systems

MDPI

Article

Electoral Districts in Chile: Optimizing Socio-Economic Homogeneity and Demographic Balance

Rodrigo Rebolledo [1], Maykol Reinoso [1], Óscar Cornejo [1,*], Carlos Obreque [2] and Felipe Baesler [3,*]

[1] Industrial Engineering Department, Faculty of Engineering, Universidad Católica de la Santísima, Concepción, Alonso de Ribera 2850, Concepción 4090541, Chile; rrebolledo@ucsc.cl (R.R.); maykol.reinoso@ucsc.cl (M.R.)
[2] Industrial Engineering Department, Faculty of Engineering, Universidad del Bío-Bío, Concepción 4081112, Chile; cobreque@ubiobio.cl
[3] College of Engineering, Architecture and Design, Universidad San Sebastián, Concepción 4081339, Chile
* Correspondence: ocornejo@ucsc.cl (Ó.C.); felipe.baesler@uss.cl (F.B.)

Abstract: This article addresses the problem of unequal representation in Chile, where the current districting does not effectively consider its socio-economic diversity. An innovative methodology is proposed that uses the socio-economic dissimilarity distance (SED) obtained using a cluster analysis to create more homogeneous electoral districts. This SED is incorporated into a mathematical programming model for (re)districting and seat allocation, taking into account criteria such as the demographic balance, contiguity and compactness. The application of this methodology in the Santiago Metropolitan Region shows a significant improvement in both the socio-economic homogeneity and demographic balance of the districts. This research has relevant implications for electoral justice in Chile, as it proposes a way to improve the representativeness and ensure that the needs of each social group are reflected in the decision-making process.

Keywords: political districting; socio-economic criteria; cluster analysis; contiguity; malapportionment

Academic Editor: Maurice Yolles

Received: 20 December 2024
Revised: 12 January 2025
Accepted: 14 January 2025
Published: 30 January 2025

Citation: Rebolledo, R.; Reinoso, M.; Cornejo, Ó.; Obreque, C.; Baesler, F. Electoral Districts in Chile: Optimizing Socio-Economic Homogeneity and Demographic Balance. *Systems* **2025**, *13*, 85. https://doi.org/10.3390/systems13020085

1. Introduction

Democratic electoral systems, including that in Chile, base their legitimacy on citizen representation, with the main objective being demographic balance, which seeks to ensure that each person's vote has the same value or weight, known colloquially as "one person, one vote", and that parliament is as representative as possible with respect to the population. Even so, ref. [1] mentions that it is necessary to consider international experiences in order to minimize the distortions identified in the Chilean case.

At the global level, electoral (re)districting considers three fundamental criteria: population, geography and socio-economic homogeneity [2]. The population criterion basically seeks balance in terms of the population density per seat so as to achieve equal importance in the vote of each citizen. The geography criterion mainly seeks to respect natural and non-natural boundaries in the allocation of territorial units into districts. The socio-economic homogeneity criterion seeks to ensure that the population of each district is the least dissimilar in socio-economic terms.

Proper redistricting should avoid gerrymandering, which is the manipulation of districts to favor a particular group [3–5].

Regarding the above, ref. [6] presents different names that are used to refer to the problem of aggregating areas into homogeneous regions; some of these are restricted clustering [7], regional regrouping [8] and region building [9]. In the local sphere, the first electoral districting in Chile was established in 1925., and then during the military dictatorship in 1989, the binominal system was implemented. It was in 2015 that the Binominal System Reform was implemented, which put an end to the binominal system and created a new electoral map, which greatly improved the previous system, mainly in terms of its representativeness. However, distortions are still evident, although to a lesser extent [10].

In the present work, the SED between communes in the same region is measured, which is used to group communes into clusters and obtain electoral districts formed of a similar population in socio-economic terms, which can help to achieve more effective representation [11]. Therefore, this work can be classified into the second level mentioned by [12], that is, an electoral geographic study that aims to define the way in which the electoral reality is distributed and to establish relationships between that electoral reality and other social aspects, such as culture and the level of socio-economic development, among other factors.

One of the most appropriate techniques used for this purpose is cluster analysis, which has as its primary objective the search for groups in a set of individuals, obtaining a distance between each of them based on variables with similar values for each of these groups [13]. In this case, the individuals under study correspond to the communes for which an SED is obtained. This is incorporated into a mathematical programming model of districting and seat allocation for the Chamber of Deputies in Chile [14] in order to incorporate the socio-economic criterion. The choice of variables for elaborating the SED between communes depends directly on the information available at the communal level. It should be considered that this distance will always have an inherent degree of subjectivity associated with the knowledge of the researcher, the available data and the references consulted; it is mentioned that spatial differentiation, i.e., the distribution of a population with certain characteristics across space, has always existed but is currently more clearly and broadly observed in the spaces of developing cities. This more visible clarity (if present) can make it possible to identify more clearly, both visually and statistically, the different socio-economic realities present in the regions of a country [15].

The purpose of this study is to provide a reference methodology for obtaining districts formed of more socio-economically homogeneous communes based on the socio-economic dissimilarity distance (SED). This distance is then incorporated into a mathematical programming model for redistricting and seat allocation, which includes multiple criteria of the redistricting process, such as demographic balance, contiguity, compactness, respect for natural and non-natural boundaries and geographical units.

The level of detail of the data is at the communal level since the objective is to determine the communes that should be grouped into the same district, and it also coincides with the minimum geographical units used in the model for the districting and allocation of seats.

The Chilean democratic electoral system bases its legitimacy on citizen representation, seeking a parliament that reflects the country's diversity. However, the current electoral system does not explicitly consider socio-economic diversity, which may result in the unequal representation of different social groups.

This article proposes developing and validating a methodology for optimizing electoral districting in Chile, considering socio-economic homogeneity and demographic balance. To this end, we propose the creation of electoral districts made up of more socio-economically homogeneous communes by means of a socio-economic dissimilarity distance (SED).

The SED, obtained through a cluster analysis, will be integrated into a mathematical programming model for districting and seat allocation.

The methodology will be applied to the Metropolitan Region of Santiago, composed of 52 communes, 7 districts and 47 seats.

It is hoped that this research will contribute to better representation of the population, where the needs and interests of each social group are reflected in the decision-making process.

2. Methods

This section details the methods used to construct the socio-economic dissimilarity distance (SED) and the mathematical programming model for districting and seat allocation.

Electoral System, (Re)Districting and Socio-Economic Homogeneity

An electoral system is defined as a set of rules, norms, procedures and/or mechanisms designed for voting and obtaining public office in a country or region [16]. Since the end of the military dictatorship, the electoral system has been the binominal system. The binominal system divided Chile's territory into constituencies (Senate) and districts (Chamber of Deputies), where the constituencies were made up of one or more districts, and in turn, the districts were made up of one or more communes. In each constituency and district, two senators and two deputies were elected, respectively (hence the name binominal). This system violated the principle of representativeness, as a person from a low-population district had up to eight times more voting power than a person from a high-population district. In addition, it radically disadvantaged small parties that were not able to join a coalition [17]. Law 20.840 was enacted in 2015, which implemented a new proportional electoral system that strengthened the representation of the population in the National Congress. This implied redistricting, i.e., the creation of new constituencies and districts. The number of districts was reduced from 19 to 15 (1 for each region), and later in 2017, the region of Ñuble was created, increasing this number to 16 districts, and each of these corresponded to between 2 and 5 senators, increasing the number of senators from 38 to 50. The number of districts decreased from 60 to 28, and each district corresponded to between 3 and 8 deputies, increasing the number of deputies to 155. In addition, the law replaced the binominal system with a proportional and inclusive system that strengthened the representativeness of the population in the National Congress through the D'Hondt system, which is a mathematical method for the allocation of seats into the lists of each constituency and district.

An electoral district is defined as a division of territory within whose boundaries a political event takes place [2] or a territorial area that will be the basis of representation of a social group before Congress [11]. The redistricting process should be seen as an ongoing task in countries or regions in order to form and adapt districts based on the changing needs and/or realities of the population, such as demographic growth or decline, in order to maintain adequate representation of the population [18]. When too much time is allowed to pass without a redistricting process, it is common to observe a deterioration in representativeness indicators given that while the electoral system remains static, the population within it is dynamic [19]. A common indicator for measuring representativeness is malapportionment (Map). This index, adapted from the proportionality index [20], calculates the difference between the percentages of the population and seats in a district.

$$Map = \frac{1}{2}\sum_{i=1}^{D}|\%E_i - \%P_i| \tag{1}$$

where D is the number of districts, $\%E_i$ is the percentage of seats in the district, i, $i = 1, 2, \ldots, D$, and $\%P_i$ is the percentage of the population in the district i, $i = 1, 2, \ldots, D$.

The search for equitable representation involves considering various criteria in electoral districting processes, such as [1,4,11,19] and [21–26]:

(a) Contiguity: It should be possible to go from any point in the district to another without leaving the district;
(b) Compactness: The shape of the district must have a regular form;
(c) Demographic balance: The population or number of voters per seat must be balanced between districts;
(d) Respect for boundaries: Natural or administrative boundaries or constrictions must be respected;
(e) Geographical units: Consideration should be given to the administrative units of the region to be districted;
(f) Socio-economic homogeneity: Districts should be as homogeneous as possible for better representation of the population

A homogeneous geographic region consists of a set of contiguous areas that show a high degree of similarity with respect to a group of attributes, such as attributes characterizing its demographic structure, employment structure, housing structure or household structure; attributes reflecting other socio-economic aspects; historical links; previous regionalizations, transport patterns; and means of communication, among others [27,28]. According to [2], socio-economic characterization is conceived of through the generation of indicators of the marginalization, social backwardness and economic participation of territorial areas. Ideally, as many socio-economic attributes as possible should be included, trying to group homogeneous units by most of these attributes [11]. However, the choice of these variables depends very closely on the data availability [29]. In most cases, the availability of data is inverse to the scale of analysis, i.e., at a large scale, such as at the country level, more data and information are available than at smaller scales. In [30], this is confirmed, mentioning that the most interesting information is available for "uninteresting" scales.

Electoral districting works that have considered socio-economic criteria are scarce, limited to [31] only considering income, which is solved by means of a heuristic. In [32], as a criterion, personal income was calculated as the sum of the standard deviations of the income of each district divided by the average total income, among other criteria. In addition, it was solved using a tabu search with adaptive memory to determine a feasible solution to the problem. In [29], a framework for obtaining homogenous territorial clusters based on a Pareto frontier is discussed that includes multi-criteria related to the territorial endogenous resources, economic profile and socio-cultural features.

In [33], a framework for political redistricting applied to the state of Arizona is presented, which, through four steps, creates an initial redistricting map and successively optimizes the compactness, competitiveness and population balance using a multilevel algorithm and a local neighbor search.

At the Chilean level, a new demographic equity criterion is presented in [34] that encourages better representation of the candidates during the campaign and once they are elected. A multi-objective combinatorial optimization problem considers this new criterion in addition to malapportionment, balance (minimizing the maximum malapportionment of the districts) and compactness, with these being applied to the Chilean case of the 2015 Reform. Equally, in [35], two mathematical programming models of feasibility with and without defined district centers are defined. They are applied considering the regional boundaries and without the regional boundaries, allowing communes from different regions to form a district. This study seeks to minimize the maximum population imbalance between the districts without considering the seats per district, i.e., it seeks for each district to have an equal population number, and contiguity is incorporated by means of shorter paths between each commune and possible district centers. In both of these aforementioned works, socio-economic criteria are not considered in defining the districts.

With regard to the creation and objectivity of socio-economic indicators, ref. [22] mentions that socio-cultural criteria are complex and subjective, and it is therefore advisable to establish methods that allow them to be systematized. This degree of subjectivity is one of the main reasons for the disagreements that occur in the study of urban fragmentation, in its definitions, in the limits of the term and in the methods and techniques used for the measurement of its indicators [36]. One of the reasons for this is that its analysis is not exclusive to any specific discipline. Hence, each discipline examines and studies it according to its language and methodologies, also considering that most of the variables associated with this subject are multidimensional [15]. For example, Chile's multidimensional poverty rate depends on five dimensions. Most of these variables are under permanent discussion, and their estimates remain open and subject to change.

An important limitation of this research is the degree of subjectivity inherent in the selection of the socio-economic variables for the construction of the socio-economic dissimilarity distance (SED).

(a) The choice of variables depends on the knowledge of the researcher, the available data and the references consulted.
(b) At larger scales of analysis (e.g., at the national level), the availability of data is greater than that at smaller scales (e.g., at the community level). This may influence the SED.
(c) The multidimensionality of socio-economic variables, such as those that make up the multidimensional poverty rate in Chile and their constant changes, makes it difficult to measure them accurately.

3. Methodology

The methodology for optimizing electoral districting is implemented in two stages:

Stage 1. The cluster analysis and determination of the SED:

1. Hierarchical cluster analysis: Ward's method with the squared Euclidean distance is used to obtain the optimal number of clusters (k) in each region;
2. Non-hierarchical cluster analysis (k-means): This is applied using the number of clusters (k) obtained in the previous step to identify the segmenting variables in each region;
3. Calculation of the SED: A new cluster analysis (hierarchical and k-means) is performed using only the segmenting variables to obtain the distances between the communes and the clusters to which they belong.

Stage 2. Integration of the SED into the mathematical programming model:

1. Incorporation of the SED into the model: The SED is included as a constraint in the model of districting and seat allocation, limiting the value of the TSED;
2. Model resolution: The model is solved with different limits for the TSED, seeking a balance between socio-economic homogeneity and other districting criteria;
3. Analysis of the results: The results obtained are evaluated in terms of TSED, demographic balance (MALR) and other relevant indicators.

Cluster analysis is a set of multivariate techniques used to classify a set of individuals into homogeneous groups. It belongs to the set of techniques whose purpose is to classify individuals. This technique seeks the maximum homogeneity in observations of the same group and the greatest difference between groups [37]. On the other hand, a cluster analysis can allow us to find associations and structures in the individuals under study which are not evident at first but can be useful once these groups have been detected. These results can contribute to the formal definition of a classification scheme for a set of individuals, suggestions for statistical models for describing populations and the assignment of new individuals into classes for diagnosis and identification, among other applications, using the characteristics of the observations that serve to classify them [38]. Regarding the various methods available for performing a cluster analysis, Ward's method is one of the most widely used to determine the number of groups and subsequently apply a non-hierarchical method [39]. It is recommended when groups of a similar size are desired [40].

In this paper, the cluster analysis can be applied to regions with a population greater than 1.93% of Chile's total population, a value obtained by dividing the minimum number of deputies per district (3) by the total number of deputies (155). Regions with a population less than or equal to 1.93% (the regions of Arica and Parinacota, Tarapacá, Atacama, Aysén and Magallanes) are assigned a single district, so it is not necessary to carry out the socio-economic homogeneity study.

The magnitude of the variables under study is given as a percentage, as this is a good way of "pre-standardizing" the data. Moreover, this is common in works of this type, as noted in [29], where percentages of the primary and secondary activities, among others, are used. It is worth mentioning that the variables used to measure socio-economic homogeneity are not exactly defined since they vary according to the criteria and purposes of the study and the data availability. The economic variables used in this work to perform the cluster analysis are the percentage of companies in each category in each commune with respect to the total number of companies in that commune and the percentage of dependent workers reported in each category in each commune with respect to the total number of dependent workers in that commune. To calculate these figures, we used the number of enterprises and the number of dependent workers reported, disaggregated by commune and category [41]. The items and the respective acronyms for the economic variables are presented in Table 1.

Table 1. Variable and acronym source: prepared by the authors based on data from [41].

Variable	Acronyms Number of Companies	Acronyms for Number of Dependent Workers
Agriculture, livestock, forestry and fisheries	ANEM	ANDE
Mining and quarrying	BNEM	BNDE
Manufacturing industry	CNEM	CNDE
Electricity, gas, steam and air-conditioning supply	DNEM	DNDE
Water supply; wastewater disposal, waste management and decontamination	ENEM	ENDE
Construction	FNEM	FNDE
Wholesale and retail trade; repair of motor vehicles and motorbikes	GNEM	GNDE
Transport and storage	HNEM	HNDE
Accommodation and food service activities	INEM	INDE
Information and communications	JNEM	JNDE
Financial and insurance activities	KNEM	KNDE
Real estate activities	LNEM	LNDE
Professional, scientific and technical activities	MNEM	MNDE
Administrative and support service activities	NNEM	NNDE

Table 1. *Cont.*

Variable	Acronyms Number of Companies	Acronyms for Number of Dependent Workers
Public administration and defense; compulsory social security schemes	ONEM	ONDE
Teaching	PNEM	PNDE
Human health care and social work activities	QNEM	QNDE
Arts, entertainment and recreational activities	RNEM	RNDE
Other service activities	SNEM	SNDE
Activities of households as employers; undifferentiated activities of households	TNEM	TNDE
Activities of extraterritorial organizations and bodies	UNEM	UNDE

Defining the economic variables in this way allows us to identify communes that while differing in population have economic variables that behave similarly. Otherwise, it would not be easy to group communes with considerable population differences given that the values of their economic variables would have differences influenced by the populations associated with each commune. In addition to economic variables, social variables are also used. The social variables considered are presented in Table 2 and were obtained from the census conducted in Chile in 2017 [42] and from the Socioeconomic Characterization Survey CASEN 2017 [43].

Table 2. Social variable source: prepared by the authors, based on 2017 census [32] and CASEN 2017 [43].

Variables	Acronym	Source
Percentage of the population belonging to indigenous or aboriginal peoples	PPIO	CENSO 2017
Average years of schooling	PAES	CENSO 2017
Percentage of population living in urban areas	PPZU	CENSO 2017
Income poverty rate	TPIN	CASEN 2017
Multidimensional poverty rate	TPMU	CASEN 2017

For the development of the SED, the methodology described in [44] for conducting a cluster analysis was used as the basis. Initially, a hierarchical cluster analysis of the agglomerative type (using Ward's method with the squared Euclidean distance) was performed to obtain the number of clusters to be used in the non-hierarchical cluster analysis (k-means), which was used to determine the segmenting variables in each region of Chile. Finally, both analyses were performed again to obtain the distances between the communes and the clusters of membership of each of them, based only on the previously identified segmenting variables.

3.1. Formulation of the Mathematical Programming Model

We used the model given in [14] as a basis and fixed the number of seats and districts for a region. In this case, new variables and restrictions were added to incorporate socio-economic homogeneity.

Parameters and sets:

I: total number of communes;

D: total number of districts;

SED_{ij}: the socio-economic dissimilarity distance between communes i and j;

$TSED_{\max}$: the maximum tolerated value of $TSED$.

Decision variables:

x_{ij}: takes the value 1 if commune i is assigned to district center j, and 0 otherwise; $i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, I$;

t_{ij} : takes the value 1 if communes i and j belong to the same district, i.e., they are assigned to the same district center, and 0 otherwise; $i, j = 1, 2, \ldots, I$.

The socio-economic homogeneity criterion, measured by the distance determined by the cluster analysis, is incorporated into the model for the district and seat allocation. To incorporate the socio-economic distances, we use the form defined by [37] (Duque, Anselin and Rey (2012)) for the Max-p-Regions problem, where the distance of a group (in this case, a district) is considered as the sum of the distances between all of the elements (in this case, communes) in it. Thus, the total socio-economic dissimilarity distance (hereafter, the TSED) is the sum over all of the districts, i.e.,

$$TSED = \sum_{(i,j) \in D \ / \ i<j} SED_{ij} \tag{2}$$

where D is the set of the pairs of communes that are assigned to the same district.

In constraint (3), if two communes i and j are assigned to the same district center k, i.e., they belong to the same district, this forces the variable t_{ij} to take the value 1, and constraint (4) limits the $TSED$ to the maximum tolerable value.

$$x_{ik} + x_{jk} - 1 \leq \ t_{ij};\ i, j, k = 1, 2, \ldots, I \tag{3}$$

$$\sum_{i=1}^{I} \sum_{\substack{j=1 \\ i<j}}^{I} SED_{ij} t_{ij} \leq TSED_{\max} \tag{4}$$

Finally, constraints (5) and (6) show the nature of the variables.

$$x_{ij} \text{ as a binary variable, } \ i, j = 1, 2, \ldots, I \tag{5}$$

$$t_{ij} \text{ as a binary variable, } \ j = 1, 2, \ldots, I \tag{6}$$

4. Results and Discussion

In this section, the main results of applying the methodology presented above to each region separately are detailed, followed by the presentation of more detailed information on the procedure applied with the example of the Metropolitan Region of Santiago.

4.1. The Number of Optimal Clusters and Segmenting Variables by Region

The number of clusters formed per region is the number of socio-economic realities that can be identified in each of them. The segmenting variables are those whose value is significantly different when compared with those for communes in other clusters, so it can be concluded that these variables are significantly relevant when forming the clusters. The number of clusters and segmenting variables by region can be seen in Table 3.

Table 3. Number of current districts, number of optimal clusters and segmenting variables by region. Source: own elaboration.

Region	Number of Current Districts	Number of Clusters	Segmenting Variables
Antofagasta	1	2	PPZU, TPIN, ENEM, ENDE, GNEM, HNEM, HNDE, JNEM, ONEM, SNEM, SNDE

Table 3. *Cont.*

Region	Number of Current Districts	Number of Clusters	Segmenting Variables
Coquimbo	1	2	PPZU, PAES, TPIN, ANEM, ANDE, ENEM, FNEM, FNDE, GNDE, HNDE, JNEM, JNDE, LNEM, MNEM, MNDE, NNEM, NNDE, PNEM, SNEM
Valparaíso	2	2	PPIO, PPZU, PAES, TPIN, ANEM, ANDE, CNEM, CNDE, DNEM, ENEM, FNEM, FNDE, GNEM, GNDE, JNEM, KNEM, KNDE, LNEM, LNDE, MNEM, MNDE, NNDE, PNEM, QNEM, RNEM, SNEM, TNEM
Metropolitana	7	3	PPIO, PPZU, PAES, TPIN, TPMU, ANEM, ANDE, BNEM, CNEM, CNDE, DNEM, ENEM, FNEM, FNDE, GNEM, GNDE, HNEM, INDE, JNEM, KNEM, KNDE, LNEM, LNDE, MNEM, MNDE, NNEM, NNDE, ONEM, ONDE, PNEM, QNEM, RNEM, SNEM. UNEM
O'Higgins	2	3	PPIO, PPZU, PAES, TPIN, ANEM, ANDE, CNEM, CNDE, ENEM, ENDE, FNEM, FNDE, GNEM, GNDE, JNEM, JNDE, MNEM, MNDE, NNEM, PNEM, PNDE, QNEM, RNEM, SNEM
Maule	2	3	PPIO, PPZU, PAES, ANEM, ANDE, CNEM, DNEM, ENEM, FNEM, FNDE, GNEM, HNEM, INEM, INDE, JNEM, JNDE, KNEM, LNEM, MNEM, NNEM, ONDE, PNEM, QNEM, QNDE, RNEM, SNEM,
Ñuble	1	2	PPIO, PPZU, PAES, TPIN, TPMU, ANEM, ANDE, BNEM, CNEM, CNDE, FNEM, GNEM, HNEM, HNDE, JNEM, JNDE, MNEM, NNEM
Biobío	2	2	PPIO, PPZU, PAES, TPIN, TPMU, ANEM, ANDE, CNEM, CNDE, DNEM, DNDE, ENEM, FNEM, FNDE, GNEM, HNEM, INEM, MNEM, NNEM, NNDE, ONEM, ONDE, PNEM, PNDE, SNEM, UNEM, UNDE
Araucanía	2	3	PPIO, PPZU, PAES, TPIN, TPMU, ANEM, BNDE, ENEM, FNEM, INDE, KNEM, LNEM, LNDE, MNEM, MNDE, NNEM, ONEM, ONDE, PNEM, QNEM, SNDE
Los Ríos	1	2	PPZU, PAES, TPIN, TPMU, CNDE, INDE, JNEM, JNDE, LNEM, LNDE, MNEM, MNDE, NNEM, QNEM, QNDE
Los Lagos	2	2	PPZU, PAES, TPIN, TPMU, ANEM, ANDE, BNDE, CNEM, ENEM, GNDE, INDE, JNEM, JNDE, KNEM, KNDE, LNEM, MNEM, MNDE, NNEM, ONEM, QNEM, QNDE, SNEM

The number of clusters formed in each region varies between two and three, which means that in each region, at most, three socio-economic realities can be distinguished. This does not mean that the number of districts to be formed must be equal to the number of clusters since it is possible to distribute communes belonging to the same cluster into two or more districts. The important thing is to try, as far as possible, to form districts with communes with a similar socio-economic reality.

4.2. Results of the Cluster Analysis for the Santiago Metropolitan Region

The Metropolitan Region of Santiago has a population of 7,122,808 inhabitants (according to the 2017 census) and a surface area of 15,403 km^2. It is made up of a total of 52 communes, and its capital is Santiago (the capital of Chile). It borders the Valparaíso region, the O'Higgins region and the province of Mendoza (Argentina). It currently has a total of 7 districts and 47 deputies.

When performing the hierarchical analysis of the Santiago Metropolitan Region, three clusters can be observed in the dendrogram, as shown in Figure 1.

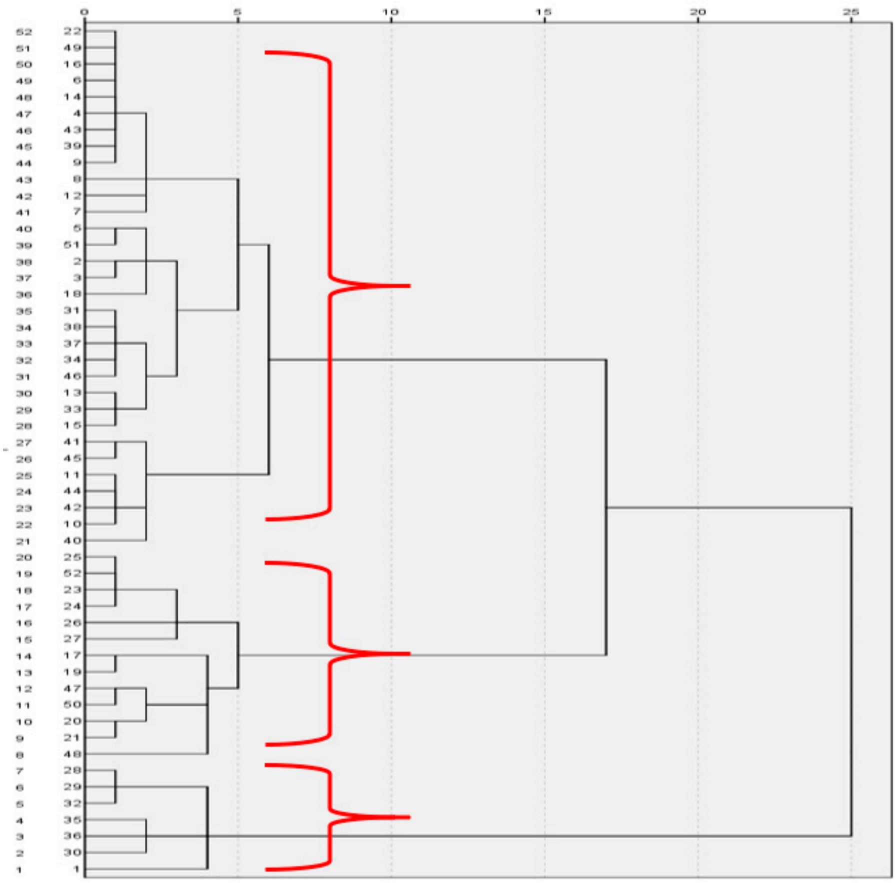

Figure 1. Dendrogram of the communes of the Metropolitan Region of Santiago. Source: own elaboration.

After applying a non-hierarchical k-means analysis, an ANOVA table is obtained to identify the segmenting variables, considering a significance of less than 0.05 (see Table 4).

Table 4. ANOVA table for the Santiago Metropolitan Region with three clusters. Source: own elaboration.

ANOVA					
Variables	F	Sig.	Variables	F	Sig.
PPIO	**22.24**	**0.00**	**JNEM**	**42.15**	**0.00**
PPZU	**30.93**	**0.00**	JNDE	0.56	0.57
PAES	**62.49**	**0.00**	**KNEM**	**35.49**	**0.00**
TPIN	**12.70**	**0.00**	**KNDE**	**23.25**	**0.00**
TPMU	**14.85**	**0.00**	**LNEM**	**48.96**	**0.00**
ANEM	**44.03**	**0.00**	**LNDE**	**3.92**	**0.03**
ANDE	**55.95**	**0.00**	**MNEM**	**136.27**	**0.00**
BNEM	**4.64**	**0.01**	**MNDE**	**19.54**	**0.00**
BNDE	1.78	0.18	**NNEM**	**16.66**	**0.00**
CNEM	**25.97**	**0.00**	**NNDE**	**4.05**	**0.02**
CNDE	**12.11**	**0.00**	**ONEM**	**4.19**	**0.02**
DNEM	**4.55**	**0.02**	**ONDE**	**3.56**	**0.04**
DNDE	1.15	0.32	**PNEM**	**4.49**	**0.02**

Table 4. *Cont.*

ANOVA					
Variables	**F**	**Sig.**	**Variables**	**F**	**Sig.**
ENEM	**25.94**	**0.00**	PNDE	1.06	0.35
ENDE	1.16	0.32	**QNEM**	**47.80**	**0.00**
FNEM	**10.26**	**0.00**	QNDE	1.08	0.35
FNDE	**6.64**	**0.00**	**RNEM**	**13.20**	**0.00**
GNEM	**23.97**	**0.00**	RNDE	2.05	0.14
GNDE	**7.20**	**0.00**	**SNEM**	**34.57**	**0.00**
HNEM	**43.88**	**0.00**	SNDE	1.79	0.18
HNDE	2.69	0.08	**UNEM**	**4.95**	**0.01**
INEM	0.71	0.50	UNDE	2.96	0.06
INDE	**4.91**	**0.01**			

The clusters obtained and the communes belonging to each of them for the Metropolitan Region of Santiago, when considering only the segmenting variables (identified in Table 4), are presented in Figure 2, together with the difference in the average value of the social variables of the communes in each cluster, see Table 5. Cluster one differs in that it is composed of the communes with the highest values for the variables PPZU and PAES and the lowest values for the variables PPIO, TPIN and TPMU. Cluster two has the lowest values for the variables PPZU and PAES and high values for the variables TPIN and TPMU. Cluster three has high values for the variables PPZU, TPIN and TPMU.

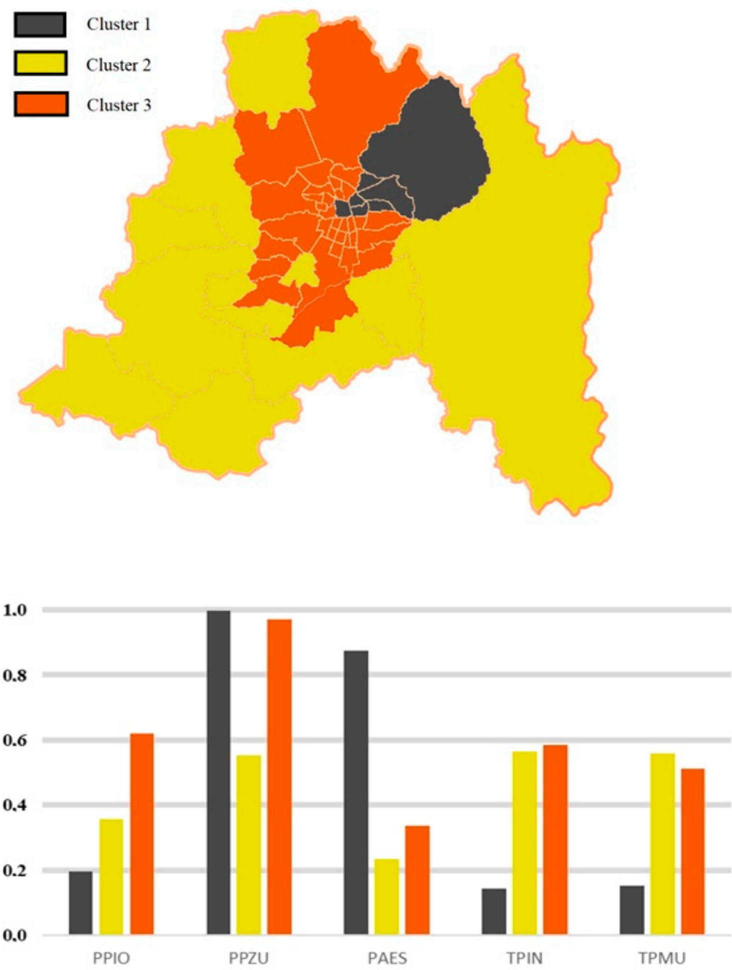

Figure 2. Spatial distribution and the average value of standardized social variables for each cluster in the Santiago Metropolitan Region. Source: own elaboration.

Table 5. Shows the center of the clusters of the region for each variable (the average of the variable based on the communes that make up the cluster), highlighting in bold the segmenting variables.

Variables	Cluster			Variables	Cluster		
	1	2	3		1	2	3
PPIO	**0.1960**	**0.3572**	**0.6195**	JNDE	0.0861	0.0759	0.0220
PPZU	**0.9963**	**0.5537**	**0.9700**	**KNEM**	**0.5116**	**0.1099**	**0.0769**
PAES	**0.8750**	**0.2331**	**0.3362**	**KNDE**	**0.5183**	**0.0417**	**0.0888**
TPIN	**0.1419**	**0.5632**	**0.5847**	**LNEM**	**0.6982**	**0.2322**	**0.1914**
TPMU	**0.1515**	**0.5570**	**0.5109**	**LNDE**	**0.2908**	**0.0729**	**0.1158**
ANEM	**0.0350**	**0.3745**	**0.0313**	**MNEM**	**0.8370**	**0.1370**	**0.2418**
ANDE	**0.1720**	**0.6224**	**0.0930**	**MNDE**	**0.4636**	**0.0424**	**0.1227**
BNEM	**0.2232**	**0.3854**	**0.0947**	**NNEM**	**0.7667**	**0.3556**	**0.5217**
BNDE	0.0167	0.0250	0.1161	**NNDE**	**0.4556**	**0.1791**	**0.2712**
CNEM	**0.1429**	**0.2925**	**0.6076**	**ONEM**	**0.2143**	**0.4167**	**0.0303**
CNDE	**0.1506**	**0.2216**	**0.4738**	**ONDE**	**0.1523**	**0.3138**	**0.0796**
DNEM	**0.3175**	**0.1667**	**0.1380**	**PNEM**	**0.6190**	**0.4815**	**0.6785**
DNDE	0.2063	0.1015	0.0550	PNDE	0.1087	0.2078	0.1977
ENEM	**0.0909**	**0.4848**	**0.1625**	**QNEM**	**0.7000**	**0.0964**	**0.1710**
ENDE	0.0292	0.0931	0.1467	QNDE	0.0897	0.1420	0.0507
FNEM	**0.1911**	**0.2489**	**0.5029**	**RNEM**	**0.6984**	**0.3843**	**0.4596**
FNDE	**0.5849**	**0.2555**	**0.4943**	RNDE	0.2308	0.1244	0.0308
GNEM	**0.2088**	**0.4103**	**0.6792**	**SNEM**	**0.3235**	**0.3444**	**0.6386**
GNDE	**0.5023**	**0.2052**	**0.5401**	SNDE	0.0880	0.2201	0.1007
HNEM	**0.1185**	**0.6052**	**0.7088**	TNEM	0.0000	0.0000	0.0000
HNDE	0.0688	0.2053	0.2174	TNDE	0.0000	0.0000	0.0000
INEM	0.1126	0.1771	0.1553	**UNEM**	**0.0000**	**0.1667**	**0.0000**
INDE	**0.3492**	**0.1950**	**0.1373**	UNDE	0.1250	0.0000	0.0000
JNEM	**0.5506**	**0.0955**	**0.1970**				

The general result of the cluster analysis is the relationship between the variables PPZU (the percentage of the population living in urban areas) and PAES (average years of schooling), i.e., the higher the percentage of the population living in urban areas, the higher the average years of schooling of the population. In the values taken for the variables in clusters one and three, it can be noted that the communes in both clusters have a similar high percentage of their population living in urban areas. Still, cluster one is composed of communes with a lower rate of income and multidimensional poverty. In contrast, cluster three is composed of communes a large part of whose population lives in urban areas, with a higher rate of both income and multidimensional poverty. Another important relationship that can be distinguished in this region (and in general in all regions of the country) is that the higher the average level of schooling of the population, the lower the poverty rate (income and multidimensional). In this case, cluster one has the highest value for the PAES variable (the percentage of the population living in urban areas) and the lowest for the TPIN (income poverty rate) and TPMU (multidimensional poverty rate) variables.

4.3. Results of Applying the Mathematical Programming Model to the Santiago Metropolitan Region

Next, an application of the SED to the Santiago Metropolitan Region is presented for the mathematical programming model for districting and seat allocation seen in Section 3.1. The TSED is incorporated into the model as a constraint, restricting its value to a tolerable maximum. First, the model is solved without restricting the TSED. Then, results are obtained by solving the model and restricting the value of TSED with upper bounds lower than the tolerable maximum. On the other hand, the malapportionment at the regional level (hereafter MALR) is used to measure the demographic balance, which is calculated considering only the population and the current number of seats in the region.

The application considers the current number of districts (7) and deputies (47). The cases are presented are detailed below:

(a) Base case: Current district design of the Metropolitan Region;
(b) Case 1: Solution of the model without restricting the TSED value;
(c) Case 2: Solution of the model restricting the TSED value to be less than or equal to 420;
(d) Case 3: Solution of the model restricting the TSED value to be less than or equal to 380;
(e) Case 4: Solution of the model restricting the TSED value to be less than or equal to 360;
(f) Case 5: Solution of the model restricting the TSED value to be less than or equal to 348.

It can be seen in Figure 3 that it is possible to improve both the TSED and MALR. From cases 1 to 5, the MALR is reduced by 64.7% on average, which is a significant improvement in the demographic balance. In case 5, it was possible to improve the TSED by 25.7%, which was the best value found.

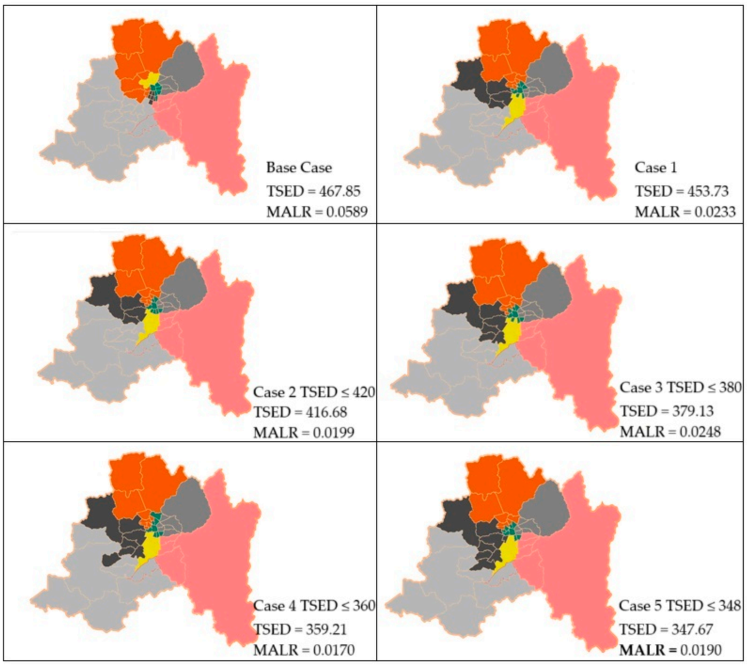

Figure 3. District maps of cases in the Santiago Metropolitan Region. Source: own elaboration.

Table 6 presents the average, maximum, minimum and range values for the TSED and MALR, calculated on the basis of the districts formed in each case. Apart from the improvement in the TSED, a decrease in the maximum and the range is achieved. As for the MALR, in all cases (1 to 5), a significant decrease in the maximum, minimum and range, respectively, is achieved.

Table 6. Statistics and total values for TSED and MALR. Source: own elaboration.

		Base Case	Case 1	Case 2	Case 3	Case 4	Case 5
TSED	Average	66.84	64.82	59.53	54.16	51.32	49.67
	Max	235.48	148.84	148.84	97.70	75.42	71.36
	Min	20.54	9.88	9.88	15.31	15.31	22.61
	Range	214.94	138.96	138.96	82.39	60.11	48.75
	Total	467.85	453.73	416.68	379.13	359.21	347.67
MALR	Average	0.0168	0.0067	0.0057	0.0071	0.0049	0.0054
	Max	0.0347	0.0128	0.0101	0.0191	0.0101	0.0107
	Min	0.0104	0.0011	0.0027	0.0011	0.0001	0.0015
	Range	0.0244	0.0117	0.0074	0.0181	0.0100	0.0092
	Total	0.0589	0.0233	0.0199	0.0248	0.0170	0.0190

5. Conclusions

The methodology proposed in this article, based on the socio-economic dissimilarity distance (SED) and a mathematical programming model, allows for the creation of more socio-economically homogeneous electoral districts with a better demographic balance.

The objective of this article is to provide a reference methodology for obtaining districts formed by more socio-economically homogeneous communes based on the socio-economic dissimilarity distance (SED). Subsequently, this distance is incorporated into a mathematical programming model for districting and seat allocation, which includes multiple criteria of the districting process, such as demographic balance, contiguity, compactness, respect for natural and non-natural boundaries and geographic units. This methodology is developed in two parts. First, the SEDs between each pair of communes are determined using the socio-economic variables that were found to be significant through the cluster analysis. The degree of detail of the data is at the communal level since the aim is to determine the communes that should be grouped into the same district, and it also agrees with the minimum geographic units used in the model for the district and seat allocation.

This methodology was applied to the Santiago Metropolitan Region, which is currently composed of 52 communes and has 7 districts and 47 seats for male and female deputies. The results of the cluster analysis identified three clusters that correspond to three defined socio-economic realities in the region.

A relationship is observed between the variables PPZU (the percentage of the population living in urban areas) and PAES (average years of schooling): the higher the percentage of the population living in urban areas, the higher the average years of schooling of the population. In the values taken by the variables in clusters one and three, it can be noted that the communes in both clusters have a similarly high percentage of their population living in urban areas. However, conglomerate one is composed of communes with a lower rate of income and multidimensional poverty. In contrast, cluster three is composed of communes a large part of whose population lives in urban areas, with a higher rate of both income and multidimensional poverty.

Another important relationship that can be distinguished in this region (and in general in all regions of the country) is that the higher the average level of schooling of the population, the lower the poverty rate (income and multidimensional). In this case, cluster one has the highest value for the PAES variable and the lowest values for the TPIN (income poverty rate) and TPMU (multidimensional poverty rate) variables.

The second part of this methodology incorporates the SED into a mathematical programming model for district and seat allocation, which is applied to the Santiago Metropolitan Region, where an improvement is obtained in both the total SED (TSED) and the demographic balance, measured using the regional malapportionment (MALR), with respect to the current district.

From cases 1 to 5, the MALR is reduced by 64.7% on average, which is a significant improvement in the demographic balance. In case (5), the TSED was improved by 25.7%, which was the best value found by the model. This shows that the incorporation of the socio-economic homogeneity criterion into the model of districting and seat assignment leads to a better population balance and more homogeneous districting in socio-economic terms.

Currently, some senators with different political tendencies have presented reform to the current political system which proposes that the Board of Directors of the Electoral Service must update the allocation of deputy seats among the established districts every ten years—in other words, the reallocation of seats to the districts. Although this is an advance with respect to having static allocation of the seats across time, this study would allow for both the districts and the seats assigned to them to be modified.

Author Contributions: Conceptualization: R.R., M.R. and Ó.C. Methodology: R.R., M.R., Ó.C., F.B. and C.O. Software: R.R., M.R. and C.O. Validation: R.R., M.R., Ó.C. and C.O. Formal analysis: R.R., M.R., Ó.C., F.B. and C.O. Investigation: R.R., M.R., Ó.C., F.B. and C.O. Data curation: R.R. and M.R. Writing—original draft preparation: R.R., M.R. and Ó.C. Writing—review and editing: R.R., M.R., Ó.C., F.B. and C.O. Visualization: R.R., M.R., Ó.C. and C.O. Supervision: R.R. and Ó.C. Project administration: R.R., M.R. and C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Vicerrectoría de Investigación y Doctorados de la Universidad de San Sebastián-Fondo USS-FIN-25-APCS-08.

Data Availability Statement: The data presented in this study are openly available in reference numbers [41–43].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guevara, G. *Proporcionalidad Territorial en el Mapa Electoral Chileno: Una Aproximación a la Geografía de la Representación a Partir de la Elección Parlamentaria 2005 (Tesis de Pregrado)*; Universidad de Chile: Santiago, Chile, 2007.
2. Soto, M.P. *Sistematización de Factores Socioeconómicos en el Proceso de Generación de Escenarios de Redistritación Electoral (Tesis de Postgrado)*; Universidad de Quintana Roo: Quintana Roo, Mexico, 2017.
3. Garg, N.; Gurnee, W.; Rothschild, D.; Shmoys, D. Combatting gerrymandering with social choice: The design of multi-member districts. In Proceedings of the 23rd ACM Conference on Economics and Computation, Boulder, CO, USA, 11–15 July 2022; pp. 560–561.
4. Borodin, A.; Lev, O.; Shah, N.; Strangway, T. Big City vs. the Great Outdoors: Voter Distribution and How It Affects Gerrymandering. *IJCAI* **2018**, 98–104.
5. Chen, J.; Cottrell, D. Evaluating partisan gains from Congressional gerrymandering: Using computer simulations to estimate the effect of gerrymandering in the US House. *Elect. Stud.* **2016**, *44*, 329–340. [CrossRef]
6. Duque, J.C.; Anselin, L.; Rey, S.J. El problema de las regiones max-p. *Rev. Cienc. Reg.* **2012**, *52*, 397–419.
7. Legendre, P. Constrained clustering. *Dev. Numer. Ecol.* **1987**, 289–307.
8. Maravalle, M.; Simeone, B. A spanning tree heuristic for regional clustering. *Commun. Stat. Theory Methods* **1995**, *24*, 625–639. [CrossRef]
9. Byfuglien, J.; Nordgard, A. Region-building—A comparison of methods. *Nor. J. Geogr.* **1973**, *27*, 127–151. [CrossRef]
10. Chávez, G. *Aplicación del Districting Problem Para la Reforma al Sistema Binominal (Tesis de Pregrado)*; Universidad Católica de la Santísima Concepción: Concepción, Chile, 2016.
11. Escandón, M.; López, L. Integridad de comunidades y redistritación urbana. *Política Cult.* **2003**, *19*, 45–63.
12. Reynoso, V. Notas para una geografía electoral del estado de Sonora. *Rev. Cienc. Soc. Humanidades* **1991**, *11*, 87–116.
13. Fernández, O. El análisis de Clúster: Aplicación, interpretación y validación. *Rev. Sociol.* **1991**, *37*, 65–76. [CrossRef]
14. Rebolledo, R.; Ulloa, A.; Cornejo, Ó.; Obreque, C.; Baesler, F. Optimizing Districting and Seat Allocation for Enhanced Representativeness in Chile's Chamber of Deputies. *Mathematics* **2024**, *12*, 3878. [CrossRef]
15. Celemín, J. Asociación espacial entre fragmentación socioeconómica y ambiental en la ciudad de mar del plata, Argentina. *EURE* **2012**, *38*, 33–51. [CrossRef]
16. Aguirre, P. Sistemas Políticos y Electorales Contemporáneos Canadá. 2001. Available online: https://portalanterior.ine.mx/documentos/DECEYEC/canada.htm (accessed on 18 September 2024).
17. Navia, P.; Rojas, P. Representación y tamaño de los distritos electorales en Chile, 1988–2002. *Rev. De Cienc. Política* **2005**, *25*, 91–116. [CrossRef]
18. López, L.; Soto, E. Federalismo y redistritación electoral en México. *Política Cult.* **2008**, *29*, 125–147.
19. Baños, M.; Palacios, C. Evolución territorial de los distritos electorales federales uninominales, 1977–2010. *Investig. Geográficas* **2014**, *84*, 81–95.
20. Loosemore, J.; Hanby, V. The Theoretical Limits of Maximum Distortion: Some Analytic Expressions for Electoral Systems. *Br. J. Political Sci.* **1971**, *1*, 467–477. [CrossRef]
21. Reynoso, D. Distritos y escaños en Brasil: Antecedentes, diagnóstico y consecuencias partidarias del malaporcionamiento. *Política Gob.* **2001**, *8*, 167–191. Available online: http://repositorio-digital.cide.edu/handle/11651/1954 (accessed on 18 September 2024).
22. López, L. Redistritación electoral en México: Logros pasados y restos futuros. *Investig. Geográficas* **2006**, *61*, 99–113. [CrossRef]
23. Vilalta, C. Los procesos de distritación electoral y el uso del criterio de comunidad de interés. *Estud. Demográficos Urbanos* **2002**, *49*, 151–173. [CrossRef]

24. Gurnee, W.; Shmoys, D.B. Fairmandering: A column generation heuristic for fairness-optimized political districting. In *SIAM Conference on Applied and Computational Discrete Algorithms (ACDA21)*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2001; pp. 88–99.
25. Lewenberg, Y.; Lev, O.; Rosenschein, J.S. Divide and conquer: Using geographic manipulation to win district-based elections. In Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems, Sao Paulo, Brazil, 8–12 May 2017; pp. 624–632.
26. Levin, H.A.; Friedler, S.A. Automated congressional redistricting. *ACM J. Exp. Algorithmics* **2019**, *24*, 1–24. [CrossRef]
27. Fischer, M. Taxonomía regional. Una comparación de algunos jerárquicos y estrategias no jerárquicas. *Cienc. Reg. Econ. Urbana* **1980**, *10*, 503–537. [CrossRef]
28. Orr, D. Congressional Redistricting: The North Carolina Experience. 1970. Available online: https://arizona-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=01UA_ALMA21436993870003843&context=L&vid=01UA (accessed on 18 September 2024).
29. Fragoso, R.; Rego, C.; Bushenkov, V. Clustering of Territorial Areas: A Multi-Criteria Districting Problem. *J. Quant. Econ.* **2016**, *14*, 179–198. [CrossRef]
30. Velázquez, G. Viejas inequidades, nuevos dimensionamientos: "centros" y "periferias" como elementos para el análisis geográfico del bienestar de la población argentina luego del censo de 2001. G. Velázquez & N. Formiga (coords.), Calidad de vida, diferenciación socioespacial y condiciones sociodemográficas. *Aportes Para Su Estud. En La Argent.* **2008**, *23*, 47.
31. Bourjolly, J.M.; Laporte, G.; Rousseau, J.M. Decoupage Electoral Automatise: Application A L'Lle De Montreal. *INFOR Inf. Syst. Oper. Res.* **1981**, *19*, 113–124. [CrossRef]
32. Bozkaya, B.; Erkut, E.; Laporte, G. A tabu search heuristic and adaptive memory procedure for political districting. *Eur. J. Oper. Res.* **2003**, *144*, 12–26. [CrossRef]
33. Swamy, R.; King, D.M.; Ludden, I.G.; Dobbs, K.W.; Jacobson, S.H. A practical optimization framework for political redistricting: A case study in Arizona. *Socio-Econ. Plan. Sci.* **2024**, *92*, 101836. [CrossRef]
34. Álvarez-Miranda, E.; Campos-Valdés, C.; Quiroga, M.M.; Moreno-Faguett, M.; Pereira, J. A Multi-Criteria Pen for Drawing Fair Districts: When Democratic and Demographic Fairness Matter. *Mathematics* **2020**, *8*, 1404–1430. [CrossRef]
35. Reyes, C.; Lía, D. Diseño de Algoritmos de Distritaje y Su Aplicación al Caso Chileno Universidad de O'Higgins. 2024. Available online: https://repositorio.uoh.cl/handle/611/994 (accessed on 18 September 2024).
36. Capron, G.; Conzález Arellano, S. Las escalas de la segregación y de la fragmentación Urbana. Trace. *Trav. Rech. Dans Les Amériques DuCent.* **2006**, *49*, 65–75.
37. Núñez-Colín, C.A.; Escobedo-López, D. Uso correcto del análisis clúster en la caracterización de germoplasma vegetal. *Agron. Mesoam.* **2011**, *22*, 415–427. [CrossRef]
38. Villardón, J.L.V. *Introducción al Análisis de Clúster*; Departamento de Estadística, Universidad de Salamanca: Salamanca, Spain, 2007; 22p.
39. Gómez, M. El análisis Clúster en investigación de marketing: Metodología y crítica. *ESIC Mark.* **2002**, *113*, 9–33.
40. Vega-Dienstmaier, J.; Arévalo-Flores, J. Clasificación mediante análisis de conglomerados: Un método relevante para la psiquiatría. *Rev. Neuro-Psiquiatr.* **2014**, *77*, 31–39. [CrossRef]
41. Servicio de Impuestos Internos. Estadísticas de Empresas por Comuna y Rubro Económico. 2019. Available online: http://www.sii.cl/sobre_el_sii/estadisticas_de_empresas.html (accessed on 18 September 2024).
42. Instituto Nacional de Estadísticas. Resultados Censo 2017—Personas. 2019. Available online: http://resultados.censo2017.cl/ (accessed on 18 September 2024).
43. Ministerio de Desarrollo Social y Familia. Estimaciones de Tasa de Pobreza por Ingresos y Multidimensional por Comunas, 2017. 2019. Available online: http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/datos_pobreza_comunal.php (accessed on 18 September 2024).
44. Malhotra, N.K. *Investigación de Mercados*, Quinta ed.; Pearson Educación: Guadalajara, Mexico, 2008.