



## **Workshop “Big data, redes y analítica social”**

**Facultad de Informática - Universidad Complutense de Madrid**

**24 de noviembre de 2017**

Las tecnologías digitales, en particular las redes sociales digitales, están propiciando la aparición de espacios para la coordinación entre individuos y para la creación y transmisión de contenidos creados y diseñados por ciudadanos. En tanto que, en muchos casos, estas actividades tienen lugar de forma autónoma fuera del marco de organizaciones estructuradas, la aparición de estas tecnologías está generando un importante proceso de cambio social, una transformación a la que algunos autores se refieren como “desintermediación”, en la que la participación social en actividades de diverso tipo (económicas, políticas, culturales) podría, al menos en parte, dejar de estar mediada por organizaciones estructuradas tradicionales para canalizarse a través de los espacios posibilitados por las nuevas tecnologías.

Aparte del interés que por sí mismo puede tener el estudio de estos procesos de transformación social, la creciente penetración de las redes sociales digitales y la gran cantidad de información que los ciudadanos transmiten en ellas está suscitando asimismo el interés en su análisis como complemento de algunas herramientas metodológicas tradicionales de la investigación empírica en las ciencias sociales, como las encuestas con trabajo de campo. Por otro lado, esta cada vez más masiva difusión de las redes digitales, con su ingente cantidad y diversidad de contenidos, está planteando nuevos retos a las técnicas tradicionales de almacenamiento, tratamiento y análisis de la información. Esto ha llevado al surgimiento, en años recientes, del paradigma *big data*, desarrollado para posibilitar la extracción de conocimiento socialmente relevante a partir de las enormes acumulaciones de datos altamente diversos que generan estas redes digitales.

Este encuentro sobre *Big data, redes y analítica social* está organizado por el proyecto de investigación Santander-UCM titulado “¿Nos dirigimos hacia una sociedad desintermediada? Un estudio de la participación social mediante análisis de redes, big data y aprendizaje automático”. El equipo multidisciplinar de este proyecto, una mixtura de científicos sociales, del tratamiento de la información y del análisis de datos, tiene por objetivo reunir en este evento a prestigiosos investigadores, provenientes de áreas diversas como la Física, la Sociología o la Informática, en el marco de una discusión abierta sobre los aspectos técnicos de las metodologías de big data y de análisis de redes, el papel de estas tecnologías en la metodología de investigación social y las aplicaciones actuales y perspectivas futuras en este campo en fuerte desarrollo.



## Programa

8:30 – 9:00: **Apertura**

9:00 – 11:00: **Primera sesión**

**Estrella Gualda** (Universidad de Huelva): *Social Big Data y retos técnicos transdisciplinares: aplicaciones en clave reticular.*

**María Luz Congosto** (Universidad Carlos III de Madrid): *Twitter, herramienta social: origen y presente.*

**Guillermo Jiménez** (Universidad Complutense de Madrid): *Análisis de Redes Sociales sin redes sociales.*

11:00 – 11:30: **Pausa**

11:30 – 13:30: **Segunda sesión**

**Alex Arenas** (Universidad Rovira i Virgili): *Modelizando la “Darknet” o red oscura.*

**Modesto Escobar** (Universidad de Salamanca): *El análisis reticular de coincidencias. Usos del 'big data' para analizar los 'small data'.*

**Rosa Benito** (Universidad Politécnica de Madrid): *Midiendo la polarización social a través de Twitter.*

13:30 – 14:30: **Mesa redonda**

14:30 – 14:45 **Clausura**

### Inscripción:

La inscripción es gratuita, pero es necesario registrarse a través del siguiente formulario web <https://goo.gl/PbxYLN>

### Localización:

Sala de Grados de la Facultad de Informática (1ª planta)

C/ Profesor José García Santesmases, 9  
Ciudad Universitaria  
28040 - MADRID  
<http://informatica.ucm.es/contacto>



## **Abstract de las conferencias**

***Modelizando la "Darknet" o red oscura (Alex Arenas):*** Si bien las propiedades estadísticas y de robustez de Internet ya no cambian significativamente a lo largo del tiempo, Darknet, una red dedicada a mantener el tráfico anónimo, aún experimenta cambios para mejorar la seguridad de sus usuarios. Presentaremos un estudio de la estructura de Darknet y encontramos que su topología es bastante peculiar, caracterizada por una distribución no homogénea de conexiones, típica de redes libres de escala, longitudes de caminos muy cortas, alta clusterización y la falta de un núcleo de nodos altamente conectados. Proponemos un modelo para reproducir tales características, demostrando que los mecanismos utilizados para mejorar la ciberseguridad son responsables de la topología observada. Inesperadamente, revelamos que su peculiar estructura hace que Darknet sea mucho más resistente que Internet, utilizado como punto de referencia para la comparación en un nivel descriptivo, con fallos aleatorios, ataques dirigidos y fallos en cascada.

***Midiendo la polarización social a través de Twitter (Rosa Benito):*** Presentamos varios estudios empíricos a partir de grandes cantidades de datos obtenidos de diversas conversaciones de la red social Twitter y de llamadas telefónicas (CDRs, registros digitales de llamadas) para analizar los patrones de comportamiento humano utilizando técnicas de redes complejas y de análisis de datos. En particular, nos centramos en el estudio de conversaciones políticas como campañas electorales y temas que pueden producir polarización social. Presentamos varios indicadores desarrollados en el grupo para medir la influencia de usuarios o el apoyo relativo en conversaciones de twitter, así como para medir el grado de polarización política. También presentamos un modelo de formación de opinión basado en la estructura de la red social de retuits. Por otro lado, utilizando grandes cantidades de datos provenientes de CDRs hemos analizado los patrones de movilidad de usuarios a diferentes escalas en países europeos y migraciones en países en desarrollo.

***Twitter, herramienta social: origen y presente (María Luz Congosto):*** Desde el año 2009 en el que un grupo de entusiastas de la cultura libre se organizaron para crear una campaña de comunicación en Twitter contra la llamada Ley Sinde hasta los últimos acontecimiento del 2017 en Cataluña, en el que el papel de esta plataforma social ha sido sumamente relevante, la forma de utilizar Twitter se ha ido transformándose. De una posición inicial, con una participación social minoritaria, con pocos Influencers, y enfocada a difundir en positivo campañas de comunicación se ha derivado una ampliación de la audiencia con la aparición de figuras muy populares, muchas veces con un discurso agresivo, más enfocado en la descalificación del contrario que a la defensa de sus propias ideas. España ha vivido durante este periodo muchos sucesos relevantes como el 15M, las protestas por los recortes, los escándalos de corrupción, más de una docena de elecciones entre generales, autonómicas y europeas y finalmente el proceso independentista. Todo ello ha dejado su traza en Twitter y nos muestra cómo va evolucionando el comportamiento social en las redes virtuales online.

***El análisis reticular de coincidencias. Usos del 'big data' para analizar los 'small data' (Modesto Escobar):*** Ante la emergencia de los big data generados tanto por la masiva digitalización, como por el creciente acceso a la información de la llamada segunda revolución digital, los científicos sociales nos enfrentamos a una serie de retos para entender la realidad social mediante nuevas colecciones de datos, métodos alternativos de muestreo, codificaciones automáticas y potentes análisis estadísticos de la información. En esta conferencia se propone un análisis de la información basado en datos dicotómicos. La idea es construir matrices binarias formadas por



1) conjuntos espaciales o temporales, 2) escenarios y 3) eventos o características de cuyos atributos también cabe dar cuenta. El tratamiento de esta estructura se basa en la metodología de las redes de doble modo combinada con los instrumentos estadísticos útiles para la selección, localización y representación de los nodos y sus vínculos. El análisis reticular de coincidencias propuesto se construye con un paquete en R disponible en el CRAN (netCoin) y se puede ejecutar en línea de modo gráfico e interactivo. Será explicado usando una variedad de ejemplos que van desde el análisis de colecciones de fotos, el análisis de contenido de textos, la representación de correspondencia epistolar, conciertos y exposiciones... hasta el análisis de múltiples respuestas en los cuestionarios.

**Social Big Data y retos técnicos transdisciplinares": Aplicaciones en clave reticular (Estrella Gualda):** Se presenta el campo de los "Social Big Data" como aquel donde interaccionan los Social Media, el Data Analysis y los Big Data (Bello-Orgaz, Jung y Camacho, 2016). En esta área inter/transdisciplinar se abordan métodos y técnicas que trabajan con datos procedentes de los medios sociales para extraer conocimiento de los mismos con el fin de poder ofrecer información útil. La información extraída a partir de los medios sociales permite en las Ciencias Sociales el avance de la investigación documental aplicando una diversidad de métodos, técnicas y procedimientos, una parte de los cuales ya se habían empleado en ámbitos como la estadística, la sociología visual, el análisis de redes sociales, el análisis del discurso, etc. Algunos aspectos que marcan la diferencia tienen que ver con el estudio de la transmisión de información en los medios sociales en directo (*streaming*) y con la ingente cantidad y diversidad de datos que se manejan (*big data*), aspectos estos que retan a la Ciencia. Junto a la gran dimensión y magnitud de los SBD, también son importantes aspectos como su heterogeneidad estructural, su dinamismo, las diferentes fuentes de los que proceden o el hecho de que gran parte de los mismos son datos no estructurados (ej. vídeo), lo cual dificulta aspectos como el análisis automático. En cualquier caso, se trate de las 3V, de 4V, de 5V, de 6V lo que caracteriza a los Big Data (gran volumen de datos, variedad, velocidad, veracidad, variabilidad y complejidad y valor), con su aparición, como resultado, se producen cambios en el proceso de investigación. Emergen nuevas tareas y se reformulan aspectos importantes de la labor investigadora, por ejemplo, en procesos como el muestreo, la recogida y almacenamiento de la información, los procesamientos y el análisis. A partir de datos extraídos de Twitter, y sobre todo, basándonos en el análisis de redes sociales, se presentan algunos de los desafíos técnicos que hemos ido encontrando a través del análisis de diferentes *datasets* de carácter social y político (refugiados y migraciones; incendios y emergencias; StarWars; #prays for...; Cataluña, el referéndum, el 155, etc.). Se plantean algunas preguntas, se discuten algunas soluciones y se reflexiona sobre el desarrollo de un campo de *Social Big Data* transdisciplinar orientado a resolver problemas y contestar preguntas más que a consolidar límites disciplinares. Paralelamente, a partir del análisis de varios casos, se aportan ideas respecto a la idea de sociedad desintermediada, reflexionando sobre cómo parece estructurarse la participación social en una red como Twitter.

**Análisis de Redes Sociales sin redes sociales (Guillermo Jiménez):** El Análisis de Redes Sociales puede parecer una disciplina que es relativamente actual debido a la gran cantidad de datos sobre interrelaciones personales que existen actualmente online. Sin embargo, muchas de las técnicas y propiedades que se estudian en esta disciplina se encuentran bajo el paraguas de lo que se conoce como el estudio de Redes Complejas o *Complex networks*, una disciplina con mayor recorrido, que se aplica a diferentes dominios y que usa datos de fuentes ajenas a las redes sociales. En esta charla se motivará el análisis de datos desde la perspectiva del análisis de redes complejas y se presentarán ejemplos reales de la aplicación de las técnicas propias del análisis de redes a fuentes de datos que no necesariamente han sido extraídas de redes sociales,

como el análisis de un partido de fútbol, la identificación de investigadores relevantes y grupos de investigación en una Facultad o el desarrollo de sistemas de recomendación a partir de las interacciones usuarios-productos.



## Conferenciantes

**Alex Arenas:** Alex Arenas (Barcelona, 1969) es Doctor en Física desde 1996, incorporándose ya



desde un año antes a la docencia en el Departamento de Ingeniería Informática y Matemáticas (DEIM) de la Universitat Rovira i Virgili, donde es permanente desde 1997. En el 2000 realizó una estancia en el grupo de Matemática Aplicada del Prof. Alexandre Chorin, en el Lawrence Berkeley Lab. (LBL) de la University of California, Berkeley. Tras esta visita, comenzó una fructífera colaboración con Berkeley, y desde 2007 es investigador visitante del LBL. Arenas ha publicado más de 160 artículos interdisciplinares en revistas científicas de primera línea como Nature, Nature Physics, PNAS, Physics Reports y Physical Review Letters, recibiendo más de 9000 citas. Es uno de los pocos europeos que ejercen como editor asociado de la más importante publicación científica en física del mundo, el American Physical Society journal, Physical Review, donde está a cargo de la sección de Redes Complejas y Física Interdisciplinar. En 2011 recibió el premio de la James Mc Donnell Foundation por el estudio de sistemas complejos. También ha sido reconocido como ICREA Academia-Institució Catalana de Recerca i Estudis Avançats, un premio catalán que promociona a los científicos más excelentes de Cataluña. Es editor del Journal of Complex Networks, así como de Network Neuroscience. Desde su elección en 2012 es miembro del Comité Directivo de la Complex Systems Society. Es el líder del grupo de investigación ALEPHSYS.

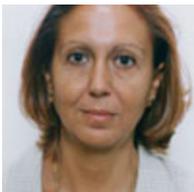
**Rosa María Benito:** Rosa María Benito Zafrilla es Catedrática de Universidad del área de Física



Aplicada en la Universidad Politécnica de Madrid (UPM) desde 1994. Es responsable del Grupo de Investigación interdisciplinar, Grupo de Sistemas Complejos de la UPM. Su investigación se enmarca dentro del estudio de los sistemas dinámicos no lineales, de la teoría del caos y la complejidad. En particular, en los últimos años ha desarrollado varios trabajos sobre la comprensión y caracterización de la estructura y dinámica de diferentes sistemas complejos mediante el uso de la Teoría de Redes Complejas y la

Ciencia de los Datos. Es de destacar un nuevo formalismo propuesto para modelar la topología de redes complejas, los estudios sobre el comportamiento individual y colectivo de los usuarios en redes sociales online como Twitter y los patrones de movilidad humana a través de datos de telefonía móvil. Ha liderado muchos proyectos de investigación tanto a nivel nacional como internacional. Posee más de 140 publicaciones en revistas científicas. Es responsable del Programa de Doctorado en Sistemas Complejos de la UPM, y ha dirigido un total de 12 Tesis Doctorales. Ha sido galardonada por la UPM con el Premios de Excelencia Docente (2009) y el de Innovación Educativa (2010).

**María Luz Congosto:** M<sup>a</sup> Luz Congosto es Doctora en Telemática por la Universidad Carlos III y



Licenciada en Informática por la Universidad Politécnica de Madrid. Desde el año 2008 es investigadora en la Universidad Carlos III especializándose en el análisis de datos sociales, preferentemente en Twitter. Utiliza análisis de redes y visualización para descubrir patrones de comportamiento, propagación de mensajes y caracterización de usuarios. Para sus experimentos ha creado la plataforma t-hoarder que permite monitorizar a largo plazo la evolución de

eventos en Twitter. Con esta herramienta ha construido un barómetro social con las reacciones de los ciudadanos ante escándalos políticos y Metroaverías que analiza la percepción de la calidad del metro de Madrid a través los mensajes de los usuarios este servicio. Del año 2000 al 2008 realizó labores de divulgación tecnológica y colaboró con Fundación Telefónica creando



plataformas de publicaciones digitales. Previamente trabajó durante 20 años en grandes proyectos de innovación en el entorno de las Telecomunicaciones para la operadora Telefónica (TESYS, MORE, INFOVÍA, EOC...). En paralelo fue profesora asociada durante dos años en la Universidad Politécnica de Madrid y cuatro en la Universidad Carlos III I.

**Modesto Escobar:** Modesto Escobar es doctor en sociología por la Universidad Complutense de



Madrid y catedrático de sociología del Departamento de Sociología y Comunicación de la Facultad de Ciencias Sociales de la Universidad de Salamanca. Ha publicado los libros *La identidad social del parado; El análisis gráfico/exploratorio; Autoempleo y actividades empresariales de las mujeres en España; El análisis de segmentación: técnicas y aplicaciones de los árboles de clasificación*, y *Análisis de datos con Stata*. Es autor también, entre otros

artículos, de “Deviation, inequality, polarization: some measures of social diversity”, “Works councils in Spain”, “Redes semánticas en textos periodísticos: propuestas técnicas para su representación” y “La calidad democrática. Una propuesta para su medición por expertos”. Ha impartido cursos de técnicas de investigación en el Centro de Investigaciones Sociológicas, en el Centro de Estudios Avanzados en Ciencias Sociales (Instituto Juan March de Estudios e Investigaciones), en el Centro de Estudios Andaluces, en el Centro de Investigación y Documentación Educativa, en el Instituto de Cooperación Iberoamericana, así como en Universidades españolas como la Autónoma de Madrid, la Pública de Navarra, y en empresas como Pegaso y OTR. Asimismo, ha sido profesor visitante de las Universidades de Warwick, Harvard, Oxford, Federal de Pernambuco, Buenos Aires y Manchester.

**Estrella Gualda:** Estrella Gualda, Doctora en Sociología por la Universidad Complutense de



Madrid, es Profesora Titular de Sociología en la Universidad de Huelva, donde es Directora del Grupo de Investigación “Estudios Sociales E Intervención Social” ([www.eseis.es](http://www.eseis.es)). Es Vicepresidenta de la Asociación Andaluza de Sociología y miembro del Consejo Científico del Research Centre for Spatial and Organisational Dynamics en la Universidade do Algarve (Faro, Portugal, [www.ualg.pt](http://www.ualg.pt)). A lo largo de su carrera ha publicado diversos libros, capítulos de

libros y artículos científicos en el campo de la sociología, exclusión social, migraciones y minorías, estudios transfronterizos, redes sociales, capital social y medios sociales. Actualmente está trabajando en varios estudios de caso a partir de Twitter desde un enfoque de Social Big Data. En este campo, algunas de sus publicaciones recientes son: un artículo sobre “Women and Refugees in Twitter: Rethorics on Abuse, Vulnerability and Violence from a Gender Perspective” (2017, *Journal of Mediterranean Knowledge*, con Gallego y Rebollo); un capítulo sobre la “Polarización de los discursos sobre los refugiados en Twitter: nuevos desafíos para la integración” (2017, con Rebollo); un artículo sobre “Refugee Crisis in Twitter: Diversity of Discourses at a European Crossroads” (2016, *Journal of Spatial and Organizational Dynamics*, con Rebollo) o un capítulo sobre “Spanish General Elections, Microdiscourses Around #20D and Social Mobilisation on Twitter: Reality or Appearance?” (2016, *Media and Metamedia Management*, en Springer). También se encuentra implicada en la Acción COST internacional “Comparative Analysis of Conspiracy Theories in Europe”, en la que participan investigadores de 33 países diferentes (<https://conspiracytheories.eu/>), donde intenta conectar los campos de Social Big Data, Twitter y el Análisis de Redes Sociales.



**Guillermo Jiménez:** Guillermo Jiménez Díaz es Doctor en Informática (2008) por la Universidad Complutense de Madrid (UCM). Su carrera profesional ha transcurrido íntegramente en la Facultad de Informática de la UCM y es miembro del Grupo



de Aplicaciones de Inteligencia Artificial (GAIA), donde ha trabajado como investigador a tiempo completo en las principales líneas de investigación del grupo: enseñanza basada en casos, enseñanza basada en juegos y, las más recientes, sistemas recomendadores y realidad aumentada. Durante los

últimos años sus intereses se han centrado en los sistemas de recomendación, ya sea en la recomendación de objetos de aprendizaje, la recomendación para grupos y tareas de matchmaking y la recomendación de problemas para jueces en línea. Sus objetivos a medio/largo plazo son profundizar en la investigación de la aplicación de técnicas de análisis de redes sociales como medio para mejorar las tareas de recomendación, añadiendo una componente más social a los sistemas recomendadores. Para más información se puede consultar su página web (<http://gaia.fdi.ucm.es/people/guille/>) o contactar con él por correo electrónico ([gjimenez@ucm.es](mailto:gjimenez@ucm.es)).



U N I V E R S I D A D  
**COMPLUTENSE**  
M A D R I D

### **Organización**

**Financiación:** Proyecto de investigación Santander – UCM PR26/16-21B titulado “¿Nos dirigimos hacia una sociedad desintermediada? Un estudio de la participación social mediante análisis de redes, big data y aprendizaje automático”.

**Organizadores principales:** J. Tinguaro Rodríguez, Adrián Riesco, José Manuel Robles.

**Comité de organización local:** Daniel Gómez, Javier Montero, Guillermo Villarino, Karina Rojas, Rafael Caballero, Mirko Antino, Stefano De Marco.