
Respondre preguntes de recerca analitzant xarxes

PID_00257255

David Laniado
Noèlia Viles Cuadros

Temps mínim de dedicació recomanat: 2 hores



David Laniado

Noèlia Viles Cuadros

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Montserrat Garcia Alsina (2018)

Índex

Introducció	5
Objectius	6
1. Mètriques per a l'anàlisi de mitjans socials	7
1.1. Mètriques per caracteritzar un node	7
1.2. Estructura de les xarxes: caracterització global de la xarxa	8
1.2.1. Grandària	8
1.2.2. Coeficient de clusterització o transitivitat	9
1.2.3. Patrons de barreja	10
1.3. Mètriques de centralitat: influència a la xarxa	14
1.3.1. Centralitat de proximitat (<i>closeness centrality</i>)	15
1.3.2. Centralitat de intermediació (<i>betweenness centrality</i>)	15
1.3.3. Comparació entre mètriques de centralitat	16
1.3.4. Pagerank	17
1.3.5. Centralitat de nucli (k-core)	19
2. Calcular mètriques amb Gephi	21
Bibliografia	23

Introducció

Fins ara hem estudiat alguns dels aspectes que descriuen una xarxa i com la visualitzem. En aquest mòdul, ens endinsarem en el tema de les mètriques de les xarxes socials. Les mètriques ens faciliten l'anàlisi de les xarxes socials per interpretar el que està passant, i com són les relacions entre els nodes que componen una xarxa.

En les tres primeres parts del mòdul treballarem les mètriques, tant les dels nodes, com les del conjunt de les xarxes. En l'última part, veurem com calcular les mètriques amb l'eina de representació i visualització de xarxes Gephi que hem estudiat en el mòdul anterior.

Objectius

Els objectius que ha d'assolir l'estudiant després de treballar els materials d'aquest mòdul didàctic són els següents:

- 1.** Familiaritzar-se amb els conceptes d'anàlisi de xarxes socials.
- 2.** Saber plantejar preguntes i cercar les respostes mitjançant l'anàlisi d'una xarxa.
- 3.** Seguir un tema en una xarxa.
- 4.** Analitzar el debat.
- 5.** Entendre i gestionar les interrelacions i el valor dels nodes en entorns de *social media*.
- 6.** Reconèixer i avaluar de forma curiosa la rellevància i significació de la informació.
- 7.** Saber recuperar dades publicades a les xarxes socials.
- 8.** Saber comunicar les conclusions de manera clara i sense ambigüitats.
- 9.** Comprendre i emetre informes professionals i específics en l'àmbit del *social media*.
- 10.** Desenvolupar una estratègia de curació de continguts i gestió de dades en el *social media*.
- 11.** Aplicar l'ús de les TIC.

1. Mètriques per a l'anàlisi de mitjans socials

En aquest apartat treballarem algunes de les mètriques que fem servir per analitzar les xarxes socials. Aquests tenen els mateixos nodes i la relació entre ells, per tal de caracteritzar una xarxa.

Estudiarem, en primer lloc, les mètriques que caracteritzen un node, prenent com a base la distància entre nodes, els camins i la longitud d'un camí. En segon lloc, treballarem l'estructura de les xarxes, amb les quals podrem caracteritzar-les de manera global. Per últim, estudiarem les mètriques de centralitat mitjançant les quals veurem les influències que hi ha a les xarxes.

1.1. Mètriques per caracteritzar un node

Un node el podem descriure també a partir de les connexions que té amb d'altres nodes, que ens indiquen les relacions que s'hi estableixen. Això ens permet conèixer el paper de cada node. A tal efecte, treballarem tres conceptes: camins, longitud d'un camí i distància entre nodes.

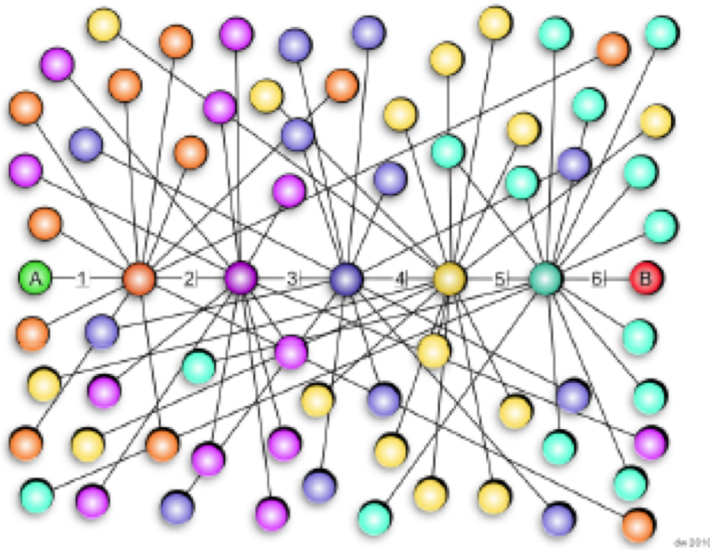
a) **Camins.** Un camí entre dos nodes d'un graf és un recorregut que els uneix mitjançant d'altres nodes.

b) **Longitud d'un camí** és el número de connexions que s'han de travessar per arribar des del node d'origen al node de destí. Són el conjunt de camins mínims. Això permet generalitzar el concepte de proximitat i de centralitat (Diaz-Guilera, 2014). Aquestes mètriques de centralitat, les veurem més endavant.

c) **Distància entre nodes.** Es defineix distància entre dos nodes a la longitud del camí més curt que uneix aquests nodes. En altres paraules, per distància s'entén sempre la distància mínima entre els dos nodes.

Aquesta distància reflecteix les relacions que es poden establir en una xarxa, partint del fet que les relacions socials són transitives. El psicòleg social Milgram va fer un experiment per saber quanta gent feia falta perquè dues persones seleccionades a l'atzar es connectessin en un país de 250 milions de persones. Va prendre una mostra de 296 persones i va concloure que només feia falta una mitjana de 5,2 persones per a connectar dues persones (Mirallbell, 2008). Posteriorment, s'han realitzat altres experiments i les xifres oscil·len entre 5 i 10 persones (figura 1).

Figura 1. Experiment de Milgram



Font: Wikipedia

Aquest experiment ha servit per treballar el concepte de «Món petit». Aquest concepte s'empra per referir-se a la distància mitjana relativament petita entre nodes d'una xarxa social. Permet saber en quin grau cada membre d'un grup pot conèixer un membre d'un altre grup dins d'una xarxa.

1.2. Estructura de les xarxes: caracterització global de la xarxa

La xarxa la podem caracteritzar a partir dels aspectes següents:

- grandària,
- coeficient de clusterització,
- patrons de barreja i
- enllaços febles.

1.2.1. Grandària

La grandària de la xarxa està basada en les distàncies entre nodes de manera estadística. Per això utilitzem dos conceptes: distància mitjana i diàmetre.

a) Distància mitjana. La distància mitjana d'una xarxa és la mitjana de les distàncies entre totes les parelles de nodes a la xarxa.

b) Diàmetre. Es diu diàmetre d'una xarxa la distància més llarga entre dos nodes. Representa el número màxim de passos que s'han de recórrer per arribar des d'un node a un altre a la xarxa.

1.2.2. Coeficient de clusterització o transitivitat

El coeficient d'aglomeració (o de clusterització) d'una xarxa és una mètrica que mesura la tendència dels seus nodes a agrupar-se.

Aquest coeficient es coneix també com a «transitivitat» d'una xarxa. De fet, indica la proporció de casos en què, donades una connexió A-B i una connexió B-C, existeix també una connexió A-C (és a dir, la proporció de relacions transitives).

La definició es basa en el concepte de **triplets de nodes**, que són grups de tres nodes connectats entre ells per dues o tres arestes:

- Si estan connectats per dues arestes, el triplet s'anomena triplet obert, en què una de les tres arestes possibles entre els tres nodes no està connectada.
- Si els nodes estan connectats per tres arestes, el triplet s'anomena triplet tancat, en què totes les tres arestes possibles entre els tres nodes estan connectades.

El coeficient de clusterització, doncs, es calcula com la proporció de triplets tancats respecte a tots els triplets (oberts o tancats). En altres paraules, és la proporció de triangles tancats respecte a tots els triangles possibles.

Un alt *clustering* és una propietat distintiva de les xarxes socials respecte a diferents tipus de xarxes, per exemple biològiques o tecnològiques (Newmann i Park, 2003). També és una de les característiques típiques de les xarxes «de món petit», que fa que hi hagi múltiples camins per a connectar dos nodes, i una alta densitat de connexions entre nodes.

A part d'aquest coeficient, els clústers també es poden mesurar mitjançant tres mètriques més:

- coeficient de clusterització local
- mitjana del coeficient de clusterització local
- coeficient de clusterització global

Vegem aquestes mètriques.

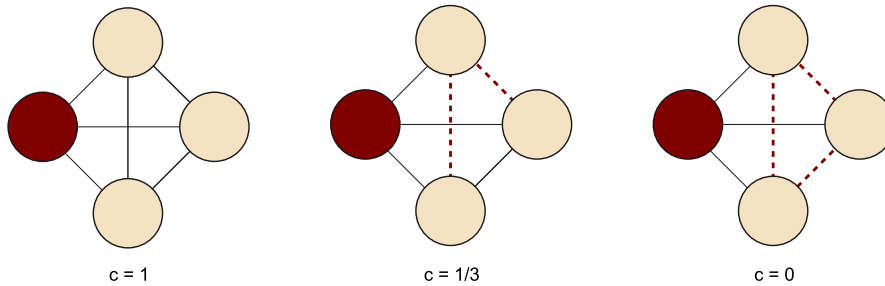
1) Coeficient de clusterització local

La clusterització local d'un node donat és el percentatge de triplets tancats en l'entorn del node. Representa la tendència dels nodes, connectats amb un node donat, a formar una clíca, en què tots estan connectats amb tots. Recordem que una clíca és un grup de nodes on cada node està connectat amb tots els altres.

Veiem un exemple de coeficient de clusterització local al voltant del node blau (figura 2), on observem que els seus tres veïns poden tenir els següents tipus de connexió:

- Tots connectats entre ells (coeficient de clusterització = 1).
- No tenir cap connexió entre ells (coeficient de clusterització = 0).
- Tenir-ne alguna (si en tenen una de les tres possibles, tenim coeficient de clusterització = 1/3).

Figura 2. Exemple de coeficient de clusterització local



Font: Wikipedia

2) Mitjana del coeficient de clusterització local

Podem caracteritzar una xarxa sencera a partir del coeficient de clusterització local. Per fer això, es pot fer servir la mitjana dels valors de tots els nodes de la xarxa. Aquesta mètrica es basa en els valors individuals. És una alternativa a una altra mesura, el coeficient de clusterització global.

3) Coeficient de clusterització global

El coeficient de clusterització global es calcula a nivell de tota la xarxa, sense fer mitjanes. És el percentatge de triplets tancats sobre tots els triplets existents a la xarxa.

1.2.3. Patrons de barreja

En una xarxa, els nodes tenen diferents característiques, com poden ser, per exemple, els atributs demogràfics dels usuaris corresponents (gènere, estat, etc.). En molts casos pot ser rellevant investigar quines tendències de relació observem entre els nodes per conèixer els patrons de connexió de les persones. És a dir, podem veure si interactuen amb nodes semblants, diferents o si hi ha absència de tendència.

La mètrica que mesura la preferència d'un node d'una xarxa per unir-se amb altres que li són similars s'anomena assortativitat.

- Si els nodes tendeixen a interactuar més amb nodes semblants, es parla de «homofília» (des del grec antic, preferència pel semblant), i correspon amb el coeficient de barreja neutral.
- Si els nodes tendeixen a interactuar més amb nodes diferents, es parla de «heterofília» (preferència pel diferent). A aquest patró li correspon el coeficient de barreja negatiu.
- Els nodes que no mostren cap tendència, tenen un coeficient de barreja neutral.

Vegem exemples d'homofília, heterofília, i la mètrica d'assortativitat, i els respectius coeficients.

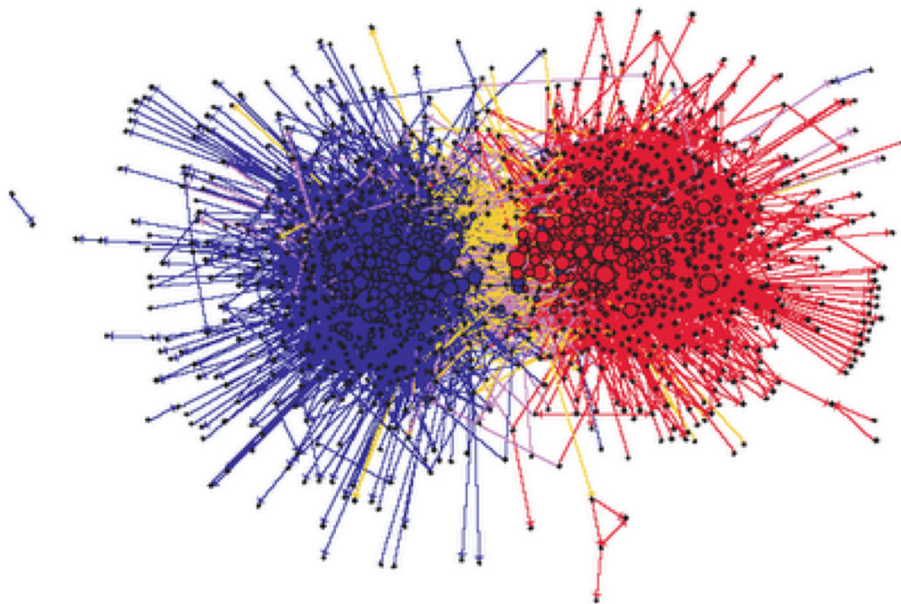
1) Exemples d'homofília i heterofília a les xarxes

Vegem imatges que ensenyen patrons de barreja segons orientació política als Estats Units. Trobem un exemple de patró d'homofília, un altre patró de neutralitat, i, per últim un exemple de patró d'heterofília.

a) Homofília (coeficient de barreja positiu)

A la següent imatge veiem l'orientació política a la blogosfera de Estats Units (figura 3). Observem preferència per part dels usuaris a connectar-se amb usuaris del mateix partit (coeficient de barreja positiu) (Adamic i Glance, 2005).

Figura 3. Xarxa d'enllaços entre blogs d'orientació republicana (vermells) i demòcrata (blau)

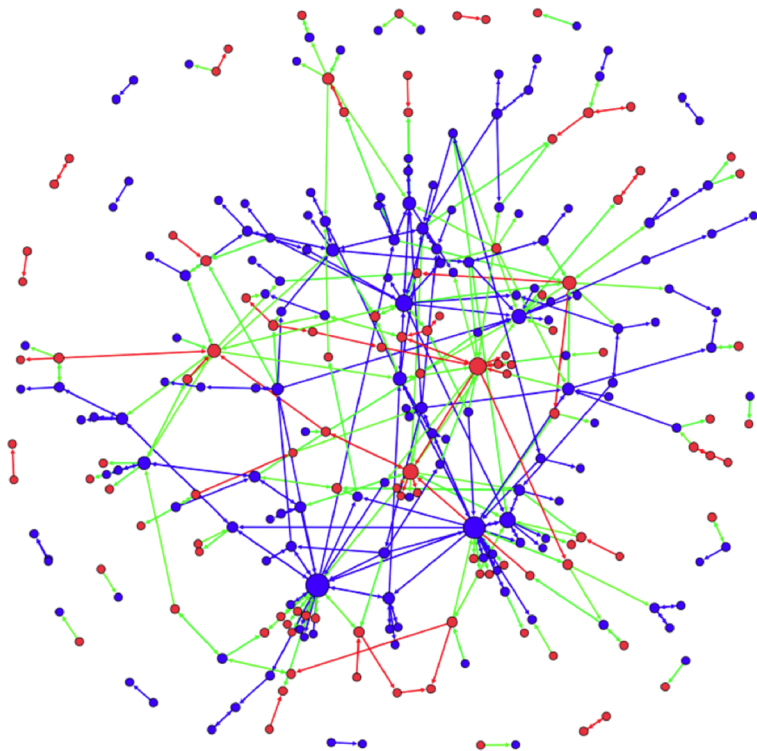


Font: Adamic i Glance (2005)

b) Patró neutral (coeficient de barreja neutral)

L'exemple següent mostra la relació i el suport dels editors de Wikipedia amb els partits Demòcrata (nodes blaus) i Republicà (nodes vermells). Observem preferència neutral entre editors de la Wikipedia afiliats als partits Demòcrata i Republicà. Els enllaços blaus representen respostes entre demòcrates, els vermells respostes entre republicans, i els verds respostes entre persones de partits diferents. La xarxa no té preferència ni per a la interacció dins d'un mateix partit, ni entre partits (coeficient de barreja neutral) (Neff i altres, 2013).

Figura 4. Xarxa de respostes entre editors de Wikipedia que donen suport explícitament el partit Republicà (nodes vermells) i el partit Demòcrata (nodes blaus) a la xarxa

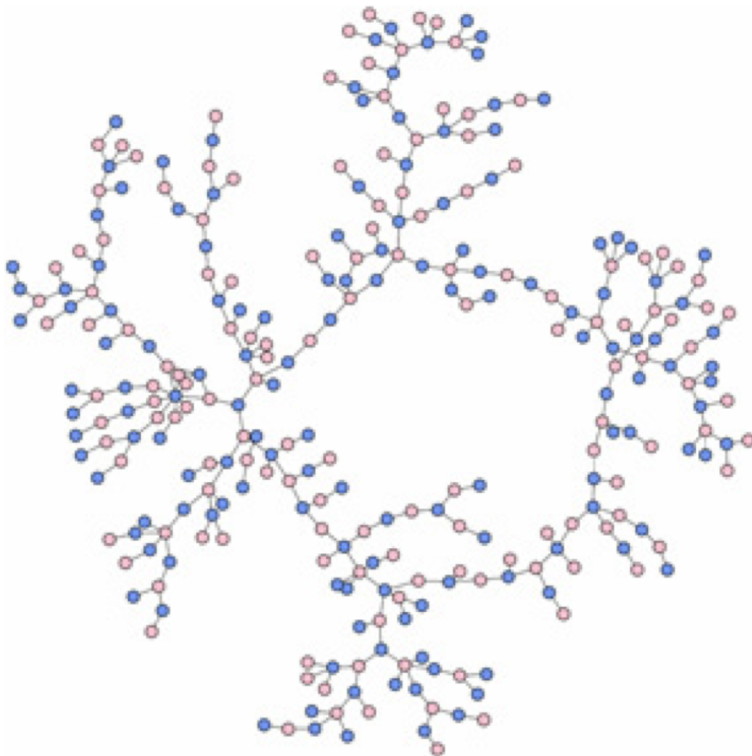


Font: Neff i altres (2013)

c) Heterofília (coeficient de barreja negatiu)

En aquest tercer exemple (figura 5) veiem una xarxa que mostra les relacions romàntiques i sexuals entre adolescents dels dos gèneres. El color representa el gènere, els nois són representats en blau, i les noies en rosa. La xarxa mostra evidència de preferència per a interaccions amb l'altre gènere (coeficient de barreja negatiu). Per tant, és una xarxa amb heterofília, ja que hi ha preferència per la interacció amb nodes diferents (persones de l'altre gènere) (Bearman, 2004).

Figura 5. Xarxa de relacions romàntiques i sexuals entre adolescents, preferència per a interaccions amb l'altre gènere (coeficient de barreja negatiu)



Font: Bearman (2004)

2) Assortativitat

Per a donar una mesura matemàtica dels patrons de barreja, es pot fer servir l'assortativitat. Així tenim xarxes assortatives i xarxes dissortatives.

- Una **xarxa és assortativa** quan hi ha evidència estadística d'una preferència per les interaccions entre nodes similars.
- Una **xarxa és dissortativa** quan, al contrari, hi ha preferència per les interaccions entre nodes diferents.

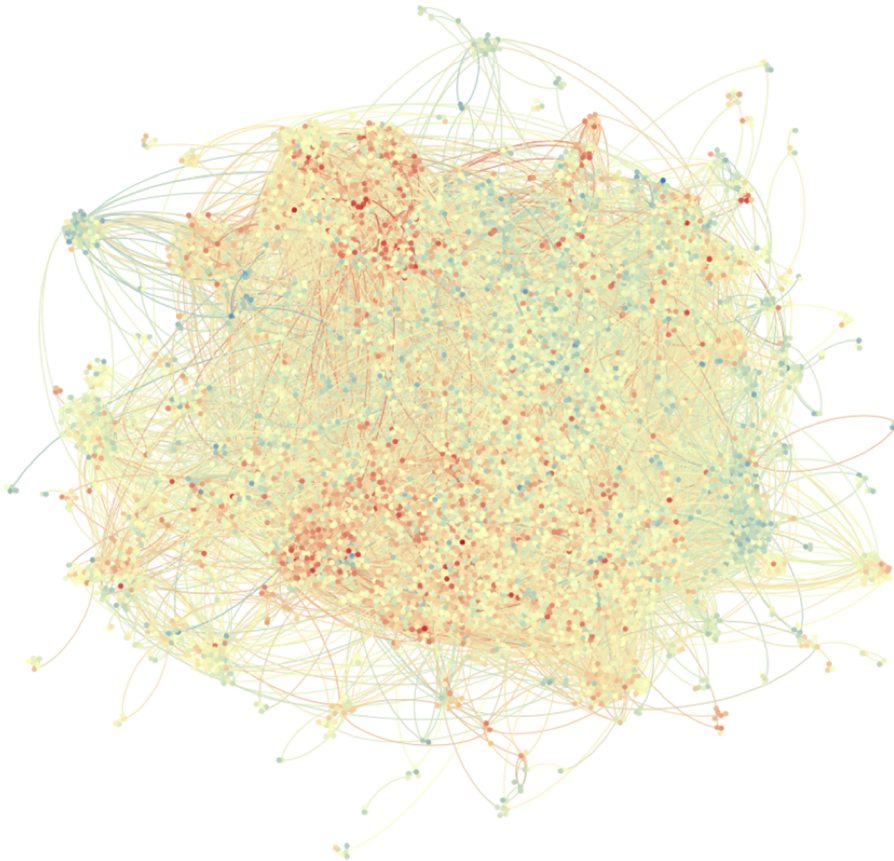
Matemàticament, l'assortativitat es defineix com la correlació de Pearson entre els valors dels dos nodes que constitueixen cada enllaç, respecte a una variable donada (Newman, 2002).

Típicament es pot calcular l'assortativitat respecte al grau. En una xarxa assortativa, els nodes amb moltes connexions tendeixen a connectar-se preferentment amb altres nodes que tenen moltes connexions. Contràriament, els nodes amb poques connexions tendeixen a connectar-se preferentment entre ells.

En la següent xarxa (figura 6), hi veiem un exemple d'assortativitat segons la tendència a expressar ràbia en les discussions de la Wikipedia. El color dels nodes representa el nivell de ràbia contingut en els seus missatges. Els colors més càlids representen paraules que expressen ràbia. Es pot observar que els

nodes del mateix color tendeixen a tenir més connexions entre ells (coeficient de barreja positiu). Veiem que els editors que fan servir més paraules associades amb ràbia tendeixen a interactuar més entre ells (Iosub i altres, 2014).

Figura 6. Assortativitat emocional en la xarxa de respostes a les pàgines de discussió de Wikipedia (coeficient de barreja positiu)



Font: Iosub i altres (2014)

1.3. Mètriques de centralitat: influència a la xarxa

Sabem que en una xarxa social les persones s'influencien, la informació i les idees es difonen i els comportaments es contagien. Però, com podem identificar els nodes que tenen més influència? O en altres paraules, com podem quantificar la influència de diferents nodes?

És clar que quan estudiem una xarxa social com un graf, estem estudiant un objecte matemàtic que representa un model molt simplificat de la realitat. Igualment, la topologia de la xarxa ens pot donar moltes pistes sobre el potencial d'influència d'actors diferents, segons la seva posició respecte als altres nodes.

Hi ha moltes maneres de ser influents, i per tant hi ha moltes mètriques que ofereixen diferents maneres de mesurar la influència des de diferents perspectives.

La mètrica de centralitat més senzilla és el grau, com ja hem vist al mòdul «Visualitzar i entendre interaccions a les xarxes socials», i indica el nombre de connexions d'un node donat.

A continuació veurem altres mètriques que ens poden ajudar a quantificar diferents aspectes de la influència en una xarxa.

1.3.1. Centralitat de proximitat (*closeness centrality*)

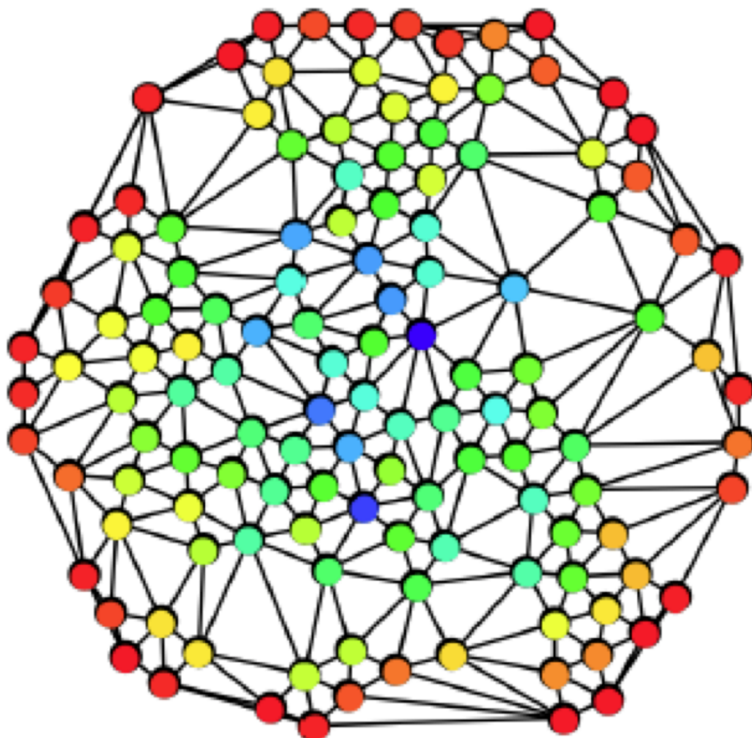
La centralitat de proximitat expressa la centralitat d'un node, és a dir, com és la seva proximitat als altres nodes de la xarxa. Es calcula com el recíproc de la suma de les distàncies entre el node i tots els altres nodes. Es multiplica aquest valor pel número de nodes, com a normalització per tal de permetre la comparació de xarxes diferents, de tal manera que la mètrica representa el recíproc de la distància mitjana.

Aquesta mètrica quantifica la centralitat d'un node com la seva possibilitat d'arribar en pocs passos a qualsevol altre node de la xarxa.

1.3.2. Centralitat de intermediació (*betweenness centrality*)

La centralitat d'intermediació mesura el potencial d'un node per connectar altres nodes entre ells (figura 7). És una altra mètrica basada en la topologia de la xarxa i en els camins més curts entre nodes, però des d'una perspectiva molt diferent.

Figura 7. Els colors més freds corresponen a la centralitat d'intermediació més alta



Un node es considera més central segons el número de camins més curts entre altres nodes que passen per ell. És a dir, mesura quantes vegades el node es troba en el camí més curt entre una parella de nodes.

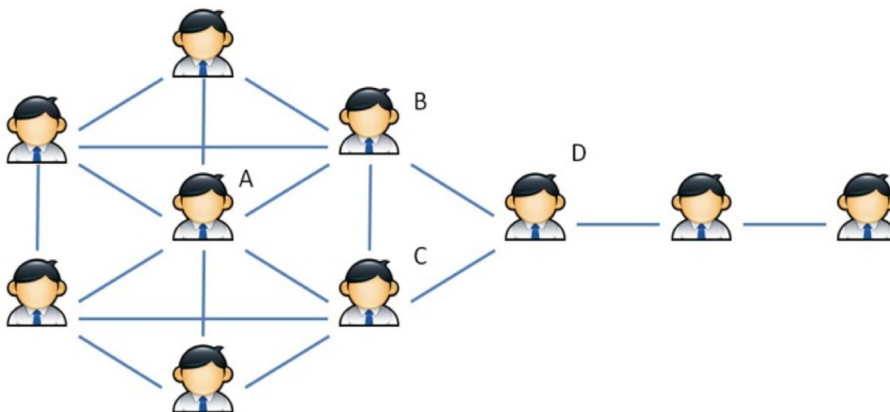
Per a calcular la *betweenness* dels nodes d'una xarxa, s'identifiquen tots els camins més curts entre cada parella de nodes a la xarxa, i s'assigna un punt a cada node travessat per cada camí. En cas d'haver-hi més camins de la mateixa longitud (mínima), es divideix la contribució d'aquella parella de nodes entre els diferents camins.

La centralitat d'intermediació pot típicament donar resultats força diferents de les altres mètriques. Això passa per dues raons: en primer lloc, es basa en un criteri diferent, i en segon lloc, no valora les connexions amb altres nodes en general, sinó que valora especialment aquelles connexions que connecten comunitats diferents, o parts de la xarxa que d'altra manera es quedarien desconnectades o llunyanes entre elles.

1.3.3. Comparació entre mètriques de centralitat

En aquesta figura es pot veure una xarxa molt senzilla que farem servir com exemple per a veure la diferència entre les diferents mètriques. Quin node creieu que és el més central?

Figura 8. Exemple de xarxa



Font: cedida per Yana Volkovich

Lògicament, no hi ha una resposta unívoca. Segons el criteri que escollim, podrà resultar més central un node o un altre. Abans de continuar llegint, us convidem a reflexionar sobre quin node és el més central segons les mètriques de grau, intermediació i proximitat.

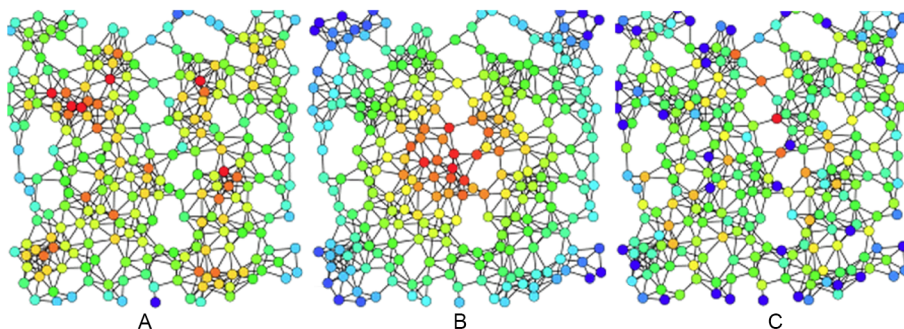
a) **Grau:** per calcular el grau hem de comptar el número de connexions de cada node, ja que el valor més alt el té el node A, que està connectat amb uns altres sis nodes.

b) Proximitat: en aquest cas hem de buscar el node amb les distàncies més curtes per a assolir els altres nodes. Podem observar que tot i que el node A tingui moltes connexions, es troba a diversos passos de separació de tota la part dreta de la xarxa. Si sumem les distàncies cap a tots els nodes, en trobem un total de 15; els nodes B i C en canvi es troben més al centre respecte a les diverses parts de la xarxa, i de fet tenen un valor total de 14, així que la seva centralitat de proximitat (el recíproc) és major.

c) Intermediació: per a trobar qui és més influent segons la centralitat d'intermediació, hem de mirar quin node es troba en posició especial per a connectar diferents parts de la xarxa, i en aquest sentit trobem que el node D té la posició més rellevant, ja que és l'únic que connecta els dos nodes de la dreta amb la resta de la xarxa, i per conseqüència molts camins més curts passen a través d'ell. B i C també es troben en posició estratègica en aquest sentit, però com que són dos, es reparteixen d'alguna manera el flux, així que tenen un valor de centralitat d'intermediació més baix, ja que cap dels dos és imprescindible per a connectar les dues parts de la xarxa.

A continuació (figura 9) veurem una altra figura que ens ajuda a veure la diferència entre les mètriques que hem vist fins aquí. Els colors més càlids representen els nodes més centrals, segons diferents mètriques. De nou, podeu jugar a endevinar quina mètrica correspon a cada figura, i comprovar les vostres respostes amb el títol de la figura.

Figura 9. Els colors més càlids corresponen a centralitat més alta, segons diferents mètriques. A: *degree*; B: *closeness*; C: *betweenness*



Font: Wikipedia

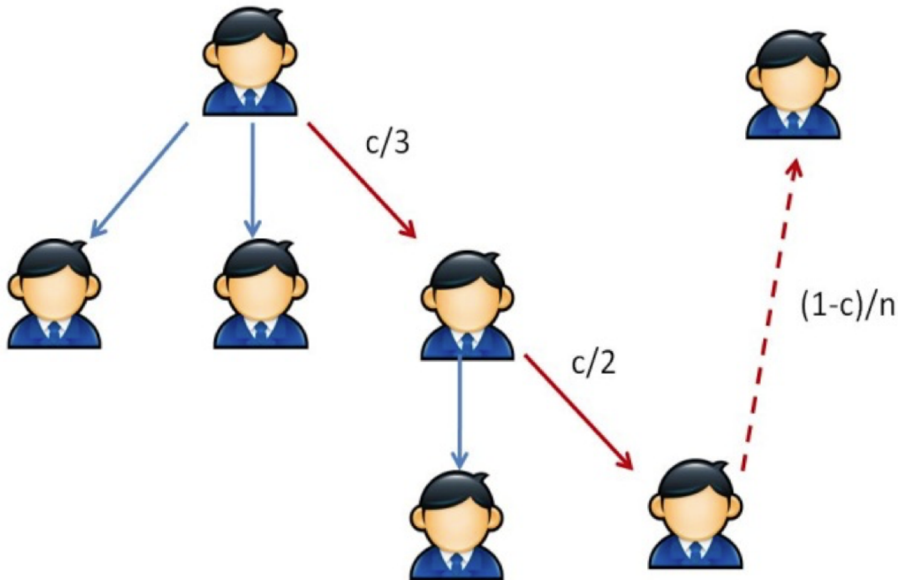
1.3.4. Pagerank

El pagerank és una mètrica de rellevància que va ser introduït pels fundadors de Google, per a identificar de manera automàtica els llocs més rellevants de la xarxa d'hipervincles a la web. El pagerank es fa servir típicament per xarxes dirigides (Page i altres, 1999).

A la figura 10, hi veiem un exemple que ensenya el funcionament de l'algoritme PageRank. Observem que des de cada node, amb probabilitat c es segueix un dels enllaços sortints (amb la mateixa probabilitat). És a dir, que la

probabilitat de seguir un enllaç concret és c/n , on n és el número d'enllaços sortints, mentre que amb la probabilitat $c-1$ es salta a un altre node de la xarxa escollit de manera aleatòria.

Figura 10. Exemple que ensenya el funcionament de l'algoritme PageRank



El PageRank és diferent del grau d'entrada. Mentre que el grau d'entrada d'un node representa el número de connexions entrants des d'altres nodes, però sense considerar la importància d'aquests nodes, el pagerank obvia aquest problema definint el valor de cada node en funció dels valors dels nodes des dels quals rep connexions. Com que el valor de cada node depèn recursivament dels valors dels altres nodes, es tracta d'un algoritme iteratiu.

Matemàticament, el pagerank representa la probabilitat de trobar-se un node donat, seguint recorreguts aleatoris en una xarxa dirigida. El model simula el procés en què es comença des d'un node aleatori i es passa d'un node a un altre agafant un camí aleatori entre els possibles a cada pas. També, en cada pas, hi ha la possibilitat de saltar a qualsevol altre node de la xarxa, escollit de manera aleatòria (amb una probabilitat $1 - c$, usualment al voltant del 15%).

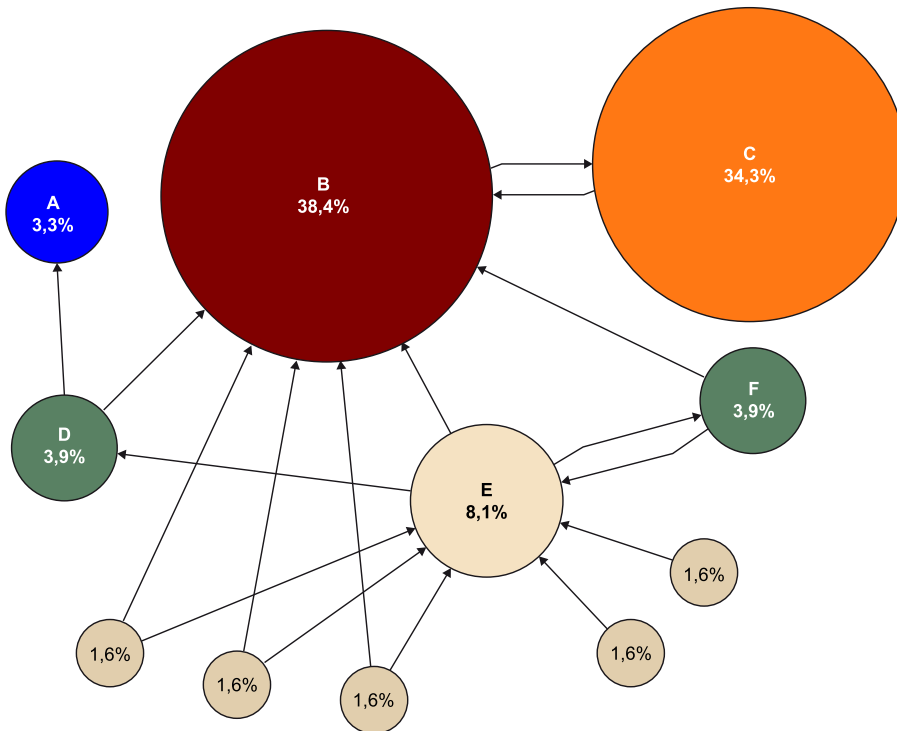
La fórmula per calcular el pagerank d'un node és la següent:

$$PR^*(i) = c \sum_{j \rightarrow i} \frac{1}{d_j^*} + PR^*(j) + \frac{1-c}{N^*} \quad (1)$$

on:

- $PR^*(x)$ és el PageRank del node x .
- d_j^* és el grau de sortida del node j .
- N^* és el número de nodes.

Figura 11. Exemple que ensenya els valors donats per l'algoritme PageRank



1.3.5. Centralitat de nucli (k-core)

Finalment, considerarem la centralitat de nucli, que es basa en la descomposició d'una xarxa en capes cada cop més centrals (k-shells).

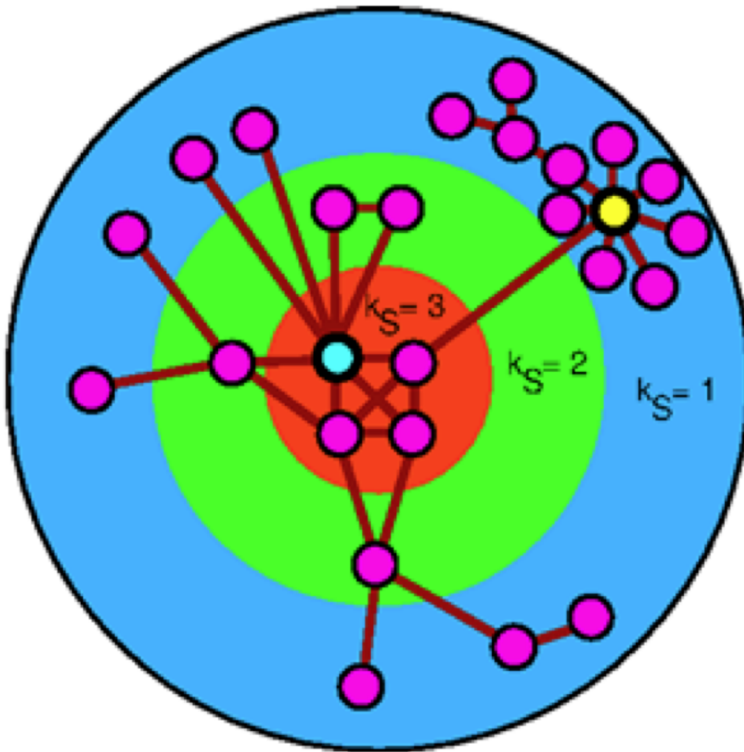
Aquesta mètrica és molt útil per a identificar els nodes que estan en el nucli de la xarxa, descartant *hubs* aïllats que tenen moltes connexions amb nodes pocs rellevants, però no amb els nodes centrals. Per exemple, és possible filtrar *spammers* o nodes que podrien aparèixer com a centrals, però en realitat només ho són gràcies a comptes falsos, etc.

El k-core de nivell k es defineix com el conjunt dels nodes en què cada node té com a mínim k connexions amb altres nodes del conjunt. Un node té k-índex k (o cau al k-shell de nivell k) quan és part del k-core de nivell k, però no del de nivell k + 1.

Per a identificar els k-shells, es comença per la capa més perifèrica de la xarxa, i es treuen tots els nodes que tenen només una connexió, o que tenen només una connexió un cop trets tots els que tenen només una connexió, amb un procés iteratiu. Així, s'identifiquen tots els nodes amb k-índex = 1. Després es fa el mateix per a identificar els nodes de k-índex = 2, i així successivament fins a arribar a la capa més interna, que representa el k-core màxim.

Podem veure un exemple en la figura 12, on veiem que el node groc té un alt grau, però totes les seves connexions són amb nodes de k -índex =1, i té només una connexió amb el nucli de la xarxa, és a dir, que també té k -índex = 1 (Kitsak i altres, 2010).

Figura 12. Exemple de descomposició d'un graf en k-shells

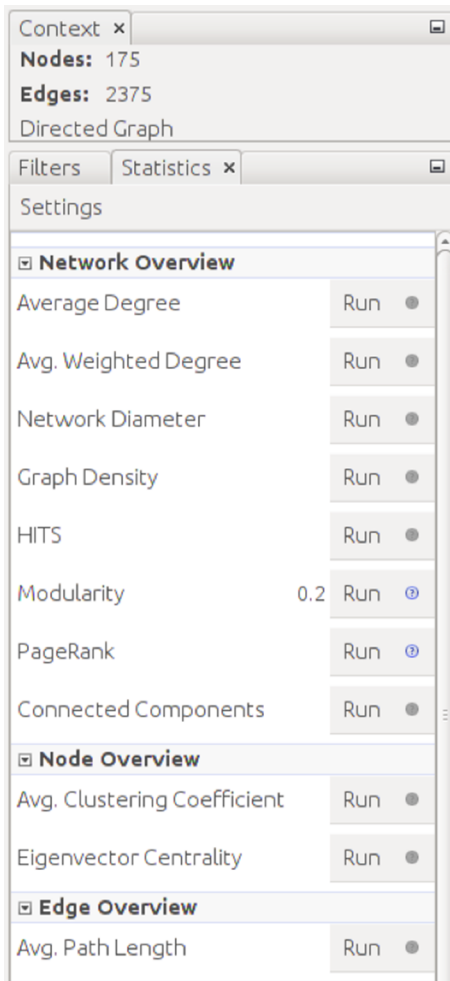


Font: Kitsak i altres (2010)

2. Calcular mètriques amb Gephi

Gephi ofereix la possibilitat de calcular diferents mètriques, tant de xarxa com de nodes:

Figura 13. Apartat d'estadístiques per a calcular mètriques amb el programa Gephi

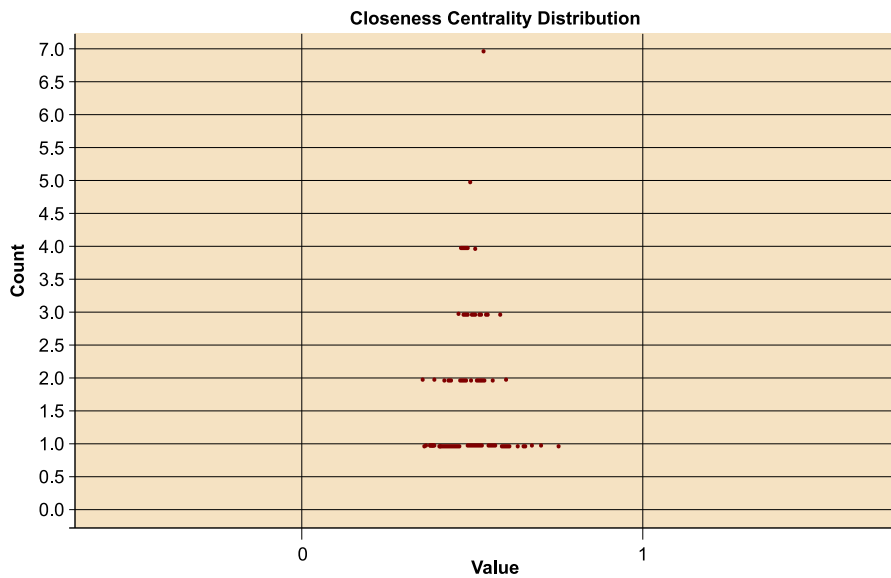


Font: Kitsak i altres (2010)

A l'apartat «Statistics» hi ha un llistat de mètriques que es poden executar, com mostra la figura 13.

Quan s'executa alguna mètrica basada en les distàncies entre nodes (com per exemple: Network Diameter, Avg. Path Length), Gephi calcula automàticament totes les altres mètriques basades en les distàncies entre nodes. És a dir, Gephi no es limita només a aquestes mètriques «globals» que caracteritzen la xarxa sencera, sinó que també calcula mètriques de centralitat basades en la distància entre nodes, incloent-hi la centralitat de proximitat (*closeness*) i d'intermediació (*betweenness*).

Figura 14. Distribució dels valors de *closeness centrality* dels nodes obtinguts calculant el diàmetre de la xarxa



D'aquesta manera, una vegada calculada una mètrica relacionada amb distàncies, tenim també aquestes mètriques de centralitat disponibles, tant al «Data Laboratory», com a la visualització de la xarxa. Aquí podem fer servir aquestes mètriques com a atributs de tipus «Ranking» per a establir la mida dels nodes i de les etiquetes etc., tal com havíem vist a l'assignatura Visualitzar i entendre interaccions a les xarxes socials amb el grau.

Bibliografia

Adamic, L. i Glance, N. (2005). The political blogosphere and the 2004 U.S. election: divided they blog. *Proceedings of the 3rd international workshop on Link discovery*, 36-43.

Barabási, Albert-László i Albert, Réka (15 d'octubre de 1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512. arXiv:cond-mat/9910332. Bibcode 1999Sci...286..509B. doi:10.1126/science.286.5439.509. MR 2091634. PMID 10521342.

Bashky, E. (2012). *Rethinking Information Diversity in Networks*. <<http://www.facebook.com/notes/facebook-data-team/rethinking-information-diversity-in-networks/10150503499618859>>

Bearman, P. S., Moody, J., i Stovel, K. (2004). Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology*, 4(110), 44-91.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., i Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 20, 469-493.

Erdos, P. i Rényi, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 5, 17-61.

Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. P., i Eguiluz, V. M. (2012). Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. *PLoS ONE*, 7(1), e29358.

Granovetter, M. S. (1977). The strength of weak ties. *Social networks*, 347-367.

Iosub, D., Laniado, D., Castillo, C., Morell, M. F., i Kaltenbrunner, A. (2014). Emotions under discussion: Gender, status and communication in online collaboration. *PLoS one*, 9(8), e104880.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., i Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature physics*, 6(11), 888.

Laniado, D., Volkovich, Y., Kappler, K., i Kaltenbrunner, A. (2016). Gender homophily in online dyadic and triadic relationships. *EPJ Data Science*, 5(1), 19.

Laniado, D., Volkovich, Y., Scellato, S., Mascolo, C., i Kaltenbrunner, A. (2017). The Impact of Geographic Distance on Online Social Interactions. *Information Systems Frontiers*, 1-16.

Leskovec, J., Kleinberg, J., i Faloutsos, C. (2017). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2.

Milgram, S. (1967). The small world problem. *Psychology today*, 2(1), 60-67.

Neff, J., Laniado, D., Kappler, K., Volkovich, Y., Aragón, P., i Kaltenbrunner, A. (2013). *Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia*. arXiv preprint arXiv:1210.6883.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2), 026126.

Newman, M. E. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 167-256.

Newman, M. E., i Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(3), 036122.

Page, L., Brin, S., Motwani, R., i Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. *Stanford InfoLab*.

Ruthven, A. (1995, 4 d'abril). Kevin Bacon is the Center of the Universe. rec.arts.movies. Google groups. Retrieved 2009-07-19.

Watts, D. i Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442. Bibcode 1998Natur.393..440W. doi:10.1038/30918. PMID 9623998.

