Networks and Spatial Economics
https://doi.org/10.1007/s11067-019-09476-3

Discovering the Hidden Community Structure of Public Transportation Networks

László Hajdu[1] · András Bóta[2,3] · Miklós Krész[4,5] · Alireza Khani[6] · Lauren M. Gardner[3,7]

Published online: 20 August 2019

Abstract
Advances in public transit modeling and smart card technologies can reveal detailed contact patterns of passengers. A natural way to represent such contact patterns is in the form of networks. In this paper we utilize known contact patterns from a public transit assignment model in a major metropolitan city, and propose the development of two novel network structures, each of which elucidate certain aspects of passenger travel behavior. We first propose the development of a transfer network, which can reveal passenger groups that travel together on a given day. Second, we propose the development of a community network, which is derived from the transfer network, and captures the similarity of travel patterns among passengers. We then explore the application of each of these network structures to identify the most frequently used travel paths, i.e., routes and transfers, in the public transit system, and model epidemic spreading risk among passengers of a public transit network, respectively. In the latter our conclusions reinforce previous observations, that routes crossing or connecting to the city center in the morning and afternoon peak hours are the most "dangerous" during an outbreak.

Keywords Network modeling · Public transportation · Community structure · Infrastructure security

1 Introduction and Background

Networks can be used to represent public transportation systems from various unique perspectives. Traditionally nodes and edges represent the physical infrastructure of a transportation system, e.g., routes and stops/stations (Háznagy et al. 2015), while

✉ András Bóta
andras.bota@umu.se

Extended author information available on the last page of the article.

recent developments in data collection and modeling allow researchers to accurately map contacts between individuals traveling together. Detailed commuting patterns can be recorded from smart card data (Sun et al. 2013, 2014; Bao et al. 2017) or can be the output of activity-based travel models (Brockmann et al. 2006; Song et al. 2010; de Montjoye et al. 2013; Chen et al. 2016; Gardner et al. 2012; Saberi et al. 2018). The resulting passenger contact patterns are both spatial and temporal in nature, and include travel times on specific vehicles and contact with other travelers (Ramadurai and Ukkusuri 2010; Illenberger et al. 2012; Khani et al. 2015). Such detailed travel patterns can then be used for various planning objectives including the estimation of the capacity of infrastructure (Wang et al. 2011), calculating environmental impact (Carlsson-Kanyama and Lindén 1999) or designing surveillance and containment strategies during an epidemic outbreak (Pendyala et al. 2012; Rey et al. 2016). While these methods allow researchers to map contacts between known individuals, the data collection and processing required to recreate a real-world contact network presents many challenges in terms of accuracy and computational complexity, among other issues such as privacy (Huerta and Tsimring 2002; Hoogendoorn and Bovy 2005; Balcan 2009; Salathé 2010; Funk et al. 2010; Nassir et al. 2012).

In this work we specifically focus on the community structure of public transit ridership patterns. Observations on social interactions reveal that people tend to form groups according to their lines of interest, occupation, etc. This concept is known as homophily in the social sciences (Eagle et al. 2009; Yuan and Gay 2006; Chin et al. 2012), while in network modeling we refer to this phenomenon as community structure (Fortunato 2010). One of the most frequently used definitions of this concept was proposed by Newman and Girvan (2004): "In an arbitrary network a *community* is a set of nodes where the density of connections between the nodes of the community is greater than the density of connections between communities". Community detection is a diverse field with many applications in various fields of science. Newman's definition has been extended or replaced by newly proposed algorithms, yet it is still the most intuitive way of describing communities.

The majority of community detection algorithms consider communities as disjoint sets of nodes. Popular approaches include modularity maximization methods (Newman and Girvan 2004; Blondel et al. 2008), information theoretic approaches (Rosvall and Bergstrom 2008), statistical inference (Peixoto 2014; Aicher et al. 2014) and spectral techniques (Krzakala et al. 2013; Newman 2013). Other algorithms allow overlaps between neighboring groups (Bóta and Krész 2015; Lancichinetti et al. 2009; Wu et al. 2012). Community detection algorithms can also be applied to weighted networks, where a value on each edge represents similarity (or distance) between nodes (Bóta and Krész 2015; Aicher et al. 2014). All weighted networks can be transformed into unweighted forms by setting a threshold and omitting edges with weights below the threshold. This technique has been used in real-world applications (Palla et al. 2005; Bóta and Kovács 2014). A selection of excellent reviews on community detection can be found in Fortunato (2010), Xie et al. (2013), and Leskovec et al. (2010).

In this work we propose the development of two novel network structures, namely the *transfer network*, and the *community network*. The transfer network captures the movement patterns of atomic passenger groups, i.e., groups of people who travel

together on one or more vehicles, mathematically defined as maximal complete subgraphs of the contact network. The community network captures the similarity of travel patterns between passengers, is derived directly from the transfer network, and is constructed using a novel link-based metric called the *connection strength*. The community network is intended to reveal groups of travelers who may also be more likely to come into contact outside of their daily travel routine e.g.: colleagues traveling to work or children going to school (Yuan and Gay 2006; Chin et al. 2012). A description of each proposed network structure as well as the contact network can be found in Table 1.

We further demonstrate potential applications of each of these novel network structures. First, the transfer network is implemented to detect the most frequent vehicle trip combinations in the system (bus and train transfers). While there are existing methods in the literature to measure the capacity of public transit systems (Manuel et al. 2006; Wen-Tai and Ching-Fu 2011), our approach provides a more refined view of passenger movement between vehicle trips by tracking the movements of groups of passengers traveling together. Identifying the most relevant vehicle trip combinations can aid public transit authorities in timetable planning and optimizing vehicle assignments.

Second, the community network is applied to evaluate a diffusion process in a transit network, specifically infectious disease spread among passengers. This application complements our previous work (Bóta et al. 2017a), to identify the components of the public transportation system most vulnerable to a bio-security threat. The community network proposed here can further improve the performance of such models due to its novel representation of passenger interactions. We use the proposed network structure to simulate how a disease might spread among the vehicles of the public transportation system and the suburbs of the city, and identify the vehicle trips most likely to carry infected passengers. All applications are evaluated on a real-world case study, the public transit system of Twin Cities, MN, using output from an activity based travel demand model (Khani et al. 2015).

Table 1 Network definitions used in this paper, including the contact network and the proposed transfer and community networks

	Contact network	Transfer network	Community network
Nodes	Passengers	Atomic passenger groups	Passengers traveling on at least two vehicle trips
Edges	Passengers are connected of they are physically present on the same vehicle trip at the same time (undirected)	Atomic passenger groups are connected if they share a passenger (directed)	Passengers are connected if they are traveling together on at least two vehicle trips (undirected)
Attributes	Contact duration and start time	Number of transfer passengers between atomic groups	Connection strength measuring the similarity of the travel patterns between passengers

The rest of the paper is structured as follows. In Section 2 we introduce the travel demand model, and how it is used to create passenger contact network. In Section 3 we introduce the transfer network, and illustrate how it can be used to detect frequent vehicle trip combinations. In Section 4 we introduce the community network, including the connection strength value, and illustrate how it can be used to model infectious disease spread. Section 5 contains our conclusions and future research directions.

2 Model Inputs

The inputs of our work are based on the transit assignment model published in Khani et al. (2015). In this section, we present a description of the assignment model and provide a short summary of the properties of the passenger contact network built from it.

2.1 Travel Demand Model

As described in our previous work (Bóta et al. 2017a), public transportation data in this study was obtained from the transit system in Twin Cities region in Minnesota, where 187 routes serve 13,700 stops in the region. Transit network and schedule data were created from General Transit Feed Specification (GTFS), including near 0.5 Million stop-times on a weekday in 2015. Transit passenger trips were obtained from Metropolitan Councils's activity-based demand model (Cambridge Systematics Inc. 2015), and contained more than 293,000 linked trips (i.e. a passenger trip from an origin to a destination that may include zero or more transfers). The assignment of transit demand to transit network was done using the FAST-TrIPs model (Khani et al. 2015). FAST-TrIPs is a schedule-based transit assignment model that generates hyperpaths using a defined logit route choice model (Khani 2013), assigns individual passengers to the paths using the hyperpath probabilities, and simulates them using a mesoscopic transit passenger simulation module. Since a calibrated transit route choice model was not available for the Twin Cities network, a route choice model from Austin, TX was borrowed from a previous study by the authors (Khani et al. 2014). The model specifies the following route choice utility function:

$$u = t_{IV} + 1.77t_{WT} + 3.93t_{WK} + 47.73X_{TR}$$

where t, t_{IV}, t_{WT}, t_{WK} and X_{TR} represent path utility, in-vehicle time, waiting time, walking time, and number of transfers in a transit path, respectively. The output of the transit assignment model contains individual passengers' trajectories, including their walking from the origin to the transit stop, boarding the transit vehicle, alighting and walking to the transfer stop, boarding the next transit vehicle, etc. and finally alighting and walking to the destination. By post-processing the transit assignment model's outputs, the amount of time each pair of passengers are on-board the same transit vehicle were calculated on a daily basis.

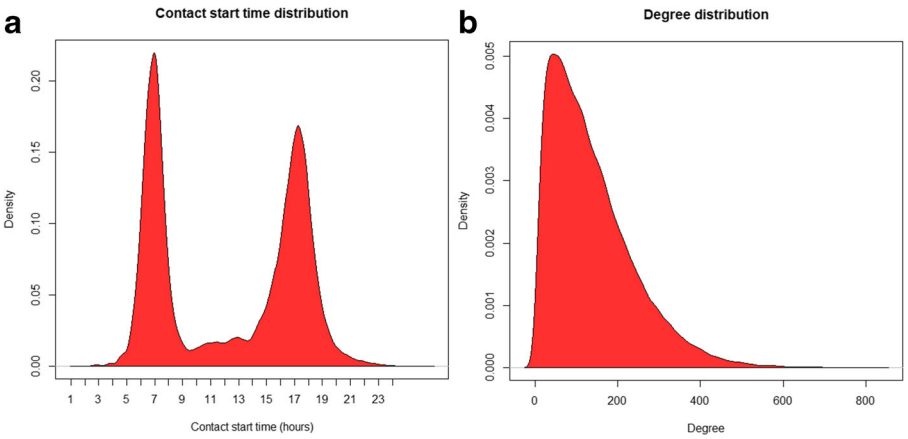

Fig. 1 The distribution of **a** contact start times and **b** the degree distribution of the contact network

2.2 Contact Network

We define a network structure denoted as the *contact network* based on the outputs of the travel demand model. The nodes of the contact network are passengers, and edges connect passengers if they were traveling on the same vehicle at the same time. The relationship is symmetric, therefore the network is undirected. All passenger movements take place in a temporal setting, which is indicated by two values assigned to the edges of the network: the contact start time on an edge indicates the start time of the contact between the two connected passengers, while a contact duration value indicates the length of the contact in minutes. Since each edge represents a connection between a pair of passengers traveling on a specific vehicle, the id of the vehicle can also be assigned to the edges of the graph. The vehicle id identifies a vehicle trip: a single vehicle traveling on a specific route with a specific start time. The basic properties of the network were shown in Bóta et al. (2017a) along with a detailed analysis regarding the networks potential role in the spreading of epidemics. A short review on the description of the network as appeared in Bóta et al. (2017a) follows.

The contact network corresponding to the Twin cities dataset has 94475 nodes and 6287847 contacts between them. Figure 1a shows the density plot of the contact start times. The distribution has two peaks, one around 7 AM and another around 5 PM, which corresponds to the morning and evening weekday commute. The average number of contacts per person is 136 during the observed day, while the maximum is 827. Figure 1b shows the degree distribution of the graph.

Figure 2 shows a subgraph of the network structure. Nodes represent passengers and edges connect them if they ride on the same vehicle trip. The dynamics of the network are omitted in this example, i.e. edges are aggregated over the entire day. The black node in the middle represents a passenger with a high number of contacts, who traveled together with all the other nodes shown on this subgraph. The other nodes are colored according to the vehicle trip they rode on, while darker edges indicate longer contact durations.

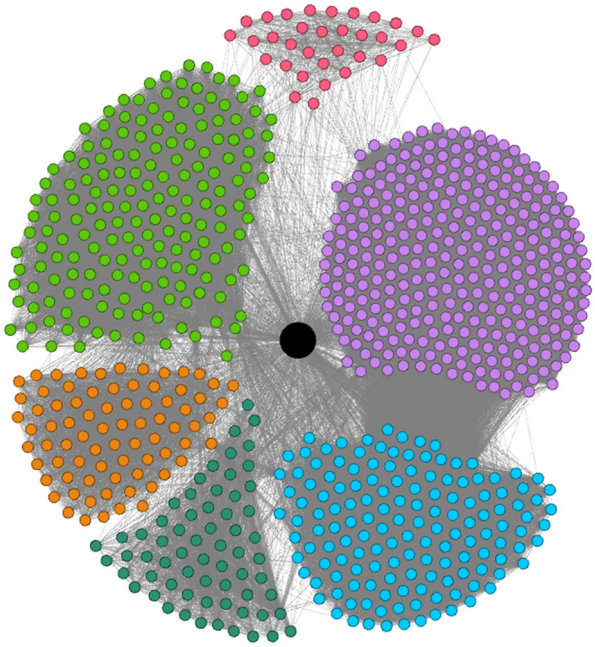

Fig. 2 A static subgraph of the contact network. A passenger with a high number contacts is represented by the black node in the middle, connected to all other contacts. The colors indicate of the vehicle trips the passengers first met on. The figure was made with Gephi using the Fruchterman-Reingold algorithm

3 Transfer Network

The first objective of this paper is to detect the movement patterns of passenger groups.

To do this, we define a novel network structure denoted as the *transfer network*. The transfer network is a directed network, where nodes represent the atomic passenger groups of the contact network and groups are connected if at least one member of a group transfers from one vehicle to another. The weight of the edges denote the number of transfer passengers.

In order to build the transfer network we identify the subgraph corresponding to each vehicle trip, detect atomic passenger groups – defined as maximal cliques – on each of the resulting subgraphs of the contact network and connect the atomic passengers groups according to direction of transfer between vehicle trips. We weight the connections based on the number of transferring passengers. Section 3.1 defines the construction of the transfer network in more detail.

The transfer network proposed in this paper provides a refined way to identify the most frequent vehicle trip combinations in the public transit system, which can aid decision makers in defining timetables and optimizing vehicle assignments. Section 3.2 discusses this application.

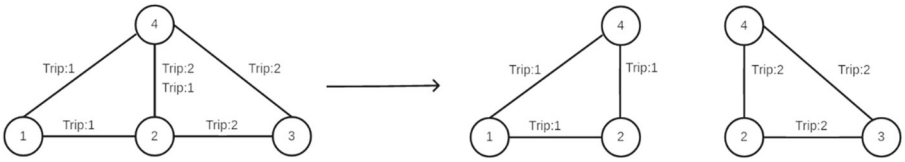

Fig. 3 Partitioning of an example graph along vehicle trips. Each vehicle trip has a corresponding subgraph where nodes are passengers who used the given vehicle trip

3.1 Transfer Network Construction

The three steps required to build the transfer network are discussed in the subsequent subsections. Section 3.1.1 defines how the subgraphs corresponding to the vehicle trips are constructed, Section 3.1.2 defines atomic passenger groups as cliques and show how they can be detected in an efficient way and Section 3.1.3 shows how the transfer network can be built from the passenger groups.

3.1.1 Graph Partitioning

We define the subgraph corresponding to each vehicle trip as follows. Let G be the original contact network, $V(G)$ the vertices and $E(G)$ the edges of the network. We divide the original network into subgraphs along vehicle trips. Let T be the set of all vehicle trips in the network and let $t_i \in T$ denote the i-th trip. We define G_{t_i} as the subgraph corresponding to trip t_i where, $V(G_{t_i})$ and $E(G_{t_i})$ are the vertices and the edges of trip G_{t_i}. We create a subgraph G_{t_i} for all trips $t_i \in T$. Since a single passenger may travel on multiple trips, the node corresponding to the passenger may appear in multiple subgraphs. Figure 3 shows an example of the partitioning process.

3.1.2 Clique Detection

In graph theory, a *clique* is defined as a fully connected subgraph of a given graph. A *maximal clique* is a clique, that is not a subgraph of any other clique. Finding the set of all maximal cliques is a well-studied NP-hard problem in graph theory. An arbitrary n-vertex graph may have up to $3^{n/3}$ maximal cliques, but this number is much lower in many complex networks, including the contact network studied in this paper. Among the available methods in clique detection we adopt the Bron-Kerbosch (BK) algorithm (Bron and Kerbosch 1973), which has proven its reliability in real-life applications (Eppstein et al. 2010). We apply the BK algorithm[1] to detect the maximal cliques of all subgraphs G_{t_i} corresponding to all vehicle trips $t_i \in T$.

The analysis can be further extended if we consider the graphs as weighted according to the contact durations available on the edges. Introducing a weight threshold τ we can prune the edges of the graphs by omitting all edges with contact durations

[1]We implemented the Bron-Kerbosch algorithm with pivoting and degeneracy ordering in the outer most level of recursion in C++. We ran the algorithm on a PC with I7 4790 CPU (3.6 Ghz) and 16GB of RAM.

Table 2 Runtimes of the Bron-Kerbosch on graphs G_0, G_5, G_{15}, G_{30}. Raw denotes runtime in seconds on the original contact network, while partitioned denotes the total runtime on all subgraphs corresponding to vehicle trips

Graph	Passengers	Edges	Trips	Time raw (s)	Time partitioned (s)
G_0	94475	6435482	8002	5285	367
G_5	91894	5203557	7934	3830	258
G_{15}	63714	1630522	6969	1195	50
G_{30}	26154	218233	4083	129	6

below τ. This allows us to redefine what a connection means in the network: passengers are only considered to be connected if they spend at least a certain amount of time on the same vehicle. We define three thresholds to represent the strength of connection between passengers $\tau_5 = 5$ minutes, $\tau_{15} = 15$ minutes and $\tau_{30} = 30$ minutes. We chose these thresholds to represent short, medium and long duration contacts between passengers. We prune the edges of the contact network by these values resulting in graphs G_5, G_{15} and G_{30}. In additional, let G_0 and τ_0 represent the original (uncut) graphs and the corresponding threshold. We run the Bron-Kerbosch clique detection algorithm on these graphs and analyze and compare results. The speed of the detection algorithm is amplified by the fact, that all subgraphs corresponding to vehicle trips are interval graphs. In order to show this speedup, we also run the clique detection algorithm on the original contact network and compare results. Table 2 shows graph size, runtime of the clique detection method on the original contact network and the total runtime of the detection method on all subgraphs for G_0, G_5, G_{15} and G_{30}. We can see a significant speedup for all graphs in this analysis. For the larger G_0 and G_5 graphs the total runtime on all subgraphs corresponding to vehicle trips is 14 times less than on the unpartitioned network.

3.1.3 Graph Building

Let F denote the *transfer network* as a directed graph where every node $v \in V(F)$ is a clique from G_{t_i} for all t_i. Edges connect nodes v and u if the corresponding cliques do not lie in the same vehicle trip, yet they have at least one common passenger. More formally, for nodes u and v in the transfer network and their corresponding cliques c_v and c_u in the contact network, u and v is connected if $c_v \cap c_u \geq 1$ and if $c_v \subseteq G_{t_v}, c_u \subseteq G_{t_u}$ then $G_{t_v} \neq G_{t_u}$.

The *direction of edges* correspond to the direction of the transfer between t_u and t_v, that is whether the passengers of $c_v \cap c_u$ move from t_u to t_v or the opposite. We establish the direction by looking at the contact start times for the individuals in $c_v \cap c_u$ in the following way. For all edges of the transfer network $e_{uv} \in E(F)$, let c_{uv} denote the set of corresponding passengers in G as $c_{uv} = c_u \cap c_v, c_v \subseteq G_{t_v}, c_u \subseteq G_{t_u}, G_{t_v} \neq G_{t_u}$. Let $\alpha_{xy}^{t_i}$ denote the contact start time between passengers x and y on vehicle trip t_i. For all passengers $p, q \in c_{uv}$, if $min(\alpha_{pq}^{t_u}) < min(\alpha_{pq}^{t_v})$, then the direction of the edge $e_{uv} \in E(F)$ is from u to v, else it is from v to u. If $|c_v \cap c_u| = 1$, then let $c_v \cap c_u = \{p_0\}$ and $p \in c_u \setminus \{p_0\}, q \in c_v \setminus \{p_0\}$. Then

$min_p(\alpha_{p_0p}^{t_u}) < min_q(\alpha_{p_0q}^{t_v})$ decides the direction of the edge as above. We assign integer values $w_v = |c_v|$ to all $v \in V(F)$ and $w_{u,v} = |c_v \cap c_u|$ for all $e(u, v) \in E(F)$. Values w_v and $w_{u,v}$ represent the amount of passengers corresponding to both the nodes and edges of the transfer network.

3.2 Detecting Frequent Vehicle Trip Combinations

The *transfer network* allows us to identify the most frequent vehicle trip combinations passengers travel on in the public transportation system. Detecting the most frequent combinations helps decision makers in defining timetables and optimizing vehicle assignments.

As before, let T be the set of vehicle trips, and $t_i \in T$ the i-th vehicle trip. We define vehicle trip pairs as follows: for all t_i and t_j, t_{ij} is a vehicle trip pair where $i \neq j$ and $t_i, t_j \in T$, while the set of all vehicle trip pairs is denoted by T^p. We assign a number m_{ij} to each vehicle trip pair $t_{ij} \in T^p$ indicating the amount of passenger traffic between the corresponding trips. To calculate this value let the set $V(C_{t_{ij}})$ contain all nodes of the transfer network corresponding to all cliques $c_{t_{ij}} \in C_{t_{ij}}$ where $c_{t_{ij}} \in G_{t_i} \cup G_{t_j}$, and take the subgraph induced by $V(C_{t_{ij}})$. The number m_{ij} is the sum of all edge weights of the induced subgraph. This approach offers a more refined view of passenger traffic between vehicle trips because it represents the movements of passenger groups as opposed the behavior of individual passengers.

To give another dimension to our analysis we compared the frequent vehicle trip combinations in both G_0, G_5, G_{15} and G_{30}, that is only considering passengers to be in contact with each other if they travel together for more than $\tau_5 = 5$ minutes, $\tau_{15} = 15$ minutes and $\tau_{30} = 30$ minutes in addition to the unfiltered network G_0. Table 3 shows the five most frequent vehicle trip combinations t_{ij} and the amount m_{ij} of passenger traffic between them.

Almost all trip combinations in Table 3 follow a similar pattern. Results for G_0 and G_5 are nearly identical. The most frequent combinations for G_{15} connect two additional suburbs (Oakdale and Stillwater) to G_0 and G_5, but otherwise are identical.

Table 3 The five most frequent vehicle trip combinations in G_0, G_5, G_{15} and G_{30}. The first number indicates the route number followed by the start time of the sepcific vehicle trip

Graph	t_i	t_j	m_{ij}	Graph	t_i	t_j	m_{ij}
G_0 1.	614/5:25	675/5:30	38	G_{15} 1.	68/6:19	94/7:04	21
2.	68/5:46	94/6:25	35	2.	94/18:32	68/18:52	19
3.	68/6:19	94/7:04	27	3.	68/5:46	94/6:24	18
4.	901/6:18	415/7:02	25	4.	614/5:16	675/5:35	16
5.	415/17:12	901/17:33	23	5.	294/5:38	94/6:44	15
G_5 1.	614/5:25	675/5:30	37	G_{30} 1.	19/17:22	850/18:03	8
2.	68/5:46	94/6:25	35	2.	54/18:02	68/18:23	4
3.	68/6:19	94/7:04	27	3.	71/6:07	61/6:38	4
4.	901/6:18	415/7:02	25	4.	5/0:30	901/2:07	4
5.	415/17:12	901/17:33	23	5.	902/19:25	68/19:52	4

The vehicle trips pairs with the greatest amount of passenger traffic between them are mostly in the morning or afternoon peak hours. Almost all of the trip combinations link one of the outlying suburbs to the city center, indicating the daily commuting patterns of workers or students. Just the urban area of Twin Cities covers 2646 km^2 with a population of more than three million. This means that commuters trying to reach the city center from one of the outlying suburbs must travel on two or sometimes three separate routes to get to their destination. The pattern – also shown on Fig. 4 – is the following. Commuters start the journey from one of the *smaller outlying suburbs* like Ridgedale (route 614), Inver Grove and other suburbs south of St. Paul (route 68) or Oakdale and Stillwater (route 294). Then they change vehicles in the transport hub of one of the *major outlying suburbs* (St. Paul, Minnetonka, Mall of America in Bloomington), and take an express service to the *city center* (routes 94, 675, 850, etc..).

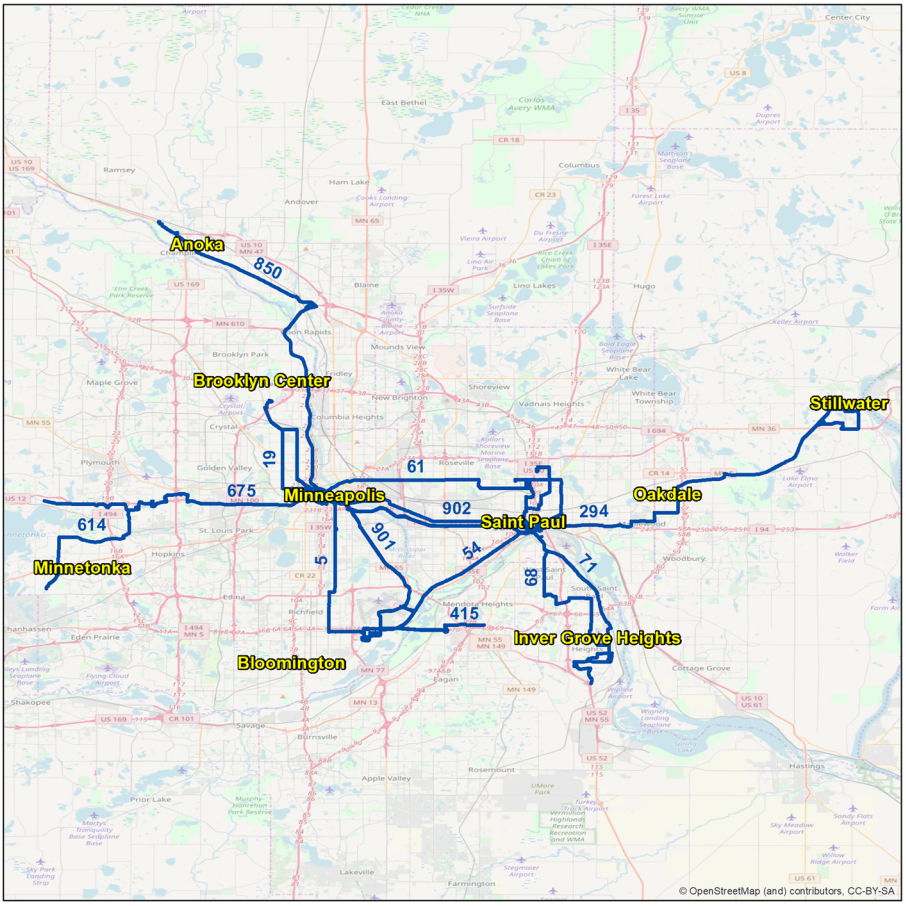

Fig. 4 Most frequent vehicle trips combinations in Twin Cities, MN for G_0, G_5, G_{15} and G_{30}

While the frequent combinations for G_{30} are similar, there are differences. While the suburbs these routes connect are different, we can see almost the same patterns as before: travel from one of the smaller outlying suburb to a major one and then to the city centre (combinations 71–61 and 5–901). One specific route (850) is one of the longest express bus route in the city and it includes a long section where the bus doesn't stop at all. One exception to the pattern is route 54, which connects St. Paul International airport to St. Paul and route 68 connecting to south St. Paul. In terms of time in addition to peak hours, we see late night services as well.

We can summarize, that graphs G_0, G_5 and G_{15} behave similarly, showing the movements of passenger groups traveling through the public transit system. The most frequent trip combinations are the ones connecting distant suburbs to the city center during the morning and afternoon peak hours. G_{30} captures an alternative set of routes corresponding to people who travel together for longer periods for different reasons, like a long service (route 850), the scarcity of services late at night (route 5 at 0:30) or traveling from the airport to a major transport hub (route 54).

4 Community Network

Next we propose a novel network structure, the *community network*, which expands the definition of passenger connectivity to be a function of both the number of transfers passengers make together as well as the total amount of time they spend together while traveling. This contrasts the contact network, which simply quantifies passenger connectivity passengers using contact duration on individual vehicle trips. We define a *community of passengers* as a set of passengers who have common travel patterns, e.g., vehicle trips and/or transfers. In order to build communities, we define and quantify a novel *connection strength* metric between passengers indicating the similarity of their travel patterns and create a new network structure, the *community network* based on this value. Section 4.1 outlines this process in more detail.

This network can serve as the basis of a community detection algorithm, but in this paper we take a different approach. We define communities as those connected by edges whose values lie above a predetermined threshold. In Section 4.2 we demonstrate this feature by identifying the commuting patterns of the members of the largest passenger community of the network.

We propose an application of the community network in Section 4.3. We show how the connection strength value can be used to model infectious disease transmission among passengers of a public transit system. Tying to our previous work in Bóta et al. (2017a) we seek to identify the vehicle trips most likely to carry infected passengers during an outbreak.

4.1 Community Network Construction

In order to detect the communities of the public transportation network we construct a weighted network structure called *community network*. The community network connects passengers using on a novel link-based metric, the *connection strength*. The connection strength s defines the edge weights in the community network, and takes

into account the number of transfers a pair of passengers makes together. Thus, if the passengers meet on multiple different vehicle trips, their connection strength s will increase. This method is based on the assumption that passengers who not only travel together but also transfer together have a stronger connection than travelers who are simply present on the same vehicle trip at the same time. Below we explain how this link metric is derived.

Let H denote the new network where the nodes are the passengers, and the set of nodes V(H) includes all passengers who traveled on at least two vehicle trips with another passenger. Thus, the nodes in the community network correspond to the edges of the transfer network. We define the connections and weights between passengers as follows. Nodes u and v in the community network are connected if they are both present in at least two different cliques c_1, c_2 in two different subgraphs corresponding to vehicle trips, *i.e.* they traveled together on at least two vehicle-trips ($u, v \in c_1 \subseteq G_{t_1}$ and $u, v \in c_2 \subseteq G_{t_2}$). Let g_{uv} denote the number of instances where u and v are members of the same clique, that is $g_{uv} = |C_{uv}|$ where $c_{uv} \in C_{uv}$ if $u, v \in c_{uv}$. Let T_{uv} be the set of *vehicle trips* where both u and v are present: $t_{uv} \in T_{uv}$ if $u, v \in t_{uv}$, and let $g_{uv}^{t_i}$ be the number of the cliques in vehicle trip t_i where u and v are both present: $g_{uv}^{t_i} = |C_{uv}^{t_i}|$ where $c_{uv}^{t_i} \in C_{uv}^{t_i}$ if $u, v \in c_{uv}^{t_i} \in G_{t_i}$. Thus, the connection strength s_{uv} between passengers u and v can be formalized as follows:

$$s_{uv} = \frac{g_{uv} * (g_{uv} - 1)}{2} - \sum_{t_i \ in \ T_{uv}} \frac{g_{uv}^{t_i} * (g_{uv}^{t_i} - 1)}{2} \tag{1}$$

Using this definition of connection strength, if a pair of passengers traveled together in g_{uv} different atomic passenger groups, then they would have $\frac{g_{uv}*(g_{uv}-1)}{2}$ different edges between them because the affected nodes form a clique in the transfer network. This way the first part of the equation rewards the movements between different vehicle trips. Since based on our community definition traveling on the same vehicle trip doesn't indicate strong connection between the passengers, in the second part of the equation we penalize any instance where u and v travels together on the same vehicle trip. The value of the penalty will be the sum of the edges in every vehicle trip where the passenger pair appears more than one time, and the number of the edges in a vehicle trip is counted in the same way as in the first part of the equation.

Algorithm 1 Construction of the community network.

```
V(H) ← Every passengers from E(F) sections
E(H) ← ∅
For in E(F)sections Do
    For u v in every passenger pairs Do
        If e(u, v) ∉ E(H)
            E(H) ← e(u, v)
        Else
            s_uv ++
        End if
    End for
End for
```

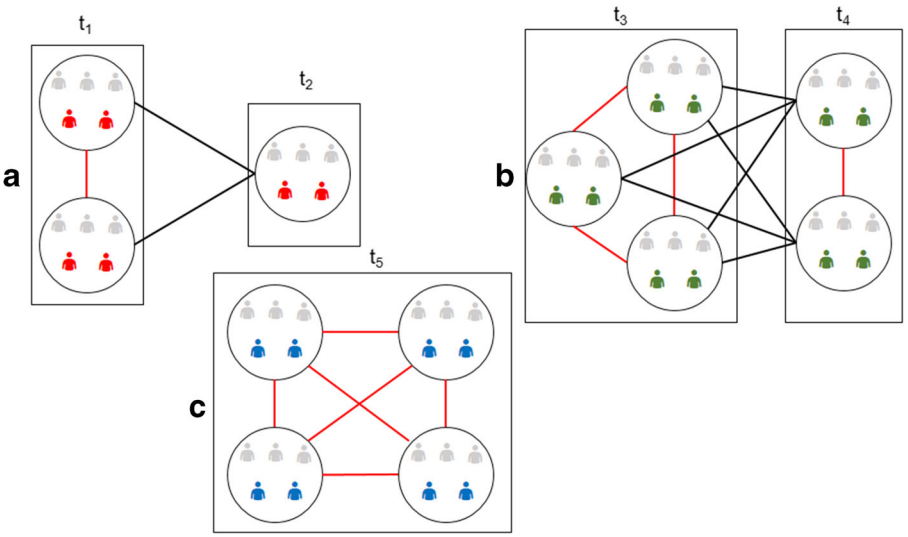

Fig. 5 Examples of the connection strength between pairs of passengers in three different travel scenarios.. Rectangles represent vehicle trips and circles represent cliques. Edges marked with black increase the connection strength between highlighted passengers. Red edges penalize connection strength because these are on the same vehicle trip. **a** two passengers travel together in two cliques on vehicle trip t_1 and one clique on vehicle trip t_2, therefore $g_{uv}^{t_1} = 2,\ g_{uv}^{t_2} = 1,\ g_{uv} = 3$ and $s = 2$. **b** two passengers travel together in three cliques on t_3 and two cliques on t_4 making $g_{uv}^{t_3} = 3,\ g_{uv}^{t_4} = 2,\ g_{uv} = 5$ and $s = 6$. **c** two passengers travel together on a single vehicle trip in four cliques making $g_{uv}^{t_1} = 4,\ g_{uv} = 4$ and $s = 0$

Algorithm 1 shows the construction of the *community network*, while Fig. 5 illustrates a few examples for computing s between passengers. On Fig. 5a two passengers travel together in two cliques on vehicle trip t_1 and one clique on vehicle trip t_2, therefore $g_{uv}^{t_1} = 2,\ g_{uv}^{t_2} = 1,\ g_{uv} = 3$ and $s = 2$. A different situation is shown on Fig. 5b where two passengers travel together in three cliques on t_3 and two cliques on t_4 making $g_{uv}^{t_3} = 3,\ g_{uv}^{t_4} = 2,\ g_{uv} = 5$ and $s = 6$. Figure 5c presents a trivial case, when two passengers travel together on a single vehicle trip in four cliques making $g_{uv}^{t_1} = 4,\ g_{uv} = 4$ and $s = 0$.

4.2 Passenger communities

In this section we illustrate examples of passenger communities in the public transportation system of Twin Cities MN. The first example seen on Fig. 6 shows subgraphs of the community network constructed from G_{30}, *i.e.* the contact network only containing edges where contact duration is above 30 minutes. Figure 6a depicts the entire community network. Most of the communities on this network are of size two or three, but there are several larger communities with strong connections between the members. Figure 6b shows a subgraph where edges with weights $s < 5$ are omitted as well as all nodes with degrees below two. The remaining subgraph contains the largest group of the network, while the largest individual community is depicted on Fig. 6c.

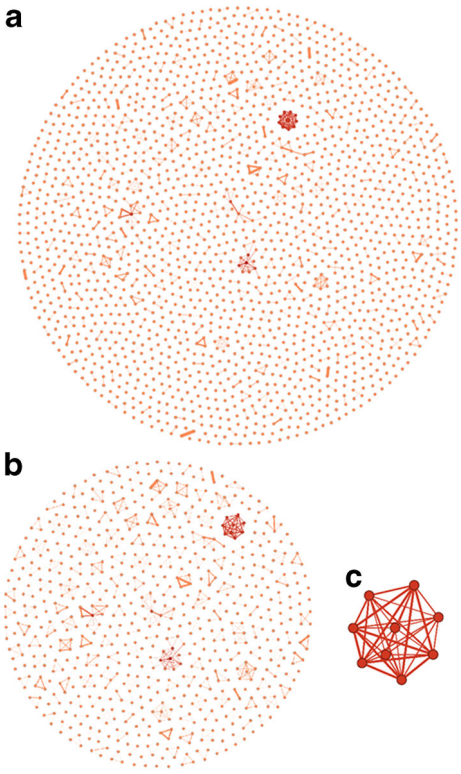

Fig. 6 The community network of Twin Cities, MN. **a** the whole community network, **b** a subgraph with edge weights greater than 5, **c** the largest passenger group of the network

Figure 6c shows the largest group in the network. The group contains nine passengers who traveled together on two different vehicle trips, while the overall time they spent using the public transportation network was almost 1.5 hours. The passengers embarked on route 805 in the morning between 7:08 and 7:16 near Blaine and traveled together to Northtown. They disembarked at 7:48 and waited together for the second vehicle 852 arriving at 8:12 and traveled together to downtown Minneapolis for almost an hour until 9:12 and 9:16. The travel path of the community, shown on Fig. 7, indicates a commuting pattern from one of the suburbs to the city center of Minneapolis.

Both contact and community networks reveal contact patterns between passengers of a public transportation system. In the contact network, the weight between individual passengers is defined based on the contact duration on individual vehicle trips. In contrast, the community network provides a more refined way to represent connection strength which takes into account the amount of transfers passengers take together.

The underlying concept behind community structure in networks is homophily, which is a well-studied concept of the social sciences (Eagle et al. 2009; Yuan and Gay 2006; Chin et al. 2012). Homophily states that people tend to form groups

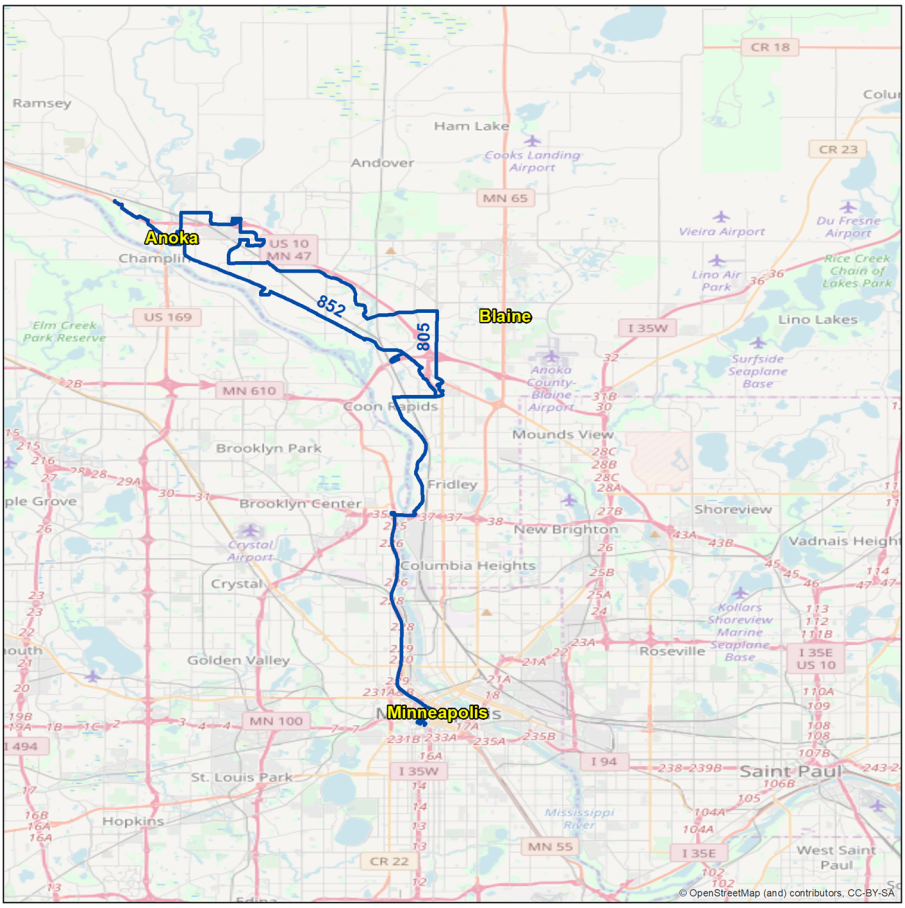

Fig. 7 The travel path of the passenger community on Fig. 6c traveling from a suburb to the city center

according to their lines of interest, occupation, etc. Physical proximity – on public transportation for example – is one of these indicators. Therefore, strong connections on the community network can help us uncover connections in other areas of life like workplace or school or other common interests. It should be noted, that physical proximity does not guarantee another type of connection, it simply increases the likelihood of occurrence for it.

4.3 Epidemic Spreading Risk Application

One application of identifying the communities within the transit network, as described in this work, is infrastructure security. Understanding passenger communities enables more efficient and accurate tracking of infectious disease spread, were one to be naturally or maliciously introduced into the public transit system. One of the challenges in modeling epidemic spreading is accurately mapping the relationships

between individuals traveling on the same vehicle. A traditional contact network as defined in Section 2 as well as in Bóta et al. (2017a, b) and Sun et al. (2013) simply revelas the set of passengers who were present on the same vehicle and the amount of time they spend on the same vehicle. As shown in Bóta et al. (2017b), this limits the options for how infection spreading probabilities can be defined to be simply a function of contact duration, without the possibility to take into account physical proximity, communication etc. between people.

In contrast, the weights of the community network reveals a deeper level of connections between travelers, as passengers with strong connections in this network may also be connected in other areas of life. Passenger groups identified in the community network are more likely to be traveling within close physical proximity of each other and interacting with each other (Eagle et al. 2009; Yuan and Gay 2006; Chin et al. 2012). As a consequence, the probablity of disease transmission between the travelers belonging to the same community is greater than between two passengers simply sharing the same vehicle without any other connection. This is especially true for large public transportation vehicles like trains or trams, where simply being present on the same vehicle may not imply any kind of connection at all.

In this section we expand upon our previous work in Bóta et al. (2017a) which examined epidemic spreading risk in the same public transportation network (Twin Cities, MN) with the goal of identifying the vehicle trips most likely to carry infected passengers. The analysis was presented in two parts. First, the passenger contact network, in the same form as in Section 2, was used to model a variety of outbreak scenarios. The scenarios differed in the number and distribution of initially infected passengers and the level of infectiousness, represented as the risk of spreading between passengers. The spreading risk was defined as the contact duration multiplied by a constant value. The scenarios were compared in terms of their impact on the network and confirmed previous observations in the literature, that is that the most central vehicle trips in the public transportation network are also the most susceptible to infection. The second part of the analysis focused on a newly proposed vehicle trip network, which represents the public transit network as a network of vehicle trips instead of passengers. We showed that centrality metrics on the vehicle trip network provided a more efficient way to detect the set of vehicle trips most susceptible to disease, and this estimation can be done much faster than running simulations on the contact network.

In the rest of this section we present an alternative way to model the risk of disease spreading between passengers, which exploits the community network structure introduced in this work. We define the infection transmission probability between a pair of passengers to be a function of both the connection strength value used in the community network and contact duration. We believe this multidimensional transmission probability more accurately represents the level of connectivity between pairs of passengers. Using these new transmission functions, we implement similar spreading scenarios to the ones in Bóta et al. (2017a) and identify and rank the vehicle trips susceptible to disease spreading. Below we define the new transmission probability function and the infection model, then present the new vehicle-trips identified to be at highest risk.

4.3.1 Experiment Setup

In order to simulate an epidemic outbreak on the contact network we use the well-known discrete compartmental susceptible-infected (SI) model. In the SI infection process all nodes adopt one of two available states: susceptible (S) or infected (I). Real values denoted as edge infection probabilities are assigned to the links of the network and denoted as $w_e \in [0, 1]$. The infection spreads in a network when susceptible nodes adopt an infected state. This is done in discrete time steps in an iterative manner starting from an initially infected set of nodes. In each iteration each infected node tries to infect its susceptible neighbors according to the transmission probability of the link connecting them. If the attempt is successful, the susceptible node is transformed into an infected one in the next iteration. In this work we limit the number of iterations to five, representing a complete work week with recurrent commuting patterns. The inputs of this model are: 1. a contact network of individuals, 2. an assignment of weights to the links of the network which represent the infection transmission probabilities and 3. the set of initially infected nodes, e.g., individuals.

We use the original structure of the contact network, which connects all pairs of passengers that travel on a vehicle-trip together as the basis of this experiment. The link weights are computed in the following way. Since the nodes and links of the community network are a subset of the nodes and links of the contact network, if a link between two nodes is present in the community network we will assign its connection strength value to the corresponding link in the contact network. We do this in all cases where such an assignment can be made. In order to account for extreme outliers of connection strength we cap all values at 100, then rescale all values to the interval of 0.1 and 0.8 using a standard feature scaling method. The 0.8 upper bound is set because a transmission probability of 1.0 is assumed to be unrealistic. For all links of the contact network that do not have a corresponding link in the community network we assign a uniform infection value of 0.05. In contrast to the duration-based value used in Bóta et al. (2017a), this enables us to capture the increased spreading risk between passengers sharing similar travel patterns, while still allowing disease to spread between travelers simply sharing a vehicle. Future work will explore the model's sensitivity to the uniform infection assignment using in this work.

Following the procedure in Bóta et al. (2017a), we randomly select 100 passengers from the network to be initially infected. Due to the probabilistic nature of the simulation model, we run the SI infection model $k = 10000$ times to quantify the likelihood of each nod being in an infectious state at the end of the fifth time step.

4.3.2 Vehicle Trip Ranking

The infection model constructed above provides the likelihood of infection for each passenger in the contact network at the end of the simulation, i.e., after five days. As in Bóta et al. (2017a), we compute a similar infection value for vehicle trips by summing the probability of infection for all passengers on a given vehicle trip. While

Table 4 Ranking of vehicle trip where infection is most likely to appear

Rank	Route number	Start time
1.	25	6:52
2.	17	6:13
3.	25	5:11
4.	14	16:54
5.	5	5:36
6.	25	7:16
7.	18	6:07
8.	11	15:52
9.	11	17:14
10.	61	6:07

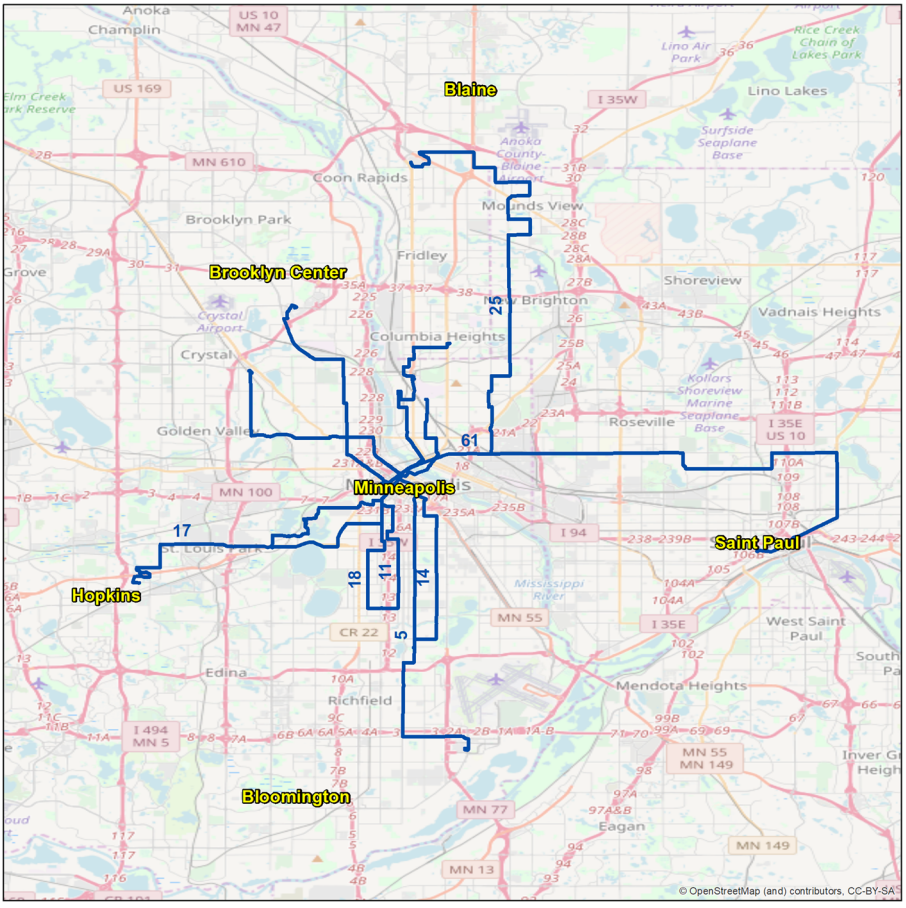

Fig. 8 Vehicle trips most likely to carry infected passengers in the public transit system of Twin Cities, MN

this does not represent a probability value in a strict sense, this value is proportional to the risk of getting infected on a given vehicle trip.

The routes which contain the highest risk vehicle trips, and corresponding travel times identified in this study are presented in Table 4. Figure 8 shows the routes on the map of the city. Coinciding with the findings in Section 3 and also in Bóta et al. (2017a), the vehicle trips are in the morning and afternoon peak hours. Also coinciding with the results in Bóta et al. (2017a) all of the high risk routes identified go through the city center.

Since the goal of Section 3.2 – identifying the most frequent trip combinations – and the goal of this section – identifying the most risky vehicle trips in the case of an epidemic outbreak – are different, the difference between the selected vehicle trips are not surprising. We observe some similarities between the "highest risk" vehicle trips identified here, and the most frequent vehicle trip combinations (identified in Section 3.2). For example, route 18 connects Bloomington with the city center, while route 61 connects St. Paul to the same destination. We have seen in Section 3.2, that these target-destination pairs are present in the most frequent combinations as well. As these routes connect the most important parts of the city, some similarity is expected.

The set of high risk vehicle trips identified here partially correspond to those identified in Bóta et al. (2017a). Specifically, the methodology proposed in Bóta et al. (2017a) identifies vehicle trips on the routes 7, 9, 11, 14, 17, 25, 61 and 94 as most at risk. These are routes crossing or connecting to the city center in the peak hours, so even though they do not completely match those in Table 4, the pattern they present is the same. The similarity in the set and type of routes identified in both studies points to the critical role of the network structure in modeling outbreak risk. Future research will continue to build on this application, and further explore the robustness and sensitivity of the proposed methodology. The relevant sensitivity analysis is, however, outside the scope of this work.

5 Limitations

This study is subject to certain limitations. We note, that the transit demand model used as an input of both works was used to generate commuting patterns for single workday. This limitation is most critical for the epidemic application due to the implicit assumption that daily travel patterns remain constant during a five day work week. While this assumption is supported by a recent study (Sun et al. 2013), more long term observations potentially involving weekends and public holidays would improve the quality of the results presented in this paper.

Further, while we present an alternative metric to quantify disease transmission risk between passengers (compared to contact duration alone), without any real-world observations on an outbreak validating the results of this work remains a challenge, and will be the focus of future research. One solution would be to use epidemic data recorded in another major city (Sun et al. 2014), but adapting such data sources to different environment presents its own set of challenges.

6 Conclusions

In order to better understand passenger commuting habits in a public transportation system, in this paper we analyzed the contact patterns from a public transit assignment model in a major metropolitan city. We did this by defining two novel network structures to detect and quantify the movement patterns of passengers. The transfer network tracks the movements of atomic passenger group while the community network links passengers with similar travel patterns. We presented applications for both networks, specifically to identify the most frequently used travel paths, i.e., routes and transfers, and model epidemic risk posed by passengers of a public transit network, respectively.

Our main findings correspond to existing literature, while providing a more detailed analysis of passenger commuting habits. We have shown that the most frequent vehicle trip combinations follow a similar pattern. The trip pairs identified using the transfer network identified peak morning and afternoon trips that connect the outlying suburbs with the CBD. The transfer network was demonstrated as a tool to efficiently detect the most frequently used vehicle trip combinations in the public transportation system.

Our results also reinforce previous observations in revealing components of the transit system at risk from an epidemic outbreak. For this purpose, we have shown how the connection strength value can be used to estimate physical contact between passengers and to identify the vehicle trips most likely to carry infected passengers: the routes crossing or connecting to the city center in the peak hours.

In the future we plan to extend the epidemic application and more thoroughly investigate how connection strength can be used to improve risk analysis in a public transportation setting.

Acknowledgements The authors are grateful to Metropolitan Council of Twin Cities for sharing the activity-based travel demand model with researchers at the University of Minnesota. Any limitations of this study remains the sole responsibility of the authors. László Hajdu acknowledges the National Research, Development and Innovation Office (NKFIH) for funding the project "Graph Optimisation and Big Data" (Grant No: SNN-117879) and the support of the EU-funded Hungarian grant EFOP-3.6.3-VEKOP-16-2017-00002. Miklós Krész acknowledges the European Commission for funding the InnoRenew CoE project (Grant Agreement #739574) under the Horizon2020 Widespread-Teaming program and the support of the EU-funded Hungarian grant EFOP-3.6.2-16-2017-00015. András Bóta acknowledges the Olle Engkvist Byggmästare Foundation.

Funding Information Open access funding provided by Umea University.

References

Aicher C, Jacobs AZ, Clauset A (2014) Learning latent block structure in weighted networks. J Complex Networks 3(2):221–248. https://doi.org/10.1093/comnet/cnu026

Balcan D (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci USA 106:21,484–21,489. https://doi.org/10.1073/pnas.0906910106

Bao J, Xu C, Liu P, Wang W (2017) Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests. Netw Spat Econ 17(4):1231–1253. https://doi.org/10.1007/s11067-017-9366-x

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008(10):P10,008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bóta A, Gardner L, Khani A (2017a) Identifying critical components of a public transit system for outbreak control. Netw Spat Econ 17(4):1137–1159. https://doi.org/10.1007/s11067-017-9361-2

Bóta A, Gardner L, Khani A (2017b) Modeling the spread of infection in public transit networks: a decision-support tool for outbreak planning and control. In: Transportation research board 96th annual meeting

Bóta A, Kovács L (2014) The community structure of word association graphs. In: Proceedings of the 9th international conference on applied informatics, pp 113–120. https://doi.org/10.14794/ICAI.9.2014.1.113

Bóta A, Krész M (2015) A high resolution clique-based overlapping community detection algorithm for small-world networks. Informatica 39(2):177–187

Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462–465. https://doi.org/10.1038/nature04292

Bron C, Kerbosch J (1973) Algorithm 457: Finding all cliques of an undirected graph. Commun ACM 16(9):575–577. https://doi.org/10.1145/362342.362367

Cambridge Systematics Inc. (2015) 2010 travel behavior inventory: model estimation and validation report prepared for metropolitan council

Carlsson-Kanyama A, Lindén AL (1999) Travel patterns and environmental effects now and in the future: implications of differences in energy consumption among socio-economic groups. Ecol Econ 30(3):405–417. https://doi.org/10.1016/S0921-8009(99)00006-3

Chen N, Gardner L, Rey D (2016) A bi-level optimization model for the development of real-time strategies to minimize epidemic spreading risk in air traffic networks. Transportation Research Record: Journal of the Transportation Research Board 2569:62–69. https://doi.org/10.3141/2569-07

Chin A, Xu B, Yin F, Wang X, Wang W, Fan X, Hong D, Wang Y (2012) Using proximity and homophily to connect conference attendees in a mobile social network. In: 2012 32nd international conference on distributed computing systems workshops. IEEE, pp 79–87. https://doi.org/10.1109/ICDCSW.2012.56

Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci 106(36):15,274–15,278. https://doi.org/10.1073/pnas.0900282106

Eppstein D, Löffler M, Strash D (2010) Listing all maximal cliques in sparse graphs in near-optimal time. In: International symposium on algorithms and computation. Springer, pp 403–414. https://doi.org/10.1007/978-3-642-17517-6_36

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3-5):75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Funk S, Salathé M, Jansen VAA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 7:1247–1256. https://doi.org/10.1098/rsif.2010.0142

Gardner L, Fajardo D, Waller S (2012) Inferring infection-spreading links in an air traffic network. Transportation Research Record: Journal of the Transportation Research Board 2300(1):13–21. https://doi.org/10.3141/2300-02

Háznagy A, Fi I, London A, Németh T (2015) Complex network analysis of public transportation networks: a comprehensive study. In: 2015 international conference on models and technologies for intelligent transportation systems (MT-ITS). IEEE, pp 371–378. https://doi.org/10.1109/MTITS.2015.7223282

Hoogendoorn S, Bovy P (2005) Pedestrian travel behavior modeling. Netw Spat Econ 5(2):193–216. https://doi.org/10.1007/s11067-005-2629-y

Huerta R, Tsimring LS (2002) Contact tracing and epidemics control in social networks. Phys Rev E Stat Nonlin Soft Matter Phys 66(056):115. https://doi.org/10.1103/PhysRevE.66.056115

Illenberger J, Nagel K, Flötteröd G (2012) The role of spatial interaction in social networks. Netw Spat Econ 13(3):1–28. https://doi.org/10.1007/s11067-012-9180-4

Khani A (2013) Models and solution algorithms for transit and intermodal passenger assignment (development of fast-trips model). PhD Dissertation, University of Arizona, Tucson AZ, USA

Khani A, Beduhn T, Duthie J, Boyles S, Jafari E (2014) A transit route choice model for application in dynamic transit assignment. Innovations in Travel Modeling. Baltimore, MD

Khani A, Hickman M, Noh H (2015) Trip-based path algorithms using the transit network hierarchy. Netw Spat Econ 15(3):635–653. https://doi.org/10.1007/s11067-014-9249-3

Krzakala F, Moore C, Mossel E, Neeman J, Sly A, Zdeborová L, Zhang P (2013) Spectral redemption in clustering sparse networks. Proc Natl Acad Sci 110(52):20,935–20,940. https://doi.org/10.1073/pnas.1312486110

Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure in complex networks. New J Phys 11(3):033,015. https://doi.org/10.1088/1367-2630/11/3/033015

Leskovec J, Lang KJ, Mahoney M (2010) Empirical comparison of algorithms for network community detection. In: Proceedings of the 19th international conference on World Wide Web. ACM, pp 631–640, https://doi.org/10.1145/1772690.1772755

Manuel C, Roberto C, Michael F (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transp Res B Methodol 40(6):437–459. https://doi.org/10.1016/j.trb.2005.05.006

de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: the privacy bounds of human mobility. Sci Rep 3:1376. https://doi.org/10.1038/srep01376

Nassir N, Khani A, Hickman M, Noh H (2012) An intermodal optimal multi-destination tour algorithm with dynamic travel times. Transportation Research Record: Journal of the Transportation Research Board 2283:57–66. https://doi.org/10.3141/2283-06

Newman ME (2013) Spectral methods for community detection and graph partitioning. Phys Rev E 88(4):042,822. https://doi.org/10.1103/PhysRevE.88.042822

Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026,113. https://doi.org/10.1103/PhysRevE.69.026113

Palla G, Derényi I., Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818. https://doi.org/10.1038/nature03607

Peixoto TP (2014) Efficient monte carlo and greedy heuristic for the inference of stochastic block models. Phys Rev E 89(1):012,804. https://doi.org/10.1103/PhysRevE.89.012804

Pendyala R, Kondhuri K, Chiu YC, Hickman M, Noh H, Waddell P, Wang L, You D, Gardner B (2012) Integrated land use-transport model system with dynamic time-dependent activity-travel microsimulation. Transportation Research Record: Journal of the Transportation Research Board 2203:19–27. https://doi.org/10.3141/2303-03

Ramadurai G, Ukkusuri S (2010) Dynamic user equilibrium model for combined activity-travel choices using activity-travel supernetwork representation. Netw Spat Econ 10(2):273–292. https://doi.org/10.1007/s11067-008-9078-3

Rey D, Gardner L, Waller ST (2016) Finding outbreak trees in networks with limited information. Netw Spat Econ 16(2):687–721. https://doi.org/10.1007/s11067-015-9294-6

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105(4):1118–1123. https://doi.org/10.1073/pnas.0706851105

Saberi M, Rashidi TH, Ghasri M, Ewe K (2018) A complex network methodology for travel demand model evaluation and validation. Netw Spat Econ, pp 1–23. https://doi.org/10.1007/s11067-018-9397-y

Salathé M (2010) A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci USA 107:22,020–22,025. https://doi.org/10.1073/pnas.1009094108

Song C, Qu Z, Blumm N, Barabási AL (2010) Limits of predictability in human mobility. Science 327:1018–1021. https://doi.org/10.1126/science.1177170

Sun L, Axhausen KW, Lee DH, Cebrian M (2014) Efficient detection of contagious outbreaks in massive metropolitan encounter networks. Sci Rep 4:5099. https://doi.org/10.1038/srep05099

Sun L, Axhausen KW, Lee DH, Huang X (2013) Understanding metropolitan patterns of daily encounters. Proceedings of the National Academy of Sciences 110(34):13,774–13,779. https://doi.org/10.1073/pnas.1306440110

Wang R, Nakamura F, Okamura T, Warita H (2011) How the risk-taking personality influences commute drivers' departure time choices. Procedia Soc Behav Sci 16(Supplement C):814–819. https://doi.org/10.1016/j.sbspro.2011.04.500

Wen-Tai L, Ching-Fu C (2011) Behavioral intentions of public transit passengers-the roles of service quality, perceived value, satisfaction and involvement. Transp Policy 18(2):318–325. https://doi.org/10.1016/j.tranpol.2010.09.003

Wu ZH, Lin YF, Gregory S, Wan HY, Tian SF (2012) Balanced multi-label propagation for overlapping community detection in social networks. J Comput Sci Technol 27(3):468–479. https://doi.org/10.1007/s11390-012-1236-x

Xie J, Kelley S, Szymanski BK (2013) Overlapping community detection in networks: the state-of-the-art and comparative study. ACM Comput Surv 45(4):43. https://doi.org/10.1145/2501654.2501657

Yuan YC, Gay G (2006) Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. J Comput-Mediat Commun 11(4):1062–1084. https://doi.org/10.1111/j.1083-6101.2006.00308.x

Affiliations

László Hajdu[1] · András Bóta[2,3] · Miklós Krész[4,5] · Alireza Khani[6] · Lauren M. Gardner[3,7]

László Hajdu
hajdul@inf.u-szeged.hu

Miklós Krész
kresz@jgypk.szte.hu; miklos.kresz@innorenew.eu

Alireza Khani
akhani@umn.edu

Lauren M.Gardner
gardner@jhu.edu

1 Institute of Informatics, University of Szeged, Árpád tér 2, 6720 Szeged, Hungary

2 Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden

3 Research Centre for Integrated Transport Innovation, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

4 Department of Applied Informatics, University of Szeged, Boldogasszony sgt. 6, 6720 Szeged, Hungary

5 Innorenew CoE, Livade 6, 6310 Izola, Slovenia

6 Department of Civil, Environmental and Geo- Engineering, University of Minnesota Twin Cities, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA

7 Department of Civil Engineering, Johns Hopkins University, 3400 N. Charles St, Baltimore, MD 21218, USA