УНИВЕРСИТЕТ ИТМО

Role discovery in node-attributed public transportation networks: the model description

Yuri V. Lytkin[1], Petr V. Chunaev[2]✉, Timofey A. Gradov[3], Anton A. Boytsov[4], Irek A. Saitov[5]

[1,2,3,4,5] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] jurasicus@gmail.com, https://orcid.org/0000-0001-8140-010X
[2] chunaev@itmo.ru✉, https://orcid.org/0000-0001-8169-8436
[3] timagradov@yahoo.com, https://orcid.org/0000-0003-2537-4087
[4] aboytsov@itmo.ru, https://orcid.org/0000-0001-8343-2519
[5] xanilegendx@gmail.com, https://orcid.org/0000-0002-2805-1323

Abstract

Modeling public transport systems from the standpoint of the theory of complex networks is of great importance to improve their efficiency and reliability. An important task here is to analyze the roles of nodes and weighted links in the network, respectively modeling groups of public transport stops and their linking routes. In previous works, this problem was solved based on only topological and geospatial information about the presence of routes between stops and their geographical location which led to the problem of uninterpretability of the discovered roles. In this article, to solve the problem, the model additionally considers information about the social infrastructure around the stops and discovers topological, geospatial, and infrastructure roles jointly. The public transport system is modeled using a special weighted network — with node attributes where nodes are non-overlapping groups of stops united by geospatial location, node attributes are vectors containing information about the social infrastructure around stops, and weighted links integrate information about the distance and number of transfers in routes between stops. To identify the model, it is sufficient to use only open urban data on the public transport system. Role discovery for stops is carried out by clustering network nodes in accordance with their topological and attributive features. An extended model of the public transport system and a new approach to solving the problem of discovering the roles of stops, providing interpretability from the topological, geospatial and infrastructural points of view, are proposed. The model was identified on the open data of Saint Petersburg about metro stations, trolleybus and bus stops as well as organizations and enterprises around the stations and stops. Based on the data, balanced parameters for grouping stops, assigning link weights and constructing attribute vectors are found for further use in the role discovery task. The results of the study can be used to identify transport and infrastructure shortcomings of real public transport systems which should be considered to improve the functioning of these systems in the future.

Keywords

node-attributed network, public transportation network, role discovery, network node classification, network topology, social infrastructure

340

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

УДК 004.94

Выделение ролей в сетях общественного транспорта с атрибутами узлов: описание модели

Юрий Всеволодович Лыткин[1], Петр Владимирович Чунаев[2✉],
Тимофей Алексеевич Градов[3], Антон Алексеевич Бойцов[4], Ирек Аликович Саитов[5]

[1,2,3,4,5] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] jurasicus@gmail.com, https://orcid.org/0000-0001-8140-010X
[2] chunaev@itmo.ru✉, https://orcid.org/0000-0001-8169-8436
[3] timagradov@yahoo.com, https://orcid.org/0000-0003-2537-4087
[4] aboytsov@itmo.ru, https://orcid.org/0000-0001-8343-2519
[5] xanilegendx@gmail.com, https://orcid.org/0000-0002-2805-1323

Аннотация

Предмет исследования. Моделирование систем общественного транспорта с позиций теории сложных сетей имеет большое значение для повышения их эффективности и надежности. При этом важной задачей является анализ ролей узлов и взвешенных связей в сети, моделирующих группы остановок общественного транспорта и связывающие их маршруты. В настоящее время данная задача решена на основе только топологической и геопространственной информации о наличии маршрутов между остановками и их географическом положении. Такое ограничение приводит к проблеме неинтерпретируемости выделенных ролей. Для решения проблемы определения ролей в сетях в предложенной модели транспорта дополнительно учтена информация о социальной инфраструктуре вокруг остановок, а также осуществлено выделение совместно топологических, геопространственных и инфраструктурных ролей. **Метод.** Система общественного транспорта смоделирована с помощью специальной взвешенной сети — с атрибутами узлов. При этом узлы — непересекающиеся группы остановок, объединенные по геопространственному положению, атрибуты узлов — векторы, содержащие сведения о социальной инфраструктуре вокруг остановок, а взвешенные связи интегрируют информацию о расстоянии и количестве пересадок в маршрутах между остановками. Для идентификации модели достаточно использовать только открытые городские данные о системе общественного транспорта. Выделение ролей остановок выполнено путем кластеризации узлов сети в соответствии с их топологическими и атрибутивными признаками. Основные результаты. Предложена обобщенная модель системы общественного транспорта. Представлен новый подход решения задачи выделения ролей остановок, обеспечивающий интерпретируемость с топологической, геопространственной и инфраструктурной точек зрения. Модель идентифицирована на открытых данных Санкт-Петербурга об остановках подземного, троллейбусного и автобусного транспорта, а также организациях и предприятиях вокруг остановок. На основе данных найдены сбалансированные параметры объединения остановок, назначения весов связей и построения векторов атрибутов для последующего использования в задаче выделения ролей. Практическая значимость. Результаты исследования могут быть использованы для определения транспортных и инфраструктурных недостатков реальных систем общественного транспорта, которые следует учитывать для улучшения функционирования этих систем в будущем.

Ключевые слова

сеть с атрибутами узлов, сеть общественного транспорта, выделение ролей, классификация узлов сети, топология сети, социальная инфраструктура

Introduction

In recent years, network theory has found its way into a variety of fields of science and technology. A network is a collection of *nodes* some of which are connected by *links*. Being so simply constructed and versatile simultaneously, networks become very useful in analyzing, modeling, and studying all sorts of complex systems such as online and offline social networks, computer and technological networks, biological and brain networks, transportation networks, etc.

The study of public transportation systems from a network theory perspective started rather recently [1, 2]. Most works on this topic are aimed at analyzing the topological structure of *public transportation networks* (or PTNs) of different cities (e.g., in Poland [2], Hungary [3], China [4]) with regard to various modes of transportation like bus [2] or subway [5]. Usually, in these cases the underlying network is defined with bus stops or subway stations as nodes and some rule to assign links between these stops and stations. The links are mainly unweighted although there are studies considering PTNs as *weighted* [6] (with references therein), where weighted bus PTNs are analyzed by means of common network characteristics.

In addition to the PTN topology, it is also usual to consider the geospatial information about the nodes in the network. A popular approach that utilizes the geography of nodes is combining sets of closely situated nodes into groups called *supernodes* [4, 7]. Such approach is motivated by the fact that people usually take walks between closely

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

341

positioned stops to make a connection, instead of sticking to a strict path through the network. Therefore, such supernode networks are more precise at modeling how people use public transport. From another perspective, some studies (e.g., [8]) consider PTNs as geospatial ones so that the spatial configuration and topology of the network are used for the identification of macroscopic and mesoscopic statistical network characteristics.

Furthermore, another notable source of information that can be used in the public transportation system analysis is social infrastructure surrounding stations and stops that, in a sense, may provide "semantics" to a PTN. For instance, it can be used to analyze and model transport accessibility [9] or as an additional component for measuring PTN transportation efficiency [7].

As far as we know, the union of weighted geospatial networks (supernodes and weighted links) and node semantics (social infrastructure in our case) have not been considered in the PTN studies (as the so-called *node-attributed networks*), although it may certainly enrich our knowledge about processes of PTN formation. This is confirmed by the case of node-attributed networks modeling online social networks where not only connections between social actors (network topology) but also actors' content (profile information, posts, etc.) are taken into account within different tasks such as community detection, link prediction, outlier identification, etc., e.g., [10–12]).

To get closer to the objective of our study, let us also mention that in the recent times *role discovery* (especially topological feature-based [13]) has become a popular topic, most notably in the domain of *non-attributed social* network analysis [13–19]. In the network context, roles refer to clusters or classes of nodes where the nodes from the same cluster are structurally similar to each other in some way. The problem of role discovery is related to another network clustering problem called *community detection* in non-attributed [20–22] and node-attributed [10–12] social networks where the clustering mainly aims to separate densely interconnected parts (called communities) of the network by means of network topology or both network topology and attributes (semantics). By contrast, role discovery aims to distinguish between various structural and other characteristics of different nodes. For instance, in a social network there can be multiple communities of people, and in each community there are people of various roles, i.e., leaders, influencers, etc., with possible transitions between roles and interaction preferences (see the recent studies on the topic, e.g., in [23–25]). Let us here specifically mention the study [26] as it seems the first attempt to enrich role discovery methodology in social online networks by the content generated by social actors ("semantics"). Although the authors do not explicitly model online social networks as *node-attributed* networks, the experimental results in [26] show that the semantics helps to identify social network roles more effectively.

In this study, we consider the experience of studies in social network analysis connected with role discovery in non- and node-attributed social networks to model and analyze PTNs. Furthermore, we are motivated by the survey [27] where PTNs are considered from the network perspective of complexity, static and dynamic resilience,

and it is emphasized that the study of PTN node roles (in particular, based on topological features — besides the well-known *hubs*, for example) is still limited although may offer useful insights into identifying the most critical nodes of PTNs. Namely, we propose an approach for solving the novel problem of role discovery for weighted node-attributed PTNs that can discover roles both in terms of network topology and node infrastructural attributes — semantics. In short, the main contributions of this paper are the following:

1. We model a PTN as a weighted node-attributed network where nodes are *supernodes*, i.e., groups of public transport stops and stations grouped with respect to their geospatial position, and node attributes are numerical vectors storing information about social infrastructure around the supernodes. The weighted links in the network integrate information about the travelling distance and the number of hops in the transportation routes between the supernodes.

2. We point out some of the common misconceptions and errors in previous analyses of the PTNs which we believe stem from the misunderstanding of some interpretations of different PTN models.

3. We propose a new approach for role discovery in weighted node-attributed networks. This approach uses semantics (i.e., node attributes) as well as structure (i.e., network topology). In the context of PTNs, this approach allows to discover meaningful roles in terms of both topological structure of stops and stations and social infrastructure around them. At the same time, the approach is not topic-specific and can be applied in other domains like social network analysis.

4. We test the framework on the newly collected open public transportation data of Saint Petersburg, Russia. It is shown to be capable of discovering different roles of public transport stops in terms of both structure and social infrastructure and extracting useful information about the overall PTN's transport and social infrastructure efficiency.

Let us additionally mention that with respect to previous studies, we

— define the supernodes formally as equivalent classes to avoid ambiguity, with the choice of reasonable thresholds;

— choose a trade-off between hop-based and distance-based routes to balance between the travelling distance and the number of hops corresponding to a given route between two nodes in a PTN;

— define the problem of social infrastructure role discovery and propose a procedure for constructing social infrastructure attributes in our model;

— scrupulously select and analyze commonly used topological features of network nodes in the context of PTN models.

Related work

Modeling public transportation networks. The study of PTNs using network (graph[1]) theory began in [1, 2].

---

[1] Here and throughout the paper we use the terms network and graph interchangeably.

342

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

The main aim of such studies is usually to analyze the topology of the given city's PTN in order to extract useful information about the state and structure of that city's public transportation system.

The two most popular ways of constructing a PTN (both were introduced in [1]) are *L-space* and *P-space* models. In both cases the nodes of the network represent various public transportation stops and stations. What these models differ in is the way of assigning the links between the nodes. According to the *L-space* model, a link is assigned between two nodes that correspond to two consecutive stops on some route. Thus, the topology of an *L-space* model is visually similar to a normal scheme of a public transportation system that one can find on an information stand near a bus stop. By contrast, in the *P-space* model, a link is put between all stops that are connected by some route (not just the consecutive ones). Therefore, in the *P-space* model and link are interpreted as a possibility of travelling directly between two nodes. (Note that as a result, the *P-space* model is normally much denser than the corresponding *L-space* model.) The difference between *L-space* and *P-space* is explained in Fig. 1.

These models have been used in virtually all the papers dealing with PTNs and were applied to analyze various cities in Poland [2], Hungary [3], China [4], among others. Such analysis is especially easy to conduct since the data needed to build a basic PTN is nowadays available publicly for most big cities around the world (Fig. 2). Usually, authors aim to check some graph-theoretic and network-theoretic properties of the constructed graphs, i.e., degree distribution, clustering coefficient, scale-free property, and so on. A comprehensive comparison of such properties between different cities around the world can be found in [28] along with interpretations of these properties in a sense of public transportation quality.

Another natural source of information for constructing a PTN is the geospacial component, i.e., the coordinates of the stops. As we mentioned previously, a conventional PTN (with separate stops as nodes) does not account
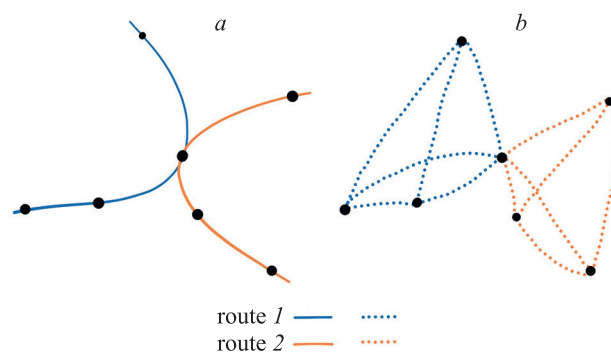

Fig. 1. Difference between *L*-space (*a*) and *P*-space (*b*)

for passengers' possibility to make walking connections between closely situated stops while moving around a city. Additionally, such approaches are not capable of combining different modes of transportation (like bus, trolleybus, tramway, and subway) in a single network. To overcome these issues, one can consider groups of nearby stops and stations as *supernodes* (Fig. 3), thus transforming the conventional node structure into the supernode structure (note that the node links are naturally transformed into the supernode links given the defined node-to-supernode mapping). Such approach was used in [4, 7].

To further improve a public transportation model, one can also assign link weights, see, e.g., [4, 6, 7]. In [7], the authors propose to assign weights to the links of the *L*-space network by counting the number of routes operating of each given link. Such weights can therefore represent the amount of passenger flow via each link. By contrast, the authors of [4] propose to assign link weights (both in *L*-space and *P*-space) as the minimal travel distance between the nodes along the corresponding route. Such approach is more suitable in terms of determining the optimal routes and connections while travelling around a city.

It should be noted that the choice of the network model as well as the method of assigning link weights greatly influences what one can then do with the resulting network
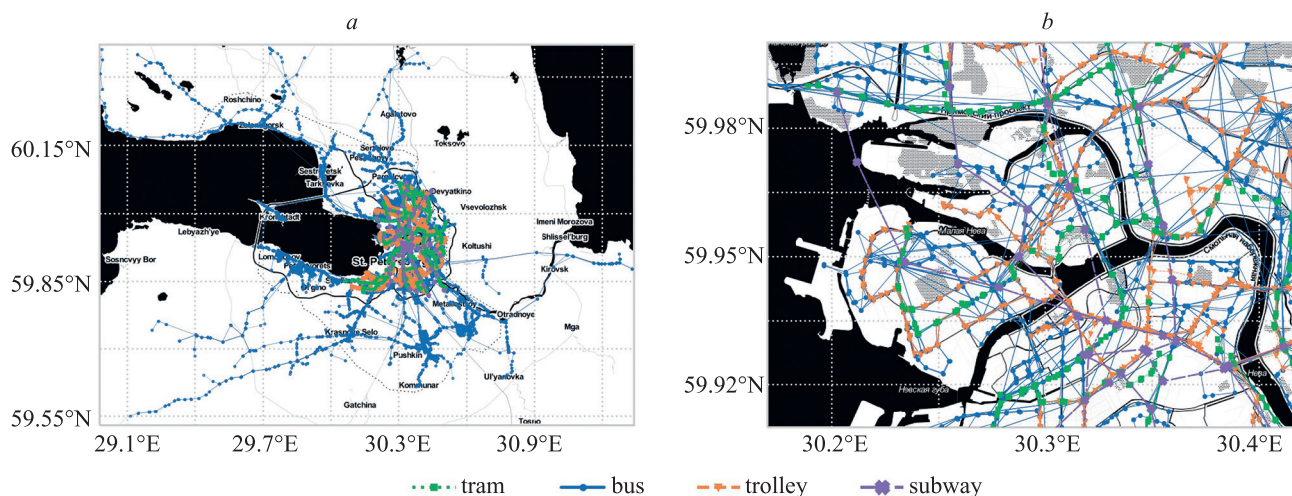
Fig. 2. The map[1] of area surrounding Saint Petersburg (*a*) and the city center (*b*), indicating stops and routes of different modes of transportation

---

[1] The maps are generated by Cartopy, a Python open package. Available at: https://scitools.org.uk/cartopy (accessed 26.09.2022).

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2
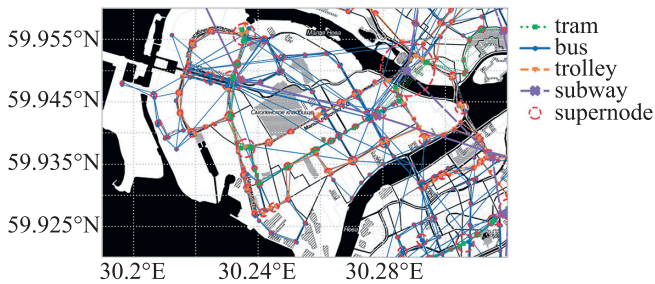
343

Fig. 3. The map[1] of Vasileostrovsky District in Saint Petersburg, indicating stops and routes for different modes of transportation as well as the supernodes (groups of nearby stops)

_____

[1] The maps are generated by Cartopy, a Python open package. Available at: https://scitools.org.uk/cartopy (accessed 26.09.2022).

model. For instance, when using the *L*-space model (as it was done in [7]), one should be careful in interpreting the shortest paths through the network, as these generally do not correspond to how passengers choose to travel in practice, since, for example, the number of connections is not minimized when using such paths while normally a passenger would want to make as little connections as possible (Fig. 4). Such misinterpretation of shortest paths may lead to subsequent misinterpretation of various centrality measures, such as betweenness centrality and closeness centrality.

The *P*-space seems to be better suited for such shortest path analysis although choosing the method of link weight assignment is still very important here. Assigning equal weights to links resolves the issue of minimizing the number of connections since in this case a shortest path through the *P*-space network is precisely the path requiring the minimal number of connections. At the same time, such shortest paths can be excessively long in terms of travelling distance. However, setting travelling distances as link weights (as in [4]) brings back the issue of the number of connections since a shortest path in terms of travelling distance can involve a suboptimal number of connections. Therefore, an intermediate approach is needed, taking into account both the number of hops in a shortest path, and the travelling distance corresponding to it. Such approach is used in our paper (Fig. 5).

There also exist methods of assigning link weights based on the flow of passengers during a certain part of the day [29, 30] resulting in a dynamic structure of the PTN. It should be mentioned, however, that such data is usually
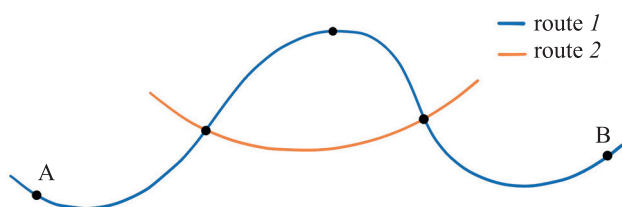
quite hard to obtain while in this paper we aim to construct the model using only the openly available data.

Finally, social infrastructure is also an available and important source of information when constructing a PTN since it sheds light on why people actually travel to a given destination (there can be, for instance, a school, a hospital, or a sightseeing spot nearby). The infrastructural component was used in [7] where the authors assigned node weights depending on a number of factors, such as the number of social infrastructure objects of certain types (recreation, emergency, education, and transportation), the total number of passengers accessing the node, etc. All these factors were then weighted producing a single value which was chosen as the node weight.

This method is useful when trying to access importance (as a unidimensional characteristic) of each node from the infrastructural standpoint. At the same time it does not capture any information about the role of the node, i.e., its unique infrastructural characteristics. Therefore, in this paper we adopt a more general multidimensional approach assigning not weights but attribute vectors to nodes.

Role discovery in public transportation and other networks. The main idea behind the role discovery is to group nodes by their connectivity patterns where each group represents some topological role, such as hub, bridge, near-clique, etc. Topological roles indicate which functions nodes serve in the network [13].

Initially role discovery was the point of interest in sociology, used to study the interactions between social actors and assign roles to actors, but networks in these studies were very small [31, 32]. In general, role discovery can be applied to any network, and the main difference across networks will be in the interpretation of roles. Lately, this concept was studied and implemented for biological networks [33], web graphs [34], and many others [35].

Fig. 4. In the *L*-space model, all consequent stops in each route are connected with a link. As a result, a shortest path in the *L*-space graph generally does not indicate an optimal route for a passenger. For instance, while travelling from point A to point B, the optimal travelling route is route *1* while the shortest path through the graph involves changing to route *2* midway

Fig. 5. In the *P*-space model, both hop-based and distance-based link weights result in shortest paths that are not indicative of optimal routes for passengers. When using hop-based link weights (i.e., each link having weight 1), a shortest path is the one with the least number of connections but it can be arbitrarily long in distance. The contrary holds for distance-based weights: a shortest path is indeed shortest in distance but it can involve arbitrarily many connections in the process. A fused approach (considering both distance and hops) mitigates such problems

344

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

The process of role discovery usually consists of several steps. Firstly, centrality measures (or other chosen features) are chosen and calculated for every node in the network. Following this, nodes are clustered by using vectors of centrality measures. As a result, nodes are grouped by similarity among centrality measures which shows how similar nodes are in terms of topology.

To the best of our knowledge, no purposeful attempts have been made to state and solve the problem of role discovery in the above-mentioned sense for PTNs. Indeed, the survey [27] (where PTNs are considered from the network perspective of complexity, static and dynamic resilience) emphasizes that the study of PTN node roles (in particular, based on topological features — besides the well-known *hubs*, for example) is still limited although may offer useful insights into identifying the most critical nodes of PTNs.

Nevertheless, we can mention, e.g., the study [8], where the geospatial configuration of a PTN is analyzed and some conclusions about the roles of the PTN nodes (by means of importance) are made. Furthermore, the topic-related work [36] aims to detect and analyze node clusters in the intercity transportation networks. The authors propose using a distance measure based on the K shortest paths between a pair of nodes to measure the proximity between all node pairs, and then use the hierarchical clustering method in order to obtain the clusters. The resulting clusters correspond to the groups of nodes that are in close proximity of each other. However, this work is more in line with the problem of community detection than role discovery since these clusters do not reflect different roles of these nodes in the network.

Another notable attempt at geospatial PTN clustering is the work [37] the authors of which introduce a problem of node-attributed spatial graph partitioning. This problem aims at obtaining clusters of nodes that are densely interconnected, homogeneous with respect to their attributes and also meet a certain size constraint in terms of the geographical coordinates of the nodes. Even though this problem can indeed be formulated in terms of PTNs and also accommodate the presence of node-attributed social infrastructure vectors; it is however more in line with community detection in node-attributed networks [10–12] rather than role discovery [13] since in general the nodes of a certain role (like transition hubs, for instance) do not need to be in close proximity of each other.

One should note that the richest experience on the role discovery task is nevertheless in the field of social network analysis where non- and node-attributed networks are deeply studied within the task [13–19]. One can find a comprehensive overview of role discovery approaches in [13] where graph-based, feature-based, and hybrid definitions of roles and methods for their discovery from social network data are discussed. Let us also mention several further studies on the topic.

In [16], a novel role discovery approach is proposed for extracting *soft* roles of social actors with similar behavioral and functional characteristics in online social networks. The study [24] is focused on the problem of research role identification (i.e., principal investigator, sub-investigator or research staff) for large research institutes in which

similar yet separated teams coexist. Furthermore, [25] states and proposes a framework for solving the multiple-role discovery task and conduct an experimental study of their framework on several real-world online document/social networks. Finally, let us mention the study [26] that seems the first attempt to enrich role discovery methodology in social online networks by the content generated by social actors, e.g., posts. In the paper, a novel method which integrates both user behavior and his/her content to identify roles is proposed. Although the authors do not explicitly model online social networks as *node-attributed* networks, the experimental results in [26] show that the semantics helps to identify various roles more effectively and to get more insights on how the network is functioning.

As we have already mentioned, in our study we take into account the experience of studies in social network analysis connected with role discovery in non- and node-attributed social networks to model and analyze PTNs.

Description of the model and the role discovery task

The model of a node-attributed public transportation network. We now proceed to describing the node-attributed PTN model that we are going to use for role discovery later. The data needed to construct such model will be described in detail in a future work, but for now we note that only the general public transportation and social infrastructure data, which is available for the majority of cities around the world, is needed here. Below, we illustrate our model with the PTN data for Saint Petersburg, Russia[1] (will be described and studied in detail in the future work) in order the make it clearer for the reader.

Formally, the model can be defined as a tuple:

$$G = (V, E, A),$$

where V is the set of nodes, $E \subseteq V \times V \times \mathbb{R}$ is the set of undirected weighted links, and $A: V \to \mathbb{R}^n$ is a mapping that defines the set of node-attributed vectors. In what follows, we will define each component of this graph.

Supernodes (nodes of the node-attributed network). The first step is to combine the public transportation stops and stations into supernodes, i.e., groups of nodes that are located close to each other, thus making it possible to make a transition between them on foot. Suppose that $S = \{s_1, \ldots, s_N\}$ is the set of public transportation stops (N in total). To combine them into supernodes, we first need to calculate the pairwise distances between each pair $s_i, s_j \in S$. This can be done using their geographical coordinates. The distances are calculated using the well-known Haversine formula:

$$d(s_i, s_j) = 2r_0 \arcsin\sqrt{\Theta(\varphi, \lambda)}, \qquad (1)$$

where

$$\Theta(\varphi, \lambda) = \sin^2\frac{\varphi_j - \varphi_i}{2} + \cos\varphi_i \cos\varphi_j \sin^2\frac{\lambda_j - \lambda_i}{2},$$

---

[1] The data along with all preprocessing and analysis procedures is available in the Github repository. Available at: https://github.com/AlgoMathITMO/public-transport-network (accessed 26.09.2022).

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

345

$d(s_i, s_j)$ is the distance between stops s_i and s_j; r_0 is the radius of Earth; $\varphi_l, \lambda_l, l \in \{i, j\}$, are latitudes and longitudes of the two points, respectively.

The most common way of grouping the closely situated stops is by using a *distance threshold* [4, 7]: all stops that are closer to each other than some constant d_0 are added to a common supernode. Since this construction is not an equivalence relation, in order to define the supernodes correctly, we also close this relation *transitively*. When this is done, the supernodes are defined as *equivalence* classes with respect to this closed relation, i.e., two stops $s_i, s_j \in S$ belong to the same supernode $\hat{s}$ *if and only if*

$$\exists n_1 = s_i, n_2, \ldots, n_K = s_j \in S:$$
$$\forall k < K \, d(n_k, n_{k+1}) \leq d_0.$$

We denote the set of all supernodes as $\hat{S}$ and use it as the set of nodes V of the graph G. In some practical cases we will also need coordinates of supernodes. For these cases we define coordinates of a supernode as simply the mean of latitude and longitude over all stops belonging to the given supernode.

Note that in general there can be nodes inside a single supernode with distance greater than d_0, provided there is a sequence of nodes

$$n_1 = s_i, n_2, \ldots, n_K = s_j \in S,$$

such that each pair n_k, n_{k+1} is closer than d_0. This can potentially result in some supernodes being arbitrarily large. This issue cannot be resolved in a symmetrical way, and we have no choice but to allow it (even though it has not been discussed in any of the previous papers, we assume that the authors of those papers also faced this issue), but we stress that an appropriate value of d_0 should therefore be chosen carefully, taking into account the sizes of the resulting supernodes (Fig. 6). Some of the characteristics of supernodes that can be considered here is the supernode size (i.e., the number of nodes inside it) or the supernode diameter (i.e., the maximal distance between two nodes inside it).

For instance, in Fig. 6 we see that when $d_0 > 0.1$ (i.e., the distance of 100 meters), the maximal supernode diameter gets beyond 1 km which is not really acceptable as a walking distance between the stops. Therefore, for our study we take $d_0 = 0.1$.

Weighted links of the node-attributed network. The second step is to define the set of links E. This is done traditionally using the information about different routes that comprise the public transportation system. Suppose that R is the set of all public transportation routes where each route is defined as a sequence of stops from S:

$$r = (s_{i_1}, \ldots, s_{i_k}),$$

where k is the route length, and each s_{i_j} is a stop from S. Since each stop $s \in S$ is mapped uniquely to a supernode $\hat{s} \in \hat{S}$, these routes can be easily converted into the sequences of supernodes:

$$\hat{r} = (\hat{s}_{i_1}, \ldots, \hat{s}_{i_l}),$$

where $l \leq k$ and $\hat{s}_{i_j} \in \hat{S}$.

Recall that in the P-space model, links are defined as all pairs of stops (not necessarily consecutive) on all the routes, i.e.,

$$\{(s_i, s_j) \in S^2 | \exists r \in R: s_i, s_j \in r\}.$$

A P-space link, therefore, means that there exists a route connecting the given pair of stops.

In order to assign weights to these links, consider an arbitrary route $r = (s_{i_1}, \ldots, s_{i_k})$ and take two arbitrary stops $s_{i_j}, s_{i_l} \in r, i_j < i_l$. Since there exists a sub-route

$$(s_{i_j}, s_{i_{j+1}}, \ldots, s_{i_l}) \subseteq r,$$

we can define a *route distance* between s_{i_j} and s_{i_l} with respect to the route r as follows:

$$rd_r(s_{i_j}, s_{i_l}) = \sum_{k=j}^{l-1} d(s_{i_k}, s_{i_{k+1}}),$$
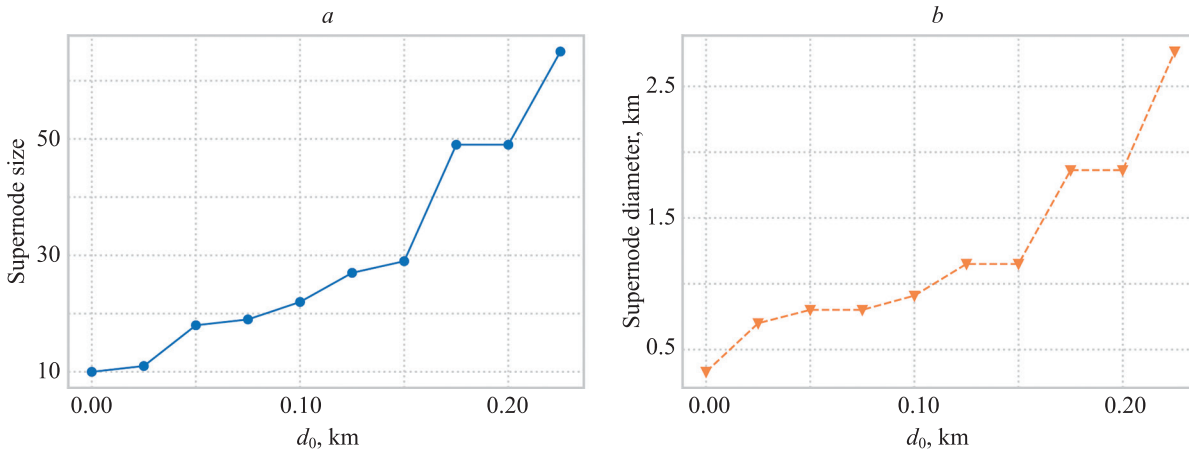
Fig. 6. Maximal supernode size (*a*) and diameter (*b*) for different values of d_0. Even for relatively small values ($d_0 > 0.15$) these characteristics grow quite rapidly resulting in some supernodes having diameter as large as 2 km and more

346

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

where d is the distance defined in eq. (1). Notice that there can be several routes connecting the same pair of stops s_i, s_j, and the corresponding route distances $rd_r(s_i, s_j)$ can vary. We thus define the route distance between two nodes s_i, s_j as the minimal route distance between them across all the available routes:

$$rd(s_i, s_j) = \min_{r \in R} rd_r(s_i, s_j).$$

Route distances were used as link weights in [4], but, as it was discussed above, such approach to assign link weights brings up an issue that a shortest path between two nodes with respect to route distances (being optimal in terms of travel distance) can be suboptimal in terms of the number of connections made while travelling via this path. Using unweighted links solves the problem of minimizing the number of connections but can result in shortest paths that are inadequate in terms of travelling distance.

This issue is illustrated in Fig. 7. In both cases we have two routes between the same pair of stops, and route A is obtained by minimizing the travel distance, while route B is obtained by minimizing the number of hops. In the first case (Fig. 7, *a*) we see that route B, while having less transfers than route A, is about 10 times longer than the latter, therefore it is much less convenient for a passenger. The second case (Fig. 7, *b*) is the opposite: route A is shorter (albeit marginally) than route B, but has 10 times more transfers, and it is very unlikely that a passenger will decide to take route A over route B.

Therefore, an intermediate approach should be adopted. Here we propose the following weighing scheme where weight $w(s_i, s_j)$ is:

$$w(s_i, s_j) = \alpha rd(s_i, s_j) + 1 - \alpha.$$

Here α is the dimensionless coefficient, the term $1 - \alpha$ can be thought of as multiplied by a 'hop-weight' of a link which is always equal to 1. This approach makes it possible to balance between the travelling distance and the number of hops corresponding to a given path between two nodes. We use these values as link weights E in our model:

$$E = \{(\hat{s}_1, \hat{s}_2, w(\hat{s}_1, \hat{s}_2)) | \hat{s}_1, \hat{s}_2 \in \hat{S}\}.$$

In order to choose an appropriate dimensionless value of α, consider the two borderline cases, namely $\alpha = 0$ and $\alpha = 1$. In the first case we get an unweighted graph (each link having weight 1), thus the shortest paths have the minimal possible number of hops. For an arbitrary pair of nodes s_i, $s_j \in S$ denote such minimal number of hops as $H_{\min}(s_i, s_j)$. In the latter case (i.e., $\alpha = 1$) we get a graph weighted with geographical distances along the links, thus the shortest paths in this case are minimal in terms of travel distance. Denote these minimal travel distances as $D_{\min}(s_i, s_j)$, $s_i, s_j \in S$.

Now, for an arbitrary $\alpha \in (0,1)$ notice the shortest paths are sub-optimal in terms of both the number of hops (denote these as $H_\alpha(s_i, s_j)$) and travel distance (denote these as $H_\alpha(s_i, s_j)$). Therefore, we can consider *mean percentage difference* between these values and their corresponding minima, i.e.,

$$MPD_H(\alpha) = \frac{100\,\%}{|V|(|V| - 1)} \sum_{u,v \in V} \frac{H_\alpha(u, v) - H_{\min}(u, v)}{H_{\min}(u, v)}$$

for hops, and

$$MPD_D(\alpha) = \frac{100\,\%}{|V|(|V| - 1)} \sum_{u,v \in V} \frac{D_\alpha(u, v) - D_{\min}(u, v)}{D_{\min}(u, v)}$$

for distances.

These values can be used to determine the optimal value of α. For instance, in Fig. 8 we see that for $\alpha = 0.2$ both MPD_H and MPD_D are less than 10 % which means that on average both the number of hops and travel distance are no more than 10 % greater than their corresponding minima.
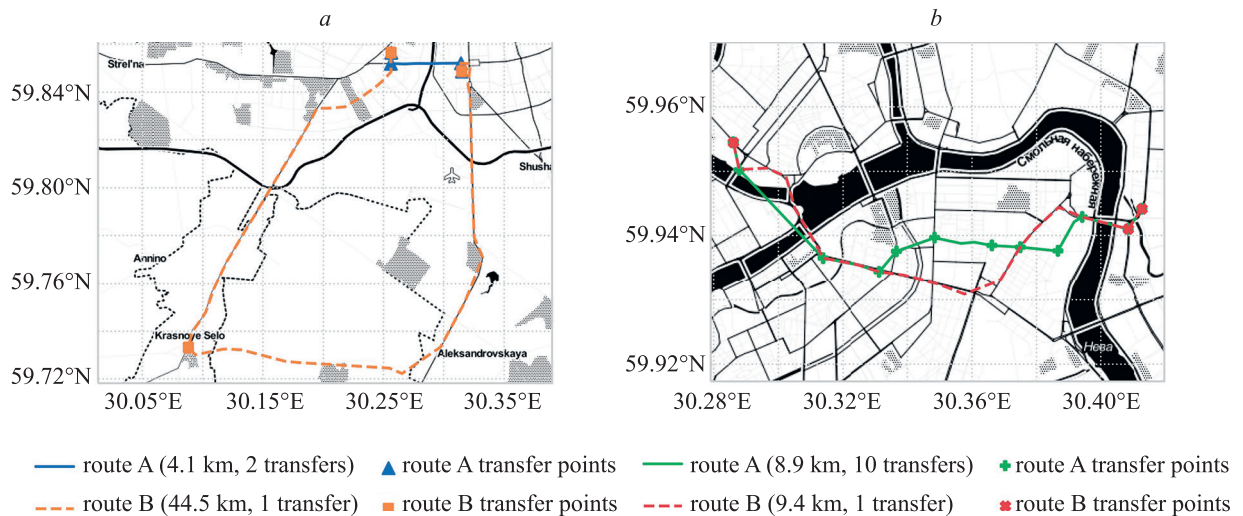
Fig. 7. Minimizing the number of hops can lead to excessively long routes (*a*), while minimizing the travel distance can lead to routes requiring an excessive number of transfers (*b*)[1]

---

[1] The maps are generated by Cartopy, a Python open package. Available at: https://scitools.org.uk/cartopy (accessed 26.09.2022).

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

347

Fig. 8. Mean percentage difference of hops (Eq. 9) and distance (Eq. 10) for different values of α. When α ≈ 0.2, MPD is less than 10 % for both hops and distance

Fig. 9. The map[1] of Vasileostrovsky District in Saint Petersburg indicating supernodes and various infrastructural objects attached to them

_____

[1] The maps are generated by Cartopy, a Python open package. Available at: https://scitools.org.uk/cartopy (accessed 26.09.2022).

Attribute vectors of the node-attributed network. Finally, we want to assign each node $\hat{s} \in \hat{S}$ a multivariate value $A(\hat{s}) \in \mathbb{R}^n$ describing it in terms of social infrastructure surrounding it. This can be done using the information about various infrastructural objects $I = \{i_1, \ldots, i_m\}$ around the city. Each object i_j is a tuple (φ, λ, t), where φ, λ are latitude and longitude of the object, and t is a categorical marker of the type of this object (i.e., be it a shop, a hospital, a sightseeing place, etc.). The set of different infrastructural object types $T = \{t_1, \ldots, t_n\}$ is usually pre-defined.

To construct node attributes, we first assign each infrastructural object to some stop. The most natural way of doing this is to assign each infrastructural object to a stop that is closest to it. We note however that such approach is not the most accurate since there are generally multiple ways of getting to a given destination (for instance, one can take multiple routes to work or school), and these can involve getting off a bus at different stops. To account for this, we propose using a *distance window* d_1 when assigning infrastructural objects to stops. To do so, take an infrastructural object i and suppose that d_{min} is the minimal distance from i to a stop. We then assign the object i to all stops s such that

$$d(i, s) \leq d_{min} + d_1,$$

where $d(a, b)$ is the distance between geographical points eq. (1). In this study we take $d_1 = 0.2$, i.e., the distance of 200 meters (Fig. 8).

Denote $I_S \subseteq I$ as the set of all infrastructural objects assigned to a stop s. When this is done, we construct a multivariate value v_s corresponding to the given stop s by counting the infrastructural objects of different types assigned to this stop, i.e., $v_s \in \mathbb{N}^n$ and

$$(v_s)_j = \#\{i \in I_s | i = (\varphi, \lambda, t), t = t_j\}.$$

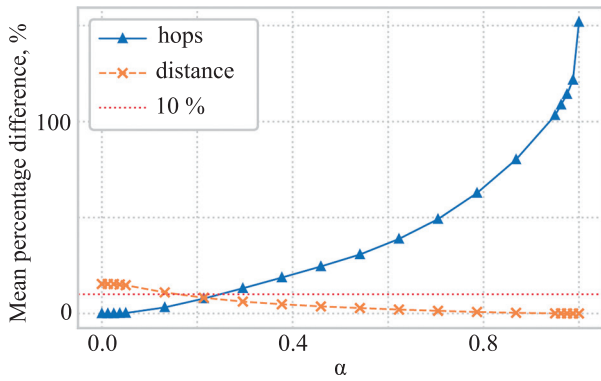These values are used as note attributes in our network model, i.e., $A: \hat{s} \mapsto v_{\hat{s}}$. Such attributes reflect the characteristics of each node in terms of what kind of social infrastructure this node is surrounded by (Fig. 9). The definition of our public transportation model is thus complete.
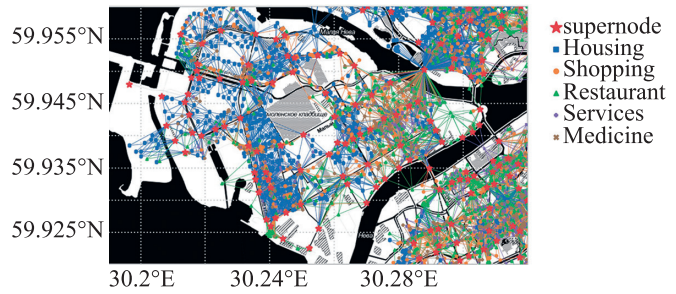
Role discovery task for the node-attributed public transportation network. The task of role discovery originated in the field of social network analysis, but has found its way into a variety of different domains of science. This task usually involves clustering of network nodes, not in a sense of connectivity structure (the so-called *community detection*), but rather in terms of topological features of nodes (for instance, various centrality measures, more on that below). Thus, the goal is to obtain clusters not of densely connected nodes, but rather of nodes having similar structural characteristics.

The basic approach to this task is therefore to extract some features of the network nodes and then use machine learning algorithms (i.e., KMeans [38]) to extract clusters based on these features. Even though originally only topological features were used in this approach, the basic framework can naturally be extended to include also node-attributed vectors (that too can be used as a separate set of node features). One can then combine these two sets of features in some way and perform clustering simultaneously, or alternatively obtain two separate clustering (with respect to topological features and node attributes) and then analyze their relationship, for instance, using a contingency table.

In this theoretical study we adopt the latter approach, i.e., we perform separate clustering with respect to topological features (derived from the network structure) and infrastructure features (using the supernode attributes) and then compare the two.

The reason for this is that these two feature sets have their own interpretations, thus interpreting clusters with respect to only one of the feature sets is much more intuitive than if one uses, for instance, concatenated features.

Conclusions

In this paper, we introduced a novel weighted node-attributed PTN model (using information about a city's social infrastructure to construct the node attributes) and illustrated its construction with the data collected about public transport stops and stations of Saint Petersburg, Russia. Moreover, we pointed out some of the common misconceptions and errors in previous analyses of the PTNs

348

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

which we believe stem from the misunderstanding of some interpretations of different PTN models.

It is also worth mentioning that the most common method of constructing supernodes (i.e., just grouping together all the closely located stops) is not without drawbacks. Additional research should be conducted regarding this problem.

Furthermore, we proposed an approach for solving the novel problem of role discovery in a PTN. The approach uses both structural (i.e., network topology) and semantic (i.e., social infrastructure around the nodes) aspects of a node-attributed PTN. The approach aims at extracting useful information about the properties and overall efficiency of a city's public transportation system from both the structural and infrastructure standpoints. In general, the proposed approach to role discovery in node-attributed networks can be applied beyond the scope of PTNs and to any other kind of network (e.g., social, biological, technical, etc.), given the appropriate set of node attributes.

Recall that a companion study to this one will be the next one where the model and the approach are applied to discover roles in the PTN of Saint Petersburg, Russia.

References

1. Sen P., Dasgupta S., Chatterjee A., Sreeram P.A., Mukherjee G., Manna S.S. Small-world properties of the indian railway network. *Physical Review E*, 2003, vol. 67, no. 3, pp. 036106. https://doi.org/10.1103/physreve.67.036106
2. Sienkiewicz J., Hołyst J. Statistical analysis of 22 public transport networks in Poland. *Physical Review E*, 2005, vol. 72, no. 4, pp. 046127. https://doi.org/10.1103/physreve.72.046127
3. Haznagy A., Fi I., London A., Nemeth T. Complex network analysis of public transportation networks: A comprehensive study. *Proc. of the 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2015, pp. 371–378. https://doi.org/10.1109/mtits.2015.7223282
4. Yang X.-H., Chen G., Chen S.-Y., Wang W.-L., Wang L. Study on some bus transport networks in china with considering spatial characteristics. *Transportation Research Part A: Policy and Practice*, 2014, vol. 69, no. 1, pp. 1–10. https://doi.org/10.1016/j.tra.2014.08.004
5. Zhang J., Zhao M., Liu H., Xu X. Networked characteristics of the urban rail transit networks. *Physica A: Statistical Mechanics and its Applications*, 2013, vol. 392, no. 6, pp. 1538–1546. https://doi.org/10.1016/j.physa.2012.11.036
6. Wang L.-N., Wang K., Shen J.-L. Weighted complex networks in urban public transportation: Modeling and testing. *Physica A: Statistical Mechanics and its Applications*, 2020, vol. 545, pp. 123498. https://doi.org/10.1016/j.physa.2019.123498
7. Shanmukhappa T., Ho I.W.-H., Chi K.T. Spatial analysis of bus transport networks using network theory. *Physica A: Statistical Mechanics and its Applications*, 2018, vol. 502, pp. 295–314. https://doi.org/10.1016/j.physa.2018.02.111
8. Wang Y., Deng Y., Ren F., Zhu R., Wang P., Du T., Du Q. Analysing the spatial configuration of urban bus networks based on the geospatial network analysis method. *Cities*, 2020, vol. 96, pp. 102406. https://doi.org/10.1016/j.cities.2019.102406
9. Lantseva A., Ivanov S. Modeling transport accessibility with open data: Case study of st. Petersburg. *Procedia Computer Science*, 2016, vol. 101, pp. 197–206. https://doi.org/10.1016/j.procs.2016.11.024
10. Bothorel C., Cruz J., Magnani M., Micenková B. Clustering attributed graphs: Models, measures and methods. *Network Science*, 2015, vol. 3, no. 3, pp. 408–444. https://doi.org/10.1017/nws.2015.9
11. Chunaev P. Community detection in node-attributed social networks: A survey. *Computer Science Review*, 2020, vol. 37, pp. 100286. https://doi.org/10.1016/j.cosrev.2020.100286
12. Atzmueller M., Günnemann S., Zimmermann A. Mining communities and their descriptions on attributed graphs: a survey. *Data Mining and Knowledge Discovery*, 2021, vol. 35, no. 3, pp. 661–687. https://doi.org/10.1007/s10618-021-00741-z
13. Rossi R.A., Ahmed N.K. Role discovery in networks. *IEEE Transactions on Knowledge and Data Engineering*, 2015, vol. 27, no. 4, pp. 1112–1131. https://doi.org/10.1109/tkde.2014.2349913
14. Ahmed N.K., Rossi R.A., Willke T.L., Zhou R. Revisiting role discovery in networks: From node to edge roles. *ArXiv*, 2016, arXiv:1610.00844. https://doi.org/10.48550/arXiv.1610.00844
15. Martínez V., Berzal F., Cubero J.-C. An automorphic distance metric and its application to node embedding for role mining. *Complexity*, 2021, vol. 2021, pp. 1–17. https://doi.org/10.1155/2021/5571006
16. Gupte P.V., Ravindran B., Parthasarathy S. Role discovery in graphs using global features: Algorithms, applications and a novel evaluation

Литература

1. Sen P., Dasgupta S., Chatterjee A., Sreeram P.A., Mukherjee G., Manna S.S. Small-world properties of the indian railway network // Physical Review E. 2003. V. 67. N 3. P. 036106. https://doi.org/10.1103/physreve.67.036106
2. Sienkiewicz J., Hołyst J. Statistical analysis of 22 public transport networks in Poland // Physical Review E. 2005. V. 72. N 4. P. 046127. https://doi.org/10.1103/physreve.72.046127
3. Haznagy A., Fi I., London A., Nemeth T. Complex network analysis of public transportation networks: A comprehensive study // Proc. of the 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). 2015. P. 371–378. https://doi.org/10.1109/mtits.2015.7223282
4. Yang X.-H., Chen G., Chen S.-Y., Wang W.-L., Wang L. Study on some bus transport networks in china with considering spatial characteristics // Transportation Research Part A: Policy and Practice. 2014. V. 69. N 1. P. 1–10. https://doi.org/10.1016/j.tra.2014.08.004
5. Zhang J., Zhao M., Liu H., Xu X. Networked characteristics of the urban rail transit networks // Physica A: Statistical Mechanics and its Applications. 2013. V. 392. N 6. P. 1538–1546. https://doi.org/10.1016/j.physa.2012.11.036
6. Wang L.-N., Wang K., Shen J.-L. Weighted complex networks in urban public transportation: Modeling and testing // Physica A: Statistical Mechanics and its Applications. 2020. V. 545. P. 123498. https://doi.org/10.1016/j.physa.2019.123498
7. Shanmukhappa T., Ho I.W.-H., Chi K.T. Spatial analysis of bus transport networks using network theory // Physica A: Statistical Mechanics and its Applications. 2018. V. 502. P. 295–314. https://doi.org/10.1016/j.physa.2018.02.111
8. Wang Y., Deng Y., Ren F., Zhu R., Wang P., Du T., Du Q. Analysing the spatial configuration of urban bus networks based on the geospatial network analysis method // Cities. 2020. V. 96. P. 102406. https://doi.org/10.1016/j.cities.2019.102406
9. Lantseva A., Ivanov S. Modeling transport accessibility with open data: Case study of st. Petersburg // Procedia Computer Science. 2016. V. 101. P. 197–206. https://doi.org/10.1016/j.procs.2016.11.024
10. Bothorel C., Cruz J., Magnani M., Micenková B. Clustering attributed graphs: Models, measures and methods // Network Science. 2015. V. 3. N 3. P. 408–444. https://doi.org/10.1017/nws.2015.9
11. Chunaev P. Community detection in node-attributed social networks: A survey // Computer Science Review. 2020. V. 37. P. 100286. https://doi.org/10.1016/j.cosrev.2020.100286
12. Atzmueller M., Günnemann S., Zimmermann A. Mining communities and their descriptions on attributed graphs: a survey // Data Mining and Knowledge Discovery. 2021. V. 35. N 3. P. 661–687. https://doi.org/10.1007/s10618-021-00741-z
13. Rossi R.A., Ahmed N.K. Role discovery in networks // IEEE Transactions on Knowledge and Data Engineering. 2015. V. 27. N 4. P. 1112–1131. https://doi.org/10.1109/tkde.2014.2349913
14. Ahmed N.K., Rossi R.A., Willke T.L., Zhou R. Revisiting role discovery in networks: From node to edge roles // ArXiv. 2016. arXiv:1610.00844. https://doi.org/10.48550/arXiv.1610.00844
15. Martínez V., Berzal F., Cubero J.-C. An automorphic distance metric and its application to node embedding for role mining // Complexity. 2021. V. 2021. P. 1–17. https://doi.org/10.1155/2021/5571006
16. Gupte P.V., Ravindran B., Parthasarathy S. Role discovery in graphs using global features: Algorithms, applications and a novel evaluation strategy // Proc. of the IEEE 33rd International Conference on Data

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

349

strategy. *Proc. of the IEEE 33rd International Conference on Data Engineering (ICDE)*, 2017, pp. 771–782. https://doi.org/10.1109/icde.2017.128

17. Revelle M., Domeniconi C., Johri A. Persistent roles in online social networks. *Lecture Notes in Computer Science*, 2016, vol. 9852, pp. 47–62. https://doi.org/10.1007/978-3-319-46227-1_4

18. Rossi R.A., Gallagher B., Neville J., Henderson K. Modeling dynamic behavior in large evolving graphs. *Proc. of the Sixth ACM International Conference on Web Search and Data Mining (WSDM'13)*, 2013, pp. 667–676. https://doi.org/10.1145/2433396.2433479

19. Vega D., Meseguer R., Freitag F., Magnani M. Role and position detection in networks: Reloaded. *Proc. of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2015, pp. 320–325. https://doi.org/10.1145/2808797.2809412

20. Yang Z., Algesheimer R., Tessone C.J. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 2016, vol. 6, no. 1, pp. 30750. https://doi.org/10.1038/srep30750

21. Fortunato S. Community detection in graphs. *Physics Reports*, 2010, vol. 486, no. 3-5, pp. 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

22. Souravlas S., Sifaleras A., Tsintogianni M., Katsavounis S. A classification of community detection methods in social networks: a survey. *International Journal of General Systems*, 2021, vol. 50, no. 1, pp. 63–91. https://doi.org/10.1080/03081079.2020.1863394

23. Bartal A., Ravid G. Member behavior in dynamic online communities: Role affiliation frequency model. *IEEE Transactions on Knowledge and Data Engineering*, 2020, vol. 32, no. 9, pp. 1773–1784. https://doi.org/10.1109/tkde.2019.2911067

24. Ni W., Guo H., Liu T., Zeng Q. Automatic role identification for research teams with ranking multi-view machines. *Knowledge and Information Systems*, 2020, vol. 62, no. 12, pp. 4681–4716. https://doi.org/10.1007/s10115-020-01504-w

25. Liu S., Toriumi F., Nishiguchi M., Usui S. Multiple role discovery in complex networks. *Studies in Computational Intelligence*, 2022, vol. 1016, pp. 415–427. https://doi.org/10.1007/978-3-030-93413-2_35

26. Liu Y., Du F., Sun J., Silva T., Jiang Y., Zhu T. Identifying social roles using heterogeneous features in online social networks. *Journal of the Association for Information Science and Technology*, 2019, vol. 70, no. 7, pp. 660–674. https://doi.org/10.1002/asi.24160

27. Zhang L., Lu J., Fu B., Li S. A review and prospect for the complexity and resilience of urban public transit network based on complex network theory. *Complexity*, 2018, vol. 2018, pp. 1–36. https://doi.org/10.1155/2018/2156309

28. Shanmukhappa T., Ho I.W.-H., Tse C.K., Leung K.K. Recent development in public transport network analysis from the complex network perspective. *IEEE Circuits and Systems Magazine*, 2019, vol. 19, no. 4, pp. 39–65. https://doi.org/10.1109/mcas.2019.2945211

29. Xu Q., Mao B., Bai Y. Network structure of subway passenger flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2016, vol. 2016, no. 3, pp. 033404. https://doi.org/10.1088/1742-5468/2016/03/033404

30. Feng J., Li X., Mao B., Xu Q., Bai Y. Weighted complex network analysis of the beijing subway system: Train and passenger flows. *Physica A: Statistical Mechanics and its Applications*, 2017, vol. 474, pp. 213–223. https://doi.org/10.1016/j.physa.2017.01.085

31. Faust K., Wasserman S. Blockmodels: Interpretation and evaluation. *Social Networks*, 1992, vol. 14, no. 1, pp. 5–61. https://doi.org/10.1016/0378-8733(92)90013-w

32. Batagelj V., Mrvar A., Ferligoj A., Doreian P. Generalized blockmodeling with Pajek. *Metodološki zvezki*, 2004, vol. 1, no. 2, pp. 455–467. https://doi.org/10.51936/ofaw1880

33. Luczkovich J., Borgatti S., Johnson J.C., Everett M.G. Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of theoretical biology*, 2003, vol. 220, no. 3, pp. 303–21. https://doi.org/10.1006/jtbi.2003.3147

34. Ma H., Zhou D., Liu C., Lyu M.R., King I. Recommender systems with social regularization. *Proc. of the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11)*, 2011, pp. 287–296. https://doi.org/10.1145/1935826.1935877

35. Golder S.A., Donath J. Social roles in electronic communities. *Internet Research*, 2004, vol. 5.

Engineering (ICDE). 2017. P. 771–782. https://doi.org/10.1109/icde.2017.128

17. Revelle M., Domeniconi C., Johri A. Persistent roles in online social networks // Lecture Notes in Computer Science. 2016. V. 9852. P. 47–62. https://doi.org/10.1007/978-3-319-46227-1_4

18. Rossi R.A., Gallagher B., Neville J., Henderson K. Modeling dynamic behavior in large evolving graphs // Proc. of the Sixth ACM International Conference on Web Search and Data Mining (WSDM'13). 2013. P. 667–676. https://doi.org/10.1145/2433396.2433479

19. Vega D., Meseguer R., Freitag F., Magnani M. Role and position detection in networks: Reloaded // Proc. of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2015. P. 320–325. https://doi.org/10.1145/2808797.2809412

20. Yang Z., Algesheimer R., Tessone C.J. A comparative analysis of community detection algorithms on artificial networks // Scientific Reports. 2016. V. 6. N 1. P. 30750. https://doi.org/10.1038/srep30750

21. Fortunato S. Community detection in graphs // Physics Reports. 2010. V. 486. N 3-5. P. 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

22. Souravlas S., Sifaleras A., Tsintogianni M., Katsavounis S. A classification of community detection methods in social networks: a survey // International Journal of General Systems. 2021. V. 50. N 1. P. 63–91. https://doi.org/10.1080/03081079.2020.1863394

23. Bartal A., Ravid G. Member behavior in dynamic online communities: Role affiliation frequency model // IEEE Transactions on Knowledge and Data Engineering. 2020. V. 32. N 9. P. 1773–1784. https://doi.org/10.1109/tkde.2019.2911067

24. Ni W., Guo H., Liu T., Zeng Q. Automatic role identification for research teams with ranking multi-view machines // Knowledge and Information Systems. 2020. V. 62. N 12. P. 4681–4716. https://doi.org/10.1007/s10115-020-01504-w

25. Liu S., Toriumi F., Nishiguchi M., Usui S. Multiple role discovery in complex networks // Studies in Computational Intelligence. 2022. V. 1016. P. 415–427. https://doi.org/10.1007/978-3-030-93413-2_35

26. Liu Y., Du F., Sun J., Silva T., Jiang Y., Zhu T. Identifying social roles using heterogeneous features in online social networks // Journal of the Association for Information Science and Technology. 2019. V. 70. N 7. P. 660–674. https://doi.org/10.1002/asi.24160

27. Zhang L., Lu J., Fu B., Li S. A review and prospect for the complexity and resilience of urban public transit network based on complex network theory // Complexity. 2018. V. 2018. P. 1–36. https://doi.org/10.1155/2018/2156309

28. Shanmukhappa T., Ho I.W.-H., Tse C.K., Leung K.K. Recent development in public transport network analysis from the complex network perspective // IEEE Circuits and Systems Magazine. 2019. V. 19. N 4. P. 39–65. https://doi.org/10.1109/mcas.2019.2945211

29. Xu Q., Mao B., Bai Y. Network structure of subway passenger flows // Journal of Statistical Mechanics: Theory and Experiment. 2016. V. 2016. N 3. P. 033404. https://doi.org/10.1088/1742-5468/2016/03/033404

30. Feng J., Li X., Mao B., Xu Q., Bai Y. Weighted complex network analysis of the beijing subway system: Train and passenger flows // Physica A: Statistical Mechanics and its Applications. 2017. V. 474. P. 213–223. https://doi.org/10.1016/j.physa.2017.01.085

31. Faust K., Wasserman S. Blockmodels: Interpretation and evaluation // Social Networks. 1992. V. 14. N 1. P. 5–61. https://doi.org/10.1016/0378-8733(92)90013-w

32. Batagelj V., Mrvar A., Ferligoj A., Doreian P. Generalized blockmodeling with Pajek // Metodološki zvezki. 2004. V. 1. N 2. P. 455–467. https://doi.org/10.51936/ofaw1880

33. Luczkovich J., Borgatti S., Johnson J.C., Everett M.G. Defining and measuring trophic role similarity in food webs using regular equivalence // Journal of theoretical biology. 2003. V. 220. N 3. P. 303–21. https://doi.org/10.1006/jtbi.2003.3147

34. Ma H., Zhou D., Liu C., Lyu M.R., King I. Recommender systems with social regularization // Proc. of the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11). 2011. P. 287–296. https://doi.org/10.1145/1935826.1935877

35. Golder S.A., Donath J. Social roles in electronic communities // Internet Research. 2004. V. 5.

36. Yue H., Guan Q., Pan Y., Chen L., Lv J., Yao Y. Detecting clusters over intercity transportation networks using k-shortest paths and hierarchical clustering: a case study of mainland China // International Journal of Geographical Information Science. 2019. V. 33. N 5. P. 1082–1105. https://doi.org/10.1080/13658816.2019.1566551

350

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

36. Yue H., Guan Q., Pan Y., Chen L., Lv J., Yao Y. Detecting clusters over intercity transportation networks using k-shortest paths and hierarchical clustering: a case study of mainland China. *International Journal of Geographical Information Science*, 2019, vol. 33, no. 5, pp. 1082–1105. https://doi.org/10.1080/13658816.2019.1566551

37. Bereznyi D., Qutbuddin A., Her Y., Yang K. Node-attributed spatial graph partitioning. *Proc. of the 28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'20)*, 2020, pp. 58–67. https://doi.org/10.1145/3397536.3422198

38. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. V. 1. Statistics*, 1967, pp. 281–297.

37. Bereznyi D., Qutbuddin A., Her Y., Yang K. Node-attributed spatial graph partitioning // Proc. of the 28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'20). 2020. P. 58–67. https://doi.org/10.1145/3397536.3422198

38. MacQueen J. Some methods for classification and analysis of multivariate observations // Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. V. 1. Statistics. 1967. P. 281–297.

Authors

Yuri V. Lytkin — PhD (Physics & Mathematics), Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57155292900, https://orcid.org/0000-0001-8140-010X, jurasicus@gmail.com

Petr V. Chunaev — PhD (Physics & Mathematics), Senior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 36522457300, https://orcid.org/0000-0001-8169-8436, chunaev@itmo.ru

Timofey A. Gradov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57221121540, https://orcid.org/0000-0003-2537-4087, timagradov@yahoo.com

Anton A. Boytsov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0001-8343-2519, aboytsov@itmo.ru

Irek A. Saitov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57215429754, https://orcid.org/0000-0002-2805-1323, xanilegendx@gmail.com

Авторы

Лыткин Юрий Всеволодович — кандидат физико-математических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57155292900, https://orcid.org/0000-0001-8140-010X, jurasicus@gmail.com

Чунаев Петр Владимирович — кандидат физико-математических наук, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 36522457300, https://orcid.org/0000-0001-8169-8436, chunaev@itmo.ru

Градов Тимофей Алексеевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57221121540, https://orcid.org/0000-0003-2537-4087, timagradov@yahoo.com

Бойцов Антон Алексеевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0001-8343-2519, aboytsov@itmo.ru

Саитов Ирек Аликович — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57215429754, https://orcid.org/0000-0002-2805-1323, xanilegendx@gmail.com

Научно-технический вестник информационных технологий, механики и оптики, 2023, том 23, № 2
Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2023, vol. 23, no 2

351