

# Detección de comunidades en redes complejas

Rodrigo Aldecoa García

10 de abril de 2013

El uso de las redes para modelar sistemas complejos es creciente en multitud de ámbitos. Son extremadamente útiles para representar interacciones entre genes, relaciones sociales, intercambio de información en Internet o correlaciones entre precios de acciones bursátiles, por nombrar sólo algunos ejemplos. Analizando la estructura de estas redes, comprendiendo cómo interaccionan sus distintos elementos, podremos entender mejor cómo se comporta el sistema en su conjunto. A menudo, los nodos que conforman estas redes tienden a formar grupos altamente conectados. Esta propiedad es conocida como *estructura de comunidades* y esta tesis doctoral se ha centrado en el problema de cómo mejorar su detección y caracterización. Como primer objetivo de este trabajo, se encuentra la generación de métodos eficientes que permitan caracterizar las comunidades de una red y comprender su estructura. Segundo, pretendemos plantear una serie de pruebas donde testar dichos métodos. Por último, sugeriremos una medida estadística que pretende ser capaz de evaluar correctamente la calidad de la estructura de comunidades de una red. Para llevar a cabo dichos objetivos, en primer lugar, se generan una serie de algoritmos capaces de transformar una red en un árbol jerárquico y, a partir de ahí, determinar las comunidades que aparecen en ella. Por otro lado, se ha diseñado un nuevo tipo de *benchmarks* para testar estos y otros algoritmos de detección de comunidades de forma eficiente. Por último, y como parte más importante de este trabajo, se demuestra que la estructura de comunidades de una red puede ser correctamente evaluada utilizando una medida basada en una distribución hipergeométrica. Por tanto, la maximización de este índice, llamado *Surprise*, aparece como la estrategia idónea para obtener la partición en comunidades óptima de una red. *Surprise* ha mostrado un comportamiento excelente en todos los casos analizados, superando cualitativamente a cualquier otro método anterior. De esta manera, aparece como la mejor medida propuesta para este fin y los datos sugieren que podría ser una estrategia óptima para determinar la calidad de la estructura de comunidades en redes complejas.