

Análisis de Redes Sociales

Guillermo Jiménez Díaz (gjimenez@ucm.es)
Alberto Díaz (albertodiaz@fdi.ucm.es)

30 de septiembre de 2015

Prefacio

Estos son los apuntes de la asignatura Análisis de Redes Sociales, impartida en la Facultad de Informática de la Universidad Complutense de Madrid por los profesores Guillermo Jiménez Díaz y Alberto Díaz, del Departamento de Ingeniería del Software e Inteligencia Artificial.

Este material ha sido desarrollado a partir de distintas fuentes, destacando como referencia principal el libro *Network Science* de Laszlo Barabasi, el libro *Social and Economic Networks* de Mathew O. Jackson y el material de la asignatura *Social Network Analysis* impartido por Lada Adamic a través de Coursera.

Tema 0: Presentación de la asignatura

0.1 Ejemplos motivadores

0.1.1 Las redes y la captura de Saddam Hussein

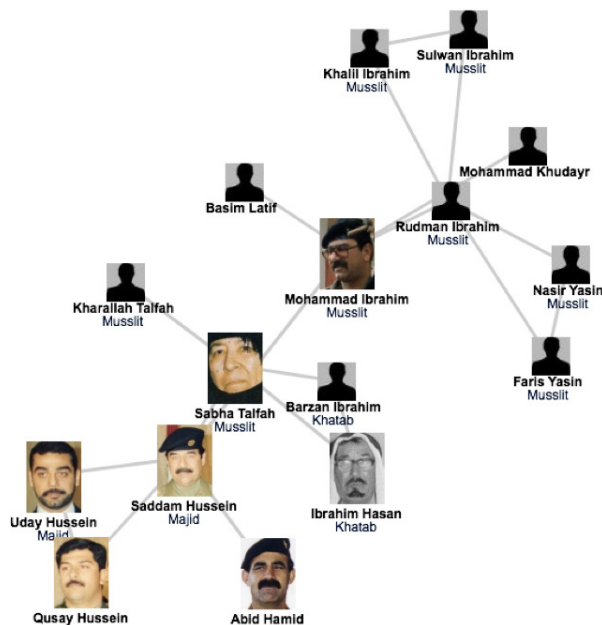


Figure 1: La red de Saddam Hussein

La captura de Saddam Hussein ilustra algunos de los aspectos claves de las redes que estudiaremos en esta asignatura:

- Muestra el poder predictivo de las redes, que permiten extraer información clave incluso a los no expertos (los soldados en este caso)
- Resalta la necesidad de diseñar mapas precisos de las redes a estudiar (en muchos casos, el proceso de diseño es muy complicado y costoso)

- Ejemplifica el hecho de que la elección de la red a emplear marca la diferencia (los militares americanos tardaron meses en darse cuenta que la red jerárquica que representaba la organización oficial iraquí era inútil para encontrar a Saddam)
- Destaca la estabilidad de algunas redes de confianza (la captura de Hussein no se basó en las técnicas clásicas de Inteligencia sino en sus conexiones sociales antes de la invasión, extraídas de viejas fotos de su álbum familiar)

0.1.2 Predicción de la epidemia de la gripe aviar (H1N1) en 2009



Figure 2: Predicción de la difusión del contagio de la gripe aviar. Extraída de [esta charla de TED](#).

La gripe aviar de 2009 es la primera pandemia cuya evolución fue predicha meses antes de que alcanzara su punto álgido:

- Haciendo uso de las redes de transporte a nivel mundial, se determinó correctamente que alcanzaría su pico en Octubre de 2009, en lugar de en Enero-Febrero (picos habituales de la gripe estándar).
- Se demostró que la vacunación masiva realizada en Noviembre de 2009 fue inútil por ser demasiado tardía, hecho comprobado a posteriori.
- El cambio fundamental consistió en considerar **el rol de las redes en la propagación de los virus**, en contraste con los modelos epidémicos clásicos.
- Hoy en día, la predicción de epidemias es una de las aplicaciones más activas del Análisis de Redes. No sólo se centra en virus biológicos sino también electrónicos (ej: predicción de la infección de 300.000 teléfonos móviles en China en 2010).

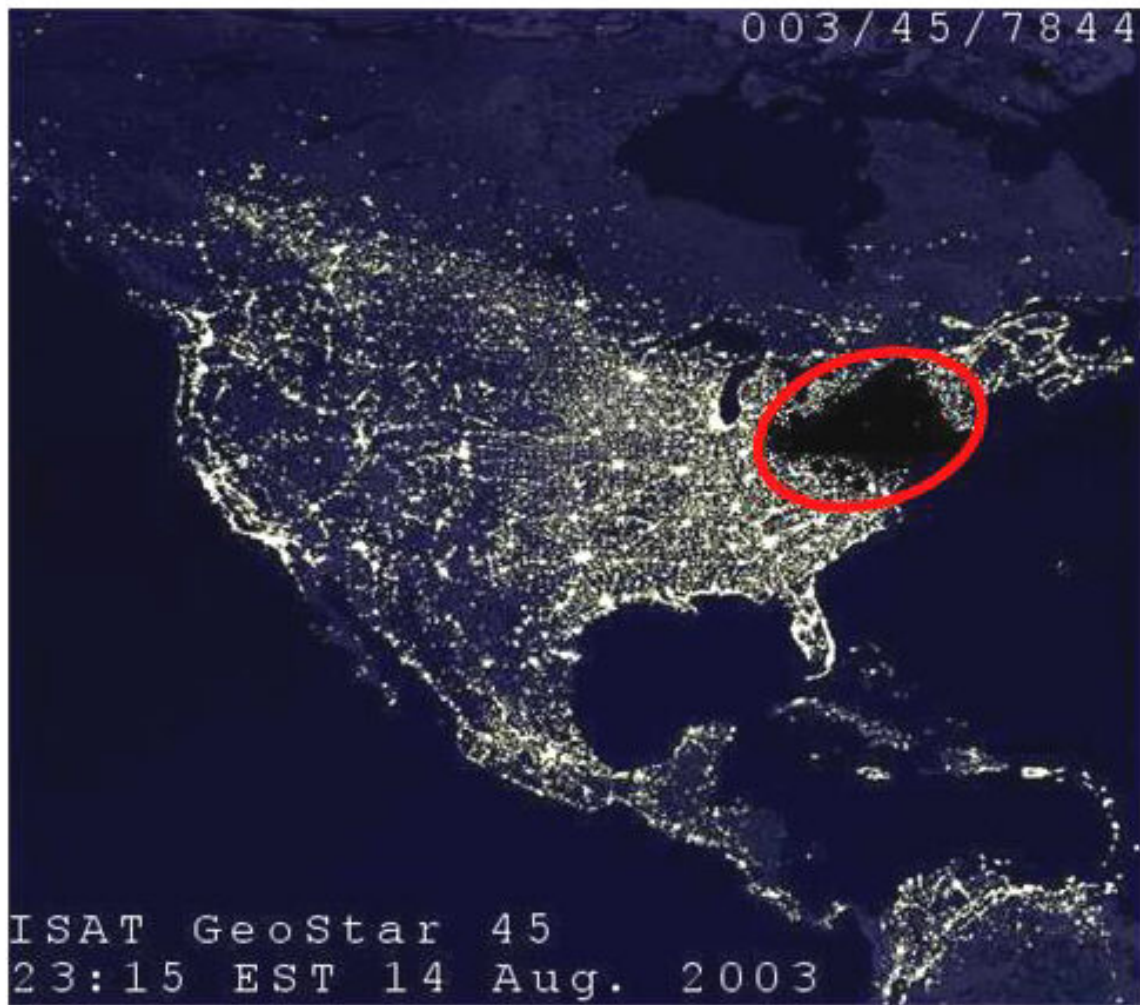


Figure 3: Apagón de 2003 debido a una sobrecarga y un fallo en cascada

0.1.3 El apagón de la costa noroeste de EEUU el 14/08/2003

- Este apagón es un ejemplo típico de un **fallo en cascada**: cuando una red actúa como sistema de transporte, un fallo local en un nodo provoca una transferencia de carga a otros nodos. Si la carga extra es excesiva para los nodos vecinos, éstos pueden fallar y redistribuirla a otros nodos a su vez (cayeron 256 plantas y 508 unidades generadoras).
- La magnitud del fallo depende de la posición en la red y la capacidad de los nodos afectados (eliminados) en el primer momento y en los siguientes.
- Los fallos en cascada son habituales en los Sistemas Complejos (ej: tráfico en Internet). La actual crisis financiera mundial es un ejemplo provocado por la crisis de crédito en los EEUU.
- También pueden tener efectos positivos. Ej: tratamientos del cáncer.
- La estructura de la red afecta a la robustez del sistema. Se pueden establecer herramientas cuantitativas que evalúen la relación entre la estructura de la red y los procesos dinámicos que se producen en ella, así como su impacto en los fallos.
- Esos fallos no son caóticos e impredecibles, siguen una serie de leyes reproducibles.

0.2 Sistemas complejos

Existen un gran número de sistemas complejos que son difíciles de comprender y analizar. Generalmente, detrás de cada sistema complejo estudiado siempre hay un diagrama de conexiones, una red, que define las interacciones entre sus componentes. No seremos capaces de entender los sistemas complejos a menos que podamos mapear y comprender las redes que los soportan.

A pesar de las diferencias aparentes en componentes e interacciones, las redes que regulan los distintos sistemas complejos existentes en nuestro mundo son similares, siguen unas leyes comunes y presentan mecanismos reproducibles.

Ejemplos de sistemas complejos:

- El cerebro humano es una red de neuronas
- La sociedad es una red de conexiones familiares, profesionales y de amistad entre individuos.
- Un sistema de comunicación es un conjunto de dispositivos de comunicación que interaccionan a través de Internet o de enlaces wireless.
- La red eléctrica se compone de generadores y líneas de transmisión entre ellos.
- El comercio y la economía se puede representar como una red de intercambio de bienes y servicios entre personas, empresas y países.
- Internet es una red de páginas enlazadas entre sí.
- El genoma humano se puede representar como una red de interacción entre los genes que lo conforman.

0.3 La emergencia de las redes y los sistemas complejos

Se puede decir que el análisis de las redes sociales es la ciencia del siglo XXI debido a su auge en investigación esto se debe a varios hechos:

0.3.1 Necesidad (urgente) de entender la complejidad

Cada vez está más aceptado el hecho de que, a pesar de su dificultad, no podemos permitirnos no entender el comportamiento de los sistemas complejos. Varios de los avances más significativos para entender la complejidad obtenidos en la última década provienen de la Teoría de Redes.

0.3.2 Disponibilidad de Datos

- Red de Actores de Cine, 1998
- La World Wide Web, 1999
- Red Neuronal del gusano C.Elegans, 1990
- Redes de Citas de Artículos Científicos, 1998
- Genoma Humano, 2001
- Red de Interacciones entre Proteínas, 2001
- Extracción de datos en principales redes sociales (Facebook, Twitter...), actualidad.

0.3.3 Universalidad

La arquitectura de las distintas redes que están apareciendo en varios dominios de la ciencia, naturaleza y tecnología es más similar de lo que podría esperarse en principio.

0.4 ¿Para qué sirve el Análisis de Redes Sociales?

El Análisis de Redes Sociales (ARS), ciencia de redes (Network Science) o análisis de sistemas complejos es multidisciplinar, ya que usa herramientas propias de la teoría de grafos, la física, estadística, informática, biología y Teoría de redes sociales, entre otros.

Entre otras cosas, el ARS sirve para:

- Entender la estructura y el comportamiento de un sistema complejo.
- Extraer información y realizar predicciones.
- Visualización de información (en forma de red).

- Detección de nodos de especial relevancia en la dinámica de la red (influenciadores y concentradores).
- Detección de vulnerabilidades y posibles fallos en cascada debido a las interconexiones entre nodos (por ejemplo, en una red eléctrica por sobrecarga de una central o en internet, debido a un ataque por denegación de servicio en un router).
- Entender cómo la estructura de la red afecta a la robustez de la misma.
- Detección de grupos y comunidades dentro de la misma.
- Conocer los procesos dinámicos que aparecen en ella: difusión de información, contagio emocional. . .

0.5 El Análisis de Redes Sociales y la Teoría de Grafos

Como veremos más adelante, el ARS hace uso de la teoría de grafos matemática, aunque tiene ciertas diferencias:

- El ARS es más empírico que la teoría de grafos.
- El ARS se centra en los datos y en la utilidad de los mismos.
- Usa, al igual que la teoría de grafos, conceptos y modelos matemáticos para describir las propiedades de la red.
- Trabaja con grandes cantidades de información por lo que genera grafos enormes.

0.6 Aplicaciones del Análisis de Redes Sociales

- **Salud:** Medicina y genómica, redes de enfermedades humanas y genes y proteínas implicados, investigación sobre el cerebro humano y redes neuronales, predicción de epidemias.
- Lucha contra el **terrorismo**
- **Economía y Marketing:** predicción económica; recomendaciones y anuncios. Detección de influenciadores, personas con una importante conexión con otras personas, como foco para la propagación de campañas.
- **Gestión:** estructura de organizaciones, identificación de líderes de opinión, grupos óptimos. . .

0.7 Objetivo general del curso

En esta asignatura se pretende enseñar a identificar las principales propiedades de una red, los métodos y modelos que se utilizan para identificarlas y caracterizarlas a partir de datos reales, la detección de grupos y comunidades así como de nodos centralizadores, y los modelos relacionados con el flujo de información a través de una red social.

0.8 Contenidos del curso

1. Introducción.
2. Propiedades básicas de las redes
3. Centralidad
4. Modularidad y detección de comunidades
5. Visualización y poda de redes
6. Modelos de redes: aleatorias, libres de escala y pequeños mundos.
7. Dinámica de las redes: Propagación y difusión.
8. Robustez.

Las clases teóricas se impartirán en el aula 12 los viernes y las clases prácticas en el laboratorio 11 los miércoles. En algunas sesiones prácticas se impartirán seminarios para facilitar la realización de las prácticas.

0.9 Evaluación de la asignatura

La evaluación final del curso se compone de dos partes:

0.9.1 Prácticas (30%)

Las prácticas son **obligatorias** y es necesario que estén **aprobadas** para que sean tenidas en cuenta en la nota final:

- Entregadas en plazo
- Satisfaciendo los requisitos del enunciado
- Obteniendo una nota entre 5 y 10

Las prácticas se realizarán en grupos de 3 o 4 alumnos. Serán 2-3 prácticas que se irán ajustando a los conceptos teóricos que se vayan contando en clase.

Las prácticas no entregadas o suspensas se tendrán que volver a entregar en septiembre.

0.9.2 Proyecto (70%)

El proyecto final consistirá en el análisis de un conjunto de datos real elegido por el grupo de alumnos. Se realizará con los mismos grupos que las prácticas y será presentado en defensa pública.

Algunos proyectos realizados durante el curso pasado fueron los siguientes:

- Análisis de la red de coautorías de profesores de la Facultad de Informática.

- Estudio de los modelos de contagio sobre la red de trenes de ADIF
- Análisis de la red de pases de la final de la Champions de 2014 (Real Madrid vs. Atlético de Madrid).
- Simulador temporal del comportamiento de los nodos de una red. Podéis verlo en funcionamiento [en este Vídeo de Youtube](#)
- Análisis de enlaces y hashtags de Twitter.

El proyecto es obligatorio y debe cumplir los mismos requisitos que las prácticas para poder aprobar la asignatura: entregado en plazo, cumpliendo los requisitos del enunciado y obteniendo una nota mayor o igual que 5.

Al igual que con las prácticas, los proyectos no entregados o suspensos se tendrán que volver a entregar en septiembre.

0.10 Bibliografía

0.10.1 Principal

- [Network Science](#). Laszlo Barabasi.
- [Networks, Crowds and Markets](#). David Easley and Jon Kleinberg, Cambridge University Press. 2010.
- Social and Economic Networks. Matthew O. Jackson. Princeton University Press. 2008.
- [Apuntes del curso pasado en GitHub](#)

0.10.2 Complementaria

- The Structure and Dynamics of Networks. Mark Newman, Albert-László Barabási, and Duncan J. Watts. 2006.
- [Linked: The New Science of Networks](#). Albert-Laszlo Barabasi, Jennifer Frangos.