

COM 2014  
1  
2



Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Planilla a completar para presentación de Cursos de Posgrado

1.- DEPARTAMENTO de COMPUTACIÓN

2.- NOMBRE DEL CURSO: Algoritmos sublineales para grandes conjuntos de datos

3.- DOCENTES:

RESPONSABLE/S: Grigory Yaroslav

COLABORADORES: .....

AUXILIARES: .....

4.- CARRERA de DOCTORADO

5.- AÑO: 2014

CUATRIMESTRE/S: curso invierno

6.- PUNTAJE PROPUESTO PARA CARRERA DE DOCTORADO: ½ (medio punto)

7.- DURACIÓN (anual, cuatrimestral, bimestral u otra): Semanal

8.- CARGA HORARIA SEMANAL:

Teóricas: .....

Problemas: .....

Laboratorio: .....

Seminarios: .....

Teórico – Práctico: 3horas

Salida a Campo: .....

9.- CARGA HORARIA TOTAL: 15 horas

10.- FORMA DE EVALUACIÓN: Examen individual escrito

11.- PROGRAMA ANALÍTICO:

*[Handwritten signature]*  
Dr. JOSÉ OLABE IPARRAGUIRRE  
SECRETARIO DE POSGRADO  
FCEN - UBA

30

Introducción. Testeo de propiedades de distribuciones: uniformidad, cierre, distribuciones monótonas. Algoritmos sublineales para grafos. Modelos densos y esparsos. Testeo de bipartición, aproximación de max-cut, ausencia de triángulos, grado promedio, testeo de conectividad, aproximación del número de componentes conexas, aproximación del peso del "minimum spanning tree", aproximación de cobertura de nodos, algoritmos distribuidos. Algoritmos de "Sketching" y de "streaming". Estimación del número de elementos distintos, norma L2, transformación de Johnson-Lindenstrauss, filtro de Bloom, MinHash.

Testeo de propiedades de datos con valores en los reales. "Heavy hitters", "Count-Min".  
Testeo de monotonía y aproximación de distancia. Técnicas de reducción de dimensiones para el testeo de propiedades.

## 12.- BIBLIOGRAFÍA:

Noga Alon, Yossi Matias, Mario Szegedy: The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci.* 58(1): 137-147 (1999)

Broder, Andrei Z. (1997), "On the resemblance and containment of documents", Compression and Complexity of Sequences: Proceedings, Positano, Amalfitan Coast, Salerno, Italy, June 11-13, 1997, IEEE, pp. 21-29, doi:10.1109/SEQUEN.1997.666900.

Johnson, William B.; Lindenstrauss, Joram (1984), "Extensions of Lipschitz mappings into a Hilbert space", Conference in Modern Analysis and Probability (New Haven, Conn., 1982), Contemporary Mathematics 26, Providence, RI: American Mathematical Society, pp. 189-206, doi:10.1090/conm/026/737400, MR 737400.

Funda Ergün, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, Mahesh Viswanathan, Spot-Checkers. *Journal of Computer System and Sciences* 2000, STOC 1998.

Oded Goldreich, Dana Ron, Property testing in bounded degree graphs. *Algorithmica* 2002, STOC 1997.

Bernard Chazelle, Ronitt Rubinfeld, Luca Trevisan, Approximating the Minimum Spanning Tree Weight in Sublinear Time. *SIAM Journal of Computing* 2005, ICALP 2001.

Noga Alon, Michael Krivelevich, Testing k-colorability. *SIAM J. Discrete Math.* 15 (2002), 211-227.

Noga Alon, Testing subgraphs in large graphs. *Random Structures and Algorithms* 21 (2002), 359-370.

Noga Alon, Eldar Fischer, Michael Krivelevich, Mario Szegedy, Efficient testing of large graphs, *Combinatorica* 20 (2000), 451-476.

Oded Goldreich, Dan Ron, Approximating average parameters of graphs, *Random Struct. Algorithms* 32(4): 473-493 (2008)

Michal Parnas, Dan Ron, Approximating the minimum vertex cover in sublinear time and a connection to distributed algorithms, *Theor. Comput. Sci.*, 381(1-3):183--196 (2007)

Huy N. Nguyen and Krzysztof Onak, Constant-Time Approximation Algorithms via Local Improvements, *FOCS* (2008)

Krzysztof Onak, Dana Ron, Michal Rosen, and Ronitt Rubinfeld, A near-optimal sublinear-time algorithm for approximating the minimum vertex cover size, In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1123--1131 (2012)

Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev, Testing properties with respect to  $L_p$  distances, In submission to *STOC 2014*.

Graham Cormode, S. Muthukrishnan: An Improved Data Stream Summary: The Count-Min Sketch and Its Applications. *LATIN 2004*: 29-38



Universidad de Buenos Aires  
Facultad de Ciencias Exactas y Naturales

Referencia Expte. N° 499.003 vinc 45

Buenos Aires, 22 SEP 2014

**VISTO:**

la nota presentada por el Dr. Diego Fernandez Slezak, Director del Departamento de Computación, mediante la cual eleva la información y el programa del curso de posgrado **Algoritmos sublineales para grandes conjuntos de datos**, que fue dictado en el marco de la ECI 2014 teniendo como responsable al Dr. Grigory Yaroslav

**CONSIDERANDO:**

- lo actuado por la Comisión de Doctorado,
- lo actuado por la Comisión de Postgrado,
- lo actuado por este Cuerpo en la sesión realizada en el día de la fecha,
- en uso de las atribuciones que le confiere el Artículo 113° del Estatuto Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD DE  
CIENCIAS EXACTAS Y NATURALES  
RESUELVE:**

**Artículo 1°:** Dar validez al dictado del curso de posgrado **Algoritmos sublineales para grandes conjuntos de datos** de 15 hs. de duración.

**Artículo 2°:** Aprobar el programa del curso de posgrado **Algoritmos sublineales para grandes conjuntos de datos**, obrante a fs 3 del expediente de la referencia.

**Artículo 3°:** Aprobar un puntaje máximo de medio (0,5) punto para la Carrera del Doctorado.

**Artículo 4°:** Comuníquese a la Dirección del Departamento de Computación, a la Biblioteca de la FCEyN y a la Secretaría de Postgrado (con fotocopia del programa incluido). Comuníquese a la Dirección de Alumnos (sin fotocopia del programa). Cumplido Archívese.

RESOLUCION CD N° 2167  
SP/ga 03/09/2014

Dr. JOSÉ OLABE IPARRAGUIRRE  
SECRETARIO DE POSGRADO  
FCEN-UBA

Dr. JUAN CARLOS REBORADA  
DECANO