**o EPJ Data Science**

**a SpringerOpen Journal**

**REGULAR ARTICLE**

**Open Access**

# Personalized routing for multitudes in smart cities

Manlio De Domenico[1*], Antonio Lima[2], Marta C González[3] and Alex Arenas[1]

*Correspondence:
manlio.dedomenico@urv.cat
[1]Departament d'Enginyeria
Informàtica i Matemàtiques,
Universitat Rovira i Virgili, Av.da
Països Catalans, 26, Tarragona,
43007, Spain
Full list of author information is
available at the end of the article

**Abstract**

Human mobility in a city represents a fascinating complex system that combines social interactions, daily constraints and random explorations. New collections of data that capture human mobility not only help us to understand their underlying patterns but also to design intelligent systems. Bringing us the opportunity to reduce traffic and to develop other applications that make cities more adaptable to human needs. In this paper, we propose an adaptive routing strategy which accounts for individual constraints to recommend personalized routes and, at the same time, for constraints imposed by the collectivity as a whole. Using big data sets recently released during the Telecom Italia Big Data Challenge, we show that our algorithm allows us to reduce the overall traffic in a smart city thanks to synergetic effects, with the participation of individuals in the system, playing a crucial role.

**Keywords:** personalized routing; collective behavior; smart city; potential energy landscape; big data

## 1 Introduction

Rapid development of wireless communication and mobile computing technologies call new research that explores the responses of urban systems to the flow of instant information. Thus, the analysis of spatial signals becomes an increasingly important research theme.

The required four steps to model trips consist of calculating trip generation, trip distribution, modal split and route assignments. The sources to inform these steps traditionally have come from travel diaries and census data [1]. However, the presence of new information and communication technologies (ICT) provide big data sources that are allowing novel research and applications related to human mobility. Recent studies have advanced the knowledge on trip generation by studying the number of different locations visited by individuals through mobile phones and quantifying their frequent return to previously visited locations. These have demonstrated that the majority of travels occur between a limited number of places, with less frequent trips to new places outside an individual radius [2, 3]. In the domain of trip distributions, new models have helped us to predict number of commuting trips when lacking data for calibration [4].

An important topic is to explore route assignments in the context of smart multimodal systems [5, 6], where individual daily trips follow recommendations based on personal and global constraints. This is of special interest towards efficient cities, where individuals could be automatically routed reducing the probability of traffic congestion and at the

same time reducing the environmental impact. From the individual's point of view, for instance, one might want to choose a trip which minimizes the amount of traffic along the route, or to avoid routes across areas with high criminality level, or to favorite routes across more touristic areas, *etc.* On the other hand, the choices of certain routes at individual level, without accounting for the *state of the system*, often leads to traffic congestion [7, 8] which, in turn, is responsible for increasing pollution while decreasing the quality of the environment, with evident impact on the community.

In this work we model the trips in an urban system as interacting particles with data-driven origin-destination pairs that can be routed in their trips. Their route choices are based in a time-varying potential energy landscape that seeks to satisfy individual's and community's requirements simultaneously. Main streams methods for distributed routing seek to avoid congestion by global travel time reduction based on optimization methods [7, 9]. More recently, adaptive path optimization on networks (London underground network and global airport network) related the problem to physics of interacting polymers [10]. In this work we go one step forward in that direction and use a framework based on potential energy landscapes to integrate diverse layers of constraints to favor certain routes and to study the effects of the level of adoption of the proposed recommendations. In this work our main focus is to explore a new framework of analysis to study routing strategies for urban mobility, while the road network constrains are left to further studies.

## 2 Data-driven routing of human mobility

We consider a geographic area of interest (e.g., a city, a district, *etc.*) and we discretize it into a grid $\mathcal{G}$ with size $L \times L'$. In the following, for sake of simplicity, we will consider squared grids with size $L$.

We model individuals moving within the grid as a complex system of interacting sentient particles whose goal is to move between two geographic points according to certain criteria. Each criterion is encoded by a matrix $\mathbf{C}$, with the same dimension of the grid, where each entry indicates the state of the corresponding cell in $\mathcal{G}$. In the same spirit of physical models of an electromagnetic surface, we use the convention that $C_{ij} > 0$ indicates a *repelling* cell, i.e., a geographic area that should be avoided. Similarly, $C_{ij} < 0$ indicates an *attracting* cell, i.e., a geographic area that should be involved for routing. Areas where $C_{ij} = 0$ are considered as neutral.

The origin of a constraint can be of different nature. In fact, there are constraints at individual level, i.e., the ones corresponding to requirements of the single user (*e.g.*, avoid areas with high criminality level), and at global level, i.e., the ones corresponding to the requirement of the whole community (e.g., keep minimum the pollution level). Moreover, there are *static (or quasi-static) constraints* corresponding to restrictions that do not change over time or change over large temporal scales, and *dynamic constraints* corresponding to rapid changes within the system itself, like the traffic flow or the weather. On one hand we should account for individuals' goals and requirements, while on the other hand it is crucial to satisfy constraints imposed for the wealth of the community.

In the following, we consider the set of all constraints, static and dynamic at individual and collective level, and we assign to each of them a time-varying matrix $\mathbf{C}^{(\alpha)}(t)$, where $\alpha = 1, 2, \ldots, M$ and $M$ is the total number of constraints. In the case of static constraints, the matrix is considered constant over time. Moreover, the entries of each matrix are rescaled to the range $0 \leq C_{ij}^{(\alpha)} \leq 1$, for all values of $i, j$ and $\alpha$ to assign a relative importance to each

constraint and to settle on a common scale. Finally, the total constraint matrix is defined by the linear combination of such constraints at each time step:

$$\mathbf{C}(t) = \sum_{\alpha=1}^{M} w_\alpha(t)\mathbf{C}^\alpha(t), \quad \sum_{\alpha=1}^{M} w_\alpha(t) = 1, \tag{1}$$

where the coefficients $w_\alpha(t)$ are empirical and define a trade-off between individual's and global constraints. It is worth remarking that these coefficient might vary over time because, depending on the circumstances (special events, incidents, *etc.*), it could be necessary to change their value to satisfy different priorities.

We define another matrix, $\mathbf{D}_\ell(t)$ ($\ell = 1, 2, \ldots, N(t)$), encoding the starting and destination cells of each individual in the system, where the starting point is considered to be a repelling or neutral area and the destination point is an attractor. The number of individuals $N(t)$ is allowed to change over time. The matrix $\mathbf{D}_\ell(t)$ might change over time because, in principle, the individual might change destination during his or her travel, and for simplicity we assume that $-1 \le D_{ij} \le 0$ for each individual. It is worth remarking that attracting cells are in general associated to destinations and should be encoded in the set of matrices $D_\ell(t)$, whereas repelling cells are associated to constraints and should be encoded only in the set of matrices $C^\alpha(t)$.

We interpret the set of matrices $\mathbf{C}(t)$ and $\mathbf{D}_\ell(t)$ as potential energy landscapes and the routing of individuals is performed by means of a gradient descent, where each user moves along geodesics while reducing his or her potential energy until he or she reaches the destination. For simplicity, we assume no dependence on time for matrices $\mathbf{D}_\ell(t)$. We consider the case of a gravitational field in two dimensions permeating the areas encoded by $\mathbf{D}_\ell(t)$. More specifically, let $(i_\ell, j_\ell)$ and $(i_\ell^{(d)}, j_\ell^{(d)})$ denote the cells of the underlying grid and destination point of the journey of individual $\ell$, respectively, and let $r = \sqrt{(i_\ell^{(d)} - i_\ell)^2 + (j_\ell^{(d)} - j_\ell)^2}$ indicate their distance. The potential energy landscape is defined by

$$D_\ell(r) = \begin{cases} -\frac{\Omega}{\sqrt{r}}, & (i_\ell, j_\ell) \ne (i_\ell^{(d)}, j_\ell^{(d)}), \\ -\frac{\Omega}{\sqrt{1-10^{-4}}}, & (i_\ell, j_\ell) = (i_\ell^{(d)}, j_\ell^{(d)}), \end{cases} \tag{2}$$

where $\Omega$ is a constant factor, defining the scale of the potential which should guarantee that the potential is strong enough in each cell. In our simulations, we considered $\Omega = 30L\sqrt{2}$.

The choice of the value of the potential at the destination is somehow arbitrary and, as a rule of thumb, it should be a number smaller than the potential of the neighbors (whose distance is $r = 1$ or $r = \sqrt{2}$, the latter if movements along diagonals are allowed), but not so small to avoid a potential well so deep that the rest of the landscape is almost flat.

To guarantee the convergence of the gradient descent even in presence of constraints or noise resulting in potential wells, we weight the overall landscape for each particle by

$$V_\ell(r, t) = \gamma(t)D_\ell(r) + \big(1 - \gamma(t)\big)C(r, t), \tag{3}$$

where $C(r, t)$ is the potential energy landscape corresponding to constraints encoded by matrix $\mathbf{C}(t)$. The weighting factor $\gamma(t)$ should be a function ranging between 0 and 1 accounting for the importance given to the constraints with respect to the destination. The key to ensure the convergence of the gradient descent, while accounting at the same time

for the constraints, is to make this function changing over time from an initial value up to 1. A candidate function is given by

$$\gamma(t; a, b) = 1 - (1 - b)e^{-at}, \tag{4}$$

where $a$ is a non-negative number whose inverse $\tau = a^{-1}$ defines the time scale for convergence to 1 and $b$ is the relative importance to be assigned at time $t = 0$ to constraints and destination. A reasonable choice is to balance the two potential energy landscapes to allow the particles to be routed according to the constraints *and* the destination up to a time scale $\tau$, above which the influence of the destination becomes more important. Small values of $b$ might give more importance to the constraints rather than destination, leading to a routing less oriented to the final destination during the first time steps. Therefore, we require $\gamma(0) \geq 1 - \gamma(0)$ leading to $b \geq 0.5$.

We rewrite Eq. (3) to put in evidence the terms corresponding to different constraints. Let $C_\sigma(r)$ and $C_\delta(r, t)$ denote the potential due to all static and dynamical constraints, respectively, which are not related to the state of the other particles of the system. For instance, $C_\sigma(r)$ might encode the landscape corresponding to crimes, supposed to change over very long time scales, while $C_\delta(r, t)$ might encode the areas where it is raining, snowing or being affected by other meteorological events. On the other hand, we make the realistic assumption that not all individuals follows the routing provided by the smart system. While the information about the traffic of all individuals can be available by sensors properly disseminated across the grid, it is not possible to predict the behavior of a certain fraction $p$ of individuals. To account for such a fraction $p$ of individuals, we consider a set of $N(1 - p)$ individuals moving along shortest paths between pairs of origin and destination, sampled from real data as discussed further in the text, and a set $Np$ of individuals moving randomly in the city, i.e., following random walks instead of shortest paths. We indicate by $F_{\text{in}}(r, t)$ the potential corresponding to the flow of individuals *within* the system, i.e., those ones following suggestions from the smart system, and by $F_{\text{out}}(r, t)$ the potential corresponding to the flow of individuals *out* of the system. The latter is modeled by a noisy flow in terms of random walking individuals, although other mobility models can be used. In order to preserve conservation of the flow, we rescale each term by the number of particles in the most visited cell, i.e., by a weight $m(t) = \max[\mathbf{F}(t)]$, being $\mathbf{F}(t)$ the matrix accounting for the flow of individuals in the city at time $t$, with $\sum_{\text{cell} \in \mathcal{G}} \mathbf{F}(t) = N(t)$. The matrix $\mathbf{F}(t)$ is not weighted by the factor $[1 - \gamma(t)]$ as in the case of $C_\sigma(r)$ and $C_\delta(r, t)$, because it would wash out the contribution of $\mathbf{F}(t)$ to the potential landscape for increasing time. This choice makes our model more realistic: in fact, while it is possible to decide to traverse an undesirable area to balance the time spent looking for alternatives, it is not possible to traverse those areas which are congested or overcrowded. Therefore, the potential energy landscape accounting for the traffic flow should not be weighted by the function $1 - \gamma(t)$, whose existence is justified only to introduce a trade-off between the needing to reach the destination and the time spent to achieve this goal while accounting for personalized constraints. Finally, Eq. (3) maps to

$$V_\ell(r, t) = \gamma(t)D_\ell(r) + \big(1 - \gamma(t)\big)\big[C_\sigma(r) + C_\delta(r, t)\big]$$
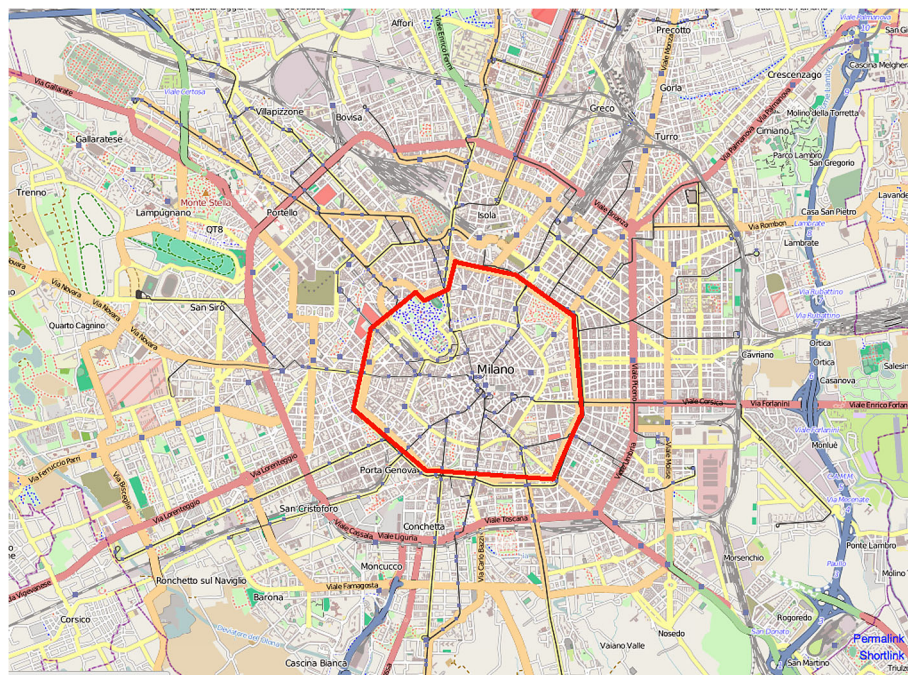$$+ m(t)\big[(1 - p)F_{\text{in}}(r, t) + pF_{\text{out}}(r, t)\big]. \tag{5}$$

**Figure 1 City of Milan.** Area of Milan (Italy) considered in the present study. The area is divided into a squared grid with 10,000 cells of size 55,225 m$^2$.
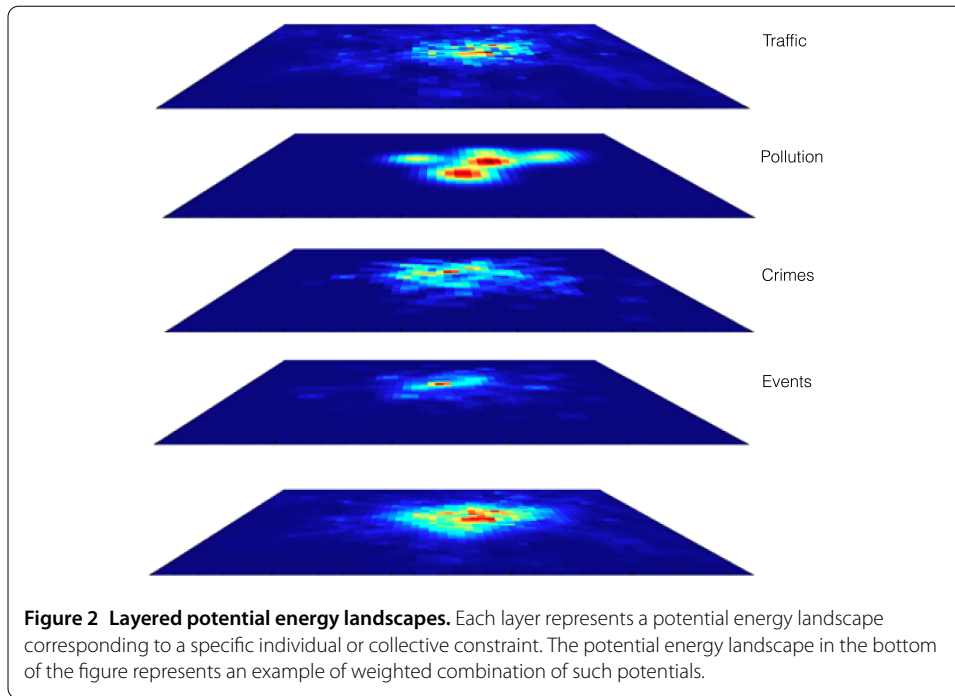
This model is rather general, accounting for the presence of traffic and, simultaneously, for personalized and collective, static and dynamic, constraints. However, in this study we focused only on static constraints and we aggregated time-varying constraints for simplicity. It is worth remarking here that the potential landscape $V_\ell(r,t)$ experienced by individual $\ell$ still changes over time, because of the traffic flow term. Moreover, if agents are distributed in the grid according to the underlying population distribution and they move along shortest-path adapting over time in the evolving potential landscape, it is not possible to perform quantitative predictions about the state of the full system at a given time without numerical simulations.

## 3 Overview of the dataset

Most of the datasets used in this work were acquired as part of the Telecom 'Big Data Challenge' and all of them are related to the city of Milan, Italy (see Figure 1).

The constraints encoded by matrices $\mathbf{C}^\alpha(t)$ can be represented as different 'layers' of the city, as shown in Figure 2. The weighted combination of such layers, as in Eq. (1), allows to build the potential energy landscapes $C_\sigma(r)$, $C_\delta(r,t)$, $F_{\text{in}}(r,t)$ and $F_{\text{out}}(r,t)$ influencing the overall landscape defined by Eq. (5).

For simplicity, we considered four static layers obtained from the provided datasets and here we explain how the layers were generated. The 'pollution' layer was generated from readings of 7 sensors scattered around the city, taken hourly over the course of 2 months. Because these sensors are very sparse in space, we smoothed their readings conveniently. The 'events' layer was generated by looking at the number of tweets coming from each grid of the city. It contains 100,000 geolocated tweets generated over a 30-day period. Lastly, the 'crime' layer was generated from a list of crimes, manually curated, and sourced from

**Figure 2 Layered potential energy landscapes.** Each layer represents a potential energy landscape corresponding to a specific individual or collective constraint. The potential energy landscape in the bottom of the figure represents an example of weighted combination of such potentials.

newspaper articles.[a] It contains 1276 crimes happened during the course of 12 months in Milan and reported by newspapers and local media.

Finally, we used data about the total number of calls and texts generated in Milan by all users of a mobile carrier, over a period of two months. We used the aggregated fraction of calls and texts between areas of the city, aggregated over the whole 2-month period, to determine the distribution of trip origin and destination, as detailed in the next session.

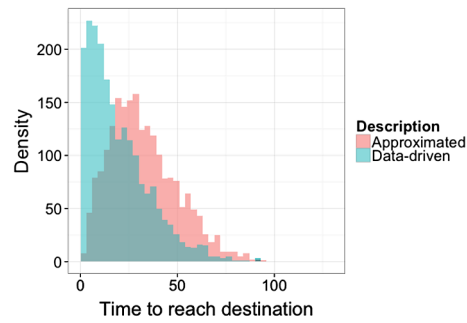## 4 Simulation of personalized routing

We performed massive simulations of personalized routing in Milan to gain insights about which factors influence the time required to complete a journey.

We started by exploring different ways to sample origin and destination cells for each individual in the city. The simplest strategy would be to choose both origin and destination with uniform probability on the grid. Of course, this strategy can not be realistic for several reasons. On one hand, the population is never uniformly distributed over metropolitan areas like Milan, where there is a high concentration of individuals in the 'core' of the city, while the population density decreases for increasing distance from the city centre [11]. In fact, assuming a uniform distribution of origins implicitly considers a population uniformly distributed. On the other hand, the choice of a random destination, regardless of the origin, is not representative of real urban mobility, where individual's journeys show a high degree of spatio-temporal regularity, with a few highly frequented locations [12–14] and high predictability of the underlying trajectories [3, 15, 16].

For this reason, we employed a data-driven approach accounting for intrinsic correlations in human mobility and leading to a more realistic distribution of origin-destination pairs. As a proxy for the population distribution, we have used the human activity measured by calls and texts generated by mobile phones. The calls dataset also provided information about the distribution of calls across all the pairs of grids; we exploited this infor-

**Figure 3 Simulation of origin and destination.**
Different strategies for the simulation of origin and destination affect the distribution of time required to complete a journey. 'Approximated' indicates a random geographical area, whereas 'data-driven' indicates a simulation where origin is sampled from the population distribution and destination is sampled according to the origin-destination probability matrix.

**Figure 4 Simulation of personalized routing in Milan.** Data-driven simulations where a certain fraction of individuals traveling in the city are routed by our system. The average time required to complete a journey decreases for increasing synergy, i.e., for increasing adoption of the personalized routing. The distribution of journey duration is shown in the ideal case, i.e., the non-physical scenario where each individual travels without constraints of any type, such as traffic, *etc*.
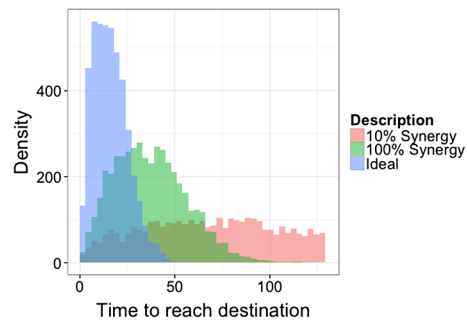
mation to sample a realistic ensemble of origin-destination pairs and to build an origin-destination probability matrix. Although this is a strong assumption, recent works [17–19] show how one of these quantities can be used to measure the other. Our simulations, summarized in Figure 3, show that the time required to complete a journey is, on average, faster when a data-driven strategy is employed vs. the one approximated by random origin and destinations.

We capitalized on this result to perform data-driven simulations by varying the fractions of individuals traveling by adopting our routing system. For each individual, we calculated again the time required to complete his or her journey, sampled according to the origin-destination probability matrix. To understand how the efficiency of our re-routing algorithm is affected by the fraction of individuals adopting the recommended routes, we define this fraction $(1 - p)$ as the *synergy* of the system and we calculate the time required to reach the destination for each individual. The remaining fraction $p$ of individuals does not follow the recommended routes. We found that the underlying synergy has a non-negligible effect on the way individuals experience mobility the city. Our results, shown in Figure 4, put in evidence that the average time required to complete a journey decreases for increasing synergy, i.e., for increasing adoption of the personalized routing. This result was expected: when only a small fraction of individuals moves along the routes suggested by our system, it is not possible to calculate efficient trajectories because the only information available to the system is about the traffic generated by other people, while the information about their origin and destination is unknown. Conversely, when a large number of individuals adopts the suggested routes the potential energy landscape is less subjected to noisy fluctuations and a more efficient calculation of trajectories can be performed. For comparison, we show in the same figure the distribution of journey duration in the non-physical scenario where each individual travels without constraints of any type, such

as traffic, *etc.* This optimal case, shown in figure for comparison, is a free-flow scenario where every person goes to their destination undisturbed by other people. Individuals' routes were sampled according to origin-destination matrix also in this case. While it is not possible to fit the distribution of the ideal journey duration, our results show that a 100% synergy produces a distribution close to the ideal one. It is worth remarking that this analysis would be able to quantify the benefits of synergy for urban traffic if information on the individual adoption of routing technology could be available to researchers.

Our routing system also allows to monitor mobility of the city from a new point of view. Interpreting individuals as particles moving in a thermodynamical system, it is possible to calculate the 'temperature' of the city. For each particle $\ell$ we calculate the mean speed at time $t$ by

$$v_\ell\left(t; t_\ell^{(0)}\right) = \frac{\sqrt{[i_\ell(t) - i_\ell(t_\ell^{(0)})]^2 + [j_\ell(t) - j_\ell(t_\ell^{(0)})]^2}}{t - t_\ell^{(0)}}, \tag{6}$$

i.e., as the ratio between the distance travelled up to time $t$ and the time required to travel. Here $t_\ell^{(0)}$ indicates the time at which the particle has been injected into the system, i.e., the time at which the individuals leaves the origin of his or her route. The temperature of this system can be defined as the mean squared speed $\langle v_\ell^2 \rangle_\ell$. This measure is better understood in terms of permeability (or connectivity) of the city, as defined in urban studies allowing us to quantify how fast individuals flow through the city. Therefore, we define the permeability by
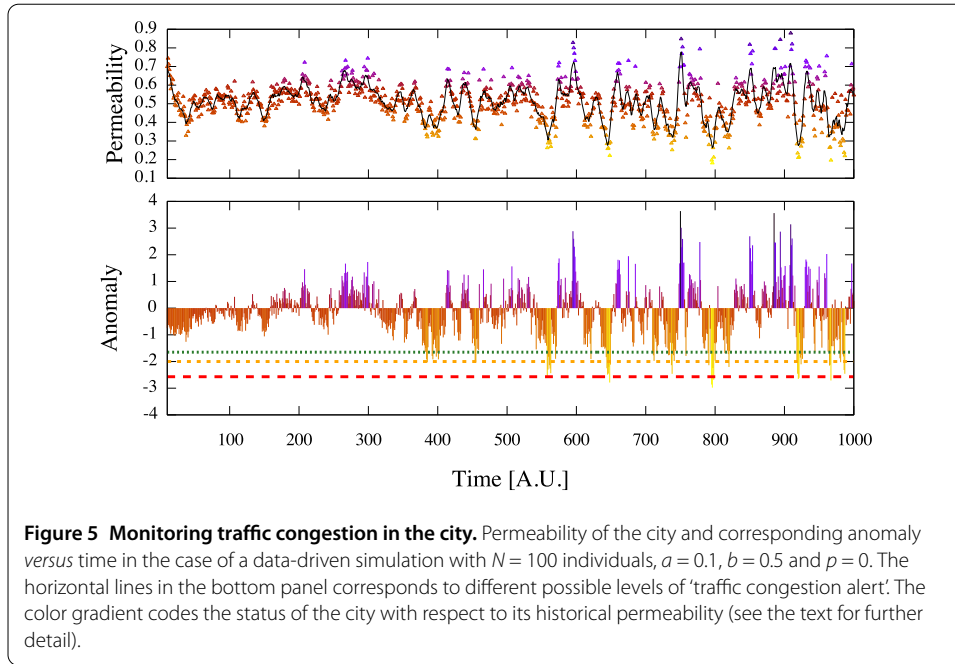
$$\mathcal{P}(t) = \left\langle v_\ell^2(t) \right\rangle_\ell = \frac{1}{N_{\mathrm{in}}(t)} \sum_{\ell=1}^{N_{\mathrm{in}}(t)} v_\ell^2(t), \tag{7}$$

where the sum and the average are limited to individuals adopting the routing system, because of the lack of information about origin-destination of the others. Nevertheless, $\mathcal{P}(t)$ is indirectly affected by the traffic generated by $N_{\mathrm{out}}(t)$ individuals, therefore it is a robust measure of permeability. Higher the value of $\mathcal{P}(t)$ faster the flow of individuals trough the city and, conversely, lower the value of $\mathcal{P}(t)$ and slower the movements in the city, i.e., higher the probability that there are congested areas or, in the worst case, 'frozen' cells in the grid. In the upper panel of Figure 5 we show how the permeability changes over time for a data-driven simulation with $N = 100$ individuals, $a = 0.1$, $b = 0.5$ and $p = 0$, i.e., for 100% synergy. The color gradient codes the status of the city with respect to its historical permeability. The existence of congested areas is more evident when the time series of anomaly $\mathcal{A}(t)$ is observed. The anomaly is defined as the departure of $\mathcal{P}(t)$ from the historical average $\mu_{\mathcal{P}}(t)$ with respect to the historical standard deviation $\sigma_{\mathcal{P}}(t)$

$$\mathcal{A}(t) = \frac{\mathcal{P}(t) - \mu_{\mathcal{P}}(t)}{\sigma_{\mathcal{P}}(t)}, \tag{8}$$

where

$$\mu_{\mathcal{P}}(t) = \frac{1}{t} \sum_{\tau=1}^{t} \mathcal{P}(\tau), \tag{9}$$

**Figure 5 Monitoring traffic congestion in the city.** Permeability of the city and corresponding anomaly *versus* time in the case of a data-driven simulation with $N = 100$ individuals, $a = 0.1$, $b = 0.5$ and $p = 0$. The horizontal lines in the bottom panel corresponds to different possible levels of 'traffic congestion alert'. The color gradient codes the status of the city with respect to its historical permeability (see the text for further detail).

$$\sigma_{\mathcal{P}}(t) = \sqrt{\frac{1}{t}\sum_{\tau=1}^{t}\mathcal{P}(\tau)^2 - \left[\frac{1}{t}\sum_{\tau=1}^{t}\mathcal{P}(\tau)\right]^2}.$$
(10)

In the bottom panel of Figure 5 we show the anomaly changing over time. The traffic experiences large fluctuations for large values of $t$, positive and negative ones, alternating periods of high permeability with a few periods of low permeability. This is due to a few overcrowded cells that are quickly and automatically uncrowded by the system itself. Therefore, it is possible to monitor the traffic of the city by looking at the permeability and its anomaly over time, programming different alert levels such as low ($-2 \leq \mathcal{A}(t) < -1.7$), medium ($-2.6 \leq \mathcal{A}(t) < -2$) or critical $\mathcal{A}(t) < -2.6$.

## 5 Discussion and conclusions

We have presented a strategy to route individuals between pairs of points of interest according to constraints of different type. Our method accounts for the simultaneous interplaying between personalized constraints, as avoiding specific areas of the city because of personal choices, and collective constraints, from pollution reduction in certain areas of the city to the presence of adverse atmospherical conditions requiring targeted intervention. We have shown that the synergy plays a fundamental role in designing a smart city: only when all individuals take part in the routing system and move according to the recommended routes, the overall traffic in the city is closer to the most ideal mobility scenario. In the presence of real time information, our method allows to monitor the state of the city in real time, automatically identifying areas that are experiencing a temporary congestion and giving authorities the possibility to intervene timely.

Finally, the potential applications of our routing strategy are multiple. For instance, for certain values of the parameters (i.e., $a = b = 0$, leading to $\gamma(t) = 0$), we obtain a routing strategy from an origin and without a fixed destination, while accounting for specified constraints. This case could be useful to perform automated routing of objects or individuals

through the city. For instance, it would be possible to route cars or drones which are collecting data about the city (as Google cars) and to route people in charge of social services like cleaning the streets or performing targeted intervention, as disseminating salt in areas with snow. An additional application could be in the field of social security, to route police cars in areas with high crimes rate. Finally, our framework can help decision-makers to real-time application of urban mobility policies in responses to crisis, e.g. the emergence of hotspots of infection in specific areas of the city (or a larger area) can be incorporated into the model to avoid people passing through dangerous areas before physical quarantine is employed.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

MDD, AL, MCG and AA devised the study and wrote the manuscript. MDD and AL performed the data analysis and data-driven simulations. All authors reviewed and approved the complete manuscript.

**Author details**

[1]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av.da Països Catalans, 26, Tarragona, 43007, Spain.  [2]School of Computer Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. [3]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

**Endnote**

    a   http://www.linkiesta.it/reati-a-milano.

**References**

1. Hazelton ML (2008) Statistical inference for time varying origin-destination matrices. Transp Res, Part B, Methodol 42(6):542-552
2. Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. J R Soc Interface 10(84):20130246
3. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018-1021
4. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484:96-100
5. De Domenico M, Solé-Ribalta A, Gómez S, Arenas A (2014) Navigability of interconnected networks under random failures. Proc Natl Acad Sci USA 111(23):8351-8356
6. Gallotti R, Barthelemy M (2014) Anatomy and efficiency of urban multimodal mobility. Sci Rep 4:6911
7. Youn H, Gastner MT, Jeong H (2008) Price of anarchy in transportation networks: efficiency and optimality control. Phys Rev Lett 101(12):128701
8. Wang P, Hunter T, Bayen AM, Schechtner K, González MC (2012) Understanding road usage patterns in urban areas. Sci Rep 2:1001
9. Delling D, Goldberg AV, Pajor T, Werneck RF (2011) Customizable route planning. In: Experimental algorithms. Springer, Berlin, pp 376-387
10. Yeung CH, Saad D, Wong KM (2013) From the physics of interacting polymers to optimizing routes on the London underground. Proc Natl Acad Sci USA 110(34):13717-13722
11. Makse HA, Havlin H, Stanley H (1995) Modelling urban growth. Nature 377:19
12. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779-782
13. Lima A, De Domenico M, Pejovic V, Musolesi M (2013) Exploiting cellular data for disease containment and information campaigns strategies in country-wide epidemics. arXiv:1306.4534
14. Salnikov V, Schien D, Youn H, Lambiotte R, Gastner M (2014) The geography and carbon footprint of mobile phone use in cote d'ivoire. EPJ Data Sci 3(1):3
15. Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. Nat Phys 6(10):818-823
16. De Domenico M, Lima A, Musolesi M (2013) Interdependence and predictability of human mobility and social interactions. Pervasive Mob Comput 9(6):798-807
17. Crandall DJ, Backstrom L, Cosley D, Suri S, Huttenlocher D, Kleinberg J (2010) Inferring social ties from geographic coincidences. Proc Natl Acad Sci USA 107(52):22436-22441

18. Farrahi K, Emonet R, Cebrian M (2014) Epidemic contact tracing via communication traces. PLoS ONE 9(5):95133
19. Palchykov V, Mitrović M, Jo H-H, Saramäki J, Pan RK (2014) Inferring human mobility using communication patterns. Sci Rep 4:6174