Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Improving the connectivity of a pedestrian network through Euclidean and shortest path distances comparison

Théophile Emmanouilidis
Institut de Géographie, Université de Lausanne
Dorigny – Anthropole CH-1015
Lausanne, Switzerland
theophile.emmanouilidis@unil.ch

Abstract

This paper analyses the structure of a large pedestrian network in Lausanne (Switzerland), namely by comparing Euclidean and shortest path distances. A centrality and a sinuosity indexes are computed to run spatial analysis at various scales. Such a calculation produces different visualisations that can help to analyse and characterize the connectivity of buildings inside a district, but also between districts. The method also suggests that the comparison between Euclidean and shortest path distances can be used to improve any pedestrian network by creating promising shortcuts, or alternatively to detect major network topological errors.
Keywords: Spatial Analysis, Mobility of Persons, Trajectories Analysis.

1 Introduction

In 2008, the city of Lausanne (135’000 inhabitants, 16km2) promoted a GIS project on schools and transportation. The goal was to visualize the location of all the schoolchildren and to compute their shortest path distance to school. Consequently; a public transportation subsidy program was developed based on the obtained results. The idea of the project was to promote walking and use of public transport instead of using private one. The datasets created during this project will be used in this paper and are property of the “Service du cadastre de Lausanne” [1].

Even if network analysis algorithms, such as routing, have been implemented for many years in GIS, researchers are still working on developing new routing algorithms [2] or optimizing them for large-scale networks [3]. The resulting network analysis tools have been used and customized by researchers to study pedestrian network structure as well as pedestrians behaviour. Based on the comparison of Euclidean and shortest path distances, the first goal of this paper is to provide a method that evaluates the connectivity of a pedestrian network. The second goal is to improve the network connectivity by detecting potential shortcut locations.

2 Datasets

2.1 The full street network (G_f)

From now on we will consider this network (fig.1) as a graph $G_f = (V_f, E_f)$ where V_f is the set of n nodes, made out of road intersections, dead ends and building locations (i.e habitation, administrative, commercial and industrial constructions), and E_f is the set of m edges connecting these elements together.

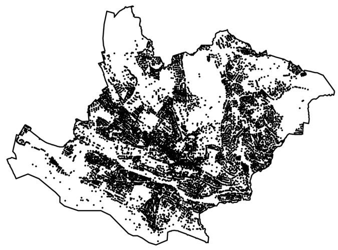
Figure 1: Pedestrian network of Lausanne

G	
n: Number of nodes	27'219
m: Number of edges	31'482
l: Total length (km)	734

2.2 Address points dataset (A)

The address points dataset (A) contains all the building locations of the town. White zones (fig.2) mainly depict recreation areas such as lakeside (bottom left) or forests (upper part of the city) and have very few constructions.

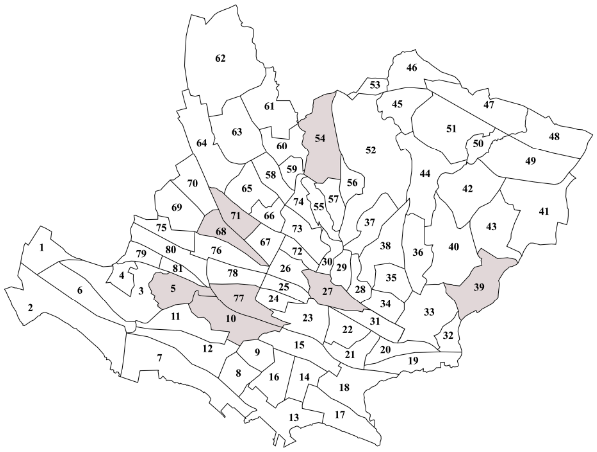
Figure 2: Building locations



2.3 Official district areas

The city is statistically divided into k=1,…,81 district areas (fig.3) Z_k covering areas of varying surface area S_k=$S(Z_k)$ and counting $n(A_k)$=$n(A(Z_k))$ address points.

Figure 3: Official statistical districts (k =1,…,81)



This paper will be mainly illustrated by the greyed districts described in the following table.

Table 1: characteristics of the main districts

Name	k	$n(A_k)$	S_k (km^2)
“Montoie”	5	195	0.26
“Marc-Dufour”	10	293	0.28
“Rue Centrale"	27	441	0.22
“Plaisance”	39	255	0.29
“Bellevaux”	54	288	0.36
“Av. d'Echallens”	68	177	0.15
“Montétan”	71	161	0.14
“Tivoli”	77	121	0.20

3 Methodology

Our method is based on the comparison between shortest path distances[1] (d_{ij}) and Euclidean distances (D_{ij}) between pairs of address points. This is done in ArcGIS by creating an OD cost matrix. The distances are symmetric, with a null diagonal, and metric; that is, $d_{ij} = d_{ji}$; $d_{ii}=0$; $d_{ik}+d_{kj}>=d_{ij}$ *and similarly for* D_{ij}. With these data, two indexes will be computed at the scale of each address point (local analysis) and at the scale of each district (regional analysis):

Centrality index:

Local: $$\bar{d}_i(Z_k) := \frac{1}{n(A_k)} \sum_{j\in Z_k} d_{ij}$$

The centrality index is defined as the average shortest-path distance from a place i to all the others locations inside the district Z_k. This index constitutes a local value and can easily be mapped as graduated colours points.

Regional values will allow us to compare the districts. They result from the aggregation of local values. The coefficient $1/\sqrt{S_k}$ reduces the sensitivity of the index to the district's size but not to its shape.

Regional: $$\bar{d}(Z_k) := \frac{1}{n(A_k)\sqrt{S_k}} \sum_{i\in Z_k} \bar{d}_i$$

Sinuosity index:

$$\delta_{ij} := \frac{d_{ij}-D_{ij}}{D_{ij}}$$

By construction, $\delta_{ij} \geq 0$. Note that this index is an alternative to the one commonly used in hydrology defined as: $\delta_{ij} = d_{ij}/D_{ij}$[5].

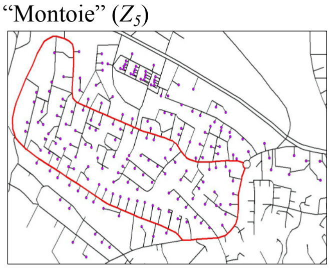
Local: $$\bar{\delta}_i(Z_k) := \frac{1}{n(A_k)} \sum_{j\in Z_k} \delta_{ij}$$

Regional: $$\bar{\delta}(Z_k) := \frac{1}{n(A_k)} \sum_{i\in Z_k} \bar{\delta}_i$$

3.1 Visualisation and interpertation of the indexes in ArcGIS. The case of the “Montoie” district

The pedestrian network of “Montoie” (fig.4) has the particularity of having one main street that loops inside the district (shown in red). The district has 195 address points, creating 195*(195-1)/2=18’915 pairwise pedestrian distances.

Figure 4: pedestrian network of “Montoie” (Z_5)



3.1.1 $d_{ij}(Zk)$ and $\delta_{ij}(Z_k)$

In ArcGIS d_{ij} appears as straight lines that are linking all the address points of a district. This representation becomes quickly unreadable if all lines are displayed. For this reason it’s more convenient to display only lines originated from one point only. For example, we can choose the most central point of the “Montoie” district that is *argmin*($\bar{d}_i$) (fig.5), the address point *i* minimizing $\bar{d}_i$, or the least central location (*argmax*($\bar{d}_i$); fig.6). This way we can visualize d_{ij} values (from green to red) through straight lines. In fig.6, we can clearly see that *argmax*($\bar{d}_i$) (the church of the district!) has its lowest d_{ij} values oriented on its left and right (green lines). Its highest values point towards the north of the district (red lines) and it suggests that the connectivity between the lower and upper part of the district could be improved by adding new pedestrian paths. The arc (in purple) has a radius of 380m and is centered on the *argmax*($\bar{d}_i$) point. We can notice that the values of d_{ij} still vary a lot around this Euclidean distance. The *argmin*($\bar{d}_i$) (fig.5) can be considered as the most central address point. It is located at the entrance of the district and does not match the centroid of the district’s shape shown as a black dot; by contrast, this centroid corresponds by a few meters to the location of *argmin*($\bar{D}_i$) (not shown here). As the values illustrated in fig.5 and fig.6 have their own data classification (natural breaks (Jenks), 5 classes) they

[1] d_{ij} is calculated through the Dijkstra algorithm [4] implemented in ArcGIS.

ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

cannot be compared directly. However, the variability of the coloured lines is more important in fig.6.

Figure 5: d_{ij} of $argmin(\overline{d_i})$ (= 344m)

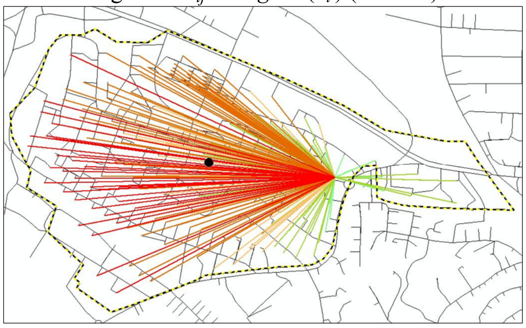
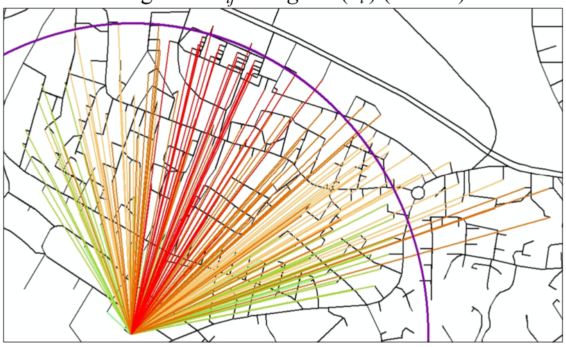


Figure 6: d_{ij} of $argmax(\overline{d_i})$ (=608m)



The same representation can be performed with the sinuosity index δ_{ij}. As previously, we choose to display all δ_{ij} from the two locations i solutions of the problems $argmin(\overline{\delta_i})$ (fig.7) and $argmax(\overline{\delta_i})$ (fig.8). Note that $argmin(\overline{d_i}) \neq argmin(\overline{\delta_i})$ and $argmax(\overline{d_i}) \neq argmax(\overline{\delta_i})$. The two patterns show that δ_{ij} values are grouped in different directions. $argmax(\overline{\delta_i})$ is the deepest dead-end of the district and its δ_{ij} values point toward its only exit to the rest of the neighborhood. The south can only be accessed after walking through sinuous paths. $argmin(\overline{\delta_i})$ is located at the entrance of the district. The pattern indicates that the buildings of the upper part can be accessed through almost straight lines. The south is accessible through a more sinuous path (like a "S") that generates higher values of δ_{ij}. The concentric circles have a radius of 100 and 200 meters. They can provide a quick visual interpretation of $\overline{\delta_i}$, especially when the network of the study area has only one or two main roads and few shortcuts. In fig.7, the first 100m are reached almost in straight lines (main streets drawn as purple lines) which is not the case in fig.8. Within a range of 200m, the shape of the main streets is pretty different. In fig.8, we can notice that most of the main streets remain within this range. Walking in such a network means that the shortest path distance will increase while the Euclidean distance remains the same or even decreases. Therefore the value of $\overline{\delta_i}$ increases.

Figure 7: δ_{ij} of $argmin(\overline{\delta_i})$(= 0.27)

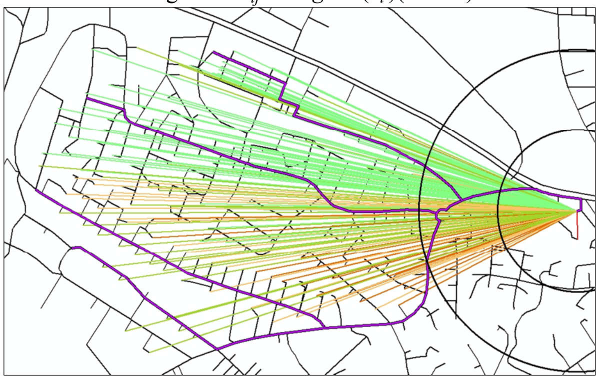
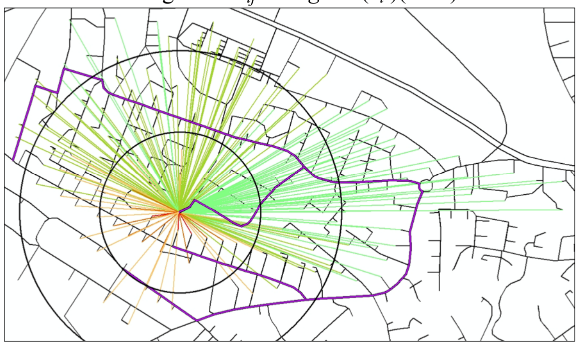


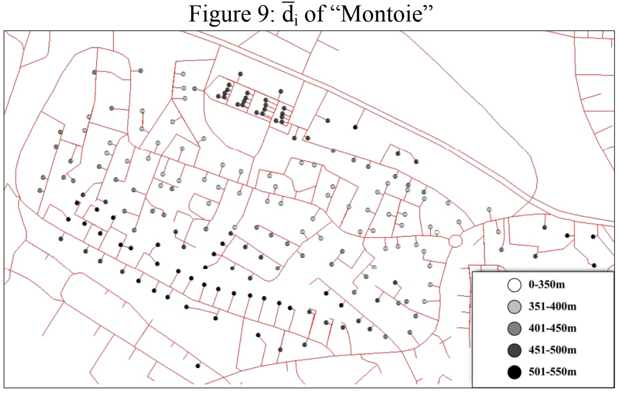
Figure 8: δ_{ij} of $argmax(\overline{\delta_i})$(=3.2)



3.1.2 $\overline{d_i}(Z_k)$ and $\overline{\delta_i}(Z_k)$

The lower part of the district contains almost all the highest values (in black) of $\overline{d_i}$ (fig.9) and $\overline{\delta_i}$ (fig.10). Such a configuration tends to confirm a lack of connectivity between the lower and the upper part of the district. The lower ones of $\overline{d_i}$ are distributed along a street. We can consider this street as being central because, for many buildings, it constitutes the only way to access the others address points. Note that the mapping of $\overline{\delta_i}$ (fig.10) detects dead-ends or clusters of points that cannot straightforwardly access bigger clusters.

Figure 9: $\overline{d}_i$ of "Montoie"



Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

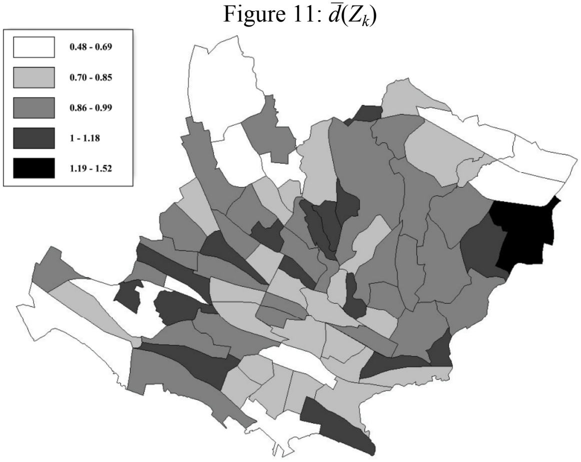
Figure 10: $\bar{\delta}_i$ of *Montoie*

3.1.3 $\bar{d}(Z_k)$ and $\bar{\delta}(Z_k)$

Let us now examine the distribution of the within-distances $\bar{d}(Z_k)$ (fig.11) and "within-sinuosities" $\bar{\delta}(Z_k)$ (fig.12) across the districts k=1,…81.. Those quantities are sensitive to the morphology and to the size of the district.

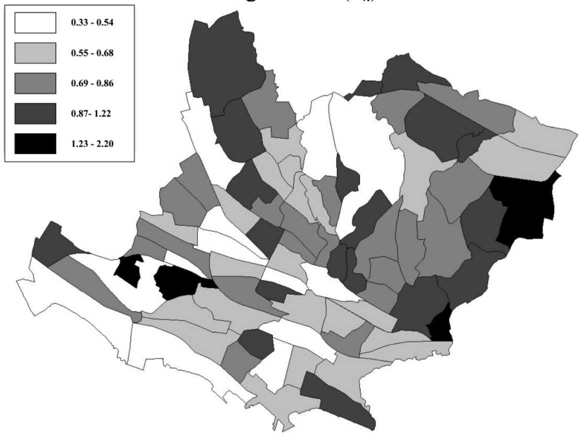
Low $\bar{d}(Z_k)$ values means that the buildings within a district are better connected than the ones located in districts with high $\bar{d}(Z_k)$ values. In fig.11, we can see that the lowest values are mainly located along the city border. The centre and the southern part of the city have low value while they tend to be similar in the eastern part of the city. The district's shape of the west are more elongated and have pretty different values of $\bar{d}(Z_k)$.

Figure 11: $\bar{d}(Z_k)$



In fig.12, the low values of $\bar{\delta}(Z_k)$ correspond to districts where buildings tend to be linked together as straight lines. On the other hand, high values of $\bar{\delta}(Z_k)$ betray a more sinuous network identifying possible candidates for the construction of new pedestrian shortcuts. Note that we have chosen to illustrate the district of "Montoie" as it has high values of $\bar{d}(Z_k)$ and $\bar{\delta}(Z_k)$.

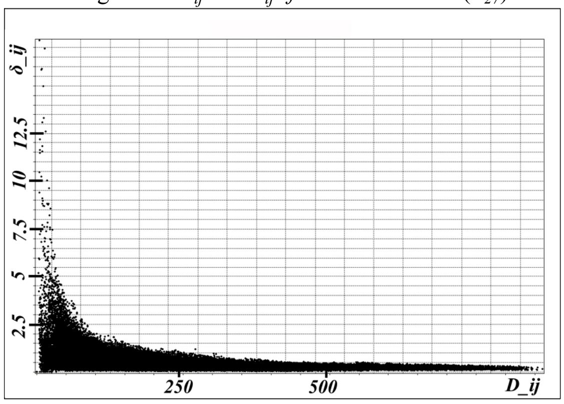
Figure 12: $\bar{\delta}(Z_k)$



3.2 Scatterplots of $d_{ij}(Z_k)$ and $\delta_{ij}(Z_k)$

For all the districts, the scatterplots of δ_{ij} and D_{ij} show similar patterns: values of δ_{ij} decreases quickly as D_{ij} increases (fig.13).

Figure 13: δ_{ij} and D_{ij} *of "Rue Centrale"* (Z_{27})

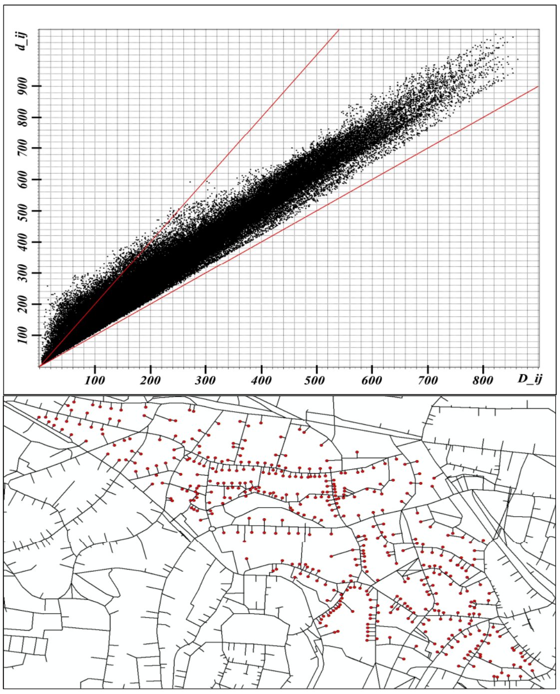


Scatterplots of D_{ij} and d_{ij} (further referred to as "SDd") return various and interesting patterns. The pattern would show linearity if $d_{ij} \approx D_{ij}$, meaning that the buildings are well connected. On the other hand, a more important scattering occurs when there is a lack of connectivity in the district, creating non-linear pattern. For a better visualization, two red lines representing $d_{ij} = D_{ij}$ and $d_{ij} = 2D_{ij}$ have been added on each scatterplot.

In our dataset of 81 districts, we can say that about 50% of the SDd show a "clean" linear pattern as shown in fig.14.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

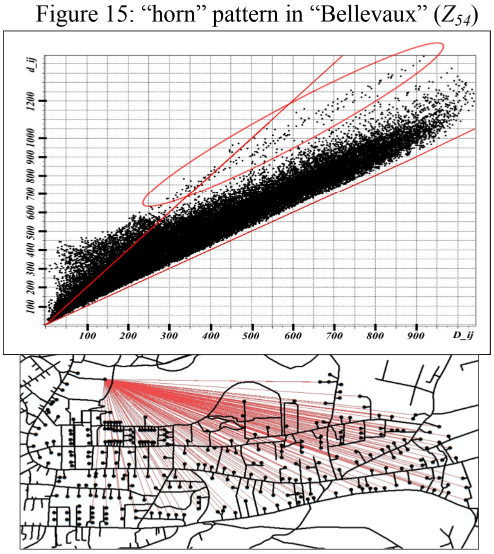
Figure 14: *SDd and network of "Rue Centrale"* (Z_{27})



For this specific case, a linear regression has been computed obtaining the expected values (in meters): $d_{ij}^* = 1.17D_{ij} + 46.3$. This means that pedestrian shortest-path distance between (far enough) different locations are on average a multiple of their Euclidean distance, with coefficient 1.17. (r^2=.96 , n=194'040 pairs of points, p=0.000).

A "horn" can appear over a main linear pattern (fig.15). This pattern is often created by an isolated point of the network. The points within the selected area correspond to the distances (D_{ij} and d_{ij}) between pairs of buildings but all originated from the same point (fig.15 bottom).

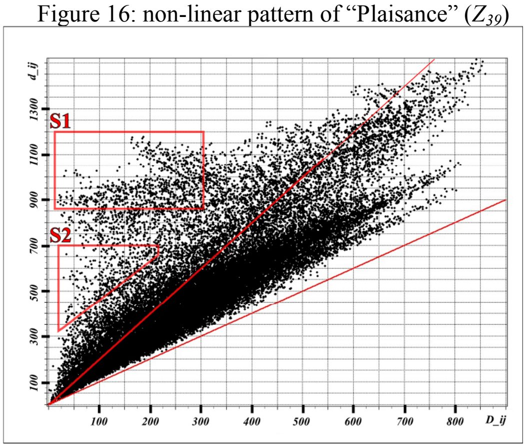
Figure 15: "horn" pattern in "Bellevaux" (Z_{54})



3.3 Detecting potential shortcuts "areas"

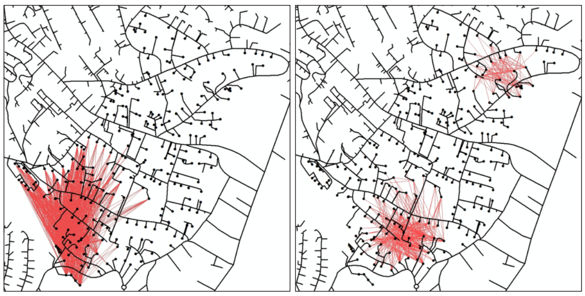
More diffuse or non-linear patterns require deeper analysis. This can be illustrated through the case of the "Plaisance" district (fig.16).

Figure 16: non-linear pattern of "Plaisance" (Z_{39})



As the graphs are interactive in ArcMAP, it is possible to select points of the scatterplot to show their corresponding geographic features on the map. In our case, one point corresponds to a line linking two buildings. For example, the points that are inside the selected area "S1" of fig.16 involve links between buildings located in the southwest of the district (fig.17 left), while "S2" concerns links between buildings located in the south and in the northeast (fig.17 right). These "areas" are good candidates for new pedestrian shortcuts.

Figure 17: Possible locations for new shortcut(s) (Z_{39})



All the SDd showing an important dispersion denotes a lack of connectivity between buildings. The most interesting cases are the ones having low value of D_{ij} and high value of d_{ij}. Please note that this lack of connectivity can relate the actual situation, but can also result from topological errors of the network (i.e. an existing path that hasn't been drawn in the GIS). This means that this scatterplot analysis can help in determining shortcuts locations, as well as detecting major network topological errors.

3.4 Creating shortcuts

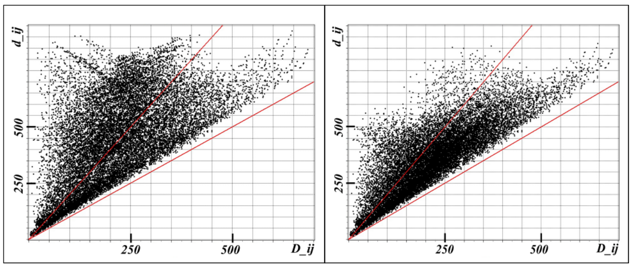
Based on the previous analysis of the district of "Montoie", we have created 3 small shortcuts linking the lower and upper part of the district (fig.18).

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 18: new shortcuts in “Montoie”

The impact of the shortcuts on the network connectivity can easily be noticed by comparing the *SDd* without (fig.19 left) and with (fig.19 right) shortcuts. In the first case, the “*V*” pattern confirms the poor network connectivity. With the shortcuts this “*V*” pattern is collapsed (fig.19 right): many points are grouped under the $d_{ij} = 2D_{ij}$ line. Numerically, the shortcuts lower the value of $\bar{d}(Z_5)$ from 442m to 360m (-22.7%) and $\bar{\delta}(Z_5)$ from 1.39 to 0.9 (-54%).

Figure 19: SDd of “Montoie” with and without shortcuts



3.5 Connection between two districts

The same analysis can be used to evaluate the connectivity between two adjacent districts Z_k and Z'_k. The distances D_{ij} and d_{ij} are calculated with $i \in Z_k$ and $j \in Z'_k$. Fig.20 (left) and fig.21 (left) illustrate two pairs of districts. In the first one, Z_{68} and Z_{78} are linked together through multiple paths. In the second one, Z_{10} and Z_{77} are separated by railways that limit the number of connections between them. The resulting *SDd* confirm that first case has a good connectivity (linear pattern) while the diffuse pattern of the second case betray a bad connectivity between Z_{10} and Z_{77}. A deeper analysis of this *SDd*, the computation and the mapping of $\bar{d}_i$ and $\bar{\delta}_i$ would help in identifying new shortcut(s) location(s).

Figure 20: network of $Z_{68} \cup Z_{78}$ and corresponding SDd

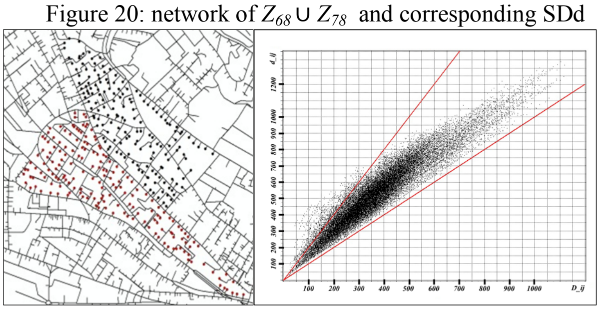
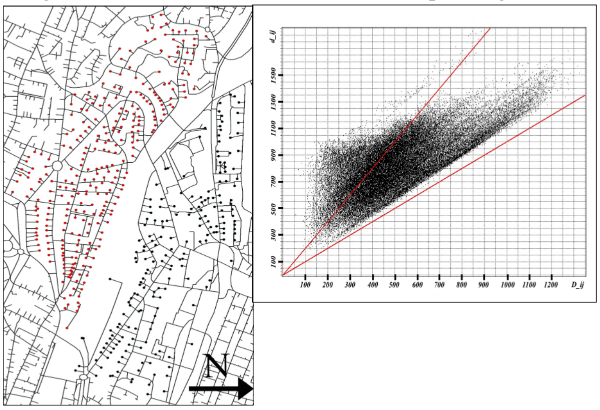


Figure 21: network of $Z_{10} \cup Z_{77}$ and corresponding SDd



4 Conclusion and work in progress

We believe that the exposed method can help in evaluating and improving an existing or a planned new pedestrian network, as well as detecting major topological network errors. The scatterplot approach can provide a pertinent overview of the network connectivity and can also suggest possible locations for new shortcuts.

Also we would like to improve our analyses by going beyond the current official districts. This can be done by aggregating the districts together, or by creating an area around the investigated location. Numerical optimization will have to be performed to overcome the problem of demanding computing time requirement, and of dealing with large matrices d_{ij} which can be reduced to a triangular form by symmetry. Our analyses are based on the building dataset (*A*), but same indexes can be computed using all (V_f) or subsets of network nodes. Weighted versions of the indexes are possible, for example by using the number of inhabitants living in each building as weight.

References

[1] Data source: “Service du Cadastre de la ville de Lausanne”, *http://www.lausanne.ch/cadastre*, 2011.

[2] B.Jiang, Xintao Liu. Computing the Fewest-turn Map Directions based on the Connectivity of Natural Roads. *International Journal of Geographical Information Science, 25(7):1069-1082,* July 2011.

[3] P. Sanders and D. Schultes. Highway hierarchies hasten exact shortest path queries. In Brodal G.S. and Leonardi S. (eds.), *Proceedings of the 13th Annual European Symposium, 568–579,* Palma de Mallorca, 2005.

[4] Dijkstra, E. W. A note on two problems in connexion with graphs. In : *Numerische Mathematik 1: 269–271,*1959.

[5] Leopold, Luna B., Wolman, M.G., and Miller, J.P. *Fluvial Processes in Geomorphology*. W.H. Freeman and Co., San Francisco, 1964.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke