

UNIVERSIDAD NACIONAL DE LUJÁN
DEPARTAMENTO DE CIENCIAS BÁSICAS

PROGRAMA OFICIAL

DENOMINACIÓN DE LA ACTIVIDAD: **11807 - Análisis y Minería de Grafos y Redes**

TIPO DE ACTIVIDAD ACADÉMICA: **Asignatura**

CARRERA: **Licenciatura en Sistemas de Información**

PLAN DE ESTUDIOS: **17.13**

DOCENTE RESPONSABLE: **Dr. Gabriel H. Tolosa, Profesor Asociado**

EQUIPO DOCENTE:

- **Mg. Santiago Banchemo, Profesor Adjunto**
- **A.S. Andrés Giordano, Ayudante de Segunda**

ACTIVIDADES CORRELATIVAS PRECEDENTES:

PARA CURSAR: **11088 (Base de Datos Masivas (Gestión y Análisis))**
 11089 (Sistemas Inteligentes)
 11090 (Recuperación de Información)

PARA APROBAR: **11088 (Base de Datos Masivas (Gestión y Análisis))**

CARGA HORARIA TOTAL

HORAS SEMANALES: **4**

HORAS TOTALES: **64**

DISTRIBUCIÓN INTERNA DE LA CARGA HORARIA:

CLASES TEÓRICAS: **50%**

CLASES PRÁCTICAS: **50%**

PERÍODO DE VIGENCIA DEL PRESENTE PROGRAMA: **2019-2020**

PROGRAMA OFICIAL

CONTENIDOS MÍNIMOS O DESCRIPTORES (Según RES.HCS. N°478/12)

Minería de grafos y ciencia de redes. Fundamentos sobre grafos y medidas principales. Modelos de evolución de redes. Medidas aplicadas a redes sociales. Análisis de contenido, enlaces y uso en la web. Conceptos sobre visualización de datos.

FUNDAMENTACIÓN, OBJETIVOS, COMPETENCIAS

La ciencia de las redes tiene como fundamentos el estudio de las propiedades de conjuntos de elementos, de naturaleza diversa que se relacionan a través de diferentes clases de interacciones y de definen patrones de interés, modelados como un grafo. La caracterización de estos esquemas de relaciones, la extracción de información desde conjuntos de datos incompletos o ruidosos permiten - por ejemplo - estudiar desde cómo se propaga una enfermedad hasta la robustez de un sistema ante posibles ataques (dirigidos o aleatorios).

Muchos de los conceptos en ciencia de redes se basan fuertemente en la teoría de grafos, aportado su naturaleza empírica a través del estudio de grandes conjuntos de datos, de naturaleza diversa, disponibles en Internet. A partir de estos datos se construyen modelos, funciones y se investiga acerca de su utilidad práctica. Es decir, la disciplina avanza de las propiedades teóricas de conjuntos de datos "linkeados" hacia la extracción de patrones de comportamiento de sus nodos (y lo que éstos representan). En los últimos años, múltiples disciplinas (ciencias de la computación, biología, física, administración, etc.) han estudiado bajo este concepto las interacciones de diferentes individuos, cada una desde su perspectiva y con objetivos particulares. Uno de los ejemplos más claros es el análisis de redes sociales llevado al mundo digital, donde algunos de los conceptos desarrollados por la sociología tienen aplicación directa en múltiples ámbitos como por ejemplo, la detección de comunidades de individuos afines o el descubrimiento de nodos de la red con roles estratégicos (por ejemplo, de influencia en el resto).

Esta asignatura brinda una introducción amplia y los fundamentos de la denominada ciencia de las redes (Network Science) junto con un conjunto de aplicaciones que permite no solamente comprender cómo se comportan los individuos bajo ciertas condiciones sino – además – aplicar y ejercitar múltiples conceptos de ciencias de la computación sobre conjuntos de datos "linkeados", los que requieren estructuras de datos y algoritmos eficientes.

OBJETIVOS GENERALES

Se espera que al completar el curso los estudiantes:

- Comprendan los alcances de la disciplina, junto con criterios que les permitan determinar sus ámbitos de aplicación.
- Cuenten con los fundamentos teóricos sobre los modelos de evolución de redes y su aplicación práctica.
- Implementen las estructuras de datos necesarias de almacenamiento y recuperación de grafos masivos.
- Adquieran criterios para utilizar la minería de grafos para la extracción de información valiosa en diferentes dominios.

PROGRAMA OFICIAL

- Aumenten sus capacidades para la implementación de módulos de software, en particular a partir de implementar técnicas complejas de recorrido en grafos.

Complementariamente, se propone que también incrementen sus habilidades para:

- Redactar informes de desarrollo, reportes técnicos o trabajos de investigación siguiendo objetivos y metodología concreta.
- Comunicar sus conocimientos, resultados de investigación a pares y/o superiores en presentaciones públicas.

CONTENIDOS

Unidad 1 – Introducción a la ciencia de las redes

Motivación y conceptos sobre ciencia de las redes. Fundamentos sobre grafos aplicados a ciencia de redes. Caminos y conectividad. Distancias y búsquedas. Grafos ponderados. Coeficiente de clustering. Ejemplos y aplicaciones sobre conjuntos de datos reales.

Unidad 2 – Modelos en grafos

El fenómeno de los mundos pequeños. Teoría de los seis grados de separación. Modelos de redes: aleatorias y libres de escala. Modelos de evolución de redes. Propagación en grafos. Modelos de dispersión de epidemias.

Unidad 3 – Análisis de Redes Sociales

Caracterización de las redes sociales digitales. Medidas de centralidad, homogeneidad y reputación. Comunidades: estructuras y modelos. Formación de opinión, coordinación y cooperación.

Unidad 4 – Minería de la Web

Extracción de Información. Clustering de documentos y detección de duplicados. Búsquedas por similitud basada en enlaces. Minería de uso: análisis de query logs. Sistemas de recomendación.

Unidad 5 – Visualización

Algoritmos de diseño para grafos. Aplicación a grafos planos. Árboles. Visualización de matrices. Clusters. Modelos para redes sociales. Herramientas de visualización de grafos. Agregación de datos y mashups.

Unidad 6 – Aplicaciones

Crawling focalizado. Análisis de sentimientos. Motores de búsqueda. Introducción a Big Data en la web. Procesamiento en tiempo real (streams).

PROGRAMA OFICIAL

METODOLOGÍA

El desarrollo del curso es de carácter teórico/práctico, con aplicación de los conceptos a las actividades propiamente de laboratorio. En las clases teóricas se plantean los conceptos, modelos, ejemplos y aplicaciones del área y demás temas propuestos en este programa.

En las clases prácticas se realizan implementaciones de los modelos desarrollados como así también de experimentos de programación y análisis de datos. Se trabaja tanto con software propio como con librerías públicas existentes y ampliamente utilizadas para la práctica de la disciplina.

Complementariamente, los estudiantes deben preparar una exposición sobre la base de la lectura e investigación de un tema propuesto por el equipo docente. Esta actividad introduce en la lectura de literatura netamente de investigación y se propone como motivadora para la discusión en clase con todo el grupo. La última actividad de evaluación consiste en un trabajo final de curso sobre algún tema del programa. Éste puede ser de carácter teórico/práctico (modelización y experimentación) o del desarrollo de una aplicación que se diferencie por implementar los conceptos presentados en la cursada.

ACTIVIDADES PRÁCTICAS

En las actividades prácticas se considera tanto la resolución de problemas como la ejercitación de laboratorio. Con las mismas se pretende reforzar los conceptos planteados en clase ya que permiten la exploración y aplicación concreta de la mayoría de los temas.

En las tareas de laboratorio se deben realizar pequeñas aplicaciones orientadas a diferentes problemas del área como recorrido de grafos (pequeños vs masivos), cálculo de métricas, uso de estructuras de datos eficientes, análisis de redes sociales, experimentos de propagación de información, entre otros. Complementariamente, se utilizan herramientas libres existentes a modo demostrativo o cuando el tema lo requiere.

Las aplicaciones se deben programar en lenguaje Python (se utilizará R cuando se requiera) y los alumnos deben demostrar sus habilidades en la programación como así también en el análisis de la situación propuesta previo a la construcción de la solución. En ambos casos, cuentan con el soporte del equipo docente.

Para el trabajo final, los alumnos deben presentar su propio proyecto, el cual se discute con los docentes. En éste deben realizar una investigación relacionada a algunos de los tópicos desarrollados en la asignatura o bien una propuesta con estudio experimental de algún enfoque alternativo a las técnicas existentes. En cualquiera de los casos, se debe elaborar un documento con formato de artículo de investigación (*paper*) donde se expongan los objetivos, antecedentes, la propuesta, la metodología utilizada y los resultados obtenidos.

PROGRAMA OFICIAL

REQUISITOS DE APROBACION Y CRITERIOS DE CALIFICACIÓN:

La evaluación consta de 1 (un) examen parcial y un trabajo final integrador (descrito en el apartado anterior) obligatorio. El examen parcial se aprueba con nota 4 (cuatro) o superior mientras que el integrador con 7 (siete) o superior.

CONDICIONES PARA PROMOVER (SIN EL REQUISITO DE EXAMEN FINAL), DE ACUERDO AL ART.23 DEL REGIMEN GENERAL DE ESTUDIOS RESHCS-LUJ:0000996-15

- a) Tener aprobadas las actividades correlativas al finalizar el turno de examen extraordinario de ese cuatrimestre.
- b) Cumplir con un mínimo del 80% de asistencia para todas las actividades.
- c) Aprobar todos los *trabajos prácticos* previstos en este programa, pudiendo recuperarse hasta un 25% del total por ausencias o aplazos.
- d) Aprobar el 100% de las evaluaciones previstas con un promedio no inferior a seis (6) puntos sin recuperar ninguna.
- e) Aprobar una evaluación integradora de la asignatura con calificación no inferior a siete (7) puntos.

CONDICIONES PARA APROBAR COMO REGULAR (CON REQUISITO DE EXAMEN FINAL)
DE ACUERDO AL ART.24 DEL REGIMEN GENERAL DE ESTUDIOS RESHCS-LUJ:0000996-15

- a) Estar en condición de regular en las actividades correlativas al momento de su inscripción al cursado de la asignatura.
- b) Cumplir con un mínimo del 70% de asistencia para todas las actividades.
- c) Aprobar todos los trabajos prácticos previstos en este programa, pudiendo recuperarse hasta un 40% del total por ausencias o aplazos.
- d) Aprobar el 100% de las evaluaciones previstas con un promedio no inferior a cuatro (4) puntos, pudiendo recuperar el 50% de las mismas. Cada evaluación solo podrá recuperarse en una oportunidad.

EXAMENES PARA ESTUDIANTES EN CONDICIÓN DE LIBRES

- a) Para aquellos estudiantes que, habiéndose inscriptos oportunamente en la presente actividad hayan quedado en condición de libres por aplicación de los artículos 22, 25, 27, 29 o 32 del Régimen General de Estudios, podrán rendir en tal condición la presente actividad.
 - b) Para aquellos estudiantes que no cursaron la asignatura y se presenten en condición de alumnos libres en la Carrera, por aplicación de los artículos 10 o 19 del Régimen General de Estudios, podrán rendir en tal condición la presente actividad. Además, quince días antes de la fecha de sustanciación de mesa, el alumno deberá entregar la resolución de todas las actividades prácticas vigentes en la última cursada.
-

PROGRAMA OFICIAL

BIBLIOGRAFÍA

SUGERIDA

- A. Barabasi. Network Science. Cambridge University Press; 1 edition, 2016. <http://networksciencebook.com/>
- M. van Steen. Graph Theory and Complex Networks: An Introduction, 2010.
- G. Caldarelli. Scale-Free Networks: Complex Webs in Nature and Technology. Oxford University Press, 2007.
- P. Carrington, J. Scott, S. Wasserman (Eds). Models and Methods in Social Network Analysis (Structural Analysis in the Social Sciences). 1st Edition, 2005.
- Zinoviev, D. Complex Network Analysis in Python. Pragmatic Bookshelf; 1 edition, 2018.

DE CONSULTA

- D. Easley, J. Kleinberg. Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010.
- T. Lewis. Network Science: Theory and Applications. Wiley; 1 edition, 2009.
- M. Newman, A. Barabási, D. Watts. The Structure and Dynamics of Networks: (Princeton Studies in Complexity) 1st Edition. Princeton University Press, 2006.
- S. Chakrabarti. Mining the Web. Discovering Knowledge from Hypertext Data.. Morgan-Kaufmann Publishers. 2003.
- A. Barabasi, J. Frangos. Linked: The New Science of Networks. Perseus Books Group; 1st edition, 2002.

RECURSOS ADICIONALES

El equipo docente mantiene un sitio web de la asignatura (<http://www.labredes.unlu.edu.ar/>) en el cual se publica el material regular y las novedades. Todos los años se actualiza una lista de artículos de investigación, tutoriales y white papers que se utilizan durante la cursada. Además, se atienden durante todo el año consultas por correo electrónico y/o sesiones de chat.

UNIVERSIDAD NACIONAL DE LUJÁN
DEPARTAMENTO DE CIENCIAS BÁSICAS

PROGRAMA OFICIAL

DISPOSICIÓN DE APROBACIÓN: CD