Studies in Big Data 15

Ana Paula Ferreira Dias Barbosa Póvoa
João Luis de Miranda
Editors

Operations Research and Big Data

IO2015-XVII Congress of Portuguese Association of Operational Research (APDIO)

IO2015
XVII Congress of APDIO

Springer

Measures in Sectorization Problems

Ana Maria Rodrigues[1] and José Soeiro Ferreira[2]

[1] INESC TEC and ISCAP – Instituto Politécnico do Porto, Porto, Portugal
amr@inesctec.pt
[2] INESC TEC and Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
jsf@fe.up.pt

Abstract. Sectorization means dividing a whole into parts (sectors), a procedure that occurs in many contexts and applications, usually to achieve some goal or to facilitate an activity. The objective may be a better organization or simplification of a large problem into smaller sub-problems. Examples of applications are political districting and sales territory division. When designing/comparing sectors some characteristics such as contiguity, equilibrium and compactness are usually considered. This paper presents and describes new generic measures and proposes a new measure, desirability, connected with the idea of preference.

1 Introduction

Sectorization problems (SP) occur in many contexts and applications in which a large region should be divided into smaller regions that meet specific conditions such as, for example, the division of a region into political districts ([Hes65], [Ric13]) or the definition of sales territories ([Gon11], [Hes71], [Kal05]). The division of a geographical area into smaller regions facilitates routing in the collection of municipal solid waste ([Rod15a], [Mou09]) or snow removal ([Muy02], [Sal12]). Sectorization may also appear in the description of regions for health care ([Ben13]), emergency accident ([Ber77]), for policing ([Dam02], [Lun12]) and location of schools ([Tak03], [Car04]), electricity distribution areas ([Ber03]) and maintenance operations [Per08]. The detection of communities in social networks is another recent application: finding groups of individuals (vertices) within which connections are dense but sparser between them ([New04]). Clustering and Sectorization present different motivations ([Kal05]): the first aims the dissimilarity between groups and the second (in territorial design, for instance) has just the opposite intention. Being the applications of SP so abundant and important, their study becomes of great relevance to society and also to science, since SP do not have an easy solution, and the solution methods used are highly dependent on applications, as can be attested by the above-mentioned references. These methods

A.P.F.D. Barbosa Póvoa and J.L. de Miranda (eds.), *Operations Research and Big Data,*
Studies in Big Data 15, DOI: 10.1007/978-3-319-24154-8_24

range from exact, using mathematical programming, to heuristic approaches, in a single or multi-criteria framework. But even those that follow "exact" models often end up in approximate methods.

Commonly, there are a set of common features to be preserved in the construction or evaluation of sectors, such as: Equilibrium (identical proportion of the parts in relation to the whole), Compactness (sectors showing regular shapes, circles or squares, avoiding "tentacles" ...) and Contiguity (elements of each sector arranged in single "body", thus avoiding breaking the sector into small portions). These three characteristics are not unique, although they account for most relevant applications. Therefore, the evaluation of the quality of the sectors, obtained by any method, is a difficult task, due to the differently defined metrics involved.

This paper discusses and proposes some general and new measures that can be used to assess the quality of sectors and subsequent comparison of the results obtained. In a more advanced stage of the ongoing research work, the authors intend to integrate these general measures with a sectorization approach based on an analogy with electromagnetism forces [Rod15b] and with a multi-criteria method described in [Fer13]. That may result in a "universal method" for many real sectorization problems, if appropriate adjustments are provided.

2 Measures

The most common measures or criteria used in Sectorization contemplate, in some way, conceptions or definitions of Equilibrium, Compactness and Contiguity. Other criteria such as integrity, in political districting, or racial balance in school districting, are used in specific applications. In this paper, in addition to advancing general definitions for the most common measures, another one, related with desirability, is proposed.

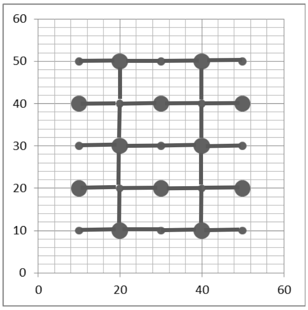


Fig. 1 Group with 25 elementary units to construct 4 sectors

The authors consider that the definitions should be simple, transparent and clear to decision makers and easily adapted to cover most current real situations. It may be questionable the use of complex measures when, often, the models need to be simplified and/or decisions must be taken in the absence of clear facts or data.

For the sake of illustration of the measures, a small example with 25 elementary units to be aggregated into 4 sectors is presented in Fig. 1. Suppose that a quantity is assigned to each point, in this case a quantity 1 assigned to a small grey circle and a quantity 2 assigned to a large grey circle. The total quantity equals 37 (12 large circles plus 13 small circles). Lines between circles establish relations between points.

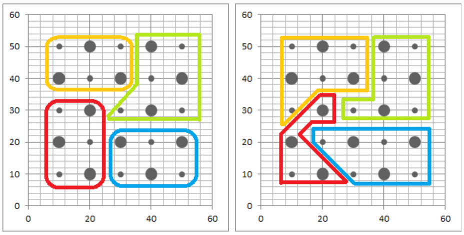


Fig. 2 Two sectorizations named Left and Right

Suppose that each of the four sectors has a limit quantity of 10, enough to accommodate the quantity 37, and that after applying different methods, two distinct sectorizations were obtained, “Left” and “Right” (see Fig. 2).

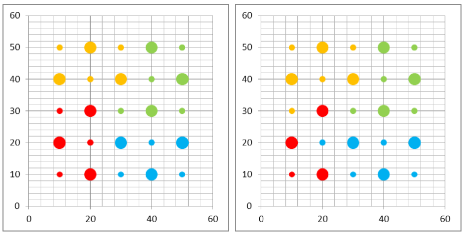
The example presented in Fig.1 with the initial map and the two proposed groups of sectors in Fig. 2 are used to illustrate the different measures.

2.1 *Equilibrium*

To evaluate the equilibrium, that is, the similarity between the quantities (number of electors, amount of work, quantity of waste to collect,…) in each sector, the proposed measure is the coefficient of variation (CV_q). It is calculated as follows, for a group of k sectors with quantities q_i, $i=1,...,k$:

$$CV_q = \frac{s'_q}{\overline{q}} \text{ , where } \overline{q} = \frac{\sum_{i=1}^{k} q_i}{k} \text{ and } s'_q = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(q_i - \overline{q})^2} .$$

Balanced sectors should have a CV_q as close to zero as possible. The use of the coefficient of variation allows the comparison between sectors obtained with different original sets.



If only the equilibrium is considered, the Left solution is better than the Right, CV_q(*Left*) <CV_q(*Right*)).

Fig. 3 Two solutions – Left: $CV_q = 0.054$ and Right: $CV_q = 0.162$

2.2 *Compactness*

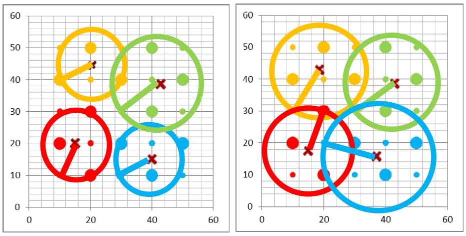
Let us associate the concept of *compactness* to the idea of *concentration* or *density*. Higher concentration should avoid sparse sectors.

Each sector i (i=1,…,k) has a value of compactness d_i that is defined by: $d_i = \frac{\sum_j q_{ij}}{dist(o_i, p_i)}$, q_{ij} represents the quantity assigned to the point j in sector i, and $dist(o_i,p_i)$ is the distance (Euclidean) between the centroid of the sector i, o_i, and the point of the same sector, p_i, that is farthest from o_i.

Depending on the particular application other more suitable metrics may be considered. In fact, in a broad sense, compactness may be defined as the total quantity (in a sector) divided by the number of elements of that sector.

Higher values of d_i represent higher values of compactness, which means a “higher density" in sector i. It is desirable that, regarding the same *map*, different sectors appear with similar values of compactness. To quantify that similarity it is proposed the coefficient of variation of compactness given by CV_d: $CV_d = \frac{s'_d}{\bar{d}}$, where $\bar{d} = \frac{\sum_{i=1}^{k} d_i}{k}$ and $s'_q = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(q_i - \bar{q})^2}$. A good sectorization must have a CV_d close to zero.

Once again the example is revisited and showed in Fig.4.



Regarding compactness, the Right solution is better than the Left (CV_d(Right)> CV_d(Left)).

Fig. 4 Left CV_d=0.7606 (d_1=0.8050; d_2=0.6472; d_3=0.8050; d_4=0.8050) and Right CV_d=0.1192 (d1=0.5824; d_2=0.6472; d_3=0.6685; d_4=0.5093)

2.3 *Contiguity*

The contiguity of sectors is another recurrent and important feature to considerer when evaluating the quality of sectors. Depending on the application it is desirable that each sector form just one body, or more *strongly*, that the interceptions between sectors are as small as possible. Two types are considered: Strong Contiguity and Weak Contiguity.

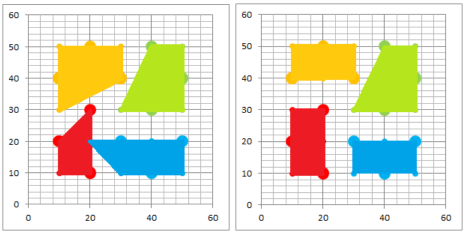


Fig. 5 *Left* sectors with strong contiguity; in *Right* sectors, red and blue touch each other

Strong Contiguity - A sectorization presents strong contiguity if the area of each smaller convex polygon, containing all elements of each sector, does not overlap the others. This condition may be even stronger: the areas should not touch each other. Only the Left solution shown in Fig.5 has absolute strong contiguity.

Weak Contiguity - is concerned with the shape of the sectors, and with the links between elementary units within each sector. If the subgraph induced by the elements (vertices) which represent the sector and links (arcs and edges) is connected, that is, there is a walk between any pair of elements of the sector without leaving it, the sector presents weak contiguity.

The evaluation of the weak contiguity of the k sectors is calculated using the adjacency matrices obtained from the k subgraphs $G_i' = (V'_i, E'_i)$ ($i=1,...,k$), where V'_i and E_i represent the set of vertices and the set of edges of subgraph G'_i, respectively. The number of vertices of each sector i is represented by: $|V_i|=n_i$, $i=1,...,k$. For each subgraph, G'_i also considers the symmetric matrix given by $M^i=[m^i_{wj}]_{w,j=1,...,ni}$ with principal diagonal with zeros

$$M^i = \begin{bmatrix} 0 & m^i_{12} & m^i_{13} & \cdots & m^i_{1ni} \\ m^i_{21} & 0 & m^i_{23} & \cdots & m^i_{2n_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m^i_{n_i1} & m^i_{n_i2} & m^i_{n_i3} & \cdots & 0 \end{bmatrix}$$

$$\text{where } m^i_{wj} = \begin{cases} 1, \text{ if in sector i exists a walk between w and j} \\ \qquad 0, \text{ otherwise} \end{cases}$$

If for all $j \in \{1,...,n_i\}$, $\sum_{w=1}^{n_i} m^i_{wj} = n_i - 1$ or, which is equivalent, for all $w \in \{1,...,n_i\}$ the condition $\sum_{j=1}^{n_i} m^i_{wj} = n_i - 1$ is verified, then sector i is contiguous. But different levels of contiguity must be considered.

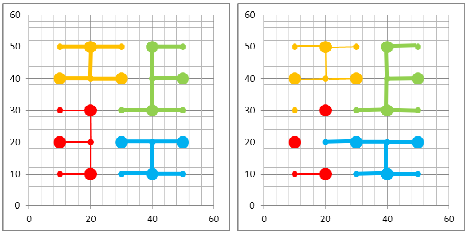
The next expression is used, as a measure of contiguity (c_i) for each sector i:

jsf@fe.up.pt

$$c_i = \frac{\sum_{j=1}^{n_i}\left(\sum_{w=1}^{n_i} m_{wj}^{i}\right)}{n_i\left(n_i - 1\right)}$$. This is not enough to characterize the level of contiguity. The quality of the sectors must combine the contiguity of all sectors produced. To evaluate the resulting contiguity, depending on the objective, some measures can be considered such as the difference between maximum and minimum values of contiguity or the value of contiguity of the worst sector. The weighted average ($\bar{c}$) of *isolated contiguities*, is proposed:

$$\bar{c} = \frac{\sum_{i=1}^{k} c_i \cdot n_i}{N}$$. $\bar{c}$ is a value that is always between 0 and 1. Using weights the final contiguity is proportional to the size of the sector.

From the perspective of contiguity, a good sectorization must have a $\bar{c}$ value as close to 1 as possible.



When only (weak) contiguity is consi-dered, Left is preferred.

Fig. 6 Left $\bar{c}$ =1 (c_1=1; c_2=1; c_3=1; c_4=1) and Right $\bar{c}$ = 0.7867 (c_1 = 0.71429; c_2=1; c_3=0.1667; c_4=1)

2.4 Desirability

Suppose the situation in which it would be desirable that some specific groups of elements belong to the same sector. It is just a preference, and not a mandatory situation. For some reason, eventually past experiences, the decision maker prefers that some groups of points stay together: served by the same vehicle, vote in the same district or attend the same school, depending on the application.

For instance, after a sectors construction, the *degree of desirability* should measure how close the solution is from the previous preferences of the decision maker.

In this context, let F represent the number of groups identified by the decision maker as having some affinity and f_i, $i=1,\dots,F$, the number of elementary units in each of the predefined F groups.

If the sectorization involves the creation of k sectors, the maximum number (H) of elementary units out of the F groups, defined by the decision maker, is

$H = \sum_{i=1}^{F} \left(f_i - \min\left\{ f_i - 1; \left\lceil \frac{f_i}{k} \right\rceil \right\} \right)$. Then, the measure of desirability (*Des*) is defined by: $Des = 1 - \frac{NE}{H}$, where *NE* is the number of elements out of the groups (after sectorization). In the following, we illustrate the definition by considering the example of Fig. 7, where $k=4$, $f_1=6$ and $f_2=4$.

$$H = f_1 - \min\left\{ f_1 - 1; \left\lceil \frac{f_1}{k} \right\rceil \right\} + f2 - \min\left\{ f_2 - 1; \left\lceil \frac{f_2}{k} \right\rceil \right\} = 6 - \min\{5;2\} + 4 - \min\{3;1\} = 7$$

This means that, in the worst case (a sectorization completely disrespecting the preferences), the number of outsiders is equal to 7.

Left and Right sectorizations have *NE* equal to 2 and 4, respectively (Fig.7).

Finally, $Des = 1 - \frac{2}{7}$ (Left) and $Des = 1 - \frac{4}{7}$ (Right). In summary, Left sectorization is closer to the previous desire of decision maker.

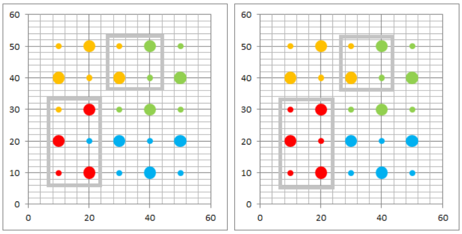


Fig. 7 Grey rectangles represent groups defined by the decision maker. Left *Des*= 0.7143 and Right *Des*=0.4286

Values of *desirability* near to 1 indicate that groups selected are mostly respected by the sectorization process. On the other hand, values near to 0 indicate that the sectorization made is not respecting the prior preferences.

If only desirability is considered, Left is better than Right.

3 Conclusion

Sectorization problems (SP) occur in many contexts and real applications, as described. Solving SP is a complex task, as mentioned in many publications, and it is usually indispensable to contemplate one or more characteristics of the sectors, to evaluate the quality of the solutions. Typical properties are connected with the ideas of equilibrium, compactness and contiguity.

This paper also considered these properties, while proposing new measures or criteria and aiming at defining them as generic and transparent as possible, so that they can be used in most practical SP. Desirability is another and new measure introduced. The idea is to embrace possible desirabilities (or preferences) of a decision maker.

This work is part of the authors' ongoing research about SP, which encompasses the integration of these generic measures with a general sectorization approach based on an analogy with Electromagnetism [Rod15b] and with a multi-criteria soft method described in [Fer13]. The final intention is to devise a "universal method" capable of dealing with many real SP, if appropriate adjustments are provided.

Acknowledgments. This work was partially financed by National Funds through the FCT-Portuguese Foundation for Science and Technology within project "Project SEROW/PTDC/EGE-GES/121406/2010" and by the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by National Funds, through the FCT within "Project NORTE-07-0124-FEDER-000057".

References

[Ben13] Benzarti, E., Sahin, E., Dallery, Y.: Operations management applied to home care services: Analysis of the districting problem. Decision Support Systems 55, 587–598 (2013)

[Ber03] Bergey, P.K., Ragsdale, C.T., Hoskote, M.: A Simulated Annealing Genetic Algorithm for the Electrical Power Districting Problem. Annals of Operations Research 121, 33–55 (2003)

[Ber77] Bertolazzi, P., Bianco, L., Ricciardelli, S.: A method for determining the optimal districting in urban emergency services. Computers & Operations Research 4, 1–12 (1977)

[Car04] Caro, F., Shirabe, T., Guignard, M., Weintraub, A.: School redistricting: embedding GIS tools with integer programming. Journal of the Operational Research Society 55, 836–849 (2004)

[Dam02] D'Amico, S.J., Wang, S.-J., Batta, R., Rump, C.M.: A simulated annealing approach to police district design. Computers & Operations Research 29, 667–684 (2002)

[Fer13] Ferreira, J.S.: Multimethodology in Metaheuristics. Journal of the Operational Research Society 64, 873–883 (2013)

[Gon11] González-Ramírez, R., Smith, N.R., Askin, R.G., Miranda, P.A., Sánchez, J.M.: A Hybrid Metaheuristic Approach to Optimize the Districting Design of a Parcel Company. Journal of Applied Research and Technology 9, 19–35 (2011)

[Hes65] Hess, S., Weaver, J., Siegfeldt, H., Whelan, J., Zitlau, P.: Nonpartisan Political redistricting by computer. Operations Research 13, 998–1006 (1965)

[Hes71] Hess, S.W., Samuels, S.A.: Experiences with a sales districting model: criteria and implementation. Management Science 18, 41–54 (1971)

[Kal05] Kalcsics, J., Nickel, S., Schröder, M.: Towards a unified territorial design approach - applications, algorithms and GIS integration, Top, 13, 1–56. Springer (2005)

[Lun12] Lunday, B.J., Sherali, H.D., Lunday, K.E.: The coastal seaspace patrol sector design and allocation problem. Computational Management Science 9, 483–514 (2012)

[Mou09] Mourão, M.C., Nunes, A.C., Prins, C.: Heuristic methods for the sectoring arc routing problem. European Journal of Operational Research 196, 856–868 (2009)

[Muy02] Muyldermans, L., Cattrysse, D., Van Oudheusden, D., Lotan, T.: Districting for salt spreading operations. European Journal of Operational Research 139, 521–532 (2002)

[Per08] Perrier, N., Langevin, A., Campbell, J.F.: The sector design and assignment problem for snow disposal operations. European Journal of Operational Research 189, 508–525 (2008)

[Ric13] Ricca, F., Scozzari, A., Simeone, B.: Political Districting: from classical models to recent approaches. Annals of Operations Research 204, 271–299 (2013)

[Rod15a] Rodrigues, A.M., Ferreira, J.S.: Waste collection routing - limited multiple landfills and heterogeneous fleet. Networks 65, 155–165 (2015)

[Rod15b] Rodrigues, A.M., Ferreira, J.S.: Sectors and Routes in Solid Waste Collection, to be published in Springer-Verlag - CIM Series in Mathematical Sciences (2015)

[Sal12] Salazar-Aguilar, M.A., Langevin, A., Laporte, G.: Synchronized arc routing for snow plowing operations. Computers & Operations Research 39, 1432–1440 (2012)

[Tak03] Takashi, T., Yukio, S.: Evaluation of School Family System Using GIS. Geographical Review of Japan 76, 743–758 (2003)

jsf@fe.up.pt