

RESOLUCIÓN EXACTA DEL MODELO DEL MÁXIMO PROMEDIO PARA EL PROBLEMA DE LA DISPERSIÓN MÁXIMA

¹Sandoya Fernando

Resumen. El problema de la dispersión equitativa, o de la dispersión máxima consiste en que dado un conjunto de nodos y una función de "distancias" entre estos nodos, se trata de escoger un subconjunto de ellos de tal manera que este sea lo más "diverso" posible, en términos de alguna métrica considerada, aquí se resuelve un problema denominado del MÁXIMO PROMEDIO en el cual el número de elementos a seleccionarse también es una variable de decisión. Otros modelos han sido extensamente estudiados por otros autores Martí y Duarte [1], Glover [2], Prokopyev [3] y Resende y Martí [4]. En este artículo se estudia el modelo del MÁXIMO PROMEDIO (MAXMEAN), que en general es un problema difícil de optimización combinatorio, del tipo Strongly NP-hard. También se estudia la utilización del modelo MAXSUM para la resolución del MAXMEAN, y se determina un procedimiento exacto más eficiente para resolverlo. Para comprobar los modelos se desarrollaron instancias de prueba aleatorias para diferentes tamaños del problema. Todos los problemas en su formulación como modelo de programación matemática MIP o MIQCP fueron implementados en GAMS y resueltos con CPLEX 12.0 y con CONOPT3.

Palabras Claves. Diversidad máxima, Optimización combinatoria, Programación cuadrática.

Abstrat. The equitable dispersion problem, or the maximum dispersion problem is that given a set of nodes and a function of "distance" between these nodes, it comes to choosing a subset of them so this is as "different" possible in terms of some metric considered. Here we solve a problem called the maximum average in which the number of elements to be selected is also a decision variable. Other models have been extensively studied by other authors: Martí and Duarte [1], Glover [2], Prokopyev [3] and Resende and Martí [4]. In this paper we study the model MAXIMUM AVERAGE (MAXMEAN), which in general is a difficult combinatorial optimization, type Strongly NP-hard. To test the models developed random test instances for different sizes of the problem. All the problems in its formulation as a mathematical programming model or MIQCP MIP were implemented in GAMS and solved with CPLEX 12.0 and CONOPT3.

Key words. Combinatorial Optimization, Maximum Diversity, Quadratic Programming.

Recibido: Agosto, 2010

Aceptado: Septiembre, 2010

1. INTRODUCCIÓN

El problema de la dispersión equitativa, o de la dispersión máxima consiste en que dado un conjunto $N = \{1, 2, \dots, n\}$ de nodos y una función de "distancias" d_{ij} entre estos nodos, se trata de escoger un subconjunto M de cardinalidad m ($m \leq n$) de tal manera que M sea lo más "diverso" posible, donde el término "diverso" se explica en términos de una métrica considerada; es decir, la dispersión debe relacionarse con algún tipo de medida, en la literatura se han planteado 3 modelos distintos, denotados como: MAXMIN, MAXSUM, y MAXMINSUM, últimamente Prokopyev [3] ha planteado dos problemas adicionales: denominados del MÁXIMO PROMEDIO y MINDIFF, de los cuales en uno de ellos, el del MÁXIMO PROMEDIO, el número de elementos a seleccionarse también es una variable de decisión, y en los restantes cuatro modelos este es un valor predeterminado, estos autores enuncian los problemas pero no plantean sus soluciones.

Los modelos MAXMIN y MAXSUM han sido extensamente estudiados por Martí y Duarte [1],

Glover [2], Prokopyev [3] y Resende y Martí [4].

En este artículo se estudia el modelo del MÁXIMO PROMEDIO (MAXMEAN), que en general es un problema difícil de optimización combinatorio, del tipo *Strongly NP-hard*.

También se estudia la utilización del modelo MAXSUM para la resolución del MAXMEAN, y se determina un procedimiento exacto más eficiente para resolverlo.

Para comprobar los modelos se desarrollaron instancias de prueba para diferentes valores de n y m , en estos problemas de prueba se generaron los valores de d_{ij} , usando primero una distribución uniforme (0,20). En vista de las soluciones que se observaron en el modelo MAXMEAN para estos valores de d_{ij} , se consideró luego instancias en las cuales los valores de d_{ij} son positivos y negativos, generándolos por medio de una distribución uniforme (-10,10)

Todos los problemas en su formulación como modelo de programación matemática MIP (*Mixed integer Programming*) con variables binarias fueron implementados en GAMS y resueltos con CPLEX 12.0, y además se trató el problema MAXSUM en su formulación de programación matemática MIQCP (*Mixed Integer Quadratically Constrained Programming*) con variables binarias usando el solver en período de prueba CONOPT3.

¹ Sandoya Fernando, M.Sc., Profesor de la Escuela Superior Politécnica del Litoral (ESPOL); Coordinador de la carrera Ingeniería en Logística y Transporte ICM – ESPOL.
(e-mail: fsandoya@espol.edu.ec)

2. FORMULACIONES

Los modelos considerados son: MAXSUM IQCP, MAXSUM IP, MAX-MEAN, MAX-MIN SUM y MIN-DIFF. La formulación de estos modelos es la siguiente:

2.1 MAXSUM IQCP

El problema de la máxima diversidad originalmente fue planteado en términos de seleccionar el conjunto de elementos que maximice la suma de las distancias entre los nodos seleccionados, esta puede ser la forma más "obvia" de plantear el problema, esta línea de pensamiento dio origen al modelo de la Máxima Suma o MAXSUM, que se formula como un problema de programación cuadrática binaria de la siguiente manera:

Para $i = 1, 2, \dots, n$, la variable x_i toma el valor 1 si el elemento i -ésimo es seleccionado y 0 sino; el modelo MAXSUM IQCP es entonces formulado como un problema de programación cuadrática binaria:

$$(\text{MaxSum IQCP}) \text{Max} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} x_i x_j \quad (1)$$

$$\text{s.t.} \sum_{i=1}^n x_i = m \quad (2)$$

$$x_i = 0, 1 \quad i = 1, 2, \dots, n \quad (3)$$

2.2 MAXSUM IP

El modelo MAXSUM IQCP puede ser fácilmente linealizado introduciendo nuevas variables variables z_{ij} de la siguiente manera:

$$(\text{MaxSum IP}) \text{Max} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} z_{ij} \quad (4)$$

s.t. (2), (3) y:

$$z_{ij} \geq x_i + x_j - 1, \quad z_{ij} \leq x_i, \quad z_{ij} \leq x_j, \quad 1 \leq i < j \leq n; \quad (5)$$

$$z_{ij} \in \{0, 1\}, \quad 1 \leq i < j \leq n \quad (6)$$

2.3 MAX MIN

Seleccionar un subconjunto de dispersión máxima de cardinalidad m también puede ser entendido como la selección de los m elementos

para los cuales se maximiza la menor de las distancias d_{ij} entre los elementos seleccionados:

$$(\text{MaxMin}) \max_{M \subseteq M, |M|=m} \left\{ \min_{i,j \in M} d_{ij} x_i x_j \right\} \quad (7)$$

$$x_i \in \{0, 1\}; \quad 1 \leq i \leq n \quad (8)$$

El modelo MAXMIN puede ser linealizado y planteado como un problema de programación entera con variables binarias, como se demuestra en [8].

2.4 MIN DIFF

Para formular el problema Min-Diff introducimos el concepto del diferencial $d(i)$ como:

$$d(i) = \sum_{j \in M, j \neq i} d_{ij} \quad (9)$$

Así el problema MIN DIFF puede ser formulado como:

$$(\text{MinDiff}) \text{Min} \left\{ \max_{i \in M} d(i) - \min_{i \in M} d(i) \right\} \quad (10)$$

$$\text{s.t.} \sum_{i=1}^n x_i = m \quad (11)$$

$$x_i = 0, 1 \quad i = 1, 2, \dots, n \quad (12)$$

Y que puede ser reformulado como un problema MIP con variables binarias de la siguiente manera:

$$\min_{t,r,s,x} t \quad (13)$$

$$\text{s.t.} \quad t \geq r - s, \quad i = 1, \dots, n; \quad (14)$$

$$r \geq \sum_{i \in M} d(i) - U_i (1 - x_i) + M^- (1 - x_i), \quad i = 1, \dots, n; \quad (15)$$

$$s \leq \sum_{i \in M} d(i) - L_i (1 - x_i) + M^+ (1 - x_i), \quad i = 1, \dots, n; \quad (16)$$

$$\sum_{i=1}^n x_i = m, \quad x_i \in \{0, 1\} \quad i = 1, \dots, n; \quad (17)$$

Donde U_i y L_i cotas superior e inferior convenientemente escogidas para $d(i)$, y

M^-, M^+ son constantes grandes, estos valores pueden ser determinados y ajustados adecuadamente, como se observa en [5].

2.5 MAXMEAN

El modelo MAXMEAN, o del máximo promedio, implica seleccionar cierto número de elementos del conjunto original, de tal manera que este represente la máxima diferencia promedio. De esta manera también es variable de decisión la cantidad de elementos a seleccionarse. El modelo puede representarse como:

$$(\text{MaxMean}) \text{Max} \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} x_i x_j}{\sum_{i=1}^n x_i} \quad (18)$$

$$\text{s.t.} \sum_{i=1}^n x_i \geq 2 \quad (19)$$

$$x_i = 0, 1 \quad i = 1, 2, \dots, n \quad (20)$$

Tal como se lo presenta este es un problema de optimización binaria fraccional, pero puede ser linealizado utilizando ciertas transformaciones, así en su formulación linealizada el problema MAXMEAN se presenta como:

$$\text{Max} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} z_{ij} \quad (21)$$

$$\text{s.t.} \quad y - z_i \leq 1 - x_i, \quad z_i \leq y, \quad z_i \leq x_i, \quad (22)$$

$$z_i \geq 0, \quad i = 1, \dots, n$$

$$y - z_{ij} \leq 2 - x_i - x_j, \quad z_{ij} \leq y, \quad z_{ij} \leq x_i, \quad (23)$$

$$z_{ij} \geq 0, \quad 1 \leq i < j \leq n$$

$$\sum_{i=1}^n x_i \geq 1, \quad \sum_{i=1}^n z_i = 1, \quad x_i \in \{0, 1\}^n \quad (24)$$

3. RESOLUCION EXACTA DEL PROBLEMA MAXMEAN

La siguiente propiedad indica que si d es una distancia, la solución al problema MAXMEAN es que todos los puntos deben ser seleccionados, con lo cual la solución del problema en este caso se vuelve trivial.

PROPIEDAD: En el problema MAXMEAN si d_{ij} satisfacen la desigualdad triangular, son simétricas y no negativas, es decir d es una distancia, entonces $m = n$.

Demostración:

Se demostrará que $\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}}{n}$ es una estrictamente creciente en función de n

(1) Como ejemplo observemos el caso $n = 3$
Por demostrar:

$$\frac{d_{12} + d_{13} + d_{23}}{3} > \frac{d_{12}}{2}$$

Como: $\frac{1}{2} d_{12} < d_{12} \leq d_{13} + d_{23}$

$$\frac{3}{2} d_{12} < d_{12} + d_{13} + d_{23}$$

$$\frac{d_{12} + d_{13} + d_{23}}{3} > \frac{d_{12}}{2}$$

Lo mismo para $\frac{d_{12} + d_{13} + d_{23}}{3} > \frac{d_{12}}{2}$ y

$$\frac{d_{12} + d_{13} + d_{23}}{3} > \frac{d_{23}}{2}$$

(2) Para $n = k+1$
Por demostrar:

$$\frac{\sum_{i=1}^k \sum_{j=i+1}^k d_{ij}}{k+1} > \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}}{k}$$

Es decir por demostrar:

$$\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij} + \sum_{i=1}^k d_{i,k+1}}{k+1} > \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}}{k}$$

Es decir:

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij} + \sum_{i=1}^k d_{i,k+1} > \sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij} + \frac{1}{k} \sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}$$

Luego hay que demostrar:

$$\sum_{i=1}^k d_{i,k+1} > \frac{1}{k} \sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}$$

Como: $d_{12} \leq d_{1,k+1} + d_{2,k+1}$

$$d_{13} \leq d_{1,k+1} + d_{3,k+1}$$

$$d_{1k} \leq d_{1,k+1} + d_{k,k+1}$$

⋮

$$d_{k-1,k} \leq d_{k-1,k+1} + d_{k,k+1}$$

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij} < (k-1) \sum_{i=1}^k d_{i,k+1}$$

$$\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}}{k-1} < \sum_{i=1}^k d_{i,k+1}$$

Como: $\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}}{k} < \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^k d_{ij}}{k-1}$ ya

está!!

OBSERVACION: La propiedad anterior no es cierta si d_{ij} son positivas y simétricas pero no satisfacen la desigualdad triangular, es decir se pueden encontrar contraejemplos en los cuales para los cuales $d_{ij} \geq 0$ y simétrica y sin embargo $m < n$. Aunque estos ejemplos serían la excepción, pues al desarrollar simulaciones con los d_{ij} generados aleatoriamente en todas las simulaciones m resultó ser igual a n , tal como se observa en la tabla 2, en la cual se presentan los resultados para valores de $n = 10, 15, 20, 25$ y 30 , para cada uno de los cuales se generaron 10 instancias aleatoriamente.

Ejemplo: Sea $n = 4$ y las distancias d_{ij} dadas en la tabla 1:

TABLA I

Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima

Ejemplo con $d_{ij} \geq 0$ y simétricas con el cual la solución del MAXMEAN es $m < n$

	1	2	3	4
1	0	20	18	1
2	20	0	20	2
3	18	20	0	1
4	1	2	1	0

En este caso $m = 3$, y se seleccionan los puntos 1, 2 y 3

El mejor promedio es 19.33, considerando $X_1 = X_2 = X_3 = 1, X_4 = 0$.

En cambio seleccionando todos los puntos; es decir, $X_1 = X_2 = X_3 = X_4 = 1$ el promedio es 15.5.

TABLA II

Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima

Resultados del modelo MAXMEAN con d_{ij} seleccionados aleatoriamente U[0,20]

n	m	
	Mínimo	Máximo
10	Mínimo	10
	Máximo	10
15	Mínimo	15
	Máximo	15
20	Mínimo	20
	Máximo	20
25	Mínimo	25
	Máximo	25
30	Mínimo	30
	Máximo	30

SOLUCION DEL PROBLEMA MAXMEAN EN EL CASO NO TRIVIAL

Se considera ahora la generación de instancias con d_{ij} seleccionadas aleatoriamente con una distribución U[-10,10], y se resuelve el modelo MAXMEAN según la formulación (21) – (24), usando C-PLEX 12 con los parámetros por defecto. Los experimentos fueron desarrollados en una Laptop Intel Core Solo 1.40 GHz, 3 Gb RAM. Los resultados son mostrados en la tabla 3.

TABLA III
 Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima
 Soluciones exactas para el modelo MaxMean para tamaños variando entre $n = 10$ a $n = 30$ nodos

TAMAÑO	PARAMETRO	TIEMPO CPU (SEGUNDOS)			VALOR OBJETIVO			M		
		MAXIMUM	MEDIAN	MINIMUM	MAXIMUM	MEDIAN	MINIMUM	MAXIMUM	MEDIAN	MINIMUM
	$n = 10$	0.330	0.264	0.172	11.75	7.206	6.258	6	4	3
	$n = 15$	3.002	1.5715	0.905	14.63	11.942	7.707	9	7	3
	$n = 20$	72.009	49.1605	25.532	18.505	14.0915	11.034	11	7.5	5
	$n = 25$	2087.09	1191.711	658.721	19.215	16.923	14.503	15	9	5
	$n = 30$	> 5 hours	> 5 hours	> 5 hours	-	-	-	-	-	-

4. RELACION ENTRE EL PROBLEMA MAXSUM Y EL PROBLEMA MAXMEAN

Nótese que se podría resolver alternativamente el problema MAXMEAN a través de la resolución del problema MAXSUM de la siguiente manera: Si resolvemos óptimamente el problema MAXSUM para todos los valores de m posibles; es decir, para $m = 2, 3, \dots, n$, y dividimos estos valores óptimos para su respectivo valor de m , tendríamos todos los valores de los mejores promedios, de ahí al seleccionar el máximo valor de ellos se tendrá el óptimo del problema MAXMEAN. Curiosamente en los resultados computacionales, esta estrategia de resolver m problemas MAXSUM resultó más eficiente que resolver el problema MAXMEAN formulado en (21)-(24), tal como se observa en la tabla 4.

SOLUCION DEL PROBLEMA MAXMEAN UTILIZANDO EL MODELO MAXSUM

Se considera ahora la generación de instancias con d_{ij} seleccionadas aleatoriamente con una distribución $U[-10,10]$, y se resuelve el modelo MAXMEAN siguiendo la estrategia de resolver m problemas MAXSUM en su formulación MAXSUM IQCP (1) –(3) usando el solver de prueba CONOPT3 y MAXSUM IP (4) – (6) usando C-PLEX 12 con los parámetros por defecto. Los experimentos fueron desarrollados en una Laptop Intel Core Solo 1.40 GHz, 3 Gb RAM.

En los experimentos computacionales, al resolver los m problemas MAXSUM, y dividiendo su óptimo para m , se observó en todas las instancias generadas que estos problemas tienen sólo un óptimo local, tal como se observa en las figuras 1, 2 y 3.

Tomando en cuenta la forma que toma el máximo promedio para distintos valores de m , con un solo máximo local, se puede explotar esta característica para generar una estrategia de resolver los problemas MAXSUM para valores crecientes de m , dividir su valor para m y parar el momento en que para el siguiente valor de m este valor del promedio decrece, con lo cual se reduciría aún más el tiempo de ejecución utilizando este procedimiento.

En cambio en la tabla 4 se observan los tiempos en los cuales se obtuvieron los resultados, cuando se utilizó este procedimiento resolviendo los problemas MAXSUM con la formulación MIP y utilizando C-PLEX12, y resolviendo los problemas MAXSUM con su formulación IQCP utilizando CONOPT3. También se reporta el GAP para el cual el solver CONOPT3 dio el resultado como óptimo, nótese que en este caso también se resolvieron m problemas pero del tipo MAXSUM. Hay que destacar el hecho de que a pesar que con CONOPT3 no se garantiza que se obtenga la solución óptima exacta, en el 40% de los experimentos si se alcanzó el óptimo.

FIGURA 1
Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima
Valores óptimos del problema MAXSUM divididos para m para una instancia con $n=30$

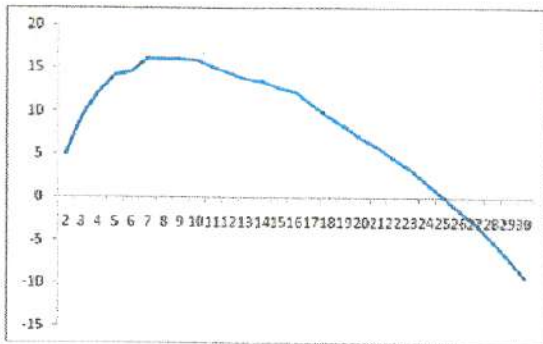


FIGURA 2
Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima
Valores óptimos del problema MAXSUM divididos para m para una instancia con $n=40$

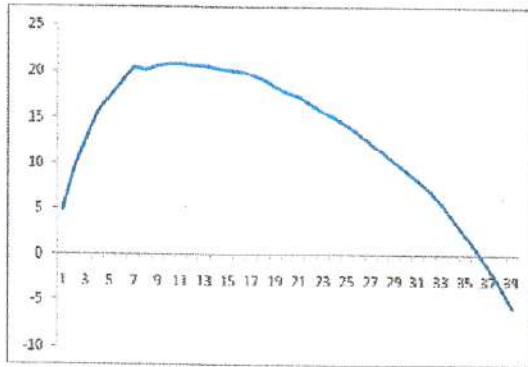


FIGURA 3
Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima
Valores óptimos del problema MAXSUM divididos para m para una instancia con $n=50$

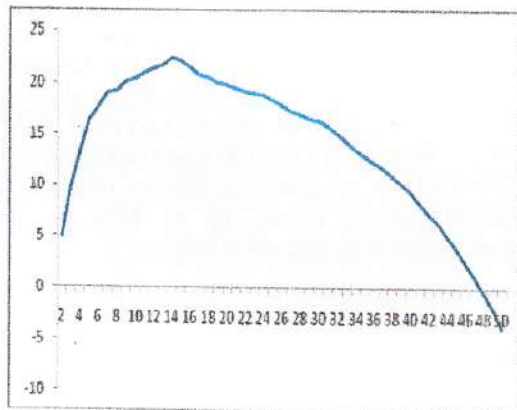


TABLA IV
Resolución exacta del modelo del máximo promedio para el problema de la dispersión máxima
Tiempos para el problema MAXMEAN al resolverlo con los tres métodos planteados

n	método	tiempo CPU	m	GAP
20	MAXMEAN	17.74	7	-
	MAXSUM IP	8.506	7	-
	MAXSUM IQCP	0.675	8	1.19%
25	MAXMEAN	788.448	7	-
	MAXSUM IP	41.884	7	-
	MAXSUM IQCP	0.783	7	1.93%
30	MAXMEAN	57723.918	7	-
	MAXSUM IP	99.408	7	-
	MAXSUM IQCP	0.757	7	5.72%
40	MAXMEAN	∞	12	-
	MAXSUM IP	745.025	12	-
	MAXSUM IQCP	1.611	10	0.95%
50	MAXMEAN	∞	14	-
	MAXSUM IP	122576.938	14	-
	MAXSUM IQCP	3.66	13	1.86%

5. CONCLUSIONES Y RECOMENDACIONES

En esta investigación se pudo determinar que el problema MAXMEAN sólo tiene sentido estudiarlo para distancias inter elemento d_{ij} negativas y positivas, pues si las d_{ij} son todas negativas la solución es siempre $m = 1$, mientras que si todas las d_{ij} son positivas es casi seguro que $m = n$.

Por otro lado resulta curioso el hecho que el modelo MIP planteado requiere un tiempo excesivo de ejecución incluso para tamaños del problema pequeños, y sólo puede ser resuelto en un tiempo prudencial para valores de n inferiores a 25. Sin embargo se puede resolver de manera exacta el mismo problema resolviendo m problemas del tipo MAXSUM en un tiempo mucho menor. Incluso, tomando en cuenta la forma que toma el máximo promedio para distintos valores de m , con un solo máximo local, podría utilizarse la estrategia de parar el momento en que para un m mayor el valor del promedio decrece, con lo cual se reduciría aún más el tiempo de ejecución utilizando este procedimiento.

La utilización de otro tipo de solvers que están en etapa de prueba, tales como CONOPT3 que utiliza procedimientos de optimización numérica no lineal, reduce drásticamente los tiempos de ejecución, en base a los tiempos alcanzados para el tamaño máximo considerado en la versión demo disponible, $n = 50$. Se especula que se

podrían resolver con la versión profesional problemas de tamaño muy grande, la desventaja es que no está garantizado que este solver alcance la solución óptima, a pesar que el programa reporta GAP cero, sin embargo si lo hizo el 40% de las veces en los experimentos computacionales.

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

- [1]. DUARTE A, MARTÍ R. (2007). "*Tabu Search and GRASP for the Maximum Diversity Problem*", European Journal of Operational Research; 178, 71-84.
- [2]. GLOVER F, KUO CC, DHIR KS. (1998), "*Heuristic Algorithms for the Maximum Diversity Problem*", Journal of Information and Optimization Sciences 1998; vol. 19, no. 1, 109-132.
- [3]. PROKOPYEV OA, KONG N, MARTINEZ-TORRES DL. (2009). "*The equitable dispersion problem*". European Journal of Operational Research; 197: 59-67.
- [4]. RESENDE MGC, MARTÍ R, GALLEGO M, DUARTE A. (2010). GRASP with path relinking for the max-min diversity problema. Computers and Operations Research; 37: 498-508.
- [5]. C-PLEX 12.0 User's Manual (2009), GAMS documents.