Uma imagem com Tipo de letra, logótipo, símbolo, Gráficos

Descrição gerada automaticamente

Uma imagem com Saturação de cores, Lilás

Descrição gerada automaticamente

**Comparing similarity between two territory partitions in political districting problems**

**Indices and practical issues**

**Sene Conté**

Thesis to obtain the Master of Science Degree in

**Computer Science and Engineering**

Supervisors: Prof. Luís Manuel Silveira Russo Prof. José Rui De Matos Figueira

**Examination Committee**

Chairperson: Prof. Pedro Tiago Gonçalves Monteiro

Supervisor: Prof. Luís Manuel Silveira Russo

Member of the Committee: Prof. Vasco Miguel Gomes Nunes Manquinho

**(Month Year)**

**Declaration**

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

**Abstract**

This study addresses the complex problem of assessing similarity between two partitions of a district map, where each partition is modeled as a connected, undirected, and planar graph. Comparing similarity between such partitions is crucial in applications like urban planning, political districting, and geographic analysis, where variations in partitioning can have significant impacts on representation and resource distribution. Traditional similarity indices, such as the Rand Index, Fowlkes-Mallows Index, and Jaccard Index, are used as foundational measures. However, these indices are extended to incorporate weighting factors for districts, allowing for a more accurate comparison by accounting for variations in district characteristics, such as population density or geographic importance. By using weighted similarity indices, this research aims to achieve a refined measurement of similarity that captures the nuanced impact of each zone. This approach offers an enhanced framework for partition analysis, improving upon existing techniques that often overlook regional disparities. Ultimately, the study presents a comprehensive solution for similarity assessment in partitioned maps, applicable to various fields that rely on accurate and equitable territorial analysis.

**Keywords**

Districting; Weighted Pair Counting; Similarity Indices; Urban Planning; Territorial Maps; Perturbations

**Resumo**

Este estudo aborda o problema complexo de avaliar a semelhança entre duas partições de um mapa distrital, onde cada partição é modelada como um grafo conectado, não-direcionado e planar. Comparar a semelhança entre estas partições é crucial em áreas como o planeamento urbano, a demarcação de distritos eleitorais e a análise geográfica, onde diferentes partições podem ter impactos significativos na representação e na distribuição de recursos. Utilizam-se índices de semelhança tradicionais, como o Índice de Rand, o Índice Fowlkes-Mallows e o Índice de Jaccard, como medidas fundamentais. No entanto, estes índices foram estendidos para incorporar fatores de peso dos distritos, permitindo uma comparação mais precisa ao considerar variações nas características dos distritos, como densidade populacional ou importância geográfica. Através do uso de índices de semelhança pesadas, esta investigação visa obter uma medição refinada da semelhança, capturando o impacto específico de cada zona. Esta abordagem oferece uma estrutura melhorada para a análise de partições, superando técnicas existentes que frequentemente ignoram as disparidades regionais. Por fim, o estudo apresenta uma solução abrangente para a avaliação de semelhança em mapas particionados, aplicável a diversas áreas que dependem de uma análise territorial precisa e equitativa.

**Palavras Chave**

Distritos; Contagem de Pares Pesadas; Índices de Similaridade; Planeamento Urbano; Mapas Territoriais; Perturbações.

# 

# **Introduction**

* 1. **Political Districting**

Political districting is the complex task of dividing a region into distinct territorial units or districts, usually for electoral, administrative, or service provision purposes. The goal is to create balanced, representative districts that satisfy certain criteria, such as equal population, compactness, contiguity, and, importantly, a fair distribution that prevents gerrymandering. Despite these objectives, achieving optimal districting is computationally challenging, especially for large-scale regions, which often leads researchers to employ heuristic methods like Local Search (LS) and Simulated Annealing (SA) to find feasible, if not exact, solutions. These challenges underscore the critical need for well-defined districting approaches, as highlighted in Silva’s work on integrality and contiguity in political maps​​ [1].

Districting challenges are inherently multi-criteria, aiming to create “homogeneous” zones that satisfy demographic, geographical, and sometimes even ecological constraints. For example, districting applications range widely, from defining electoral boundaries to organizing police patrol areas or public service zones. Each of these cases involves grouping elementary units into contiguous clusters that reflect a balanced mix of attributes, thus forming a range of viable solutions that often represent a compromise between conflicting objectives. The search for optimal solutions in such scenarios is generally replaced with an effort to find non-dominated solutions, which are judged based on criteria like homogeneity, geographical continuity, and other region-specific constraints​ [2].

In police districting, the objective is to design patrol sectors that distribute workload evenly across areas, considering factors like response times and operational efficiency. Initially, police sectors were manually drawn based on visible geographical features, but modern approaches incorporate Geographic Information Systems (GIS) and predictive analytics to optimize workload balance and patrol efficiency. Recent research emphasizes the importance of automation and the integration of data-driven methods, which can help create well-balanced, contiguous patrol zones and reduce response times across sectors​ [3].

Political districting’s challenges and solutions reflect broader socio-political implications, as equitable district boundaries directly impact democratic representation, service accessibility, and public trust in governance. Advances in mathematical modeling and computational heuristics, such as evolutionary algorithms and local search methods, continue to shape this field, helping decision-makers explore a diverse array of districting solutions that can balance population distribution with geographical constraints. These tools play a significant role in making districting an increasingly transparent and representative process across applications.

A critical aspect of political districting, particularly when revising or comparing existing maps, is the comparison of alternative partitions. Comparing two territory partitions is essential for evaluating changes in district boundaries, assessing their alignment with demographic shifts, and ensuring that new configurations meet fairness and representativeness standards. Pereira et al. (2009) introduced indices for comparing partitions, including compatibility, inclusion, and distance measures. These indices provide a systematic way to assess how closely an alternative map resembles an existing one, offering insight into the differences between configurations that may have socio-political or operational implications. This method is particularly valuable in evaluating proposed political districting maps against established ones to ensure continuity or manage gradual transitions in boundary adjustments [4].

In this thesis, we focus on the comparative analysis of political districting solutions, specifically examining how two territorial partitions can be evaluated for similarity. By leveraging the methodologies discussed above, we aim to develop a robust framework for comparing political district configurations. This comparative approach will contribute to a deeper understanding of how district boundaries evolve and how new proposals impact existing social, political, and operational contexts.

* 1. **Problem Modeling**

In political districting (PD), the challenge of partitioning a given territory into distinct, cohesive zones, or districts, is critical for ensuring fair representation and efficient public service provision. Each district, designed to meet demographic, geographic, and socio-economic criteria, must balance population, maintain geographical contiguity, and optimize compactness. Beyond initial partitioning, a significant aspect of PD involves comparing alternative district maps to assess similarity. Such comparisons are essential when proposing new maps, as they allow decision-makers to evaluate how well new districts align with or diverge from existing configurations, taking into account both demographic and spatial factors.

To model this problem, we represent the territory as a contiguity graph , where is the set of vertices, each representing an indivisible elementary unit of territory (such as a municipality), and is the set of edges, where an edge represents adjacency between two units and . This graph is structured to be connected, undirected, and planar, which ensures that the entire territory forms a single contiguous entity, that each adjacency is reciprocal, and that it can be represented without overlapping edges, reflecting realistic territorial boundaries.

Within this framework, a **zone** is defined as a subset of contiguous elementary units, while a **partition**  is a collection of zones that collectively cover the territory without overlap, forming a cohesive district map. In comparing two different partitions, we denote one as and the other as .

For each elementary unit there is one and only one zone such that belongs to . For the sake of simplicity, an elementary unit, is also represented by its index . *Figure 1*(a) represents a territory composed of 16 elementary units, divided into four zones, . *Figure 1*(b) shows the contiguity graph G corresponding to the territory of *Figure 1*(a). To facilitate meaningful comparisons, we assign characteristic values to each elementary unit based on a specific property (e.g., population, area, or economic output). For each unit , the **characteristic function** returns the value of a chosen attribute. Using this, we define the **zone sum ,** representing the total characteristic value across all units in a zone , as . The **total characteristic** for the entire territory is then , enabling us to calculate the **zone weight .** This weight provides a basis for weighted comparisons, allowing zones with higher demographic or economic significance to contribute proportionally to the similarity assessment.

The following definitions establish the basis for comparing two partitions in terms of their zones and their constituent units. First, an **attribute** in a contiguity graph is a real-valued function defined on with non-negative values. For any subset , the characteristic sum denotes the overall attribute value for that subset, and is the overall attribute for as a whole. This function is fundamental for evaluating and comparing the similarity of characteristics between zones and partitions.

Two zones and are considered **equal** with respect to an attribute if each is included within the other for the attribute under consideration, denoted . Similarly, two partitions and are considered **equal** with respect to the attribute if, for every pair of zones , either or the two zones do not overlap in terms of the characteristic. This definition of equality ensures that both partitions are equivalent in terms of their overall structure and the distribution of the attribute across zones.

A grid of circles and numbers

Description automatically generated with medium confidence

Figure 1. A territory and the associated contiguity graph.

A pair of squares with lines and letters

Description automatically generated with medium confidence

Figure 2. Four different partitions, Y, Y', Y'', and Y’’’.

**1.3 Motivation**

In the realm of political districting, ensuring fair and representative territorial partitions is critical to democratic processes, public administration, and resource allocation. The need to compare and evaluate territorial partitions arises frequently, particularly in scenarios involving electoral redistricting, urban planning, or public service zoning. A partition comparison allows decision-makers to assess the extent to which a proposed or alternative configuration aligns with historical or current partitions. This is particularly important in preserving continuity, minimizing disruptions, and ensuring equity across districts. The research of Tavares Pereira et al. emphasizes the socio-economic importance of comparing partitions to address discrepancies and optimize decision-making frameworks [4].

The challenge of comparing partitions extends beyond political applications. The Adjusted Rand Index (ARI) and other pair-counting methods have become standard tools in cluster analysis and unsupervised machine learning, where they serve as benchmarks for validating clustering algorithms. These indices provide insights into the alignment of clusters across diverse applications, from biology to computational linguistics​ [5]. In particular, the evolution of indices to accommodate asymmetric and weighted scenarios reflects the growing complexity of real-world partition comparison problems.

Weighted comparisons are indispensable in many fields where certain attributes carry more significance than others. In political districting, for instance, population balance and geographical contiguity often take precedence over compactness. Similarly, in sports, weight measures are used to rank players and teams. For instance, let’s take an example of European Football, The European Golden Shoe, also known as European Golden Boot, is an award that is presented each season to the leading goal scorer in league matches from the top division of European national leagues. It has been calculated using a weighting in favor of the highest ranked leagues. Between 1968 and 1991, the award was given to the highest goal scorer in any European league, regardless of the strength of the league in which they played. But following some incidents, since 1996-97 season, the award has been awarded based on a point system that allows players in tougher leagues to win even if they score fewer goals than a player in a weaker league. The weightings are determined by the league’s clubs in European competitions. Goals scored in the top five leagues (see [6]) according to the UEFA coefficient list are multiplied by factor of two, goals scored in the leagues ranked 6 to 22 are multiplied by factor of 1.5 and goals scored in the leagues ranked 22 and above are multiplied by factor of 1 [7].

The use of weights in neural networks are pivotal for learning, optimization, and decision-making. They determine the influence of each input feature on the model's predictions, effectively guiding the learning process. Han et al. (2015) highlighted the profound impact of learning both weights and connections for improving the efficiency of neural networks. Their work demonstrated that by pruning low-weight connections, those deemed less significant—the computational cost of neural networks can be reduced by orders of magnitude without compromising accuracy. For example, in the case of AlexNet, they achieved a ninefold reduction in parameters, making neural networks more accessible for deployment on resource-constrained devices like mobile systems. This illustrates the power of weights in maintaining a balance between performance and computational efficiency, a principle that resonates with the need for weighted analysis in comparing territorial partitions [8].

Beyond efficiency, weights also enable neural networks to address real-world challenges such as imbalanced data distributions. In class-imbalanced learning, the dynamically weighted loss functions proposed by Wei-Dong et al. ensure that underrepresented classes receive greater attention during training. This approach not only reduces bias but also enhances the network's confidence calibration, leading to fairer and more reliable predictions. This methodology underscores the adaptability and significance of weights in handling diverse scenarios, from image classification to complex societal problems​ [9].

These diverse applications illustrate the profound impact of weighted comparisons across domains. The ability to incorporate weights into pair-counting indices not only enriches the analysis but also makes it more adaptable to specific contexts and objectives. By extending and refining these tools for political districting, this research aims to provide a robust framework for comparing territorial partitions, addressing both the intrinsic challenges of districting and the broader applicability of weighted similarity measures.

**1.4 Organization of the document**

|  |  |
| --- | --- |
| **Notation** | **Description** |
|  | Territory, where each represents an indivisible elementary unit. |
|  | A zone, defined as a set of contiguous elementary units (e.g., municipalities). |
| *γ* | A partition or district map of the territory |
|  | Characteristic function returning the value of a specific characteristic (e.g., population, area) for an elementary unit |
|  | Total characteristic value for a zone , calculated as |
|  | Total characteristic value for the entire territory , given by |
|  | Weight of a zone , calculated as . |
|  | Contiguity graph representing territory , where is the set of vertices (elementary units) and is the set of edges (borders between adjacent units). |
|  | Set of vertices in the contiguity graph, representing elementary units. |
| } | Set of edges in the contiguity graph, where represents a border between adjacent units and |

Table 1. Table of Concepts

Pair Counting Indices

Rand Index

Fowlkes Mallows Index

Jaccard Index

Properties of the indices.

Weighted Pair Counting

Mathematical formulation

Validation of formulas

Comparing with traditional pair counting (optional)

Implementation

Algorithm

Data structure

Perturbation algorithm

Complexity analysis

Solution

Experimental results

Data Set Generation and Preprocessing

Results

Conclusion

Conclusions

System Limitation and Future Work

Bibliography

# Bibliography

|  |  |
| --- | --- |
| [1] | J. P. R. G. P. Da Silva, "Master’s Thesis: Integrality, Contiguity, and Absence of Holes in forming Political Districting maps," Lisbon, 2023. |
| [2] | F. T. Pereira, V. Mousseau and J. R. Figueira, "Multiple Criteria Districting Problems, Models, Algorithms, and Applications," *ResearchGate,* 2014. |
| [3] | F. Liberatore, M. Camacho-Collados and B. Vitoriano, "Police Districting Problem," *Springer Nature,* p. 9–29, 2020. |
| [4] | F. T. Pereira, J. R. Figueira, V. Mousseau and B. Roy, "Comparing two territory partitions in districting problems," *Socio-Economic Planning Sciences,* pp. 72-88, 2009. |

Appendix