



---

# *PRACTICA 4*

---

Jesús Campos Márquez



2018/2019

MODELOS DE BUSQUEDA Y HEURISTICAS DE BUSQUEDA  
Universidad de Huelva



# Índice

## **1. Definición teórica y estudios de los algoritmos**

- 1.1. Algoritmo Genético Básico**
- 1.2. Algoritmo Genético CHC**
- 1.3. Algoritmo Genético Multimodal**

## **2. Resultados de los algoritmos**

- 2.1. Algoritmo Genético Básico**
- 2.2. Algoritmo Genético CHC**
- 2.3. Algoritmo Genético Multimodal**

## **3. Resultados Globales**

## **4. Conclusiones**

## 1. Definición teórica de los algoritmos

- Algoritmo Genético Básico (AG)

Se ha optado por al algoritmo genético generacional, que básicamente consiste en emplazar a la antigua población por la nueva generada tras sus cruces y posibles mutaciones. (Figura 1.1)

Su representación la vamos a basar en Orden ya que la población son permutaciones de ciudades.

Para construirlo, se genera una población inicial, de la cual se hará una selección, mediante ruleta proporcional (dotando de mayor presión selectiva a aquellos mejores individuos), que serán los padres para generar un nuevo individuo (hijo), que consta de partes de los dos padres.

Para ello se ha utilizado el operador de Cruce OX. Además, se ha añadido una probabilidad de mutación del 1%, debido a que una probabilidad demasiado grande puede generar cambios muy bruscos en la población, perdiendo aquel sujeto que pueda ser un buen individuo y un elitismo también del 10%, aunque estos parámetros han sido estudiados y se compararan gráficas de los mismos

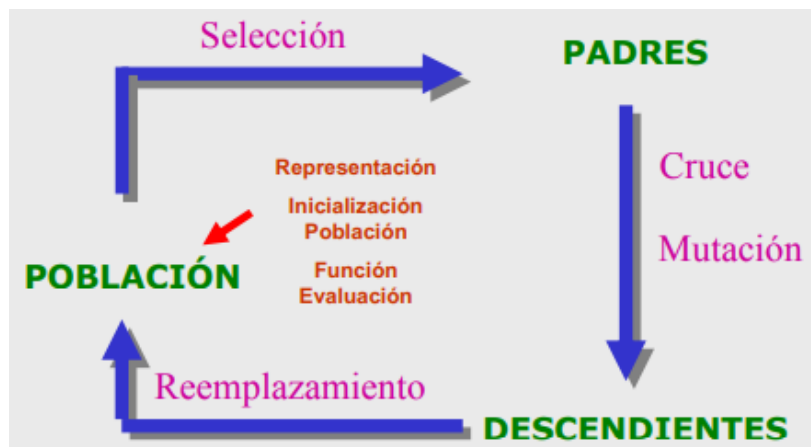


Figura 1.1

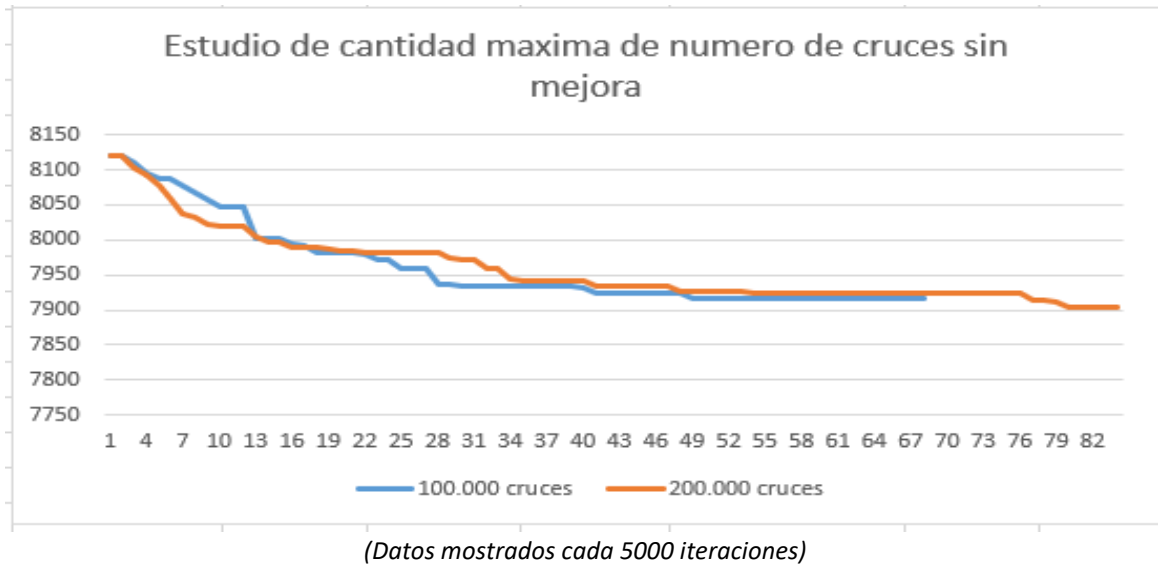
Para los siguientes estudios se han establecido como parámetros iniciales:

- Probado con poblaciones iniciales aleatorias y 1 Greedy
- 100.000 cruces sin mejora
- 10 de tamaño de la elite respecto al tamaño del problema.
- 10 de tamaño de mutación y 1 de probabilidad de mutación
- 30 de población

## MB Y HB

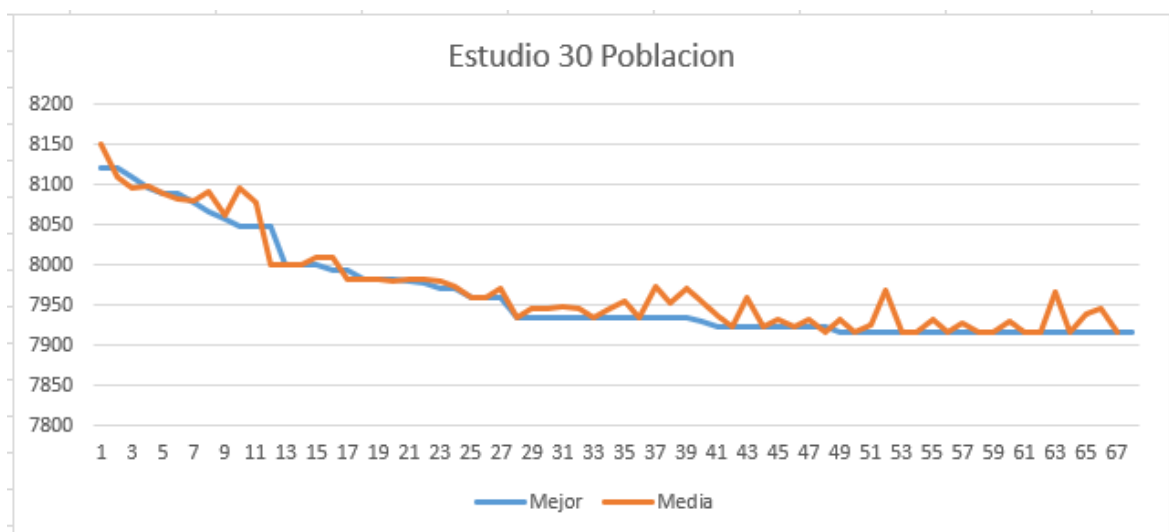
Y se han calculado las medias de 3 ejecuciones del ch130.

- Estudio del Criterio de parada/convergencia: La parada se hará efectiva cuando haya un numero de cruces que generen hijos que no mejoren a la solución actual. Para ello se ha generado un estudio en el que vemos los valores medios de coste con 100.000 cruces sin mejora y con 200.000 cruces sin mejora:



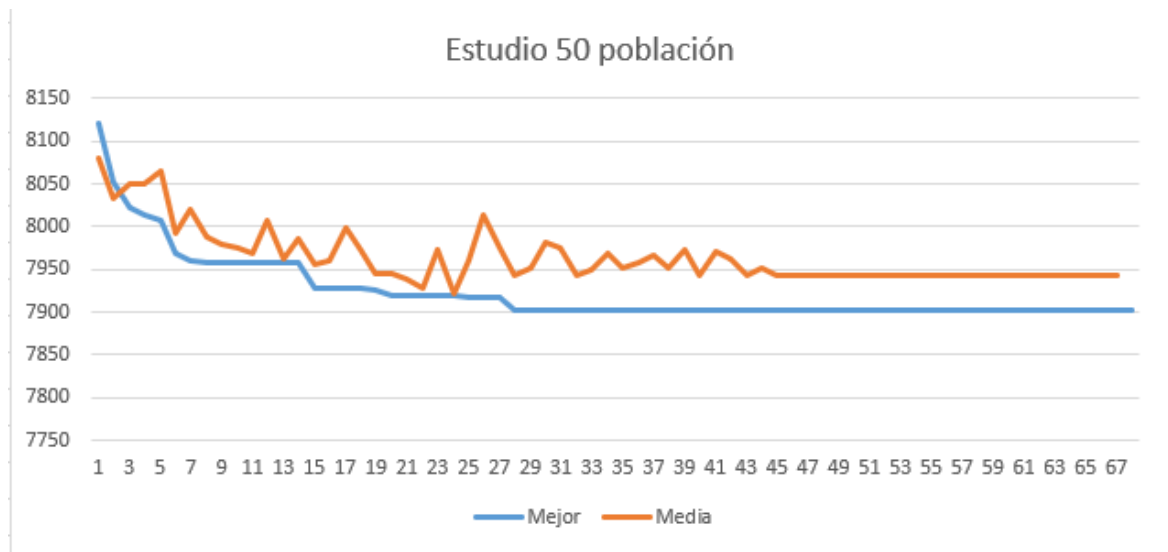
Como podemos ver, con 200.000 cruces sin mejora llega a mejores soluciones, aunque sin ser muy significativas lo tomaremos para el problema como mejor parámetro final ya que tampoco aumenta considerablemente el tiempo de ejecución del algoritmo.

- Estudio de población inicial: Para decidir si empezar con una población en un intervalo de [30,100], para ello se muestran las siguientes gráficas:

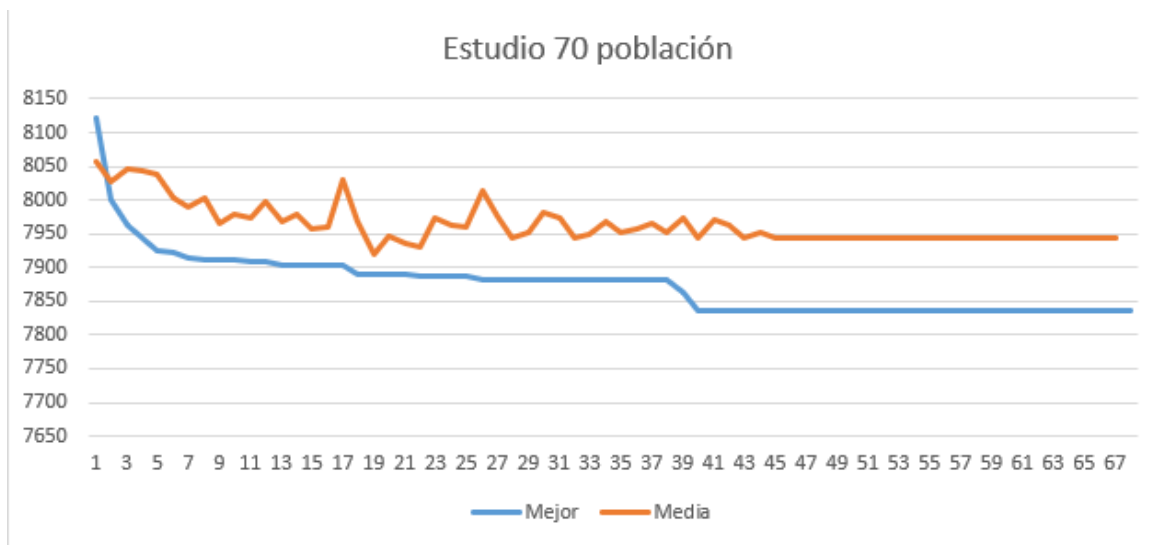


## MB Y HB

Donde vemos que la media se aproxima bastante al mejor coste, y esto es que la población se parece mucho al coste del mejor individuo.



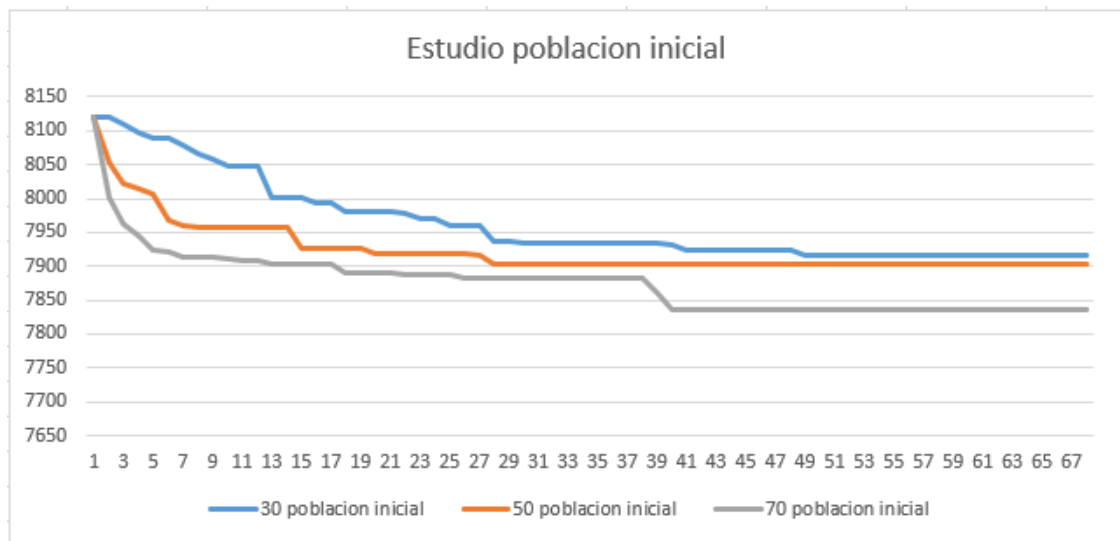
Donde también podemos ver como la media de la población se aproxima al mejor, aunque se ven saltos debidos a los aleatorios de las nuevas generaciones.



Por último, vemos el estudio para una población inicial de 70, donde también la media de las poblaciones se aproxima a la mejor población, aunque no tanto como en la población de 30, y esto se debe a la diversidad que hay en una población más grande.

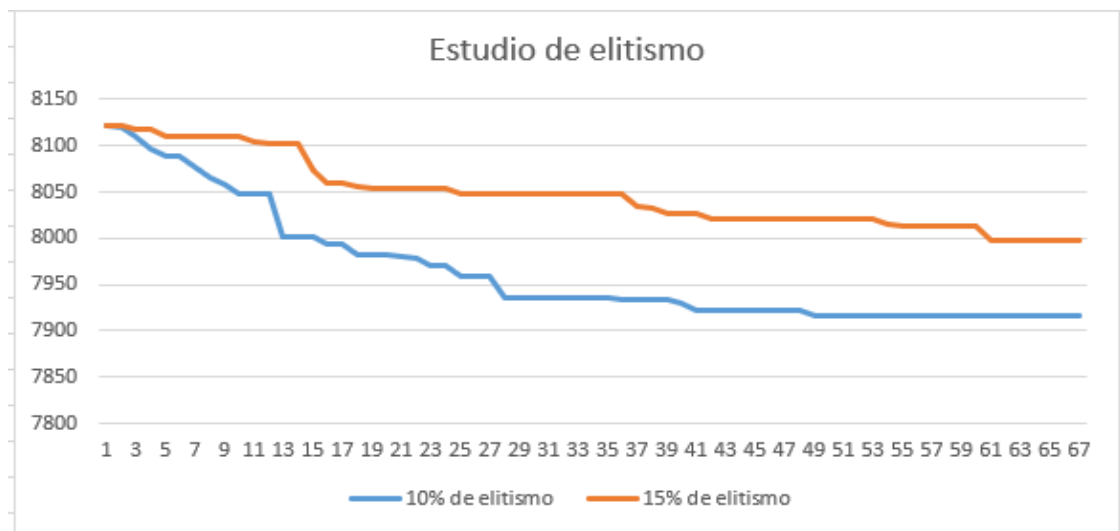
Para decidir qué parámetro vamos a utilizar vamos a comparar los resultados de los mejores en una misma gráfica:

## MB Y HB



Como se puede observar, el mejor criterio para la población es mantenerlo en unos 70 de población inicial, ya que se obtienen resultados mejores en menos iteraciones.

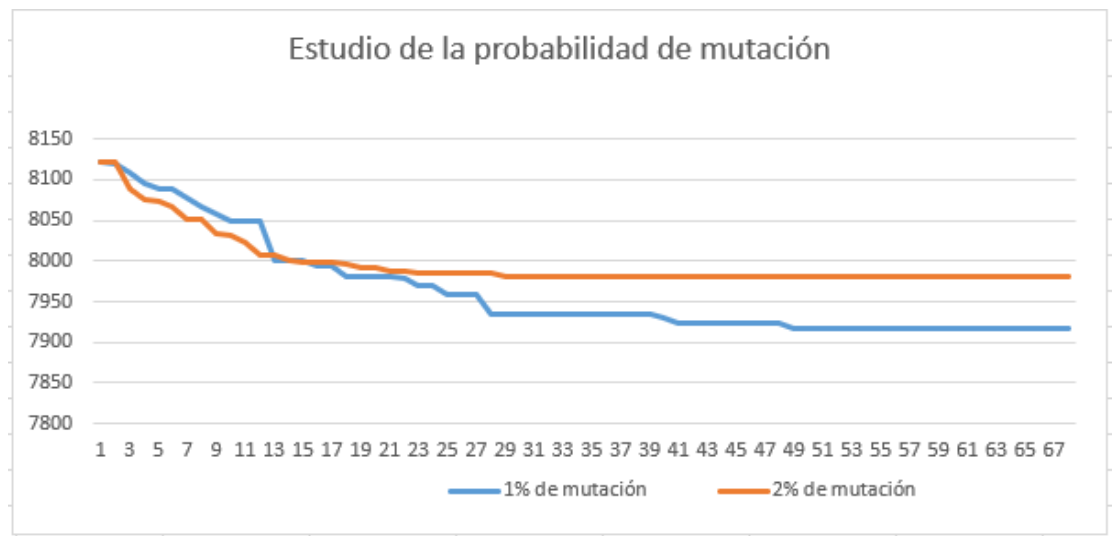
- Estudio del porcentaje de elitismo: El porcentaje de elitismo marca cuantos individuos de la anterior generación permanecerá para la nueva por ser un buen individuo:



Podemos ver que con un 10% de elitismo el algoritmo llega a una mejor solución que con un 15%, y esto es debido a que es capaz de mezclarse con individuos menos buenos para llegar a hacer una mejor solución.

## MB Y HB

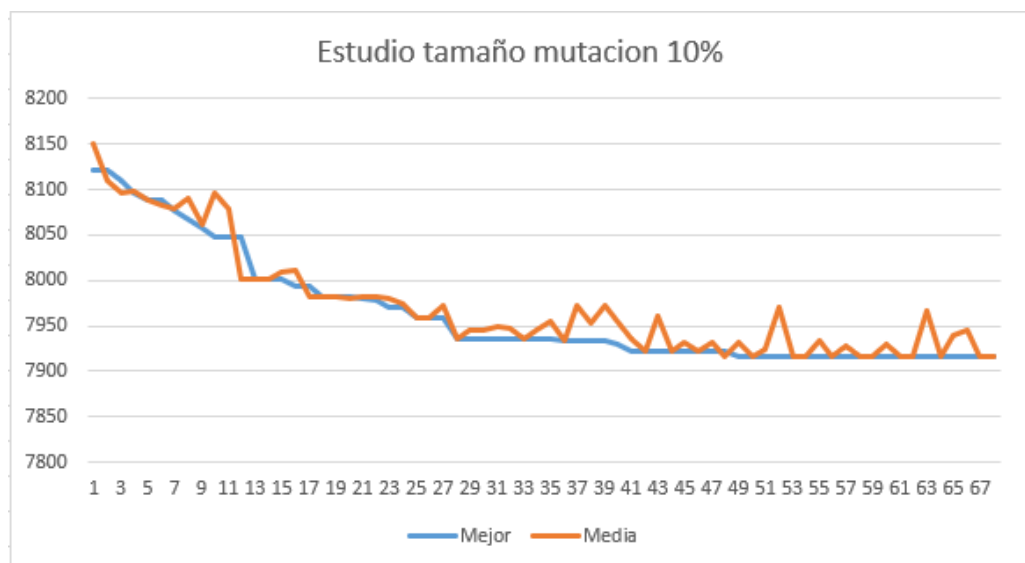
- Estudio sobre el porcentaje de mutación: El porcentaje de mutación es para que ciertas veces el hijo generado cambie considerablemente, y con ello llegar a una



mejor solución que la actual. La idea no es mutar demasiadas veces ya que buenos hijos podrían verse afectado y por lo tanto reducir la capacidad de mejora:

Observamos pues que con un 1% llega a mejores soluciones, y esto se debe a que como decíamos anteriormente, no se hagan cambios en los hijos generados muy frecuentemente.

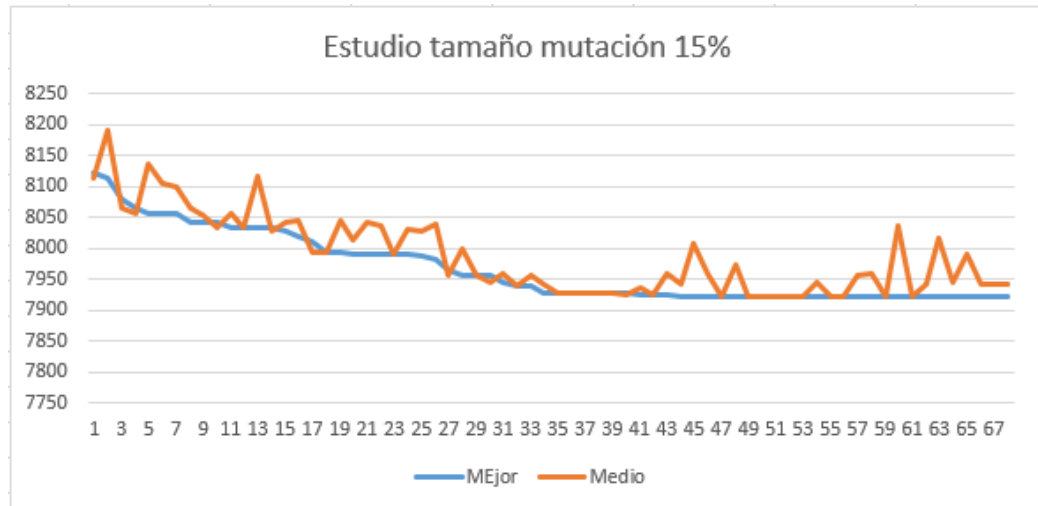
- Estudio sobre el tamaño de mutación: La mutación está compuesta por una reorganización de los cromosomas, para ello se obtiene un tamaño de (tamaño del problema)\*(tamaño de mutación). Vemos el estudio:



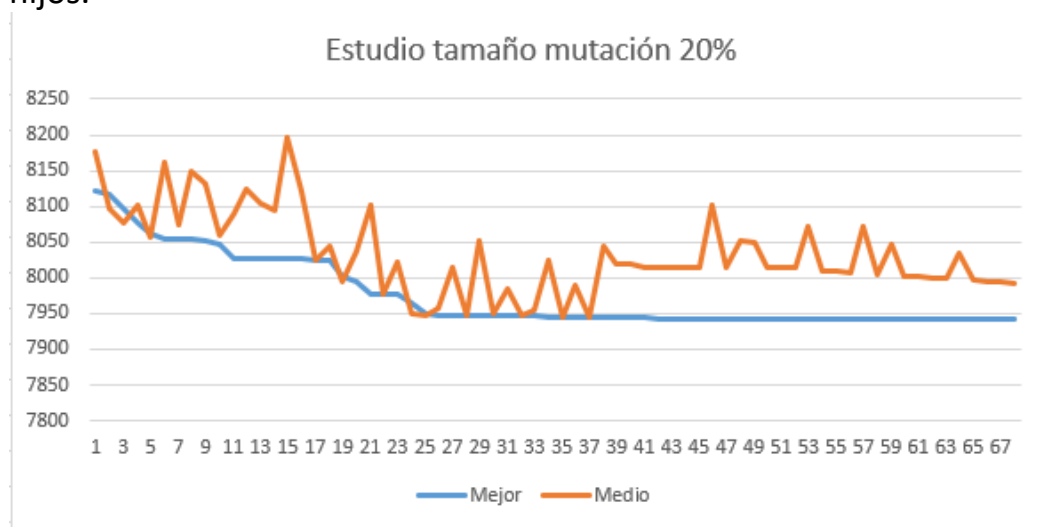


## MB Y HB

En la gráfica podemos ver que, con un tamaño de mutación relativamente pequeño, el coste medio de la población se acerca bastante al mejor.



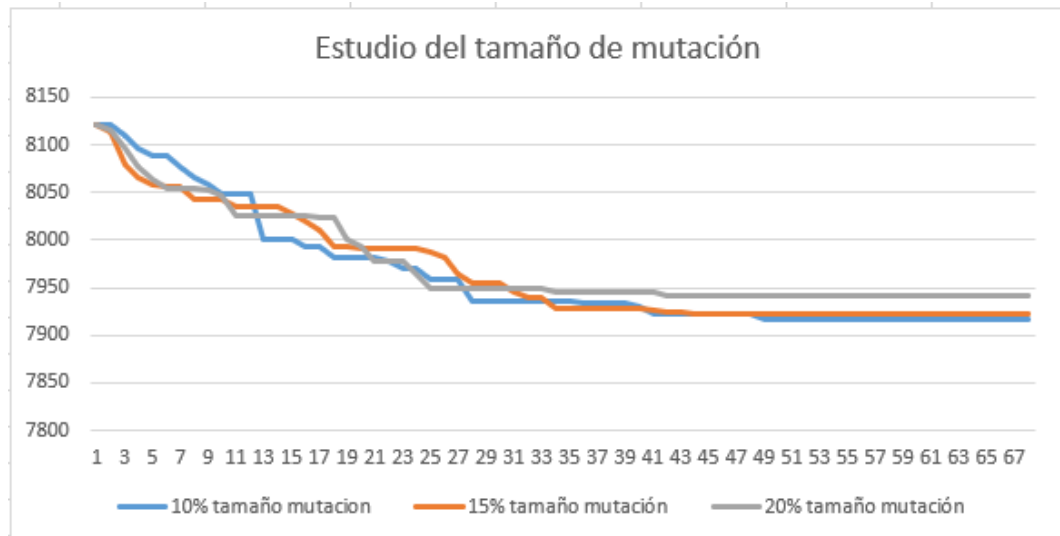
Con un tamaño de mutación mayor, se puede ver que la media también se acerca pero con demasiados picos, ya que se están haciendo cambios más bruscos en los hijos.



Por último vemos el estudio con un tamaño de mutación del 20%, donde claramente se ve que aunque se acerca la media al mejor, hay diferencias significativas entre ellos, y esto se debe a que la cantidad de cromosomas a cambiar es muy grande, y para un tamaño del problema relativamente pequeño, esto es demasiado, y aunque también se genere una mayor diversidad, su consecuencia es que en el cruce entre ellos se pueden generar hijos demasiado malos.

## MB Y HB

Por último, para ver que parámetro a elegir se ha generado una grafica que compara los mejores entre los estudios realizados:



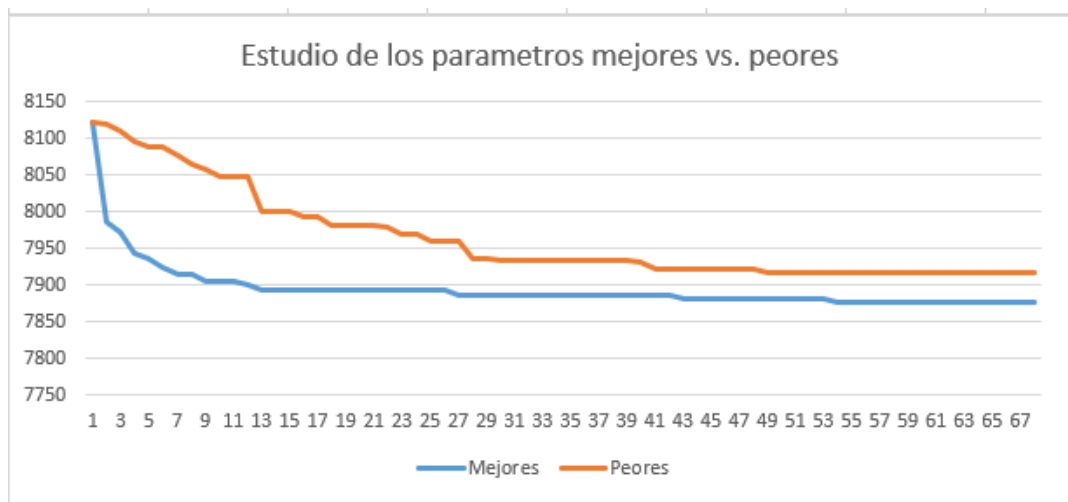
La grafica muestra que con un tamaño de 10% de mutación va muy ligeramente mejor que con un 15%, pero mejor que con un 20% y esto se debe como se decía en el estudio anterior a que al hacer cambios muy bruscos se genera mayor diversidad, pero también hijos peores para un problema pequeño como el ch130, sin embargo, puede ser que para problemas mayores una mayor mutación de mejores resultados.

Concluimos, por tanto, que al haber diferencia entre un 10% y un 20%, siendo el primero mejor, decidimos coger un tamaño de mutación del 10%.

- Estudio sobre todos los mejores aplicados a la vez sobre todos los peores:

Para comprobar que los parámetros que han salido como mejores funcionan todos a la vez, dando buenos resultados, se ha comparado con aquellos que dan los peores:

## MB Y HB



Podemos ver por tanto que los parámetros elegidos como mejores, aunque consigan solo pequeñas mejoras por separado, efectivamente funcionan mejor en todo su conjunto, por lo que trabajaremos a partir de ahora con ellos para generar las tablas de costes.

- Algoritmo genético CHC

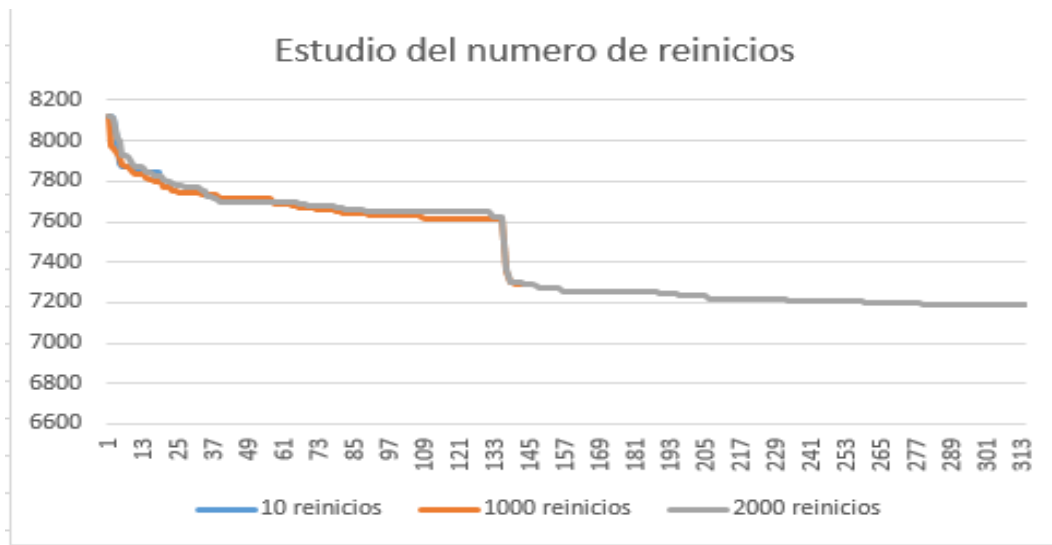
Su objetivo es seleccionar una elite que preserve los mejores individuos, y con un operador de cruce para que los hijos producido sean diferentes a sus padres, y para ello, se ha implementado mediante un cruce basado en Orden, con distancia de Haming para la prevención de 'incesto' y con un umbral de este del  $(\text{tamaño del problema})/4$ .

Cuando no se generan nuevos hijos el umbral desciende hasta que llega a y en ese caso la población se reinicia manteniendo los dos mejores actuales, que será combinado con otros individuos de la población generados aleatoriamente, con esto conseguimos una alta diversidad y una convergencia elevada, debido este último al elitismo.

Además, para establecer el número de reinicios, se ha realizado su correspondiente estudio para establecer cuál puede ser un buen parámetro de parada:

## MB Y HB

- Estudio del número de reinicios para obtención de mejores soluciones CH130:



(Datos mostrados cada 5000 iteraciones).

Podemos ver que a mayor número de reinicios mejor obtención de costes obtenemos. Por lo tanto, para los datos de las tablas cogeremos 2000 reinicios.

- Algoritmo Genético Multimodal

Al algoritmo Genético se le ha introducido el concepto de nicho, para recuperar el espacio abandonado en la exploración del óptimo del problema, es decir, el concepto de nicho nos permite buscar en zonas diferentes del espacio de búsqueda.

Se ha implementado mediante el método de clearing, en la que la selección se realizará únicamente sobre aquellos individuos dominantes de la población sobre su radio de nicho (cuyo radio es de un 10% del tamaño del cromosoma, para que no haya suficiente diversidad).

Para ello, se ordena la población con las mejores soluciones al inicio, se comparan los individuos entre sí, y para aquellos que estén dentro de su radio de nicho (mediante distancia de hamming) se eliminarán. Con ello, se obtienen los individuos dominantes de cada nicho.

## 2. Resultados de los algoritmos

Se mostrarán tablas gráficas del estudio realizada para cada algoritmo:

- Algoritmo Genético Básico

	St70		Ch130		A280	
#Ejecución y Seed	Coste	#Ev	Coste	#Ev	Coste	#Ev
1	809	13910847	8414	16505502	3117	30553977
2	819	13048125	7881	18466188	3261	11326011
3	814	16031049	7330	26181195	2974	22431105
4	738	13025445	7861	25432842	3049	33428785
5	836	13758072	7196	22085460	3003	19753108
Media	803,2	13954707,6	7736,4	21734237,4	3080,8	23498597,2
Desviación Típica	33,8431677	1098836,1	436,52473	3784132,29	102,227981	7895911,57

Del algoritmo a primera vista podemos observar cómo da unas soluciones bastante buenas al recordar ciertas soluciones de otros algoritmos como la búsqueda local o como el Greedy estándar que daba las mejores soluciones hasta el momento, esto podríamos decir que **se debe a que de una población que ya tiene una buena solución a partir del Greedy es capaz de mejorarla.**

Aunque vemos una cantidad enorme de evaluaciones, su principal inconveniente es la cantidad de parámetros que se tienen que ajustar para llegar a obtener mejores soluciones.

- Algoritmo Genético CHC

	St70		Ch130		A280	
#Ejecución y Seed	Coste	#Ev	Coste	#Ev	Coste	#Ev
1	718	4011182	7200	3882870	2911	4466908
2	735	3907476	7226	3912670	2999	4453250
3	744	4019088	7155	3905770	2776	4524678
4	715	3916034	7610	3945884	2909	4477190
5	733	3901480	6977	3865954	2821	4510682
Media	729	3951052	7233,6	3902629,6	2883,2	4486541,6
Desviación Típica	10,8995413	52587,1191	207,309045	27265,9359	77,7287592	26901,3205

Da resultados muy buenos, mejorando al Genético Básico, su diferencia es que utiliza un sistema para mezclar aquellos que se parezcan menos y así obtener una mayor diversidad, logrando alcanzar buenas soluciones. También se debe a que el algoritmo empieza con un individuo de la población calculado por Greedy para lograr una solución

## MB Y HB

mejor y a que el mecanismo de reinicios deja las dos mejores soluciones encontradas hasta el momento como parte de la siguiente población inicial.

- Algoritmo Genético Multimodal

#Ejecución y Seed	St70		Ch130		A280	
	Coste	#Ev	Coste	#Ev	Coste	#Ev
1	809	7027229	8597	11544049	3176	7000139
2	819	7023379	8089	7000139	3283	7000139
3	814	7049699	7646	10844259	2992	7000139
4	828	7001609	7992	13766269	3127	7000139
5	836	7057539	7465	7000139	3022	7000139
Media	821,2	7031891	7957,8	10030971	3120	7000139
Desviacion Típica	9,70360758	19930,5428	391,670984	2656131,73	105,585984	0

Sus resultados son, relativamente buenos, pero no llegan a los anteriores, esto se debe a que con el clearing implementado, se quitan buenos individuos solo por su distancia de hamming, es decir, la distancia de hamming no es una buena forma de comparar para este problema en concreto por que se comparan cromosoma a cromosoma, y no por rangos de ellos, que sería lo más eficaz para el problema del tsp.

Vemos también que para el a280 tienen todas las mismas evaluaciones en todas las ejecuciones y esto es debido a que no es capaz de encontrar mejores soluciones y por lo tanto no reinicia el número de cruce sin mejora.

### 3. Resultados globales y comparacion entre algoritmos

St70 (675)				
Modelo	Media	Mejor	Desviacion Típica	Ev
Greedy	847,1	815	473,4996938	1
Búsqueda local	1036,8	957	58,04101998	112001
AG Básico	803,2	738	33,8431677	13954707,6
AG CHC	<b>729</b>	<b>715</b>	<b>10,89954127</b>	3951052
AG Multimodal	821,2	809	9,703607577	7031891
Ch130 (6110)				
Modelo	Media	Mejor	Desviacion Típica	Ev
Greedy	7752,2	7358	473,4996938	1
Búsqueda local	11132,8	10305	599,2633478	208001
AG Básico	7736,4	7196	436,5247301	21734237,4
AG CHC	<b>7233,6</b>	<b>6977</b>	<b>207,3090447</b>	3902629,6
AG Multimodal	7957,8	7465	391,6709844	10030971
A280 (2579)				
Modelo	Media	Mejor	Desviacion Típica	Ev
Greedy	3128,9	2975	94,44622809	1
Búsqueda local	2784,6	<b>2784</b>	<b>1,8</b>	448001
AG Básico	3080,8	2974	102,2279805	23498597,2
AG CHC	<b>2883,2</b>	<b>2776</b>	<b>77,72875916</b>	4486541,6
AG Multimodal	3120	2992	105,5859839	7000139

Tabla de comparación global

La tabla anterior muestra los resultados globales, con los mejores resultados tras 5 ejecuciones de cada algoritmo y con cada fichero. Siendo los datos en negrita y amarillo los mejores resultados obtenidos. Además, para poder hacer una buena comparación se utilizarán resultados de algoritmos de la practica anterior y que se han utilizado en este practica para mejorar resultados, como son la búsqueda del mejor de todos (Búsqueda local o BL) y el Greedy estándar (utilizado como una solución inicial dentro de la población).

Los resultados muestran un claro ‘Vencedor’ el cuál es el Algoritmo genético CHC, gracias a su diversidad, generando así soluciones mejores.

Por otra parte, podemos ver que para problemas en los que los puntos están muy juntos y de forma lineal, como pasa con el A280, los genéticos no funcionan demasiado bien, aunque, aun así, son capaces de dar soluciones aceptables en un tiempo de computo también aceptable.

Para el algoritmo genético multimodal, se ve que da unos resultados similares a los del algoritmo genético básico, pero al final peores, y esto podemos decir que se debe a que el clearing, basado en distancia de haming, de la forma en la que la

diferencia es cromosoma a cromosoma, es bastante mala para este algoritmo, pudiéndose mejorar con una distancia basada en bloques de cromosomas iguales.

Por último, fijamos la vista en la búsqueda local del fichero A280, en el que da mejor solución, y esto es a que el algoritmo para el problema en el que los puntos están muy juntos y muy linealmente, es capaz de buscar mejor en su espacio de soluciones, que los genéticos usando una heurística más compleja.

#### 4. Conclusiones

Tras el estudio de los algoritmos con los distintos datasets, tras el estudio de los distintos parámetros y tras ver la comparativa global de los algoritmos con Búsqueda del mejor de todos y Greedy de la práctica 2, puede decirse que el CHC dentro de los algoritmos genéticos implementados, da los mejores resultados, aunque, echando la vista hacia la practica 3 de las hormigas, no son capaces de llegar a esas soluciones, y esto es porque los genéticos son buenos algoritmos para solucionar problemas, pero van mejor con cierto tipo de ellos, siendo los algoritmos basados en hormigas los mejores para el tsp, debido a que han sido diseñado para problemas de grafos en general.

Por otra parte, las gráficas mostradas nos dan una vista sobre la capacidad de convergencia de los algoritmos, como decíamos anteriormente, y nos daban la capacidad de decidir los parámetros para encontrar mejores soluciones.

Por último, podemos decir que en general, para los algoritmos genéticos su mayor dificultad radica en la capacidad de diseñar unos parámetros buenos para encontrar mejores soluciones, y en elegir el algoritmo exacto para el tipo de problema a tratar, siendo la elección de estos algoritmos genéticos el ultimo remedio y la 'arma' más grande para resolver el problema.