



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico I

Distanciamiento social

Algoritmos y Estructura de Datos III
Segundo Cuatrimestre de 2020

Grupo 29

Integrante	LU	Correo electrónico
Cerdeira, Elías Nahuel	692/12	eliascerdeira@gmail.com
Panichelli, Manuel	72/18	panicmanu@gmail.com

Reentregar



Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://www.fcen.uba.ar>

Índice

1. Introducción	1
2. Metodología	1
2.1. Fuerza Bruta	1
2.2. Backtracking	2
2.3. Programación Dinámica	3
3. Experimentación	4
3.1. Métodos	4
3.2. Control	5
3.3. Podas	5
3.3.1. Factibilidad - LowM	6
3.3.2. Optimalidad	6
3.4. Grupos - Solapamiento	8
3.5. Caché	8
3.6. Tiempos vs Complejidad teórica	8
3.6.1. BT	9
3.6.2. FB	10
3.6.3. DP	10
4. Conclusiones	11

1. Introducción

En tiempos de crisis sanitaria como los que atraviesa el país y el mundo en este momento, es necesario poder tomar decisiones que permitan resguardar la salud de la población y mantener el nivel de productividad económica que mitigue los efectos sobre los niveles de desocupación y pobreza. El contexto de *distanciamiento social* plantea el problema de *negocio por medio* (NPM) que busca determinar qué locales conviene abrir, dentro de un conjunto de locales de zonas comerciales, en función del *beneficio* económico que aportan, asignado por el gabinete económico, y un valor de *contagio*, asignado por un grupo de expertos de la salud.

Bien la motivación

Formalmente, dado una secuencia de $n \geq 0$ locales comerciales en orden $L = [1, \dots, n]$, el beneficio y contagio $b_i, c_i \in \mathbb{N}_{\geq 0}$ de cada local $i \in L$ y el límite de contagio $M \in \mathbb{N}_{\geq 0}$, el NPM consiste en determinar cuál es el máximo beneficio comercial obtenible sin sobrepasar el límite de contagio establecido por los especialistas. Para simplificar, se considera *solución* a toda secuencia de locales L' que sea subsecuencia de L y se dice que es *factible* si la suma de los valores de contagio de los locales de L' no supera el límite de contagio, es decir, (i) $\sum_{l \in L'} b_l \leq M$, y si no cuenta con dos locales colindantes, es decir, si (ii) $(\forall i \in [2, \dots, n-1] \wedge L_i \in L') : (L_{i-1} \notin L' \wedge L_{i+1} \notin L')$. Además se espera que siempre sea posible al menos poder abrir un local, es decir, $\exists i \in [1, \dots, n] : L_i \leq M$.

no sería C_{i}?

se podría chequear un lado solo, pero ok!

i es un local
o es un índice?

confuso este ejemplo
L' es una secuencia?
L_1' qué sería?

A continuación se exhiben algunos ejemplos con sus correspondientes respuestas esperadas. Si se cuentan con $n = 4$ locales. Sea $M = 40$, los beneficios $b = [10, 20, 30, 40]$ y los contagios $c = [10, 10, 10, 10]$. Las soluciones factibles son $L'_1 = 1, L'_2 = 2, L'_3 = 3, L'_4 = 4, L'_5 = 1, 3, L'_6 = 1, 4, L'_7 = 2, 4$ y la solución óptima es la 7 con beneficio máximo de 60. Por otro lado, si se tiene $M = 20$, los beneficios $b = [10, 15, 30, 15]$ y los contagios $c = [15, 25, 10, 5]$. Las soluciones factibles son $L'_1 = 1, L'_2 = 3, L'_3 = 4, L'_4 = 1, 4$ y la solución óptima es la 2 con beneficio máximo de 30.

El objetivo de este trabajo es implementar una solución para NPM utilizando tres técnicas algorítmicas distintas y evaluar la efectividad de cada una para distintos conjuntos de instancias. En primer lugar se utiliza *fuerza bruta* (FB) que consiste en enumerar todas las soluciones posibles, de manera recursiva, y luego buscar entre las soluciones factibles aquella que sea óptima. Para el algoritmo de *backtracking* se introducen podas para reducir el número de nodos del árbol recursivo. Finalmente, se utiliza almacenamiento en memoria para evitar reprocesar resultados de subproblemas ya calculados. Este proceso se conoce como *memoización* y el algoritmo resultante es el de *programación dinámica* (PD).

En la sección de metodología se introducen y explican los algoritmos de cada una de las técnicas utilizadas en el trabajo junto con las respectivas demostraciones de correctitud y complejidad. Luego, se exponen los experimentos realizados con sus resultados y la respectiva discusión. Por último, se detallan las conclusiones finales del trabajo.

2. Metodología

2.1. Fuerza Bruta

La técnica de **fuerza bruta** consiste en recorrer **todo** el espacio de soluciones en busca de aquellas factibles u óptimas. En este caso, el conjunto de soluciones está compuesto por todas las subsecuencias de L . Por ejemplo, si $L = [1, 2, 3]$ con $b = [20, 15, 30]$, $c = [10, 15, 20]$ y $M = 30$ todas las subsecuencias posibles son $[], [1], [2], [3], [1, 2], [1, 3], [2, 3], [1, 2, 3]$ y las soluciones factibles son $[1], [2], [3], [1, 3]$.

La idea del algoritmo 1 es generar de forma recursiva las soluciones recorriendo todos los locales, pudiendo elegir incluir o no cada local en cada paso. Al contar con todo el espacio de soluciones, se recorre cada una para determinar si es factible y, en caso de serlo, se devuelve su beneficio total. Luego se selecciona alguna de las óptimas dentro del conjunto de soluciones factibles.

En la [fig. 1](#) se puede observar un ejemplo del árbol de recursión generado para la instancia detallada más arriba. Cada nodo intermedio representa una solución parcial, es decir, la solución al problema considerando hasta el local i -ésimo. Las hojas representan todas las soluciones posibles. La solución óptima se colorea con verde, las demás soluciones factibles, con azul y las soluciones no factibles, con rojo. La solución al problema analizado se obtiene llamando a $FB(L, n, M, vecindad, 0)$ donde *vecindad* es un vector de *boolean* inicializado en cada posición en *False*.

La correctitud del algoritmo proviene del hecho de que se recorre *todo* el espacio de soluciones, dado que para cada local de L se generan dos ramas, si se considera para la subsecuencia o no. Y al llegar a las hojas se filtran las soluciones que no sean óptimas o factibles.

OK!

consejo:
agregar \label en la figura
y \ref para referenciarla

como hay 2^n hojas en el árbol binario completo y cada hoja tiene complejidad $O(n)$, podemos decir que el algoritmo tiene complejidad $O(n \cdot 2^n)$

La complejidad en el peor caso es $O(n \times 2^n)$, lo que se desprende de que el árbol utilizado para recorrer el espacio de soluciones es un árbol binario completo de altura $n + 1$. Además, en el caso base se recorre todo el vector de vecindad para generar el beneficio acumulado, lo que se realiza en $O(n)$. En el resto de los casos, solamente se realiza el llamado recursivo y se utiliza la función *max* lo que se considera de complejidad constante $O(1)$. Se puede observar que frente a cualquier instancia el algoritmo se comportará de igual manera, dado que genera el mismo tipo de árbol con la misma cantidad de nodos, es por ello que el conjunto de instancias del peor caso es igual al del mejor caso.

buena conclusión

Algorithm 1 Algoritmo de *Fuerza Bruta* para NPM.

```
1: function FB(locales, n, M, vecindad, i)
2:   if Ya recorri todos los locales then
3:     for j in [0..n] do
4:       Voy acumulando el beneficio y el contagio
5:       Verifico el vector de la vecindad
6:       if Hay dos vecinos consecutivos then
7:         return instancia inválida
8:       if Si el local actual pertenece then
9:         Acumulo su beneficio y contagio
10:      if Pasé el contagio permitido then
11:        return instancia inválida
12:      return el beneficio acumulado
13:   else
14:     return el máximo entre FB poniendo al local i y no poniéndolo.
```

"Voy acumulando" suena un poco raro (cuándo transcurre la acción?) más recomendable usar presente / pasado simple en pseudocódigo

lo mismo, en qué momento del algoritmo exactamente transcurre la acción de acumular?

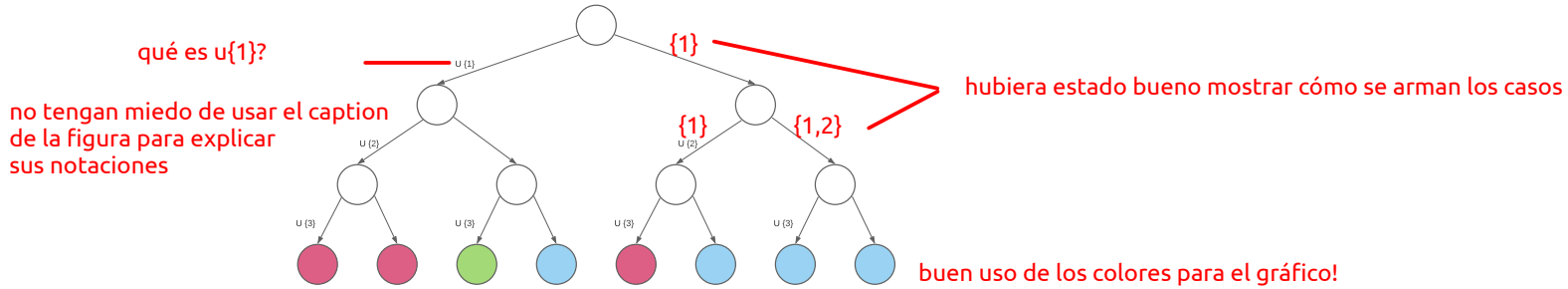


Figura 1: Ejemplo de ejecución del Algoritmo 1 para $L = [1, 2, 3]$ con $b = [20, 15, 30]$, $c = [10, 15, 20]$ y $M = 30$. En rojo las soluciones no factibles, en azul las factibles y en verde la solución óptima $\{3\}$. $\{3\}$ es solución óptima en este caso? por qué no $\{1,3\}$?

2.2. Backtracking

El algoritmo de *backtracking* recorre un árbol para generar las soluciones posibles de manera similar a FB. La diferencia radica en la presencia de *podas*, es decir, reglas que permiten evitar explorar todo el espacio de soluciones omitiendo recorrer ciertas ramas basado en algún criterio. Se dividen en dos tipos: *factibilidad* y *optimalidad*.

- **Factibilidad.** Dada una solución parcial S' representada por un nodo intermedio n_0 que cuenta con un contagio acumulado $m = \sum S'_c$ y el vector de vecindad en estado V' . Como para todo local L_i tenemos valores de contagio positivos, si $m > M$ entonces no tiene sentido continuar extendiendo S' porque en todos los casos el contagio final excederá M . Además, si en V_i tenemos las últimas dos posiciones del vector indicando que están presentes L_{i+1} y L_{i+2} , cualquiera de las soluciones finales tendrá dos vecinos colindantes y no será factible. En cualquiera de los dos casos, se puede evitar continuar recorriendo el subárbol formado a partir del eje n_0 y así reducir el número de operaciones. Se puede observar el pseudocódigo en el algoritmo 2 **Ok!**
- **Optimalidad.** Dada una solución parcial S' representada por un nodo intermedio n_i con beneficio $b = \sum S'_b$. Se cuenta con la solución factible con beneficio máximo B hasta el momento. En cada nodo se calcula b_r , beneficio que se puede obtener al agregar a todos los locales por recorrer. Finalmente, si $b + b_r \leq B$ cualquier decisión que se tome a continuación en el subárbol llevará a una solución al menos tan buena como la ya conocida. Por lo tanto, se

puede podar esa rama y evitar cálculos innecesarios. En el algoritmos 2 se actualiza la variable de *maxBeneficio* cada vez que se encuentra una solución factible en la línea 11 y se evalúa la regla de la poda en la línea 8. **Ok!**

Algorithm 2 Algoritmo de *Backtracking* para NPM.

```
1: function BT(locales, i, M, vecindad, beneficioacumulado, maxBeneficio)
2:   if Se excedió el contagio máximo then
3:     return instancia inválida
4:   if Los dos vecinos anteriores están en la solución parcial then
5:     return instancia inválida
6:   if Ya recorrí todos los locales then
7:     return beneficioacumulado
8:   if Si beneficioacumulado + lo máximo que puedo agregar ≤ maxBeneficio then
9:     return instancia inválida
10:  Calculo el maxLocal entre poner o no el siguiente local, actualizando el M y el beneficioAcumulado
11:  Actualizo maxBeneficio si corresponde
12:  return maxLocal
```

muy coloquial el pseudocódigo. Puede haber alguna que otra frase para líneas específicas, pero lo ideal es que no sean todas frases, porque queda poco claro el comportamiento. Es importante que quede claro qué estructuras usan y qué operaciones hacen sobre ellas

La solución al problema se obtiene llamando *BT(locales, n, M, vecindad, 0, 0)*. Como en el peor caso no se aplica ninguna poda, se recorren todos los nodos del árbol de soluciones, que ya vimos que son $O(2^n)$. En cada uno se hacen las podas de factibilidad que son de orden constante, y la de optimalidad que es $O(n)$ (ya que se recorren en el peor caso todos los locales). El resto de las operaciones se hacen en orden constante. Por lo tanto, la complejidad temporal es $O(n \times 2^n)$. **Bien**

Existen instancias donde el algoritmo va a recorrer el árbol de soluciones completos, por ejemplo todas las que presenten contagios $c = [k, k, \dots, k]$, beneficios $b = [l, l, \dots, l]$ y $M > k * n$ con $k, l \in \mathbb{N}_{\geq 0}$. En este caso la solución óptima es agregar todos los locales de posición impar, pues todos presentan iguales contagio y beneficio y además, el valor de M permite agregar virtualmente todos los locales.

ojo que no es la única solución óptima siempre. En caso de $|b|$ par, también se podrían tomar todos los locales de posición par

Por otro lado, el mejor caso ocurre cuando la solución óptima se encuentra rápido o cuando se corta rápidamente porque se excede la cota de contagios. Las instancias de tipo $b = [1, \dots, 1, B]$ con $B > n$ y/o con $c = [C, \dots, C, C']$ con $C > M$ y $C' < M$. Así se encuentra una solución óptima en la primera rama y luego se realizan las correspondientes podas por optimalidad y/o factibilidad lo que garantiza que ningún otro nodo se ramifique. En estos casos el algoritmo se comportará de forma cuadrática. **Buen análisis!**

2.3. Programación Dinámica

Los algoritmos de *programación dinámica* entran en acción cuando existe superposición entre los subproblemas de un problema recursivo. La idea consiste en almacenar el cálculo del primer llamado a cada subproblema y que, en los subsiguientes llamados, se obtenga el valor *memoizado*. En este caso se define la siguiente función recursiva que resuelve el problema:

$$npm_pd(i, M) = \begin{cases} -\infty & \text{si } M < 0 \\ 0 & \text{si } i = 0 \\ \max\{npm_pd(i-1, M - c_i) + b_i, npm_pd(i-1, M)\} & \text{sino} \end{cases} \tag{1}$$

Importante!! Esto no es equivalente a lo que tienen en el código. El bool vecino es clave para el funcionamiento de su algoritmo. En esta función están aceptando casos donde hay locales contiguos abiertos

Coloquialmente se puede definir a *npm_pd(i, M)* como el máximo beneficio de una subsecuencia de locales $[L_i, \dots, L_n]$ que tenga como límite el contagio M . Se observa fácilmente que *npm_pd(n, M)* resuelve NPM ya que representa el máximo beneficio de una subsecuencia de L con límite de contagio M . A continuación, se observa que la recursión efectivamente representa lo que se acaba de enunciar coloquialmente.

Correctitud

- a. Si $M < 0$ entonces claramente ninguna subsecuencia va a cumplir la cota de contagio ya que todos los valores son enteros positivos. Así, la respuesta es $npm_pd(n, M) = -\infty$, el neutro en cuanto a *max*, lo que hará que no se considere.

Ejemplo:

dados b = [2,3] c = [2,3] M = 6 la solución óptima es abrir el local 2, beneficio total = 3

este algoritmo hace: npm_pd(2,6)= max{npm_id(1,3)+3, npm_pd(1,6)} =

max {max{npm_id(0,1)+2+3, npm_id(0,3)+3}, max{npm_pd(0,4)+2,npm_pd(0,6)} } = max{max{0+2+3, 0+3}, max{0+2, 0} } = max{5, 2} = 5

- b. Si $i = 0$ entonces quiere decir que buscamos la subsecuencia de beneficio máximo dentro de una secuencia de locales vacía. En este caso, como el beneficio de la lista vacía es nulo se tiene que $npm_pd(n, M) = 0$.
- c. En el resto de los casos, se desea buscar una subsecuencia de $L^i = [L_1, \dots, L_i]$ que tenga beneficio máximo con límite de contagio M . De existir una subsecuencia que cumpla las condiciones, puede tener o no al i -ésimo local. Basta entonces con contemplar al local i -ésimo, y luego seguir buscando soluciones de forma recursiva. Si no lo consideramos, seguimos la búsqueda a partir del local siguiente: $npm_pd(i - 1, m)$. Si consideramos al i -ésimo elemento, el resto de la solución tendrá límite de contagio $M - c_i$, y el beneficio total tendrá también b_i , lo que es lo mismo que obtener $npm_pd(i - 1, m - c_i) + b_i$. Por último, la mejor solución será el máximo entre ambas.

Como se aseguran de no tener locales abiertos contiguos?

Memoización

Se puede ver que la función recursiva 1 toma dos parámetros: $i \in [1, \dots, n]$ y $m \in [0, \dots, M]$. Los casos con $i = 0$ y $m < 0$ son casos base y se resuelven de manera *ad-hoc* en tiempo constante. Por lo tanto, la cantidad de posibles combinaciones con la que se puede llamar a la función está determinada por la combinación de ambos. Es así que resulta haber $O(n * M)$ casos posibles, de manera que si se puede implementar una memoria que recuerde y almacene el resultado de los casos ya resueltos se puede calcular una sola vez cada uno y asegurar, de esa manera, resolver el problema sin resolver más de $O(n * M)$ subproblemas. El algoritmo 3 muestra la idea aplicada a la función 1.

Algorithm 3 Algoritmo de Programacion dinamica para NPM.

```
1: function PD(locales, i, M, vecino, memoria)
2:   if Se excedió el contagio máximo then
3:     return instancia inválida
4:   if i = 0 then
5:     return 0
6:   if Aún no calculé el subproblema then
7:     Calculo los beneficios obtenidos al poner o no al siguiente local
8:     Guardo ambos valores en memoria en forma de tupla
9:   if La solución obtenida por poner al local no cumple con la restricción de vecindad then
10:    return el resultado de la otra rama
11:  return el máximo valor guardado en memoria
```

La idea de hacer un pseudocódigo es poder explicar las partes importantes del código sin tener que atender a los detalles no importantes (por ejemplo, cosas específicas del lenguaje). En este caso, la forma en la que PD hace recursión sí es importante. Hubiera estado bueno que muestren cómo se calcula el paso recursivo.

La complejidad del algoritmo queda determinada por la cantidad de posibles estados o subproblemas que se resuelven y el costo de resolver cada uno. Como se mencionó anteriormente, se resuelven a lo sumo $O(n * M)$ subproblemas. Las operaciones de las líneas 2 a 6 y 8 a 11 se realizan en tiempo constante $O(1)$ y en la línea 7 se realiza el llamado recursivo. Así es como se deduce que el algoritmo tiene complejidad $O(n * M)$ en el peor caso. La memoria se puede implementar como una matriz de acceso y escritura constante y su inicialización tiene un costo $O(n * M)$, es por eso que tanto el peor como mejor caso tendrán complejidad $O(n * M)$. Bien

3. Experimentación

En la experimentación se intentará mostrar las fortalezas y debilidades de cada técnica, comparándolas entre sí para diferentes tipos de instancias. Además, cada instancia tendrá una cantidad de locales $n \in [1, \dots, 30]$, lo que permitirá apreciar la evolución del tiempo de ejecución en función de n , y compararlo con la complejidad teórica descrita en la Sección 2. Las ejecuciones fueron realizadas utilizando el lenguaje de programación C++, en una computadora con un Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz (32K cache lvl 1, 256K lvl 2 y 6MB lvl 3) y 16GB de RAM.

3.1. Métodos

Las técnicas a considerar son las siguientes:

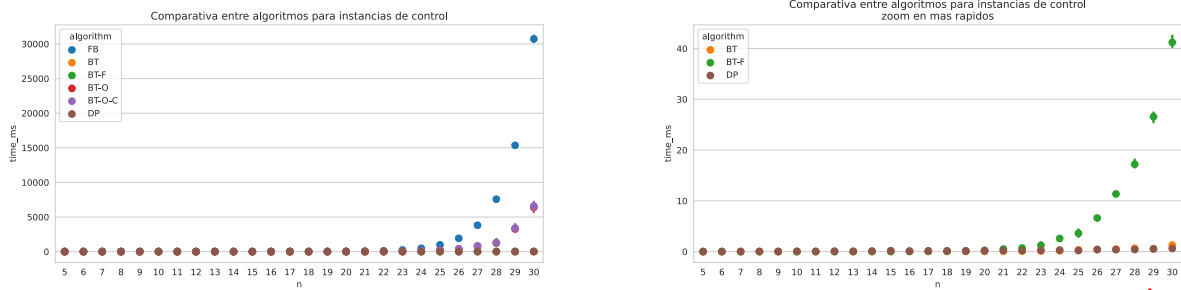
- **FB**: algoritmo 1 de Fuerza Bruta, $O(2^n)$ no era $O(n * 2^n)$
- **BT**: algoritmo 2 de Backtracking con ambas podas (optimalidad cacheada y factibilidad) $O(2^n)$ no es $O(n * 2^n)$ por la poda de optimalidad?

- **BT-O**: algoritmo de Backtracking solo con podas de optimalidad, con factibilidad en las hojas para todos los locales. $O(n \times 2^n)$
- **BT-O-C**: algoritmo de Backtracking con podas de optimalidad cacheadas, se calculan antes de iniciarlo, pero las podas de factibilidad se hacen en las hojas recorriendo todos los locales. luego su complejidad será $O(n + n \times 2^n) \subseteq O(n \times 2^n)$
- **BT-F**: algoritmo de Backtracking solo con podas de factibilidad, por lo tanto su complejidad será $O(2^n)$
- **DP**: algoritmo 3 de Programación Dinámica, $O(n \times M)$

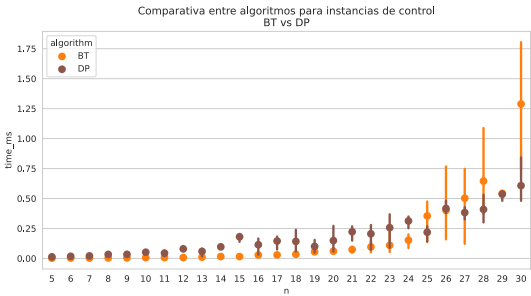
3.2. Control

Para poder comparar las instancias buenas y malas con casos *promedio*, vamos a usar instancias de *control*, generadas tomando aleatoriamente el contagio y beneficio de cada local de rangos fijos. El límite de contagio también se establece aleatoriamente.

aleatoriamente de qué manera?
entre qué valores?



(a) Entre todos los algoritmos para cada n , con instancias de control (b) Zoom en los algoritmos más rápidos para ver las diferencias



(c) BT vs DP

Figura 2: Comparativa para dataset de control

Como primera comparación de los algoritmos, en la Figura 2 se puede ver a grandes rasgos la *performance* de cada técnica. Se observa que BT y DP son los que presentan mejores tiempos de ejecución. Lo que puede resultar llamativo es que, viendo la Figura 2c, a pesar de que BT tenga mayor variación, no parecen mostrar grandes diferencias en tiempos de ejecución. Esto puede deberse a que para el caso general, lo que suele definir si una técnica tiene un buen desempeño es si poda por factibilidad o no, lo cual también explicaría los tiempos de BT-F.

Estaría bueno tener mayor info sobre los casos de control para realizar hipótesis más concisas

3.3. Podas

Se estudiarán ambas podas por separado: factibilidad y optimalidad. Se compararán los resultados con FB a modo de control. Es importante notar que no es posible ver un caso *malo* para la poda de factibilidad, ya que se hace en $O(1)$.

cuáles son los rangos fijos aleatorios? Eso influye mucho en la performance. Por ejemplo, si se toma M aleatorio entre 1 y 10, costos aleatorios entre 20 y 100, su poda de factibilidad va a ser super efectiva ya que el M va a ser fácilmente excedido

Pero no tanto si se toma, por ejemplo, un M aleatorio entre 100.000 y 200.000, costos aleatorios entre 1 y 10, donde el M no se excederá con tanta facilidad

bien logrado el zoom

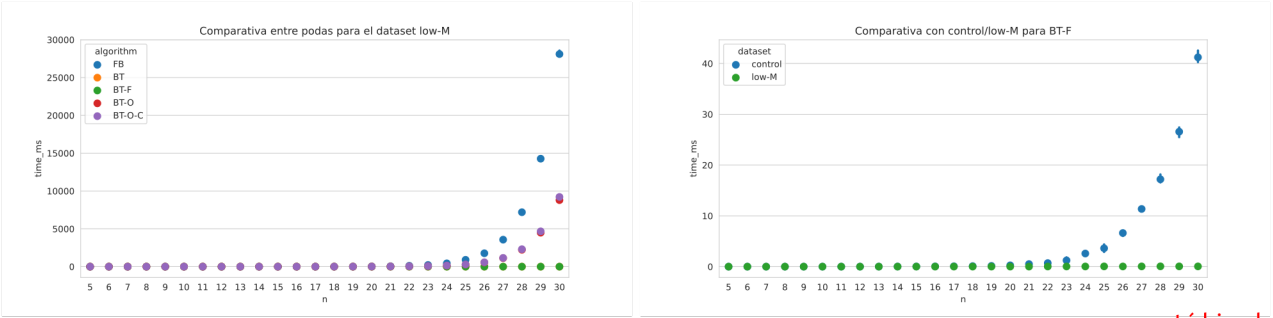
3.3.1. Factibilidad - LowM

Ya que la poda por factibilidad corta la ejecución de una rama en la que se excedió el límite de contagio, una clase de instancias en las que presente mejor desempeño es aquella en la que el límite de contagio es muy bajo y se alcanza con un par de locales (o incluso uno).

Como se puede ver en la Figura 3a, los algoritmos que tienen poda por factibilidad son muchísimo más rápidos que los que no para este tipo de instancias. Además, en la Figura 3b se ve la comparación con el control, lo que muestra que es un caso más rápido que el promedio.

buena idea! Estaría bueno también que expliquen un poquito más cómo se generan las instancias LowM

No se ve a BT en el gráfico :(



(a) Comparativa entre podas

(b) Comparativa con control

Figura 3: Comparativa para dataset de control

está bien la idea y logran encontrar una diferencia, pero hay que mencionar que la magnitud del eje Y cambió en comparación con el gráfico anterior. Son diferencias muy chicas.

3.3.2. Optimalidad

La poda por optimalidad realiza recortes en instancias que, observando los locales restantes, no pueden llegar a obtener un mejor resultado que el máximo resultado obtenido hasta el momento. Se proponen entonces dos clases de instancias distintas:

- **Uniformidad:** en caso de que todos los locales presenten el mismo beneficio, no van a existir ciertas ramas que produzcan mejores resultados que otras. En estas circunstancias, nunca se aprovechará la poda por optimalidad, pero siempre se estará pagando el *overhead* de computar el beneficio restante.
- **Uno para dominarlos a todos:** en estas instancias existe un local distinguido, que tiene beneficio mayor a la suma de todos los demás. De esta forma, cada vez que se llegue a él, posteriormente se deberían realizar las podas de todas las ramas.

Algo a tener en cuenta es que la posición del local distinguido determina la relevancia de la poda. Si se encuentra primero, su beneficio se va a tener en cuenta luego de haber visto todas las hojas, por lo que va a ser lo mismo que el caso uniforme. En cambio, si está del tercero en adelante, se debería almacenar rápidamente su beneficio, haciendo que se poden el resto de los casos. Por lo tanto, correremos solo para $n = 30$ variando todas las posiciones posibles.

Interesante, pero será una situación común dado un grupo de locales aleatorios?

Se espera que la implementación *cacheada* presente mejores tiempos de ejecución respecto de la que lo computa en el momento.

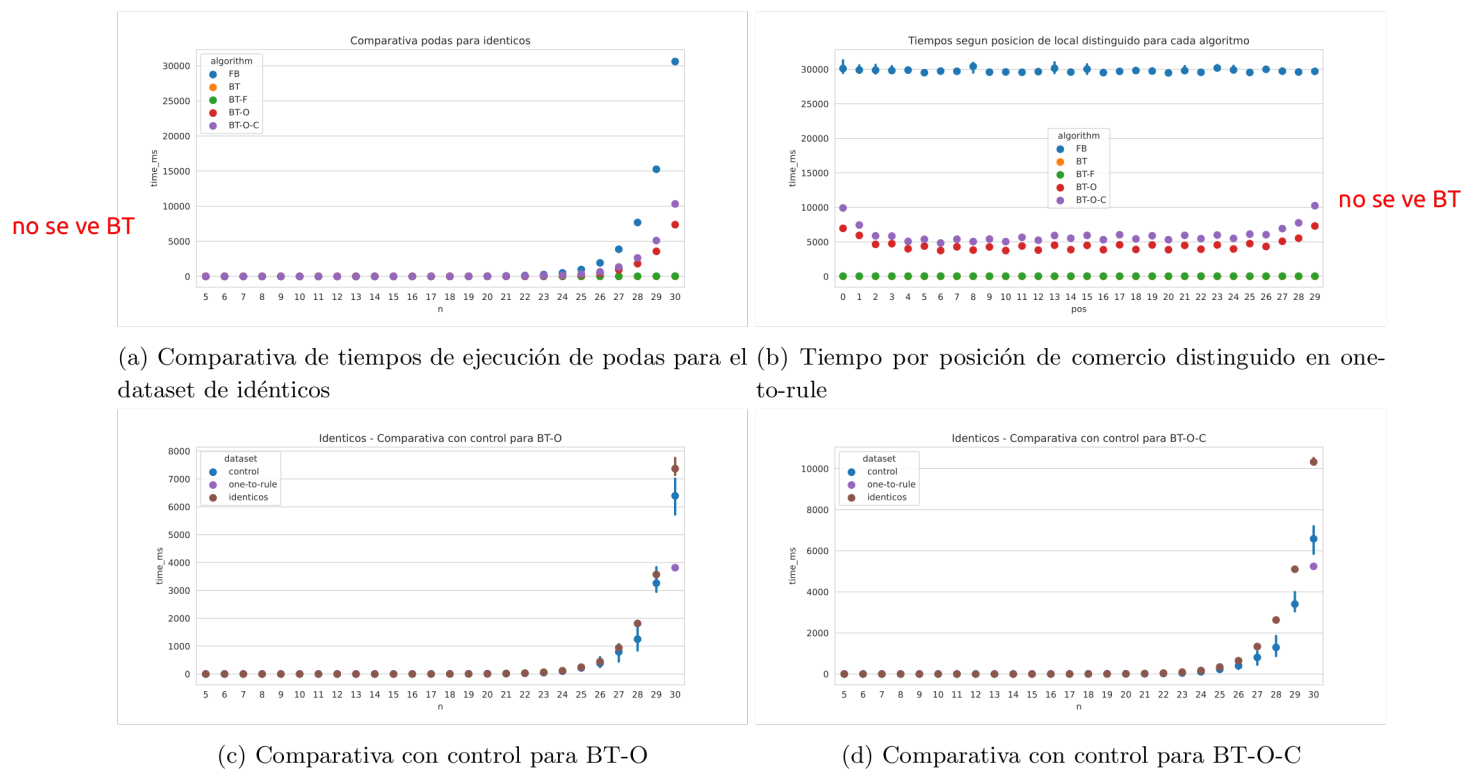


Figura 4: Optimalidad: one-to-rule e identicos

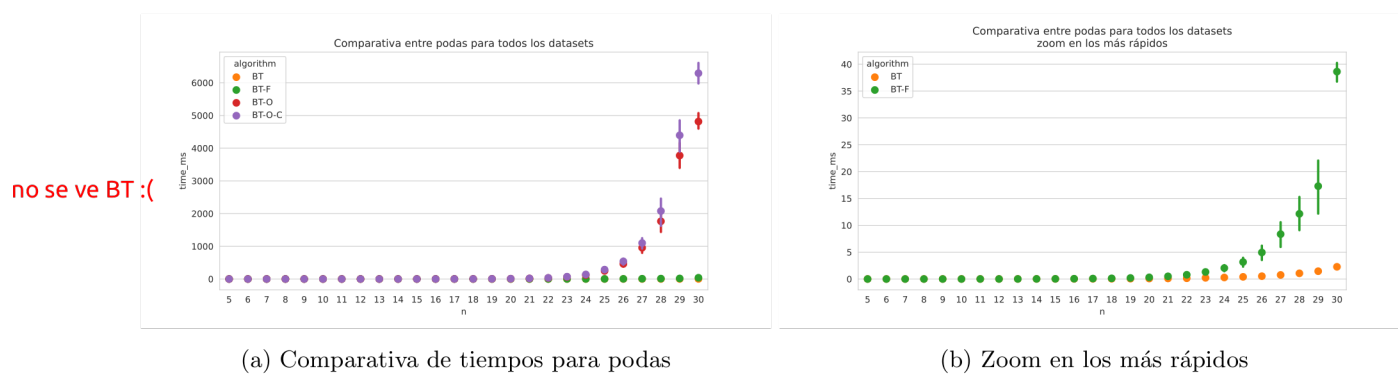


Figura 5: Comparativa de tiempos para todas las podas y todos los datasets.

A pesar de ser un caso favorable, Fig 4a muestra que el hecho de no estar aplicando podas por factibilidad hace que aquellos algoritmos que solo aplican podas por optimalidad sean peores que BT y BT-F, pero al menos siguen siendo mejor que FB. Además, en 4b vemos como según la posición en la que aparezca el local distinguido, cambia el tiempo de ejecución de la misma manera para BT-O y BT-O-C y, como era de esperar, no cambia para los demás algoritmos que se mantienen relativamente constantes. Finalmente, al comparar con el control se observa lo esperado, tiene un tiempo de ejecución peor al del mejor caso (one to rule), pero mejor que el peor (idénticos).

Por último, se observa en Fig 5 una comparación de todos los algoritmos. De aquí surgen dos observaciones interesantes. En primer lugar, contrario a lo esperado, la variante con caché de la poda de optimalidad termina siendo más lenta que la que simplemente lo calcula en cada nodo intermedio. Esto puede deberse a que instancias de tamaño (n) pequeño como las evaluadas en este trabajo, el costo de ir a buscar el dato cacheado a memoria es mayor a $O(30)$, lo cual tendiendo al infinito no debería suceder. A futuro se podría realizar una ampliación de la experimentación que pusiera a prueba esta hipótesis.

Por otro lado, también puede verse que el mejor de todos fue BT como era esperable y, si bien las podas de optimalidad no le brindaron mucha mejora con respecto a BT-F, se nota cada vez más a medida que aumenta el n .

por qué un acceso a memoria es $O(30)$? al hacer `cacheB[i]`, no hacen una pasada lineal, se accede directamente a esa posición de memoria

quizás sería mas interesante hacer crecer n de manera mas rapida, porque desde $n = 1$ hasta $n = 20$ todos los algoritmos parecen dar igual. podrian probar $n = [2,4,6...]$ o tambien $n = [5,10,15,20...]$ por dar algunos ejemplos.

3.4. Grupos - Solapamiento

Al margen del solapamiento que se da por la naturaleza del problema, se pueden construir instancias que hagan que sea aún mayor. Intuitivamente, a mayor solapamiento más valores van a estar memoizados, y por lo tanto podremos podar más ramas, así reducir el tiempo de ejecución.

Para modelar estas instancias, se introdujo la noción de *grupos*. Se dividen los locales de cada instancia en alguna cantidad de grupos, para los cuales todos los locales tienen el mismo nivel de contagio. Así, a menor cantidad de grupos, mas locales con el mismo nivel de contagio hay, lo que se espera que lleve a mayor solapamiento, pues permitiría que por una mayor cantidad de caminos se llegue al mismo nivel de contagio restante, ya que se cuenta con el mismo nivel de contagio acumulado.

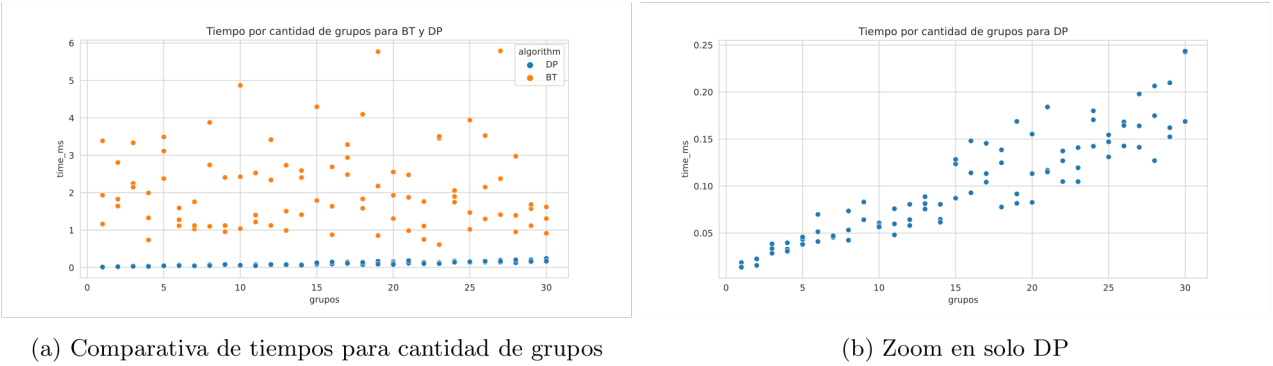
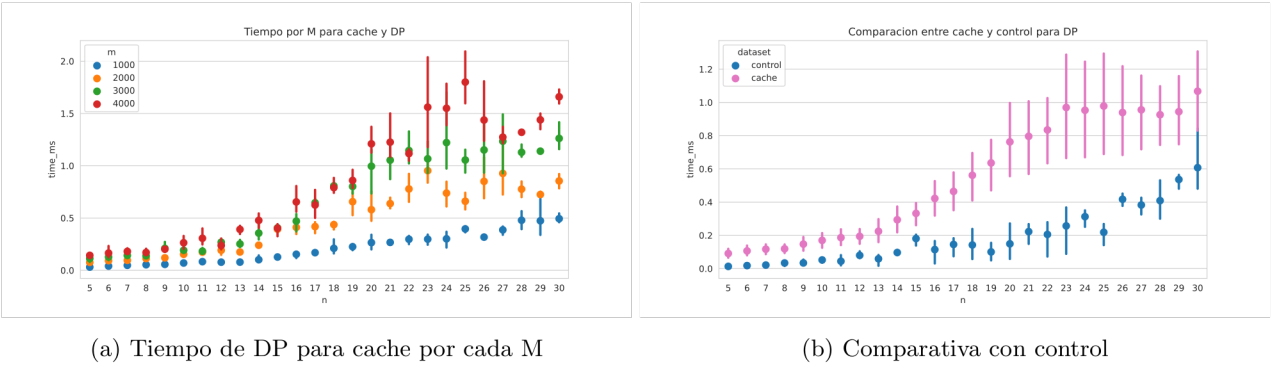


Figura 6: Grupos

Como se puede ver en Fig 6, los grupos no afectan tanto a BT, pero a mayor cantidad de grupos mayor tiempo de ejecución para DP, como era esperado. **Muy buenos graficos y conclusión!**

3.5. Caché

Para ver el lado negativo de la memoización, si se cuenta con una estructura lo suficientemente grande como para que no entre toda en caché, para instancias en las que no hay solapamiento debería notarse el *overhead* de ir a memoria.



Efectivamente, a mayor M, mayor el tamaño de la estructura de memoización y, por lo tanto, mayor el tiempo de ejecución. Comparándolo con el control se aprecia que es peor que el caso promedio. **control sería FB en este caso? me parece mejor que lo aclaren**

3.6. Tiempos vs Complejidad teórica

Para verificar experimentalmente las complejidades propuestas, se corren los algoritmos contra instancias de diferentes tamaños y se comparan sus tiempos de ejecución con la complejidad teórica. Luego se calcula el grado de correlación mediante el coeficiente de Pearson.

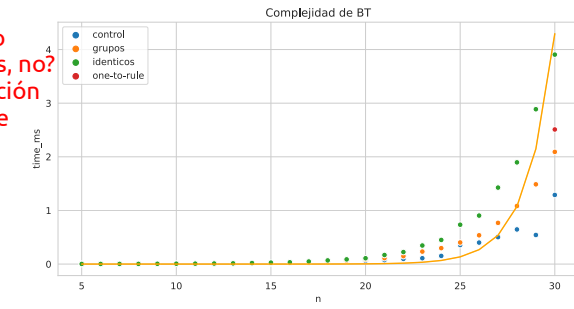
Super interesante idea!

Pero muy confusa la explicacion. Al introducir un concepto nuevo, está bueno una definición clara, ejemplos y frases cortas y concisas.

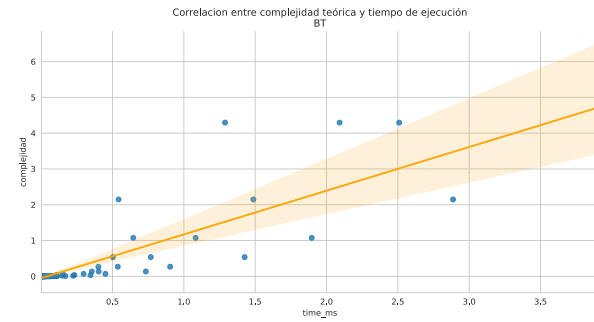
por ejemplo:
"Llamamos 'grupo' a un conjunto de locales con mismo contagio. Por ejemplo, decimos que $c = [1, 1, 3, 1, 1, 3]$ tiene dos grupos. Cuantos menos grupos, mayor será el solapamiento de subproblemas. Dicho solapamiento se verá reflejado en una peor performance en BT, no tanto para DP."

3.6.1. BT

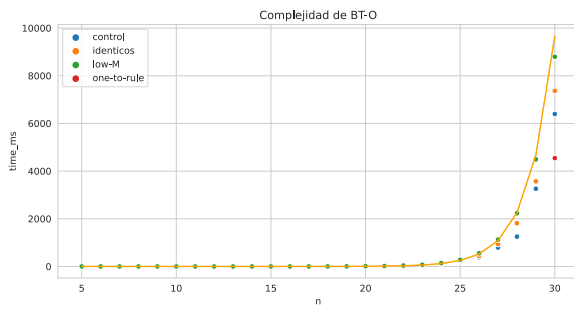
control en este caso es locales aleatorios, no? usan la misma notación para BT y medio que confunde



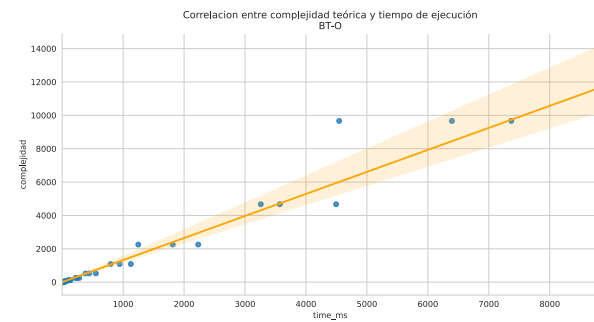
(a) BT - Tiempos vs $O(2^n)$



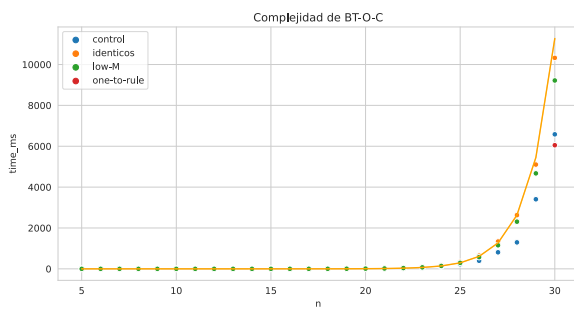
(b) BT - Pearson: 0.828



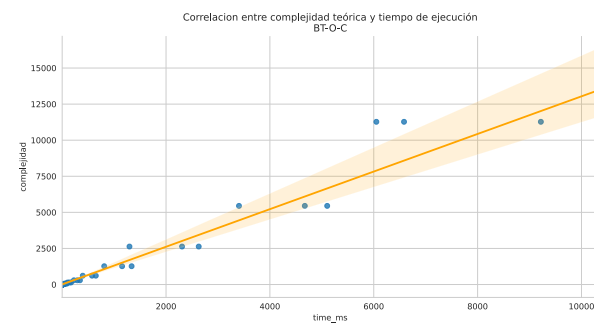
(c) BT-O - Tiempos vs $O(n \times 2^n)$



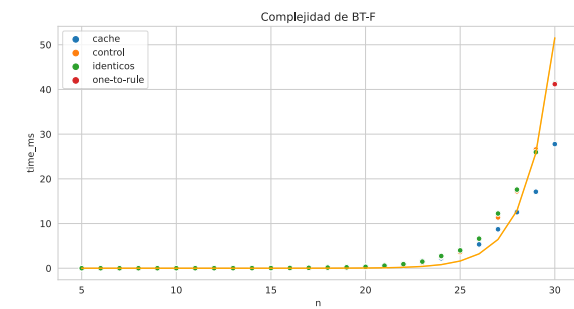
(d) BT-O - Pearson: 0.947



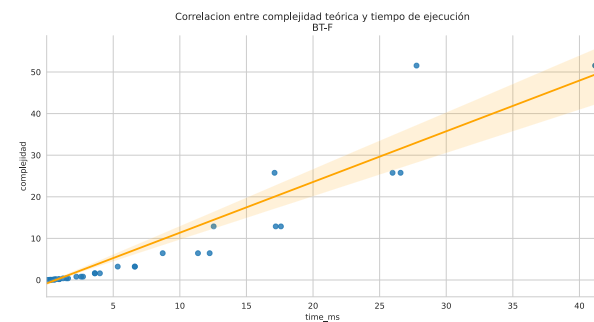
(e) BT-O-C - Tiempos vs $O(n \times 2^n)$



(f) BT-O-C - Pearson: 0.955



(g) BT-F - Tiempos vs $O(2^n)$



(h) BT-F - Pearson: 0.967

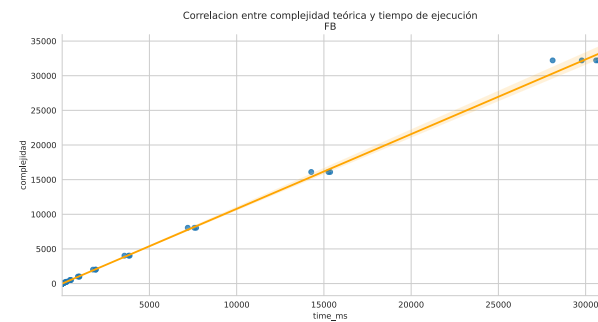
Los tiempos de ejecución de las variantes de BT se asemejan a la curva de la complejidad teórica propuesta.

Ok! está bien, pero qué tal el Pearson? dio bien? dio mal? que es un Pearson alto, qué es uno bajo?

3.6.2. FB



(a) FB - Tiempos vs $O(2^n)$

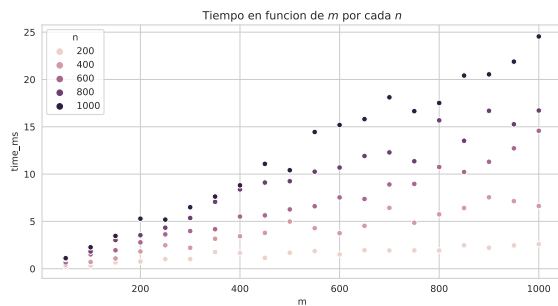


(b) FB - Pearson: 0.999

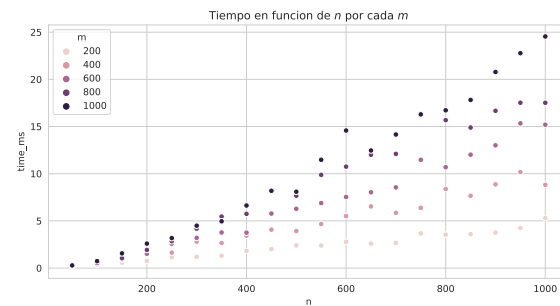
Como se puede ver, al ser el algoritmo que menos varía según el caso, ya que siempre recorre *todo* el espacio de búsqueda, su tiempo de ejecución se asemeja mucho en todos los casos a la complejidad teórica en peor caso. **Buenísimo!**

3.6.3. DP

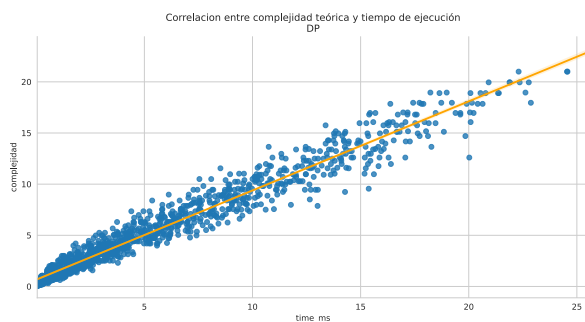
En este caso no es tan simple como para los demás, ya que la complejidad depende de dos parámetros en lugar de sólo uno. Por eso se graficó cada una por separado en función de la otra. En todos los casos se debería tener una relación **Ok!** lineal.



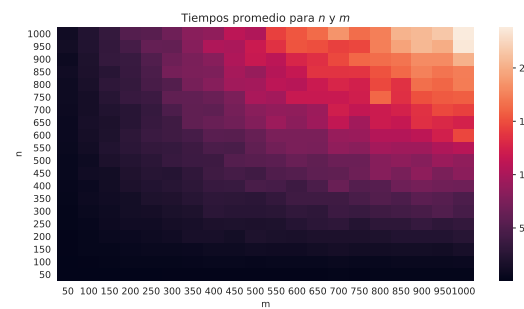
(a) Tiempo en función de m para cada n



(b) Tiempo en función de n para cada m



(c) Comparación con $O(n \times M)$, Pearson: 0.978



(d) Tiempos para cada n y m

Interesante la comparación!

Interesante el heatmap!!

Figura 10: Tiempos para DP

Efectivamente, como se puede ver en 10a y 10b hay una relación lineal entre n y m , la cual esta fuertemente correlacionada (0,978) con la complejidad teórica. Otra forma de visualizarlo es con el *heatmap*.

4. Conclusiones

En este trabajo se presentaron tres técnicas distintas que permiten resolver el problema de NPM. El algoritmo de *Fuerza Bruta* es poco eficiente temporalmente ya que para un número relativamente pequeño de locales ($n = 30$) el tiempo de ejecución se hace excesivamente elevado para todos los casos. Esto se ve mejorado en *Backtracking* mediante el uso de *podas*, donde la máxima performace se alcanza con una combinación de podas por factibilidad y optimalidad, permitiendo reducir el tiempo significativamente, en especial para los tipos de instancias favorables para las podas. Finalmente, el algoritmo de *Programación Dinámica* es el que presenta un menor impacto en su tiempo de ejecución aumentando considerablemente el n , pero lo hace con un gran costo adicional en memoria. Además, pasa a depender del límite de contagio M , lo que hace que para valores elevados no sea la mejor opción.

También se evaluaron distintos tipos de instancias, aleatorias, one-to-rule, según grupos ...

Una línea de trabajo a futuro es ver si para algún tamaño n de la secuencia de locales empieza a presentar mejores resultados precalcular los valores utilizados en la poda por optimalidad. Además, podrían implementarse otras estructuras para realizar la memoización en el caso de programación dinámica, tal podría ser el caso de mitigar el efecto de matrices ralas, que en la implementación utilizada en este trabajo llevarían a un gran desperdicio de memoria y tiempo de ejecución en su inicialización.

Están usando matrices ralas?

Sería interesante también explorar alguna potencial mejora en la poda por optimilidad utilizada, dado que se decidió usar un algoritmo bastante simple en este trabajo. En lugar de contabilizar todos los locales que falta por recorrer, se podría pensar alguna estrategia golosa que permita reducir los tiempos de ejecución que quizás cobren relevancia frente a instancias más complejas.