

Estudio experimental del problema $2 + p$ -COL aleatorio

Roberto Cruz Rodes Andrés D. Galvis Santamaría
Universidad de Antioquia Universidad de Antioquia

Liliam R. Suaza Jiménez
Universidad de Antioquia

Recibido May. 2, 2006 Aceptado Ago. 14, 2007

Abstract

The random $2 + p$ -COL problem smoothly interpolates between P and NP by mixing together the polynomial 2-COL problem and the NP-Complete 3-COL problem. In this article we report an extensive experimental study of the phase transition behaviour and computational cost of this problem. A new parameter for the experimental analysis of the phase transition was introduced. The obtained results for the end of the phase transition agree with theoretical values obtained by other authors. The analysis of the computational cost, which was performed at the end of the phase transition, showed a polynomial behaviour up to $p \approx 0.7$, and an exponential one for $p \geq 0.8$.

Keywords: random $2 + p$ -COL, phase transition, computational cost, NP-Complete problems

MSC(2000): Primary: 68Q17

Resumen

El problema $2 + p$ -COL aleatorio interpola suavemente desde el problema polinomial 2-COL aleatorio al problema NP-Completo 3-COL aleatorio. En este artículo se presenta un extenso estudio experimental del comportamiento de la transición de fase y el costo computacional de este problema. Se introdujo un nuevo parámetro para el estudio experimental de la transición de fase. Los resultados para el fin de la transición de fase se ajustan a los valores obtenidos teóricamente por otros autores. El estudio del costo computacional, realizado en la frontera de la fase, mostró un escalamiento polinomial para $p \leq 0,7$ y exponencial para $p \geq 0,8$.

Palabras y frases claves: $2 + p$ -COL aleatorio, transición de fase, costo computacional, problemas NP-Completo

1 Introducción

Los problemas NP-completos conforman la clase de problemas computacionales cuyas soluciones no se pueden obtener utilizando algoritmos polinomiales. Investigaciones relacionadas con la complejidad algorítmica buscan explicar por qué los problemas NP-completos son difíciles de resolver. En estas investigaciones se ha encontrado que las soluciones de estos problemas exhiben transiciones de fase análogas a las que aparecen en los sistemas físicos. Dichas transiciones ocurren al variar algunos parámetros de los problemas y traen como consecuencia un cambio drástico tanto en el costo computacional de la solución como en la naturaleza de la misma. Los conocimientos más interesantes, sobre el comportamiento de la transición de fase, se han obtenido en las investigaciones del problema $2 + p$ -SAT

aleatorio [8, 9]. Este problema es una interpolación entre el problema polinomial 2-SAT ($p = 0$) y el problema NP-completo 3-SAT ($p = 1$).

En [11] Walsh propone se investiguen otros modelos que interpolan entre un problema polinomial y uno NP-Completo. Uno de los problemas propuestos se deriva del k -COL aleatorio y es el $2 + p$ -COL aleatorio ($0 \leq p \leq 1$). El problema k -COL aleatorio consiste en un grafo $G(n, \alpha n)$ con n vértices, $m = \alpha n$ aristas generadas uniforme e independientemente y una colección de k colores. Se requiere asignar a cada vértice un color, de modo tal que todo par de vértices que forman una arista del grafo tengan colores diferentes. El problema 2-COL es polinomial, mientras que el k -COL es NP-completo para todo $k \geq 3$.

El problema 2-COL aleatorio no exhibe una transición de fase con umbral; la probabilidad de que el grafo sea no coloreable, cuando n tiende al infinito, es positiva para todo $\alpha > 0$ y aumenta gradualmente con α hasta alcanzar el valor 1 cuando $\alpha \approx \frac{1}{2}$. El problema 3-COL aleatorio sí presenta una transición de fase con umbral. En [2], Achlioptas y Friedgut mostraron la existencia de una sucesión $\alpha_3(n)$, tal que para todo $\epsilon > 0$, la probabilidad de que el grafo $G(n, (\alpha_3(n) - \epsilon)n)$ sea 3-coloreable, tiende a 1 cuando n tiende a infinito, mientras que la probabilidad de que el grafo $G(n, (\alpha_3(n) + \epsilon)n)$ sea 3-coloreable tiende a 0 cuando n tiende al infinito. Se conjetura que la sucesión $\alpha_3(n)$ converge a cierta constante α_3 alrededor de la cual la probabilidad de que el grafo sea 3 coloreable cambia abruptamente de 1 a 0. La mejor cota inferior conocida para α_3 es 1,923 y se debe a Achlioptas y Moore [1]. La mejor cota superior es 2,495, y fue obtenida independientemente tanto por Kaporis y otros [7] como por Fontoulakis y McDiarmid [5].

Semejante al problema $2 + p$ -SAT aleatorio, el problema $2 + p$ -COL aleatorio interpola suavemente desde el problema polinomial 2-COL aleatorio al problema NP-Completo 3-COL aleatorio. $G_{2+p}(n, \alpha n)$ representa un grafo aleatorio con n vértices y $m = \alpha n$ aristas generadas uniforme e independientemente, donde pn vértices disponen de tres colores para ser coloreados, mientras que los restantes $(p - 1)n$ vértices disponen de dos de los tres colores. Inicialmente se fijan los dos colores que pueden tener estos vértices.

En [12], Walsh analiza experimentalmente la transición de fase del problema $2 + p$ -COL aleatorio. Este problema fue transformado en uno SAT con cláusulas de 2 y 3 literales; en su análisis utilizó el algoritmo *zchaff*. El problema SAT resultante no es representativo del $2 + p$ -SAT aleatorio, las variables no están distribuidas uniformemente en las cláusulas de 2 y 3 literales. Walsh usó valores de $p = 0,0, 0,2, 0,4, 0,6, 0,8$ y $1,0$ y valores de $n = 100, 200$ y 300 , observó que la transición de soluble a insoluble comienza a agudizarse en $p = 0,8$; ésta observación la concluyó de los resultados experimentales que obtuvo. Respecto al costo computacional, observó un cambio de polinomial a exponencial en $p = 0,8$; los cálculos fueron realizados con valores de p entre $0,0$ y $1,0$ con pasos de $0,1$ y valores de $n = 50, 100, 150, 200, 250, 300$ en la frontera de la fase, es decir cerca del menor valor de α donde el 100 % de las instancias son no coloreables.

Walsh sugiere utilizar otros algoritmos y explorar soluciones con otros valores de p y de n a fin de precisar lo que ocurre alrededor de $p = 0,8$ y dilucidar si este comportamiento es algoritmo dependiente o no.

Para analizar la transición de fase en el problema $2 + p$ -COL aleatorio, Walsh introdujo los parámetros:

$$\begin{aligned}\delta_{2+p} &= \sup\{\alpha \mid \Pr[G_{2+p}(n, \alpha n) \text{ es } 2+p \text{ coloreable}] \rightarrow 1 \text{ cuando } n \rightarrow \infty\} \\ \gamma_{2+p} &= \inf\{\alpha \mid \Pr[G_{2+p}(n, \alpha n) \text{ no es } 2+p \text{ coloreable}] \rightarrow 1 \text{ cuando } n \rightarrow \infty\}\end{aligned}$$

El problema 2-COL aleatorio no presenta una transición de fase con umbral, en este caso $\delta_2 = 0$ y $\gamma_2 \approx \frac{1}{2}$. Esta es una diferencia importante con el $2 + p$ -SAT aleatorio en donde el 2-SAT si presenta umbral para la transición de soluble a insoluble. La existencia de un subproblema 2-COL aleatorio embebido en el problema $2 + p$ -COL aleatorio para todo $p < 1$ implica que la probabilidad de que un grafo sea no coloreable es positiva para cualquier $\alpha > 0$, luego $\delta_{2+p} = 0$ para $p < 1$. Por otra parte, como el problema 3-COL aleatorio presenta una transición de fase con umbral, utilizando las cotas inferior y superior establecidas, se obtiene que $1,923 < \delta_3 \leq \gamma_3 < 2,495$. Los valores de δ_{2+p} y γ_{2+p} marcan respectivamente el inicio y el final de la transición de fase de problema $2 + p$ -COL aleatorio. El inicio de la fase se mantiene fijo en $\delta_{2+p} = 0$ para todo $p < 1$ y se produce un salto a $\delta_3 > 1,923$ para $p = 1$. Teniendo en cuenta que este problema contiene al 2-COL con $(1 - p)n$ vértices, Walsh obtuvo la siguiente cota superior para el final de la transición de fase:

$$\gamma_2 \leq \gamma_{2+p} \leq \min\left(\frac{\gamma_2}{1 - p}, \gamma_3\right) \quad (1)$$

De esta cota se concluye que el problema $2 + p$ -COL aleatorio no se comporta como el 2-COL a partir de $p \geq 1 - \frac{1}{2\gamma_3}$. Utilizando la mejor cota superior conocida hasta ahora para γ_3 se observa que este comportamiento es para valores de $p \geq 0,8$.

Los resultados experimentales de Walsh concuerdan con la cota superior obtenida para los valores de $p \leq 0,8$, utilizando el valor de 2.52 para γ_3 , la mejor cota superior conocida en el momento que Walsh escribió su trabajo.

Una nueva cota superior para el valor de γ_{2+p} fue obtenida por Cruz [4]. Esta nueva cota mejora la obtenida por Walsh para valores de $p > 0,73$. Es decir, el problema $2 + p$ -COL aleatorio, no se comporta como el problema 2-COL aleatorio en $0,73 \leq p \leq 1$.

En este trabajo se hace un análisis experimental de la transición de fase y del costo computacional en la solución de instancias del problema $2 + p$ -COL aleatorio para diferentes valores de p y n . El análisis experimental se realiza utilizando una adaptación del algoritmo *smallk* [3].

En la sección 2 se describe brevemente el algoritmo *smallk* y la modificación implementada. Los resultados obtenidos con este algoritmo respecto a transición de fase y costo computacional se presentan en la sección 3. Finalmente, en la sección 4 se dan las conclusiones.

2 Modificación del algoritmo *smallk*

Smallk [3] es un algoritmo completo y recursivo para colorear grafos con k colores, donde k es un entero pequeño. Utiliza heurísticas de propagación y métodos de selección dinámica de variables diseñados para evitar asignaciones de color que alteren iteraciones futuras en el proceso recursivo.

La estructura del algoritmo *smallk* es la siguiente.

```

algoritmo smallk( $G$ )
   $v$  = seleccionarVertice( $V(G)$ )
   $c$  = seleccionarColor( $G, v$ )
   $G'$  = colorearVertice( $V(G)$ )
  si (contradiccion( $G$ ))
    restaurarGrafo( $G'$ )
  fin si
  si ( $|V(G')| \neq 0$ )
    smallk( $G'$ )
  sino
    fin smallk
  fin si
fin smallk

```

Para el procesamiento de instancias del problema $2 + p$ -COL aleatorio, fue necesario adaptar el algoritmo *smallk* de tal manera que permitiera discriminar los pn vértices de tres colores de los $(1 - p)n$ vértices de dos colores.

La adaptación del algoritmo fue validada mediante una prueba piloto, en la cual se generaron conjuntos de 100 instancias del problema $2 + p$ -COL aleatorio para diferentes valores de p y n , con $p = 0,2, 0,6, 0,8, 1,0$ y $n = 100, 200$. Cada instancia generada y solucionada con el algoritmo *smallk* fue traducida en una instancia SAT con cláusulas de 2 y 3 literales, las instancias traducidas fueron solucionadas con el algoritmo *tableau modificado* [10]. Los resultados obtenidos con ambos algoritmos fueron equivalentes.

3 Resultados

El análisis experimental del problema $2 + p$ -COL aleatorio se realizó generando conjuntos de instancias con un número de vértices mayor al reportado en otros estudios [11, 12]. El total de instancias generadas osciló entre 1500 y 2000 para valores de $p = 0,2, 0,4, 0,5, 0,6, 0,7, 0,8, 1,0$ y $n = 100, 200, 300, 400, 500, 1000$.

El proceso experimental fue el mismo para el análisis de la transición de fase y el análisis de costo computacional. Para cada combinación de p y n , se definió el número inicial de m aristas del grafo de forma tal que el parámetro $\alpha = \frac{m}{n}$ tomara un valor inicial igual a 0,1. Se generó el grafo con n vértices y m aristas distribuidas aleatoriamente. De estos n vértices $(1 - p)n$ tienen dos colores fijos y los restantes tienen los tres colores disponibles. La colorabilidad del grafo se

determina utilizando el algoritmo *smallk* adaptado; si el grafo es coloreable se le añaden m aristas de forma aleatoria. El proceso se repite hasta que la instancia se haga no coloreable; el parámetro α se incrementa cada vez en 0,1.

3.1 Transición de fase

En la Figura 1. se observa como ocurre la transición de fase para diferentes valores de p con el máximo n utilizado. Nuestros resultados coinciden en los valores de p analizados por Walsh [12]; es decir, para valores de $p \leq 0,6$, la transición de coloreable a no coloreable es suave; mientras que para $p \geq 0,8$ la probabilidad de ser coloreable cae abruptamente a partir de cierto valor del parámetro α . En nuestro experimento, la transición de fase comienza a agudizarse para valores próximos a $p = 0,7$, lo cual es una diferencia notable con los resultados obtenidos por Walsh. Un comportamiento similar se observó para los diferentes valores de n analizados. Con el propósito de obtener una medida que permita establecer una

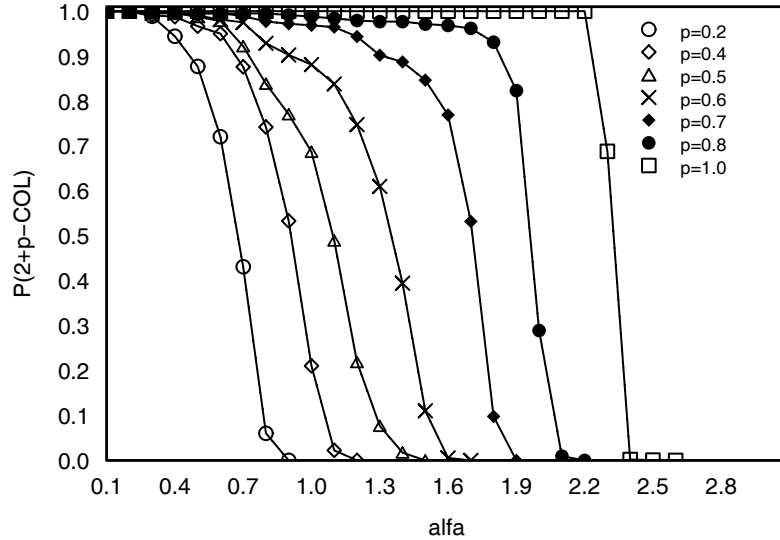


Figura 1: Transición de fase en el máximo valor de n analizado

diferencia significativa en el comportamiento de la transición para los diferentes valores de p , se introduce el parámetro:

$$\Delta(p) = \frac{\alpha_{10}(p) - \alpha_{90}(p)}{\alpha_{50}(p)} \quad (2)$$

donde $\alpha_x(p)$ representa el valor de α para el cual el $x\%$ de las instancias son coloreables. Este parámetro está inspirado en la clasificación de transición aguda o suave para propiedades monótonas en grafos aleatorios de n vértices donde cada arista se genera con probabilidad igual a ρ [6]. Sea $\mu(\rho)$ la probabilidad de

p	n	α_{50}	α_{90}	α_{10}	Δ
0.2	1000	0.676	0.466	0.789	0.4787
0.4	1000	0.910	0.668	1.059	0.4298
0.5	1000	1.095	0.726	1.284	0.5092
0.6	1000	1.380	0.912	1.510	0.4336
0.7	1000	1.707	1.313	1.800	0.2848
0.8	500	1.961	1.829	2.068	0.1219
1.0	500	2.327	2.232	2.386	0.0660

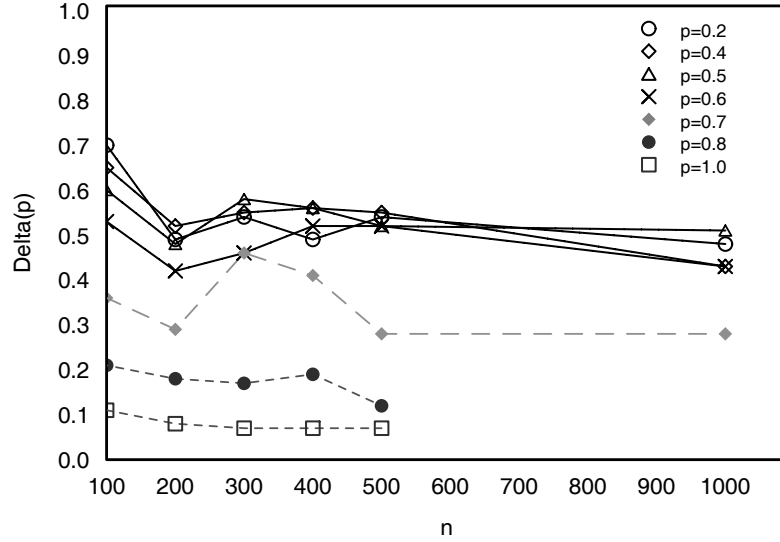
Cuadro 1: Valores del parámetro Δ para el mayor valor de n analizado en cada p

que el grafo aleatorio posea cierta propiedad A , por ejemplo la propiedad de ser $2 + p$ -coloreable. Dado $\epsilon < 0,5$, sea $\mu(\rho_0) = \epsilon$, $\mu(\rho_1) = 1 - \epsilon$, $\mu(\rho^*) = 0,5$ y

$$\delta(p) = \frac{\rho_0 - \rho_1}{\rho^*} \quad (3)$$

Si este cociente tiende a cero cuando n tiende al infinito la transición es *aguda*. Por otra parte, si el cociente se mantiene mayor que 0 cuando n tiende al infinito la transición se clasifica como *suave*. Un paralelo entre los parámetros Δ y δ , identifica a ϵ con la probabilidad de colorabilidad igual a 0,1 y a ρ_0 con el valor de α_{10} , como la razón del número de aristas sobre el número de vértices para el que se alcanza un 10% de instancias coloreables. La transición es aguda si el parámetro Δ se acerca a cero para valores crecientes de n , y es suave si el parámetro tiende a una constante.

La Tabla 1. presenta los valores del parámetro Δ para el mayor valor de n utilizado en cada valor de p . Se evidencian dos grupos, el primero para $p \leq 0,6$, donde el valor de Δ se mantiene por encima de 0,4, y el segundo, para $p \geq 0,8$ donde el valor del parámetro se acerca a cero. Este comportamiento del parámetro Δ se evidencia también para los restantes valores de n analizados como se puede ver en la Figura 2., en donde se muestra la variación del parámetro Δ como función de n para cada valor de p . Para $p = 0,7$ los resultados obtenidos no son concluyentes respecto a la inclusión de la transición en uno u otro grupo, pero se observa que alrededor de este valor se produce el cambio de la transición. En la Figura 3. se muestran los valores estimados experimentalmente para γ_{2+p} , es decir para el final de la transición de fase, y se comparan con la cota superior obtenida por Walsh [12] y la obtenida por Cruz [4]. En esta figura se utilizó el menor valor de α para el cual todas las instancias se hacen no coloreables. Como puede observarse, los resultados obtenidos concuerdan con los valores obtenidos teóricamente para $p = 0,8$ y $p = 1,0$, donde se obtuvieron valores de α por debajo de la cota sobreestimada de Walsh y cercanos a la cota superior obtenida por Cruz.

Figura 2: Comportamiento del parámetro Δ como función de n

3.2 Costo computacional

El estudio del costo computacional se realizó en el final de la transición de fase; es decir, para aquellos valores de α donde todas las instancias se hacen no coloreables. Este costo se expresa como la mediana del número de ramificaciones que necesita el algoritmo *smallk* para establecer la colorabilidad de cada instancia. El comportamiento del costo computacional obtenido en la frontera de la fase se muestra en la Figura 4. En este caso las gráficas obtenidas son consecuentes con el hecho de que el costo crece a medida que crece p . Se observa claramente que hay un cambio en el comportamiento del costo de polinomial a exponencial a partir de $p = 0,8$ lo cual concuerda con los resultados obtenidos por Walsh [12].

4 Conclusiones

En este experimento computacional se estudia la transición de fase y el costo computacional del problema $2 + p$ -COL aleatorio. En el proceso experimental se han utilizado instancias con un mayor número de variables a las utilizadas por Walsh. Se adaptó el algoritmo de coloreo *smallk*, para que permitiera resolver instancias del problema $2 + p$ -COL aleatorio.

Los valores obtenidos experimentalmente en la frontera de la fase confirman lo obtenido teóricamente por Walsh [12] y por Cruz [4] respecto a la cota superior.

Respecto a la transición de fase, nuestros resultados experimentales confirman lo obtenido por Walsh. Además se introdujo el parámetro $\Delta(p)$, el cual permitió observar una diferencia significativa del comportamiento de la transición. Para $p < 0,7$ se observó una transición de fase suave, mientras que para $p \geq 0,8$ se observó una transición de fase aguda. El resultado obtenido por el parámetro

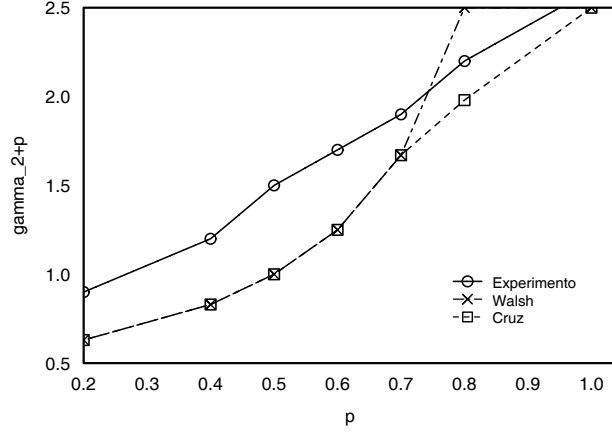


Figura 3: Valores experimentales y teóricos para el final de la transición de fase del problema $2 + p$ -COL aleatorio

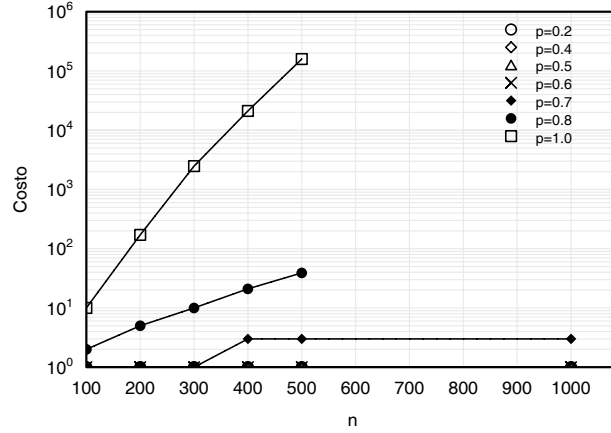


Figura 4: Costo computacional en la frontera de la fase en escala logarítmica

$\Delta(p)$ en $p = 0,7$ no permite concluir como ocurre allí la transición de fase (suave o aguda); sin embargo, nuestros resultados indican que el cambio de la transición ocurre alrededor de este valor.

El estudio del costo computacional se realizó en la frontera de la fase. Se observó un escalamiento polinomial del costo para valores de p hasta $0,7$ y exponencial para valores de $p \geq 0,8$. Estos resultados, semejantes a los obtenidos por Walsh en [12], se obtuvieron con un algoritmo cualitativamente diferente al utilizado por él, lo cual sugiere que el comportamiento del problema no es algoritmo dependiente. Esta es una inquietud manifestada por Walsh en estudios anteriores.

Finalmente, se sugiere que para obtener una descripción más precisa del cambio del costo y de la transición de fase, es necesario explorar detalladamente lo que ocurre para los valores de p entre $0,7$ y $0,8$ utilizando un mayor número de

variables.

Referencias

- [1] Achlioptas, D., Moore, C.: Almost all graphs with average degree 4 are 3-colorable. *Journal of Computer System Science*, 6, pp. 441-471, 2003
- [2] Achlioptas, D., Friedgut, E.: A sharp threshold for k -colorability. *Random Structures and Algorithms*, 14, pp. 63-70, 1999
- [3] Culberson, J.: Graph coloring resources. <http://www.cs.ualberta.ca/~joe/Coloring/>
- [4] Cruz, R.: Upper bound on the non-colorability threshold of the $2 + p$ -COL problem. *Electronic Notes in Discrete Mathematics*, 19, pp. 17-23, 2005
- [5] Fontoulakis, N., McDiarmid, C.: Upper bounds on the non-3-colourability threshold of random graphs. *Discrete Mathematics and Theoretical Computer Science*, 5, pp. 205-226, 2002
- [6] Friedgut, E.: Sharp threshold of graph properties, and k -sat problem. *Journal of the AMS*, 12, pp. 1017-1054, 1999
- [7] Kaporis, A., Kirousis, L., Stamatiou, Y.: A note on the non-colorability threshold of a random graphs. *Electronic Journal of Combinatorics*, 7, R29, 2000
- [8] Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B., Troyansky L.: $2 + p$ -SAT: Relation of Typical-Case Complexity to the Nature of the Phase Transition. *Random Structures and Algorithms*, 15, pp. 414-440, 1999
- [9] Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B., Troyansky, L.: Determining computational complexity from characteristic phase transitions. *Nature*, 400, pp. 133-137, 1999
- [10] Suaza, L., Arbeláez, C., Cruz, R.: Análisis experimental del costo computacional del problema $2 + p$ -SAT aleatorio, *Revista Investigación Operacional*, 25, No 3, 2003
- [11] Walsh, T.: From P to NP: COL, XOR, NAE, 1-in- k , and Horn SAT. *Proc. Of AAAI-2002*, 2002
- [12] Walsh, T.: $2 + p$ -COL. *Proceedings of the Computational Symposium on Graph Couloing and its Generalizations*, Cornell, 2002

Dirección de los autores

Roberto Cruz Rodes — Universidad de Antioquia, Medellín

e-mail: `rcruz@matematicas.udea.edu.co`

Andrés D. Galvis Santamaría — Universidad de Antioquia, Medellín

e-mail: `andresda@cs.udea.edu.co`

Liliam R. Suaza Jiménez — Universidad de Antioquia, Medellín

e-mail: `lsuaza@udea.edu.co`