

Analítica de datos y *machine learning* en redes complejas

Curso presencial

Teléfono: +57 1 320 8320 Ext. 2111
Email: direcontinua@javeriana.edu.co

Educación **Continua**

Generamos experiencias educativas



INTENSIDAD HORARIA
Modalidad Extendida: 16 horas
Modalidad Express: 0 horas

Horarios

Lunes 13 de Junio a Jueves 16 de Junio de 8:00am a 12:00m.

Propuesta de Valor

Las redes (o grafos) son estructuras que permiten modelar, de manera natural, relaciones binarias (llamados enlaces o arcos) entre nodos. Por eso su importancia en áreas como las ciencias de la computación y las matemáticas. Más recientemente, con la relevancia de las redes sociales en nuestra vida diaria, y la aparición de nuevas tecnologías que han permitido la construcción de redes biológicas a diferentes niveles y escalas, el análisis de redes complejas (llamadas así, principalmente, por su gran tamaño) se ha convertido en una disciplina en sí misma. Problemas fundamentales como la clasificación de nodos, la predicción de enlaces faltantes en una red o en una secuencia de ellas (e.g., dada la evolución de un sistema a lo largo del tiempo) han sido centrales en el entendimiento de nuevos fenómenos y dinámicas de los sistemas que modelan. El análisis de grandes volúmenes de datos y técnicas de aprendizaje automático (en inglés, *machine learning*) han sido claves en instrumentar procedimientos para el análisis de este tipo de redes. La oferta académica de este curso está basada en resultados de proyectos de investigación y consultoría. Los algoritmos y técnicas que aquí se presentan han sido utilizados en escenarios reales y han dado resultados validados experimentalmente en el caso de las redes biológicas. La oferta académica de cursos similares en Colombia es escasa o inexistente a pesar de la creciente importancia y relevancia de analizar redes complejas en diferentes contextos. Por ello este curso, con un enfoque *hands on*, está abierto al público en general interesado en el análisis de redes complejas usando analítica de datos y *machine learning*. Finalmente, el curso contribuye al derrotero de conocer y aplicar algoritmos y técnicas del estado del arte para identificar patrones y predecir comportamientos en redes complejas, como redes sociales y biológicas.



Objetivos

General

Ejemplificar el uso de analítica de datos y aprendizaje automático (en inglés, *machine learning*) aplicando algoritmos y técnicas del estado del arte para identificar patrones y predecir comportamientos en redes complejas, como redes biológicas y redes sociales. El curso privilegia la aplicación de los conceptos a través de casos de estudio prácticos en un computador.

Específicos

- Explicar conceptos de redes complejas necesarios para entender su importancia en la modelización de sistemas biológicos y sociales, entre otros.
- Identificar los principales problemas de aprendizaje que hay sobre redes complejas.
- Conocer algoritmos y técnicas del estado del arte, con sus respectivos fundamentos, para clasificar y predecir propiedades de nodos y enlaces en redes complejas.
- Aplicar los algoritmos y técnicas sobre conjuntos de datos reales usando el lenguaje de programación Python (y algunas de sus librerías).

Dirigido a

- Ingenieros de sistemas, electrónica, multimedia o afines; Matemáticos, físicos o biólogos; o profesionales que quieran explorar temas de redes complejas y *machine learning* con aplicaciones en biología y redes sociales.

Metodología

El curso de verano se realiza con una metodología práctica donde se privilegia:

- Exposición teórica esencial complementaria a los materiales y referencias que se entregan.
- Aplicaciones de conceptos a casos de estudio.
- Laboratorios prácticos para la réplica del material presentado.

Contenido Académicos



Temarios generales trabajados por módulo. Todos los módulos están guiados a través de una combinación entre componentes teóricos y prácticos.

Temario teórico y práctico

Módulo 1. Introducción a las redes complejas y al aprendizaje automático

- Definición de red, sus componentes y propiedades.
- Tipos de redes.
- Distribución de grado de los vértices.
- Redes libres de escala.
- Representación computacional de redes.
- Análisis computacional de redes, librerías/paquetes disponibles.
- *Machine learning* para clasificación y predicción.

Módulo 2: Co-expresión génica y predicción de funciones de genes ([artículo](#))

- Definición del problema de clasificación de nodos.
- *Tree ensembles, random forests*.
- Clasificación jerárquica multi-etiqueta (hierarchical multi-label classification).
- Redes de co-expresión génica y *Gene Ontology*.
- Predicción de funciones de genes.

Módulo 3: Respuesta a estrés salino en arroz y co-expresión de comunidades de genes ([artículo](#))

- Definición de comunidades en redes.
- *Hierarchical Link Clustering* para la detección de comunidades superpuestas.
- Regresión LASSO.
- Redes de co-expresión, genotipos y fenotipos.
- Identificación de genes que responden a estrés salino en arroz.

Módulo 4: Redes sociales y predicción interacción entre usuarios ([artículo](#))

- Definición del problema de predicción de aristas o arcos.
- Redes que evolucionan a través del tiempo.
- Descomposición espectral de una red.
- Predicción de la evolución de redes de menciones en Twitter.

Conferencistas

Camilo Rocha.

Profesor Asociado del Departamento de Electrónica y Ciencias de la Computación de la Pontificia Universidad Javeriana, Cali. Es profesor universitario desde 2002 y usualmente imparte cursos de fundamentos de computación, métodos formales, algoritmos, grafos y lógica. Es Doctor en Ciencias de la Computación y Magíster en Matemáticas de la Universidad de Illinois, e Ingeniero y Magíster en Ingeniería de Sistemas y Computación de la Universidad



de los Andes. Actualmente es el Decano de la Facultad de Ingeniería y Ciencias de la Javeriana, Cali.

En su labor como investigador ha participado de un total de 8 proyectos de investigación en convocatorias nacionales e internacionales. Como resultado de su actividad investigadora, cuenta con más de 45 publicaciones arbitradas, entre revistas indexadas y capítulos en libros con ISBN. Actualmente es co-PI del proyecto In-Silico dentro del programa de investigación Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (ÓMICAS; <https://www.omicas.co>), ganador de la segunda convocatoria de Colombia Científica en el foco de Alimentos. In-Silico desarrolla y prueba modelos computacionales para mejoramiento in silico de variedades, utilizando información de genes y circuitos génicos asociados con la productividad y tolerancia a condiciones de estrés en plantas, basados en teoría de grafos, algoritmos de análisis sobre grandes volúmenes de datos y herramientas de visualización.

Actualmente está dirigiendo 4 tesis doctorales y es revisor de diversas revistas internacionales en el campo de las ciencias de la computación. Asimismo, ha colaborado en la organización y actividad docente de cursos de formación para investigadores y profesionales del sector, y ha participado como profesor invitado en diversos cursos en universidades e instituciones de investigación, tanto nacionales como internacionales.

Miguel Romero.

Estudiante de último año del Doctorado en Ingeniería y Ciencias Aplicadas de la Pontificia Universidad Javeriana, Cali. Tiene experiencia como profesor universitario en cursos de algoritmos, grafos y cálculo diferencial. Es Economista e Ingeniero de Sistemas de la Escuela Colombiana de Ingeniería Julio Garavito, Bogotá.

Ha trabajado en lógica de reescritura, análisis de redes, aprendizaje automático, algoritmos y programación competitiva. En su labor como estudiante e investigador ha sido beneficiario de convocatorias nacionales como Jóvenes Investigadores Colciencias (2017 y 2018) y del programa de investigación Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (ÓMICAS; <https://www.omicas.co>), específicamente del proyecto In-Silico.

Descripción del Curso

Modalidad

Presencial

Horarios

Lunes 13 de Junio a jueves 16 de Junio de 8:00 am a 12:00m.

Intensidad horaria

- Modalidad Extendida: 16 horas
- Modalidad Express: 0 horas



Lugar

Instalaciones de la Pontificia Universidad Javeriana

La planta física de la Dirección de Educación Continua - DEC, cumple con altos estándares de calidad en cuanto a ubicación, comodidad, estacionamiento, seguridad y acceso. Los salones son de uso exclusivo para la actividad de Educación Continua, por lo cual los participantes en nuestros programas, cuentan con espacios adecuados para su capacitación.

La Dirección de Educación Continua-DEC dispone de diecinueve (19) salas de conferencias o seminarios con capacidad promedio para 30 personas dotadas con ayudas audiovisuales básicas. Cada sala cuenta con: Tablero acrílico, Televisor de 20 o 25", VideoBeam, Retroproyector, Proyector de Diapositivas y PC disponible para proyección de presentaciones multimedia.

La Universidad Javeriana está en capacidad de ofrecer todos sus programas de Educación Continua en cualquier ciudad del país o fuera de él.