

Detección de comunidades en redes complejas

Rodrigo Aldecoa García

May 31, 2013

L'ús de les xarxes per a modelar sistemes complexos és creixent en multitud d'àmbits. Són extremadament útils per a representar interaccions entre gens, relacions socials, intercanvi d'informació a Internet o correlacions entre preus d'accions borsàries, per anomenar només alguns exemples. Analitzant l'estructura d'aquestes xarxes, comprenent com interaccionen els diferents elements, podrem entendre millor com es comporta el sistema en el seu conjunt. Moltes vegades, els nodes que conformen aquestes xarxes tendeixen a formar grups altament connectats. Aquesta propietat és coneguda com a *estructura de comunitats* i aquesta tesi doctoral s'ha centrat en el problema de com millorar la seva detecció i caracterització. Com a primer objectiu d'aquest treball, tractarem de generar mètodes eficients que permeteixen caracteritzar les comunitats d'una xarxa i comprendre la seua estructura. Segon, pretenem plantejar una sèrie de proves on testar aquests mètodes. Finalment, suggerirem una mesura estadística que pretén ser capaç d'avaluar correctament la qualitat de l'estructura de comunitats d'una xarxa. Per dur a terme aquests objectius, en primer lloc, es generen una sèrie d'algoritmes capaços de transformar una xarxa en un arbre jeràrquic i, a partir d'aquí, determinar les comunitats que hi apareixen. D'altra banda, s'ha dissenyat un nou tipus de *benchmarks* per testar aquests i altres algoritmes de detecció de comunitats de forma eficient. Finalment, i com a part més important d'aquest treball, es demostra que l'estructura de comunitats d'una xarxa pot ser correctament avaluada utilitzant una mesura basada en una distribució hipergeomètrica. Per tant, la maximització d'aquest índex, anomenat *Surprise*, apareix com l'estratègia idònia per obtenir la partició en comunitats òptima d'una xarxa. *Surprise* ha mostrat un comportament excel·lent en tots els casos analitzats, superant qualitativament a qualsevol altre mètode anterior. D'aquesta manera, apareix com la millor mesura proposada per a aquest fi i les dades suggereixen que podria ser una estratègia òptima per determinar la qualitat de l'estructura de comunitats en xarxes complexes.