Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 11 (2016) pp. 011-020

Passenger Pattern Recognition in Railway Stations using Quantitative Network Analysis

Marian Sorin Nistor[1], Martin Zsifkovits[2], Silja Meyer-Nieberg[3], and Stefan Wolfgang Pickl[4]

Institute for Computer Science, Mathematics and Operations Research,
Universität der Bundeswehr München
85577 Neubiberg, Germany
[1]sorin.nistor@unibw.de, [2]martin.zsifkovits@unibw.de, [3]silja.meyer-nieberg@unibw.de, [4]stefan.pickl@unibw.de

***Abstract*: Security installations in public transportation systems play a more and more important role due to the increased threat of terroristic attacks on such critical infrastructures. However the placement of such installations is rather critical as their installation must not influence the passenger flow and so cause inconvenience, delays or crowds. Especially the need for an open system in public transportation demands for solutions that do not influence or delay passenger flows. For giving advice in this decision problem we observed passenger flows in the Munich central station and applied quantitative graph theory on the network data. The main aim was to better understand the status quo of passenger patterns. With its help, critical parameter constellations can be identified and investigated in detail. Furthermore we are able to identify special groups of passengers and the differences in their behavior. The method of quantitative network analysis in this field is a novel approach and to the best of our knowledge it was not applied in previous analysis.**

Keywords: network analysis, quantitative networks, passenger pattern recognition, RiKoV, RE(H)STRAIN.

I. Introduction

The protection of critical infrastructures is of high importance for every nation. Disturbances such as blackouts might lead to cascading effects due to high interconnections among several types of infrastructures. Especially after terroristic attacks to the rail-bound public transport system, its protection became a major issue in public decision making. For example, the series of attacks on stations in Madrid 2004 caused almost 200 deaths and 1.400 injured people. The recent (21.08.2015) attack of a terrorist with machineguns in a Thalys high-speed train on the border between Belgium and France did not cause human victims, but tremendous media attention. This shows the insecurity in the public and the importance of the topic at hand, as such critical infrastructures are essential for the functioning of a modern society [1].

For protecting the public from terroristic attacks in this vulnerable transportation system, risk management [2] as well as security installations and measures are coming more and more into the focus of ongoing research [3–7]. Here, the emphasis lies on the protection of train stations as one of the most vulnerable parts of the network and being the link to the trains. The effectiveness of security measures depends, however, strongly on the location where they are placed [8]. While the problem of optimal sensor placement is often considered in literature, see e.g. [9], we argue that additional factors have to be taken into account. While it is common to consider the station layout for providing optimal sensor coverage, see e.g. [10], the movement behavior of crowds on the effectiveness of the security measures or surveillance sensors themselves is only seldom investigated. Therefore, agent-based simulation models play a major role in the analysis, whereby an adequate parameterization of passenger behavior is a crucial task. However, there is hardly any literature available on how different people behave and move in public places.

In order to fill this research gap we observed passengers in the main station in Munich in order to get first insights in passenger patterns and recognize effects on the overall picture. Therewith we want to see where security installations might lead to promising screening rates and at the same time do not influence the overall crowd behavior. This specific knowledge of passenger flows and preferences of specific groups might support the creation of simulation models with advanced parameters and also the decision of optimal placement of security measures in a further step.

For instance, interdisciplinary applications such as web graphs and hypertext, biological/chemical structures, and financial networks have been examined extensively [1, 11–15], while, to the best of our knowledge, the evaluation of passenger patterns using this approach was not done before and therefore brings a totally new perspective into the field.

First visions of combining modern risk management concepts with strategic forecasting strategies on complex graph theoretic structures where discussed generating the Operations Research OR2010 conference topic "Mastering Complexity". There is still a lot of potential in this innovative combination [16].

The rest of the paper is structured as follows: In section two we will provide some insights into the observation method and the gathered data. Section three gives insights into the quantitative network analysis and in section four we will present some selected results. Finally we conclude the paper in section five and give insights into further research.

II. Real-Life Observations

For the observations at the Munich main station three students of the Universität der Bundeswehr München positioned randomly in the vicinity of the exits and tracked passengers based on a convenience sample, meaning that they chose any random person that passed them after being available for a new observation. The passengers were not in contact with the observation team and were not aware of being tracked through the station. This was especially possible through finding positions that allow for seeing most areas in the station and therefore reduce the need of walking and following people.

For each observation a printed map of the station was used and the walking path drawn into it. Additionally, the observation team noted the date, time, the observed person's sex, the estimated age and the fact if the person was alone or in companionship. Observations started with the entry of a person into the train station and ended either with their exit at a gate, exit into a train, or when a person spent more than ten minutes at a spot (e.g. track, restaurant, etc.). We conducted most of the observations (700) in the morning between seven and eleven a.m. in order to get a picture of more crushed situations. Compared to this we conducted observations (218) as a control group in the evening between five and six p.m. Generally, the station can be split into a departure and arrival area including the main hall and a shopping and food area behind. The main hall can also be used as shortcut between two streets which leads to a high frequency of walk-through traffic. The two areas are shown in Figures 1 and 2.

Figure 1. Arrival and Departure Area. Copyrights of [17]

Figure 2. Shopping and Food Area. Copyrights of [17]

The positions where people stopped were marked and the pauses recorded in seconds. Table 1 gives an overview of the observation dates and the numbers of observed passengers.

Table 1. Observation data.

Weekday	**Date**	**7-11 am.**	**5-6 pm.**
Monday	02.03.2015	100 Obs.	30 Obs.
Tuesday	10.02.2015	100 Obs.	30 Obs.
Wednesday	18.02.2015	100 Obs.	30 Obs.
Thursday	12.02.2015	100 Obs.	30 Obs.
Friday	06.02.2015	100 Obs.	37 Obs.
Saturday	14.02.2015	100 Obs.	30 Obs.
Sunday	15.02.2015	100 Obs.	31 Obs.

During the observations some special events were noticed that might influence the results. The weather conditions during the observations were cold and sometimes rainy, which might also influence the time spent in a (warmer and dry) train station compared to a hot summer day. On February 14th the station was more crowded than else when due to a soccer (home) game of Bayern Munich. Furthermore it was Valentine's Day and free roses were distributed to passengers, which led to some stops in the distribution area. On Sunday, February 15th there was a raffle (tombola) in the main hall which again caused some crowds of passengers and influenced additional stops. On Monday, March 2nd there was a rescue operation of the ambulance at train tracks 11-16 which influenced walking paths. However, such events are common in such stations and therefore the dataset is accepted.

After collecting the data the train station was split into sectors and single positions were defined as nodes. The observed path on special nodes of individuals was then coded in a Microsoft Excel file with the duration of stays.

The basic idea of the project at hand was to analyze the data set using quantitative graph theory. Therewith we are aiming to reduce the complexity in the huge data set and get very clear and precise insights in the behavior of different groups of people.

III. Quantitative Network Analysis

Networks or graphs are relational structures for modelling complex relationships between different entities [18, 19]. The entities can differ regarding the field of application. Networks representing structural objects are often referred to as complex networks [20] as they are based on the idea to only consider networks that possess unique and non-random topologies. Suppose that we are able to generate all existing networks. After this, we remove ''simple'' networks given by k-regular graphs and random graphs [21, 22]. A graph is called k-regular when every vertex has the same degree. Thus the network is quite symmetrical and is therefore not considered as complex. Similarly, random networks do not have a structure because they are random. Following this augmentation, complex networks have a fixed topology with unique graph-theoretical patterns. In this paper, we elaborate on weighted and undirected passenger networks, where we focus on quantitative graph theory [23]. For further explanation see [1].

Based on the coded observations, four groups of networks are highlighted. The group of networks based on gender (men, women), age (under 40, over 40), time of the day (morning, afternoon), and time of the week (business days, weekend). A network consists of all visited nodes by the passengers of a specific group (e.g. gender – Men/Women networks), and all edges (connections) that they create between the nodes. The not visited (isolated) nodes are not considered in these networks. Each network has different structural properties. The aim of this paper is to understand how the networks differentiate within a group based on their structural properties.

In order to perform a more sensitive analysis among the networks, for each, a ten percent frequency filter is applied. The frequency percentage is given by the probability of a person to travel from point A to point B, in a specific network. This results in another network that considers the probability of passing from one node to another higher than ten percent.

For each group of networks, a complete network is calculated. E.g. the Men and Women networks are collapsed in a gender network including all the unique nodes and edges of the two sources (see Table 7). The ten percent frequency filter is also applied for the complete networks. Finally, each group is having six networks, and the four groups totaling 24 networks.

Modeling the passenger flow as a network can be beneficial even by using the most basic structural property of a network node. This is the degree of a node which is represented by the number of its adjacent connections. In this case it represents the number of passengers from a specific group visiting a specific element of the station (e.g. the information desk). An example is presented in Figure 3 [24], where the most frequently visited ten elements of the station during the weekend can be observed. Here, as expected, the most visited element of the train station is the information desk. Surprisingly, among the top ten most frequently visited elements of the station there is not even one single train platform included. Even more, the exits are passed even less frequent than other elements, e.g. food stores, newsstand, etc. More than half of the elements included in Figure 3 are related with food. This can show that during the weekends the train station serves more as a food court than as a public transportation service (especially departure).

For a more profound quantitative analysis of the passengers flow, several network metrics are proposed to describe the topologies of the networks. These are the most common and meaningful for this type of networks [18, 19, 23, 25].The selected metrics are the network diameter [18], density [26], average path length [27], and modularity [28–30]

To determine the structural complexity of the networks, two entropy measures are used [4] based on an information-theoretic analysis approach introduced by Dehmer [31]. These will be analyzed along with a couple of classical measures, e.g. Wiener index [32], and Randić index [33]. The statistical analysis has been performed using the programming language R [34] (Release version 3.1.3).

IV. Results

As mentioned in the previous section, in this paper the passengers flow is modeled as an undirected and weighted network $N = (V, E)$. $V = \{v_i | i = 1, 2, \dots, n\}$ is the set of network nodes representing the elements of the train station. $E = \{(v_i, v_j) | v_i, v_j \in V\}$ represents the set of edges connecting the network nodes and having the number of passengers traveling from one node to another as weights w_{ij}, where $w_{ij} \geq 0$.

In the following, the network measures are described and their numerical results are interpreted. Individual networks are analyzed and compared with their complementary network in every group in particular, and with the rest of the networks in general.

A. Network diameter

The diameter of a network $\varnothing(N)$ is the greatest number of nodes to be passed in order to travel from one node to another when paths which backtrack, detour, or loop are excluded

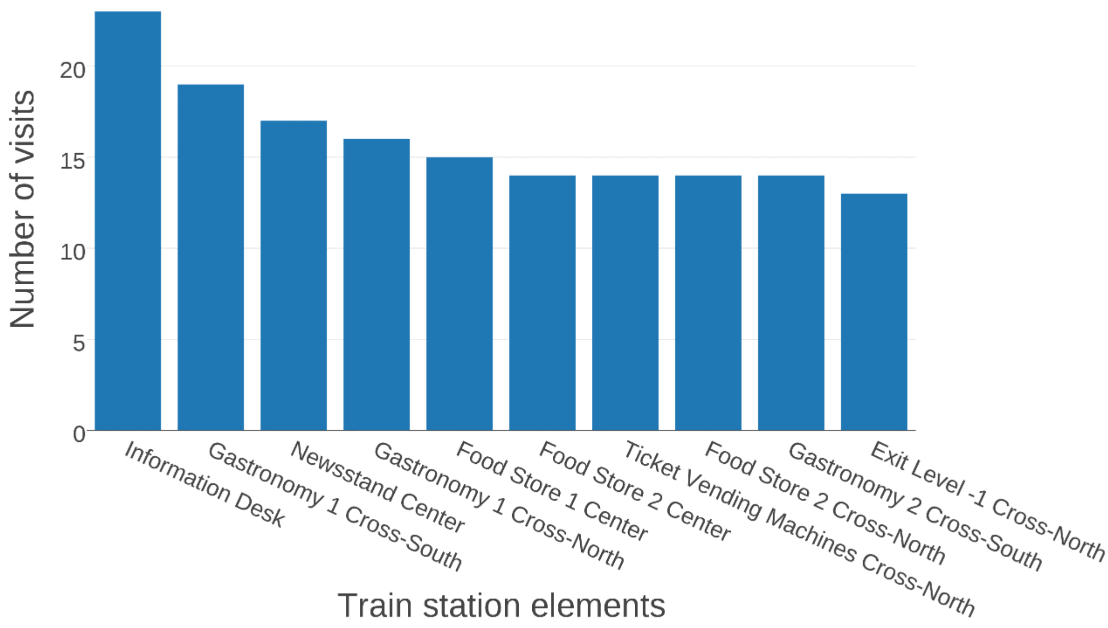

Figure 3. Bar chart representation of the most frequently visited ten train station elements over a weekend by the passengers, generated with [24]

from consideration [18]. This helps to understand how tight or loose a network is (see Figure 4).

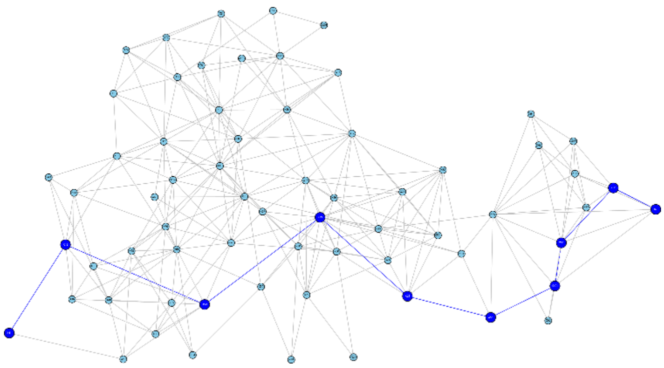

Figure 4. Visual representation of a network diameter path created by the passengers flow, generated with [28]

If $d(v_i, v_j)$ is considered the shortest path between any two elements of the station, then

$$\varnothing(N) = max_{v_i,v_j} d(v_i, v_j) \quad (1)$$

where $v_i, v_j \in V$. This is the so-called "longest shortest path" between any two elements of the station.

In general, the highest value of diameter is reached by a path network when its diameter value is given by the number of nodes. On the opposite, a full connected network has the value one. In Table 2 the diameter of the networks fluctuates between the values ten and twenty. The lowest value is reached by the *Men* network where the diameter is ten for a network with 354 connections (edges) between its station elements (nodes). This is the tightest network based on its diameter. Interesting is that the highest diameter value is reached by the *Under 40* network. This network has one more node and 30 extra edges than the previous mentioned, and contrary to expectations, this is the loosest network.

Table 2. Networks diameter results.

N	V	E	$\varnothing(N)$
Men	65	354	10
Women	66	354	14
Under 40	66	384	20
Over 40	66	315	15
Morning	66	385	14
Afternoon	66	262	19
Business days	66	389	14
Weekend	66	278	15

Notation: N – network; V – number of nodes; E – number of edges; $\varnothing(N)$ – diameter.

Observing the network diameter results from Table 2, we can spot interesting behaviors of the passengers. The men tend to browse the train station in a star movement by visiting more neighbors of each element of the station. Women tend to create more a path movement visiting the elements of the station rather linearly than the men do. When analyzing the age group, passengers under 40 years are walking more in a linear movement in the train station than the elder ones. And in the mornings people tend to browse more in a star movement than in the afternoons. For the week time group we observe the less meaningful difference compared with the other groups.

For an extended analysis, the ten percent filter mentioned in the previous section is applied. The results can be observed in Table 7. Here the situation changes very sharply for the day time and the week time groups. A very large amount of passengers tends to have a path movement in the network in the mornings, and during the business days. This is opposite in the afternoons and in the weekend days. When analyzing the gender and age, the situation tends to preserve as for the initial networks.

B. *Network density*

The ratio between the number of edges and the number of possible edges gives the density of the network [26]. This can be defined as:

$$D = \frac{2*E}{V(V-1)} \quad (2)$$

For this analysis we consider self-loops in determining the network density. Thus, the possibility that a passenger can return to the same element of the station without reaching another one in between is considered. Self-loops can be considered in the density formula as:

$$D_{s-l} = \frac{2*E}{V(V+1)} \quad (3)$$

The higher the value, the denser is the network. The opposite, a network with only a few edges is a sparse network.

A full network is a network with all its possible edges and then the density reaches the maximum value of one. The minimum value of the density is reached when the network has a path shape. In Table 3, the densest network is the *Business days* network and has 17.59 % of its potential.

The sparsest network, among the other network, is the *Afternoon* network where its density reaches only 11.85% of the maximum possible value.

Table 3. Networks density results.

N	V	E	$D_{s-l}(N)$
Men	65	354	0.165
Women	66	354	0.1601
Under 40	66	384	0.1737
Over 40	66	315	0.1425
Morning	66	385	0.1741
Afternoon	66	262	0.1185
Business days	66	389	0.1759
Weekend	66	278	0.1257

Notation: N – network; V – number of nodes; E – number of edges; $D_{s-l}(N)$ – density.

In Table 3, the most balanced density value in all groups is between men and women. Thereafter, we can observe that passengers under 40 years of age tend to visit slightly more nodes than the elder ones. Concerning the other groups, we can observe that in the mornings and during business days people create substantially denser networks compared with afternoons and weekend days networks. This can be also seen in Figure 5.

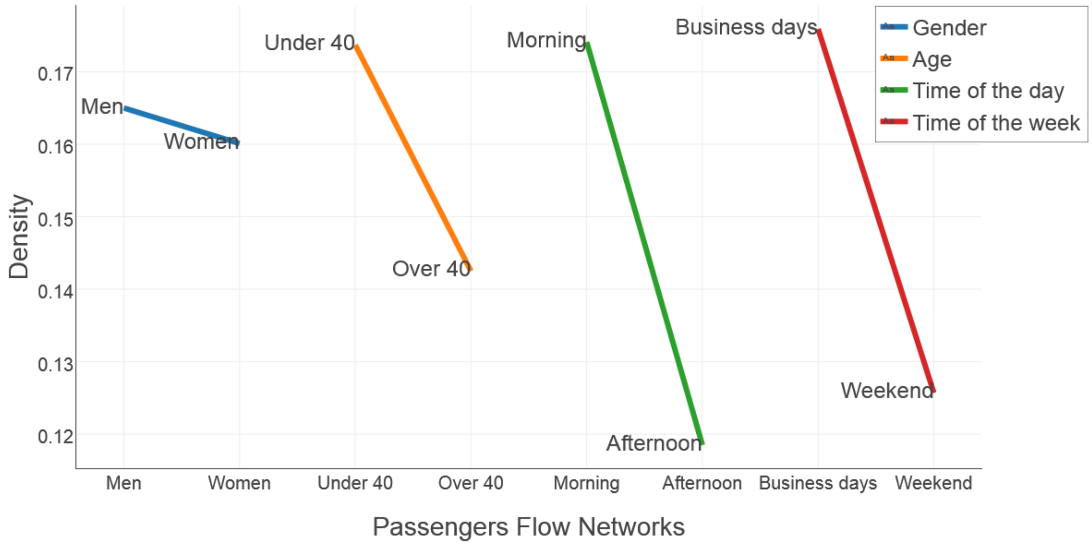

Figure 5. Visual representation of the passengers flow network density values, generated with [24]

The extended analysis for density with the ten percent filter can be observed in Table 7. Here, the density tends to be opposite than for the initial networks in the age group, day time group, and week time group. Thus, the elder passengers, the passengers in the afternoon, and the passengers during the weekend days, visit more nodes in the network. We also notice that men are visiting slightly more nodes than women, compared with the initial network case.

C. Average path length

The average path length in a network is determined based on the average of the shortest paths for all possible pairs of nodes in a network. The shortest path between a pair of nodes in a network is a path with the minimal number of nodes [27]. The average path length can be defined as:

$$\lambda = \frac{1}{V(V-1)} \sum_{(v_i,v_j)\in E;\, v_i \neq v_j} d(v_i, v_j) \quad (4)$$

The shortest path length is opposite to the diameter of a network when being determined, but the values have similar interpretation. Smaller values of the average of the shortest paths indicate efficient connections in the network. The greater values indicate in average a higher number of nodes to be passed in the network between every two nodes. It is interesting to observe in Table 4 that from all the present networks, the extreme values are found in the same group. Therefore, the *Afternoon* network created by the passengers is the most efficient, while its opposite is the *Morning* network.

Furthermore, in Table 4 we observe that there is just a small difference between men and women. Contrariwise, observing the other three groups, we can conclude that the passengers under 40 years old, the passengers in the mornings, and the passengers during the business days, create in average shorter paths in the networks between the elements of the station. The opposite holds for the complementary networks from this three groups.

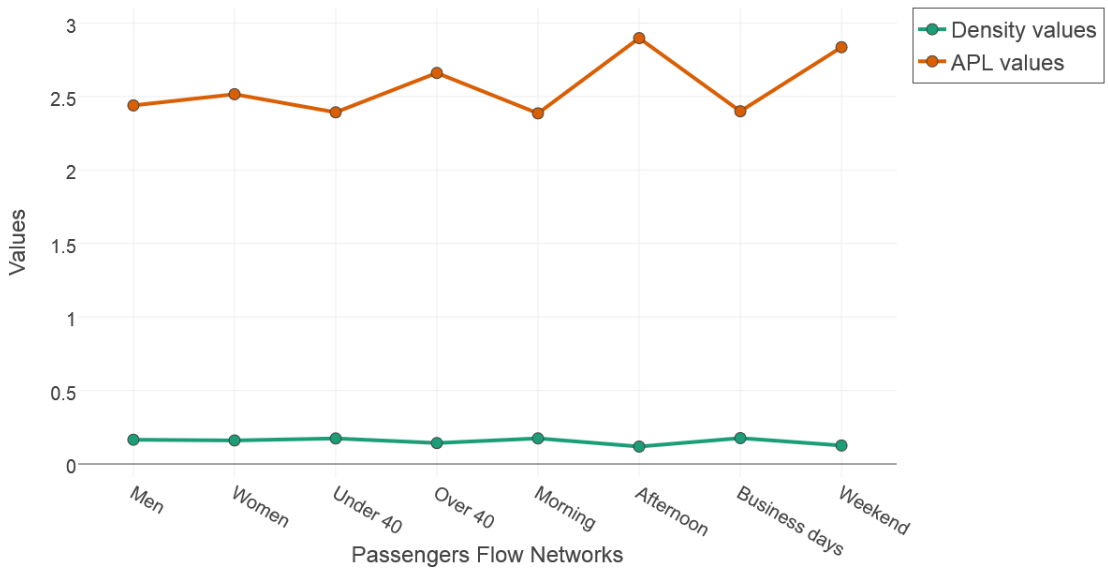

Figure 6. Visual representation of the passengers flow networks density values and the average path length values to observe their linear correlation, generated with [24]

Table 4. Networks average path length results.

N	***V***	***E***	$\boldsymbol{\lambda(N)}$
Men	65	354	2.4399
Women	66	354	2.5156
Under 40	66	384	2.3935
Over 40	66	315	2.6606
Morning	66	385	2.386
Afternoon	66	262	2.8979
Business days	66	389	2.4005
Weekend	66	278	2.8359

Notation: ***N*** – network; ***V*** – number of nodes; ***E*** – number of edges; $\boldsymbol{\lambda(N)}$ **–** average path length.

Because the observations between density and average path length are having similar changes, we measure the linear correlation (dependence) between the two measures results. We use the Pearson's correlation coefficient. Appling this measure on the two samples, the result is -0.9660123, which means that the density values are highly negatively correlated with the average path length values. Thus, when the values of density are increasing, the values of average path length are decreasing with the same intensity. The same holds for the opposite situation. This can be seen in Figure 6.

For a deeper analysis, the ten percent or higher probability filter that a passenger is visiting the adjacent nodes can be observed in Table 7. Here, men tend to have longer paths in the train station than women. A similar observation for the passengers visiting the station in the mornings, and during the business days is observed. The age group remains more balanced.

D. Network modularity

Another measure used to study the passengers flow is the network modularity. This measure helps to understand the denseness of groups of passengers in the train station. In this case, to understand how the passengers form groups based on different criteria (e.g. age, gender, time of the day, etc.) Therefore, modularity is a network property of dividing the network into groups. A group of a network is a subset connected part of that network to measure how separated the different node types are from each other [29].

For this analysis, the greedy optimization of modularity is used [28, 30]. The graph modularity [28] is defined as:

$$Q = \frac{1}{2E}\sum_{v_i,v_j\in V}\left(A_{ij} - \frac{k_{v_i}k_{v_j}}{2E}\right)\delta(c_{v_i}, c_{v_j}) \quad (5)$$

where E – number of edges; A_{ij} – adjacency matrix element of N; k_{v_i} and k_{v_j} – degree of v_i and v_j; c_{v_i} and c_{v_j} – type/component of v_i and v_j; $\delta(c_{v_i}, c_{v_j})$ is 1 if $c_{v_i} = c_{v_j}$, or 0 otherwise [28].

A high modularity means a dense number of connections between the nodes within groups of a network, and only few connections between nodes in different groups [30]. Starting with the original networks which can be seen in Table 5, the highest value of modularity among all the other networks is possessed by the *Weekend* network (see Figure 7).

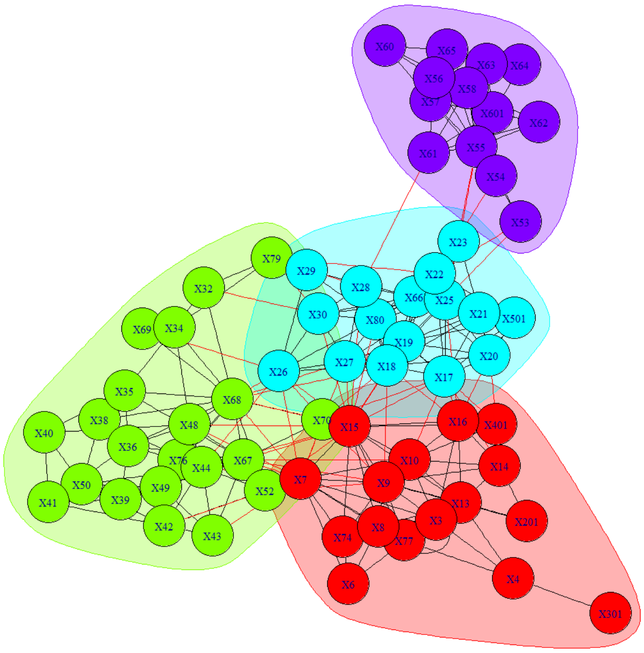

Figure 7. Visual representation of Weekend network which has the highest modularity value among the other passengers flow networks, generated with [28]

In Figures 8 we show the visual representation of the lowest modularity which can be found at the *Men* network. Thereby one can observe that the nodes of the single modules are visually more mixed than in Figure 7.

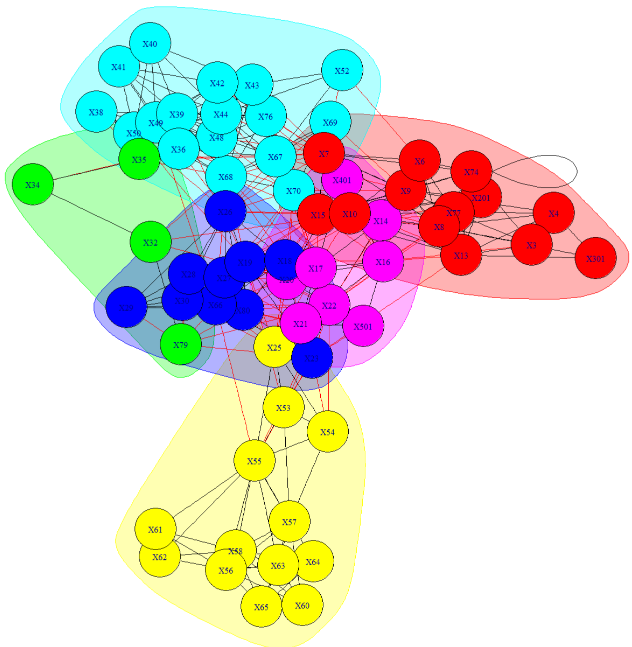

Figure 8. Visual representation of Men network which has the lowest modularity value among the other passengers flow networks, generated with [28]

In Table 5 we also observe that the highest difference between the modularity values of two complementary networks is between the *Business days* network, and the *Weekend* network. According to these observations, the passengers are more scattered in the train station during the business days. While during the weekend they tend to spend their time there more in a module (group) of nodes, and crossing less between the modules.

For the time of the day analysis, the passengers spend their time more in a modular sense in the mornings, and less grouped during the evenings. This means that the passengers tend to browse more in the station while waiting for their trains in the morning than in the afternoon when they just cross the station without too much stopping around.

Table 5. Networks modularity results.

N	V	E	$Q(N)$
Men	65	354	0.4067
Women	66	354	0.4478
Under 40	66	384	0.4269
Over 40	66	315	0.4395
Morning	66	385	0.4825
Afternoon	66	262	0.4534
Business days	66	389	0.4196
Weekend	66	278	0.5338

Notation: N – network; V – number of nodes; E – number of edges; $Q(N)$ – modularity.

In the group gender we find that women, in comparison with men, tend to be more organized in visiting the elements of the train station. For the age group we find comparable results, where passengers under 40 years tend to stay in the station more in an organized way than the ones over 40 years of age, but with a less meaningful difference.

For the networks with the ten percent filter (see Table 7) we observe that the modularity efficiency increases for all of them. Analyzing the groups, we find that the ratio is quite similar only in two groups: gender, and time of the day (compared with the original networks). In the other two groups, we observe that elder passengers, and the passengers during the weekend days, are crossing the platform more frequently. While in the case of younger passengers, and passengers traveling during the business days, they tend to browse more around their interest point in the train station.

E. Structural measures

For the available networks, we try to determine their structural complexity. This allows to understand the profound structure (e.g. the symmetry) of the networks, and how these can be analyzed in comparison with their complementary network.

The entropy of a network is defined by using a probability value for each node of the network. Entropy measures using the information-theoretic approach of Dehmer [31] are applied. The advantage of using this method is that the probability values are not determined for each subtracted module (partition), but for each node of the network. To quantify the structural information based on a given probability distribution, the local node functionals are used. They are defined as positive mappings [31].

Entropy j-spheres is an entropy measure based on information functional using the nodes' j-spheres. This measure captures the information structure of the complete neighborhood of each node of the network by taking into account the number of nodes in the available j-spheres around a node, as shown in [4].

Entropy centrality is an entropy measure based on information functional using node centrality. This measure captures the centrality properties of each node in the network, as shown in [4]. Furthermore, we interpret the connection with two classical topological descriptors, e.g., the Wiener index [32], and the Randić connectivity index [33].

Wiener index captures the structural branching in a network based on the shortest distance between nodes, as defined in [32]. Randić connectivity index captures the connectivity structural information in a network based on the degrees of the nodes, as defined in [33].

The first observation from Table 6, regarding the two entropy measures, is that these capture the structural information uniquely compared with the other classical measures.

Dehmer et al [35] found that the two entropy measures are used to capture the symmetric structures of networks. These have the property of changing their position (e.g. by rotation, reflection) and appear unchanged compared with their original state. According to Dehmer [31], the smaller the entropy value, the more symmetric the network is (and vice-versa). Therefore, the quantities can be useful when exploring the structural organization of the networks. Based on this and on the values of the two entropy measures from Table 6, we can conclude that the analyzed networks possess a very symmetric information structure. These values significantly change with the number of nodes of the network, e.g. the *Men* network and cases where networks are disconnected (see Table 7 for the ten percent filter networks).

Table 6. Networks structural measures results.

N	V	E	$I^{\lambda}_{f,lin^{V}}(N)$	$I^{\lambda}_{f,lin^{C}}(N)$	$W(N)$	$R(N)$
Men	65	354	6.0152	5.9744	5075	31.155
Women	66	354	6.033	5.9946	5396	31.7177
Under 40	66	384	6.0381	6.0043	5134	31.7727
Over 40	66	315	6.0366	5.9909	5707	31.577
Morning	66	385	6.0351	5.9957	5118	31.7577
Afternoon	66	262	6.0367	5.9826	6216	31.4252
Business days	66	389	6.0381	5.9992	5149	31.7619
Weekend	66	278	6.0366	5.9973	6083	31.3041

Notation: N – network; V – number of nodes; E – number of edges; $I^{\lambda}_{f,lin^{V}}(N)$ – entropy measure based on information functional using the j-spheres; $I^{\lambda}_{f,lin^{C}}(N)$ – entropy measure based on information functional using vertex centrality; $W(N)$ – Wiener index; $R(N)$ –Randić's connectivity index.

The two entropy measures and the Wiener index are measures based on distances in a network. Even though, the values are highly uncorrelated. The Pearson's correlation coefficient between the entropy j-spheres and Wiener index is -0.09213692, and the Pearson's correlation coefficient between the entropy centrality and Wiener index is 0.176219.

Furthermore, the Wiener index captures the branching structural information. The higher the value, the more cyclic the network is. Analyzing the values of the initial networks from Table 6, we notice that in the afternoons and during the weekends, passengers tend to create more cycles in the networks. The same holds for women and passengers over 40 years old, but with a less meaningful difference compared with their complementary network.

For a further analysis, the ten percent filer networks from Table 7 can be considered. Here, the Wiener index values increase for these particular networks. The situation changes in the mornings, and during the business days, when the passengers tend to create more cycles in the networks compared with their initial status. The same holds for the *Men* network in comparison with the *Women* network from the gender group.

For the networks in the age group, the situation preserves. Passengers over 40 years old tend to create more cycles in the network than the ones under 40 years of age.

Another observation is made regarding the Randić's connectivity index measure, which has a minimal value for a path network, and a maximal value for a star network [33]. According to the values from Table 7, all the networks have a similar general shape and that is rather a star-shape than a path-shape.

Table 7. Quantitative network analysis results.

	$\boldsymbol{N}$	$\boldsymbol{V}$	$\boldsymbol{E}$	$\varnothing(\boldsymbol{N})$	$\boldsymbol{D_{s-l}(N)}$	$\boldsymbol{\lambda(N)}$	$\boldsymbol{Q(N)}$	$\boldsymbol{I^{\lambda}_{f,lin^{V}}(N)}$	$\boldsymbol{I^{\lambda}_{f,lin^{C}}(N)}$	$\boldsymbol{W(N)}$	$\boldsymbol{R(N)}$
Gender	Men	65	354	10	0.165	2.4399	0.4067	6.0152	5.9744	5075	31.155
	Women	66	354	14	0.1601	2.5156	0.4478	6.033	5.9946	5396	31.7177
	*Both genders	66	423	4	0.1913	2.2886	0.4164	6.0302	5.9948	4909	31.9803
	Men >=10%	65	158	103	0.0737	4.3149	0.5415	6.0117	5.9759	8975	31.5761
	Women >=10%	66	151	101	0.0683	4.062	0.6	6.0346	6.0013	8713	31.7889
	*Both genders >=10%	66	196	8	0.0886	3.6765	0.5365	6.0356	6.0065	7886	32.2132
Age	Under 40	66	384	20	0.1737	2.3935	0.4269	6.0381	6.0043	5134	31.7727
	Over 40	66	315	15	0.1425	2.6606	0.4395	6.0366	5.9909	5707	31.577
	*All ages	66	423	4	0.1913	2.2886	0.4164	6.0302	5.9948	4909	31.9803
	Under 40 >=10%	66	151	99	0.0683	4.0559	0.588	6.0322	5.9953	8700	32.0624
	Over 40 >=10%	66	155	64	0.0701	4.1953	0.561	6.0347	6.0053	8999	31.854
	*All ages >=10%	66	190	8	0.0859	3.6499	0.5863	6.0362	6.0094	7829	32.1437
Time of the day	Morning	66	385	14	0.1741	2.386	0.4825	6.0351	5.9957	5118	31.7577
	Afternoon	66	262	19	0.1185	2.8979	0.4534	6.0367	5.9826	6216	31.4252
	*Whole day	66	423	4	0.1913	2.2886	0.4164	6.0302	5.9948	4909	31.9803
	Morning >=10%	66	149	160	0.0674	4.4228	0.5918	6.0343	6.0082	9487	31.9407
	Afternoon >=10%	66	159	69	0.0719	3.9058	0.5267	5.9938	5.9612	7878	31.3635
	*Whole day >=10%	66	202	9	0.0914	3.696	0.5798	6.0344	6.004	7928	31.9093
Time of the week	Business days	66	389	14	0.1759	2.4005	0.4196	6.0381	5.9992	5149	31.7619
	Weekend	66	278	15	0.1257	2.8359	0.5338	6.0366	5.9973	6083	31.3041
	*Whole week	66	423	4	0.1913	2.2886	0.4164	6.0302	5.9948	4909	31.9803
	Business days >=10%	66	144	166	0.0651	4.5133	0.571	6.0337	6.005	9681	31.8836
	Weekend >=10%	66	157	50	0.071	3.8093	0.5513	6.0361	5.9918	8171	31.7555
	*Whole week >=10%	66	201	8	0.0909	3.3879	0.5894	6.0374	6.0012	7267	31.8938

Notation: $\boldsymbol{N}$ – network; $\boldsymbol{V}$ – number of nodes; $\boldsymbol{E}$ – number of edges; $\varnothing(\boldsymbol{N})$ **–** diameter; $\boldsymbol{D_{s-l}(N)}$ **–** density; $\boldsymbol{\lambda(N)}$ **–** average path length; $\boldsymbol{Q(N)}$ **–** modularity; $\boldsymbol{I^{\lambda}_{f,lin^{V}}(N)}$ **–** entropy measure based on information functional using the j-spheres; $\boldsymbol{I^{\lambda}_{f,lin^{C}}(N)}$ **–** entropy measure based on information functional using vertex centrality; $\boldsymbol{W(N)}$ – Wiener index; $\boldsymbol{R(N)}$ –Randić's connectivity index.

* Group networks: these networks are a result of summing up the two networks from each group. These networks consist of all the unique nodes and edges of the two networks to be summed. The role of the group networks is to highlight the difference between an individual network and its group network. E.g. to highlight the difference between the measure results of the *Morning* network in comparison with the *Whole day* network.

V. Conclusions

For the possible implementation of various security measures in public places such as train stations without influencing passengers walking paths, one needs to understand their walking patterns in detail. As providers of such critical infrastructures as well as politicians and passengers tend to prefer an open system, the placement of measures is a very critical issue. One needs to maximize the rate of screened passengers and so of threats as well as minimize the change of walking paths in order to avoid crowds.

The method of quantitative network analysis in this field is a novel approach and to the best of our knowledge it was not applied in previous analysis. In this novel analysis now at hand we found some very interesting and meaningful insights in the pattern of different groups of passengers based on real-life observation data. We identified for example that women tend to spend more time for browsing in the station than men do, but on the other hand their walking path is far more structured than the one of men. Furthermore, passengers seem to spend more time with browsing through the station in the mornings than they do in the afternoon. Also age influences the structure of the paths. Within the observed station the information counter and food stores were the most frequently visited spots – even more than train tracks or exits from entering passengers.

However, we are also facing several limitations in our study. We assume the patterns to change according to different station structures. In the analysis at hand we only consider a dead-end station. Therefore, more research is needed on the field, especially in terms of additional observations. Additionally, the observations focused on passengers entering the station from entries only. We did not consider people arriving by train. This might also influence the used dataset and results. However, for the planning of security guards or smart camera systems, the findings might be very helpful as over different times of the day and days of the week the behavior changes strongly.

In further research we will apply the findings directly in agent-based simulation models with adequately parameterized agents that follow the detected patterns in order to test for several security measures available for the considered station.

Acknowledgment

The support from the German Federal Ministry of Education and Research (BMBF) (project RiKoV, Grant No.13N12304) is gratefully acknowledged.

Research of Marian Sorin Nistor, was funded by the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme FP7/2007-2013/ under REA Grant Agreement Number 317382.

References

[1] M. Dehmer, S. Meyer-Nieberg, G. Mihelcic, S. W. Pickl, and M. Zsifkovits, “Collaborative risk management for national security and strategic foresight,” *EURO Journal on Decision Processes*, vol. 3, no. 3-4, pp. 305–337, 2015.

[2] S. Meyer-Nieberg, S. W. Pickl, and M. Zsifkovits, “Operations research for risk management in strategic foresight,” *GRF Davos Planet@Risk*, vol. 3, no. 2, pp. 281–288, 2015.

[3] M. Dehmer, A. Lechleuthner, O. A. Mudimu, and S. W. Pickl, “Exploring data analysis techniques for threat estimation,” in *Proceedings of the 10th Future Security*, B. Jürgen, Andreas Meissner, and G. Jürgen, Eds, 2015, pp. 233–237.

[4] M. Dehmer, M. S. Nistor, W. Schmitz, and K. A. Neubecker, “Aspects of Quantitative Analysis of Transportation Networks,” in *Proceedings of the 10th Future Security*, B. Jürgen, Andreas Meissner, and G. Jürgen, Eds, 2015, pp. 239–244.

[5] T. D. Raty, “Survey on contemporary remote surveillance systems for public safety,” *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no. 5, pp. 493–515, 2010.

[6] S. Meyer-Nieberg, M. Zsifkovits, S. W. Pickl, and F. Brauner, “Assessing Passenger Flows and Security Measure Implementations in Public Transportation Systems,” in *Proceedings of the 10th Future Security*, B. Jürgen, Andreas Meissner, and G. Jürgen, Eds, 2015, pp. 425–432.

[7] S. Meyer-Nieberg, M. Dehmer, H. Bracker, and Schneider B, “Assessing the Vulnerability of Dynamical Systems in Public Transportation,” in *Proceedings of the 9th Future Security*, 2014, pp. 247–252.

[8] S. Meyer-Nieberg, D. Hauschild, S. Luther, and M. Zsifkovits, “Simulation-Based Analyses for Critical Infrastructure Protection: Identifying Risks by using Data Farming,” in *Operations Research Proceedings 2015*, 2015.

[9] M. Valera and S. A. Velastin, “Intelligent distributed surveillance systems: a review,” in *Vision, Image and Signal Processing, IEE Proceedings-*, 2005, pp. 192–204.

[10] A. Sforza, S. Starita, and C. Sterle, “Optimal Location of Security Devices,” in *Railway Infrastructure Security*: Springer, 2015, pp. 171–196.

[11] S. Poledna, J. L. Molina-Borboa, S. Martínez-Jaramillo, M. van der Leij, and S. Thurner, “The multi-layer network nature of systemic risk and its implications for the costs of financial crises,” *arXiv preprint arXiv:1505.04276*, 2015.

[12] T. Kinattukara and B. Verma, “Multi-Clustered Neural Network for Improved Classification of Road Images,” *Journal of Information Assurance & Security*, vol. 10, no. 1, 2015.

[13] D. Helbing, D. Armbruster, A. S. Mikhailov, and E. Lefeber, “Information and material flows in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 1, pp. xi–xvi, 2006.

[14] D. Helbing, “Global Networks Must be Redesigned,” in *Thinking Ahead-Essays on Big Data, Digital Revolution, and Participatory Market Society*: Springer International Publishing, 2015, pp. 67–73.

[15] M. Graube, P. Ortiz, M. Carnerero, O. Lázaro, M. Uriarte, J. Pfeffer, and L. Urbas, “Flexibility vs. Security in Linked Enterprise Data Access Control Graphs,” *Journal of Information Assurance & Security*, vol. 9, no. 2, 2014.

[16] S. W. Pickl, "Personal Communication", Sep. 2010.

[17] Deutsche Bahn AG, *München Hauptbahnhof.* Available: http://www.bahnhof.de/bahnhof-de/Muenchen_Hbf.html (2015, Jun. 11).

[18] F. Harary, *Graph theory*: Addison-Wesley, Reading, MA.

[19] C. Berge, *The theory of graphs*: Courier Corporation, 2001.

[20] M. E. J. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[21] S. N. Dorogotsev and J. F. Mendes, *Evolution of Networks: From biological networks to the Internet and WWW*: Oxford University Press.

[22] P. Erdös and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hungar. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[23] M. Dehmer and F. Emmert-Streib, *Quantitative Graph Theory: Mathematical Foundations and Applications*: CRC press, 2014.

[24] Plotly Technologies Inc, *Collaborative data science.* Available: https://plot.ly.

[25] F. Emmert-Streib and M. Dehmer, “Exploring statistical and population aspects of network complexity,” (en), 2012.

[26] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*: Cambridge university press, 1994.

[27] D. B. West, “Introduction to Graph Theory,” (en), *Prentiss Hall, Upper Saddle River, NJ*, 1996.

[28] Gabor Csardi and Tamas Nepusz, “The igraph software package for complex network research,” (en), *InterJournal*, vol. Complex Systems, pp. 1–9, http://igraph.org, 2006.

[29] L. O. James, R. G. Stanton, and D. D. Cowan, “Graph decomposition for undirected graphs,” in *Proceedings of the Third Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ, Boca Raton, Fla, 1972)*, 1972, pp. 281–290.

[30] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” (en), *Physical review E*, vol. 70, no. 6, p. 66111, 2004.

[31] M. Dehmer, “Information processing in complex networks: Graph entropy and information functionals,” (en), *Applied Mathematics and Computation*, vol. 201, no. 1, pp. 82–94, 2008.

[32] H. Wiener, “Structural determination of paraffin boiling points,” (en), *Journal of the American Chemical Society*, vol. 69, no. 1, pp. 17–20, 1947.

[33] X. Li and Y. Shi, “A survey on the Randic index,” (en), *MATCH Commun. Math. Comput. Chem*, vol. 59, no. 1, pp. 127–156, 2008.

[34] R Development Core Team, *R: A Language and Environment for Statistical Computing* (en). Vienna, Austria. Available: http://www.R-project.org.

[35] M. Dehmer, K. Varmuza, S. Borgert, and F. Emmert-Streib, “On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures,” (en), *Journal of chemical information and modeling*, vol. 49, no. 7, pp. 1655–1663, 2009.