

Nordic Journal of Surveying and Real Estate Research, Special Series, Vol. 4 (2009)

Received December 23, 2008

Accepted after revision June 25, 2009

Effect of Physical Constraints on Spatial Connectivity in Urban Areas

Sagi Filin, Yoav Vered

Mapping and Geo-Information Engineering

Technion - Israel Institute of Technology

Technion City, Haifa 32000, Israel

filin@technion.ac.il, veredyv@technion.ac.il

***Abstract.** Obstacle effect on proximity, connectivity, and organization of spatial data calls for derivation of measures that enable quantifying their influence. Provision of such measures is valuable for ensuring an aware planning, analysis of obstacle impact on spatial data, and the consequent placement of crossings. This paper proposes quantifying obstacle influence via their impact on connectivity and aggregation of data. As the paper shows, the derived indices enable capturing the actual obstacle effect on spatial data while accommodating datasets with different level of complexity. The information and contribution of these indices are demonstrated and analyzed, and results show how the derived measures reflect changes in spatial data arrangement.*

***Keywords:** Spatial data mining, clustering, obstacles, spatial obstruction analysis*

1 Introduction

Physical constraints affect organization and connectivity among spatial data, thus having a central role in shaping the way that space is perceived. Constraints can take a variety of forms ranging from large scale entities like rivers, freeways, or borders, to walls and fences, and temporal forms like traffic accidents or congested roads that alter spatial connectivity within designated networks. As an encounter with obstacles is common, and their presence dictates choices made, their influence must be an integral part of spatial data analysis practices. To demonstrate the influence of physical constraints, consider the data depicted in Figure 1. In its elementary form (Figure 1a), no apparent order is noticeable, and only a careful inspection

may lead to the pointset separation into two groups. This arrangement is altered, however, when an obstacle is introduced into the data (Figure 1b). In the altered arrangement, the pointset is separated into two groups, one on each side of the obstacle. A different placement (Figure 1c) yields a different partition, where all but one of the points is joined into one group. The isolated object to the left of the obstacle now appears detached and unrelated. These variations in arrangement indicate that the presence of obstacles influences connectivity, dictates data aggregation, and thus affects interpretation of space.

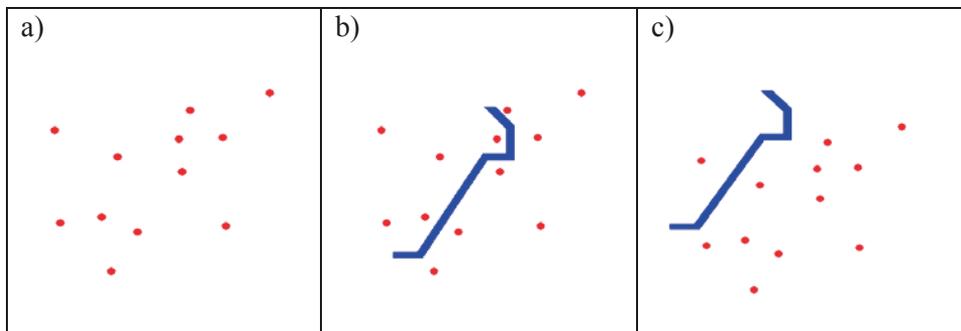


Figure 1. Effect of different configurations of entities and obstacles, a) original pointset; b) same pointset with an obstacle crossing it – the aggregation of points has changed – four points on the lower-left part and the rest towards the upper right; c) same pointset with an obstacle, the influence of the obstacle is limited

Integration of constraints into spatial data analysis refers to a variety of processes, ranging from data aggregation and geometric pattern detection to data interpretation, querying, and prediction. However, a fundamental question arises relating to measuring the actual influence of obstacles on spatial data. Returning to Figure 1, the three setups can be easily ranked according to the obstacle influence, with set 'a' having the highest connectivity level; set 'c' follows as the obstacle affects only one entity and leaves all others uninfluenced; and finally set 'b' is the most affected, as the obstacle partitions the data into two disjoint groups. Such ranking is guided by the alteration in connectivity and aggregation among points; however it is qualitative in nature, and cannot be generalized into more complex scenarios. When the number of objects and obstacles increases, ranking their significance and evaluating their effect becomes a challenging task, one that cannot be left to intuition only. The question that arises is which measures yield such observations, and how they can be formulated into quantitative metrics which can become applicable to general, more complicated, datasets.

Research related to obstacle integration into spatial data analysis has focused mostly on geometrical interpretation (via construction of visibility graphs), and on obstacle effect on clustering of spatial data both in terms of data description and analysis (e.g., Miller and Han, 2001). The influence of obstacles on clustering, known as clustering with obstructed distances (COD) problem, concerns the integration of the distance influence on the data aggregation. Only a few algorithms that handle this problem have been reported, with existing ones relating to three categories: partitioning based, hierarchical based, and density based. Tung et al. (2001) propose a partitioning K-medoids-based approach that uses "obstructed distances" to find the shortest paths on a constructed visibility graph. Estivil-Castro et al. (2000) present a graph partitioning approach that is based on clustering points through the dual relation between the Delaunay triangulation and Voronoi diagram. Zaiane and Lee (2002) propose an extension of DBSCAN (Ester et al., 1996), where clustering is performed by gradual construction of regions based on density measures. Wang et al. (2004), propose a density based approach that extends an obstacle-free version (Wang and Hamilton, 2003) to support obstacles and facilitators (crossings). In reference to querying spatial data in the presence of obstacles, Zhang et al. (2005), discuss geometric implications of obstacles on such procedures.

Most research has focused on how spatial data "responds" to the presence of obstacles in terms of data analysis, but not to the more fundamental question of how data is affected by their presence. The different manners by which obstacles influence spatial data motivate study into the evaluation of their overall effect. This paper concerns derivation of indices for ranking influence of obstacles on spatial datasets, with two measures standing out – obstacle influence on distances and on data aggregation. The paper first analyzes the effect obstacles have on spatial connectivity and the type of measures that describe them, followed by derivation of measures for distance and aggregation obstruction, and ending in a demonstration of the derived measures and analysis of the results. The application is demonstrated on a dataset consisting of neighborhoods in a large metropolitan area and physical constraints among them.

2 The effect of Constraints on spatial data

Constraints on spatial data can be introduced in various forms, including: object constraints – which are provided in attributal terms and differentiate spatial objects into groups among which connectivity is of no concern or has a reduced importance (e.g., residents of different countries), physical constraints – where physical objects (e.g., streams, highways, walls, and fences) interfere with connectivity among objects, and aggregation

constraints – which can act by limiting the number of clusters or by imposing attributal constraints on the individual objects within the clusters (Tung et al., 2001). Our focus in this study is on physical constraints and their effect on spatial data. The other noted forms are either relevant to the extraction of a pointset or can be regarded as a subclass of the problem addressed in which constraints on the connectivity are imposed.

The studied problem can be stated as follows, given a set of n objects $P = \{p_1, p_2, \dots, p_n\}$ with $p_i = \{x_i, y_i\}$, and a set of m obstacles $O = \{o_1, o_2, \dots, o_n\}$, whose outline is modeled as a non-intersecting polygon, measure the influence of O on the level of spatial connectivity of the data. Measuring obstacles interference can be applied in three different levels:

1. Effect of an individual obstacle on the connectivity.
2. Effect of a subset of obstacles ($O_a \subset O$) on the connectivity.
3. Effect of O as a whole on the connectivity.

which are associated with the following spatial queries: i) measure the actual effect of individual obstacles on the level of spatial connectivity (analysis of individual obstacles), ii) find the subset of obstacles with the biggest effect on connectivity (evaluation of a subset), and iii) measure spatial connectivity in light of existing obstacles (evaluation of O as a whole).

The actual obstacle effect on the data can be approached in several ways, with key indicators including intrinsic obstacle properties, distance obstruction, and aggregation obstruction. Intrinsic obstacle properties refer to shape parameters, with smaller obstacles likely having a lesser effect on the data as compared to elongated ones or those covering a large area. However, as Figure 1 demonstrates, shape properties do not account for the obstacle placement and are insufficient measures for spatial connectivity. Obstacle influence on distances echoes variation in proximity among entities, and since obstructed distances grow compared to unobstructed ones, reflects influence on connectivity among objects. Obstacle effect on aggregation echoes variation in data arrangement, which may result in an increase in the number of clusters or a change in entity association among clusters. The latter two influence types are consequential in nature, reflecting alteration caused by the obstruction presence (as a function of both shape and placement), and are more appropriate than those derived from the obstacle properties. We consider, therefore, the obstacle influence as reflecting change in the spatial arrangement and connectivity of the data.

3 Obstruction measures

The derived obstruction measures are guided by the following requirements, i) measurement of the level of interference should be comparative and

evaluate two states – unobstructed and obstructed by the analyzed obstacle; *ii*) measurements that apply to an obstructed environment should apply to an obstacle-free environment; *iii*) obstacles cannot improve the level of spatial connectivity among objects; and finally, *iv*) other than degenerate cases they should worsen it.

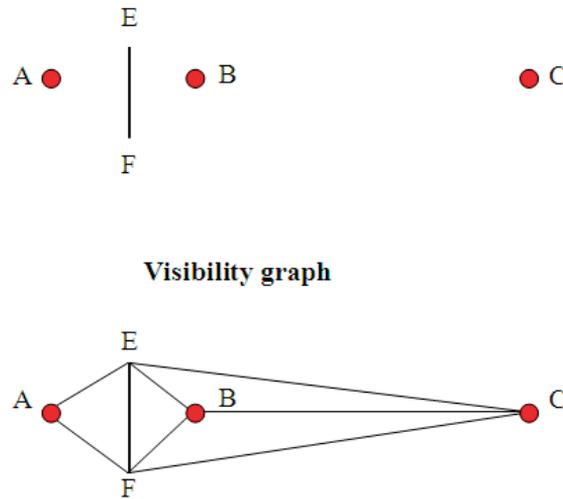


Figure 2. Obstacle effect on distance; in the obstacle-free environment (top) the distances between points A, B, and C are derived from differencing their ordinate values. In the presence of an obstacle, paths must refer to available crossings (given here by the obstacle end-points) and distances are derived from the visibility graph

3.1 Distance obstruction

Physical constraints increase distance among objects. In an obstacle-free environment, the traveling distance can either be measured by the Euclidean distance linking points or by the shortest path within a network. Focusing on Euclidean measures in an obstacle-free environment, one realization is that in the presence of obstacles, the complete-graph assumption (that facilitates the Euclidean measures) ceases to exist. Distances are measured using the visibility graph (de Berg et al., 2000) whose edges do not intersect the obstacle bounding polygon. Figure 2 illustrates this fact and shows that presence of an obstacle gives rise to more than one possible route linking entities. The distance among those entities now turns into a shortest path computation.

Distance obstruction measures

The measure proposed here is based on evaluating the matrix of distances

among all points within the dataset. We define $\mathbf{D}b$ as a distance matrix in an obstacle-free environment, with $\mathbf{D}b_{ij} = dfb(p_i, p_j)$, the distance function between points i and j ; and $\mathbf{D}a$ as a distance matrix between objects in the obstructed case, with $\mathbf{D}a_{ij} = dfa(p_i, p_j)$, the distance measure. Values for df can be either computed via a simple Euclidean distance, or the shortest path on a graph.

$\mathbf{D}r$, the obstacle influence matrix is then defined as:

$$\mathbf{D}r_{ij} = \begin{cases} 0 & i = j \\ \frac{\mathbf{D}a_{ij} - \mathbf{D}b_{ij}}{\mathbf{D}a_{ij}} & i \neq j \end{cases} \quad (1)$$

which measures the influence an obstacle has on an increase in traveling distance. Equation (1) shows the value is determined as the ratio between an increase in traveling distance and actual distance that should be traveled in the presence of an obstacle. Values in $\mathbf{D}r$ range between 0 and 1, with 0 indicating no change, and 1 being reached when $\mathbf{D}a_{ij} \gg \mathbf{D}b_{ij}$ or when the obstacle blocks connection between points.

Generally, $\mathbf{D}r$ is symmetric ($\mathbf{D}r_{ij} = \mathbf{D}r_{ji}$), provided the path computation is commutative. $\mathbf{D}r_j = \sum_i \mathbf{D}r_{ji}$ provides the total obstruction measure for the presence of the obstacle on entity j , and

$$dr_j = \frac{\mathbf{D}r_j}{(n-1)} \quad (2)$$

provides the average obstruction measure for an object. The total distance obstruction on connectivity is measured as

$$DOI = \frac{2 \sum_{i=1}^n \sum_{j=i}^n \mathbf{D}r_{ij}}{(n-1)^2} \quad (3)$$

3.2 Clustering obstruction

Obstacles also affect data aggregation. Figure 3 demonstrates the different effects an obstacle can have on spatial organization. The original dataset (Figure 3a) consists of 12 objects arranged in more or less two clusters – one consisting of five entities and the other of seven. In Figure 3b, an oblong obstacle is positioned between the two clusters. The obstacle has no direct influence on data aggregation as it does not affect the two clusters. Figure 3c

offers a different setup whereby the obstacle crosses the two individual clusters and separates each into two (four, total). In its present location, it not

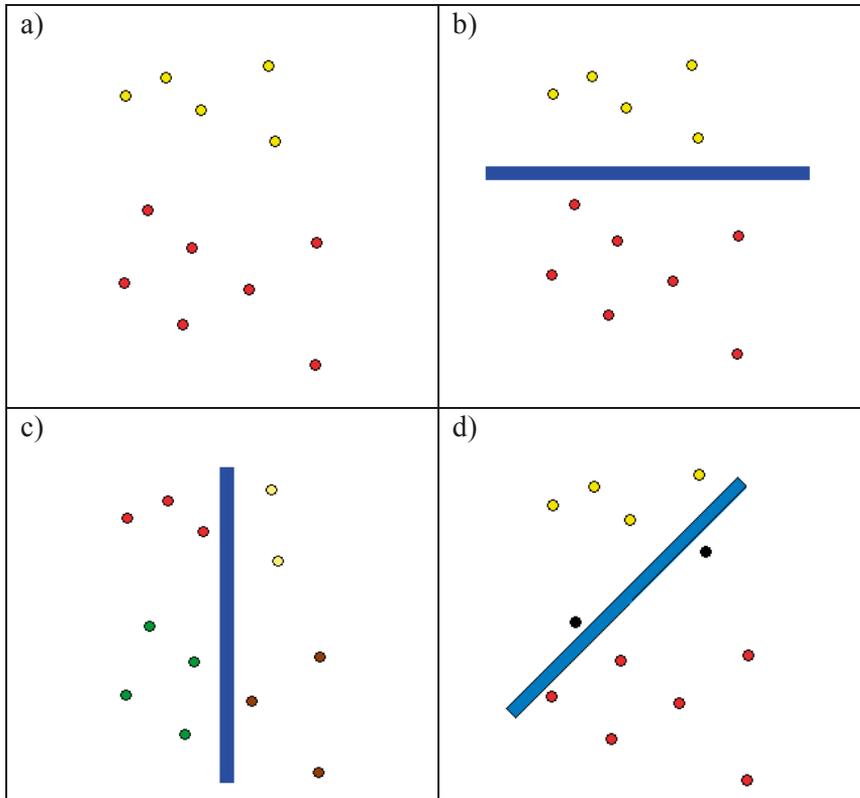


Figure 3. The effect of an obstacle on the aggregation, a) original dataset consisting of two clusters; b) an obstacle with no impact on the clustering, c) an obstacle breaking the two clusters into four, d) an obstacle placement where the two original clusters were maintained but the number of points in each has decreased

only partitions the data, but also increases the number of clusters. Finally, Figure 3d shows an obstacle that cuts out an individual object from each cluster. Here, the number of clusters does not change, but the two separated objects take the form of noise.

Aside from the practical concern of clustering with the presence of obstacles (see, Tung et al., 2001; Wang et al., 2004; Zaiane and Lee, 2002; Ester et al., 1996), all three cases raise the question of how to quantify obstacle influence on the results, including which measures better reflect change in aggregation, and how the number of clusters and cluster properties should be weighed in. They also raise the question of how noise should be

treated (namely, not incurring an artificial increase in the number of clusters), and if treated, how it can be compared to a noise-free clustering.

Clustering obstacle measures

A review of measures assessing quality of clustering results (e.g., Halkidi et al., 2002) shows that scatter and density of clusters are the main properties to quantify change in aggregation. Scatter parameters measure the level of separation among clusters (in which higher values indicate better aggregation), and density parameters measure compactness of the clustering result (with preference for compact clusters). These parameters are usually utilized for clustering optimization of datasets that were not altered, and therefore not applicable for the present case.

The proposed measure analyzes point scatter among the resulting/obstructed clusters compared to the original/unobstructed ones. The measure is based on the formation of an "association matrix", which counts mutual entities between a given cluster in the unobstructed and obstructed cases. The matrix size is defined by the number of original/unobstructed clusters (# of columns) and resulting/obstructed clusters (# of rows). Each value in the matrix records the number of mutual points associated with a cluster formed in the unobstructed case against the obstructed cluster. The COI measure is then based on selecting from each row the entry holding the largest number of mutual points, indicating how many instances from the original clusters were "unaffected" by the obstacle presence. The remaining points are considered ones scattered from the original cluster due to obstacle presence. By comparing the number of scattered points for all rows to the original number of points, a record of difference between clustering results (unobstructed and obstructed) is obtained. Strong deviation indicates stronger obstacle effect, while smaller deviation indicates preservation of the original structure.

More formally, let C be a matrix of an $m \times n$ size, where m is the number of original clusters, and n is the number of resulting clusters; c_{ij} records the number of mutual points to clusters i and j , $\sum_j c_{ij} = \#$ of points in the original cluster i , and $\sum_i c_{ij} = \#$ of points in the resulting cluster, j . Using C , the obstruction measure is computed as

$$COI = \frac{t - \sum_{i=1}^n \max(c_i)}{t} \quad (4)$$

with $\max(c_i)$, as the maximum value in row i , and t , the number of objects in the pointset. In essence, C describes the distribution (or scatter) of the entities within the obstructed clusters with respect to their original/unobstructed

clustering. If the clusters were not scattered, each row in the matrix would hold only one non-zero entry.

The proposed index generates a bounded, unitless, measure, ranging between 0 and 1. A 0 value is obtained when the obstacle does not influence data aggregation, and the measure approaches 1 for a maximal obstruction (when each entity forms an individual cluster). Referring to Figure 3, one can see that for set 'b', the COI measure is 0 (as the aggregation was not affected by the obstacle presence), for set 'c' the COI measure is $5/12$, implying the original clustering result was worsened by 40%, and finally for set 'd' the COI measure is $2/12$, relating to the two points split from their original clusters. Note that counting entity association enables considering the two points as noise and not as two newly formed individual clusters. Notice also that no distance measure is introduced to the COI but instead the analysis is performed with respect to the counts relating to the clustering results, thereby enabling representing the actual obstacle effect on data aggregation.

4 Analysis and discussion

Demonstrating the application of the proposed measure, the derived indices for the pointset in Figure 1 are analyzed. In terms of distance obstruction, indices are 5% and 15% for sets 'c' and 'b' respectively. These values reflect the realization that not all distances were affected by the obstacle (relating to distances between nearby points that were not affected by the obstruction), and that some affected points were relatively far from one another at the outset (explaining the increase of only 15%). Nonetheless, as an average increase of 15% in length to all distances between points, this increase can be regarded considerable. Sub-measures that help in creating the obstruction matrix (Equation (1)) also allow for classification of influences on individual points. For set 'c' they show that the most influenced point is the isolated one. For set 'b', they show that almost all points were influenced by the same order, thereby allowing understanding the actual impact of the obstacle on the point level.

Turning to analyze the obstacle influence on the aggregation of the data, Figure 4 shows the obstacle effect on clustering. Clustering was performed using AutoClust+ (Estivil Castro et al., 2000), a density-based method, which is based on graph partitioning of points into clusters through the dual relation between Delaunay triangulation and Voronoi diagram. For the obstructed case, arcs that intersect the obstacle are removed and the clustering is reevaluated.

As Figure 4 shows, the obstacle in set 'b' has little influence on the clustering, leaving only one point out of its original cluster, whereas the

obstacle in set 'c' partitioned the points into three clusters. Accordingly, the association matrices documenting this partitioning are

$$C_b = \begin{bmatrix} 7 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad (5)$$

$$C_c = \begin{bmatrix} 5 & 3 & 0 \\ 0 & 2 & 2 \end{bmatrix}$$

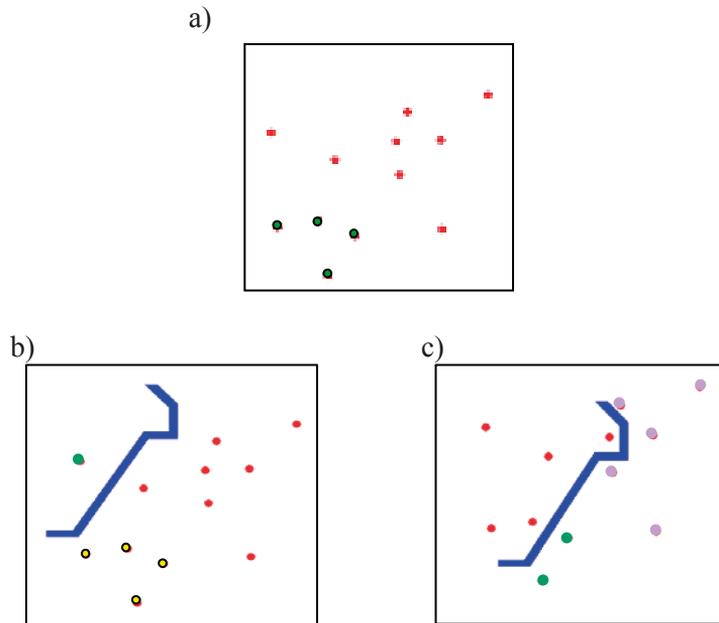


Figure 4. Effect of different obstacle placements on data aggregation

The obstruction indices for both cases are 0.08 (=1/12) and 0.42 (=5/12) for sets 'b' and 'c' respectively. The first index reflects the separation of only one point from its original association, while the second one reflects scatter of 42% of the points from their original cluster. This can also be visually seen in Figure 4, showing how the larger cluster has been split and points have changed association.

4.1 Application of the measures

The derived measures are now applied on a dataset consisting of 61 neighborhoods in the Tel-Aviv Metropolitan area (Figure 5). The different background shades designate three different municipalities in the

metropolitan area. Two streams cross the Tel-Aviv region – the Ayalon (flowing in a South-North direction) and the Yarkon (flowing from East to West). As the figure shows, both streams divide the metropolitan area into three sub-regions, and therefore have the effect of physical constraints on the connectivity among neighborhoods. We examine the effect of the obstacles on connectivity with contribution of bridges that cross the streams in alleviating the obstacles effect.

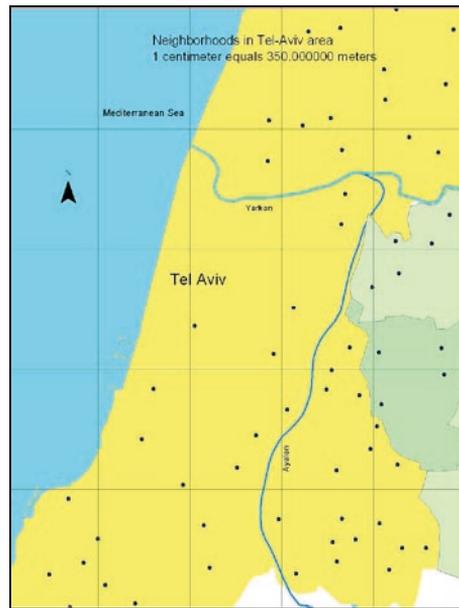


Figure 5. *Neighborhoods and obstacles in the Tel-Aviv Metropolitan area*

To evaluate the influence of obstacles on spatial connectivity, first examined is the connectivity in an obstacle-free environment, and then in the presence of obstacles. The evaluation here is of O as a whole.

4.2 Distance obstruction index

As an input to the distance obstruction evaluation, the coordinates of the 61 neighborhood centroids are given. The pair-wise distance matrix \mathbf{D}_b is computed based on Euclidean distances and the obstructed distance matrix \mathbf{D}_a is computed in terms of visibility graph distances (Figure 6). The general DOI measure is 0.10, which appears rather low, but considering the fact that points on the same side of the obstacle are not affected by its presence, one can realize that the actual obstacle effect on space is considerable. Notwithstanding, one can infer that the two streams together with crossings along them reduce connectivity by only 10 percent, which appears as an

acceptable value. The entity most affected by the obstacles (namely, have a maximal dr value) has a value of 0.17, whereas the least affected point has the measure of 0.02. The median value of the obstruction per point is 0.09 which is in agreement with the general DOI measure.

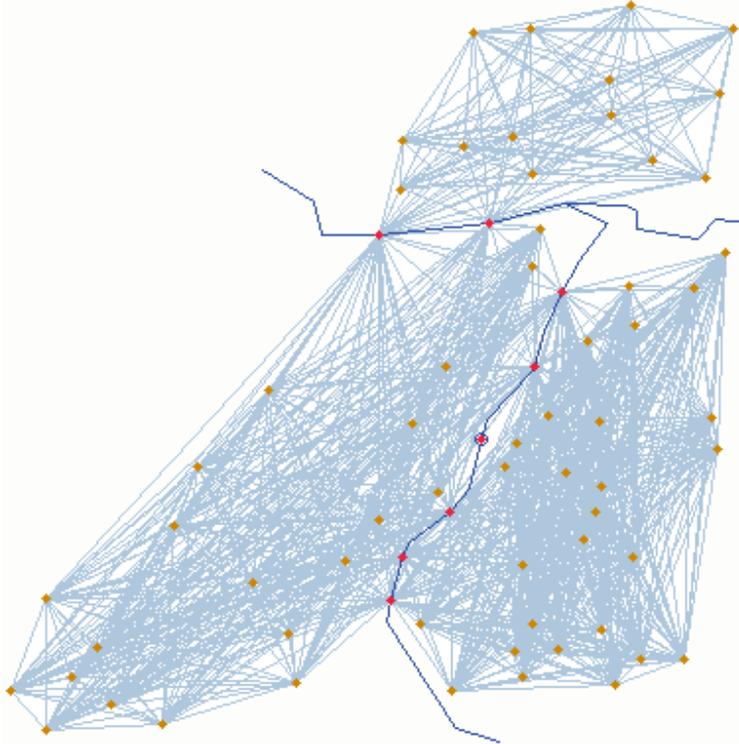


Figure 6. *Visibility graph for the neighborhoods in the presence of obstacles and crossings*

4.3 Cluster obstruction index

Clustering with no obstacles has led to generation of two clusters (Figure 7a), one consisting of 55 points and the other of six. As the clustering is density based, the algorithm follows the density of the neighborhoods, which is more or less even for the large cluster, thus explaining its spread. Results for clustering in the presence of obstacles are depicted in Figure 7b, showing six clusters. The results show that compared to the unobstructed case, the clusters have completely changed. Among the six, the neighborhoods that are to the right of the Aylon stream and those that are to the north of the Yarkon stream were aggregated into two clusters. The other four clusters are of neighborhoods that are bounded by the two streams. Among them two

clusters consist of only two points, another of four points, and the biggest of twelve points.

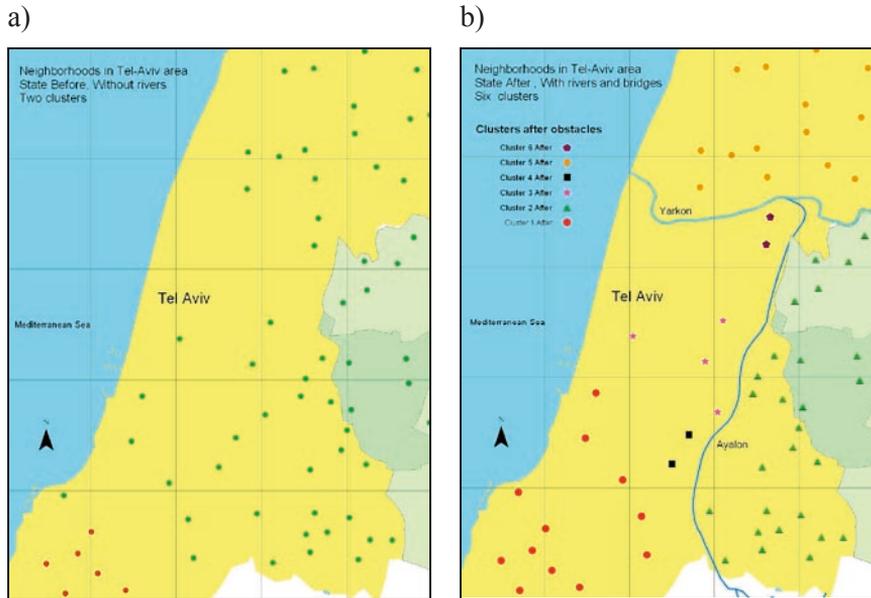


Figure 7. Clustering results for the obstacle free (a) and obstructed (b) settings

Table 1. "Association matrix" between the unobstructed and obstructed clusters

Original\Resulting	1	2	3	4	5	6	Max(c_j)
1	6	0	0	0	0	0	6
2	6	27	4	2	14	2	27

The COI computation is derived from a 2×6 matrix (Table 1), whose values reflect spatial organization of the resulting clusters with respect to the original (unobstructed) ones. As the table shows, for the original cluster consisting of six points, there is a resulting cluster containing all points (resulting cluster 1); the first column in C has this as the only non-zero value (Table 1). The other resulting clusters contain points that form the second original cluster (the larger one). As the second column in C shows, the scatter for this cluster is therefore large. Following Equation (3) the scatter for the first original cluster is zero, and is 28 for the second one. The COI measure of $28/61 \cong 0.46$, reflects the change of the clusters between unobstructed and obstructed cases. A visual comparison of the clustering results in Figure 7a, and 7b shows that such level of interference appears appropriate, as the obstacles have caused the large cluster to break into smaller pieces. At the

same time, the measure reflects the fact that some of these pieces are bigger than others. Notice that if the number of clusters with and without the obstacle were to be compared, the level of interference would have risen to $4/6 \cong 0.67$. The COI measure is, therefore, more attentive than the distribution of entities among clusters, and does not consider only a summary of the clustering results.

5 Conclusions

The contribution of the paper is in addressing a question that has not received much attention in relation to spatial data analysis, namely how the influence of a spatial obstacle can be measured. This question poses theoretical concerns relating to measuring properties that characterize obstacles and their influence, and to ways that such measures can be integrated. In answering them a more complete, and in fact correct, description of the spatial setup can be provided.

Focusing on analysis and derivation of measures for the influence of obstacles on connectivity and aggregation of data, an attempt was made to feature the actual changes, while at the same time to minimize the correlation between them. As the paper shows the derived measures are bounded and normalized, therefore allowing a uniform scale to quantify their impact irrespective of changes to the spatial data arrangement. The results indicate that the generated values provide adequate measures to feature the changes in the data.

References

- de Berg, M., M. van Kreveld, M.H. Overmars, and O. Schwarzkopf (2000). *Computational Geometry, Algorithms and Applications* (2nd ed.). Springer-Verlag
- Ester, M., H. Kriegel, J. Sander and X. Xu (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*
- Estivil-Castro, V. and I. Lee (2000). AUTOCLUST+: Automatic Clustering of Point-Data Sets in the Presence of Obstacles. In *Proceedings of International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining-TSDM2000*. Lyon, France
- Halkidi, M., Y. Batistakis and M. Vazirgiannis (2002). Clustering Algorithms and Validity Measures. *SIGMOD Record*, 31(3)
- Miller, H. and J. Han (2001). *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, London, UK
- Tung, A., J. Hou and J. Han (2001). Spatial Clustering in the Presence of Obstacles. In *Proceedings of the 17th International Conference on Data Engineering*. Heidelberg, Germany

Wang, X. and H. Hamilton (2003). DBRS: A Density-Based Spatial Clustering Method with Random Sampling. In Proceedings of the 7th PAKDD, Seoul, Korea: 563–575

Wang, X., C. Rostoker and H. Hamilton (2004). Density-Based Spatial Clustering in the Presence of Obstacles and Facilitators. Technical Report CS-2004-9. University of Regina, Department of Computer Science, Canada. ISSN: 0828-3494: 16pp

Zaiane, O.R. and C.H. Lee (2002). Clustering Spatial Data When Facing Physical Constraints. In Proceedings of the IEEE International Conf. on Data Mining, Maebashi City, Japan, 737-740

Zhang, J., D. Papadias, P. Mouratidis and M. Zhu (2005). Query Processing in Spatial Databases Containing Obstacles. *International Journal of Geographical Information Science*, 19(10):1091–1111