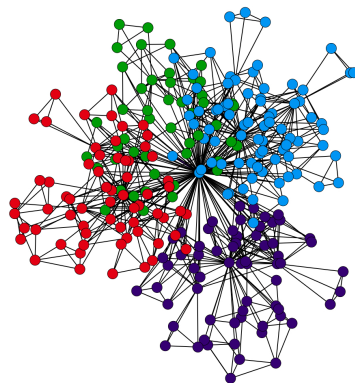


Centro Académico de Análisis de BIG DATA: **CABIDA**
Facultad de Ciencias, Universidad de la República



Construyendo futuro.

Integrante de la Red Latinoamericana de Análisis de Datos Complejos
Red LADAC
LATIN AMERICAN BIG DATA NETWORK



1. Presentación

Afirmaciones como la siguiente: “estamos en la era de BIG DATA” las encontramos sistemáticamente en los últimos tiempos tanto en el medio académico como en los medios de comunicación, las empresas y el gobierno.

Los conceptos de Data Driven Science y Big Data han llegado a ser motivos de prioridad de estado en las naciones más desarrolladas.

El volumen de datos disponibles que se pueden utilizar para mejorar las decisiones ha aumentado de forma exponencial en los últimos diez años. Internet, registros de salud, e incluso los teléfonos inteligentes se han convertido en sensores que proporcionan las fuentes de datos masivos. Por ejemplo, se puede encontrar en la página del banco BBVA lo siguiente:

“La aplicación de los análisis de Big Data ofrece una ventaja muy competitiva a las empresas. Esto no es ninguna novedad. Poco a poco más empresas asumen estas estrategias para aportar mayor valor añadido en sus decisiones cruciales.”...

De hecho, según una encuesta de Forbes Insight, elaborada para Teradata dos de cada tres directivos aseveran que estos análisis de datos ya tienen un impacto significativo y medible en los ingresos de sus empresas.”(Fuente: Innovación BBVA). Y agregan: *Uno de los puntos de inflexión está siendo el establecimiento de un diálogo entre los científicos interesados en problemas de Big Data y los líderes de negocios.* ”

Sin embargo, en general, no se define con precisión a que nos referimos con problemas de BIG DATA, y nos encontramos en muchos casos con aplicaciones rutinarias y mal desarrolladas que terminan en una gran frustración y en resultados específicos difícilmente aplicables en otros contextos. **“Big Data needs Big Theory”**(Rich Timpono Senior VP Ipsos Science Centre).

Las áreas de aplicación son muy variadas, desde la bioinformática, la matemática del genoma, las neurociencias, los sistemas biológicos complejos, la astroinformática, las finanzas, las redes de comunicaciones, las redes sociales, las transacciones comerciales, así como las ciencias sociales, el marketing y el marketing político entre otras muchas.

Sin pretender dar una definición precisa de los problemas de BIG DATA, desde nuestro punto de vista, involucran el análisis (estocástico en general) de grandes volúmenes de datos en dimensiones muy grandes o en dimensión infinita típicamente mal condicionados.

Los métodos tradicionales desarrollados por la estadística, en particular por la estadística multivariada clásica resultan insuficientes, y se requiere **desarrollar nuevas teorías y métodos estadísticos y computacionales para abordarlos en forma responsable.**

CABIDA es un centro de investigación, recientemente creado en el ámbito de la Facultad de Ciencias de la Universidad de la República, de carácter académico, interesado en las

aplicaciones de nuevos modelos matemáticos en problemas de BIG DATA, integrando la **Red de Centros de Investigación LADAC** creado para potenciar las capacidades de análisis con que contamos en nuestro país con la colaboración de cada uno de los centros adheridos a la red.

Es voluntad del Centro interactuar con otros actores en el medio local e internacional interesados en problemas de Big Data. Participarán del Centro investigadores de los los grupos de Probabilidad y Estadística del Centro de Matemática (CMAT) de la Facultad de Ciencias y del Instituto de Matemática y Estadística (IMERL) de la Facultad de Ingeniería, el grupo de Bioinformática del Instituto Pasteur, y el grupo de procesamiento de señales del Instituto de Ingeniería Eléctrica (IIE) de la Facultad de Ingeniería, así como otros investigadores interesados en problemas de BIG DATA.

En una primera etapa participarán en la red LADAC los siguientes Centros de Investigación:

- **Center for Mathematical Modeling (CMM)**. Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. En particular los Laboratorios de Computación de Alto Rendimiento, de Bioinformática y Ciencias del Genoma y de Astroinformática.
- **Cepid NeuroMat**. Universidad de Sao Paulo, Brasil.
- **Centro de Investigación en Matemática (CIMAT)**. Mexico
- **Centro de Análisis de Big Data (CABIDA)**. Facultad de Ciencias- Universidad de la República, Montevideo, Uruguay.
- **Departamento de Matemática**. Fac. de Ingeniería Química, UNL, Santa Fé, Argentina.
- **Instituto de Matemática Pura e Aplicada (IMPA)**. Rio de Janeiro, Brasil.
- **Programa de Análisis Social de la Ciudadanía Audiovisual Latinoamericana (PASCAL)** Universidad de San Martín, Argentina.

Institutos y laboratorios asociados.

- **Instituto de Neurociencia Cognitiva y Traslacional (INCyT)**. INECO, Universidad Favaloro, CONICET, Argentina.
- **Laboratorio de Investigación en Neurociencias**. Departamento de Matemática, Universidad de San Andrés, Argentina.

Entre los distintos centros que integran la red, contamos con expertos que ya han abordado problemas de BIG DATA en bioinformática, la matemática del genoma, las neurociencias, la astroinformática, las finanzas, el análisis de imágenes, el aprendizaje supervisado y semi-supervisado, el estudio de redes complejas (redes de comunicaciones, redes sociales, redes en neurociencias, internet) así como las ciencias sociales, y el marketing político entre otras, desde un punto de vista teórico y aplicado.

2. Objetivos de CABIDA

Objetivos Generales

- Nuclear a distintos grupos de investigación locales con expertise en distintos aspectos de los problemas de Big Data, con el objetivo de abordar problemas complejos.
- Estimular la interacción científica entre los integrantes del centro y con los centros de la red.

Objetivos Específicos

1. Promover y estimular a los estudiantes hacia la matemática aplicada y la bioinformática, así como hacia otras áreas de aplicación de BIG DATA. Esto incluye, la realización de cursos y seminarios específicos, la dirección de tesis de posgrado en los programas existentes y la realización de pasantías en los centros de la red.
2. Estimular el intercambio científico con expertos de la región y el mundo a través de visitas científicas a nuestro centro que incluirán el dictado de cursos y conferencias. Financiar visitas científicas en el país y hacia y desde los distintos centros de la red, aprovechándolas regionalmente.
3. Estimular la interacción entre las empresas y el sector académico, en una primera etapa a través de pasantías de estudiantes avanzados en las empresas orientados por investigadores del Centro.
4. Compartir recursos humanos y computacionales entre los distintos grupos integrantes del Centro.
5. Organizar workshops nacionales y regionales con la participación de investigadores y estudiantes de todos los centros que integran la red LADAC.
6. Presentar proyectos conjuntos a fuentes internacionales de financiación nacionales y regionales.
7. Realizar convenios con centros internacionales de Big Data a través de la red LADAC.

Estructura

CABIDA tendrá una estructura simple, con una Comisión Directiva de 5 miembros, con cinco suplentes respectivos, integrada por representantes de los grupos de investigación entre los cuales se elige un coordinador que actuará durante dos años. Dicha comisión será nombrada por el Consejo de la Facultad de Ciencias. También contará con un Consejo Asesor Internacional de excelencia.

Nuevos integrantes de CABIDA podrán incorporarse a propuesta de la Comisión Directiva. También podrán los investigadores que deseen incorporarse a CABIDA presentar su aspiración que deberá ser aprobada luego de una evaluación de la trayectoria académica del aspirante por la Comisión Directiva.

3. Plan preliminar de actividades para 2016-2017

1. Difusión de la creación y las actividades de CABIDA a nivel nacional. Creación y administración de una página web de CABIDA.
2. Elaboración de un listado de problemas abiertos en el área de Big Data.
3. Elaboración y difusión de un calendario de cursos y seminarios dictados en nuestro país por los integrantes del Centro durante 2016-2017 orientados a estudiantes de posgrado.
4. Realización de un Workshop durante 2016/2017.
5. Organización del XII ERPEM (Encuentro Regional de Probabilidad y Estadística Matemática) en el año 2017 centrado en los problemas de interés de la red LADAC.
6. Creación de núcleos regionales en problemas de bioinformática, la matemática del genoma, las neurociencias, la astroinformática, las finanzas y las ciencias sociales, liderados por uno o dos de los centros participantes en la red.
7. Inicio de ronda de contactos con empresas interesadas en problemas de Big Data. Comienzo del programa de pasantías de estudiantes en empresas con problemas de BIG DATA, orientados por investigadores de CABIDA.
8. Dirección de tesis de postgrado en problemas de Big Data.
9. Búsqueda de fuentes de financiamiento para CABIDA.