Model-based Assessment of Aspects of User-satisfaction in Bicycle Sharing Systems

Mieke Massink
CNR-Institute of Information Science and Technologies "Alessandro Faedo"
Pisa, Italy
Email: mieke.massink@isti.cnr.it

Rytis Paškauskas
CNR-Institute of Information Science and Technologies "Alessandro Faedo"
Pisa, Italy
Email: rytis.paskauskas@gmail.com

***Abstract*—Modelling of bike sharing systems as Markov Renewal Process is examined with the aim of capturing and assessing various forms of user (dis)satisfaction. A class of models with minimal assumptions about distributions of bicycle parking stations and service requests is developed in which rational commuter behaviour is taken into account. Stochastic time evolution of these models is studied, using a generalised version of Gillespie's exact stochastic integration algorithm that accounts for non-Markovian inter-event times. The model is shown to reproduce quite faithfully the trip-duration statistics of smaller and larger real bike sharing systems, such as those in London and Pisa, including the algebraic 'tails' of these distributions that are made up of longer cycling trips. The latter are related to user's difficulties to find suitable parking places therefore to a potential source of distress. The model also predicts other salient features such as a mode at 10 minutes and crossover behaviour at about 30 minutes. The framework can be extended to include measures either designed to improve, or anyway to affect, the user experience with a system, such as incentives for spontaneous vehicle redistribution. User satisfaction is difficult to assess in real systems because these naturally collect only data about trips that actually, and thus successfully, took place giving only partial and biased insight in user satisfaction.**

I. Introduction

Smart bicycle sharing is a form of public transport that provides short-term self-service bicycle hiring [1], [2]. It has evolved a long way from the early ideas dating back to the sixties. Recent popularity of bike-sharing gained momentum with the introduction of information and smart card technologies allowing vehicle tracking. Today, hundreds of cities worldwide have such programs, operating up to tens of thousands of vehicles and thousands of docking stations (e.g. Hangzhou or Paris[1]). In those cities, smart bike-sharing has become a reliable mode of public transport, welcomed by the general public for its dependability and bicycle's environmental, societal, and health benefits [3]. However, smart bike-sharing programs raise multiple issues concerning their carbon footprint [4], integration with other modes of public transport, choosing proper service features [5], and understanding the effects of user incentives [6], to mention a few.

In developed urban environments, the question that potential users of any public form of transport will be asking themselves increasingly often is 'which' mode of public transport to rely on rather than 'if' they will use public transport. This concerns smart bike-sharing too, since the majority of cycling trips in cities could also be made by a combination of walking and other modes of public transport (cf. [2, Fig. 4]), or private bike. The successful running of multiple public transport services may in the long term be determined not only by proper top-down planning, but also by the cumulative effects of 'micro-decisions' by the public, as the example of the bike-sharing system in Melbourne [7] suggests. Being able to evaluate the balance between services and policies could in the long run determine the success of some programs.

It is notoriously difficult to evaluate user satisfaction from the available data collected by a system. Typical bike-sharing data consists of static parameters of stations and fluctuating numbers of parked vehicles. In many cases, vehicle identification numbers are also available, which allow to relate the hiring and the corresponding returning events, and visualise dominant spatiotemporal flows [8], [9], [10]. Naturally, this data concerns only such trips that actually, and thus successfully took place, and raises the issue of missing information about the users who chose an alternative service. Moreover, even the successful trip data concerns only the middle part of the 'walk-cycle-walk' travel cycle. The missing links conceal the trip–objective relation, which is important to the evaluation of a system from the service efficiency perspective. Useful additional insights can be had using alternative approaches such as [11], although they are prone to similar bias issues.

The main contribution of this paper is to show that a *model based approach*, taking into account certain minimal assumptions about the user behaviour, can provide complementary insights into the performance of bike-sharing from a users' perspective. This is illustrated by showing that the aggregated cycling time distributions of real bike-sharing systems can be reproduced to a degree without the parameter fitting of real systems, or the use of privacy-sensitive user information. Furthermore, enriching the model in a step-wise manner suggests other generic insights into the multifaceted question of user satisfaction. We explore the interpretation of cycling trip durations in a manner akin to that of an 'actuary'. Statistics of human life's duration follow a certain probability law.[2] The subtle features of the probability density function (PDF), especially in the so-called 'tail' of extreme events, are oftentimes of most interest, since they have important consequences to the conditional expected life-times, and to the assessment of long

[1]Institute for Transportation and Development Policy of China, http://www.publicbike.net/defaulten.aspx; Vélib', http://www.velib.paris.fr

[2]approximated by, for example, the Gompertz-Makeham model.

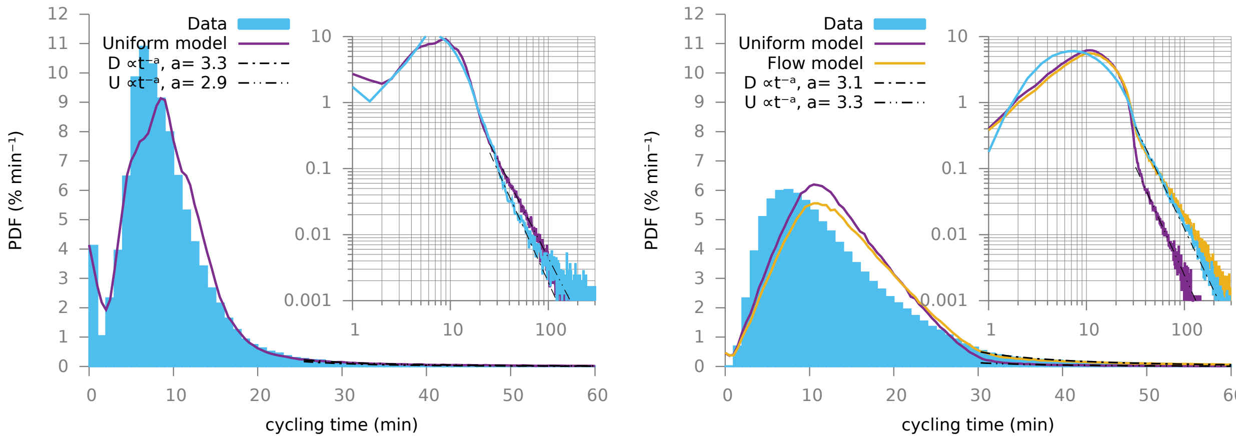

Fig. 1. Cycling duration histograms (Data) in London (right) and Pisa (left), using 831,754 trip records in October 2012, and 242,248 trips, made between June 2013 and May 2015, respectively, and the simulation results for a uniform model (dark lines) and a flow model (light lines). Maintenance trips are excluded.

term risks. In this article we will compare two data-sets, each consisting of cycling trips (excluding maintenance), made by users of bike-sharing systems in two cities: London (UK) and Pisa (Italy). As is evident from Fig. 1 (Data), the cycling time PDFs share remarkable similarities, such as a mode at about 10 minutes, or an algebraic 'tail' beyond 30 minutes (cf. [9, Fig. 3] for analogous PDF from Lyon). The latter part of the PDF is well approximated by $f(t) \propto t^{-a}$, with an exponent $a > 0$. Validity of this law spans nearly three decades and appears to be limited only by the sample size. In both cases, it predicts cycling trips whose duration exceeds the time necessary to traverse a corresponding city.

The presence of surprisingly long cycling trips and their representative algebraic 'tails' in the PDFs are found in all considered cities, and in data on individual pairs of stations. In this article we provide a theoretical explanation for such domain independence by rational behaviour. For that, we use an agent-based model with the following basic assumption: Bike-sharing users' ultimate concern is to save travel time and to arrive at a planned destination with a high probability just before the planned appointment time. This paradigm is implemented using a rigorous framework of Markov Renewal Processes (MRP) [12], [13], that accomodate the necessary behavioural complexity due to a possibility to use non-Markovian renewal PDFs. The model lends itself to both a quantitative and a qualitative interpretation of the algebraic tails, suggesting the following explanation for the observed data: agents who are aiming at arriving on time at their destinations, when they choose to use bike-sharing, must nevertheless take the risk of, either, not finding a free slot where to park, or no bicycle for hiring, at the expected location. These contingencies lead to unplanned delays. Clearly, if a person cannot predictably make it to the appointment on time, it is going to reflect on her satisfaction with the service.

We will consider both single and multiple agent models. For the latter, varying the agent and station distributions, we will obtain the 'uniform', and the 'flow' models (whose PDFs are also shown in Fig. 1) and relate several differences of these models to the factors affecting user satisfaction with the system. The main insight is that a model could be characterised by some emerging condition, such as the formation of large clusters with completely full stations.

The outline of the paper is as follows. Section II briefly recalls the essence of MRPs, provides a description of the bike-sharing model, including its justification, and describes the simulation method. The results concerning trip duration distributions are presented in Section III, whereas models for Pisa and London are discussed in Section IV. Section V concludes with some further considerations and open issues.

II. The Bike-sharing Model

A. *Markov Renewal Processes*

The Markov Renewal Process (MRP) is a generalisation of a Continuous Time Markov Chain (CTMCs) to non-Markovian events, and non-exponential distributions of inter-event times [12], [13]. In this section we briefly recall MRPs, and motivate their use for the modelling of bike-sharing.

Let $(X, T) = \{X_i, T_i; i \in \mathbb{N}\}$ be a stochastic process in $E^{\mathbb{N}} \times \mathbb{R}_+^{\mathbb{N}}$, where E is some countable set, representing the 'state space', and $\mathbb{R}_+ = [0, \infty)$ represents the time-line of evolution. A MRP is a Kolmogorov model such that for each pair of states $i, j \in E$, the conditional probability is given by

$$\Pr\{X_{n+1} = j, T_{n+1} - T_n \le t \mid X_n = i\} = Q_{ij}(t)\,, \quad (1)$$

where $Q_{ij}(t)$ is a right-continuous, non-decreasing and bounded function, with $Q_{ij}(\infty) \le 1$, and $\sum_j Q_{ij}(\infty) = 1$. A matrix $Q = (Q_{ij}(t); i, j \in E)$ with these properties is called a semi-Markov kernel of (X, T). A matrix $P = (P_{ij})$, whose elements are defined by $P_{ij} = Q_{ij}(\infty)$, is a stochastic matrix, and functions $F_{ij}(t) = Q_{ij}(t)/P_{ij}$, for each $i, j \in E$, are distributions. As a consequence, $X = (X_n; n \in \mathbb{N})$ is a Markov chain (DTMC) with state space E and transition matrix P, i.e. the conditional jump probabilities are

$$\Pr\{X_{n+1} = j \mid X_n = i\} = P_{ij}\,, \quad (2)$$

and the distribution of *sojourn time* in a state i, conditional on a subsequent jump to a state j, is given by $F_{ij}(t)$:

$$\Pr\{T_{n+1} - T_n < t \mid X_n = i, X_{n+1} = j\} = F_{ij}(t). \quad (3)$$

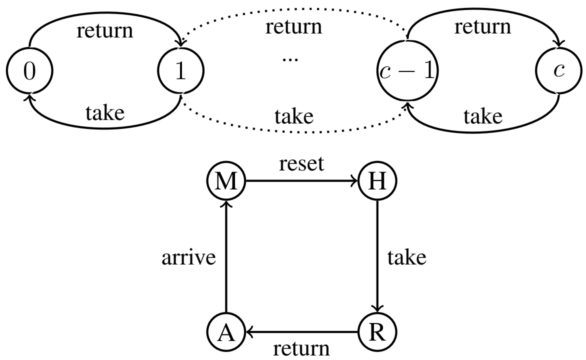

Fig. 2. The automata of a bicycle station (top) and a user-agent (bottom). The state space corresponding to a system of A agents and S stations, is a Cartesian product $E = \prod_{i=1}^{S} \{0, \ldots, c_i\} \times \{\mathrm{H}, \mathrm{R}, \mathrm{A}, \mathrm{M}\}^A$.

Several classes of models of transport (including CTMCs and M/G/1 queueing systems) can be interpreted as special cases of MRPs with a suitable choice of Q [12].

B. Motivation for the use of MRP

An MRP generalises a Markov Processes in two aspects: it provides a mechanism to use arbitrary distributions (and not only exponential ones), and it allows to use transitions, conditioned on a current state *and* on the state to be entered subsequently. These are the main features used in what follows.

We will assume that bike-sharing users are time-conscious people whose decision to use bike-sharing is determined by the concern to save travel time, and to reach their objective at the expected time with high degree of certainty. If we accept this premise then we must also accept that the speed of travel is a major factor in the competitiveness of various modes of transport. To take the speed of travel into account in a stochastic model, it is easy to show that transition rates must be functions of both the current and future states, and that the probability distributions are not exponential. Let the state space represent an 'address book' of all the stations; we may take $E = \{1, \ldots, S\}$ where S is the number of stations, and each index is uniquely associated to some address $\mathbf{x}_i$. For an arbitrary pair of indices $i \neq j$, consider a trip from $\mathbf{x}_i$ to $\mathbf{x}_j$ along a fixed path, traversed at a constant pace p, measured in minutes per kilometre. The duration of this trip is $T = p\,|\mathbf{x}_i - \mathbf{x}_j|$, where $|\cdot|$ is the length of a path, measured in kilometres. This result can be given a probability distribution $F_{ij}(t) = 1_{t > t^a_{ij}}$, where 1_A is an indicator function, equal to one if A is true and zero otherwise, and $t^a_{ij} = p\,|\mathbf{x}_i - \mathbf{x}_j|$ is the so-called *activation time*. Clearly, $F_{ij}(t)$ is not an exponential distribution for any pair of indices, and its parameter t^a_{ij} is a function of the current state i, and a possible future state j. This argument is easily generalised to stochastic travel processes. Any trip between a pair of distant locations (i, j) will take a human traveller at least some finite time $t^a_{ij} > 0$, so that $F_{ij}(t) = 0$ if $t < t^a_{ij}$[3], whereas the exponential distribution is characterised by $t^a_{ij} = 0$, allowing arbitrarily fast travelling.

[3] A bound of t^a_{ij}, obtained by substituting the speed of light for $1/p$ gives absolute certainty, but also much larger bounds, assuming much slower speeds, can be used with near certainty.

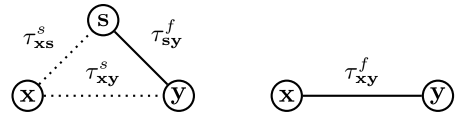

Fig. 3. Spatial (left) and temporal (right) decision criteria of participation.

C. Bike-sharing model

The bike-sharing model is a generalisation of these motivating ideas. It describes a population of agents and bicycle stations, contained in a two-dimensional rectangle representing a populated urban area.

1) Population of bicycle stations: A station is represented by a triple $(n, c, \mathbf{x})$, where n is the number of available bicycles (the *occupation number*), c is the *capacity*, and $\mathbf{x}$ is the geographical coordinate (the *address*) of a given station. An automaton of a typical station with capacity c is shown in Fig. 2. The total number of bikes is $N = \sum_{i=1}^{S} n_i$, and the total capacity is $C = \sum_{i=1}^{S} c_i$. Fixed capacity and instantaneous transaction approximations are assumed throughout.

2) Population of user-agents: An agent combines several human factors pertaining to travelling and decisions. Each agent is parameterised by two addresses that specify its origin and destination locations. The cycling and walking paces, and corresponding rates, are considered as random numbers sampled from a normal distribution. The typical human speeds of 5 km/h for walking, and 12 km/h for cycling, yield paces of 12 min/km, and 5 min/km, respectively, whereas the mean of both rates is set to one (min^{-1}). The agent states (see Fig. 2) are denoted and interpreted as follows: H= $\{$wants to hire a bike$\}$, R= $\{$wants to return a bike$\}$, A= $\{$wants to arrive$\}$, and M= $\{$wants to reset$\}$. A single 'walk-cycle-walk' travel cycle is quantified by a sequence of transitions H$\rightarrow$R$\rightarrow$A$\rightarrow$M, with transition epochs T_1, T_2, T_3, T_4. The total duration of a trip is $T_4 - T_1$, whereas the duration of its cycling part is $T_3 - T_2$. An additional 'mutation' transition M$\rightarrow$H is added to make agents' life cyclic and states recurrent, and allowing continual regeneration of their objectives.

3) Stochastic dynamics: Agents drive the system by spontaneous decisions. There are two types of agent decisions that result in *firing* or *mutation* transitions. Firing transitions are further distinguished as either 'take' or a 'return' transitions. They are synchronised in an obvious manner with two kinds of state-changes occurring at bicycle stations (see Fig. 2). The remaining 'arrive' and 'reset' transitions are *mutation* transitions. They are defined by *not* being synchronised with any station update. Both re-initialise the agent states, the first one resulting in the arrival at a destination, the second one in a complete regeneration of objectives. Agent's objectives can be initialised in various ways. In this article, we consider the so-called *spot commuter* mutation protocol. A spot commuter selects a random new pair of locations and the time until activation, both sampled from appropriate distributions. We remark that other protocols can be easily designed where, for example, the arrival time, rather than the departure time, is a relevant issue. It is important to emphasise that agents can estimate the expected arrival time using Markovian forecasting protocol and, consequently, measure the late arrival time as a difference between the expected and actual arrival epochs.

The model that is used for the travel process is a composition of a *conditional travel process* and a *station utility model*, addressing two major sources of uncertainty of travel in urban environments. The conditional travel process addresses the randomness due to various interactions with the environment. It states that, conditional to fixed end points and a fixed itinerary, the travel time is a stochastic variable that corresponds to a first passage process of a one-dimensional random walk. The so-called renewal distribution of this process is the inverse Gaussian distribution [14] which is further approximated by a delayed exponential distribution $F(t) = 1_{t>t^a}(1-\mathrm{e}^{-r^a(t-t^a)})$. Here, $t^a = p^a d$ is the activation time of an agent a, p^a is its travel pace, d is the distance, and r^a is as arrival rate, related to the diffusion property of a random walk. The station utility model, combined with the assumption of stochastic dynamics, address the decision process under uncertainty. Following the von Neumann-Morgernstern axiomatic approach to the description of such decisions, the existence of *station utility functions* with respect to hiring and returning is posited in the form $u_{ij}(\tau)$, for $i, j \in E$, with a control parameter τ, called the *decision scale* parameter. The latter is a 'motivational' parameter, describing the perceived utility of a station from an agent's point of view, and influencing the agents' decision to return a bike to a particular station. A model for station utility perception is proposed in which the probability of $\{d > x\}$ for large distances x is given by, approximately, $\Pr\{d > x\} \sim \exp -\frac{p_f^2 x^2}{2\tau^2}$. Thus, agents with a larger τ value tend to search for suitable stations in a greater area surrounding their target. However, venturing further away from the destination increases the walking fraction of the trip, so that the total trip duration is likely to increase. On the other hand, smaller values should lead to shorter trips, provided that a suitable station can actually be found within the search area.

Composition of the two models yields a renewal function $F^a_{ij}(t)$ for an agent's a arrival time T at a station j, given the current state i, as

$$\Pr\{T < t\} = F^a_{ij}(t) = 1_{t>t^a_{ij}}\left(1 - \mathrm{e}^{-u_{ij}(\tau^a)r^a(t-t^a_{ij})}\right). \quad (4)$$

4) Elective participation and posterior evaluation: Agents are provided with the capacity to decide whether to accept or reject bike-sharing as a means of achieving their objective, and to measure the effectiveness of a trip in the case of acceptance. They decide whether a bike-sharing trip is a viable alternative to walking by estimating a kind of triangle inequality. Assuming that agents know the distance from the origin ($\mathbf{x}$) to the destination ($\mathbf{y}$) and to the neighbouring stations ($\mathbf{s}$), and their physical parameters in advance, they estimate the expected travel time using a station $\mathbf{s}$ as $\tau^s_{\mathbf{xs}} + \tau^f_{\mathbf{sy}}$, and compare it against the estimated time of walking directly to the destination, $\tau^s_{\mathbf{xy}}$ ('s' and 'f' refer to walking and cycling, respectively, see Fig. 3). A station is accepted as a candidate for a trip if it satisfies the *triangle inequality*

$$\tau^s_{\mathbf{xs}} + \tau^f_{\mathbf{sy}} < \tau^s_{\mathbf{xy}}\,. \quad (5)$$

If at least one station in the network satisfies (5), a bike-sharing trip is accepted, otherwise it is rejected. Agents estimate the cycling time from the origin to destination, $\tau^f_{\mathbf{xy}}$. A trip is accepted only if

$$\tau^f_{\mathbf{xy}} < t_\mathrm{c}\,, \quad (6)$$

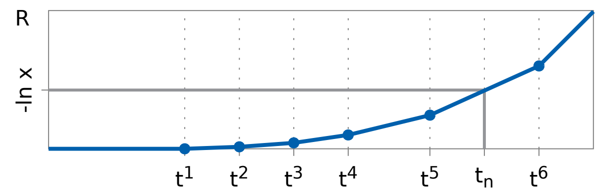

Fig. 4. The first half of the stochastic simulation algorithm determines the sojourn time (t_n) and the participating channels in the next transition (M_n).

where t_c is the *cycling tolerance* parameter.

For an acceptable trip, agents estimate the efficiency of a trip by comparing the actual trip duration $t_\text{tot} = T_4 - T_1$ with the estimated trip duration, had the agent walked the same itinerary, i.e. with $\tau^s = \tau^s_1 + \tau^s_2 + \tau^s_3$ (where indices 1,2,3 denote the consecutive parts of the 'walk-cycle-walk' travel cycle). To decide whether using bike-sharing has been a winning strategy, agents estimate the *efficiency ratio*

$$e = \frac{\tau^s}{t_\text{tot}}\,. \quad (7)$$

Thus agents can decide, retrospectively, whether bike-sharing saved time ($e > 1$) or if it was a waste of time ($e < 1$).

D. Method of simulation

The bike-sharing model describes a stochastic jump process in continuous time (see section II-B). CTMCs are oftentimes simulated using methods that produce statistically exact sample paths [15], [16]. The basic method requires two random number generations per step, one for each member of the pair (X, T). Thus, a statistically exact 'first reaction' of an M-channel CTMC with rates $\lambda_1(X_n)$, ..., $\lambda_M(X_n)$ at the nth step could be determined as follows: The *sojourn time* $t_n = T_{n+1} - T_n$ in a current state X_n is drawn from an exponential distribution $1 - \exp(-\lambda(X_n)t)$, where $\lambda(X_n) = \sum_{i=1}^M \lambda_i(X_n)$, whereas the *next transition channel* m is drawn from a discrete M-point distribution with weights $\hat\lambda_i(X_n)$, $i = 1, \ldots, M$, where $\hat\lambda_i = \lambda_i/\lambda$.[4] A practical algorithm consists of drawing two random numbers $x, y \sim U(0,1)$ from a uniform distribution U, then letting $t_n = -\frac{1}{\lambda(X_n)}\ln x$, and letting m be such that $\hat\lambda_1 \le \cdots \le \sum_{i=1}^m \hat\lambda_i \le y \le \sum_{i=1}^{m+1}\hat\lambda_i \le 1$[15], [16], [17]. Alternative exact methods [18], and other, exact and approximate variations exist to address specific issues viz. multiple time-scales (see, e.g. [17] and references therein).

Although the bike-sharing model is non-Markovian, this CTMC methodology can be adapted to yield a statistically exact simulation of it. It models each trip as an event with its own transition channel. Table I provides the estimates of event frequencies in real systems. Specifically, each agent-station pair is considered as a possible transition channel (thus M can be quite large, and vary in time). The total transition rate in a state X is obtained as a sum of rates from all channels, the rate for each being $r^a_{ij}(t;\tau^a) = 1_{t>t^a_{ij}} u_{ij}(\tau^a)p^a$. Note that although a one agent – one station pair distribution is (a delayed) exponential, a system of many stations or agents

[4] interpreted as rolling an M-faced 'loaded die' with weights $(\hat\lambda_i)$.

gives a non-exponential distribution of the form $1 - \mathrm{e}^{-R(t)}$. The phase, $R(t)$, defined as an integral over the total rate, $R(t) = \int_0^t \sum r_{ij}^a(t';\tau^a)\mathrm{d}t'$, is illustrated in Fig. 4. The main steps of the analogous 'first reaction' algorithm are as follows. A random number $x \sim U(0,1)$ is drawn and the sojourn time t_n is solved for from $-\ln x = R(t_n)$ (see Fig. 4). The number (M_n) and identities of channels, participating in the next reaction, are defined as all the channels with the activation times satisfying $t_{ij}^a \leq t_n$ ($M_n = 5$ in Fig. 4). The first reaction channel is then determined, using the second draw of a random number, and a discrete M_n-point distribution, as before.

III. Basic predictions and analysis

A. Metrics

Although the trip duration comprises the primary data that is directly measured and modelled, there are features that better describe a system's functioning than trip durations. We consider the following three metrics.

1) Trip efficiency metric: The efficiency or gain of a trip is defined by (7). It is a measure of how useful a particular trip is to an agent, as compared with the same trip made by walking: bike-sharing trips with $e > 1$ save time whereas trips with $e < 1$ are a waste of time.

2) Excluded population metric: Agent locations are sampled with the assumption that each agent is interested in using bike-sharing. However, an agent uses bike-sharing only if the triangle inequalities (5), and the cycling tolerance inequality (6), are satisfied. The fraction of all agents who, based on the union of these inequalities do not, or cannot choose bike-sharing (because of empty stations), is called the excluded population metric (EPM).

3) Congestion metric: In a queueing model approach to bike-sharing, Fricker and Gast [6] identify completely empty or full stations as problematic because they inhibit one direction of traffic. To relate their work to ours, we introduce the bicycle and the slot congestion as follows,

$$p_t^+ = \frac{1}{tS}\int_0^t \sum_{i=1}^{S} 1_{\{s_i(t)=c_i\}},\; p_t^- = \frac{1}{tS}\int_0^t \sum_{i=1}^{S} 1_{\{s_i(t)=0\}}\,, \tag{8}$$

Since both parameters range from $p_t^\pm = 0$ (no station is ever full/empty), to $p_t^\pm = 1$ (all stations are always full/empty) they give an average measure of 'problematic stations' in the sense of [6]. Note that $p_t^+ + p_t^- \leq 1$ for all t.

B. Model parameters

The results in this section concern a single agent model (except in section III-G where a multi-agent model is used) with 6×6 stations with $c_i = 20$ and an initial filling degree $N/C \approx 0.5$, placed in a 3×3 km area (similar to Fig. 9), and random initial configuration of bicycles in stations.

C. Decisions determining travel efficiency

The trip efficiency metric (7) provides insights into the effectiveness of bike-sharing trips. As shown in Fig. 5, there exists a compromise value of τ that minimises both the median of total trip times, and the scatter of their distribution, and

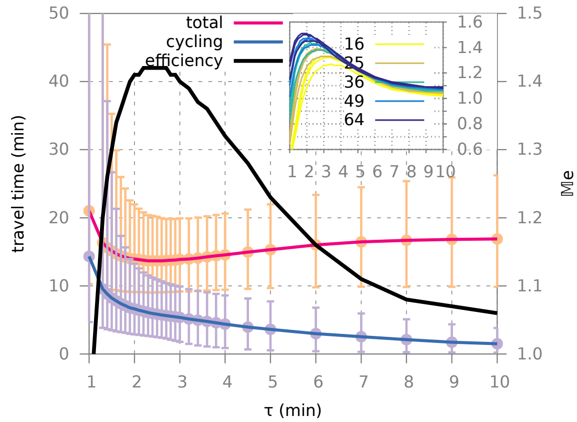

Fig. 5. The decision scale $\tau = 2.5$ yields largest efficiency, shortest total trips, and the best confidence (central 80% percentile). The median efficiency curves are qualitatively similar, irrespective of the system size (S, legend) or the station filling $N/C = 0.25, 0.5, 0.75$ (inset).

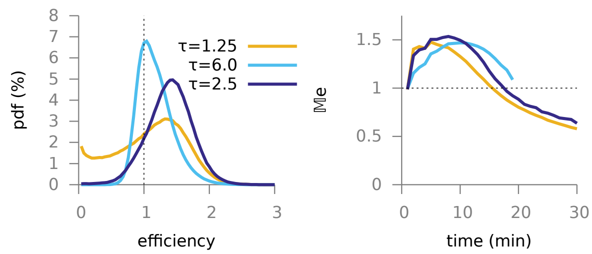

Fig. 6. The PDF of trips as a function of the efficiency (left panel) and median efficiency of trips as a function of cycling time (right panel) provide complementary information to Fig. 5 about the trip–efficiency relation.

maximises the median efficiency. The insert shows the median efficiency for different system sizes and values of N/C.

Additional insights into the trip–efficiency relationship are provided by Fig. 6, showing the PDFs of efficiency, and median efficiency as a function of cycling trip duration for three cases: the near-optimal $\tau = 2.5$ (cf. Fig. 5), and sub-optimal ones: $\tau = 1.25$, and $\tau = 6$. The sub-optimal PDFs either have a peak near $e = 1$ so that a typical trip is either not worth taking or is marginally so ($\tau = 6$), or have excessive exposure to anomalous long trips ($\tau = 1.25$). Figure 6 (right) shows that travel is most efficient in a window of cycling times between roughly 5 and 15 minutes. The latter time can be explained by the considered size of the area (3×3 km).

Values of τ below a certain threshold may result in the perceived utility of all stations becoming negligible. In that case, the model predicts very long 'cycling' trips, even longer than cycling across the entire city! Curiously, there exists a real life analogy of the negligible utility setting[5]. Occasionally, users abandon their bicycles 'on the curb', preferring to leave them unguarded (paying fees for extended usage, or even a fine) rather than taking time to park them.

[5]This was reported to us by the office running the shared bicycle system in Pisa (M. Bertini, private communication)

D. Risk-taking and conditional expectated travel time

In Fig. 7 the distribution of cycling times is shown for several values of τ. Note that for smaller values of τ the longer times are distributed asymptotically as t^{-a}, i.e. the distribution has 'an algebraic tail'. Since walking is less uncertain, the distribution of late arrival times to the destination has qualitatively similar asymptotic properties as cycling time distributions. Let us briefly summarise the implications of distributions having algebraic tails, and how the existence of such tails in the cycling distributions may suggest quantifying user dissatisfaction with the system.

Someone interested in the expectation of being late at a destination, would assume that the arrival time is a random variable T, with some distribution $F(t)$, i.e. $\Pr\{T \leq t\} = F(t)$, and would compute the conditional expectation of arrival time, given that currently at time t the destination hasn't been reached yet: $\mathbb{E}_t T = \int_t^\infty yF(\mathrm{d}y)/\int_t^\infty F(\mathrm{d}y)$. Freely adjusting the reference frame so that the expected arrival time is set to zero, then if $t < 0$ one is early, and if $t > 0$ one is running late. As a general feature of typical distributions, $\mathbb{E}_t T \approx 0$ if $t \ll 0$, meaning that one is expected to arrive 'on time' provided a long enough headway. However, if $t > 0$ then the expectation is at least t (meaning $\mathbb{E}_t T \geq t$), but the precise expression depends crucially on F. If the distribution is strictly algebraic 'in the tail', then $\mathbb{E}_t T = t + \frac{1}{a-2}t$ if $a > 2$, and $\mathbb{E}_t T = \infty$ otherwise. Figure 7 (right) shows how $\mathbb{E}_t T$ changes with t for $\tau = 2.5$ leading to on average shorter trips but occasionally long delays, and $\tau = 6$ (more tolerance), leading to on average longer trips, but more predictable delays that are closer to a normal distribution. It is useful to think in terms of the expected delay δ, defined for $t > 0$ through $\mathbb{E}_t T = t + \delta$. For comparison, the normal distribution yields $\delta \approx \sigma^2/t$ for large t, which means that the near-immediate arrival becomes more certain as the delays accumulate. This is in stark contrast with algebraic distribution, for which we have $\delta = t/(a-2)$. In this case, because $\delta \propto t$, the near-immediate arrival becomes less probable as the delays accumulate. This situation suggests that relatively long undesired trips can sometimes occur that would be perceived negatively by the users who, as we stipulated in the beginning, are time-conscious agents. The exponent a, or the ratio $(a-1)/(a-2)$ can be used to quantify this effect. A comparison of the curves in Fig. 7 (left) and the histograms in Fig. 1 (Data) suggest that the actual behaviour is consistent with moderate risk-taking, exemplified by the optimal $\tau = 2.5$.

E. Rational vs. smart decisions

The model predicts a minor mode at zero minutes' travel. This feature is present in some cities (Pisa, see below, and Lyon, see [9]) but is absent in the London data-set. Borgnat et al. [9] suggest that mistakenly hired malfunctioning vehicles are the explanation. In our model, short trips result invariably from a station of a previous hiring being selected for returning. Such trips exist in the model because all stations, including ones that have just been used for hiring, are equally valid (but not equally probable) options for returning. This is a consequence of the distance-only dependent utility model. A typical setting, whose likely outcome is a 0-minute trip, should have a common nearest station to both the origin and destination locations, placed roughly in the middle between the two objectives. In this setting, it would be 'rational' to hire a bike at the midway station, only to find afterwards that it is also rational to return it there too, even if an agent would be better off walking to the destination directly. This feature of the model is kept because it reproduces the data in Pisa while the overestimate in the London data-set is moderate, and because we have no factual data to rule out a possibility of similar decisions by real users. In the case of a future confirmation of a plausible hypothesis by Borgnat et al., it could be changed by a modification to the station utility model.

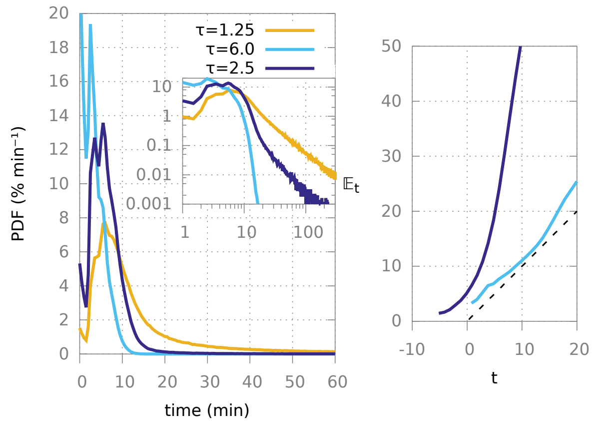

Fig. 7. An agent's propensity to risk-taking (τ) affects the tails of cycling time PDFs (left) and the conditional expectation of the (late) arrival time (right)

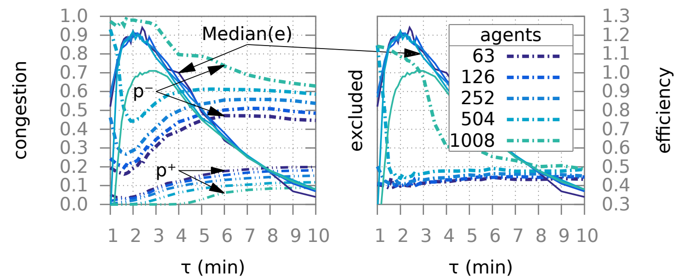

Fig. 8. Model with $N = 252$ and $C = 720$: Median efficiency is relatively insensitive to rising congestion levels (left panel). Likewise the EPM, although rigid tolerance (small τ) result in an unusable system (right panel).

F. Central mode of the distribution

The most prominent feature of the distributions is a mode, typically around 10 minutes. We found that its precise location and the shape of the distribution in the vicinity of the mode is approximated to a good degree by assuming that the cycling tolerance parameter t_c (see section II-C4) is a random number from a uniform distribution, $t_c \sim U(t_{\min}, t_{\max})$. Presence of the cutoff in the expected cycling time is an important underlying property leading to the observed distributions, affecting the characteristic 'neck' of these distributions between the mode and the crossover into the algebraic 'tails' Fig. 1. Typically in real systems, $t_{\max}$ corresponds to the free cycling allowance.

G. Travel efficiency describing the agent's perspective

In a queueing model approach to bike-sharing, Fricker and Gast [6] predict that the population of problematic stations

Characteristic	PSA	LDN	Characteristic	PSA	LDN
Stations	15	742	Approx. area km^2	15	90
Approx. fleet size	150	11,500	Average stations km^{-2}	1	8
Approx. capacity	270	19,000	Cycling $\geq$ 30 min %	6.0	7.7
Average trips h^{-1}	32	1120			

TABLE I. PISA (PSA) AND LONDON (LDN) BIKE-SHARING SYSTEMS

increases when a system has many, or few bicycles with respect to its capacity. They show the existence of an optimal station filling, whose value is not far from $N/C \approx \frac{1}{2}$. There is an apparent disagreement with our model because neither of our metrics has a similar optimal value as a function of a set of aggregated system parameters. The apparent disagreement originates in the interpretation of the term 'problematic'. It is useful to make a distinction between user and exploitation issues. For example, the efficiency (and with it user satisfaction) remains high also for systems with only a few bikes. As another example, Fig. 8 shows that the excluded population remains low for a larger value of τ. The latter would suggest an equally high service level by a network of any size, provided that agents are willing to search for a station in a large area. However, as we have seen in Fig. 5, the efficiency deteriorates markedly with increasing τ. On the other hand, an empty system leads to high EPM, which should be troublesome news to an operator, committed to increasing a system's usage, and to potential users who would like to find a bike. Exploring several metrics with our models we may begin to address issues like (1) 'Given a certain density of population presumably interested in bike-sharing and given a configuration of bicycle stations, what is the fraction of population that will find it impractical to use it?', or (2) 'Can efficiency of trips be improved by changing the number of vehicles or station capacities?' Our model shows that these objectives are not identical and possibly require a trade-off. However, they suggest several ways of attacking the problem such as considering different system configurations or changing the utility perception by introducing incentives.

IV. BIKE-SHARING IN PISA AND LONDON

The bike-sharing systems in London and Pisa differ in size by orders of magnitude, as is evident from Table I. And yet, the cycling time distributions in these cities bear distinctive similarities, as shown by the filled areas in Fig. 1, shared also by other cities [9]. It is tempting to consider them as members of a family of distributions, characterised by a major mode at about 10 minutes, algebraic tails of the distributions with exponent $a \approx 3$, containing about 7% of trips longer than 30 minutes, and a minor mode close to 0 minutes (the latter is absent in the London data). It was suggested that at least three kinds of agents would be required to model similar distributions [9]. In the following we show that a single type of agent is sufficient to generate qualitatively correct distributions with all aforementioned characteristics. This can be achieved without sophisticated parameter fitting, by using only some qualitative arguments. We emphasise that the objective of this study is not a model that copies a real system in question; rather, it is geared towards a model that incorporates quantitatively correct features of user behaviour, with the aim of providing insights into the plausible underlying reasons for the observed data. In fact, the simple models discussed section III already contain many of the salient features, although they are not sufficiently accurate for what concerns some quantitative aspects.

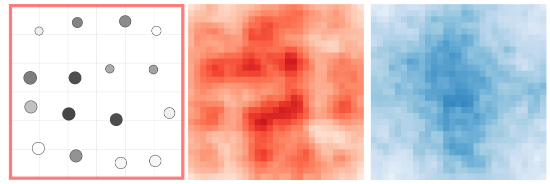

Fig. 9. The uniform Pisa model. Left panel: The map of stations and a random snapshot of their filling (circle size $\propto c$, shade $\propto n$). Middle, right: distributions of agents' origin and destination locations, respectively.

Model	τ (min)	Area + stations (km^2)	S	Capacity min/max	C	Bikes N	Trips (h^{-1})
Pisa	3.5	3×3	16	10/25	170	304	35
London	1.8	7×13	722	15/40	19,652	10,000	823

TABLE II. MODEL PARAMETERS

The curves for the uniform models, shown in Fig. 1, correspond to a multi-agent model (see Table II) with uniformly distributed agent locations. This setting generates uniform demand so that the bicycles are expected to be also uniformly distributed. As such, it is a statistically optimal system of redistribution that does not require intervention. The single necessary additional feature to obtain good *quantitative* agreement between collected data and model results such as those in Fig. 1, are the agent flows. Their introduction leads to some areas being consistently short of bicycles, and others running out of available parking slots. Such flows can be easily constructed using inhomogeneous spatial distributions, and they may be given an additional temporal dimension, for example by requiring agents to honour synchronised appointments. Such temporal 'tidal flows' are present in real systems in the form of morning and afternoon commutes in opposing directions, often with a clear spatial separation. It was found by experimentation that temporal features accentuate the effects of scarce resources and that this effect is quantitatively similar to the static flows.

The model of Pisa represents a relatively small system with a 4×4 array of stations (see Table II). Calibration with a single-agent model, as in Fig. 5, yields $\tau \approx 3.5$ (min), which is quite large as a result of sparsely distributed stations. Introduction of flows to the Pisa model was not necessary to explain the data[6]. The model of London covers a larger area with a 19×38 array of stations (see Table II). Calibration as above yields $\tau \approx 1.8$, which is smaller due to a denser network. To introduce flows, the origin and destination locations were sampled from a superposition of Gaussian distributions, and some counter-current flows for the balancing. This setup is shown in Fig. 10, and the resulting PDF is presented in Fig. 1 (the 'flow model').

Note that there are trips lasting longer than 30 minutes. Since agents use bike sharing with an *a-priori* expectation that their trip should be shorter than a certain $t_{\max}$, with $\sup t_{\max}$ being the aforementioned figure, we know that all trips 'in the tail' were not intended to be so long. This is a first qualitative indication of possible dissatisfaction on the user side. Comparing the uniform vs. flow models, the number of

[6] The discrepancy between these trial and actual distributions visible in Fig. 9 are the result of rejected trips, see sections II-C4 and III-A2

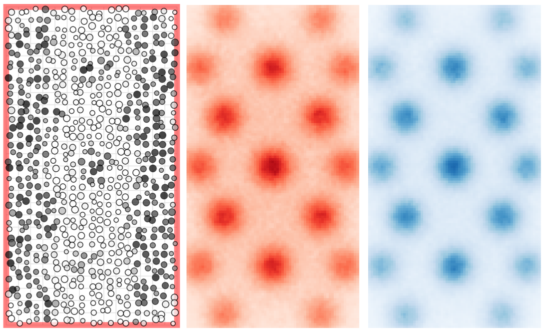

Fig. 10. The flow London model. Left panel: The map of stations and a random snapshot of their filling (circle size $\propto c$, shade $\propto n$). Middle, right: distributions of the origin and destination locations, respectively.

trips in which agents do not find a bike is 10% (Pisa) and about 1% (London). Trips in the tail are 2% (Pisa) and 2% vs. 7.7% (London). The number of completely full stations is 12% (Pisa) and 10% vs. 20% (London). Thus, longer trips are positively correlated with full stations. This suggests that full stations are indeed a plausible cause of the occurrence of (likely) undesired longer trips, and thus a source of concern for user satisfaction. A comparison between the data and the corresponding uniform model (optimal in a sense described before) demonstrates that Pisa is optimal (in that sense), whereas London is significantly sub-optimal. The slightly better-than-optimal performance observed in Pisa is explained by the anticipation of problematic traffic situations using active bike redistribution. London, which has an 8 times higher density of stations, appears to be more problematic. The model suggests that full stations occur in clusters (somewhat like those in Fig. 10) rather than being more uniformly distributed over the area. This emerging formation of clusters is most likely the main cause of the presence of a larger number of long trips. Reducing their number would likely require a redistribution strategy aiming at the breaking up of such clusters.

V. Conclusion

The inclusion of minimal, but plausible, user behaviour and flows in a general bike-sharing model based on MRPs is shown to be sufficient to explain some of the main features of the distributions of actually observed data. The approach provides complementary insight into the attractiveness of bike-sharing from a user's perspective, including that of potential users not captured by data sets. In reality, the choice of a suitable mode of transport is also related to the modal split. The latter is usually evaluated using demand elasticities, empirical parameters that are difficult to model. An agent-based approach that reproduces realistic distributions encourages to extrapolate: in future work we plan to estimate the population that did not find it profitable to use a particular system, and relate it to the demand elasticity. Moreover the evaluation of the effects of alternative configurations and the assessment of the effect of user incentives proposed to stimulate bicycle redistribution by the users will be addressed. Finally, we plan to compare our results with approximate mean-field models of bike-sharing [19].

Acknowledgement

This work is partially supported by the EU project QUANTICOL (nr. 600708), and the IT MIUR project CINA. We thank Marco Bertini and BicinCittà for sharing data of the CicloPi bike-sharing system in Pisa, Mirco Tribastone and Daniël Reijsbergen for sharing data of the bike-sharing system in London, and Dr Cecilia Oram with Sustrans Scotland for references [3], [10].

References

[1] P. D. Maio, "Bike-sharing: Its history, impacts, models of provision, and future." *J. of Public Transport.*, vol. 12, no. 4, pp. 41–56, 2009.

[2] P. Midgley, "Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas," in *19th session of the Commission on Sustainable Development*, ser. CSD19/2011/BP8. United Nations, 2011.

[3] J. Pucher, J. Dill, and S. Handy, "Infrastructure, programs, and policies to increase bicycling: an international review," *Preventive medicine*, vol. 50, pp. S106–S125, 2010.

[4] E. Fishman, S. Washington, and N. L. Haworth, "Bike share's impact on car use: evidence from the United States, Great Britain, and Australia," in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, 2014.

[5] M. H. ter Beek, A. Fantechi, and S. Gnesi, "Challenges in modelling and analyzing quantitative aspects of bike-sharing systems," in *Leveraging Applications of Formal Methods, Verification and Validation*, ser. LNCS, vol. 8802. Springer, 2014, pp. 351–367.

[6] C. Fricker and N. Gast, "Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity," *EURO J. Transport. Log.*, pp. 1–31, 2014.

[7] A. Carey, "Spoke too soon: Melbourne bike share to drag chain another year," *The Age*, Sept 4, 2014.

[8] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *IJCAI*, 2009, pp. 1420–1426.

[9] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, "Shared bicycles in a city: A signal processing and data analysis perspective," *Adv. Complex Syst.*, vol. 14, no. 3, pp. 415–438, 2011.

[10] O. O'Brien, J. Cheshire, and M. Batty, "Mining bicycle sharing data for generating insights into sustainable transport systems," *J. Transport Geogr.*, vol. 34, pp. 262–273, 2014.

[11] J. Bachand-Marleau, B. H. Y. Lee, and A. M. El-Geneidy, "Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use," *Transport. Res. Rec.: J. TRB.*, vol. 2314, no. 1, pp. 66–71, 2012.

[12] E. Çinlar, *Introduction to stochastic processes*. Prentice-Hall, 1975.

[13] V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, 1995.

[14] R. Chhikara, *The Inverse Gaussian distribution: Theory, Methodology, and Applications*. CRC Press, 1988, vol. 95.

[15] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comp. Phys.*, vol. 22, no. 4, pp. 403–434, 1976.

[16] D. F. Anderson, "A modified next reaction method for simulating chemical systems with time dependent propensities and delays," *J. Chem. Phys.*, vol. 127, no. 21, p. 214107, 2007.

[17] D. T. Gillespie, A. Hellander, and L. R. Petzold, "Perspective: Stochastic algorithms for chemical kinetics," *J. Chem. Phys.*, vol. 138, no. 17, p. 170901, 2013.

[18] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Phys. Chem. A*, vol. 104, no. 9, pp. 1876–1889, 2000.

[19] D. Latella, M. Loreti, and M. Massink, "On-the-fly PCTL fast mean-field approximated model-checking for self-organising coordination," *Science in Computer Programming*, 2015, accepted for publication.