



# Algorithms for the People Intra-City Flows Analysis by the Data on Bank Cards Payments

Sinitsyn EV<sup>1,\*</sup>, Laptev VM<sup>1</sup>, Komarova KS<sup>2</sup> and Buzunov AN<sup>2</sup>

<sup>1</sup>Ural Federal University, Russia

<sup>2</sup>Division, Corporate Clients 360, PJSC Sberbank, Ekaterinburg, Russia

\*Corresponding author: Sinitsyn EV, Ural Federal University, Russia; E-mail: [sinitsyn\\_ev@mail.ru](mailto:sinitsyn_ev@mail.ru)

## Abstract

The tasks of passenger flows forecasting as well as the detection of places where the people concentrate during intra-city flows are important for planning the development of transport infrastructure, demand for transport services and organizing of traffic control. The lessons of the COVID-19 pandemic show that such forecasting is also relevant for anti-epidemic measures. The goal of this paper is the development of probabilistic mathematical model for intra-city people flows on the base of bank card payments analysis. This model allows to obtain information about territorial and temporal intra-city people flows. Such flows can be also clustered by age - gender composition and payment purposes. The usage of this model is illustrated by the examples of two Russian cities: Ekaterinburg and Moscow.

**Keywords:** Probabilistic model of people flows; Forecasting of people/passenger flows; People flows forecasting by means of bankcards payments analysis

## Introduction

Long-term planning of transport infrastructure and urban development, tasks of public safety (for example avoiding of accidents caused by overcrowding), anti-epidemic measures are based on complete, timely information about the people flows over exact areas. Such information is also implied by the "Smart city" standards a promising direction of the urban development. Currently, several methods are used to get necessary Information [1]:

- Technical vision systems [2,3].
- Using of various sensors - "counters" [4].
- Sold tickets data.

All the above methods need either special measures, or rather expansive equipment [2,3]. Meantime, one can suggest two methods that do not have these shortcomings, since the information about the people flows over the studied territories is a by-product of technologies satisfying the regular needs of residents. These methods are the usage of data from mobile operators and the analysis of bank cards payments data. The last are concentrated in the ecosystems of large banks. For example,

in Russia one may use the data of Burbank - the largest bank in Russia, Central and Eastern Europe, and one of the leading financial institutions worldwide. The bank cards of it are used by most Russian citizens [5,6]. The people flow analysis based on the mobile operator's data is well known and is implemented in many popular navigation services. The examples of bank cards payments analysis for the same purposes are much scarcer. At the same time, such payments are an inalienable part of everyday life for any citizen and can be a source of regular necessary information to solve the above-mentioned problems. In data analysis one can use the descriptive analytics and predictive modelling [7]. A common variant of application the above-mentioned methods is namely descriptive analysis. For the purposes of predictive modelling, it is worth to use probabilistic models (check and the references in it). In this paper we are going to suggest the variant of such model that was successfully used in different analytical tasks [8-10]. Particularly, in the tasks, related with analysis of members' flows between various communities. For example, this can be the predictions of passengers, or pedestrian's flows between different city places or different cities

Received date: 18 April 2022; Accepted date: 25 April 2022; Published date: 30 April 2022

**Citation:** Sinitsyn EV, Laptev VM, Komarova KS, Buzunov AN (2022). Algorithms for the People Intra-City Flows Analysis by the Data on Bank Cards Payments. SunText Rev Econ Bus 3(2): 157.

**DOI:** <https://doi.org/10.51737/2766-4775.2022.057>

**Copyright:** © 2022 Sinitsyn EV, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[10]. Let's suppose that there are  $M$  city's regions and  $X_i$  is the number of potential participants of flows in each, so:

$$\sum_{i=1}^M x_i = N \quad (1)$$

Where  $N$  is the total number of moving citizens. The task is to determinate the probability of the distribution  $\{X_1, X_2, \dots, X_M\}^{-1}$  at the time  $t - \Phi(X_1, X_2, \dots, X_M|t)$ . Such movements can be considered as Continuous-time Markov process [11]. The task has many analogues, for example: random walks over the vertexes of an oriented graph, birth-death processes [11]. For our tasks it is much more convenient to use not absolute values  $X_i$ , but concentrations  $x_i = \frac{x_i}{N}$ .

$$\sum_{i=1}^M x_i = 1 \quad (1)$$

And probability distribution  $P(x_1, x_2, \dots, x_M|t)$  - respectively. The equation for it can be derived in a manner like [11].

$$\frac{\partial \Phi(x_1, \dots, x_M|t)}{\lambda N \partial t} = \sum_{i,j \neq i}^M \lambda_i \left(x_i + \frac{1}{N}\right) P_{ij} \Phi\left(x_1, \dots, x_i + \frac{1}{N}, \dots, x_j - \frac{1}{N}, \dots, x_M|t\right) - \sum_{i,j \neq i}^M \lambda_i x_i P_{ij} \Phi(x_1, \dots, x_M|t) \quad (2)$$

Here  $P_{ij}$  is the probability of transition from the place  $i$  to the place  $j$ ,  $\lambda_i$  is the relative, number of citizens leaving the place  $i$  and  $\lambda$  is the total number of citizens, leaving all  $M$  places at the unit of time (see section 2.1 below for details). If  $P_{ij}$ ,  $\lambda_i$  are determined on the base of information about real people flows the equation (2) describes the model process with statistical characteristics equal to that of the real one. All the above-mentioned methods are appropriate to calculate of  $P_{ij}$ ,  $\lambda_i$  for concrete city or region.

## Data and Methods

### The data on bank card's payments

To test the model, we used the data of payments by Sberbank's cards tied to the place of registration of the device through which the payment was done. For the analysis, we used the data for payments made within 7 months from July 2020 to December 2021, grouped by working and non-working days of each month and by two-hour intervals from 0 to 24 hours. Information about the age and gender of cardholders as well as the payment's purpose was also available. A cardholder was counted as moved from one place to another if she or he paid twice during one hour in the different places. Data were collected both for the number of such persons and for the amounts of money spent in each place. The territories of cities were covered by the net of hexagons (edge of each was 170 m) with known coordinates (latitude and

longitude) of centers. Such hexagons were grouped (if necessary) into larger geographical objects (city districts for example). It was these objects that were considered as places between which the citizens were moving, in equation [2].

These data were used to calculate the main parameters in (2) -  $P_{ij}$ ,  $\lambda_i$ :

$$P_{ij} = \frac{N_{i \rightarrow j}}{N_i} \quad (3a)$$

Here  $N_{i \rightarrow j}$  - number of cardholders who moved from the place (node)  $i$  to the node  $j$ ,

$$N_i = \sum_{j \neq i}^M N_{i \rightarrow j} \quad (3b)$$

$M$  - As above is the total number of the nodes under consideration.  $N$  - Is the total number of moving cardholders during the considered time -  $T$ :

$$N = \sum_i^M N_i \quad (3c)$$

And finally:

$$\lambda_i = \frac{N_i}{N} \quad (3d)$$

One can easily show that for each node:

$$\sum_{j \neq i}^M P_{ij} + P_{ii} = 1 \quad (3e)$$

It is convenient to introduce dimensionless time

$$\tau = \lambda t, \quad \lambda = \frac{N}{T} \quad (4)$$

As an example, Fig.1 shows the comparison of  $\lambda$  distribution for the nodes of two Russian cities: Moscow and Ekaterinburg (Figure 1).

$$\begin{aligned} \lambda_{max}(\text{Moscow}) &= \max_{i=1, \dots, M} \lambda_i \\ &= 0,015, \lambda_{max}(\text{Ekaterinburg}) \\ &= \max_{i=1, \dots, M} \lambda_i 0,092 \end{aligned}$$

The overall number of citizens in flows during  $T - \lambda$  (5) is equal:

$$\lambda(\text{Moscow}) = 1,11 \cdot 10^4 \text{hour}^{-1}, \quad \lambda(\text{Ekaterinburg}) = 3,42 \cdot 10^2 \text{hour}^{-1}$$

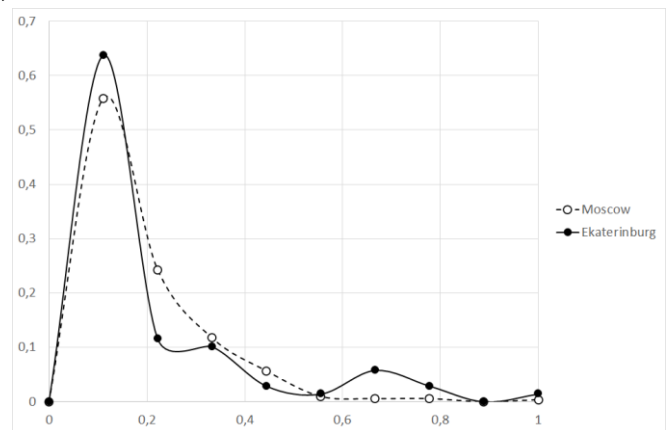


Figure 1: Relative fraction (y-line) of nodes  $i$  with  $\lambda_i - (3d)$  lying in the corresponding intervals of  $\frac{\lambda_i}{\lambda_{max}}$  values (x-line).

The probabilities  $P_{ij}$  (5) cannot be shown because of its tremendous amount. For illustration we present figure 2 the probabilities of citizen's movement on the definite distance for the same cities as on Figure 1 (Figure 2).

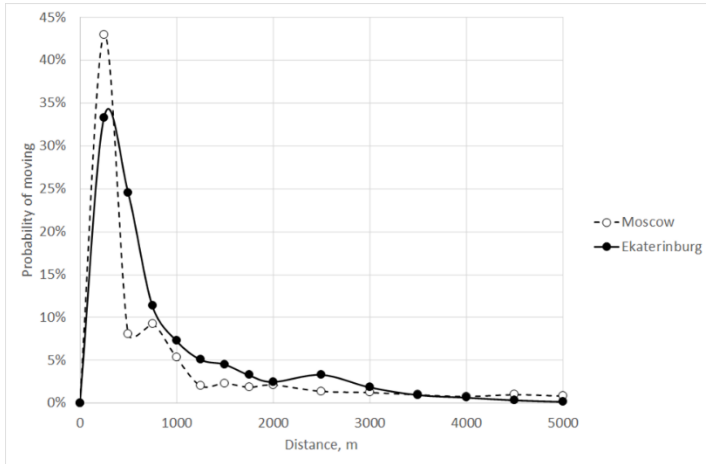


Figure 2: The probability of citizen's movement on the given distance.

### Methods of people flows prognosis

In this section we are going to discuss the possible solutions of the main equation (2) and its consequences, particularly to the possible variants of people flows predictions.

**Correlation functions of people concentrations:** Let's return to the equation (2). It can be used to determine moment-generating function or solved numerically. We'll use both variants below. The results obtained allow to predict the probability distribution  $\Phi(x_1, x_2, \dots, x_M | t)$ , for the parameters  $P_{ij}$ ,  $\lambda_i$ , determined in the section 2.1. Usually, it is quite enough to calculate first two moments of a random variable  $x_i$ : expected value  $\langle x_i \rangle$  and correlation function  $\sigma_{ij}$

$$\langle x_i \rangle = \int x_i \Phi(x_1, \dots, x_i, \dots, x_M | \tau) dx_1 dx_2 \dots dx_M, \quad (5a)$$

$$\sigma_{ij} = \int x_i x_j \Phi(x_1, \dots, x_M | \tau) dx_1 dx_2 \dots dx_M - \langle x_i \rangle \langle x_j \rangle. \quad (5b)$$

After standard routine with the usage of (2) one can transform (6a,b) to the form:

$$\frac{\partial \langle x_i \rangle}{\partial \tau} = \sum_{k \neq i}^M \{ \lambda_k P_{ki} \langle x_k \rangle - \lambda_i P_{ik} \langle x_i \rangle \}, \quad (6a)$$

$$\begin{aligned} \frac{\partial \sigma_{ij}}{\partial \tau} = & \sum_{k \neq i}^M \lambda_k P_{ki} \sigma_{kj} + \sum_{k \neq j}^M \lambda_k P_{kj} \sigma_{ik} - (\lambda_i P_{i^*} + \lambda_j P_{j^*}) \sigma_{ij} + \\ & + \frac{\delta_{ij}}{N} \sum_{k \neq i}^M (\lambda_k P_{ki} \langle x_k \rangle + \lambda_i P_{ik} \langle x_i \rangle) - \\ & - \frac{1 - \delta_{ij}}{N} (\lambda_j P_{ji} \langle x_j \rangle + \lambda_i P_{ij} \langle x_i \rangle) \end{aligned} \quad (6b)$$

Here

$$P_{i^*} = \sum_{k \neq i}^M P_{ik} = 1 - P_{ii}. \quad (6c)$$

By direct summation in (7a,b) it can be shown that:

$$\sum_{k \neq i}^M \frac{\partial \langle x_i \rangle}{\partial \tau} = \sum_{i,j}^M \frac{\partial \sigma_{ij}}{\partial \tau} = 0, \quad (7)$$

As it should be in accordance with (1).

From (6) one can find that for a large  $N$ ,  $\langle x_i \rangle$  has the zero order of magnitude in  $(1/N)$ , and  $\sigma_{ij}$  is proportional to  $(1/N)$ , so as  $N$  tends to infinity,  $\langle x_i \rangle$  is finite, and tends to zero. Moreover, it can be shown, that not only  $\sigma_{ij}$ , but also the high order moments of the random variables  $x_i$  tends to zero as  $N \rightarrow \infty$ . Thus, in the limit of large  $N$ :

$$\Phi(x_1, \dots, x_i, \dots, x_M | t) \rightarrow \delta(x_1 - \langle x_1(t) \rangle) \dots \delta(x_M - \langle x_M(t) \rangle), \quad (8)$$

Where  $\langle x_i(t) \rangle$  – are the solutions of (6a)

**The continuity equation and its solutions:** In the limit of large  $N$  it is convenient to transform equation (2) by expanding the probability  $\Phi(x_1, \dots, x_i + \frac{1}{N}, \dots, x_j - \frac{1}{N}, \dots, x_M | t)$  in Taylor's series in the vicinity of the point  $\epsilon = 1/N = 0$  up to the first order over  $\epsilon$ . In the zero order over  $\epsilon$  one gets identity. In the linear approximation (2) takes the form of continuity equation:

$$\frac{\partial \Phi(x_1, \dots, x_M | \tau)}{\partial \tau} + \sum_i^M \frac{\partial J_i}{\partial x_i} = 0, \quad (9a)$$

$$J_i = \Phi(x_1, \dots, x_M | \tau) \sum_{j \neq i}^M (P_{ji} \lambda_j x_j - \lambda_i x_i P_{ij}), \quad (9b)$$

$J = \{J_1, \dots, J_M\}$  can be treated as probability current. Multiplying the components  $J_i$  by the total number of moving citizens, we get the values of flows through the node  $i$ . The solutions of (9) that correspond to stationary states are of particular interest. In such states

$$\frac{\partial \Phi(x_1, \dots, x_M | \tau)}{\partial \tau} = 0, \quad (10a)$$

And correspondingly:

$$J_i = \Phi(x_1, \dots, x_M | \tau) \sum_{j \neq i}^M (P_{ji} \lambda_j x_j - \lambda_i x_i P_{ij}) = \text{const}_i \quad (10b)$$

It is easy to show that total current:

$$J = \sum_{i=1}^M J_i = 0. \quad (10c)$$

We'll consider the case  $J_i = 0$ . This means that in stationary state summary currents between each of nodes  $i=1,2,\dots,M$  and their surroundings are absent (or that is the same, the currents from each node  $i$  and to it are equal). In this case one gets from (10b) the equations for intensities of stationary flows from the nodes  $i=1,2,\dots,M - J_i = \lambda_i \cdot x_i$ :

$$\sum_k^M P_{ki} J_k = J_i \quad (11)$$

In accordance with (3e) matrix  $P_{ki}$  is stochastic and consequently has eigenvector  $\{j_1^s, \dots, j_M^s\}^{\text{Transposed}}$ , corresponding to eigenvalue equal to unity [12,13]. Stationary concentrations

$x_i^s$  can be defined either as  $x_i^s = j_i^s / \lambda_i$ , or directly from the equation for  $x_i$ , by means of the algorithm like PageRank [12,13].

One can transform the equation for  $x_i$  to discrete form, taking as the unit of time in (6a) –  $\lambda^{-1}$  defined by (4):

$$x_i(t + 1) = \sum_{k=1}^M M_{ik} x_k(t), \quad (12a)$$

Where:

$$M_{ik} = \delta_{ik}(1 - \lambda_i) + P_{ki} \lambda_k \quad (12b)$$

And other parameters are defined by (3). Either matrix  $P_{ki}$ , or  $M_{ik}$  can be treated as indecomposable. It is evident that (12) describes the random walk over the vertexes of an oriented graph and element  $M_{ik}$  corresponds to the edge connecting vertex  $i$  with vertex  $k$ . One can show that for each column of matrix  $\widehat{M}$  (12b) the sum of all elements is unity. Thus  $\widehat{M}$  can be treated as the stochastic matrix of the transition's probabilities for such random walk. In the limit  $t \rightarrow \infty$ , (12a) takes the form of an eigenvalue problem for matrix  $\widehat{M}$  with the eigenvalue equal to unity [12,13]. The components of the corresponding eigenvector describe the stationary state achievable in the limit  $t \rightarrow \infty$ ,

$$\lim_{t \rightarrow \infty} x_i(t + 1) = \lim_{t \rightarrow \infty} x_i(t) = x_i^s \quad (13)$$

It can be shown that this stationary state is stable.

The solution of (9) can be found by standard routine [14]:

$$\begin{aligned} & \Phi(x_1, \dots, x_M | \tau) \\ &= \exp\left(\tau \sum_{i,j \neq i} \lambda_j P_{ij}\right) \Psi[e^{\lambda_1 P_1^* \tau} u_1, \dots, e^{\lambda_M P_M^* \tau} u_M] \quad (14a) \end{aligned}$$

Here  $\Psi[\dots]$  – an arbitrary function, of  $M$  independent first integrals of the auxiliary system:

$$\frac{dx_1}{u_1(x_1, \dots, x_M)} = \dots = \frac{dx_M}{u_M(x_1, \dots, x_M)} = d\tau \quad (14b)$$

$$u_i(x_1, \dots, x_M) = \sum_{k=1}^M P_{ki} j_k - j_i = \sum_{k=1}^M \widehat{U}_{ik} j_k \quad (14c)$$

The concrete form of  $\Psi[\dots]$  is defined by the initial conditions on  $\Phi(x_1, \dots, x_M | \tau)$ :

$$\Phi(x_1, \dots, x_M | \tau = 0) = \Phi_0(x_1, \dots, x_M) \quad (14d)$$

Let's assume, for example, that the initial distribution is uniform in some hypercube of the space  $\{x_1, \dots, x_M\}$ :

$$a_i \leq x_i \leq b_i, i = 1, \dots, M \quad (15)$$

and is equal to zero otherwise. One can show, that for  $\tau \rightarrow \infty$ ,  $\Phi(x_1, \dots, x_M | \tau \rightarrow \infty)$  exponentially tends to infinity. The region where  $\Phi(x_1, \dots, x_M | \tau) > 0$  is defined by the equations:

$$\sum_{j=1}^M \widehat{U}_{ij}^r e^{-\lambda_j P_j^* \tau} a_j \leq \sum_{k=1}^M P_{ki} j_k - j_i \leq \sum_{j=1}^M \widehat{U}_{ij}^r e^{-\lambda_j P_j^* \tau} b_j, \quad (16)$$

$$\sum_{j=1}^M \widehat{U}_{ij}^r e^{-\lambda_j P_j^* \tau} b_j, \quad (16)$$

$$i = 1, \dots, M.$$

Here  $\widehat{U}^r$  is submatrix of  $\widehat{U}$  (14c) with  $\text{rang } r < M^1$ . One can show that in the limit  $\tau \rightarrow \infty$  the inequalities (15) can be satisfied only if  $x_i(\tau \rightarrow \infty) \rightarrow x_i^s$ . Thus, the probability distribution  $\Phi(x_1, \dots, x_M | \tau)$  in the limit  $\tau \rightarrow \infty$  has the form of sharp –  $\delta$  – peak

$$\Phi(x_1, \dots, x_M | \tau \rightarrow \infty) \rightarrow \delta(x_1 - x_1^s) \cdots \delta(x_M - x_M^s) \quad (17)$$

**The feedback effects:** Instead of solving multi-node problem one can use more simple approximations. For example, the most probable concentrations of citizens in the nodes, where intra-city flows are crossed can be estimated in two nodes model: X is the node of interest, and Y all other nodes (XY – model). For analysis of flows between nodes  $i$  and  $j$  one can use three nodes model:  $i$  is node X,  $j$  – Y, and all other nodes are united into node Z (XYZ – model). For instance, in Ekaterinburg standard deviation between concentrations  $x_s$  defined in the multi-node and XY models did not exceed some percent. So, such models are quite appropriate to treat some simple one – two node tasks when the mutual influence of the processes in different nodes is not essential. Moreover, XY-model is very convenient to study feedback effect – an influence of people concentration in the nodes on the probabilities of their replacements between these nodes. In the XY – model the equation (9) can be reduced to the equation with the only one variable and solved analytically. If  $x, y$  – are the concentrations of citizens in the nodes X and Y ( $x+y=1$ ) the probability distribution of concentrations  $x$  or  $y$  is:

$$\Phi(x, t) = \frac{z[W^{-1}(t+W(x))]}{z(x)} \Phi_0[W^{-1}(t+W(x))], \quad (18a)$$

where  $W^{-1}(\dots)$  is the function reversal to:

$$W(x) = \int \frac{dx}{z(x)} + c \quad (18b)$$

$c$  is a constant, that does not influence the results,  $\Phi_0[\dots]$  – the initial probability distribution:

$$z(x) = P_+(x - x_s) \quad (18c)$$

Here  $x_s$  is a singular point of (6), corresponding to stationary state  $\frac{\partial \Phi(x, t)}{\partial t} = 0$ :

$$P_+ = \lambda_y P_{yx} + \lambda_x P_{xy} \quad (18d)$$

$$x_s = \frac{\lambda_y P_{yx}}{\lambda_y P_{yx} + \lambda_x P_{xy}} \quad (18e)$$

The solution (18) allows to analyse the peculiarities of  $\Phi(x, t)$  for the tasks where transition probabilities  $P_{yx}, P_{xy}$  or flow intensities  $\lambda_y, \lambda_x$  in (18c) depends upon  $x$  ( $y$ ). The last, results in dependence of  $x_s(x)$ . In this case the equation (6) can have several singular points. All of them are the solutions of transcendental equation:

$$x = x_s(x) \quad (19)$$

Let's consider uniform initial distribution (all concentrations  $x$  are equivalently possible), so  $\Phi_0(x) = \theta(x)$ , where  $\theta(x)$  is Heaviside step function. Expanding  $z(x)$  in Taylor's series in the vicinity of any solution of (19) –  $x_*$  one finds:

$$\Phi(x, t) = e^{\beta t} \{ \theta(x - x_{min}) - \theta(x - x_{max}) \} \quad (20a)$$

Here:

$$\beta = P_+(x_*)[1 - x'_S(x_*)], \quad (20b)$$

And:

$$x_{min} = x_*(1 - e^{-\beta t}), \quad (20c)$$

$$x_{max} = x_*(1 - e^{-\beta t}) + e^{-\beta t}$$

Thus, if  $x'_S(x_*) \leq 1$ ,  $\beta$  is positive and  $P(x_*, t \rightarrow \infty) \rightarrow \infty$ , while the interval for  $x$   $[x_{min}, x_{max}]$  becomes more and more narrow. For  $x'_S(x_*) > 1$  there are no singularities in the  $P(x_*, t)$  behavior. To check the above-mentioned features, we realized the numerical solution of (3) for two opposite patterns of citizen's behavior:

- The avoidance of the places with small people concentrations.
- The tendency to visit such places.

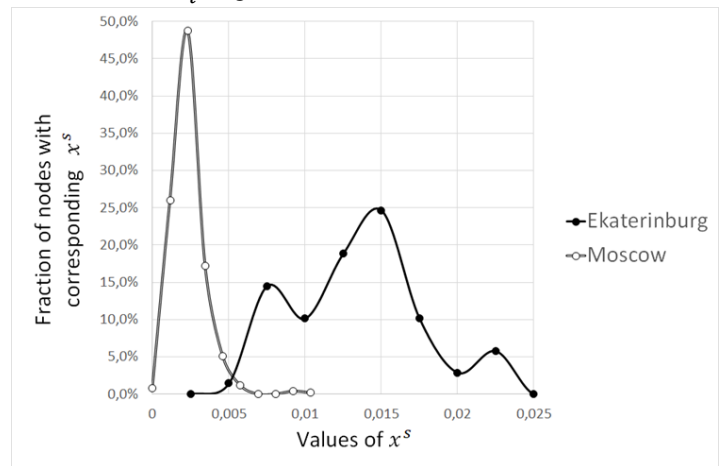
Suppose that  $X$  is the investigated place and  $Y$  all other places of the city available for visits. For simplicity, let's assume that in (18)  $\lambda_x = \lambda_y$ . Figure 3 shows the probabilities of  $XY$ -model for the two above cases and the time evolutions of the corresponding probabilities distributions (Figure 3).

**Figure 3:** The transitions ( $X \rightarrow Y, Y \rightarrow X$ ) probabilities  $P_{xy}, P_{yx}$  and corresponding probabilities distributions for different patterns of citizens behavior in the movements: a) the avoidance of the places with small people concentrations; b) the tendency to visit places with small people concentrations; a1, b1 – transitions probabilities; a2, b2 – the graphical determination of singular points – the solutions of the equation (19) –  $x_*$ ; a3, b3 – the results of computer modelling – numerical solutions of (2) for  $XY$ -model with  $N=1000$ . The equation was solved recurrently step by step from initial probability distribution to the time step –  $\tau=S$ . Time unit for each step was  $\lambda^{-1}$  defined by (4).

It is worth to mention that in the case of avoiding places with small people concentrations the equations (2), (9) have two stable singular points –  $(x_{*1}, x_{*3})$  and one unstable –  $x_{*2}$ . In the case of attractiveness of such places the stable point is only one –  $x_{*1}$ . These results match with the conclusions made above on the base of (20).

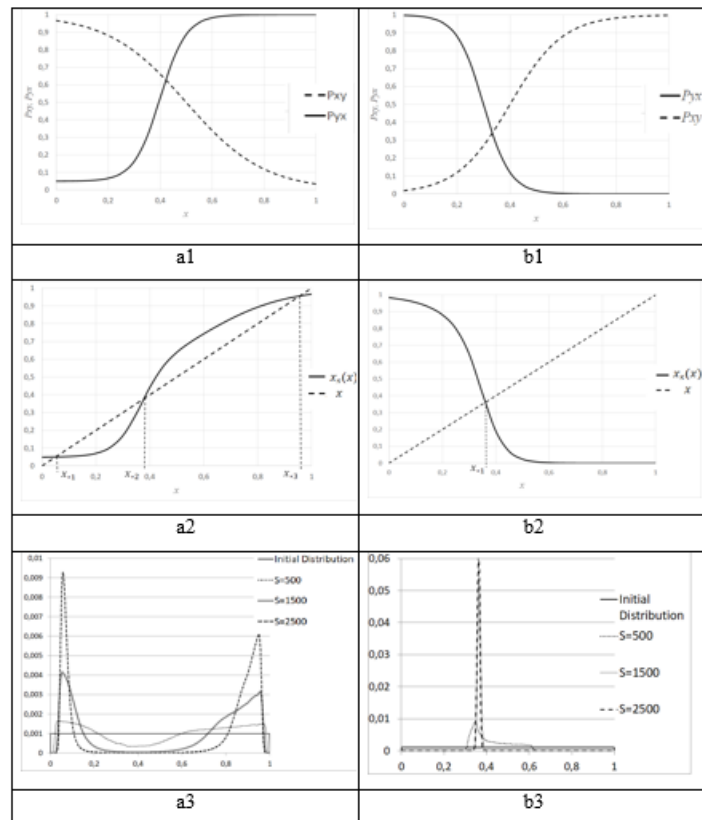
## Results

In this section we are going to apply the above models to the analysis of people flows in two Russian cities – Moscow, and Ekaterinburg (the native town of the authors). The hexagons for Ekaterinburg were grouped into 69 nodes, corresponding to established administrative division, large shopping centers were also considered as independent nodes. In Moscow we simply grouped neighboring hexagons to reduce their number to observable quantity. Figure 4 shows the distributions of stationary concentrations  $x_i^S$  (Figure 4).



**Figure 4:** The fraction of given values of  $x_i^S$  – the most probable concentrations of citizens in the nodes of Ekaterinburg and Moscow.

The statistical parameters of  $x_i^S$  are presented (Table 1). The presented data show that distribution of people concentration over the nodes is highly heterogeneous. The shift of the distribution for Moscow to the area of low concentrations compared to Yekaterinburg is likely caused by studying the people flows between larger administrative entities in the latter



city. Quite expectedly, the maximum  $x_i^S$  values correspond to the shopping centers, hotels, train stations, and airports. At the different times of the day, the maximum  $x_i^S$  is reached at different nodes. For example, in Ekaterinburg the moving of the position with maximum people concentration during the day occurs in the square 11×13 km in the central part of the city. For longer time intervals the distributions of  $x_i^S$  are more stable. For example, in Ekaterinburg the cosine distances between sets  $\{x_1^S, \dots, x_M^S\}$  defined in average on month's data did not exceed  $11^0$  [12].

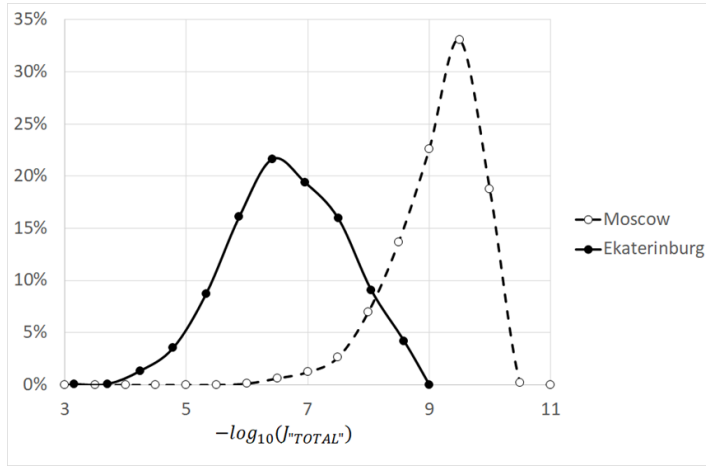


Figure 5: The probability density functions of variable  $-\log(J^{TOTAL})$  (21c)

Table 1: The statistical characteristics of  $x_i^S$ .

City	Average value – $\langle x_i^S \rangle$	Standard deviation – $\sigma_s$	Variation coefficient – $\frac{\sigma_s}{\langle x_i^S \rangle}$	Median
Moscow	1,97E-03	1,85E-03	93,6%	1,49E-03
Ekaterinburg	1,45E-02	7,93E-03	54,7%	1,30E-02

Figure 5 shows the distribution of flows intensities in the stationary states for two above mentioned cities. To estimate the total load on the city's road network we summarize the flows "To the node i":

$$J_i^{TO} = \sum_{j \neq i}^M P_{ji} \lambda_j x_j^S \quad (21a)$$

And "From the node i":

$$J_i^{FROM} = \lambda_i x_i^S \sum_{j \neq i}^M P_{ij} \quad (21b)$$

For each node. So, the total flow for the node i in both directions is:

$$J_i^{TOTAL} = J_i^{FROM} + J_i^{TO} \quad (21c)$$

It is worth to remind that absolute values of flows intensities are given by multiplying of (21c) by the total number of citizens in flows (Figure 5).

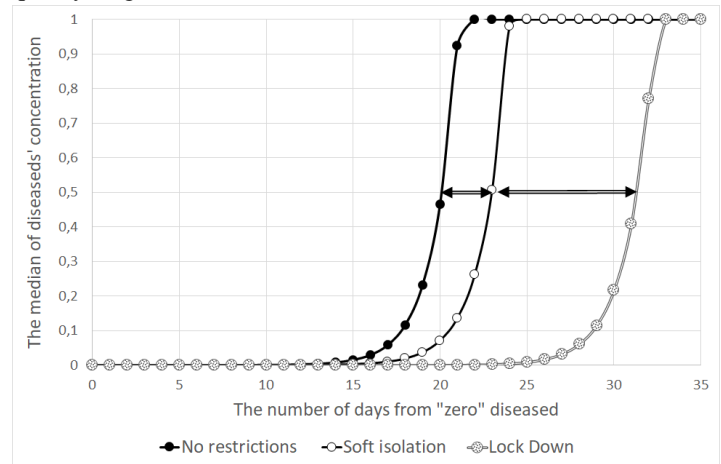
Detailed analysis of people flows shows that most citizens are moving on small distances. For example, in Ekaterinburg 88%, and in Moscow 72% of citizens move on the distances less than 1750 m. So, the analysis of pedestrian's traffic is of high practical importance. Such analysis is realized in the special service "Pedestrian Traffic", built on big data from SBER. This service uses access to more than 75% of transactions activity of clients - physical persons. The service allows to track and prognoses daily and seasonal pedestrian activity, to build a map of the most popular pedestrian routes as well as to record amount of people living in a certain area for both city center and suburbs.

The proposed model can also be used in such, unfortunately currently required, activity as planning of anti-epidemic measures. Really, using it one can trace the flows of infected persons, starting from "zero" diseased. The information about coefficient of the infection spread ( $R_t$  - the number of people, who can be infected on average by one diseased) makes it possible to estimate the statistical characteristics of the disease's spread over the concrete city by means of the proposed model. As an example, Fig. 6 shows the time evolution of a contagious disease in Ekaterinburg, under various restrictions on the people flows. Two types of restrictions were considered:

**Soft isolation:** The restrictions of the most significant flows between the administrative districts of the city, and conserving of free movement within each of them;

**Lock down:** The restrictions of all significant flows both between the administrative districts of the city and inside them.

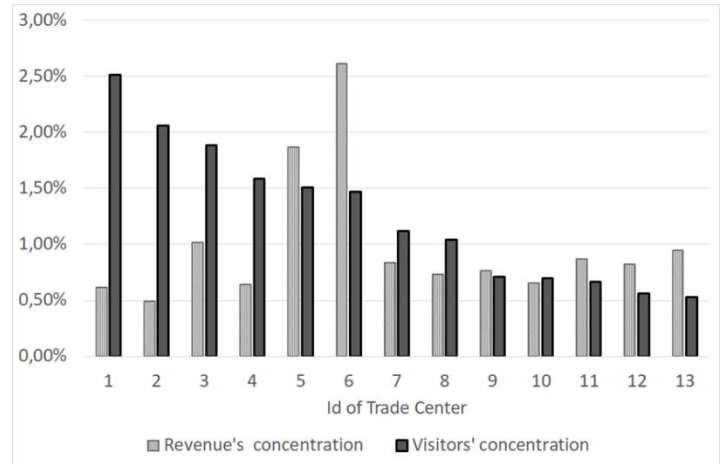
It was assumed that the infection spread coefficient was  $R_t = 1$ . Fig. 6 shows that restrictions on the people flows over the city significantly slows down the spread of the disease. This gives the healthcare system the time to prepare for receiving the patients. The proposed model allows to introduce such restrictions very selectively. So, the slowing down of the disease's spread can be made without serious problems for business activity and life quality (Figure 6).



**Figure 6:** The influence of the restrictions on people flows on the velocity of the disease's spread, calculated on the base of model, proposed in the section 2.2. Arrows show the time reserve earned by healthcare system.

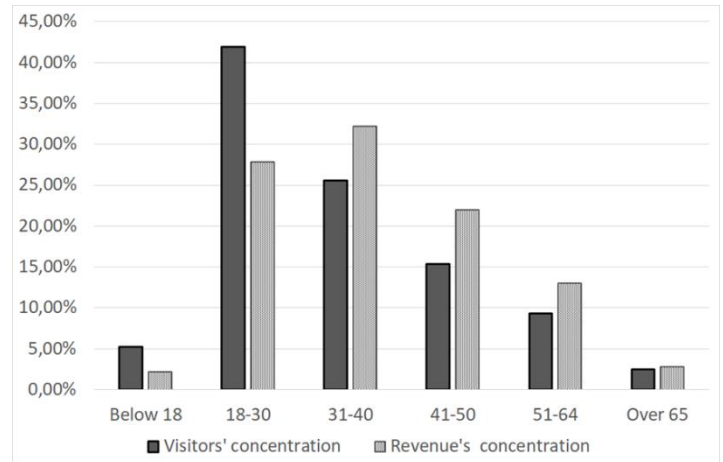
### Discussion

As it was shown in Section 2.2, the stationary states are achieved during a sufficiently long process of people random replacements over the city with the given probabilities of transitions between the nodes. This allows to interpret  $x_i^S$  as the most probable concentration of people in the nodes  $i=1, M$ . The fact that  $x_i^S$  is the most probable concentration is also indicated by the results of analytical solutions, of the model's basic equations and numerical simulation data Figure 3. Thus, one can use  $x_i^S$  as the prognosis of people concentrations in the studied nodes, and correspondingly the flows (10b), (21) can be used as the prognosis of people flows over the city [14-16]. The data for determination of model's parameters ( $P_{ij}, \lambda_i$ ) were obtained from bank transactions, so in fact such data allows to analyze the movements not all the citizens but only "buyers". This has its own pros and cons. Pros: the data contains information not only about cardholders' flows, but about the sums paid by them at each place as well. So, the model can be used to trace and prognoses the effective demand. Moreover, such demand can be tracked down to different ages and gender groups of buyers as well as groups commodities and services. Cons: the credibility of the applying forecasts based on the data about buyers' flows to the flows of the entire population is discussible and should be scrutinized in each concrete case. As an example of pros, we applied our model to determine the most probable concentration of citizens (for different ages groups) in main shopping centers of Ekaterinburg and compare the data obtained with the concentrations of money spent by buyers in these centers. The results are shown on Figure 7. Notice that shopping centers with the maximum concentration of visitors are not always the leaders in revenue's concentration. This is a food for the reflection of the centers' administrations. The same can be said about Figure 8 where the distributions of visitors and revenues connected with different ages groups are presented (Figure 7,8).



**Figure 7:** The visitors' and revenue's concentrations in thirteen main shopping centers of Ekaterinburg.

To study the potential disadvantages of the above approach, we analyzed the passenger's flows of bus public transport in the Sverdlovsk region. The data obtained within the framework of the proposed model were compared with the direct counting of the tickets sold. It was shown that the proposed model correctly reflects the trends in changings of passenger's flows by the days of a week and time of a day. The passenger concentrations at the stopping points, predicted by the model, worse agree with the actual data, although their relative values at different stopping points are reflected more or less correctly. Apparently, this is related to the above-mentioned feature of the model - the analysis of the buyers of goods along the route, and not all the passengers.



**Figure 8:** The visitors' and revenue's concentrations in the different ages' groups of the shopping center with maximum revenue in Ekaterinburg.

### Conclusions

The possible model for the analysis of pedestrian and passenger flows based on the analysis of data on payments by bank cards during the movements across the routes were considered. We



formulated the algorithms, allowing to predict the places of the most probable concentration of citizens in the process of intercity flows and the magnitudes of passenger and pedestrian flows. It should be emphasized that the proposed model creates the possibility to predict not only the flows of citizens, but also the flows of the effective demand along the routes of people flows. We express our gratitude to the employees of SBER Analytics Company for providing the data for our analysis.

## References

1. Lohan ES, Kauppinen T, Debnath SBC. A survey of people movement analytics studies in the context of smart cities. In Proceedings of 19th Conference of Open Innovations Association (FRUCT). 2016; 151-158.
2. Brusyanin DA, Vicharev SV, Sheka AS. Intellectualnaya Sistema analiza passagiropotokov s ispolsovaniem technicheskogo zreniya. Transport Urala. 2012; 2: 86-89.
3. Kilger M. A shadow handler in a video-based real-time traffic monitoring system. In Proceedings of the IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, USA. 1992; 11–18.
4. ITLINE. 2022.
5. Namiot D, Pokusaev O, Chekmarev A. Use of telecommunications operators' data in transport planning. Int J Open Information Technologies. 2019; 7: 51-59.
6. Kaisheng Z, Mei W, Bangyang W, Daniel S. Identification and prediction of large pedestrian flow in urban areas based on a hybrid detection approach. Sustainability. 2017; 9: 36-51.
7. Kuhn M, Johnson K. Applied predictive modelling, springer Science Business Media LLC New York. 2013; 42-506.
8. Farrahi K, Gatica-Perez D. Discovering routines from large-scale human locations using probabilistic topic models. Association Computing Machinery. 2011; 1: 1-27.
9. Sinitsyn EV, Tolmachev AV, Laptev VM. Model of socio-economic factors in the spread of SARS-CoV-2 across Russian regions. In Proceedings of VIII International Scientific Conference New Trends, Strategies and Structural Changes in Emerging Markets (NTSSCEM 2021), SHS Web of Conferences. 2021; 2-13.
10. Tolmachev AV, Sinitsyn EV, Brusyanin DA. Transport system modelling based on analogies between road networks and electrical circuits. R-Economy. 2019; 5: 92-98.
11. Feller W. An introduction to probability theory and its applications. John Wiley. 1957; 427-466.
12. Leskovec J, Rajaraman A, Ullman JD. Mining of massive datasets. Cambridge University Press. 2020; 182-200.
13. Gantmacher FR. Teoriya Matrits. Moskva, Nauka. 1988; 332-375.
14. Elsgolts LD. Differentsialnie uravneniya. Moskva, LKI. 2014; 257-273.
15. Arandiga F, Baeza A. A Spatial-Temporal model for the evolution of the COVID-19 pandemic in Spain including mobility. Mathematics. 2020; 8: 1677-1686.
16. Hazarie S, Soriano-Panos D, Arenas A. Interplay between population density and mobility in determining the spread of epidemics in cities. Commun Phys. 2021.