
SEGMENTACIÓN AUTOMÁTICA DE LA PROVINCIA DE BUENOS AIRES PARA EL CENSO NACIONAL ARGENTINO 2010

FLAVIA BONOMO ^{*}
DIEGO DELLE DONNE ^{**}
GUILLERMO DURÁN ^{***}
JAVIER MARENCO ^{**}

Resumen

La planificación de un censo poblacional implica diversos desafíos logísticos, uno de los cuales es determinar qué hogares debe visitar cada censista. Este problema se conoce como el problema de segmentación de viviendas, y generalmente involucra un conjunto de restricciones sobre los recorridos que cada censista puede realizar y determinados criterios de homogeneidad y uniformidad que las soluciones deben respetar. En este trabajo presentamos un enfoque basado en programación lineal entera para la resolución del problema de segmentación en las zonas urbanas y semi-urbanas de la Provincia de Buenos Aires, que fue aplicado exitosamente durante el Censo Nacional 2010 en Argentina.

PALABRAS CLAVE: Censo Poblacional, Programación lineal entera, Segmentación.

^{*}Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina y IMAS-CONICET, Argentina

^{**}Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina e Instituto de Ciencias, Universidad Nacional de General Sarmiento, Argentina y IMAS-CONICET, Argentina

^{***}Depto. de Matemática, FCEyN, Universidad de Buenos Aires, Argentina, Depto. de Ingeniería Industrial, FCFM, Universidad de Chile, Chile e IMAS-CONICET, Argentina

1. Introducción

Un *censo poblacional* es el proceso de adquisición de información demográfica de un área geográfica (habitualmente un país). El objetivo principal es relevar datos estadísticos sobre los habitantes, incluyendo información sobre sus viviendas, su nivel educativo y su realidad laboral. Típicamente, un censo se realiza a lo largo de un sólo día no laborable, y la información es recabada por *censoistas* que visitan casa por casa a todos los habitantes del área censada.

Las actividades de planificación de un censo suelen comenzarse con varios años de anticipación, dado que un censo involucra un importante movimiento logístico realizado durante unas pocas horas. Una tarea crucial dentro de la planificación de un censo es determinar qué viviendas deberá visitar cada censoista, problema que en este contexto se denomina *problema de segmentación de viviendas*. Se trata de un problema de *partición* del conjunto de viviendas en subconjuntos que cumplan cierto conjunto de restricciones, y puede ser de muy difícil resolución dependiendo del tipo de restricciones y de los objetivos de la partición.

Cuando esta partición se realiza a relativamente alto nivel –por ejemplo, para determinar distritos electorales dentro de un área geográfica– el problema es conocido como *problema de re-zonificación* (*redistricting problem*, en inglés), y para este tipo de problemas existe una abundante literatura [2, 3, 6, 8, 9, 10, 11]. Existen varios paquetes de software que permiten realizar este tipo de tareas en forma automática y semi-automática (por ejemplo, en [4] se presenta una herramienta *open source* implementada para el entorno de programación estadística R). Sin embargo, las restricciones presentes en los problemas de re-zonificación en general difieren de los requerimientos de un problema típico de segmentación de viviendas, con lo cual se hace necesario recurrir a algoritmos específicos para este último caso. Además de estas restricciones, se agregan al problema de segmentación de viviendas ciertos criterios de homogeneidad y uniformidad que deben ser respetados por las soluciones halladas (los mismos se detallan en la Sección). En los enfoques de resolución manual utilizados en la actualidad, estos criterios son prácticamente imposibles de respetar, ya que las soluciones dependen fuertemente de las decisiones de cada operador.

En este trabajo describimos el problema de segmentación de viviendas para el Censo Nacional 2010 en la Provincia de Buenos Aires (Argentina), y presenta-

mos una herramienta basada en programación lineal entera para su resolución. Esta herramienta fue utilizada exitosamente durante el proceso de planificación del Censo Nacional 2010 en esta provincia. No estamos al tanto de otros trabajos en la literatura de investigación de operaciones que hayan encarado el problema de segmentación de viviendas en el contexto de un censo. El resto de este trabajo está organizado del siguiente modo. En la Sección damos una descripción general del Censo Nacional 2010 y del problema particular de segmentación de viviendas que nos ocupa. La Sección 2.1 contiene los detalles del algoritmo propuesto y menciona las principales dificultades encontradas durante su implementación. Por último, la Sección 3.1 menciona los resultados obtenidos con esta herramienta y la Sección 3.1 cierra el trabajo con conclusiones sobre el proceso realizado.

2. Descripción del problema

El territorio de Argentina está dividido en 23 *provincias* además de la Ciudad Autónoma de Buenos Aires (a la que habitualmente se refiere como Capital Federal), que tiene un status jurídico especial. A su vez, cada provincia está dividida en *partidos* o *departamentos*. A los fines del censo, cada partido se divide en conjuntos de manzanas contiguas denominados *radios censales*. Cada radio censal contiene aproximadamente 300 hogares y, dependiendo de su densidad poblacional, entre 1 y 50 manzanas. A lo largo de este trabajo nos referiremos a los radios censales simplemente como “radios”.



Figura 1: Mapa de Argentina, resaltando la Provincia de Buenos Aires.

La Provincia de Buenos Aires (ver Figura 1) es la provincia de mayor superficie (307.571 km²) y mayor cantidad de habitantes (unos 15.3 millones de

habitantes) de Argentina. No incluye a la Ciudad Autónoma de Buenos Aires pero sí incluye al Gran Buenos Aires, un cordón de unos 9 millones de habitantes que rodea al territorio de la Ciudad de Buenos Aires y conforma junto con ella un núcleo poblacional único. Salvo el Gran Buenos Aires, la Provincia de Buenos Aires es predominantemente de carácter rural, con algunos sectores definidos dedicados al turismo (sobre la costa atlántica), la explotación minera (en el sur de la provincia) y la industria metalúrgica (en el sector noreste, sobre la costa del Río Paraná). En contraste, el Gran Buenos Aires tiene un perfil fundamentalmente urbano e industrial. La Provincia de Buenos Aires está dividida en 134 *partidos*, dentro de los cuales se ubican 16.691 radios. En este trabajo estamos focalizados en la segmentación de estos casi 17.000 radios censales.

Dado un radio, el problema de segmentación de viviendas consiste en particionar las viviendas del radio en conjuntos disjuntos de viviendas –llamados *segmentos*– de modo tal que cada segmento sea asignado a un censista. Cada segmento debe estar compuesto por viviendas *contiguas*. Dos viviendas se consideran contiguas si están en la misma manzana y son adyacentes, o bien pertenecen a esquinas de dos manzanas distintas y dichas esquinas están enfrentadas sobre una misma calle, es decir, un segmento no puede cruzar una esquina en diagonal. La Figura 2 muestra algunos ejemplos de segmentos factibles y no factibles de acuerdo con este criterio de contigüidad. Este criterio está determinado por el manual de procedimientos censales de la Dirección Provincial de Estadísticas de Buenos Aires.

Denominamos *segmentación* a la partición de las viviendas de un radio en un conjunto de segmentos. De acuerdo con el manual de procedimientos del Censo Nacional 2010, una segmentación factible debe cumplir las siguientes restricciones:

- Cada vivienda debe pertenecer a un único segmento y el conjunto de segmentos debe cubrir a todas las viviendas del radio.
- Si un *lado* de una manzana no tiene ninguna vivienda, ese lado también debe formar parte de algún segmento. Es importante notar que denominamos “lados” a los bordes rectos de una manzana. En los trazados urbanos típicos de las ciudades argentinas, habitualmente una manzana es un rectángulo que contiene cuatro lados, aunque no es infrecuente encontrar manzanas con morfologías distintas.
- Cada segmento debe contener entre 32 y 40 viviendas.
- Cada censista no debe recorrer una distancia superior a cierta cota superior L , que depende de la densidad poblacional del radio en cuestión.



Figura 2: Los segmentos mostrados en (a) y (b) son ejemplos de segmentos factibles que cruzan la calle. El segmento mostrado en (c) no es factible dado que las viviendas no son contiguas. El segmento mostrado en (d) no es factible dado que cruza una esquina en diagonal.

- Los censistas no pueden cruzar avenidas, vías del ferrocarril ni cursos de agua.
- Un segmento debe estar contenido dentro de su radio.

La última restricción es muy importante para nuestros propósitos, dado que como un segmento no puede involucrar viviendas de dos radios distintos, entonces el problema de segmentar todas las viviendas de la provincia se reduce a 16.691 instancias individuales, una por cada radio. Por otra parte, los siguientes elementos son deseables para la segmentación, en orden de preferencia:

1. Se deben privilegiar los segmentos que consten de manzanas completas.
2. Los segmentos deben consistir de lados completos y ser tan “compactos” como sea posible. Este requerimiento de *compacidad* está definido informalmente y hace referencia a la amplitud de un segmento a lo largo de las manzanas del radio al que pertenece. Por ejemplo, un segmento constituido por una o varias manzanas completas es considerado muy com-

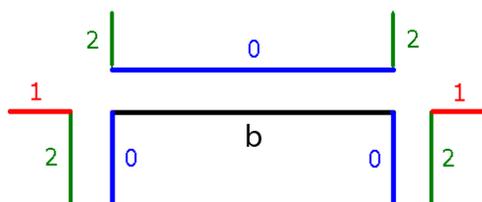


Figura 3: Niveles de adyacencia de los lados que pueden recorrerse a continuación del lado b en un segmento factible.

pacto, mientras que un segmento que recorre varias manzanas utilizando pocos lados por manzana es considerado poco compacto. El objetivo de emplear segmentos tan compactos como sea posible es el de minimizar la posibilidad de que los censistas cometan errores en sus recorridos.

3. Si no es posible contar con lados completos (por ejemplo, si un lado tiene más de 40 viviendas), se puede partir un lado y distribuirlo en más de un segmento. En este caso se debe dar preferencia a los segmentos que contengan todos los apartamentos de un mismo edificio, de modo tal que cada edificio sea atendido por un mismo censista.
4. Si el punto anterior no es posible (por ejemplo, porque un edificio tiene más de 40 apartamentos), entonces se debe dar preferencia a los segmentos que no dividan los apartamentos de un mismo piso. En otras palabras, idealmente cada piso de un edificio debe ser atendido por un mismo censista.

Como preferencia adicional, luego de recorrer un lado de una manzana es deseable o bien no cruzar la calle o bien continuar por el lado directamente enfrente al lado recorrido. La Figura 3 especifica esta preferencia. Luego de recorrer el lado b de la figura, es preferible no cruzar la calle y continuar con los lados marcados con 0 de la misma manzana o bien cruzar al lado directamente enfrente a b . Si esto no es posible, es preferible continuar el segmento por el lado de mismo sentido en la siguiente manzana (lados marcados con 1 en la figura). Si esto no es posible, se admite dar un giro al recorrido y continuar por alguno de los lados en las manzanas adyacentes (lados marcados con 2 en la figura).

2.1. Modelo de programación lineal entera

Sobre la base de la descripción anterior, se puede formular un modelo de programación lineal entera para este problema. Sea \mathcal{R} el radio a procesar, y sea $\mathbb{S}_{\mathcal{R}}$ el conjunto de todos los segmentos factibles sobre el radio \mathcal{R} . Para cada segmento $s \in \mathbb{S}_{\mathcal{R}}$ introducimos la variable binaria x_s , de modo tal que $x_s = 1$ si y sólo si el segmento s participa de la solución.

Con el objetivo de maximizar la compacidad de los segmentos seleccionados, proponemos la siguiente función objetivo. Dado un segmento $s \in \mathbb{S}_{\mathcal{R}}$, definimos su *compacidad* como $\text{comp}(s) = \frac{\text{lados}(s)}{\text{manzanas}(s)}$, donde $\text{lados}(s)$ y $\text{manzanas}(s)$ representan la cantidad de lados y manzanas, respectivamente, en el segmento s . De esta manera, un segmento que involucra una manzana rectangular completa tendrá una compacidad de 4, mientras que un segmento que involucre cuatro lados en cuatro manzanas distintas tendrá una compacidad de 1. Cabe notar que con esta definición, un segmento formado por una manzana completa logra la misma compacidad que sumar dos segmentos formados por media manzana cada uno, o bien cuatro segmentos formados por cuartos de manzana, sin embargo el primer caso es el más deseado. Para priorizar los segmentos de mayor compacidad y evitar estos casos, definimos la *valuación* de un segmento s como $\text{val}(s) = k^{\text{comp}(s)}$. Tras algunas pruebas preliminares para determinar el valor de k , se fijó este valor en 10, pues con este valor se obtuvieron los mejores resultados. Este criterio fue definido en conjunto con los responsables de la planificación del censo en la Provincia de Buenos Aires, y proporcionó resultados satisfactorios.

Sea \mathbb{V} el conjunto de viviendas de \mathcal{R} , y sea \mathbb{L}_0 el conjunto de lados sin viviendas de \mathcal{R} . Para cada $v \in \mathbb{V}$ llamamos $S_v \subseteq \mathbb{S}_{\mathcal{R}}$ al conjunto de segmentos factibles que incluyen a la vivienda v , y para cada $l \in \mathbb{L}_0$ llamamos $L_l \subseteq \mathbb{S}_{\mathcal{R}}$ al conjunto de segmentos factibles que incluyen al lado l . Con estas definiciones, el modelo de programación lineal entera de segmentación para este problema es el siguiente:

$$\text{máx} \sum_{s \in \mathbb{S}_{\mathcal{R}}} \text{val}(s) \cdot x_s$$

$$\sum_{s \in S_v} x_s = 1 \quad \forall v \in \mathbb{V} \tag{1}$$

$$\sum_{s \in L_l} x_s = 1 \quad \forall l \in \mathbb{L}_0 \tag{2}$$

$$x_s \in \{0, 1\} \quad \forall s \in \mathbb{S}_{\mathcal{R}} \tag{3}$$

Las restricciones (1) aseguran que cada vivienda es cubierta por exactamente un segmento, y las restricciones (2) aseguran que cada lado sin viviendas es recorrido por exactamente un segmento. Es importante observar que no estamos incluyendo en este modelo las preferencias en cuanto a los niveles de adyacencia utilizados en los segmentos que cruzan las calles.

Este modelo tiene en general una cantidad muy grande de variables, dado que la cantidad de segmentos factibles crece en forma exponencial a medida que aumenta la cantidad de manzanas en el radio en cuestión. Por este motivo, no resulta práctico ejecutar el modelo con todos los segmentos factibles. Una opción natural es recurrir a un algoritmo de generación de columnas (ver [5] para mayores detalles sobre esta técnica), aunque en este caso no resulta claro a priori cómo resolver satisfactoriamente el subproblema de generación de columnas (generar segmentos válidos con costos reducidos positivos, en nuestro caso) sin recurrir a heurísticas o búsquedas exhaustivas, además de que no se tendrían en cuenta las preferencias en cuanto a los niveles de adyacencia utilizados al cruzar la calle. Estas preferencias se podrían incorporar en la función objetivo a través de la función de valoración de los segmentos, aunque esto agregaría parámetros adicionales que deberían ser ajustados para obtener resultados satisfactorios.

Por estos motivos, en el presente trabajo implementamos un enfoque basado en la resolución secuencial de este modelo para conjuntos cada vez más grandes de segmentos, teniendo en cuenta explícitamente los niveles de adyacencia utilizados por los segmentos. Este esquema nos permitió hallar soluciones satisfactorias, de acuerdo con los resultados reportados en la Sección 3.1.

3. El algoritmo de segmentación

Describimos en esta sección el algoritmo que implementamos para encarar la resolución del problema de segmentación de viviendas. Decimos que un segmento se encuentra *excedido* si contiene más de 40 viviendas o su longitud supera el límite L . Decimos que un segmento es *factible* si no está excedido y tiene al menos 32 viviendas. En el contexto del algoritmo, los segmentos no son necesariamente factibles. Para $\delta \in \{0, 1, 2\}$, decimos que un segmento es δ -conexo si sus lados están conectados por adyacencias de tipo a lo sumo δ , de acuerdo con la especificación de la Figura 3.

El Algoritmo 1 ilustra el procedimiento propuesto, tomando los datos geográficos del radio como datos de entrada y el valor $\delta \in \{0, 1, 2\}$ como parámetro. Para $i \geq 1$, denominamos S_i al conjunto de segmentos no excedidos (aunque no necesariamente factibles) que involucran no más de i manza-

nas. En la i -ésima iteración, el Algoritmo 1 ejecuta el modelo de programación lineal entera de segmentación utilizando solamente los segmentos factibles S'_i del conjunto S_i . Si el modelo no tiene solución factible, se continúa iterando hasta que el modelo tenga solución factible, o S_i no varíe con respecto a S_{i-1} , o bien hasta llegar a un límite MI de iteraciones especificado de antemano. En cuanto el modelo es factible, el algoritmo termina retornando la solución obtenida por el modelo.

Algoritmo 1 Algoritmo de segmentación para nivel δ de adyacencias.

```

1:  $S_b \leftarrow \{\}$  // conjunto base
2: para toda manzana  $q$  hacer
3:    $S_b \leftarrow S_b \cup \{\text{segmentos de } q \text{ no excedidos}\}$ 
4: fin (para)
5:  $i \leftarrow 1$ 
6:  $S_i \leftarrow S_b$ 
7: repetir
8:   Ejecutar el modelo de PLE de segmentación con los segmentos factibles de  $S_i$ 
9:   si hay solución entonces
10:     Terminar (con solución)
11:   fin (si)
12:    $S_{i+1} \leftarrow S_i$ 
13:   para todo  $(s_i, s_b) \in S_i \times S_b$  hacer
14:     si  $s_i \cup s_b$  es un segmento  $\delta$ -conexo no excedido entonces
15:        $S_{i+1} \leftarrow S_{i+1} \cup \{(s_i \cup s_b)\}$ 
16:     fin (si)
17:   fin (para)
18:   si  $S_{i+1} = S_i$  entonces
19:     Terminar (sin solución)
20:   fin (si)
21:    $i \leftarrow i + 1$ 
22: hasta que  $i > MI$ 
23: Terminar (sin solución)

```

Es importante observar que la solución obtenida por este algoritmo puede no ser óptima para el modelo de PLE de segmentación utilizando todos los segmentos, dado que se ejecuta este modelo con un subconjunto de segmentos $S'_i \subseteq \mathcal{S}_{\mathcal{R}}$. Sin embargo, este procedimiento secuencial permite tener en cuenta de una manera muy natural las preferencias sobre la compacidad de los segmentos seleccionados, aumentando la posibilidad de encontrar primero las soluciones preferenciales. Vale aclarar que la solución obtenida es factible para dicho modelo.

En las líneas 2–4, el algoritmo genera el conjunto base de segmentos S_b utilizando lados de manzanas completos. En este paso, se generan para cada manzana todos los posibles segmentos no excedidos contenidos en la manzana.

Con el objetivo de hallar soluciones formadas por segmentos lo más compactos posible, se ejecuta el Algoritmo 1 secuencialmente para $\delta = 0, 1, 2$, interrumpiendo el procedimiento cuando se encuentra la primera solución factible.

Si luego de todo el proceso mencionado no se encuentra solución factible, se activa la opción de partir lados. Para ello, se introduce un parámetro P que indica el número máximo de viviendas que puede tener un lado y se fragmentan en dos o más partes los lados que superen en viviendas a este valor. Cada una de estas *partes* debe tener a lo sumo P viviendas. El algoritmo de fragmentación de lados es un procedimiento goloso, e intenta dejar en una misma parte los apartamentos de un mismo edificio. Para ello, comenzando desde una de las esquinas del lado a partir, se construye una *parte* del lado recorriendo el mismo hasta encontrar un edificio o bien hasta juntar P viviendas. En cualquiera de los dos casos, se termina la *parte* actual y se comienza una nueva. Siguiendo el mismo procedimiento se obtiene el conjunto de partes correspondientes al lado partido. Finalmente, para estos lados partidos se especifican nuevas adyacencias, de acuerdo con la Figura 4. Vale aclarar que

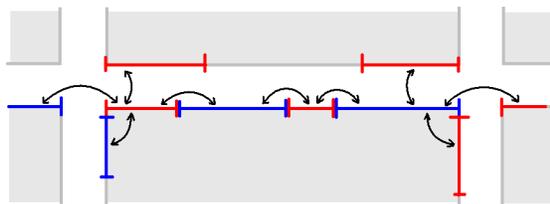


Figura 4: Adyacencias para lados partidos.

mientras más chico sea P , mayor será la cantidad de lados (i.e., partes) de las manzanas del radio y mayor será el tamaño del conjunto base S_b .

Una vez partidos los lados de acuerdo con la descripción anterior, se ejecuta el Algoritmo 1 nuevamente para $\delta = 0, 1, 2$, intentando encontrar una solución factible. Si aún así no se encuentra una solución factible, se reduce el valor de P y se repite el proceso. El Algoritmo 2 ilustra este procedimiento recibiendo como parámetro una lista PL que contiene los sucesivos valores a utilizar para el parámetro P . Vale aclarar que estos valores se utilizan en orden decreciente, ya que mientras mayor sea el valor de P , menor fragmentación tendrán las soluciones obtenidas. Si luego de todo este procedimiento no se encuentra una solución factible, el procedimiento termina informando al usuario que no se pudo encontrar una segmentación. Cabe mencionar que inicialmente el valor de P es menor o igual a 40 (veremos más adelante que estos valores dependen de la densidad poblacional del radio a resolver).

Algoritmo 2 Algoritmo de segmentación

- 1: **para** cada $P \in PL$ **hacer**
 - 2: Partir los lados que superen en viviendas el valor P .
 - 3: **para** $\delta = 0$ **hasta** 2 **hacer**
 - 4: Ejecutar el Algoritmo 1 para δ
 - 5: **si** el Algoritmo 1 encontró una solución **entonces**
 - 6: Retornar la solución y terminar
 - 7: **fin (si)**
 - 8: **fin (para)**
 - 9: **fin (para)**
 - 10: Terminar (sin solución)
-

3.1. Mejoras para radios con baja densidad poblacional

En las etapas de prueba y ajustes preliminares, el Algoritmo 2 permitió resolver una gran cantidad de radios urbanos de la Provincia de Buenos Aires. Sin embargo, este algoritmo tuvo serios problemas para resolver radios semi-urbanos con densidades poblacionales relativamente bajas. Por ejemplo, la Figura 5 muestra un radio de la Ciudad de Olavarría para el cual este algoritmo no logró encontrar una solución factible. El principal problema en este tipo de radios es que tienen un gran número de manzanas sin viviendas o muy poco pobladas, y esto hace que el algoritmo requiera alcanzar un alto número de iteraciones para obtener suficientes segmentos factibles como para poder cubrir el radio entero. Por otro lado, al tener muy pocas viviendas por manzana, la cantidad de segmentos no excedidos es excesivamente grande. Por ejemplo, el conjunto S'_7 para el radio de la Figura 5 está formado por más de 100,000 segmentos, lo que hace que tanto la generación de los segmentos como la resolución del modelo tome demasiado tiempo. La combinación de estas dos características hace que sea imposible llevar a la práctica el algoritmo implementado.

Sin embargo, el hecho de que un radio esté muy poco poblado hace que no sea necesario manejar tantos segmentos base, ya que las restricciones para combinarlos son menos prohibitivas (teniendo en cuenta los límites inferior y superior de viviendas por segmento). Así, para solucionar los problemas mencionados, agregamos al algoritmo las siguientes mejoras:

- En primer lugar, las manzanas que tienen muy pocas viviendas son consideradas como un único segmento no excedido al momento de agregar segmentos al conjunto base S_b (línea 3 del Algoritmo 1). Para ello se agrega un parámetro MP que indica el mínimo número de viviendas que debe tener una manzana para generar varios segmentos base a par-

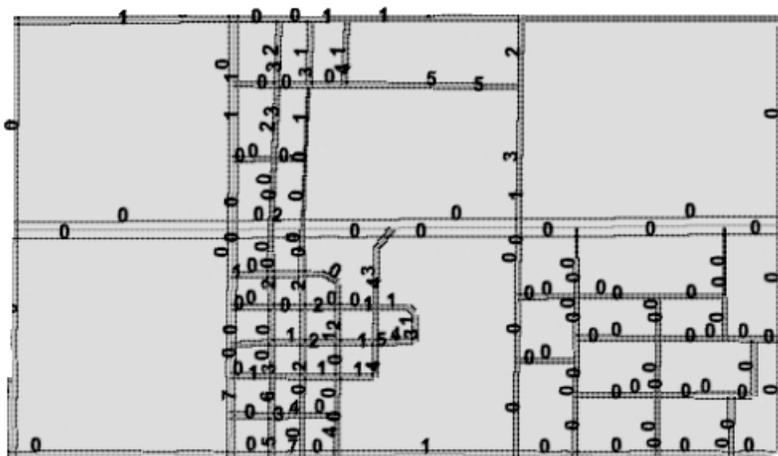


Figura 5: Radio semi-urbano de la Ciudad de Olavarría para el cual el Algoritmo 2 tiene una performance inaceptable. Los números en la figura indican la cantidad de viviendas en cada lado.

tir de ella. Es decir, si una manzana no alcanza este valor de viviendas, entonces sólo se agrega a S_b el segmento correspondiente a la manzana completa, en lugar de considerar todos los segmentos posibles de la manzana. Esto contribuye a reducir considerablemente el tamaño de S_b en radios de baja densidad de población.

- Una vez generados los segmentos base S_b , si alguno de ellos tiene pocas viviendas, se lo combina arbitrariamente con algún segmento adyacente formando un único segmento base con los dos. Para ello se agrega un parámetro MH que indica el mínimo número de viviendas que deben tener los segmentos de S_b . Este procesamiento del conjunto S_b se agrega al Algoritmo 1 a continuación de la línea 4.

Estos parámetros deben ser manejados con cuidado ya que según el radio a resolver, valores erróneos pueden incrementar los tiempos de ejecución (por ejemplo, valores muy bajos para P) o bien comprometer la factibilidad del modelo (por ejemplo, valores muy altos para MP y MH).

4. Resultados computacionales

Dada la gran cantidad de parámetros definidos y su relevancia para los tiempos de ejecución del algoritmo, fue necesario realizar una experimentación preliminar con diversos radios de prueba, para definir el mejor juego de parámetros

para cada tipo de radio. Como resultado de esta experimentación se resolvió clasificar los radios según su densidad poblacional en tres categorías: *urbanos* (hasta 10 manzanas), *semi-urbanos* (entre 11 y 30 manzanas) y *rurales* (más de 30 manzanas). La Tabla 1 muestra los valores seleccionados para los parámetros, de acuerdo con las características de cada radio.

	Parám.	Urbanos	Semi-urbanos	Rurales
Cantidad máxima de iteraciones en la generación de segmentos	<i>MI</i>	4	7	9
Número mínimo de viviendas para partir una manzana	<i>MP</i>	1	2	10
Número mínimo de viviendas en los segmentos base (si no, se agrupan con otros)	<i>MH</i>	0	1	5
Número máximo de viviendas por parte	<i>PL</i>	[32, 16, 10]	[32, 16]	[40, 32, 20]
Límite de tiempo para la ejecución del modelo de IP (seg)	<i>MT</i>	60	60	120

Tabla 1: Valores utilizados para los parámetros del algoritmo según el tipo de radio.

Los algoritmos mencionados en la sección anterior fueron codificados en C++ y los modelos de programación lineal entera fueron resueltos con Cplex 12.1 [7]. Los datos de manzanas, lados y viviendas se tomaron de la base de datos geográfica de la Provincia de Buenos Aires, implementando en un sistema de información geográfica las interfaces necesarias para exportar la información e importar y visualizar las segmentaciones obtenidas por nuestro algoritmo.

La generación de los segmentos requiere de un tiempo relativamente corto, siendo el peor caso de unos 2 minutos con los parámetros especificados en la Tabla 1. Más del 99% de los modelos de programación lineal entera se resuelve en pocos segundos, dado que la relajación lineal resulta muy ajustada y la primera solución factible hallada por Cplex suele ser óptima. En pocos casos se llega al límite de tiempo *MT* especificado en la Tabla 1 con solución sub-óptima, la cual se toma dentro del Algoritmo 1 como la segmentación del radio.

El proceso de segmentación en la Provincia de Buenos Aires para el Censo Nacional 2001 (el anterior censo realizado en Argentina) se realizó manualmente, demandando 25 operadores a tiempo completo que trabajaron durante 30 días (es decir, unas 6,000 horas-hombre). En cambio, el Censo Nacional 2010 fue la primera oportunidad en la que la segmentación de viviendas en la Provincia de Buenos Aires se realizó por medio de herramientas computacionales

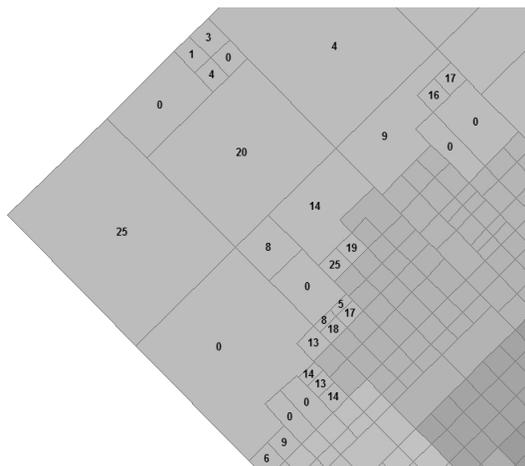


Figura 6: Radio semi-urbano para el cual el Algoritmo 2 no puede encontrar solución factible.

automáticas. Nuestro algoritmo obtuvo segmentaciones satisfactorias para el 96 % de los radios en aproximadamente 320 horas de CPU sobre un computador con procesador Intel Celeron[®] a 2.4 GHz y memoria RAM de 2 GB (el equivalente a un tiempo menor a 1 día en un cluster de 15 computadores).

En aproximadamente unos 600 radios (el 4 % del total) el algoritmo no encontró una solución factible en forma automática. Estos radios fueron resueltos utilizando la herramienta pero relajando levemente las restricciones del problema, o bien en forma manual. Por ejemplo, la Figura 6 corresponde a un radio semi-urbano muy poco poblado, en el cual las manzanas con pocas viviendas se agrupan entre sí para formar bloques de manzanas indivisibles (según la agrupación de manzanas descrita en la Sub-Sección 3.1). Este radio contiene tres manzanas contiguas con 14, 13 y 14 viviendas, respectivamente (ubicadas en la mitad inferior de la figura), las cuales están circundadas por manzanas escasamente pobladas, y por lo tanto probablemente agrupadas en un único bloque, “aislando” a estas tres manzanas. Estas manzanas suman 41 viviendas, con lo cual no es posible ubicarlas en un único segmento (superan el límite de 40 viviendas para un segmento) ni tampoco armar dos segmentos con ellas (segmentos con menos de 32 viviendas). Este radio particular se resolvió en pocos segundos relajando en una unidad el límite superior de 40 viviendas por segmento.

Muchos de los radios no resueltos por nuestro algoritmo tienen características similares al radio de la Figura 6. Es decir, se trata de radios predominantemente poco poblados pero con zonas densas en las fronteras (usualmente en las fronteras linderas con radios de mayor densidad). Al comentar estas carac-

terísticas con el equipo responsable de la planificación del censo, la respuesta obtenida fue que estos radios están incorrectamente delimitados, ya que no es deseable que un radio contenga zonas con densidades poblacionales muy distintas. Estos efectos son causados por el crecimiento poblacional.

5. Conclusiones

El algoritmo descrito en este trabajo permitió realizar el proceso de segmentación de viviendas de la Provincia de Buenos Aires dentro de los plazos que imponía la planificación del censo. Cabe destacar que el tiempo de desarrollo impuesto para la resolución del problema fue de tan sólo dos meses. A diferencia del proceso de segmentación manual empleado en el censo anterior, en el que cada operador podía introducir un sesgo en la segmentación realizada, el proceso automático a cargo del algoritmo permitió generar una segmentación con criterios uniformes para toda la provincia, contribuyendo así a una división del trabajo pareja entre los censistas. Además, los tiempos de procesamiento disminuyeron considerablemente con relación a la operatoria manual.

La performance del algoritmo demostró ser muy sensible a la parametrización utilizada. Con un conjunto adecuado de parámetros la resolución puede demorar unos pocos segundos, mientras que un juego de parámetros mal seleccionado puede comprometer la factibilidad del problema o bien llevar la cantidad de segmentos a varias centenas de miles, haciendo que la resolución demore varias horas. La clasificación de los radios de acuerdo con su densidad nos permitió manejar adecuadamente estos parámetros. Por último, el procedimiento secuencial dado por los algoritmos descritos fue fundamental para cumplir con las preferencias en cuando a las características deseables para los segmentos.

No obstante los resultados obtenidos, los algoritmos implementados en este trabajo ofrecen una solución heurística al problema. Como trabajo futuro, sería interesante profundizar en el enfoque de generación de columnas para la resolución exacta del modelo de programación entera planteado. El desafío mayor en este enfoque es la resolución del subproblema de generación de columnas.

Considerando que los radios censales tienen aproximadamente 300 viviendas y que cada censista recorre entre 32 y 40 viviendas, la cantidad de censistas necesaria para cada radio varía muy poco (i.e., entre 8 y 10 censistas aproxima-

damente). Por lo tanto, en cada instancia no parece tener sentido un objetivo que involucre la minimización de los censistas requeridos. Sin embargo, considerando que en la provincia hay casi 17.000 radios censales, éste puede ser un punto interesante a optimizar. Este aspecto del problema no fue abordado en este trabajo, pero puede ser un trabajo interesante a futuro.

En lo que hace a la opinión de los usuarios de la herramienta desarrollada, el responsable de los sistemas de información geográfica para el Censo Nacional 2010 en la Provincia de Buenos Aires, Fernando Aliaga, comenta que “el uso de esta herramienta computacional nos permitió una segmentación homogénea con criterios de compacidad uniformes, a diferencia de la segmentación manual que depende en gran medida de las decisiones de los operadores” [1]. El Censo Nacional 2010 se realizó exitosamente el 27 de Octubre de 2010, y fue calificado por las autoridades de la Provincia de Buenos Aires como un éxito organizativo [12].

Agradecimientos: Los autores agradecen al responsable de los sistemas de información geográfica para el Censo Nacional 2010 en la Provincia de Buenos Aires, Fernando Aliaga, y a su equipo de trabajo por su colaboración en varios aspectos fundamentales para el desarrollo de este trabajo. Este trabajo fue parcialmente financiado por ANPCyT PICT-2007-00518 y PICT-2007-00533, CONICET PIP 112-200901-00178, y UBACyT 20020090300094 y 20020100100980 (Argentina), FONDECyT 110797 e Instituto de Ciencias Milenio “Sistemas Complejos de Ingeniería” (Chile).

Referencias

- [1] Aliaga, F., Comunicación personal, Noviembre. 2010.
- [2] Altman, M., Is Automation the Answer: The Computational Complexity of Automated Redistricting, *Rutgers Computer and Law Technology Journal* 23(1) 81–141. 1997.
- [3] Altman, M., MacDonald, K., and McDonald, M.P., From Crayons to Computers: The Evolution of Computer Use in Redistricting, *Social Science Computer Review* 23(3) 334–346. 2005.
- [4] Altman, M. and McDonald, M.P., Bard: Better Automated Redistricting, *Journal of Statistical Software* 31(3). 2009.
- [5] Barnhart, C., Johnson, E. L., Nemhauser, G. L., Savelsbergh, M. W. P. and Vance, P. H., Branch-and-Price: Column Generation for Solving Huge Integer Programs, *Operations Research* 46 316–329. 1998.
- [6] Bozkaya, B., Erkut, E., and Laporte, G., A tabu search heuristic and adaptive memory procedure for political districting, *European Journal of Operational Research* 144(1) 12–26. 2003.
- [7] IBM ILOG, User's Manual for CPLEX. 2009.
- [8] Flesichmann, B. and Paraschis, J.N., Solving a large scale districting problem: a case report, *Comput. Oper. Res.* 15(6) 521–533. 1998.
- [9] Garfinkel, R.S. and Nemhauser, G.L., Optimal Political Districting by Implicit Enumeration Techniques, *Management Science* 16(8) B495–B508. 1970.
- [10] Helbig, R.E., Orr, P.K., and Roediger, R.R., Political redistricting by computer, *Commun. ACM* 15(8) 735–741. 1972.
- [11] Hess, S.W., Weaver, J.B., Siegfeldt, H.J., Whelan, J. N., and Zitlau, P.A., Nonpartisan Political Redistricting by Computer, *Operations Research* 13(6) 998–1006. 1965.
- [12] La voz de Tandil, Se censó más del 95 % de las viviendas en la provincia. Recuperado el 15 de Noviembre de 2010.
http://www.lavozdetandil.com.ar/ampliar_notas.php?id_n=20090. 2010.

