceedings of the IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*.

Matthias Grossglauser and D.N.C. Tse. 2002. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking* 10, 4 (2002), 477–486. DOI:http://dx.doi.org/10.1109/TNET.2002.801403

Jérôme Härri, Fethi Filali, and Christian Bonnet. 2009. Mobility models for vehicular ad hoc networks: a survey and taxonomy. *IEEE Communications Surveys Tutorials* 11, 4 (2009), 19–41. DOI:http://dx.doi.org/10.1109/SURV.2009.090403

Dirk Helbing and Péter Molnár. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* 51 (1995), 4282–4286. Issue 5. DOI:http://dx.doi.org/10.1103/PhysRevE.51.4282

David A. Hensher and Kenneth J. Button. 2008. *Handbook of Transport Modelling*. Elsevier.

John D. Herbert and Benjamin H. Stevens. 1960. A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science* 2, 2 (1960), 21–36. DOI:http://dx.doi.org/10.1111/j.1467-9787.1960.tb00838.x

Andreas Horni, Kai Nagel, and Kay W Axhausen. 2011. *High-resolution destination choice in agent-based demand models*. Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme.

Wei-Jen Hsu, Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Ahmed Helmy. 2009. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE/ACM Trans. Netw.* 17, 5 (2009), 1564–1577. DOI:http://dx.doi.org/10.1109/TNET.2008.2011128

Karin Anna Hummel and Andrea Hess. 2013. Movement Activity Estimation and Forwarding Effects for Opportunistic Networking based on Urban Mobility Traces. *Wireless Communications and Mobile Computing* 13, 3 (2013), 343–360. DOI:http://dx.doi.org/10.1002/wcm.2216

Esa Hyytiä, Pasi E. Lassila, and Jorma T. Virtamo. 2006. A Markovian Waypoint Mobility Model with Application to Hotspot Modeling. In *Proceedings of the IEEE International Conference on Communications (ICC)*, Vol. 3. IEEE, 979–986. DOI:http://dx.doi.org/10.1109/ICC.2006.254875

Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, New York, NY, USA, 239–252. DOI:http://dx.doi.org/10.1145/2307636.2307659

Ravi Jain, Dan Lelescu, and Mahadevan Balakrishnan. 2005. Model T: an empirical model for user registration patterns in a campus wireless LAN. In *Proceedings of the 11th annual international conference on mobile computing and networking (MobiCom)*. ACM, New York, NY, USA, 170–184. DOI:http://dx.doi.org/10.1145/1080829.1080848

Andreas Janecek, Danilo Valerio, Karin Anna Hummel, Fabio Ricciato, and Helmut Hlavacs. 2012. Cellular data meet vehicular traffic theory: location area updates and cell transitions for travel time estimation. In *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp)*. ACM, New York, NY, USA, 361–370. DOI:http://dx.doi.org/10.1145/2370216.2370272

Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, and Subhash Suri. 2003. Towards realistic mobility models for mobile ad hoc networks. In *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, New York, NY, USA, 217–229. DOI:http://dx.doi.org/10.1145/938985.939008

David B. Johnson and David A. Maltz. 1996. Dynamic Source Routing in Ad-hoc Wireless Networks. In *Mobile Computing*, T. Imielinski and H. Korth (Eds.). Kluwer Academic Publishers, Chapter 5, 153–181.

Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. 2015. Understanding Human Mobility from Twitter. *PLoS ONE* 10, 7 (07 2015), e0131469. DOI:http://dx.doi.org/10.1371/journal.pone.0131469

Dmytro Karamshuk, Chiara Boldrini, Marco Conti, and Andrea Passarella. 2011. Human mobility models for opportunistic networks. *IEEE Communications Magazine* 49, 12 (2011), 157–165. DOI:http://dx.doi.org/10.1109/MCOM.2011.6094021

Dmytro Karamshuk, Chiara Boldrini, Marco Conti, and Andrea Passarella. 2012. An Arrival-based Framework for Human Mobility Modeling. In *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 1–9. DOI:http://dx.doi.org/10.1109/WoWMoM.2012.6263683

Ari Keränen, Jörg Ott, and Teemu Kärkkäinen. 2009. The ONE Simulator for DTN Protocol Evaluation. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques (Simutools)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, 55:1–55:10. DOI:http://dx.doi.org/10.4108/ICST.SIMUTOOLS2009.5674

Jonghyun Kim, Vinay Sridhara, and Stephan Bohacek. 2009. Realistic mobility simulation of urban mesh networks. *Ad Hoc Netw.* 7, 2 (2009), 411–430. DOI:http://dx.doi.org/10.1016/j.adhoc.2008.04.008

Minkyong Kim, David Kotz, and Songkuk Kim. 2006. Extracting a Mobility Model from Real User Traces. In *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 1–13. DOI:http://dx.doi.org/10.1109/INFOCOM.2006.173

Tetsuo Kobayashi, Harvey J. Miller, and Walied Othman. 2011. Analytical methods for error propagation in planar space-time prisms. *Journal of Geographical Systems* 13, 4 (2011), 327–354. DOI:http://dx.doi.org/10.1007/s10109-010-0139-z

Sokol Kosta, Alessandro Mei, and Julinda Stefa. 2010. Small World in Motion (SWIM): Modeling Communities in Ad-Hoc Mobile Networking. In *Proceedings of the IEEE Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*. IEEE, 1–9. DOI:http://dx.doi.org/10.1109/SECON.2010.5508278

Thomas Kunz, Atif A. Siddiqi, and John Scourias. 2001. The peril of evaluating location management proposals through simulations. *Wireless Networks* 7, 6 (2001), 635–643. DOI:http://dx.doi.org/10.1023/A:1012375503129

Patrick Laube, Matt Duckham, and Thomas Wolle. 2008. Decentralized Movement Pattern Detection amongst Mobile Geosensor Nodes. In *Proceedings of the 5th International Conference Geographic Information Science (GIScience)*. 199–216. http://dx.doi.org/10.1007/978-3-540-87473-7_13

Kyunghan Lee, Seongik Hong, Seong Joon Kim, Injong Rhee, and Song Chong. 2009. SLAW: A New Mobility Model for Human Walks. In *Proceedings of the 28th IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 855–863. DOI:http://dx.doi.org/10.1109/INFCOM.2009.5061995

Kyunghan Lee, Seongik Hong, Seong Joon Kim, Injong Rhee, and Song Chong. 2012. SLAW: Self-similar Least-Action human Walk. *IEEE/ACM Trans. Netw.* 20, 2 (April 2012), 515–529. DOI:http://dx.doi.org/10.1109/TNET.2011.2172984

John G. Markoulidakis, George L. Lyberopoulos, Dimitrios F. Tsirkas, and Efstathios D. Sykas. 1997. Mobility modeling in third-generation mobile telecommunications systems. *IEEE Personal Communications* 4 (1997), 41–56.

Christoph P. Mayer and Oliver P. Waldhorst. 2011. On the Impact of Graph Structure on Mobility in Opportunistic Mobile Networks. In *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 882–887.

Randy L. Mayes. 2009. Developing Adequacy Criterion for Model Validation Based on Requirements. In *Proceedings of the IMAC-XXVII: Conference & Exposition on Structural Dynamics*.

Harald Meyer, Oscar Trullols-Cruces, Andrea Hess, Karin A. Hummel, José M. Barceló-Ordinas, Claudio E. Casetti, and Gunnar Karlsson. 2011. VANET Mobility Modeling Challenged by Feedback Loops. In *Proceedings of the 10th IFIP Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*. IEEE, 95–102. DOI:http://dx.doi.org/10.1109/Med-Hoc-Net.2011.5970499

Harvey J. Miller. 2008. Time Geography. In *Encyclopedia of GIS*. Springer US, 1151–1156. DOI:http://dx.doi.org/10.1007/978-0-387-35973-1_1383

Darakhshan J. Mir, Sibren Isaacman, Ramon Caceres, Margaret Martonosi, and Rebecca N. Wright. 2013. DP-WHERE: Differentially private modeling of human mobility. In *Proceedings of the IEEE International Conference on Big Data*. 580–588.

Frédéric Morlot, Salah-Eddine Elayoubi, and François Baccelli. 2010. An Interaction-Based Mobility Model for Dynamic Hot Spot Analysis. In *Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2294–2302.

Aarti Munjal, Tracy Camp, and William C. Navidi. 2011. SMOOTH: a simple way to model human mobility. In *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*. ACM, New York, NY, USA, 351–360. DOI:http://dx.doi.org/10.1145/2068897.2068957

Mirco Musolesi and Cecilia Mascolo. 2006. A community based mobility model for ad hoc network research. In *Proceedings of the 2nd International Workshop on Multi-hop Ad Hoc Networks: From Theory to Reality (REALMAN)*. ACM, 31–38. DOI:http://dx.doi.org/10.1145/1132983.1132990

Mirco Musolesi and Cecilia Mascolo. 2009. Mobility Models for Systems Evaluation. A Survey. In *Middleware for Network Eccentric and Mobile Applications*, Benoit Garbinato, Hugo Miranda, and Luis Rodrigues (Eds.). Springer, 43–62.

Samuel C. Nelson, Albert F. Harris, III, and Robin Kravets. 2007. Event-driven, role-based mobility in disaster recovery networks. In *Proceedings of the 2nd ACM Workshop on Challenged Networks (CHANTS)*. ACM, New York, NY, USA, 27–34. DOI:http://dx.doi.org/10.1145/1287791.1287798

Tijs Neutens, Tim Schwanen, and Harvey J. Miller. 2010. Dealing with Timing and Synchronization in Opportunities for Joint Activity Participation. *Geographical Analysis* 42, 3 (2010), 245–266. DOI:http://dx.doi.org/10.1111/j.1538-4632.2010.00792.x

Calvin Newport, David Kotz, Yougu Yuan, Robert S. Gray, Jason Liu, and Chip Elliott. 2007. Experimental Evaluation of Wireless Simulation Assumptions. *SIMULATION: Transactions of The Society for Modeling and Simulation International* 83, 9 (2007), 643–661. DOI:http://dx.doi.org/10.1177/0037549707085632

Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Song Chong. 2008. On the Levy-Walk Nature of Human Mobility. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 924–932. DOI:http://dx.doi.org/10.1109/INFOCOM.2008.145

Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seongjoon Kim, and Song Chong. 2009. CRAWDAD data set ncsu/mobilitymodels (v. 2009-07-23). Downloaded from http://crawdad.org/ncsu/mobilitymodels/. (2009).

Michele Rossi, Leonardo Badia, Nicola Bui, and Michele Zorzi. 2005. On group mobility patterns and their exploitation to logically aggregate terminals in wireless networks. In *Proceedings of the IEEE 62nd Vehicular Technology Conference (VTC-2005-Fall)*, Vol. 3. IEEE, 1863–1867. DOI:http://dx.doi.org/10.1109/VETECF.2005.1558429

Radhika Ranjan Roy. 2011. *Handbook of Mobile Ad Hoc Networks for Mobility Models*. Springer.

Paolo Santi. 2012. *Mobility Models for Next Generation Wireless Networks: Ad Hoc, Vehicular and Mesh Networks*. Wiley.

Matthias Schwamborn, Nils Aschenbruck, and Peter Martini. 2010. A realistic trace-based mobility model for first responder scenarios. In *Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWIM)*. ACM, New York, NY, USA, 266–274.

DOI:http://dx.doi.org/10.1145/1868521.1868564

Filippo Simini, Marta C. Gonzalez, Amos Maritan, and Albert-Laszlo Barabasi. 2012. A universal model for mobility and migration patterns. *Nature* 484 (2012), 96–100. DOI:http://dx.doi.org/10.1038/nature10856

Catherine Soanes and Angus Stevenson (Eds.). 2005. *The Oxford Dictionary of English (revised edition)*. Oxford University Press.

Chaoming Song, Tal Koren, Pu Wang, and Albert-Laszlo Barabasi. 2010. Modelling the Scaling Properties of Human Mobility. *Nature Physics* 6, 10 (2010), 818–823. DOI:http://dx.doi.org/10.1038/nphys1760

Vikram Srinivasan, Mehul Motani, and Wei Tsang Ooi. 2006. Analysis and implications of student contact patterns derived from campus schedules. In *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, 86–97. DOI:http://dx.doi.org/10.1145/1161089.1161100

Herbert Stachowiak. 1973. *Allgemeine Modelltheorie*. Springer, Vienna – New York.

Diane Tang and Mary Baker. 1999. Analysis of a metropolitan-area wireless network. In *Proceedings of the ACM/IEEE international conference on mobile computing and networking (MobiCom)*. ACM, New York, NY, USA, 13–23. DOI:http://dx.doi.org/10.1145/313451.313460

Joanne Treurniet. 2014. A Taxonomy and Survey of Microscopic Mobility Models from the Mobile Networking Domain. *ACM Comput. Surv.* 47, 1, Article 14 (2014), 32 pages. DOI:http://dx.doi.org/10.1145/2616973

Cristian Tuduce and Thomas R. Gross. 2005. A mobility model based on WLAN traces and its validation. In *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, Vol. 1. IEEE, 664–674. DOI:http://dx.doi.org/10.1109/INFCOM.2005.1497932

Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. 2011. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases* 11, 37 (02/02/2011 2011). DOI:http://dx.doi.org/10.1186/1471-2334-11-37

Dirck van Vliet. 1978. Improved shortest path algorithms for transport networks. *Transportation Research* 12, 1 (1978), 7–20. DOI:http://dx.doi.org/10.1016/0041-1647(78)90102-8

Ryan Vogt, Ioanis Nikolaidis, and Pawel Gburzynski. 2012. A realistic outdoor urban pedestrian mobility model. *Simulation Modelling Practice and Theory* 26 (2012), 113–134. DOI:http://dx.doi.org/10.1016/j.simpat.2012.04.006

Vladimir Vukadinović, Fabian Dreier, and Stefan Mangold. 2014. Impact of Human Mobility on Wireless Ad Hoc Networking in Entertainment Parks. *Ad Hoc Networks* 12 (2014), 17–34. DOI:http://dx.doi.org/10.1016/j.adhoc.2012.05.011

Vladimir Vukadinović, Ólafur Ragnar Helgason, and Gunnar Karlsson. 2009. A mobility model for pedestrian content distribution. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques (Simutools)*. 1–8. DOI:http://dx.doi.org/10.4108/ICST.SIMUTOOLS2009.5645

Duncan J. Watts. 1999. Networks, Dynamics, and the Small-World Phenomenon. *Amer. J. Sociology* 105 (1999), 493–527.

Stephan Winter and Zhang-Cai Yin. 2010. Directed Movements in Probabilistic Time Geography. *International Journal of Geographical Information Science* 24, 9 (2010), 1349–1365. DOI:http://dx.doi.org/10.1080/13658811003619150

Stephan Winter and Zhang-Cai Yin. 2011. The elements of probabilistic time geography. *GeoInformatica* 15, 3 (2011), 417–434. DOI:http://dx.doi.org/10.1007/s10707-010-0108-1

Jihwang Yeo, David Kotz, and Tristan Henderson. 2006. CRAWDAD: a community resource for archiving wireless data at Dartmouth. *SIGCOMM Comput. Commun. Rev.* 36, 2 (2006), 21–22. DOI:http://dx.doi.org/10.1145/1129582.1129588

Jungkeun Yoon, Mingyan Liu, and Brain Noble. 2003. Random waypoint considered harmful. In *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM)*, Vol. 2. IEEE, 1312–1321. DOI:http://dx.doi.org/10.1109/INFCOM.2003.1208967

Jungkeun Yoon, Brian D. Noble, Mingyan Liu, and Minkyong Kim. 2006. Building Realistic Mobility Models from Coarse-grained Traces. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services (MobiSys)*. ACM, New York, NY, USA, 177–190. DOI:http://dx.doi.org/10.1145/1134680.1134699

Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring Human Mobility with Multi-source Data at Extremely Large Metropolitan Scales. In *Proceedings of the International Conference on Mobile Computing and Networking (MobiCom)*. ACM, New York, NY, USA, 201–212. DOI:http://dx.doi.org/10.1145/2639108.2639116

Chen Zhao and M.L. Sichitiu. 2010. N-Body: Social Based Mobility Model for Wireless Ad Hoc Network Research. In *Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*. 1–9. DOI:http://dx.doi.org/10.1109/SECON.2010.5508280

Qunwei Zheng, Xiaoyan Hong, and Jun Liu. 2006. An Agenda Based Mobility Model. In *Proceedings of the 39th Annual Symposium on Simulation*. IEEE Computer Society, 188–195. DOI:http://dx.doi.org/10.1109/ANSS.2006.11

Received .; revised .; accepted .