

Compatibilização de setores censitários baseada em grafos para análise histórica de processos urbanos em escala local

Pedro H. R. Mendonça, Fabio Kon

¹Departamento de Ciência da Computação
IME - Universidade de São Paulo

{pedrorezende, kon}@ime.usp.br

Abstract. *Urban data analysis often relies on census data, but using this kind of data for historical series at a local scale is limited by the incompatibility of census tract grids. In this article, we propose an automated methodology for matching IBGE censuses through the construction of a graph with spatial join operations and non-spatial index comparisons. We adopt an additive heuristic to minimize the impact of the matching process on analytical possibilities. We deploy our methodology to match the 2010 and 2022 census grids for the São Paulo Metropolitan Region and propose metrics to evaluate the results, concluding with possible improvements.*

Resumo. *Análises de dados urbanos recorrem frequentemente a dados censitários, mas o emprego desses dados em escala local para séries históricas é limitado pela incompatibilidade das malhas de setores censitários. Neste artigo, propomos uma metodologia automatizada de compatibilização de censos do IBGE baseada na construção de um grafo a partir de operações de união espacial e comparação de índices não espaciais. Buscamos uma heurística aditiva para reduzir o impacto da compatibilização nas possibilidades de análise. Aplicamos nossa metodologia entre os censos de 2010 e 2022 para a Região Metropolitana de São Paulo e propomos métricas para avaliar os resultados, indicando possíveis aprimoramentos.*¹

1. Introdução

Estudos sobre processos urbanos e avaliação de políticas de planejamento do uso e ocupação do solo frequentemente cruzam informações com dados censitários, pois são uma fonte de dados oficial com grande cobertura temática, alcance nacional e regularidade temporal. Mas análises urbanas que dependem de séries históricas entre censos em escala local, em recortes geográficos menores que um bairro, geralmente enfrentam um desafio metodológico relacionado à compatibilidade entre as bases censitárias de censos diferentes.

A consistência de análises longitudinais sobre dados de pesquisas como os censos requer (I) que a metodologia de coleta e agregação dos dados seja historicamente consistente (equivalência de escopo) e (II) que as unidades espaciais de análise correspondam

¹Esta pesquisa é financiada pela FAPESP, processos 23/00811-0, 23/12851-7 e 24/12863-8. Agradecemos às equipes do LabCidade FAUUSP e do Instituto Peregrum pelas contribuições ao desenvolvimento da metodologia.

espacialmente às mesmas regiões geográficas (equivalência de estrutura espacial). A primeira premissa pode ser assegurada a partir da leitura da documentação de divulgação dos dados, como feito em [Barbosa 2014], e não é objeto deste artigo. Já a equivalência de estrutura espacial depende da escala geográfica e do nível de agregação dos dados divulgados.

Em escala municipal, a equivalência de estrutura espacial costuma ser facilmente atendida: os dados agregados para o mesmo município são comparáveis entre censos, pois a delimitação espacial do município raramente muda. Já na escala mais próxima, dos setores censitários, a correspondência depende de um tipo particular de união de bases de dados. Diferente da implementação de operações de união espacial (*spatial join*) em sistemas gerenciadores de bancos de dados espaciais e em bibliotecas e softwares GIS, a união entre dados de duas malhas de setores censitários depende de uma operação de união baseada não apenas no conjunto de geometrias, mas também, simultaneamente, em índices não espaciais. Fundamentamos esse problema na Seção 3.

Neste artigo, propomos um método para compatibilização de malhas de setores censitários dos censos do Instituto Brasileiro de Geografia e Estatística (IBGE) baseado na codificação e nos polígonos dos setores, com uma heurística aditiva (usando operações de soma e união). Para isso, adaptamos exemplos encontrados na bibliografia para a escala local. Propomos um algoritmo, o implementamos para as malhas dos censos de 2010 e 2022 na Região Metropolitana de São Paulo e apresentamos os resultados para avaliar casos de uso e possibilidades de melhorias futuras para a generalização da metodologia.

2. Trabalhos relacionados

A comparabilidade entre áreas censitárias está associada fundamentalmente à metodologia de cada censo nacional, e os trabalhos relacionados têm, no geral, um perfil técnico. Em escala internacional, identificamos paralelos do censo brasileiro com a estrutura dos censos dos EUA e da Argentina, que também utilizam entidades geográficas equivalentes ao setor censitário (*census tracts* e *radios censales*, respectivamente) e possuem uma extensão territorial da mesma ordem de grandeza que o Brasil.

Nos Estados Unidos, a equivalência de estrutura espacial em escala local pode ser assegurada diretamente usando as “tabelas de comparabilidade” entre setores, disponibilizadas pelo US Census Bureau [US Census Bureau 2024b]. Esses arquivos permitem, por exemplo, a elaboração de visualizações cartográficas da variação populacional [US Census Bureau 2024a].

Para o censo da Argentina, Rodríguez [Rodríguez 2021] propõe e aplica um método de “correção” dos setores entre os censos de 1991, 2001 e 2010 com operações de união espacial. O processo tem como objetivo ajustar imprecisões entre os limites dos setores, usando delimitadores naturais (ruas, rios, perímetros administrativos). O método é ajustado para cada censo, empregando critérios de validação da correção pela distância entre centroides de dois setores ou pela sobreposição visual (com intervenção humana) de bordas de polígonos. O resultado dessa metodologia é disponibilizado como um arquivo oficial do Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet), permitindo a criação de tabelas de compatibilidade sem influência de imprecisões nos ajustes dos polígonos.

Já em contexto brasileiro, trabalhos de compatibilização têm como principais

desafios a imprecisão geográfica no desenho dos setores e a ausência de arquivos digitais com cobertura espacial completa para os censos de 1991 e 2000 [Hirye et al. 2016, Lobo 2009, Reis 2013]. Yamaguchi [Yamaguchi 2017] categoriza os estudos com compatibilização de setores censitários de acordo com o uso de dados auxiliares, como imagens de satélite ou indicadores socioeconômicos. Os trabalhos relatados pela autora recorrem a técnicas de interpolação para reagregar os dados em uma nova malha regular, ou usam “tabelas de compatibilidade” (equivalentes às tabelas de comparabilidade do US Census Bureau) em escalas territoriais maiores.

As tabelas de compatibilidade relacionam operações de redesenho aos recortes censitários envolvidos, permitindo a identificação de áreas correspondentes entre censos e reconstituição da equivalência de estrutura espacial. A autora cita o trabalho desenvolvido pelo IPEA utilizando “Áreas Mínimas Comparáveis” (AMCs) a partir de tabelas de compatibilidade das malhas municipais de 1872 a 2000 [Reis et al. 2008], que repercute em outras publicações do instituto.

O princípio das AMCs é garantir que os dados sejam reagregados para uma malha geográfica consistente com as metodologias empregadas na série histórica. Ao realizar apenas operações de soma de variáveis quantitativas e mesclagem de polígonos, a metodologia de coleta e agregação dos dados originais continua íntegra. Já as metodologias de interpolação adicionam novas camadas de premissas – de que a distribuição de uma variável é homogênea dentro de uma entidade geográfica, por exemplo – para calcular os pesos de interpolação entre as malhas originais e a nova malha regular.

Em termos de implementação desses métodos, a compatibilização por interpolação requer operações entre geometrias (polígonos, pontos) e, a depender do uso de dados auxiliares, verificações manuais das feições resultantes, como no método de Rodríguez para o censo argentino. A implementação das AMCs do IPEA se baseia em operações de álgebra relacional simples, mas requer o processo adicional de criação de tabelas de compatibilidade.

Neste trabalho, buscamos viabilizar a aplicação da heurística aditiva das AMCs para a escala dos setores censitários, por sua simplicidade conceitual e por dispensar premissas adicionais. Essa abordagem é centrada na indexação dos dados a partir das malhas censitárias. Estudos que exijam a agregação dos dados para uma terceira malha geográfica estruturalmente diferente das malhas censitárias (como malhas regulares, por exemplo), costumam depender das premissas da abordagem por interpolação. Nesses casos, a heurística aditiva é uma alternativa pouco efetiva.

3. Fundamentação

O IBGE passou a adotar setores censitários a partir do Censo de 1940 com o propósito original de organização da operação censitária, orientando a coleta pelos recenseadores. Um setor censitário urbano representa uma área que agrega um conjunto de domicílios recenseável por um mesmo agente censitário. A partir do Censo de 2000, a malha geográfica de setores passou a ser divulgada em formato digital, juntamente com os dados agregados por setor censitário, viabilizando o uso amplo de dados censitários em bases de dados geoespaciais e análises demográficas em escala local. Desde então, para além da função operacional, o setor censitário ganhou uma função analítica para diversas áreas do conhecimento.

Nos censos mais recentes, a adoção e aprimoramento de tecnologias de mapeamento digital trouxe melhorias no desenho das malhas, mas, em contrapartida, tornou o desenho dos setores censitários geometricamente incompatível entre censos (Figura 1). Para o Censo de 2022, o IBGE desenvolveu um sistema de redesenho descentralizado da malha censitária com auxílio de imagens de alta resolução, o que permitiu um ganho de precisão do desenho dos setores, mas também agravou as diferenças em relação às malhas censitárias antigas.

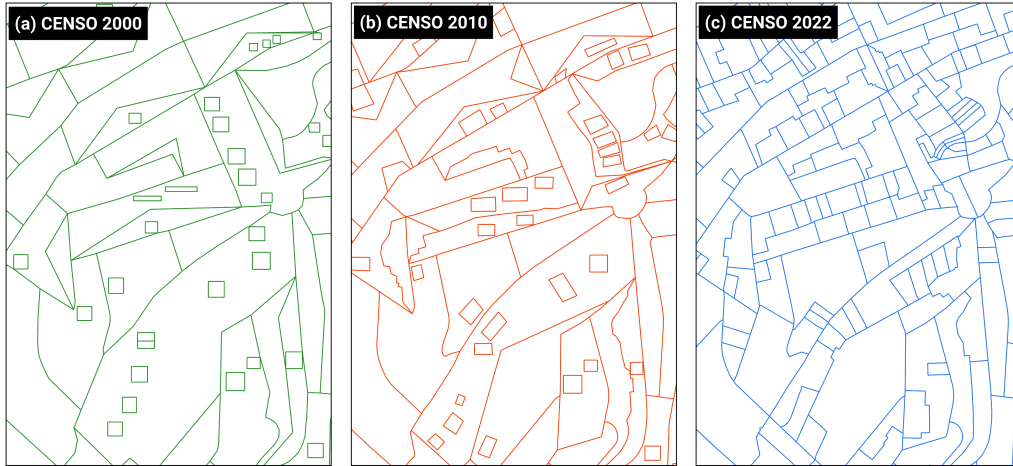


Figura 1. Setores censitários de uma mesma extensão geográfica na região da Bela Vista, centro de São Paulo, nos censos de 2000, 2010 e 2022.

O redesenho de setores censitários é tema recorrente de notas técnicas do IBGE desde os primeiros censos em que foram implementados [IBGE 1980], e é presente também em publicações dos censos dos EUA [US Census Bureau 2024b] e da Argentina [Rodriguez and de Grande 2024]. Os manuais apontam para um conjunto de operações lógicas resultantes da operação de redesenho de um setor censitário, que podemos sistematizar como:

- **Divisão:** quando um setor censitário é dividido em vários ($1 : n, n > 1$);
- **União:** quando se une mais de um setor censitário ($n : 1, n > 1$);
- **Manutenção:** quando o setor é mantido ($1 : 1$);
- **Reconfiguração:** quando não há uma única relação não ambígua de divisão, união ou manutenção ($m : n$, com $m, n > 1$).

No caso dos censos do IBGE, os manuais técnicos trazem ainda uma outra variável vinculada ao redesenho, relacionada ao tipo de setor (urbano ou rural). Neste artigo, desconsideramos aspectos qualitativos do setor censitário. Consideramos o setor censitário como uma entidade geográfica composta apenas por um polígono e um código identificador (“id”).

O id de um setor censitário é uma cadeia numérica estruturada, conforme a Figura 2, cujos primeiros 11 dígitos determinam a filiação do setor censitário a entidades geográficas superiores. Considerando que a codificação das entidades-mãe é consistente entre censos, o redesenho de setores censitários gera correspondências do tipo $r = (I, J)$ onde I e J são conjuntos de setores censitários dos censos C_I e C_J , respectivamente, em



Figura 2. Estrutura do código do setor censitário.

que todo setor $s \in (I \cup J)$ de uma mesma relação r tem sempre os 11 primeiros dígitos idênticos.

Observando particularmente os ids dos setores, isolados de seus polígonos, uma operação de redesenho entre censos gera as possíveis situações:

- **Exclusão:** um id deixa de existir
- **Criação:** um id passa a existir
- **Manutenção:** um id é mantido, e se refere ao mesmo setor censitário
- **Desassociação:** um id existe nos dois censos, mas se refere a entidades geográficas diferentes. Trata-se de um caso particular de dois setores não relacionados que sofreram exclusão e criação.

Os setores censitários são validados por reconhecimento prévio das localidades por agentes censitários, que podem modificá-los de acordo com necessidades operacionais. Situações de reconfiguração de polígonos ou desassociação de ids não são indicadas como boas práticas nos manuais técnicos, mas existem. Já a situação de manutenção pressupõe que tanto o polígono quanto o id dos setores entre dois censos representam a mesma entidade geográfica. Numa perspectiva computacional, a compatibilização dos setores censitários de dois censos pode ser dividida em dois problemas fundamentais: (i) identificar casos de manutenção, assegurando que não incluam setores com desassociação de ids, e (ii) identificar corretamente os subconjuntos de setores envolvidos em cada operação de redesenho.

4. Metodologia

A partir da revisão da bibliografia, propomos uma metodologia automatizada para compatibilização dos setores censitários, usando seus polígonos e ids para gerar AMCs de setores censitários e tabelas de compatibilidade. Para isso, empregamos operações de união espacial e comparações entre ids para popular um grafo não direcionado G , que chamamos de grafo de compatibilização. Consideremos duas malhas censitárias M_I do censo I , e M_J do censo J , compostas por setores s . Construímos $G = (V, A)$ com:

$$V = V_I \cup V_J$$

$$V_I = \{s \mid s \in M_I\}$$

$$V_J = \{s \mid s \in M_J\}$$

$$A = \text{CompatMalhas}(V_I, V_J)$$

Onde *CompatMalhas* é uma função que implementa o algoritmo 1. A notação $geom(s)$ representa o polígono do setor censitário s , e $id(s)$ representa o código identificador do setor s . As notações $geom(S)$ e $id(S)$ representam o conjunto de polígonos e de códigos identificadores de um conjunto de setores S . Cada componente do grafo resultante representa uma operação de redesenho, com um conjunto de setores de cada censo implicados nessa operação.

Algoritmo 1: CompatMalhas

Entrada: V_I, V_J
Saída: Conjunto A de arestas do tipo $s_I s_J, s_I \in V_I, s_J \in V_J$

```
1  $A, C \leftarrow \emptyset;$   
  // Arestas de manutenção  
2 para cada  $s_I \in V_I$  faça  
3   se  $id(s_I) \in id(V_J)$  então  
4      $s_J \leftarrow$  setor em  $V_J$  com  $id = id(s_I);$   
5     se  $s_I$  e  $s_J$  estiverem na mesma localização absoluta e relativa e  
      conservarem seus olgónos então  
6        $A \leftarrow A \cup \{s_I s_J\};$   
7        $C \leftarrow C \cup \{s_I, s_J\};$   
8     fim  
9   fim  
10 fim  
11 para cada  $s_J \notin C$  faça  
    // Arestas de manutenção desassociada  
12   se  $geom(s_J)$  intersecciona uma  $geom(s_I)$  de um  $s_I \in V_I$  tal que  
       $id(s_J) = id(s_I)$  então  
13      $A \leftarrow A \cup \{s_I s_J\};$   
14      $C \leftarrow C \cup \{s_I, s_J\};$   
15   fim  
    // Inclusão de arestas de divisão  
16   se  $geom(s_J)$  tem uma parcela de sua área maior que  $L_{min}$  em  $geom(s_I)$  de um  
      setor  $s_I \in V_I$  e a interseção  $geom(s_I) \cap geom(s_J)$  não é residual então  
17      $A \leftarrow A \cup \{s_I s_J\};$   
18      $C \leftarrow C \cup \{s_I, s_J\};$   
19   fim  
20 fim  
    // Redefine os polígonos de  $V_I$  e  $V_J$  com buffer  
21  $geom(V_I) \leftarrow \{buffer(geom(s), B) \mid s \in V_I\};$   
22  $geom(V_J) \leftarrow \{buffer(geom(s), B) \mid s \in V_J\};$   
23  $N \leftarrow \{s \mid s \in V_I \cup V_J, s \notin C\}$   
    // Arestas de sobreposição forçada de  $I$  para  $J$   
24 para cada  $s_I \in (V_I \cap N)$  faça  
25   se  $geom(s_I)$  intersecciona  $geom(s_J)$  de um setor  $s_J \in V_J$  e  
       $geom(s_I) \cap geom(s_J)$  não é residual então  
26      $A \leftarrow A \cup \{s_I s_J\}$   
27   fim  
28 fim  
    // Arestas de sobreposição forçada de  $J$  para  $I$   
29 para cada  $s_J \in (V_J \cap N)$  faça  
30   se  $geom(s_J)$  intersecciona  $geom(s_I)$  de um setor  $s_I \in V_I$  e  
       $geom(s_J) \cap geom(s_I)$  não é residual então  
31      $A \leftarrow A \cup \{s_I s_J\}$   
32   fim  
33 fim  
34 retorna  $A;$ 
```

O algoritmo consiste em três etapas: identificação de manutenções, identificação de divisões e sobreposição forçada. A ordem de aplicação desses procedimentos busca conservar ao máximo a definição dos ids e desenho dos polígonos, criando um fluxo “em funil” que se inicia sem manipulações geométricas, para identificação de manutenções. Os procedimentos seguintes empregam manipulações simples (interseções de polígonos para identificação de divisões) e, então, mais agressivas (interseções com buffer para sobreposição forçada) apenas entre subconjuntos de setores definidos pela conformação lógica da operação de redesenho. Setores que passam por operações de manutenção, por exemplo, não devem ser incluídos na identificação de operações de divisão.

No início da execução do algoritmo, G é um grafo vazio com vértices representando setores censitários de dois censos, nomeados pelo seu id e pelo seu respectivo censo (Figura 3). Durante a execução, correspondências entre setores cujos primeiros 11 dígitos do id não são iguais são consideradas residuais, pois são inconsistentes em hierarquia de codificação.

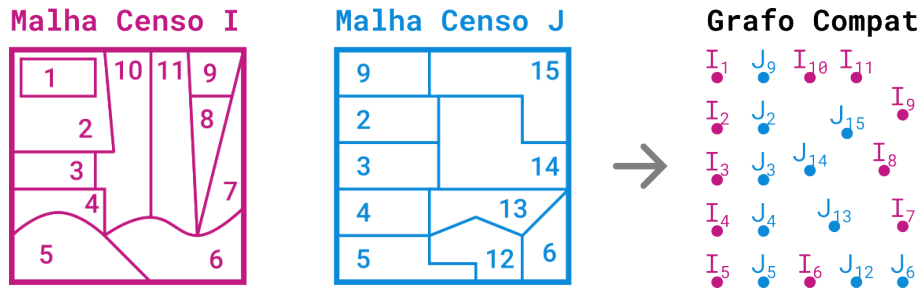


Figura 3. Criação de grafo de compatibilização, não populado.

4.1. Identificação de manutenção

A identificação dos casos de manutenção tem como objetivo identificar de forma unívoca pares de setores s_I, s_J que representam exatamente a mesma região geográfica em ambos os censos. Para assertar essa condição, empregamos alguns critérios. Verificamos a preservação absoluta da localização, ou seja, se os dois setores com mesmo id não estão geograficamente distantes. Verificamos também a localização relativa, ou seja, se é possível identificar setores vizinhos que também foram preservados entre censos. Por fim, avaliamos se os polígonos foram preservados, ou seja, se s_I e s_J não passaram por redesenho geométrico para representar uma extensão geográfica diferente. Atendidos esses critérios, adicionamos a aresta $s_I s_J$ a G , representando uma operação de redesenho 1 : 1 (Figura 4). Os setores assim classificados são isolados das demais etapas, pois não devem receber novos vínculos de correspondência.

Há casos de setores distintos que representam a mesma extensão geográfica, mas não possuem o mesmo id. Apesar de, logicamente, representarem uma manutenção, não é possível estabelecer uma correspondência inequívoca a partir apenas do polígono e do id. A correspondência entre esses setores será identificada em outras etapas.

4.2. Identificação de divisões

Nesta etapa, buscamos identificar as correspondências entre setores cujos códigos foram criados ou excluídos ou que estão desassociados, a partir da sobreposição espacial de seus

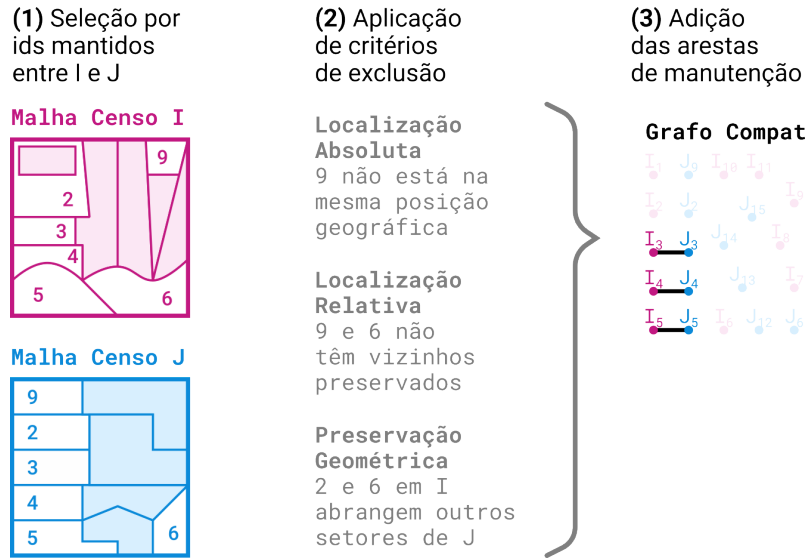


Figura 4. Esquema da etapa de identificação de manutenções.

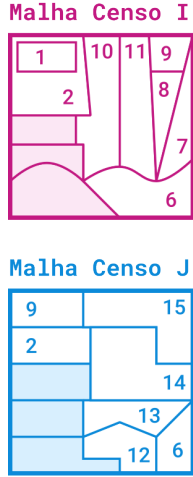
polígonos, para identificar casos em que um setor passou por divisão. Uma divisão implica uma operação de redesenho $1 : n$ em que a área dos setores divididos se sobrepõe integralmente ao setor original. Permitimos também ocorrências em que dois setores desassociados de mesmo id criem uma correspondência com $n = 1$. Nesse caso, rotulamos a divisão como “manutenção desassociada” e não isolamos os setores de outras etapas. Esse caso é contabilizado como uma manutenção na identificação de operações de redesenho.

Como há diferenças significativas na precisão do desenho dos polígonos entre duas malhas censitárias, é necessário adotar critérios para reduzir a sensibilidade da sobreposição espacial. A Figura 5 ilustra esse procedimento. O primeiro critério empregado é a remoção de polígonos residuais (“rebarbas”), feita a partir da razão área/perímetro (razão A/P). Em seguida, calculamos a razão entre a área de cada polígono interseccionado e a área do setor do censo mais atual ao qual ela pertence originalmente. Para pares de interseção em que essa razão está acima de um limite parametrizado L_{min} , ou seja, a área interseccionada representa, no mínimo, uma parcela L_{min} da área de um setor do censo mais novo, adicionamos uma aresta entre os respectivos vértices dos setores no grafo.

4.3. Sobreposição forçada

Os filtros aplicados nas etapas anteriores excluem casos excepcionais de redesenho, geralmente vinculados a operações de reconfiguração ($m : n$). Nesses casos, realizamos uma correspondência forçada. Seleccionamos os nós isolados do grafo de compatibilização e aplicamos um buffer negativo parametrizado para retrair o contorno dos polígonos em B metros. Em seguida, agrupamos os polígonos por censo e interseccionamos com a malha completa do censo complementar, removendo os polígonos residuais pelo mesmo critério usado anteriormente (razão A/P). Os pares de interseção remanescentes são adicionados ao grafo como arestas de correspondência forçada (Figura 6).

(4) Seleção de setores fora de relações de manutenção



(5) União geométrica com remoção de resíduos



(6) Aplicação de critérios de inclusão

Manutenção desassociada
2 e 6 se interseccionam entre I e J

Limite mínimo de sobreposição (L_{min})
Fragmentos (I_1, J_9) e (I_6, J_{12}) representam L_{min} da área original dos setores J_9 e J_{12}

(7) Adição de arestas de manutenção desassociada e de divisão

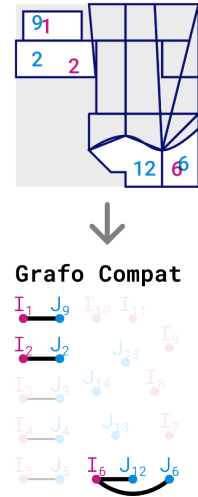
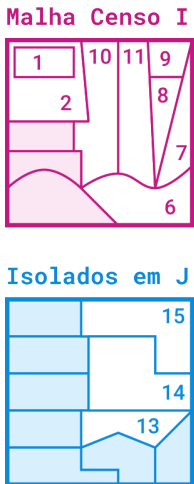
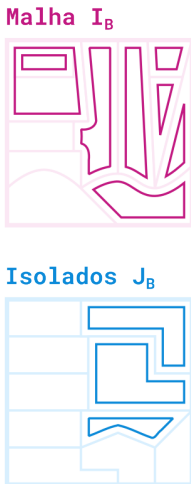


Figura 5. Esquema da etapa de identificação de divisões e manutenções desassociadas.

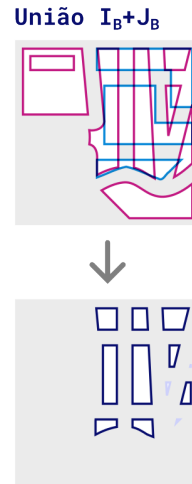
(8a) Seleção de setores fora de manutenções em I e isolados em J



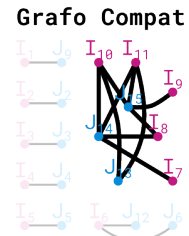
(9a) Aplicação de buffer negativo de B metros



(10a) União geométrica e remoção de resíduos



(11a) Adição de arestas de sobreposição forçada



(b) Repetir os passos usando Malha J e isolados em I

Figura 6. Esquema da etapa de sobreposição forçada.

4.4. Codificação das AMCs

As etapas acima resultam em um grafo cujos componentes representam conjuntos bipartidos de setores dos dois censos. A partir dos membros de cada componente, é possível criar tabelas de compatibilidade com relações $r = (s, K)$ em que K é o identificador de um componente e s é um setor do censo S tal que $s \in vertices(K)$. São criadas duas tabelas, uma para cada censo compatibilizado. A partir delas, criamos uma nova

malha geográfica poligonal, de AMCs de setores censitários, a partir da dissolução dos polígonos de setores pertencentes ao mesmo componente K (Figura 7). As componentes são classificadas conforme a operação de redesenho que representam ($1 : 1$, $1 : n$, $n : 1$ ou $m : n$).

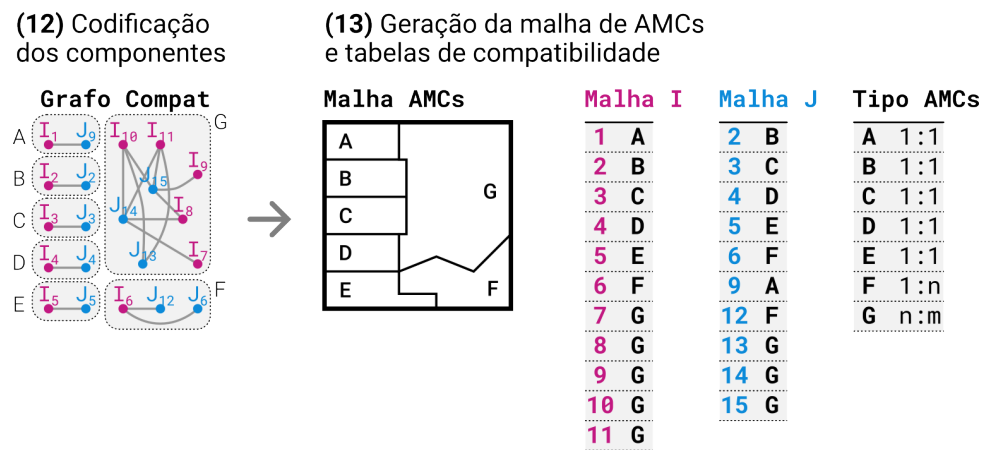


Figura 7. Esquema da codificação e criação das tabelas de compatibilidade.

5. Implementação e aplicação para os censos de 2010 e 2022

A metodologia apresentada foi implementada em Python para compatibilização dos censos de 2010 e 2022 (malha preliminar, divulgada em 2023) na Região Metropolitana de São Paulo (RMSP), a partir dos arquivos disponibilizados no serviço geoftp do IBGE. Foram usadas as bibliotecas `networkx` para construção dos grafos, `geopandas` para leitura e operação com as bases de dados espaciais dos dois censos e `shapely` para operações entre polígonos. O código e os arquivos resultantes estão disponíveis em repositório no GitHub.

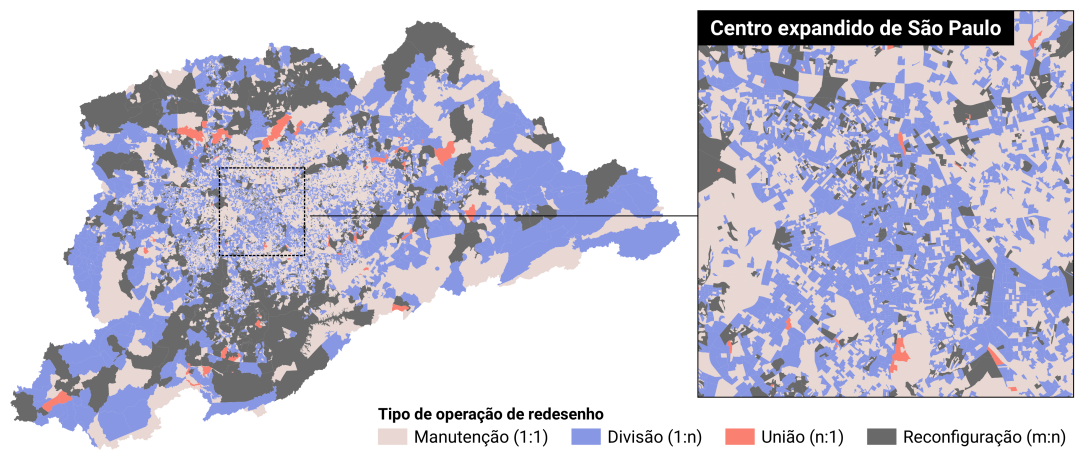


Figura 8. Resultados da compatibilização dos censos de 2010 e 2022.

O critério de preservação da localização absoluta em manutenções foi implementado para que um par de setores só seja considerado mantido se a distância entre os seus centroides for de até a metade da soma das diagonais de suas caixas envoltórias. Isso garante que, quanto menores os setores (ou seja, quanto mais detalhado o desenho local da

malha censitária), menor a tolerância a deslocamentos espaciais. O critério de preservação relativa foi adotado a partir da verificação de polígonos adjacentes de cada setor, exigindo que ao menos um vizinho também tenha sido mantido. O critério de preservação de polígonos foi implementado para excluir dos casos de manutenção os ids mantidos cujos setores do censo mais recente interseccionem mais de L_{min} da área de um outro setor com id diferente, sendo L_{min} o mesmo parâmetro usado para o critério de identificação de divisões.

5.1. Métricas de avaliação

Para avaliar os resultados da implementação e a escolha dos parâmetros, propomos as seguintes métricas:

- **Manutenções inconsistentes:** Número de arestas de manutenção (premissa 1 : 1) em operações de redesenho diferentes de 1 : 1;
- **Divisões não puras:** número de arestas de divisão (premissa 1 : n) em operações de redesenho m : n ;
- **Setores desconexos:** Número de setores que permaneceram desconexos ao final da compatibilização;
- **Isolados persistentes:** Número de setores censitários que não receberam arestas para nenhum buffer $B < 0$;
- **Redesenhos extensos:** Número de operações de redesenho do tipo m : n em que m e n são maiores que 10.

Numa implementação ideal, não há casos de manutenções inconsistentes e setores desconexos, e as ocorrências de isolados persistentes, divisões não puras e desenhos extensos são baixos. Mas, devido às particularidades de redesenho das malhas censitárias, que variam dependendo de características da região geográfica e do agente censitário responsável, esta implementação do algoritmo de compatibilização não permite assegurar a identificação inequívoca de todas as operações de redesenho na escala da RMSP. Assim, a escolha dos parâmetros razão A/P , L_{min} e B foi feita buscando minimizar as métricas apresentadas.

5.2. Resultados e discussão

A Figura 9 mostra os resultados das métricas propostas ao variarmos os parâmetros razão A/P e L_{min} . O aumento da razão A/P diminui divisões não puras, mas aumenta isolados persistentes. O mesmo ocorre para L_{min} em relação a manutenções inconsistentes contra isolados persistentes e redesenhos extensos. A combinação dos parâmetros tem efeito composto sobre as métricas de compatibilização e, por isso, escolhemos valores intermediários (razão $A/P = 0,2$ e $L_{min} = 70\%$) para a geração da malha final.

Além disso, para garantir que nenhum setor permaneça desconexo, executamos a operação de sobreposição forçada em sequência com três valores diferentes de B (-20, -10 e 0), e não descartamos os setores classificados como manutenção na última execução. A adoção de valores decrescentes em série para B ajudou a reduzir o valor final dos indicadores. A Tabela 1 mostra valores finais e métricas da execução nessa implementação, e a figura 8 mostra as AMCs classificadas por tipo de operação para a RMSP.

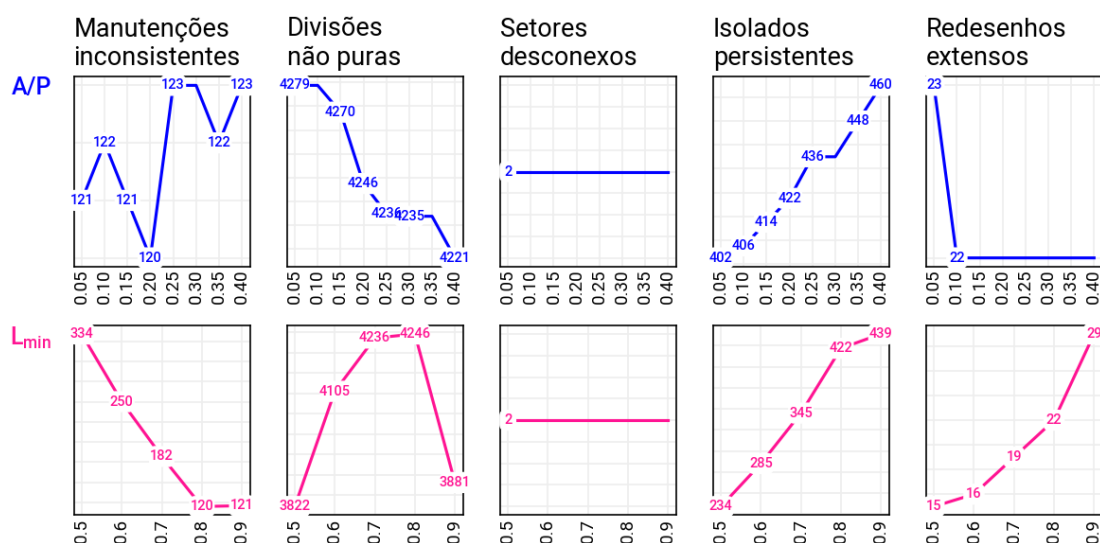


Figura 9. Variação das métricas para diferentes valores da razão A/P e de L_{min} , mantidos constantes os demais parâmetros.

Tabela 1. Métricas dos resultados da implementação para a RMSP.

| Malhas | |
|--|-----------------------|
| Número de setores da malha de 2010 | 30.815 |
| Número de setores da malha de 2022 | 47.184 |
| Correspondências | |
| Correspondências totais | 50.281 (100%) |
| Correspondências por manutenção | 18.182 (36,0%) |
| Correspondências por manutenção desassociada | 749 (1,5%) |
| Correspondências por divisão | 22.600 (44,9%) |
| Correspondências por sobreposição ($B = -20m$) | 6.537 (13,0%) |
| Correspondências por sobreposição ($B = -10m$) | 1.845 (3,7%) |
| Correspondências por sobreposição ($B = 0m$) | 422 (0,8%) |
| Operações de redesenho | |
| Número de AMCs = Operações de redesenho | 28.263 (100%) |
| Operações de manutenção (1 : 1) | 18.973 (67,1%) |
| Operações de divisão (1 : n) | 8.033 (28,4%) |
| Operações de união (n : 1) | 156 (0,5%) |
| Operações de reconfiguração (m : n) | 1.101 (3,9%) |
| Métricas | |
| Manutenções inconsistentes (% das manutenções) | 120 (0,3%) |
| Divisões não puras (% das divisões) | 4.246 (18,8%) |
| Setores desconexos (% dos setores) | 2 (0,003%) |
| Isolados persistentes (% dos setores) | 422 (0,5%) |
| Redesenhos extensos (% dos AMCs) | 22 (0,07%) |

Avaliamos que a implementação produz arquivos de compatibilização adequados, capazes de assegurar a equivalência de estrutura espacial em escala local com poucas

perdas. As operações de manutenção predominam sobre as demais, e as operações de reconfiguração e de união são incomuns, conforme esperado. A escolha da ordem das etapas no algoritmo corresponde à ordem das operações de redesenho mais frequentes (manutenção, divisão e reconfiguração), cumprindo a proposta de fluxo de classificação “em funil”.

A heurística adotada permite que as AMCs resultantes sejam empregadas em análises de processos urbanos que usam localizações precisas para uniões espaciais. É o caso do trabalho de [Mendonça et al. 2024], que utilizou uma versão preliminar deste algoritmo para classificar transformações demográficas na RMSP entre 2010 e 2022. No geral, análises baseadas em séries censitárias em escala local que dependam da união espacial entre dados com localização precisa, como lotes urbanos ou coordenadas, podem se beneficiar da implementação da metodologia proposta.

Comparado com os métodos não automatizados empregados nos trabalhos relacionados, ainda que a intervenção humana possa reduzir possíveis imprecisões, há uma redução significativa do custo de realizar essa operação manualmente. A execução do código em uma máquina intermediária (intel i7-9700F com 32GB de RAM) levou de 3 a 4 minutos para finalizar a compatibilização da RMSP, uma das áreas mais densamente urbanizadas do país. Em contrapartida, há uma possível perda por simplificação excessiva ou identificação incompleta das operações de redesenho, com caráter espacialmente concentrado. Assim, a verificação visual dos resultados é recomendável.

Ainda, é necessário considerar o contexto regionalmente diverso da malha censitária brasileira. Com a descentralização do redesenho, a precisão e consistência das malhas entre censos pode variar consideravelmente entre diferentes cidades. Em decorrência disso, a escolha dos parâmetros pode ter efeitos significativamente diferentes em aplicações sobre outros recortes espaciais. Isso é verdade também para o interior de regiões metropolitanas. Vemos na Figura 8, por exemplo, que operações de redesenho por reconfiguração se concentram em áreas periféricas, semi-rurais e rurais, gerando possíveis perdas de informação em análises de processos específicos a essas regiões. Já nas áreas em que a malha é mais densa, como em bairros verticalizados, os setores costumam ter áreas muito pequenas e são mais sensíveis a pequenos ajustes de desenho.

6. Trabalhos futuros

A partir da avaliação dos resultados, apontamos dois principais caminhos para futuros avanços na metodologia proposta.

O primeiro deles é o desenvolvimento de uma forma mais robusta para escolha da razão A/P , L_{min} e de B . A razão A/P pode ser complementada ou substituída por outras métricas de geometria, como alongação, compacidade e curvatura. Uma avaliação sistemática dos efeitos combinados dos parâmetros pode aprimorar a escolha de seus valores. Contudo, a escolha de parâmetros globais tende a balancear distorções entre áreas de malha densa (centros urbanos) e esparsa (regiões semi-rurais e rurais). Assim, entendemos que a implementação de uma escolha automatizada de parâmetros dependente do contexto geográfico é uma alternativa mais promissora.

Um segundo caminho para futuros aprimoramentos toma o exemplo de trabalhos de compatibilização que utilizam dados auxiliares. O uso de imagens de satélite, dados

de uso do solo, identificação de áreas não habitadas e a malha viária pode aprimorar os resultados da compatibilização ao diminuir a influência de áreas não habitadas – ou seja, de pouca relevância para análises demográficas.

Referências

- Barbosa, R. J. (2014). Comparabilidade das informações disponíveis nos Censos (1960-2010) e PNADs (1976, 1985, 1995 e 2005). *São Paulo: Centro de Estudos da Metrópole*.
- Hirye, M. C. d. M., Amaral, S., Monteiro, A. M. V., and Alves, D. S. (2016). Interpolação de dados censitários para análise da ocupação intraurbana em Altamira (PA) em 2000 e 2010. *Revista Brasileira de Cartografia*, 68(8):2016.
- IBGE (1980). *IX Recenseamento Geral - Instruções para delimitação dos setores censitários*. IBGE, 1 edition. Acesso em: 2025-03-02.
- Lobo, M. A. A. (2009). Método para compatibilizar setores censitários urbanos de 1991 e 2000 aplicado ao estudo da dinâmica populacional da região metropolitana de Belém (PA). *Revista Brasileira de Gestão Urbana*, 1(1):71–84.
- Mendonça, P. H. R., Lima, P. H. B. M., Costa, D. F., Canan, H. G., Benedusi, A. A., Giacomini, L. A., Azzolini, G. S., Andrade, L. G. E., Stroher, L. E., Santoro, P. F., and Rolnik, R. (2024). A expansão – com desadensamento – da Região Metropolitana de São Paulo entre 2010 e 2022. *E-metropolis*, 15.
- Reis, E., Pimentel, M., Alvarenga, A. I., and Santos, M. d. C. H. (2008). Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. *Rio de Janeiro: Ipea/Dimac*, 40.
- Reis, I. A. (2013). Compatibilização de populações entre malhas censitárias diferentes com o uso de imagens de sensores orbitais. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR - INPE*.
- Rodríguez, G. M. (2021). Comparabilidad retrospectiva en la cartografía censal digital del INDEC. estado actual, avances y desafíos en Argentina y la ciudad de Buenos Aires. *Población de Buenos Aires*, 18(30):22–33.
- Rodriguez, G. M. and de Grande, P. E. (2024). Base cartográfica de radios del censo argentino 2022: Primera versión revisada y corregida para uso en Sistemas de Información Geográfica. <https://ri.conicet.gov.ar/handle/11336/238198>. Acesso em: 2025-03-02.
- US Census Bureau (2024a). History of census tracts and blocks. <https://www.census.gov/about/history/historical-censuses-and-surveys/census-programs-surveys/geography/tracts-and-block-numbering-areas.html>. Acesso em: 2025-03-02.
- US Census Bureau (2024b). Relationship files. <https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2020.html#tract>. Acesso em: 2025-03-02.
- Yamaguchi, F. Y. (2017). Avaliação de dados de grades regulares para fins estatísticos. Master's thesis, Escola Politécnica da Universidade Federal da Bahia.