EL PROBLEMA DE HOMOGENEIDAD Y COMPACIDAD EN DISEÑO TERRITORIAL

María Beatríz Bernábe Loranca[1,3], David Pinto Avendaño[3], Marco Rodríguez Flores[3], Elias Olivares Benitez[1], Jorge Ruíz Vanoye[2], Rogelio Gonzalez V., José Luis Martínez Flores[1]

[1]Universidad Popular Autónoma del Estado de Puebla, Programa de Doctorado de Logistica y Dirección de la Cadena de Suministro, México
[2]Universidad Autónoma del Carmen, México
[3]Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, México

beatriz.bernabe@gmail.com

RESUMEN

El agrupamiento para unidades, zonas o áreas geográficas ha sido útil para resolver problemas relacionados con Diseño Territorial debido a que la propia definición exige el tratamiento de los datos espaciales bajo esquemas de clasificación con requerimientos de topología de las zonas.

El agrupamiento con restricciones espaciales tiene condiciones de complejidad computacional en Diseño Territorial. Considerando el carácter combinatório, se han utilizado herramientas de aproximación heurística para su resolución, sin embargo, a pesar de los esfuerzos, los aspectos de optimalidad y satisfactibilidad son débiles. En este punto, nos encontramos revisando estos puntos analizando el problema desde instancias pequeñas hasta donde la complejidad comienza a ser exponencial.

Con el fin de construir un método para crear grupos compactos y homogéneos, hemos recogido diversas implementaciones ya reportadas y por otro lado recuperamos un algoritmo sin el uso de heurísticos con el propósito de combinar los puntos fuertes de cada planteamiento, de este modo, es posible contar con una variante en el agrupamiento de zonas compactas y homogéneas para grupos pequeños donde la homogeneidad se resuelve para dos casos: 1) variables poblacionales y 2) número de objetos que componen los grupos. Se utiliza como caso de estudio las AGEBs (áreas geoestadísticas básicas) de la Zona Metropolitana del Valle de Toluca ZMVT. Los resultados del agrupamento para los algoritmos que se exponen en este artículo se presentan en mapas.

PALABRAS CLAVE: Agrupamiento, Compacidad, Diseño Territorial, Homogeneidad

Área principal (OA – Otras Aplicaciones en OR)

ABSTRACT

The clustering for geographical units, zones or areas have been used to solve problems related with Territorial Design due to the fact that its own description demands the processing of spatial data under schemes of classification with requirements of topology of the zones.

Due to the clustering and the spatial restrictions, the conditions of computational complexity in Territorial Design are of combinatorial nature and tools of heuristic approximation have been utilized to its solution, however, despite the efforts, the aspects of optimality and fulfillment are weak. In this point, we are looking through these aspects to analyze the problem from small instances until the complexity begins to grow.

With the goal of constructing a method to create compact and homogenous groups, we have collected diverse proposals and on the other hand we retrieved an algorithm without the use of heuristics with the objective of combining the strong points of each approach, in this way is possible to count with a variant in the clustering of compact and homogenous zones for small groups where the homogeneity is solved for two cases: 1) for population variables and 2) for the number of objects that form the groups. The AGEBs (Basic Geo-statistical Areas) of "Valle de Toluca" ZMVT are used as a case of study.

KEYWORDS. Clustering, Compactness, Homogeneity Second keyword, Territorial Design

Main area (OA - Other applications in OR)

1. Introducción

En Diseño Territorial DT, el diseño de zonas ocurre cuando pequeñas áreas o unidades geográficas deben ser agrupadas en zonas más grandes y que resulten aceptables de acuerdo a los requerimientos propios del problema. Estos criterios pueden ser la generación de zonas conexas que contengan aproximadamente la misma cantidad de habitantes, clientes, medios de comunicación, servicios públicos, etcétera. El diseño de zonas o DT aparece en diversas aplicaciones (creación de zonas escolares, zonas para análisis socioeconómico, diseño de territorios de ventas, servicios o mantenimiento, diseño geográfico de censos, problemas de asignación-localización, etc.).

Un caso bien conocido y demandado es la distritación política que persigue agrupar unidades geográficas en un número predeterminado de zonas o distritos donde los distritos electorales contengan el mismo número de habitantes (homogeneidad en los grupos para número de votantes), que sean conexos y geométricamente compactos. Por otra parte, el problema de diseño de zonas o DT es computacionalmente difícil en cuanto al tamaño del espacio solución. Aún para un pequeño número de unidades geográficas y zonas, la cantidad de posibles arreglos es muy grande.

En este trabajo abordamos el problema de diseño de zonas donde el agrupamiento implícito se caracteriza por ser compacto y homogéneo. Se atienden dos casos de homogeneidad: balanceo en el número de unidades geográficas que conforman los grupos y homogeneidad para variables descriptivas poblacionales provenientes de un censo. Se discuten aspectos de compacidad y homogeneidad proponiendo finalmente el agrupamiento compacto y homogéneo para variables poblacionales y número de objetos cuando son grupos pequeños los que se agrupan.

El presente trabajo se encuentra organizado como sigue: Introducción como la sección 1. El problema de clasificación con compacidad se aborda en la sección 2 para continuar en la sección 3 donde se trata el problema de agrupamiento homogéneo. En la sección 4 se presenta el problema de agrupamiento compacto-homogéneo y finalmente en la sección 5 se exponen las conclusiones.

2. Compacidad

Cada problema de diseño territorial presenta características diferentes con algunos criterios comunes que obedecen a la definición de DT. Por ejemplo, cuando en términos coloquiales se plantee un problema que requiera agrupar zonas donde se respeten límites de traslados, las zonas dentro de los grupos estén cercanas, los grupos tengan un equilibrio en cuanto a la población o en el número de zonas o cualquier otro aspecto subyacente, entonces es posible afirmar que tal problema se coloca como uno de diseño de zonas. Sin embargo, presentar una revisión de los trabajos DT merece un artículo extenso (Duque, 2007). Este trabajo se ubica en exponer una breve discusión de 2 criterios en DT: compacidad y homogeneidad.

El concepto de compacidad puede ser adaptado con distintos enfoques según el problema a tratar, de tal modo que los autores propongan conceptos y métodos adecuados pero sin distraerse del concepto topológico-geométrico de compacidad. Las aplicaciones de compacidad tienen que ver con disminución en los tiempos de traslado entre las unidades geográficas que conforman los grupos requeridos para visitar a los clientes o usuarios, y para este propósito se minimiza la distancia entre las unidades geográficas y el centro de la zona a la que pertenecen. Por otro lado, en el diseño de zonas electorales la compacidad es vista como una herramienta para la prevención de la manipulación electoral y las medidas propuestas buscan minimizar la desviación de la "forma" de las zonas con respecto a figuras geométricas como círculos, cuadrados o hexágonos.

Los principios de equilibrio poblacional y conexidad resultaron insuficientes para evitar la manipulación de las zonas electorales (Sherstyuk, 1998). Entonces la compacidad geométrica se ha propuesto como una condición que busca evitar la creación de zonas electorales con formas irregulares y así prevenir la generación de planes de zonificación favorables a un candidato o partido. También se considera que la compacidad favorece la claridad en la delimitación de las zonas y se ha apreciado que las zonas compactas son más fáciles de administrar por los representantes electos debido a que se disminuyen los tiempos de traslado y las dificultades de comunicación (Bélanger & Eagles, 2001).

No ha sido definida de manera precisa compacidad geométrica y muchos autores han dedicado espacio para describirla cuantitativamente en función del problema En Richard G. Niemi en 1990 pueden verse más de 20 medidas diferentes (Niemi et al, 1990).

De manera intuitiva, podemos entender compacidad geométrica cuando imaginamos que la forma de cada agrupación de zona se parezca a un cuadrado un círculo o a una figura geométrica convexa, esto significa que, en algunos casos, las medidas propuestas hasta el momento no sean totalmente convincentes (Young, 1988).

Em este punto, la compacidad la hemos definido informalmente como sigue:

Decimos que existe compacidad cuando se reúnen muy estrechamente varios objetos de modo que no queden huecos o que queden pocos de tal modo que se cumpla con la idea de que los objetos hacia un centro que representa un grupo tengan una distancia mínima

Para formalizar esta idea cuando los objetos son AGEBs, se define primero distancia entre grupos.

Si denotamos por $Z = \{1,2,..,n\}$ al conjunto de n objetos a clasificar, se trata de dividir Z en k grupos $G_{1,} G_{2,} .., G_k = Z$ con $k < n$ de tal forma que:

* $\bigcup_{i=1}^{k} G_i = Z$ * $G_i \cap G_j = \phi$ * $i \neq j$ $|G_i| \geq 1$ $i = 1,..,k$

Un grupo G_m con $|G_m| > 1$ es compacto si para cada objeto $t \in Gm$ cumple

$$\underset{i \in G_m, i \neq t}{Min}\, d(t,i) \;<\; \underset{j \in Z - G_m}{Min}\, d(t,j)$$

Un grupo G_m con $|G_m| = 1$ es compacto si su objeto t cumple $\underset{i \in Z - \{t\}}{Min}\, d(t,i) \;<\; \underset{j,l \in Gf\ \forall\ f \neq m}{Min}\, d(t,l)$

Considerando como objetos los AGEBs, en la figura 1 notamos que dos AGEBs cumplen las restricciones anteriores. El grupo azul que solamente consta de un AGEB esta bien clasificado porque la distancia entre el AGEB azul y el AGEB más cercano es mayor que la distancia mayor entre dos AGEBs de un mismo grupo. Para el grupo rosa que tiene 7 AGEBs, puede observarse que el AGEB t de este grupo ha sido bien clasificado porque la menor distancia que hay entre t y el AGEB más cercano de su mismo grupo es menor que la distancia entre este y el AGEB j mas cercano del grupo verde.

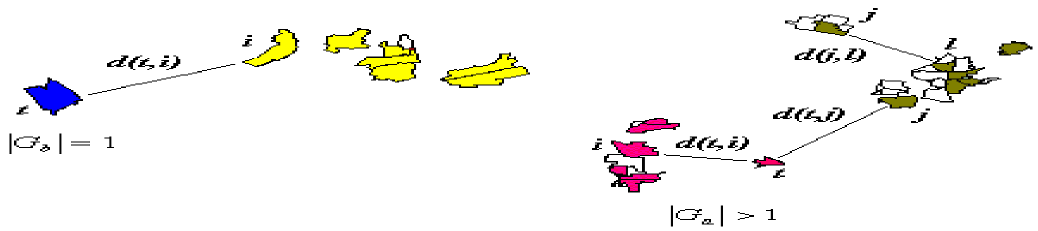

Fig. 1 Cuatro grupos compactos

Tratamiento de la compacidad (algoritmo 1)

La idea central de este algoritmo consiste en reducir el problema a la compacidad para 2 objetos de forma iterativa (aquellos con la menor distancia entre sí). Se consideran los pares de objetos más cercanos entre ellos y se colocan en un grupo nuevo si ninguno de los dos pertenece a un grupo. En otro caso, uno de ellos formará parte del grupo del objeto contrario si ya se ha asignado a un grupo y el otro no. Cuando dos objetos pertenezcan a un grupo diferente, se fusionan los dos grupos. A pesar de la gran cantidad de cálculos y comparaciones, es posible dar respuesta a esta complejidad computacional almacenando las distancias entre los objetos en una base de datos con bajo costo en espacio de memoria. Si los objetos son las AGEBs, entonces aproximadamente cada estado ocupa alrededor de 4 mega bytes de memoria.

Algoritmo 1 de agrupamiento bajo condiciones de compacidad

1. $C = \{ (x,y) \mid x \in O \wedge y \in O \wedge x \neq y \wedge (y,x) \notin C \}$
2. Ordenar ascendentemente a C por las distancias de los pares de objetos.
   $m \leftarrow 0$
   $empty \leftarrow 0$
3. Para cada par de objetos $x,y \in C$ hacer
   3.1. $ng \leftarrow |O| - (|G_1| + \ldots + |G_m|) + m - empty$
   3.2. si $|G_1| + \ldots + |G_m| = |O|$ entonces salir
   3.3. si $x \notin G_1 \cap \ldots \cap G_m \;\wedge\; y \notin G_1 \cap \ldots \cap G_m$ entonces
        $m \leftarrow m+1$
        $G_m = \{x,y\}$. Ir a 3.7
   3.4. si $x \notin G_1 \cap \ldots \cap G_m \;\wedge\; y \in G_1 \cap \ldots \cap G_m$ entonces
        $G_y = G_y \cap \{x\}$; $y \in G_y$. Ir a 3.7
   3.5. si $x \in G_1 \cap \ldots \cap G_m \;\wedge\; y \notin G_1 \cap \ldots \cap G_m$ entonces
        $G_x = G_x \cap \{y\}$; $x \in G_x$ .Ir a 3.7
   3.6. si $x \in G_1 \cap \ldots \cap G_m \;\wedge\; y \in G_1 \cap \ldots \cap G_m \wedge x \neq y \wedge ng > k$ entonces
        $G_x = G_x \cap G_y\, G_y = \varnothing$
        $empty \leftarrow empty+1$
   3.7. si $ng = k$ entonces
          Para cada $z \in [\, |O| - (|G_1| + \ldots + |G_m|) \,]$ hacer
                $m \leftarrow m+1$
                $G_m = \{z\}$

En la siguiente figura 2 puede verse una prueba para 30 grupos.

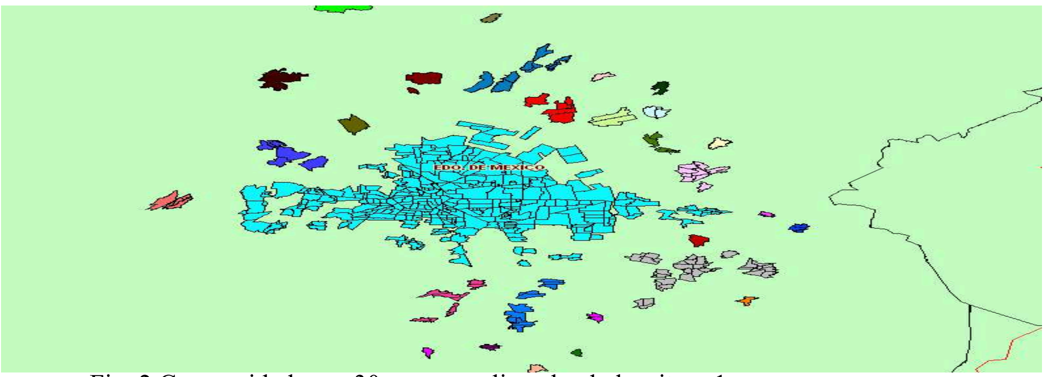

Fig. 2 Compacidad para 30 grupos aplicando el algoritmo 1

Propuesta 2 para algoritmo de compacidad

Considerando los aspectos generales de k-medias, esta propuesta algorítmica parte de un conjunto de objetos y a partir de la distancia entre ellos, consigue obtener grupos de tal manera que la distancia entre los objetos pertenecientes a un grupo sea menor que la distancia hacia un objeto de otro grupo.

El algoritmo comienza buscando *m* representantes de grupo ($m \leq k$), dividiendo el estado correspondiente en *m* cuadrados iguales, así, el centro de cada cuadrado "atrae" al objeto más cercano convirtiéndolo en un representante de grupo. Estos representantes quedan distribuidos uniformemente en el estado y se adapta el número de representantes al número de grupos para después asignar el resto de los objetos a un grupo, el cual tenga la distancia menor entre uno de sus objetos y el objeto que se va a clasificar.

Algoritmo de compacidad 2

1. Encontrar a los representantes de cada grupo.
   1.1 Formar una rejilla de m cuadrados iguales con las coordenadas (x_o,y_n) en la esquina superior izquierda y (x_e,y_s) en la esquina inferior derecha.
   $$m = \left|\sqrt{k}+1\right| \times \left|\sqrt{k}+1\right|$$
   1.2 Para cada cuadro de la rejilla hacer
   -Sea x,y las coordenadas del centroide del cuadro actual de la rejilla.
   -Crear un nuevo grupo, y asignar como representante de este grupo al objeto cuya distancia hacia el punto x,y sea mínima.
   1.3 Mientras el número de representantes de grupo sea mayor a k hacer
   -fusionar en un solo grupo a los 2 representantes más cercanos entre si
2. Asignar al resto de objetos a un grupo.
   2.1 para cada objeto O que no sea representante de grupo hacer
   - Sea C_i el objeto del grupo i que está más cercano a O
   - asignar al objeto O al grupo al que pertenece el objeto con
   $$\min d(O,c_i) \forall_{1=1,2,\ldots,k}$$

donde:
k: Número de grupos a obtener, n: Número de objetos a clasificar, d(o,o_2): Distancia entre el objeto *o* y el objeto o_2, x_o: Longitud del objeto con la longitud más baja, x_e: Longitud del objeto con la longitud más alta, y_n: Latitud del objeto con la latitud más alta, y_s: Latitud del objeto con la latitud más baja. La obtención de x_o, x_e, y_n y y_s es realizada mediante consultas SQL a la base de datos.

En este algoritmo es necesario que antes del paso 2, se ordenen los objetos. Para este problema, las AGEBs se organizan por municipio, y dentro de cada municipio por localidad dado que la distribución geográfica de las localidades dentro de un municipio contribuye a que se agrupen primero los objetos más cercanos a los representantes. En la siguiente figura 3, se muestra un ejemplo para 30 grupos compactos.

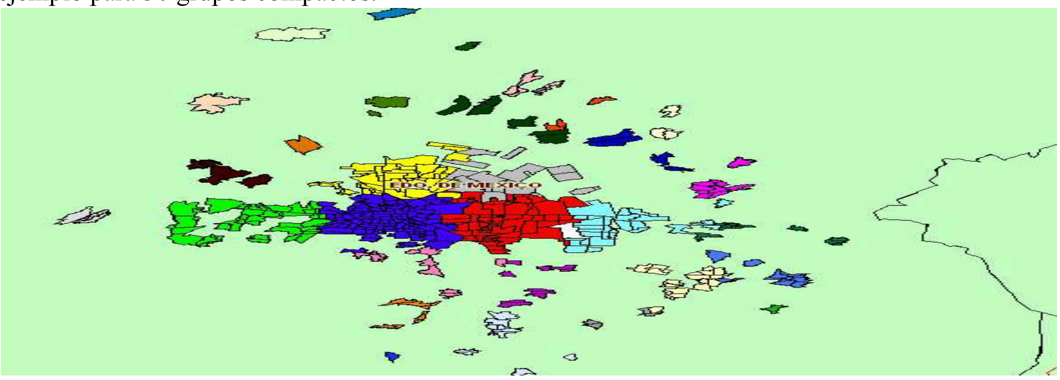

Fig. 3 Compacidad para 30 grupos bajo la aplicación del algoritmo 2

El problema adquiere mayor dificultad cuando se incluye otra medida asociada a compacidad. En este trabajo discutimos la relación entre homogeneidad y compacidad para la agrupación de áreas geoestadísticas geográficas donde se abordan por separado dos condiciones de homogeneidad: equilibrio para variables poblacionales y balanceo para el número de objetos que componen los grupos.

3. Homogeneidad

Existe imprecisión en cuanto a homogeneidad y equilibrio. Lo que se trata de encontrar es que los grupos de zonas tengan equilibrio o balanceo con respecto a una o varias propiedades de las unidades geográficas que la forman. Por ejemplo, pueden diseñarse zonas que tengan la misma carga laboral, los mismos tiempos de traslado o los mismos porcentajes de representación étnica o socio-económica. En general, no es posible lograr el equilibrio exacto, entonces es sensato calcular la desviación con respecto al arreglo ideal. Cuanto mayor sea la desviación, peor es el equilibrio de la zona o del plan de zonificación generado.

Equilibrio Poblacional

Se han sugerido varios métodos para calcular el equilibrio poblacional de las zonas pero todos ellos conducen a resultados muy parecidos, por lo cual existe poca discusión con respecto a las ventajas de aplicar un método específico.

A continuación se listan algunas fórmulas para homogeneidad:

1. La forma más sencilla de medir el equilibrio poblacional consiste en sumar los valores absolutos de la diferencia entre la población de cada zona y el promedio poblacional por zona

$$\sum \left| P_i - P \right|$$

donde: P_i = Población de la zona i, $\overline{P}$ = Promedio de población por zona dado por $\overline{P} = \sum_{k \in K} P_k / n$, n = Número de zonas a crear , K = Conjunto de todas las unidades geográficas P_k = Población de la unidad geográfica k.

2. La diferencia de población entre la zona más poblada y la menos poblada $\underset{MAX}{P_i} - \underset{MIN}{P_k}$

En algunas ocasiones esta diferencia es dividida entre el promedio poblacional $\underset{MAX}{P_i} - \underset{MIN}{P_k} / \overline{P}$

3. La división de la zona más poblada entre la menos poblada $\underset{MAX}{P_i} / \underset{MIN}{P_k}$

4. El siguiente método está dado por la función

$$\sum_{j \in J} \max \left\{ P_j - (1+\beta)\overline{P}, (1-\beta)\overline{P} - P_j, 0 \right\} / P$$

donde: J = Conjunto de todas las zonas, P_j = Población de la zona j, $\overline{P}$ Promedio de población por zona, β = Porcentaje de desviación estándar poblacional. De esta forma se busca que la población de cada zona se encuentre dentro del intervalo $\left[(1-\beta)\overline{P}, (1+\beta)\overline{P}\right]$ con 0≤β≤1. Se observa que esta función tomará el valor de 0 si la población de cada zona se encuentra en el intervalo $\left[(1-\beta)\overline{P}, (1+\beta)\overline{P}\right]$. En otro caso tomará un valor positivo igual a la suma de desviaciones con respecto a estas cotas (Eric, 2009).

Homogeneidad para variables

La homogeneidad para variables de las AGEBs se entiende cuando los valores de una población tengan las mismas características según las variables que se vayan a considerar en el estudio. La clasificación de zonas geográficas (AGEBs) en grupos homogéneos persigue que cada grupo mantenga aproximadamente la misma cantidad de una variable poblacional, por ejemplo, que cada grupo tenga aproximadamente el mismo número de habitantes.

Esta clasificación es útil cuando se quiere obtener zonas a las que se les quiere asignar cargas de trabajo o recursos equitativamente, por ejemplo, supóngase que se quiere suministrar a un estado de la republica un servicio *x* que lo provee la maquina *m*; si la maquina *m* solo puede atender *v* viviendas, entonces se necesitarían *totalViviendas/v* maquinas para suministrar a todo el estado. Una respuesta sencilla reside en dividir el estado en *totalViviendas/v* grupos que mantengan homogeneidad en cuanto a la variable poblacional "*numero de viviendas*", de este modo es posible asignarle a cada grupo (zona) una máquina.

Homogeneidad para variables poblacionales

Si denotamos por $Z=\{1, 2, \ldots, n\}$ al conjunto de n objetos a clasificar y a $V_1, V_2, \ldots V_n$ al conjunto de valores de la variable de la cual se quiere mantener homogéneos a los grupos, se trata de dividir Z en k grupos $G_1, G_2, \ldots, G_k$ con $k<n$, de tal forma que:

1) $\bigcup_{i=1}^{k} G_i = Z$
2) $G_i \bigcap G_j = \phi \qquad i \neq j$
3) $|G_i| \geq 1 \qquad i = 1, 2, \ldots, k$

Un grupo G_m es homogéneo si cumple:

$$\sum_{j \in G_m} V_j \approx \frac{1}{k} \sum_{i=1}^{n} V_i$$

Cuando solo se considera la homogeneidad, los resultados no son prácticos, principalmente por la ausencia de la compacidad.

La dispersión de los objetos es una consecuencia esperada como puede observarse en la siguiente figura 4 donde la homogeneidad responde al mismo número de viviendas.

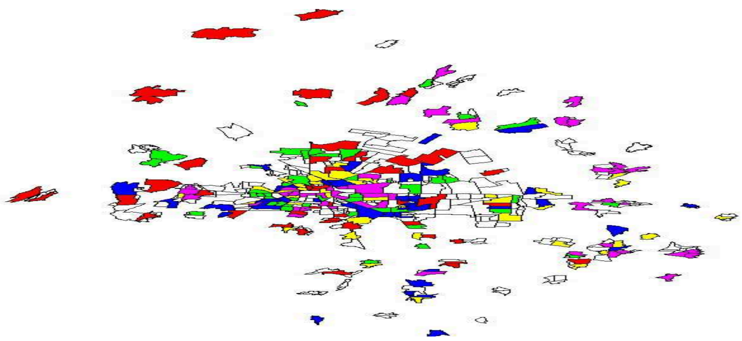

Fig. 4 Solución a la clasificación homogénea en la que los AGEBs que componen cada uno de los cuatro grupos se encuentran dispersos por todo el estado. Cada grupo tiene aproximadamente el mismo número de viviendas.

Experimentando diversas implementaciones similares, se observó que cuando a los algoritmos 1 y 2 de compacidad se le incorporan restricciones de homogeneidad, se clasifican primero los AGEBs más cercanas, en otras palabras, primero se crean grupos donde hay mayor acumulación de AGEBs que generalmente es el centro del estado. Este proceso significa dejar los objetos excedentes en el último grupo, los cuales son aquellos en las orillas del estado y por tanto este último grupo tendrá AGEBs dispersos por las orillas del estado.

En este punto, se ha diseñado un nuevo algoritmo que descarta las ideas centrales de los algoritmos 1 y 2 de compacidad. El propósito ahora es tener como prioridad a la homogeneidad tratando de conservar compacidad entre todos los grupos creados.

Cuando existe mayor acumulación de AGEBs en el centro del estado, estos AGEBs deben agruparse al final justo por su naturaleza geográfica de cierta compacidad. El objetivo es comenzar a clasificar los AGEBs que se encuentran en la periferia del estado considerado los dos AGEBs más alejados entre ellos y se crean 2 grupos homogéneos-compactos asignando a estos grupos los AGEBs más cercanos a cada uno de ellos hasta que tengan el valor promedio de grupo. Una vez creados estos grupos, nuevamente se toman del resto de AGEBs a los 2 AGEBs más alejados y se crean otros 2 grupos homogéneos-compactos de manera iterativa hasta agrupar todos los objetos.

Algoritmo de Homogeneidad para variables y Compacidad
Sea O el conjunto de todos los objetos a clasificar.
Sea k el número de grupos a obtener.
Sea G_i el i-ésimo grupo creado.
Sea d(x,y) la distancia entre el AGEB x y el AGEB y.
Sea V_i el valor del i-ésimo AGEB de la variable elegida para mantener homogeneidad.
Sea $\delta(G) = \sum_{i \in G} V_i$

1.

1.1 $avg = \frac{1}{k}\sum_{i=1}^{n} V_i$

1.2 $\xi = \frac{1}{n}\sum_{i=1}^{n} V_i$

2. si k = 0 entonces salir
si k = 1 entonces
para cada objeto x ∈ O hacer
$G_k = G_k \cap \{x\}$
fin para
fin si
si k > 1 entonces
sea x,y los objetos con el máximo d(x,y) ∀x,y ∈ O
$G_k = \{x\}$
$G_{k-1} = \{y\}$
C = { (a,b) | ((a=x V a=y) Λ b∈O) V ((b=x V b=y) Λ a∈O) Λ a≠b}
ordenar a C ascendentemente por d(a,b)
para cada par de objetos (a,b) ∈ C hacer
si a=x V b=x entonces g←a sino g←b
si a=x V a=y entonces item←V_b sino item←V_a
si $\delta(G_g) + item < avg + \xi$ entonces
$G_g = G_g \cap \{item\}$
O = O - {item}
fin si
si $\delta(G_k) > avg - \xi \ \Lambda\ \delta(G_k) < avg + \xi \ \Lambda$
$\delta(G_{k-1}) > avg - \xi \ \Lambda\ \delta(G_{k-1}) < avg + \xi$ entonces
k←k-2
ir a 2
fin si
fin para

En la figura 5 se muestra el resultado de aplicar este algoritmo al problema *población femenina por encima del promedio* sobre el estado de Toluca con k=16 y la variable *"población total"* sobre la que se mantiene homogeneidad entre 62579 y 65204.

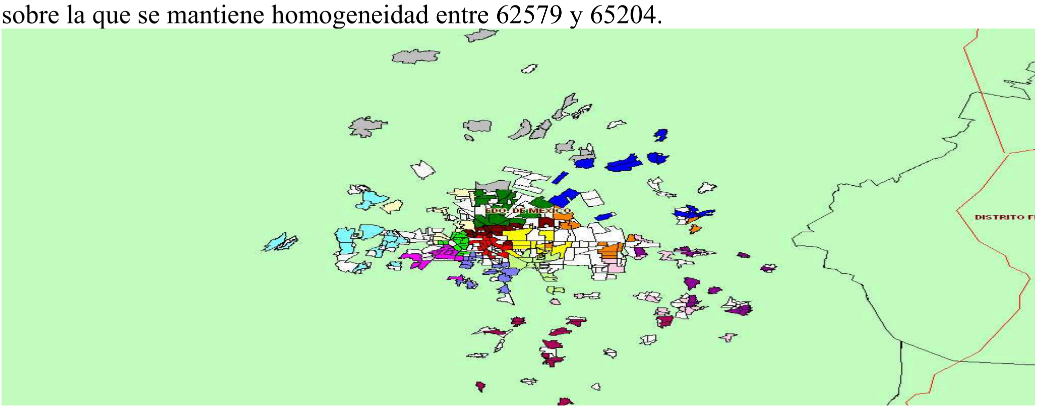

Fig. 5 División de 16 grupos a la consulta población femenina por encima del promedio, manteniendo homogeneidad en la variable *"población total"*.

4 Problema Compacidad y Homogeneidad para número de objetos

Se ha insistido informalmente que en el agrupamiento geográfico la compacidad consiste en asignar cada unidad geográfica *i* al representante de grupo (centroide) más cercano. Visto como un problema de optimización, la función objetivo es minimizar la distancia total, es decir, la suma de las distancias de cada unidad geográfica a su respectivo centroide (Bernábe et al, 2009). Tratando el problema de esta manera se logran formar grupos compactos de unidades geográficas, sin embargo, muchas aplicaciones requieren que exista equilibrio para los números de objetos que componen los grupos (grupos con la misma cantidad de elementos). Entonces, para n unidades geográficas y k grupos a formar, cada grupo deberá tener n/k miembros cuando n es par y (n/k) + 1 cuando es impar. A este problema lo denominamos homogeneidad en cuanto al número de elementos. La combinación de compacidad y homogeneidad es tratada en esta sección.

Propuesta de Solución

El particionamiento es un método bien conocido que ha sido implementado para resolver problemas de agrupamiento compacto. El particionamiento se caracteriza porque el número de grupos a formar se conoce apriori. A partir de implementaciones que ya han sido reportadas (Bernábe et al, 2009), se retoman las condiciones del particionamiento para incorporar la restricción de homogeneidad asumiendo que la complejidad computacional se eleva considerablemente. En este sentido, se ha resuelto el problema de compacidad-homogeneidad en el número de objetos para instancias de tamaño pequeño (en la práctica se ha notado que estas no deben exceder a 30 grupos).

Estudio de un algoritmo de particionamiento para lograr compacidad.

PAM (Partitioning Around Medoids) es el algoritmo de particionamiento que se ha elegido para incorporar la homogeneidad en el número de objetos que conforman los grupos (Struyf et al, 1997). Las bondades de PAM consisten en que a diferencia de k-medias, PAM construye una solución inicial de k medoides o centroides, a partir de la cual se intercambian los medoides/centroides por los demás elementos y se conservan las soluciones que disminuyen el costo total de la solución mientras que k-medias trabaja en base a una solución inicial generada aleatoriamente. A esta etapa cuando se genera la solución inicial se le conoce como "BUILD-step" y a la etapa de intercambio de objetos seleccionados (medoides) por objetos no seleccionados se le llama "SWAP-step". Esta última etapa se ejecuta sucesivamente hasta que no se puede mejorar más la solución.

Modificación en PAM para lograr compacidad y homogeneidad

Se ha construido un algoritmo que actúa sobre la función objetivo de PAM. Para n elementos a agrupar y k grupos a formar, se busca que cada grupo tenga n/k elementos o máximo n/k + 1 cuando n es impar. Para un grupo $j \in \{0,\ldots,k\text{-}1\}$ definimos $sizeE_j$ como el tamaño esperado de *j* el cual es calculado en base al principio de homogeneidad.

Sea $size_j$ el tamaño actual del grupo *j*, es decir, la cantidad de elementos que están en tal grupo. Primero se obtiene el primer grupo para el cual su tamaño actual sea menor que el tamaño esperado ($size_j < sizeE_j$) y se procede a ejecutar el procedimiento recursiveHomogeneityAdjust(), cuyo algoritmo es el siguiente:

Homogeneidad para balanceo en el número de objetos

Sea *j* el índice del grupo encontrado al que le faltan elementos.

Sea toSteal la cantidad de elementos que necesita "robar" el grupo *j* para tener el tamaño Esperado ($sizeE_j$).

Sea cost el costo del a solución.

```
Procedimiento recursiveHomogeneityAdjust(j, centroids, toSteal, cost)
    i ← getClosestCentroidTo(j);
    surplus ← size_j – sizeE_j;
    si toSteal < surplus entonces
        Stack.push(j, toSteal);
        mientras !stack.empty() hacer
            Node ← stack.pop();
            List ← orderByDistance(j, cluster_i);
            para h ← 0 TO h < node.toSteal hacer
                cluster_j.add(list.head());
                cost ← updateCost();
            end loop
            i ← temp.j;
            list.clear();
        end loop
    si no
        Stack.push(j, toSteal);
        toSteal = toSteal – surplus;
        recursiveHomogeneityAdjust(i, centroids, toSteal, cost);
    fin if
fin procedimiento
```

El algoritmo consiste en obtener el centroide *i* más cercano al centroide *j*, de tal manera que *j* "quita" los elementos que le hagan falta de *i* siempre que este grupo *i* no tenga elementos suficientes para darlos a *j* y conservar en *i* el tamaño esperado ($sizeE_i$). El algoritmo continua recursivamente obteniendo el centroide más cercano k a *i* del cual, se pueden tomar los elementos necesarios para que k tenga un tamaño mayor o igual al esperado, entonces *i* puede escoger de k los elementos necesarios para satisfacer su propia demanda y la de *j*.

Las estructuras de datos importantes para este algoritmo son el stack que almacena los centroides que requieren más elementos y la lista que guarda y ordena los elementos del grupo que puede ceder algunos de estos, el ordenamiento es de acuerdo a la distancia hacia el centroide del grupo que solicita elementos.

El problema estimado en este algoritmo es la dispersión de elementos. Esta situación obedece a aquellos casos donde un grupo que contiene muchos elementos debe ceder una gran porcentaje de objetos, entonces cede a otros varios de sus elementos al grado de que el centroide llega a estar rodeado de elementos de otros grupos, ya que un grupo no puede ceder su centroide, es decir, los elementos alrededor del centroide fueron "robados". Este inconveniente de la dispersión ocurre por lo regular para casos de 20 grupos o más.

A continuación se presentan 2 casos representativos para compacidad-homogeneidad en el número de objetos para 4 y 14 grupos. Los mapas fueron obtenidos a partir de una interfaz geográfica (Zamora, 2006).

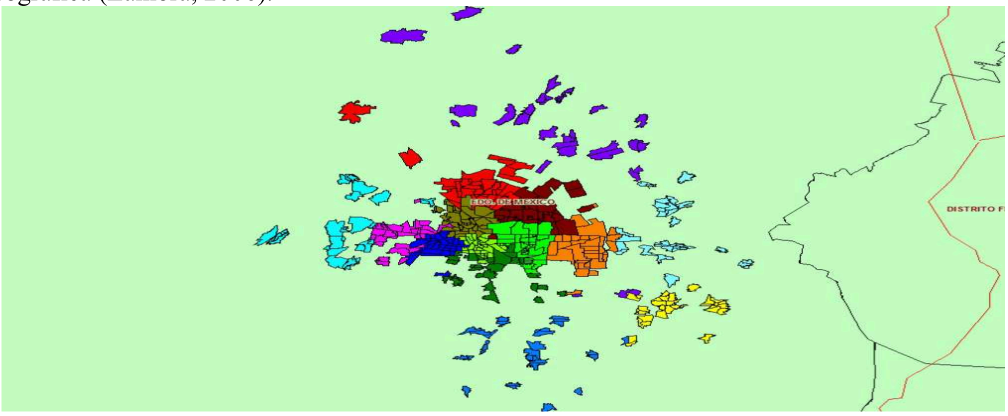

Fig. 6 Prueba para 14 grupos con PAM modificado. La homogeneidad para el número de objetos se encuentra entre 34 y 35 objetos.

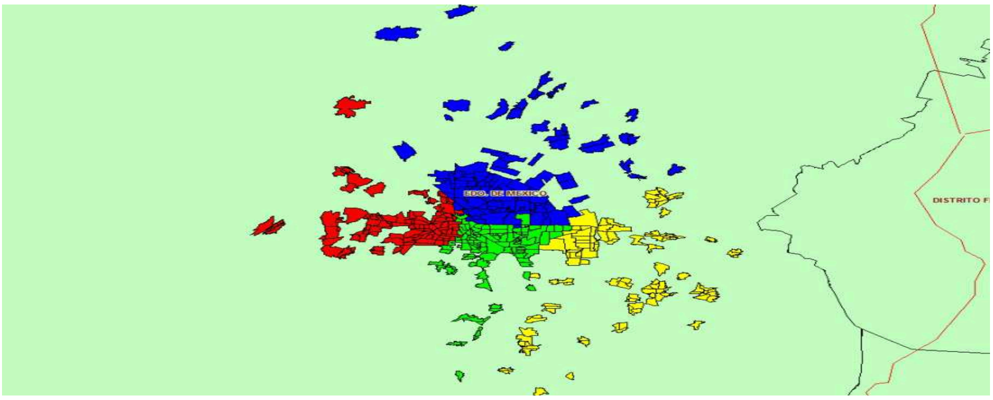

Fig 7 Corrida de 4 grupos de PAM modificado, la homogeneidad se cumple satisfactoriamente y no se observa dispersión de elementos. La homogeneidad para el número de objetos se encuentra entre 117 y 118 objetos.

5 Conclusiones

Una discusión formal sobre la incorporación de medidas de homogeneidad en algoritmos de compacidad geométrica merece un trabajo extenso. En este trabajo estamos reportando algunas de las experiencias adquiridas con el fin de construir un algoritmo compacto-homogéneo para unidades geográficas. Se han logrado buenos resultados con una homogeneidad satisfactoria para grupos de no mayores a 20. El trabajo futuro reside en responder a compacidad-homogeneidad para grupos grandes. Para ello se tienen que redoblar esfuerzos a pesar de que ya hemos construido un algoritmo híbrido de recocido simulado con búsqueda por entorno variable para homogeidad en el número de objetos (Bernábe, 2012). El conflicto aquí es que aún con el uso de heurísticos, cuando las instancias son grandes existen problemas de infactibilidad.

En la siguiente tabla se muestran los resultados del algoritmo compacto y homogéneo para número de objetos. Se observa que a pesar de que el costo de la función objetivo aumenta muy poco con PAM modificado, la homogeneidad es satisfactoria a diferencia de PAM normal. El costo computacional también se eleva pero no considerablemente.

N° de grupos	Menor número de elementos	Mayor número de elementos	Costo	Tiempo (seg)	Tipo de algoritmo
4	117	118	27.385883	4.107	PAM modificado
14	34	35	15.056902	79.536	PAM modificado
4	51	172	27.17601	0.0257	PAM
14	10	64	12.985695	3.604	PAM

Tabla 1 Resultados para 4 y 14 grupos para homogeneidad en el número de objetos

Una nueva propuesta consiste en cambiar el modelo matemático y su asociada implementación. Se ha elegido el problema del particionamiento de grafos, donde a partir de un grafo de n nodos y m aristas en k componentes de mismo tamaño, estos componentes son grafos de tamaño n/k o n/k + 1 como máximo para un número impar de nodos. La implementación de este modelo implica la generación de un grafo de acuerdo a las coordenadas de cada nodo, es decir conectar los nodos con aristas y generar la matriz de adyacencia correspondiente. Esta solución se consigue con técnicas de graficación como Triangulación de Delaunay y su dual el Diagrama de Voronoy. Cuando el grafo es obtenido, la inclusión de una meta-heurística se hace necesaria. Se ha escogido a búsqueda tabú para resolver el problema en un tiempo de cómputo razonable ya que al igual que el problema del agrupamiento geográfico este problema está clasificado como NP-difícil. Actualmente este último punto está en fase de desarrollo.

Por parte de la homogeneidad para variables (Zamora, 2006), se han tenido buenos resultados, pero no así para grupos grandes. Esta situación se encuentra en desarrollo atendiéndose con técnicas multiobjetivo.

Finalmente los resultado gráficos en mapas son el resultado de una interfaz de los algoritmos con un sistema de Información Geografico (Enrique, 2006)

Referencias

Bernábe M. B., Espinosa J. E. & Ramírez J. (2009), Evaluación de un Algoritmo de Recocido Simulado com Superficies de Respuestas, *Revista de Matemáticas Teoría y Aplicaciones,* 16(1), 159-177.

Bernábe M. B., Pinto A. D., Olivares B. E., González V. R., Martinez F. J., Vanoye R. J. (2012). A hybrid metaheuristic for the partitioning problem with homogeneity constraints on the number of objects, ICAOR 2012 (accepted, publishing on july 2012).

Bélanger Paul, Eagles Munroe (2001). The compactness of federal electoral districts in Canada in the 1980s and 1990s: an exploratory analysis. *Canadian Geographer*, 45 (4), 450-460.

Duque J. C., Ramos R., Suriñach J. (2007). Supervised Regionalization Methods: A Survey, International Regional Science Review 30 (3), 195-220.

Eric Alfredo Rincón García (2009). Diseño de zonas geométricamente compactas utilizando celdas cuadradas, Tesis Doctoral*, Posgrado de Ingeniería, Investigación de Operaciones UNAM*, 19-22.

Niemi R. G., Grofman B., Carlucci C., Hofeller T. (1990). Measuring Compactness and the Role orf a Compactness Standard in a Test for Partisian and Racial Gerrymandering. Journal of Politics, 52 (4), 1155-1181.

Sherstyuk K. (1998). How to gerrymander: A formal analysis. Public Choice, 95, 27-49.

Struyf A., Hubert M. & Rousseeuw P. J. (1997). Clustering in an object-oriented environment. *Journal of Statistical Software*, 1(4), 02-10.

Young H. P. (1988). Measuring the Compactness of Legislative Districts. Legislative Studies Quarterly, 13 (1), 105-115.

Zamora Alcocer Enrique. (2006). Implementación de un algoritmo compacto y homogéneo para la clasificación de zonas geográficas AGEBs bajo una interfaz gráfica. *Tesis de Ingeniería en Ciencias de la Computación, Benemérita Universidad Autónoma* de Puebla, Puebla, México.