

**Formulario de aprobación de curso de
posgrado/educación permanente**

Asignatura: Análisis de datos en Redes

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado



Educación permanente



Profesor de la asignatura ¹: Dr. Pablo Rodríguez Bocca, Gr. 4, Instituto de Computación

(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹:

(título, nombre, grado, instituto)

Otros docentes de la Facultad:

(título, nombre, grado, instituto)

Docentes fuera de Facultad: Dr. Diego Kiedanski, ML Engineer, Tryolabs, Uruguay

(título, nombre, cargo, institución, país)

¹ CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

[Si es curso de posgrado]

Programa(s) de posgrado: Maestría en Informática, Doctorado en Informática

Instituto o unidad: Instituto de Computación

Departamento o área: Departamento Investigación Operativa

Horas Presenciales: 60 horas

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 10 créditos

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

Público objetivo:

Curso pensado para estudiantes de posgrado, que tienen interés/necesidad en ampliar su conocimiento en el análisis de datos, el modelado predictivo, y la matemática subyacente a los grafos. Además de una aplicación directa en las áreas de inteligencia artificial e investigación operativa, los temas de la asignatura cubren un amplio rango de aplicación en distintas áreas de informática, convirtiéndola en un buen complemento para otras disciplinas.

La suma de las visiones teórica y práctica brindan herramientas genéricas para desarrollar las capacidades de análisis e investigación del estudiante.

Cupos: No tiene cupo.

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

Objetivos:

Una red es un concepto más amplio que un grafo, y en general se utiliza para modelar sistemas reales complejos como las interacciones entre personas, proteínas, neuronas, computadoras, elementos de comunicación, sistemas de transporte, etc. El análisis de redes se clasifica (al igual que con otros datos) en descriptivo, predictivo y prescriptivo. Este curso se focaliza en el análisis descriptivo y predictivo de datos en redes, que incluyen: la visualización de las redes; el cálculo de métricas y medidas sobre grafos; el modelado de su estructura y otras propiedades; y la predicción de características o procesos en nodos, enlaces y en la red completa. En resumen, se focaliza en las técnicas útiles para extraer nuevo conocimiento de estos sistemas.

El objetivo de esta asignatura es brindar una visión general del análisis de redes, describiendo sus fundamentos, técnicas, y herramientas para su aplicación práctica.

La asignatura presenta un equilibrio entre las visiones teórica y práctica. En lo teórico, se incluyen los principales conceptos del análisis descriptivo y predictivo de redes, a saber: ¿Qué es el análisis de datos en redes? Repaso de los principales conceptos de la matemática en teoría de grafos. Cálculo de características de nodos (centralidad, redes egocéntricas,...), enlaces, y estructura (cohesión local y global, componentes, etc.). Modelado generativos de grafos (redes aleatorias, etc.), y modelado predictivo (espacios embebidos, redes neuronales para grafos, etc.). Principales problemas predictivos involucrados: predicción de características y enlaces, detección de comunidades, evaluación de procesos dinámicos y difusión en grafos.

En la parte práctica: se brindan razones para analizar redes, y los principales usos de la disciplina. ¿Cómo se analizan las redes?, ¿Qué herramientas se utilizan? Casos de aplicaciones reales.

Al finalizar el curso, el estudiante podrá diseñar y realizar investigación relacionada al análisis en redes, incluyendo: la recolección de datos, el análisis, y la documentación de resultados en formato científico (publicación académica).

Conocimientos previos exigidos:

Conocimientos previos recomendados:

Probabilidad y Estadística,
Matemática Discreta,
Algoritmos y estructura de datos,
Aprendizaje Automático,
Investigación Operativa.

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

Descripción de la metodología:

[Obligatorio]

La enseñanza se realiza fundamentalmente en modalidad de taller+proyecto, o sea, centrada en trabajos de laboratorio y actividades individuales/grupales asistidos por los docentes con el objetivo de realizar un proyecto final. Se dictarán 4 horas semanales de clase teórico-prácticas, incluyendo presentaciones teóricas, clases de práctico en computador, y monitoreo/consultas como apoyo a la realización del proyecto final. Cada alumno deberá dedicar adicionalmente un promedio de 2 horas semanales para estudio adicional y 4 horas semanales para la realización del proyecto final.

En las clases prácticas se requiere la utilización de un computador personal para realizar los ejercicios en clase. Durante las clases y de forma domiciliar se realizarán diferentes actividades que estimulen la investigación, la lectura de artículos y la discusión y reflexión sobre las

problemáticas presentadas. En todo momento, se buscará involucrar a los participantes en discusiones activas.

El proyecto final de investigación requiere las siguientes actividades por parte del estudiante: a mediados del curso realizar una propuesta por escrito del proyecto (incluyendo la fuente de los datos, y una propuesta de análisis científico sobre los datos); al final del curso entregar un informe completo del trabajo (datos, código fuente y informe de resultados) así como una presentación oral del proyecto.

Detalle de horas:

- Horas de clase (teórico): **26**
- Horas de clase (práctico): **18**
- Horas de clase (laboratorio): (incluidas en práctico)
- Horas de consulta: **12**
- Horas de evaluación: **4**
 - Subtotal de horas presenciales: **60**
- Horas de estudio: **14**
- Horas de resolución de ejercicios/prácticos: **16**
- Horas proyecto final/monografía: **60**
 - Total de horas de dedicación del estudiante: **150**

Forma de evaluación:

[Indique la forma de evaluación para estudiantes de posgrado, si corresponde]

[Indique la forma de evaluación para estudiantes de educación permanente, si corresponde]

Se realizará un proyecto de investigación en la temática de la asignatura. Este proyecto podrá ser realizado en grupos dependiendo del alcance del tema elegido para el proyecto y de la curricula de estudiantes en cada edición.

No hay examen final.

Las actividades del curso (y su ponderación para la nota final) consisten en:

- (5%) Propuesta escrita de proyecto final
- (15%) Participación en clase
- (20%) Presentación oral del proyecto
- (60%) Informe en formato científico del proyecto de investigación

La asignatura se aprueba con el 60% de los puntos, teniendo un mínimo del 60% en cada actividad (propuesta, participación, presentación, e informe).

Para la participación en clase se tomará en consideración: la resolución de los ejercicios de práctico; la asistencia a clase; la lectura previa de los temas a cubrir en cada clase; y la frecuencia/calidad de las contribuciones en las clases.

Temario:

1. Introducción:

- Introducción al análisis de redes
(análisis de datos descriptivo/predictivo/prescriptivo, aprendizaje automático supervisado/no supervisado, definiciones en ciencia de redes)
- Motivación y utilidad del análisis de redes
(datos en formato de redes, visualización, uso de las redes en distintas disciplinas, repositorios de datos disponibles, descripción de problemas emblemáticos de análisis predictivo de redes, introducción a los modelos generativos, y librerías y software usuales)
- Matemática de las redes:
 - repaso de teoría de grafos
(representación de grafos, tipos de grafos, redes de varios modos, caminos, componentes, conectividad, etc., opcional algoritmos y complejidad en redes)
 - repaso opcional en inferencia estadística
(probabilidad, estadística e inferencia, estimaciones puntuales, intervalos de confianza, test de hipótesis)
- Prácticos opcionales en teoría de grafos e inferencia estadística
- Capítulos: [SAND] Cap. 1, 2 (1-48). [SANDR] Cap. 1, 2 (1-28). [NE] Cap. 1, 2, 3, 4, 5, 6 (1-167), [opcional] 9,10,11 (273-392)

2. Análisis descriptivo de datos en redes: medidas y métricas en redes

- Mapeo del grafo de la red y visualización
- Práctico 1: Software para visualizar redes (layouts, uso de etiquetas y colores).
- Medidas de centralidad de nodos
(vectores propios, intermediación, cercanía, PageRank, hubs y autoridades, etc.)
- Práctico 2: Calcular e interpretar medidas de centralidad de nodo en redes reales
- Cohesión en las redes
(escalas de cohesión, medidas de solapamiento local y global, enunciar el problema de detección de comunidades)
- La estructura de gran escala de las redes (componentes, cliques, k-cores, efecto de pequeño mundo). Distribución del grado de conectividad (distribuciones power-law y scale-free)
- Práctico 3: Análisis descriptivo y comparativo sobre diversas redes (medidas de centralidad, cohesión y distribución de grado). Aplicar sobre los datos del proyecto final.
- Capítulos: [SAND] Cap. 3, 4 (49-122). [SANDR] Cap. 3, 4 (29-67). [NE] Cap. 7, 8 (168-270)

3. Análisis predictivo de datos en redes: predicción de información faltante, modelos generativos y procesos sobre redes

- Técnicas predictivas en redes
(espacios embebidos de nodos, redes neuronales para grafos, algoritmos de detección de comunidades, aplicación a problemas de predicción sobre redes)
- Práctico 4: Introducción a los espacios embebidos de nodos. Interpretar resultados.
- Práctico 5: Introducción a las redes neuronales para grafos. Aplicación para predecir información faltante en redes.
- Práctico 6: Detectar comunidades. Evaluar medidas de cohesión y solapamiento.

- Práctico 7: Aplicar sobre los datos del proyecto final: espacios embebidos de nodos, y/o redes neuronales para grafos, y/o algoritmos de detección de comunidades.
- Modelos generativos de redes (tradicionales y profundos), estimaciones y muestro
- Práctico 8: Crear grafos aleatorios, calcular la distribución del grado y de las componentes, calcular camino más corto, etc.
- Procesos tradicionales sobre redes: robustez y fallas de componentes, difusión de información en redes (epidemias), evolución y dinámica en redes
- Práctico 9: Evaluar el efecto de mundo pequeño para la difusión de epidemias
- Capítulos: [GRL] Cap. 3, 4, 5, 6, 8, 9 (28-74, 102-122), [SAND] Cap. 4.3.3, 5, 6, 7, 8, 9 (102-111, 123-331), [SANDR] Cap. 4.4, 5, 6, 7, 8, 9, 10 (59-66, 69-195), [NE] Cap. 12, 13, 14, 15, 16, 17, 18 (397-704)

4. [opcional] Análisis prescriptivo de datos en redes: optimización combinatoria sobre redes

- Introducción a la problemática y a las metodologías existentes

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

[SAND] Kolaczyk, E.D. "Statistical Analysis of Network Data: Methods and Models". Springer Series in Statistics. Springer New York, 2009. ISBN 9780387881461. (disponible en portal Timbó)

[GRL] Hamilton, William L. "Graph Representation Learning". Synthesis Lectures on Artificial Intelligence and Machine Learning, Volume 14, 3. Morgan and Claypool. September, 2020. (pre-publicación disponible on-line: https://www.cs.mcgill.ca/~wlh/grl_book/)

[SANDR] Kolaczyk, E.D. and Csardi, G. "Statistical Analysis of Network Data with R". Use R!, Springer New York, 2014. ISBN 9781493909834. (disponible en portal Timbó)

Recomendada:

[NE] Mark Newman. Networks: An Introduction. Oxford University Press, Inc. New York, NY, USA 2010. ISBN:0199206651 9780199206650

[BA] Albert-László Barabási, Network Science. (disponible online <http://barabasilab.com/networksciencebook/>)

[NCM] David Easley and Jon Kleinberg. Networks, Crowds, and Markets, Reasoning About a Highly Connected World. Cornell University, New York. Date Published: July 2010. ISBN: 9780521195331 (disponible online)

[ASN] Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). Analyzing Social Networks. SAGE Publications Limited. ISBN: 9781446247419.

Datos del curso

Fecha de inicio y finalización: segundo semestre lectivo completo

Horario y Salón: a confirmar

Arancel: no corresponde

[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

Arancel para estudiantes inscriptos en la modalidad posgrado: no corresponde

Arancel para estudiantes inscriptos en la modalidad educación permanente: no corresponde
