
Visualitzar i entendre interaccions a les xarxes socials

PID_00257253

David Laniado
Noèlia Viles Cuadros

Temps mínim de dedicació recomanat: 5 hores



David Laniado

Noèlia Viles Cuadros

La revisió d'aquest recurs d'aprenentatge UOC ha estat coordinada per la professora: Montserrat Garcia Alsina (2018)

Índex

Introducció	5
Objectius	6
1. Introducció a l'anàlisi de xarxes socials	7
1.1. Concepte: anàlisi de xarxes socials	7
1.2. Modelització d'interaccions socials a través de grafs	9
1.2.1. Relacions dirigides i no dirigides	10
1.2.2. Relacions explícites/implícites	11
1.2.3. Relacions amb pesos	12
1.3. Representació de xarxes socials	12
1.3.1. Representació descriptiva	12
1.3.2. Representació algebraica de les xarxes	13
1.4. Estructura d'una xarxa	14
1.4.1. Components connexes	14
1.4.2. Cliques	15
1.4.3. Comunitats o clústers	16
1.4.4. Modularitat	17
1.4.5. Grau	17
2. Visualització de xarxes socials	19
2.1. Gephi: descàrrega, instal·lació, i ús	19
2.1.1. Pas 1: descarreguem i instal·lem Gephi	20
2.1.2. Pas 2: obrim Gephi i importem l'arxiu de dades	20
2.1.3. Pas 3: informe d'importació	21
2.1.4. Pas 4: visualitzem i treballem els grafs	22
2.1.5. Pas 5: previsualització	31
2.1.6. Pas 6: exportació	32
2.2. Exemple d'anàlisi dels amics de Facebook amb Gephi. Detecció de comunitats	33
2.2.1. Descàrrega i preparació de dades	33
2.2.2. Visualització i anàlisi	33
2.3. Exemple: visualització de la xarxa d'interaccions al voltant d'un <i>hashtag</i> a Twitter	38
2.3.1. Càrrega d'arxius	38
2.3.2. Centralitat	39
2.3.3. Detalls dels nodes	40
2.3.4. Comunitats d'usuaris	42
2.3.5. Millora de la visualització	44
2.3.6. Exportació de la imatge	45

3. Visualització de la difusió d'informació: viralitat.....	47
3.1. Divulgació per xarxes	48
3.2. Tipus de contingut viral	50
3.2.1. Contingut viral espontani	50
3.2.2. Contingut viral estratègic	50
3.2.3. Contingut viral circumstancial	51
4. Altres eines per visualitzar dades de Facebook.....	53
4.1. Facebook Insights	53
4.1.1. Viralitat de publicacions	54
4.1.2. M'agrada	55
4.1.3. Abast	56
4.1.4. Visites a la pàgina	56
4.1.5. Extracció de dades	56
4.2. Rfacebook: Accés a l'API de Facebook via R (opcional)	57
4.2.1. Analitzar les dades d'una pàgina de Facebook	57
4.2.2. Analitzar les dades d'un grup de Facebook	60
4.2.3. Extracció d'informació personal	61
4.2.4. Informació de la pàgina de Fans	61
4.2.5. Anàlisi de «m'agrada»	61
Bibliografia.....	65

Introducció

En tots els *social media* circula una gran quantitat de dades, que ben analitzades tenen el potencial de generar molta de la informació que precisen els gestors dels diferents tipus d'organitzacions per prendre decisions.

Visualitzar la relació i les xarxes que creen les dades facilita l'anàlisi i contribueix a la presa de decisions.

En aquest mòdul, en primer lloc estudiarem de manera introductòria el concepte d'anàlisi de xarxes socials, la seva representació i les estructures de les xarxes. En segon lloc, treballarem les eines amb les quals podrem descarregar les dades i visualitzar-les per tenir-les preparades per al seu anàlisi final. Aquest últim aspecte el treballarem a l'assignatura Respondre preguntes de recerca analitzant xarxes. En tercer lloc, veurem el fenomen de la viralitat com a exemple d'allò que permet obtenir la visualització. Per últim, veurem altres eines per visualitzar dades descarregades de Facebook.

Objectius

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

1. Identificar els elements a les xarxes i les dinàmiques de les converses, analitzant les interaccions a les xarxes.
2. Recollir i analitzar opinions.
3. Anàlisi de xarxes qualitatiu i quantitatiu.
4. Visualitzar les dades existents a les xarxes socials.
5. Conèixer eines per visualitzar dades.
6. Reconèixer i avaluar de forma curosa la rellevància i significació de la informació.
7. Saber comunicar les conclusions de manera clara i sense ambigüitats.
8. Comprendre i emetre informes professionals i específics en l'àmbit del *social media*.
9. Aplicar l'ús de les TIC.

1. Introducció a l'anàlisi de xarxes socials

Les xarxes socials s'estudien atenent la seva representació, la seva estructura i els seus components, amb l'objectiu de conèixer el funcionament de les relacions, el comportament de les comunitats existents en una xarxa, el capital social de les organitzacions o de les persones, i fins i tot per conèixer el funcionament d'un producte en el mercat, o per al desenvolupament de nous productes.

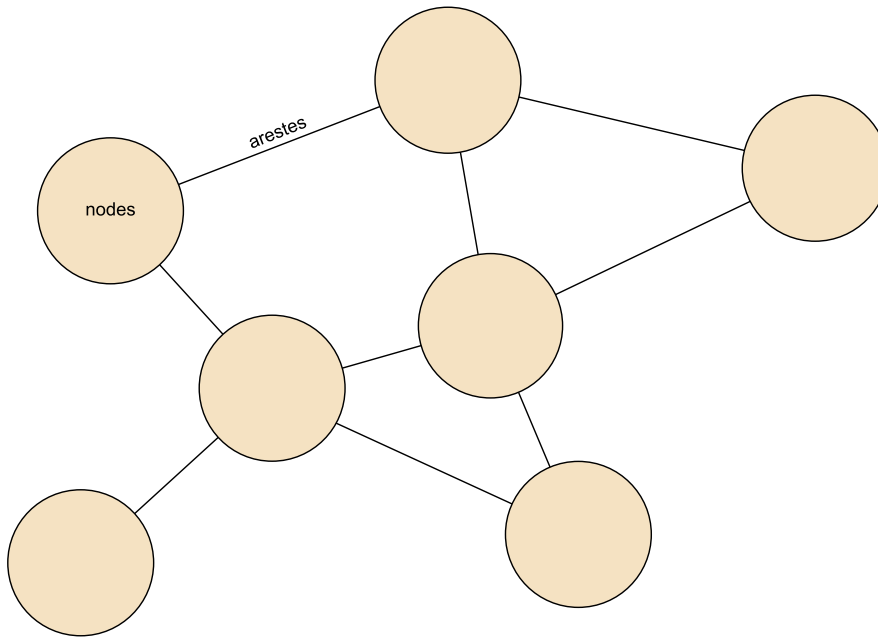
En aquest apartat estudiarem, en primer lloc, el concepte d'anàlisi de xarxes socials. En segon lloc, veurem la modelització de les interaccions socials a través de grafs i què són les relacions dirigides i no dirigides, les relacions explícites i implícites, i les relacions amb pesos. En tercer lloc, treballarem com podem representar les xarxes socials, de manera descriptiva i de manera algebraica. Per últim, coneixerem els components que conformen l'estructura d'una xarxa social. Aquests aspectes són els més bàsics per visualitzar una xarxa social. Aquest últim tema el veurem a la segona part d'aquest mòdul.

1.1. Concepte: anàlisi de xarxes socials

L'anàlisi de xarxes socials té les seves arrels a principis del segle XX, i consisteix en fer servir l'anàlisi de grafs per a investigar les relacions socials.

Un graf es defineix com un conjunt de nodes i arestes (o connexions entre els nodes): $G = (N, A)$ (figura 1). Els grafs es poden usar per modelitzar estructures basades en diferents tipus de relacions entre persones o grups.

Figura 1. Representació descriptiva d'una xarxa social



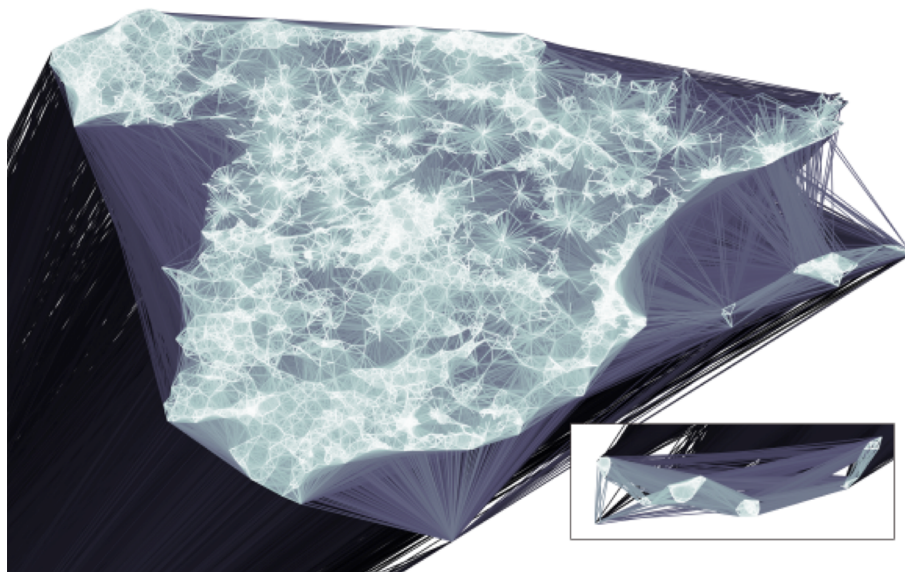
En una xarxa social, els nodes o vèrtexs representen individus o organitzacions i les arestes representen les seves relacions.

Alguns models de grafs de les xarxes socials són:

- **Grafs d'amistat:** són grafs on dues persones estan connectades si són amics (en el món real, a Facebook...).
- **Grafs d'interacció:** són grafs on es representa la interacció d'una persona amb un altra. Visualitzem una aresta d'una persona a una altra si la primera persona ha interactuat amb la segona persona.
- **Grafs d'influència:** són grafs on es representa la influència d'una persona en un altra. Visualitzem una aresta d'una persona a una altra si la primera persona pot influir la segona persona.

En els *social media* a partir de les connexions entre membres d'una xarxa es construeixen infinitat de xarxes socials. Un exemple el veiem a les connexions entre usuaris de la xarxa social en línia de Tuenti, en un mapa de l'Estat espanyol. Veiem que les mateixes connexions entre usuaris dibuixen el mapa d'Espanya (figura 2).

Figura 2. Connexions entre usuaris de la xarxa social en línia Tuenti que dibuixen el mapa d'Espanya



Font: Laniado i altres (2017).

Tradicionalment, els estudis d'anàlisi de xarxes socials en disciplines com a la sociologia feien servir enquestes per a recopilar informació sobre les connexions entre persones: per això es tractava de casos d'estudi generalment limitats a un nombre reduït de nodes, i el procés de recopilació de dades era costós. Amb la difusió d'internet i dels mitjans socials, la disponibilitat de rastres digitals fa que es generin moltes dades que faciliten aquest tipus d'anàlisi i permeten investigar les dinàmiques socials, els patrons de comportament col·lectiu, la influència entre persones i els mecanismes de difusió de la informació. D'altra banda, la tecnologia anomenada *big data* facilita el processament d'aquestes dades.

A continuació introduïrem els conceptes bàsics de teoria dels grafs, relacions, representació i d'anàlisi de xarxes socials, i la seva aplicació al cas dels *social media*.

1.2. Modelització d'interaccions socials a través de grafs

Les dinàmiques de la xarxa dicten la difusió d'informació, notícies i idees. Poden ajudar a identificar els gustos, les opinions i l'activitat d'algú. Si podem comprendre la xarxa de relacions al voltant d'una persona, tindrem un coneixement molt més profund que si els avaluem de manera aïllada.

Quan estudiem una xarxa social podem trobar persones influents, anticipar pics de demanda de productes o serveis, generar enfocaments de màrqueting més específics i predir l'activitat il·legal. A un nivell més personal, també podem crear comunitats, identificar persones vulnerables i aïllades i ajudar-les a trobar noves connexions.

El que fem amb l'anàlisi de xarxes socials és representar la realitat complexa de les dinàmiques socials a través d'un model simplificat, en què ens centrem en les relacions entre els individus. Sempre hem de tenir present que es tracta d'un model que ens proporciona una representació parcial de la realitat. No sempre serveix aquest tipus de model, i hi ha diversos aspectes que s'han de tenir en consideració quan l'apliquem.

Podem trobar diferents tipus de relacions socials que originen diferents tipus de xarxa. A continuació veurem tres grups de relacions, basades en criteris diferents: direcció, manifestació de la relació i pesos.

Les relacions les identifiquem per les línies (arestes) que uneixen els nodes.

1.2.1. Relacions dirigides i no dirigides

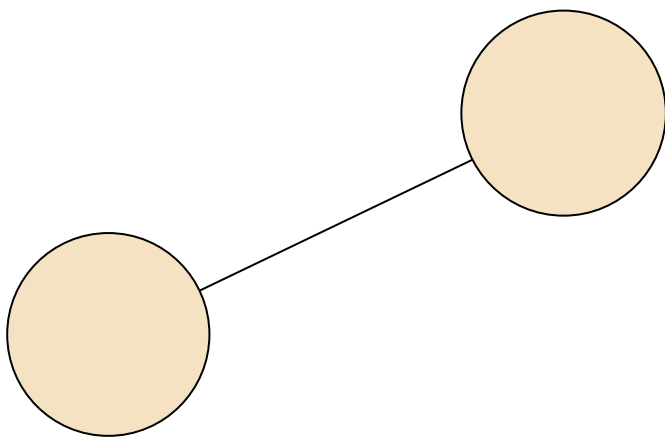
En vista de si les relacions tenen una direcció o no la tenen, veiem els següents tipus de relacions socials, que al mateix temps conformen dos tipus de xarxes:

a) **Relació no dirigida** on la relació social és recíproca (bidireccional). Alguns exemples són:

- Amistats a Facebook.
- Coautors d'articles científics.

L'aresta va cap a les dues direccions.

Figura 3



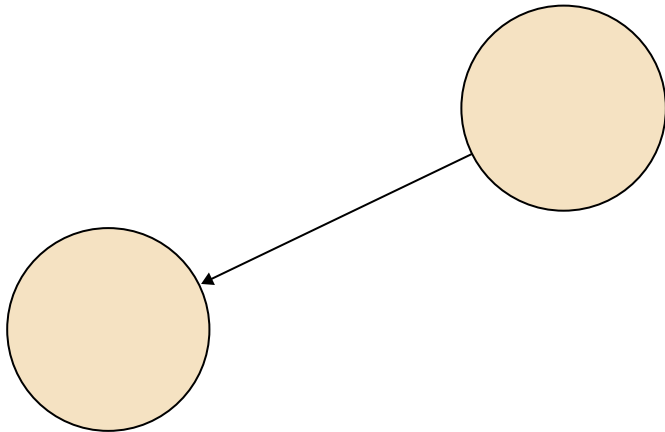
b) **Relació dirigida**, en què la relació social no és bidireccional. Alguns exemples són:

- Seguidors a Twitter (un usuari segueix un altre usuari).
- *Fan Page* a Facebook (un usuari segueix una pàgina).
- Mencions en mitjans socials (un usuari menciona un altre usuari).
- *Retweets* a Twitter (un usuari comparteix el contingut publicat per un altre usuari).

- Respostes en mitjans socials (un usuari contesta un altre usuari).

Representem aquesta relació mitjançant una fletxa d'un node cap al que té la relació. Però el segon node no té relació amb el primer.

Figura 4



Atenent el tipus de relació dirigida o no dirigida, trobem dos tipus de xarxa. Una xarxa que representa relacions no dirigides, o simètriques, es diu **xarxa no dirigida**; mentre que una xarxa de relacions que té una direcció definida es dirà **xarxa dirigida**, en què es fa distinció entre arestes que entren i que surten d'un node. Una xarxa dirigida es pot representar sempre com a no dirigida, si deixem de considerar la direcció de les interaccions i ens centrem només en el fet que dos nodes estiguin connectats d'alguna forma.

Es poden establir llinars per veure la connexió de nodes. Es pot definir una xarxa no dirigida en què dos nodes estan connectats només si s'han fet, per exemple, alguna repulada mútuament. És a dir, que es consideren només les relacions recíproques, en què la informació flueix en els dos sentits.

1.2.2. Relacions explícites/implícites

També podem distingir relacions socials explícites i implícites, d'acord als següents criteris:

a) Relació explícita. Una relació serà explícita quan els mateixos usuaris declaren la relació:

- Amistats a Facebook.
- Seguidors a Twitter.

b) Relació implícita. Una relació serà implícita quan la relació es dedueix del comportament, s'analitzen aspectes com per exemple:

- Interaccions en mitjans socials (mencions, *retweets*, respostes...).
- Coautors d'articles científics, actors que han actuat a la mateixa pel·lícula.

- Similitud (dos usuaris compren els mateixos productes, o segueixen les mateixes pàgines).

En el cas de relacions implícites, s'hi poden definir **llindars** per a considerar que dos nodes estan connectats. Per exemple, es pot crear una xarxa entre usuaris de Twitter en què hi hagi un enllaç directe del node A al node B només si A ha fet al menys N *retweets* a B, i es pot establir per exemple un llindar de $N = 3$. D'aquesta manera, tindrem una xarxa en què les connexions indiquen interaccions repetides, i no ocasionals.

1.2.3. Relacions amb pesos

Finalment, distingim entre relacions amb pesos i sense pesos.

Pes és el concepte emprat per indicar el valor de la relació entre dos nodes. Representa el nombre de camins diferents existents entre el node d'origen i el node final.

Relacions amb pesos es refereix al nombre d'interaccions entre dos nodes i s'anota sobre l'aresta. Per exemple, l'aresta $A \rightarrow B$ pot tenir com a pes el nombre de *retweets* que A ha fet a B.

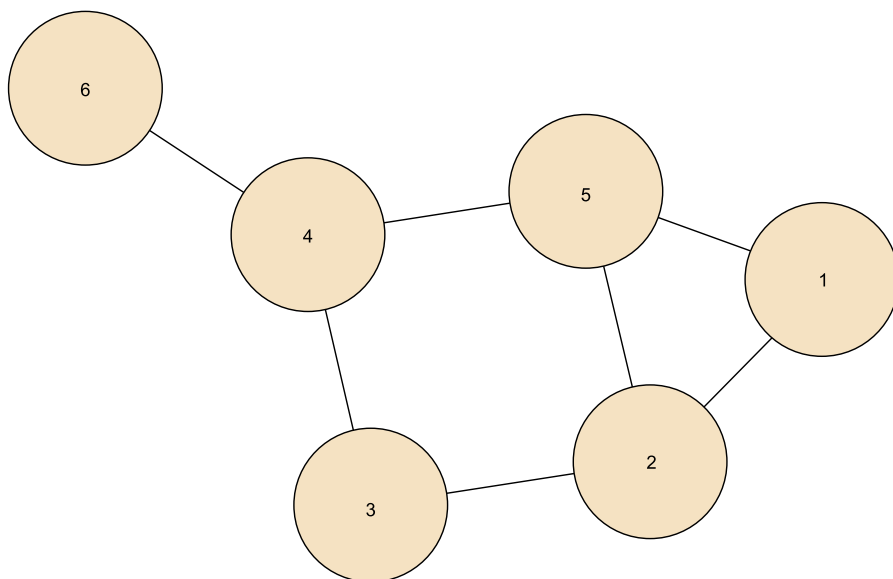
1.3. Representació de xarxes socials

Les xarxes socials, per tal de poder-les visualitzar, es poden representar de manera descriptiva i algebraica. Anem a veure en què consisteixen aquests dos modes de representar les xarxes.

1.3.1. Representació descriptiva

Les xarxes es poden representar amb els grafs, que consten de nodes i enllaços. Recordem que un graf es defineix com a un conjunt de nodes i arestes (o connexions entre els nodes): $G = (N, A)$ (figura 5).

Figura 5. Exemple de graf



Font: Wikipedia.

Els nodes són els elements principals de les xarxes i les arestes són els enllaços entre els nodes, que indiquen relació.

1.3.2. Representació algebraica de les xarxes

Hi ha diferents maneres de representar un graf a nivell matemàtic. Aquestes les veiem a continuació (figura 6).

1) **Conjunts d'arestes:** és el format més senzill, que consisteix en un llistat de parelles de nodes connectats. Típicament es pot emmagatzemar com un arxiu d'extensió csv (Comma Separated Values) amb dues columnes, en què cada línia representa una aresta i conté una parella de nodes connectats. En general, farem servir aquest format.

2) **Llista d'adjacència:** per cada node, s'indica el llistat de nodes amb el qual està connectat. A l'exemple de la figura 6 podem veure que el node 1 està connectat amb si mateix i amb els nodes 2 i 5 {1,2,5}, el node 2, al seu torn, està connectat a més amb el 3 i el 5 {3,5}, el 3 amb el 4 {4}, i el node 4, amb el 5 i el 6 {5,6}. El node 4 també té connexió amb el 3, però ja s'ha expressat abans en la representació del node 3.

3) **Matriu d'adjacència:** cada node representa una línia (i una columna), cada cel·la representa la intersecció entre dos nodes, i té un 0 si la connexió no existeix, mentre que, si hi ha una connexió, la cel·la conté el valor del pes de l'aresta corresponent. En general es fa servir el valor 1 per a indicar que hi ha una connexió. En el cas de xarxes pesades, es poden tenir valors diferents. Per exemple, es pot veure que la diagonal que va de la part superior esquerra a la

part inferior dreta, és ple de zeros menys a la primera cel·la, ja que només el node 1 té connexió amb si mateix. Com en aquest cas la xarxa no és dirigida, la matriu és simètrica.

4) Matriu d'incidència: cada línia representa una aresta, i cada columna un node; en cada línia hi ha valors diferents de 0 només en correspondència dels nodes tocats per l'aresta.

Figura 6. Diferents representacions d'un graf

Graf G(V,A)	Conjunts	Llista d'adjacència	Matriu d'adjacència	Llista d'incidència
	$V = \{1, 2, 3, 4, 5, 6\}$ $A = \{\{1,1\}, \{1,2\}, \{1,5\}, \{2,3\}, \{2,5\}, \{3,4\}, \{4,5\}, \{4,6\}\}$	$\{\{1,2,5\}, \{3,5\}, \{4\}, \{5,6\}\}$	$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

Font: Wikipedia.

1.4. Estructura d'una xarxa

Les xarxes presenten estructures diferents que es caracteritzen en relació a la grandària, els *clustering*, i els components. En aquest apartat veurem alguns d'aquests aspectes per entendre'n després la visualització, tema principal de l'assignatura. En l'assignatura *Respondre preguntes de recerca analitzant xarxes* treballarem més a fons la teoria de les xarxes socials, conceptes com grandària, *clustering* i altres components.

Els aspectes que treballarem ara són: components connexos, cliques, comunitats o clústers, modularitat i grau.

1.4.1. Components connexes

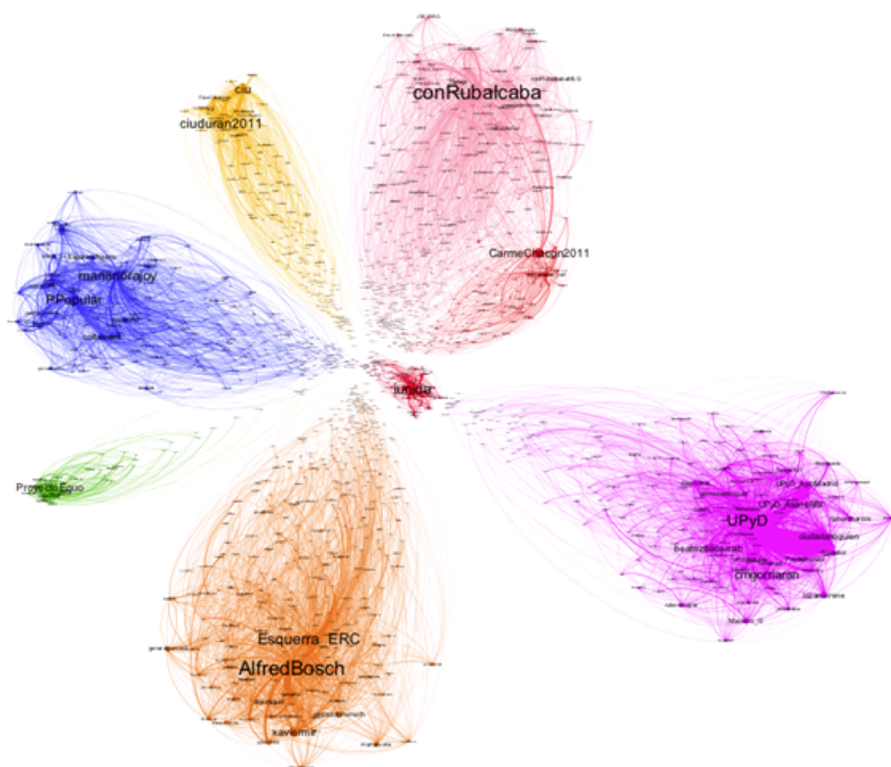
Es defineix component connex un grup de nodes en què sempre és possible trobar un camí entre qualsevol parella de nodes. Si un graf està compost per un sol component, és a dir, si és possible trobar un camí entre qualsevol parella de nodes que componen el graf, es tracta d'un graf «totalment connex».

Diem que un **component** està **fortament connex** en el cas d'un graf dirigit (no bidireccional). S'anomena component fortament connex un grup de nodes en què sempre és possible trobar un camí entre cada parella de nodes, en les dues direccions. Un exemple el veiem a la figura 7, que representa la xarxa de *retweets* entre polítics durant la campanya de les eleccions generals que va tenir lloc a Espanya el novembre de 2011 (20N). El color dels nodes representa les comunitats (clústers). No hi ha gairebé cap *retweet* entre partits diferents,

o sigui que cada partit tendeix a representar un component connex diferent, amb l'excepció d'alguna connexió. Aquestes comunitats en aquest exemple han estat identificades amb el mètode de Louvain.

El mètode de Louvain és un algorisme molt eficient i molt emprat per a la detecció de comunitats (clústers), com veurem més endavant.

Figura 7. Xarxa de *retweets* entre polítics



Retweet social graph. The color and the size of each node correspond to the cluster it belongs and its in-degree respectively.

Font: Aragón i altres (2013).

Usualment, analitzant xarxes social és típic trobar una component que inclou la gran majoria dels nodes. Aquesta component s'anomena «component gegant». En xarxes socials, es considera que és molt probable que un component ja es connecti amb la resta de la xarxa (al component gegant) abans de que arribi a la mida de 10 nodes. Així una xarxa social pot estar típicament composta per una component gegant que inclou més del 90% dels nodes, i de molts altres petits components que inclouen molt pocs nodes cadascuna.

1.4.2. Cliques

S'anomena clíca un grup de nodes en què cada node està connectat amb tots els altres. És una estructura comuna en xarxes socials, especialment en l'adolescència, ja que els individus tendeixen a formar grups cohesionats.

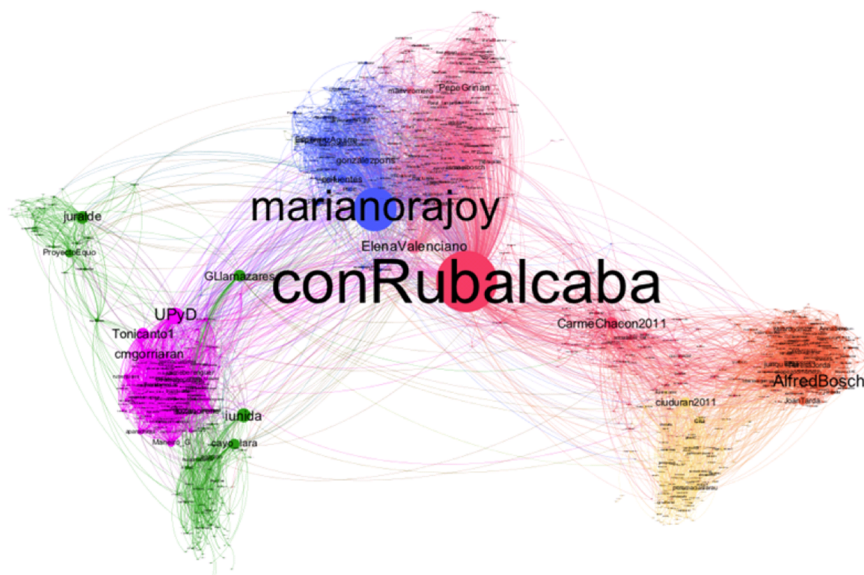
1.4.3. Comunitats o clústers

Mentre que els components connexes i les cliques representen conceptes matemàtics de teoria dels grafs que defineixen de manera unívoca un subconjunt de nodes amb una propietat ben precisa, el concepte de comunitat (o clúster) és més obert. Una comunitat és un conjunt de nodes que tenen moltes connexions entre ells.

Com que no hi ha una manera unívoca de definir comunitats en una xarxa, diferents algorismes poden identificar comunitats diferents.

Hi ha diferents tipus d'algorismes, alguns identifiquen comunitats que poden tenir solapaments, és a dir en què un node pot pertànyer a més d'una comunitat. Tanmateix, la majoria d'algorismes generen una partició, en què cada node pot pertànyer només a una comunitat. Un exemple el veiem a la figura 8, on hi ha la xarxa de respostes entre polítics durant la campanya de les eleccions generals que varen tenir lloc a Espanya el novembre 2011 (20N). El color dels nodes representa les comunitats (clústers) identificades amb el mètode de Louvain. Hi ha més interaccions entre partits diferents, però la majoria es troben dins d'un mateix partit, així que les comunitats tendeixen a coincidir amb els partits. Només Izquierda Unida i Equo es troben en el mateix clúster (verd) (Aragón i altres, 2013).

Figura 8. Xarxa de respostes entre polítics



Reply social graph. The size of each node corresponds to its in-degree and the color represents the cluster it belongs to (PSOE, PP, IU-EQUO, UPyD, CIU, ERC).

Font: Aragón i altres (2013).

1.4.4. Modularitat

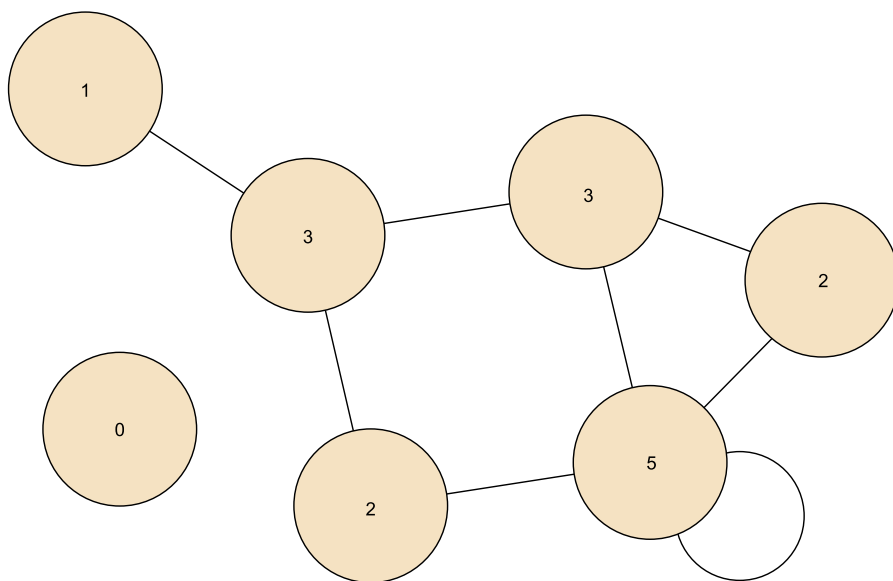
La modularitat és una mètrica que, donada una partició d'un graf en comunitats, expressa el grau de la separació entre comunitats. Més específicament, la modularitat té valors més alts com més enllaços cauen dins d'una mateixa comunitat, i com menys en cauen entre una comunitat i una altra (Newmann, 2006).

Aquesta mètrica es pot fer servir per a buscar la manera òptima de partir una xarxa en comunitats. Un algorisme molt eficient i molt emprat per a la detecció de comunitats, basat en l'optimització de la modularitat, és el mètode Louvain (Blondel i altres, 2008). El mètode descompon la comunitat d'amics en comunitats secundàries més petites. És un algorisme *greedy*, és a dir que no troba la solució òptima sinó una aproximació calculada de manera eficient. Trobar la solució òptima demanaria un gran esforç computacional amb molt temps de processament, per això es cerca una aproximació. Hi ha un element aleatori, és a dir que execucions diferents poden donar resultats diferents per al mateix graf. Aquest algorisme està implementat al programa Gephi (amb el nom «modularity»), i és el que fem servir en els nostres exemples.

1.4.5. Grau

El grau d'un node és el nombre d'arestes incidents sobre un node, i representa el nombre de connexions d'un node. És la mètrica de centralitat més senzilla. En altres paraules, el grau d'un vèrtex o node representa el nombre d'arestes que hi incideixen (figura 9).

Figura 9. Exemple de graf no dirigit on en cada node hi ha indicat el seu grau



Font: Wikipedia.

En el cas d'una xarxa dirigida (no bidireccional), es poden distingir el grau d'entrada i el grau de sortida:

- El **grau d'entrada** representa el nombre d'arestes que tenen com a destí el node considerat, o dit d'una altra manera, el nombre d'arestes que entren en un node.
- El **grau de sortida** representa el nombre d'arestes que tenen el node com a origen, o dit d'una altra manera, el nombre d'arestes que surten d'un node.

En general, el grau d'entrada representa l'atenció rebuda per un node: per exemple, el nombre de seguidors, el nombre de persones que han compartit (o han fet un *retweet*) algun contingut generat pel node considerat, o que han contestat un missatge. El grau de sortida, al contrari, tendeix a representar l'atenció que es dona a altres nodes, com ara el nombre de comptes que un node segueix, etc. Per això en molts casos el grau d'entrada és una millor mesura de la rellevància d'un node que el grau de sortida.

2. Visualització de xarxes socials

Visualitzar dades és, en molts casos, la millor manera per a poder entendre el que ens volen dir i explicar-les intuïtivament (influències, nodes destacats, grups, entre d'altres exemples).

Existeixen moltes eines per gestionar i visualitzar dades. N'hi ha d'específiques per veure el conjunt de dades que configuren les xarxes socials. Per a veure les xarxes disposem de diversos programaris amb capacitat de gestionar un gran volum de dades. Exemples d'aquestes eines són: Gephi, Pajek, CiteSpace, TouchGraph i biblioteques desenvolupades en R (igraph, sna, TNET, statnet i NetworkX).

En aquest capítol veurem primer com instal·lar i treballar amb Gephi. A continuació veurem exemples de visualització i anàlisi de xarxes a Facebook i Twitter amb el programa Gephi. En el primer cas, visualitzarem dades de Facebook, obtingudes amb Netvizz (Sherlock, 2014). En el segon cas veurem dades de Twitter, obtingudes amb Flocker. Recordem que Netvizz i Flocker són eines que hem vist a l'assignatura *Big data i anàlisi de xarxes socials: conceptes i eines* per a la recopilació de dades. En tercer lloc, veurem una de les prestacions de la visualització, que és la detecció de la viralització dels continguts. Per últim, veurem com podem analitzar xarxes amb Facebook Insights i Rfacebook. Aquesta última és una eina més avançada que requereix programació.

2.1. Gephi: descàrrega, instal·lació, i ús

En aquesta secció treballarem la visualització d'una xarxa social amb el programa de visualització de xarxes anomenat Gephi que hem esmentat a l'assignatura *Big data i anàlisi de xarxes socials: conceptes i eines*.

Recordem que Gephi és una plataforma interactiva de codi obert (*open source*) per a la visualització i exploració de tot tipus de xarxes i sistemes complexos amb gràfics dinàmics i jeràrquics.

Seguidament, us descrivim els passos que heu de realitzar per a obtenir una xarxa amb els atributs més comuns a partir d'un fitxer ja generat. Podeu fer servir una xarxa de contactes propis. Si feu servir la xarxa de Facebook, tingueu en compte que recentment Facebook ha restringit l'accés a les dades i aquesta opció ja no està disponible, així que podreu fer el mateix amb una altra xarxa baixada de Netvizz com hem vist a l'assignatura *Big data i anàlisi de xarxes socials: conceptes i eines*, com ara la xarxa de «m'agrada» entre pàgines al voltant d'una pàgina de partida.

En cas de no disposar d'una xarxa, podeu emprar una de les que gephi posa a la vostra disposició, des de la pàgina web: <https://github.com/gephi/gephi/wiki/Datasets>. Quan es descarrega el programa, en porta incorporats alguns exemples. Treballarem en aquest apartat amb diferents exemples.

Assumim que ja teniu seleccionat un fitxer en format GDF obtingut amb Netvizz, i els passos a seguir són els que descrivim a continuació.

2.1.1. Pas 1: descarreguem i instal·lem Gephi

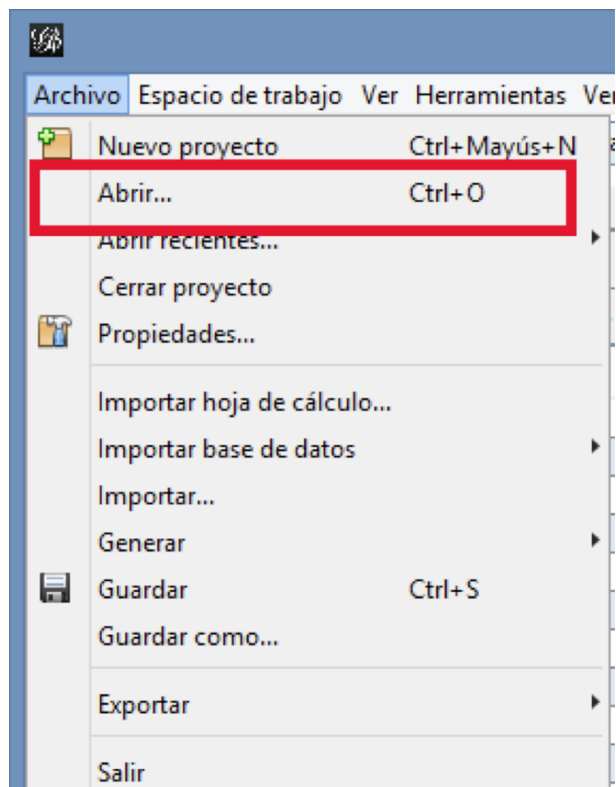
Us podeu descarregar el programa des del següent enllaç: <http://gephi.org/users/download/>.

El programa necessita que tingueu instal·lat Java JRE. El més probable és que ja el tingueu instal·lat. En cas contrari, podeu trobar instruccions a la pàgina web de Gephi, i seguir les instruccions d'instal·lació, d'acord al vostre sistema operatiu (Windows, Mac OS o Linux): <https://gephi.org/users/install/>.

2.1.2. Pas 2: obrim Gephi i importem l'arxiu de dades

A la barra de menú premeu «obrir» i seleccioneu l'arxiu que ha generat Netvizz o Flocker, o el que heu seleccionat del web de Gephi. O be, un dels que ve per defecte amb el programa, com per exemple Les Miserables.

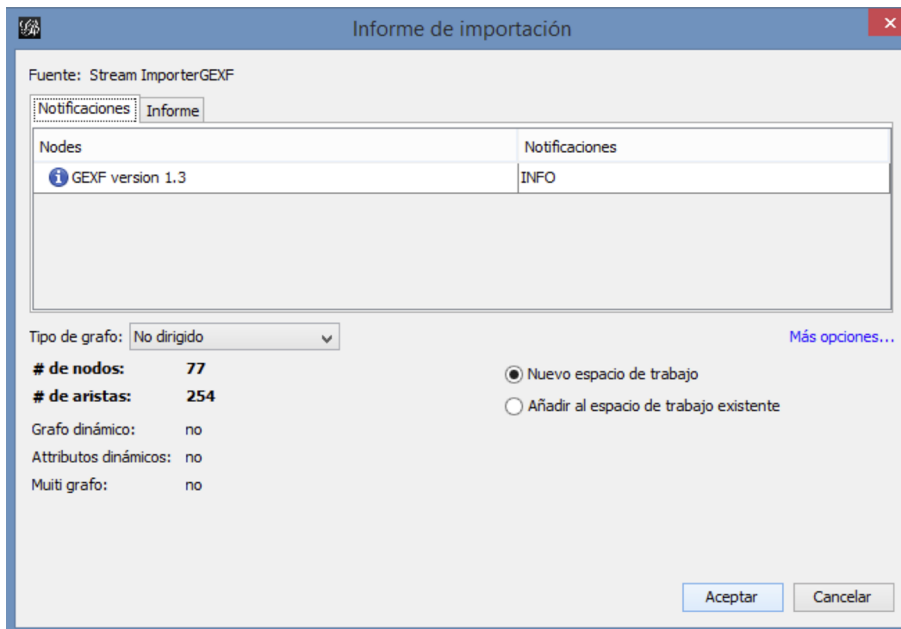
Figura 10



2.1.3. Pas 3: informe d'importació

Per poder visualitzar les dades, hem d'importar l'informe. Quan s'obre aquesta finestra apareix l'informe sobre el nombre de nodes, el nombre d'arestes, i el tipus de graf.

Figura 11



Obtindreu una finestra similar a aquesta i premerem «acceptar» per a validar i veure les dades que hem importat.

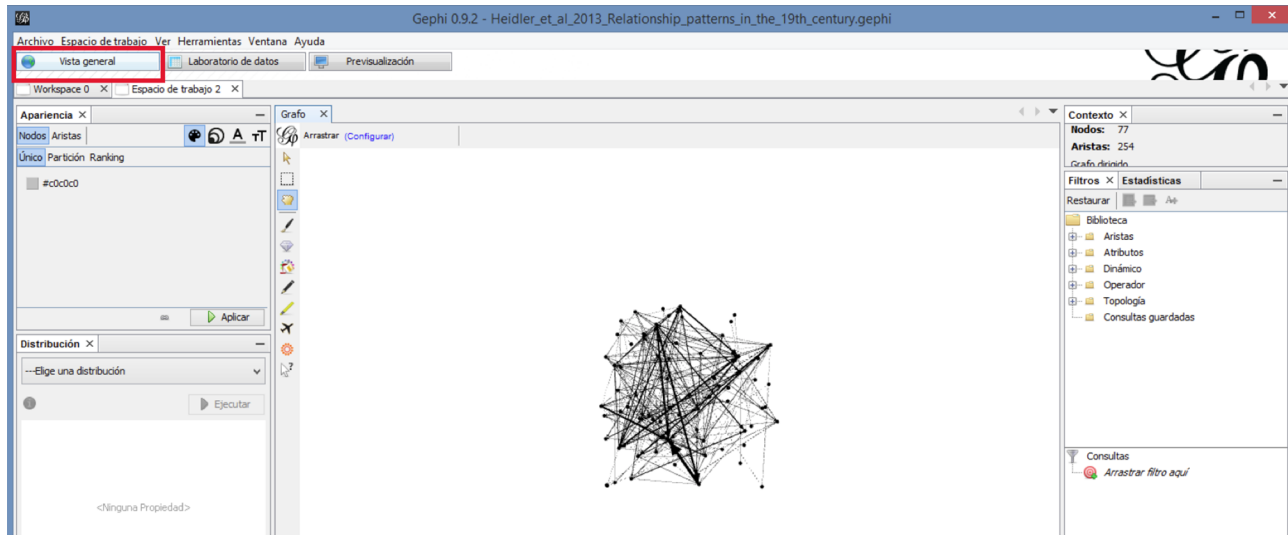
Figura 12

Id	Label	Interval	Modularity Class
11	Valjean		1
48	Gavroche		8
55	Marius		6
27	Javert		7
25	Thenardier		7
23	Fantine		2
58	Enjolras		8
62	Courfeyrac		8
64	Bossuet		8
63	Bahorel		8
65	Joly		8
24	MmeThenardier		7
26	Cosette		6
41	Eponime		7
57	Mabeuf		8
59	Combeferre		8
61	Feuilly		8
0	Myriel		0
66	Grantaire		8
68	Gueulemer		7
69	Babet		7
70	Claquesous		7
16	Tholomyes		2
60	Prouvaire		8
71	Montparnasse		7

2.1.4. Pas 4: visualitzem i treballem els grafs

Amb la «vista general» podem veure una xarxa, la densitat de la qual pot variar depenent del nombre de nodes.

Figura 13

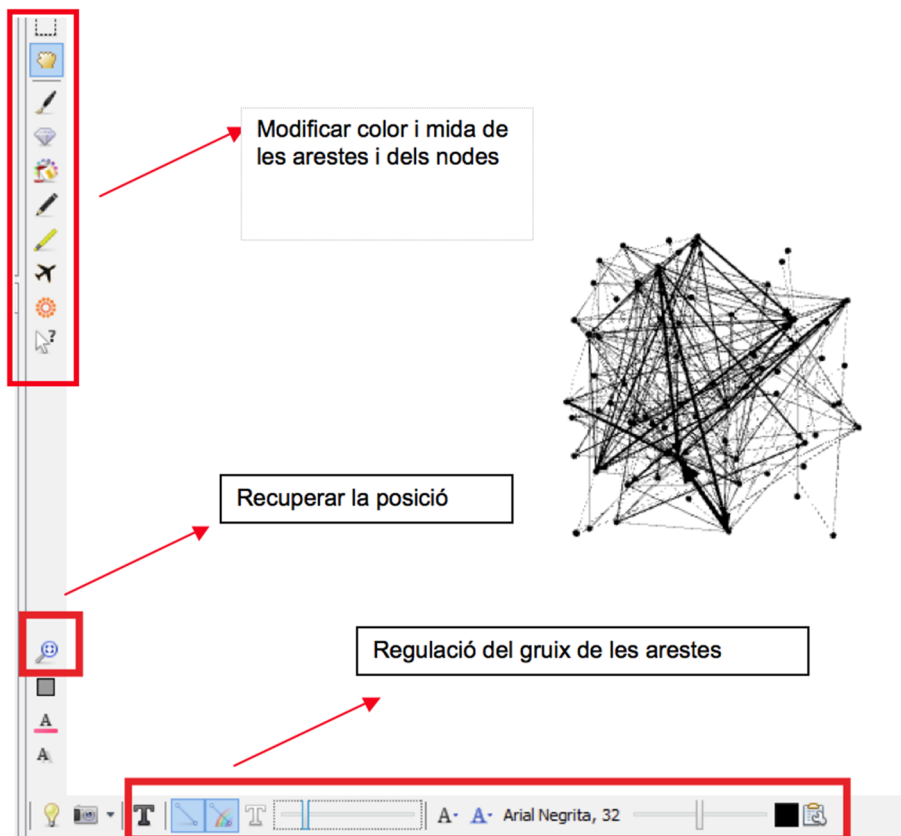


El gràfic que en resulta pot ser confús, perquè els nodes i les arestes queden molt superposades. Per això n'hem de treballar la visualització.

1) Distribució i visualitzacions del gràfic

Amb el ratolí ens podem moure pel gràfic, i amb la roda del ratolí podem ampliar o disminuir la mida del gràfic. Amb el botó dret podem arrossegar el ratolí per recórrer les arestes i els nodes del gràfic. Podem també regular els gruixos de les arestes, amb l'eina que hi ha a la part inferior. Si perdem el gràfic de vista, podem recuperar-lo clicant la icona de lupa.

Figura 14

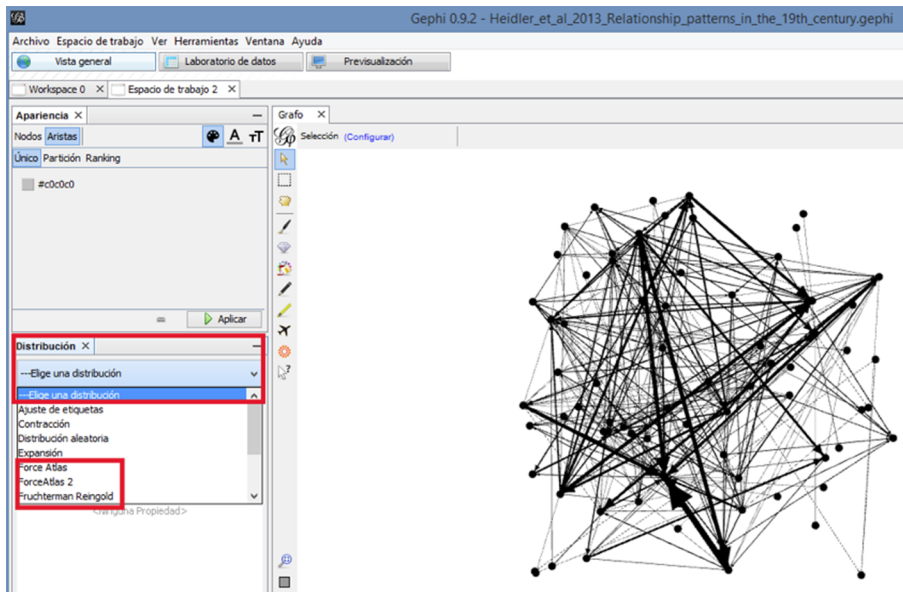


Si els nodes grans cobreixen els petits el disseny no és del tot clar. Per tant, hem de treballar més la visualització, tenint en compte la mida dels nodes. Utilitzarem algorismes i treballarem amb les mides, per tal que els nodes no es solapin.

2) Treballem la visualització: algorismes

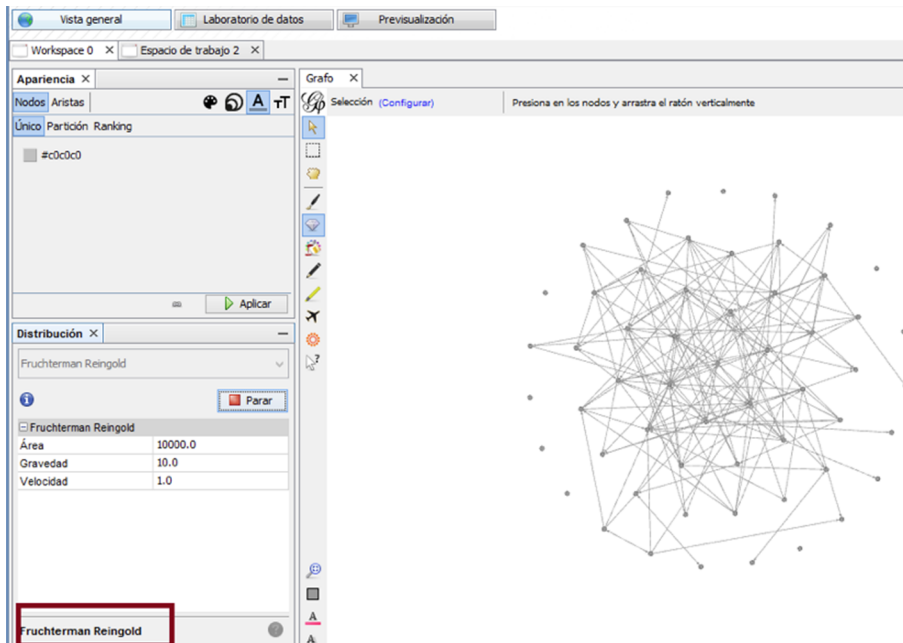
A la part esquerra de la pantalla hi ha el menú de **distribució**. En aquest menú seleccionem l'algorisme amb el que volem fer l'espacialització del gràfic, per tal de donar-hi més espai. Podem fer servir dos algorismes, per veure les diferències: Fruchterman Reingold i Force Atlas2.

Figura 15



Fruchterman Reingold ofereix una visualització que disposa els nodes d'una manera gravitacional (atracció-repulsió). És possible distingir les comunitats (parts més densament connectades de la xarxa). S'ha de pitjar «executar» i deixar que funcioni fins que s'estabilitzi el gràfic, i després pitjar «parar». Podem engrandir el gràfic amb la roda del ratolí. Per tornar a centrar el gràfic hem d'utilitzar la lupa.

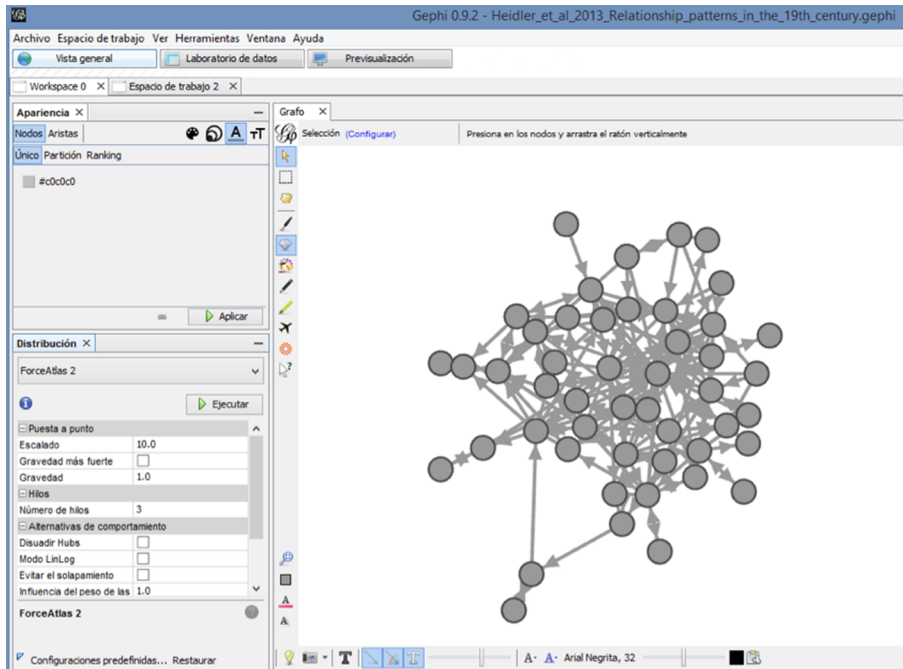
Figura 16



L'algortme de disseny **Force Atlas 2** serveix per a dispersar grups i donar espai al voltant de nodes més grans. Un cop seleccionat l'algortme, podeu deixar els valors que venen per defecte, i a continuació hem de prémer «executar» i deixar actuar la funció fins que el gràfic s'estabilitzi.

Podem aplicar Force Atlas 2 directament sense aplicar Fruchterman Reingold abans, però és millor aplicar abans aquest algorisme, per a desenredar la xarxa abans (Grandjean, 2016).

Figura 17



Podem engrandir la visualització jugant amb les mesures que ens ofereix el menú de cada algorisme («escalat» en el cas del ForceAtlas2, i «àrea» en el cas de l'algorisme Fruchterman Reingold. Sempre que apliquem els canvis, haurem de prémer «executar», i quan estigui estabilitzat, prémer «parar».

3) Configurar mida i color dels nodes

El mòdul d'«aparença» ofereix diverses possibilitats per configurar el color dels nodes. Per això hem de seleccionar l'apartat «partició» i seleccionar els colors a la paleta. Quan finalitzem, hem de prémer «aplicar» per veure'n el resultat.

Figura 18

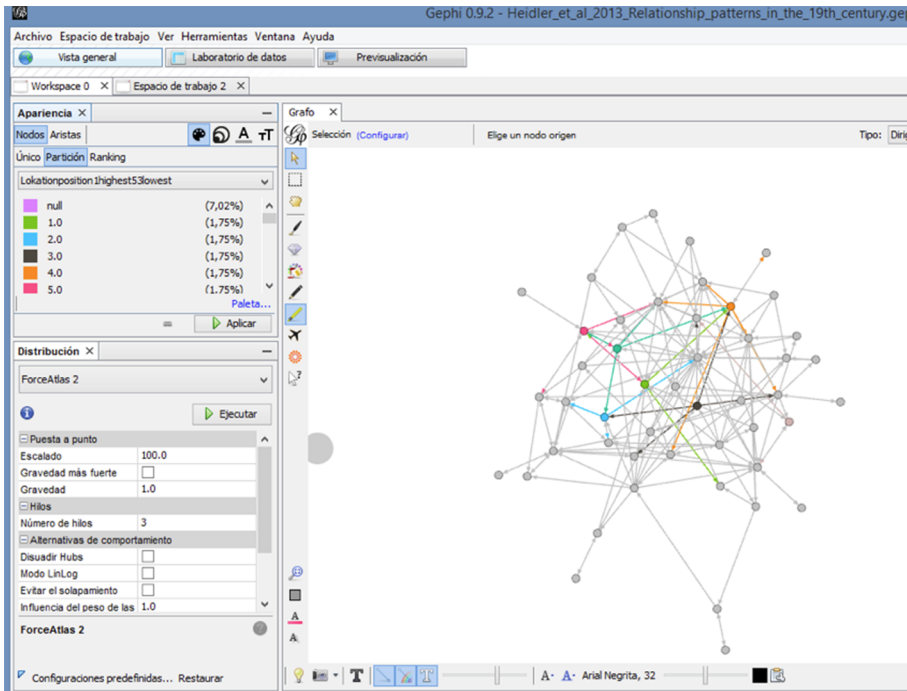
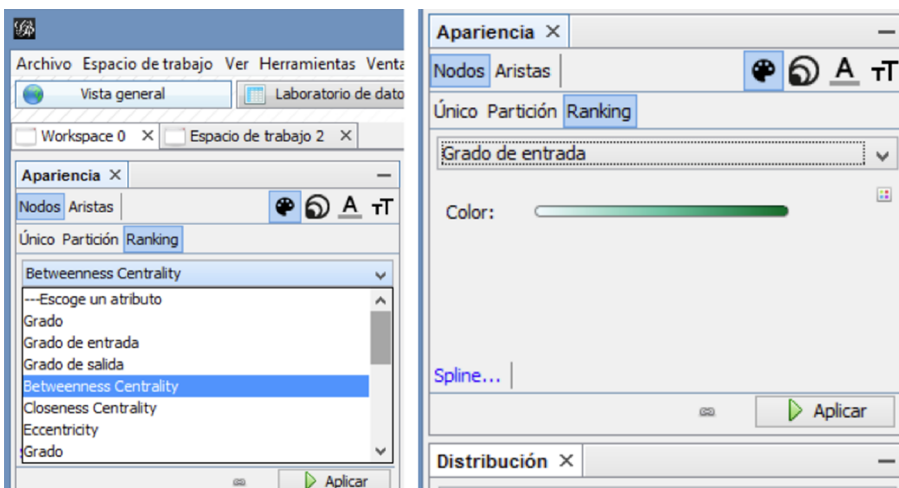


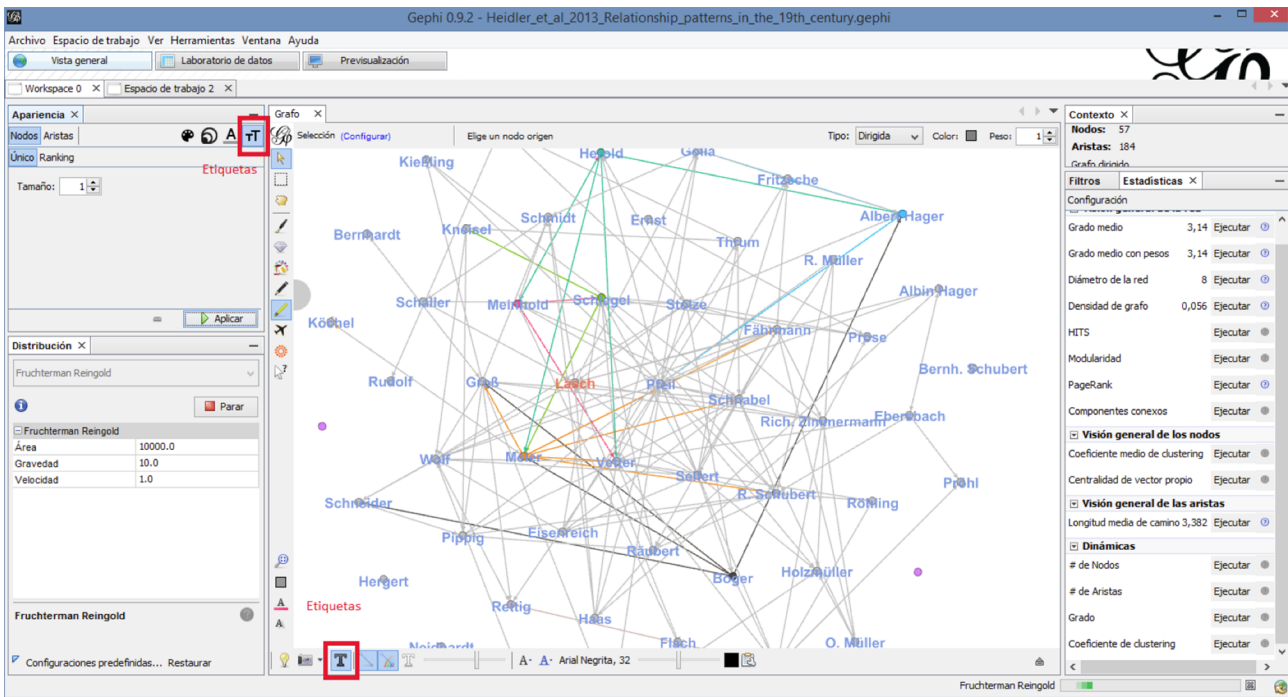
Figura 19



4) Mostrar etiquetas

Podem explorar la xarxa en detall i veure etiquetes dels nodes, els colors i la mida que indiquen els nodes centrals. En cas necessari podem treballar les etiquetes per canviar color de les lletres i mida.

Figura 20



5) Configurar les mesures de centralitat

A l'apartat d'«estadística» podem veure diferents mesures: grau mig, modularitat, etc. Podeu jugar amb la visualització. Els conceptes i la seva aplicació els veurem a l'assignatura Respondre preguntes de recerca analitzant xarxes.

Figura 21

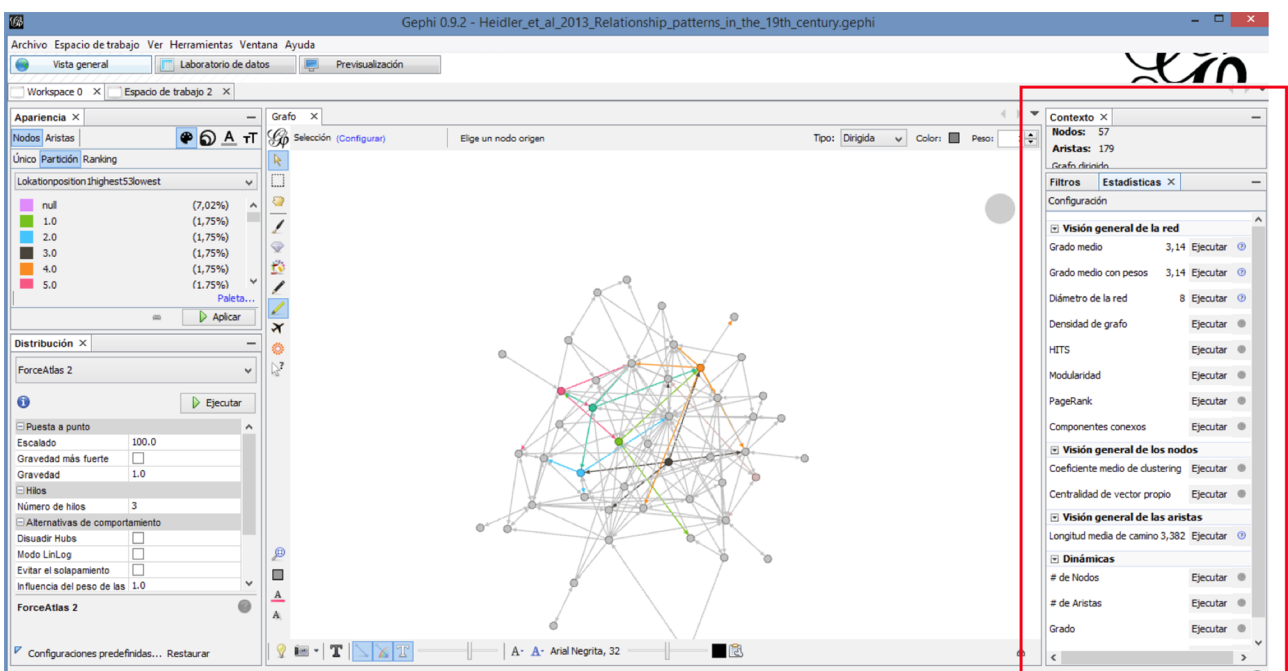
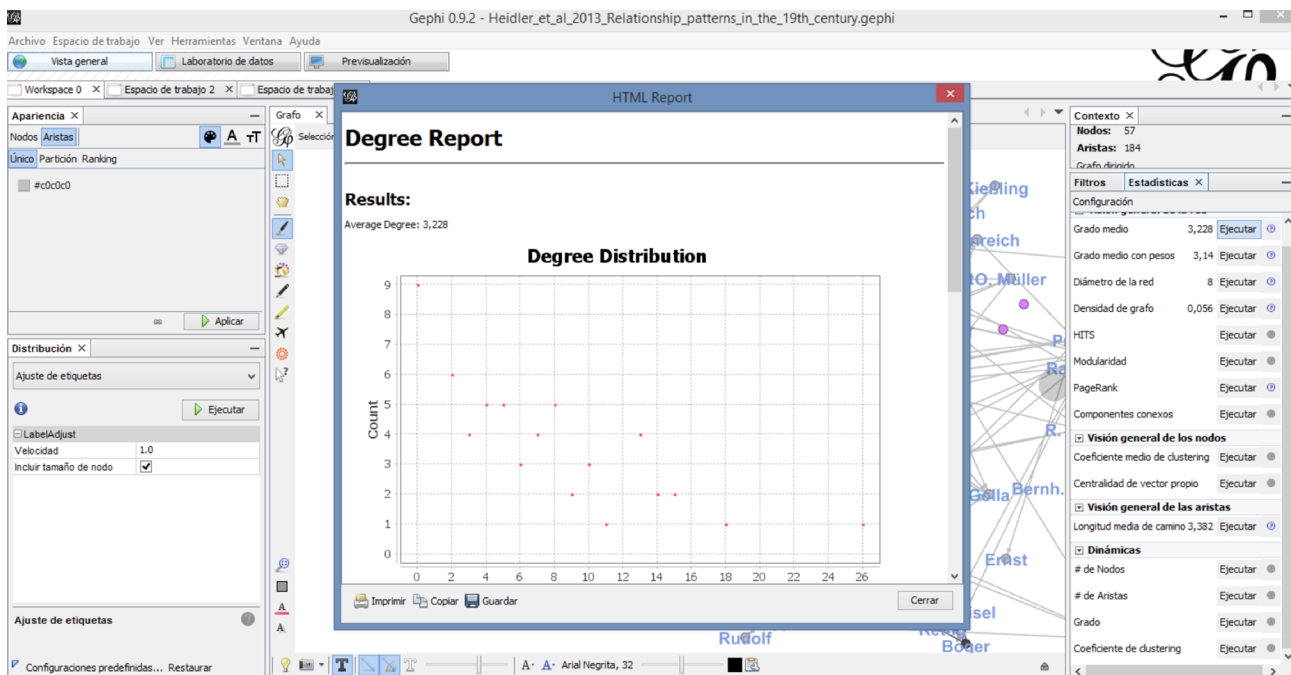


Figura 22



Podem afegir més informació al gràfic donant nous atributs i nous colors als nodes. En el «laboratori de dades», hem de seleccionar la taula de les dades, i ordenar-les segons el seu pes (*weight*), que pot ser variat. Per això s'ha de calcular el **grau ponderat** dels nodes. També podem assenyalar-ne la direcció, tenint en compte l'origen i el destí, mitjançant una fletxa en la pantalla de vista general. Per tant, el grau que s'ha de calcular ha de diferenciar entre les connexions d'entrada i de sortida. Per aconseguir-ho, hem de prémer el «**Grau Mitjà Ponderat**» en l'apartat «Estadística» per tal de calcular aquests valors per a cada node.

6) Detecció de comunitats

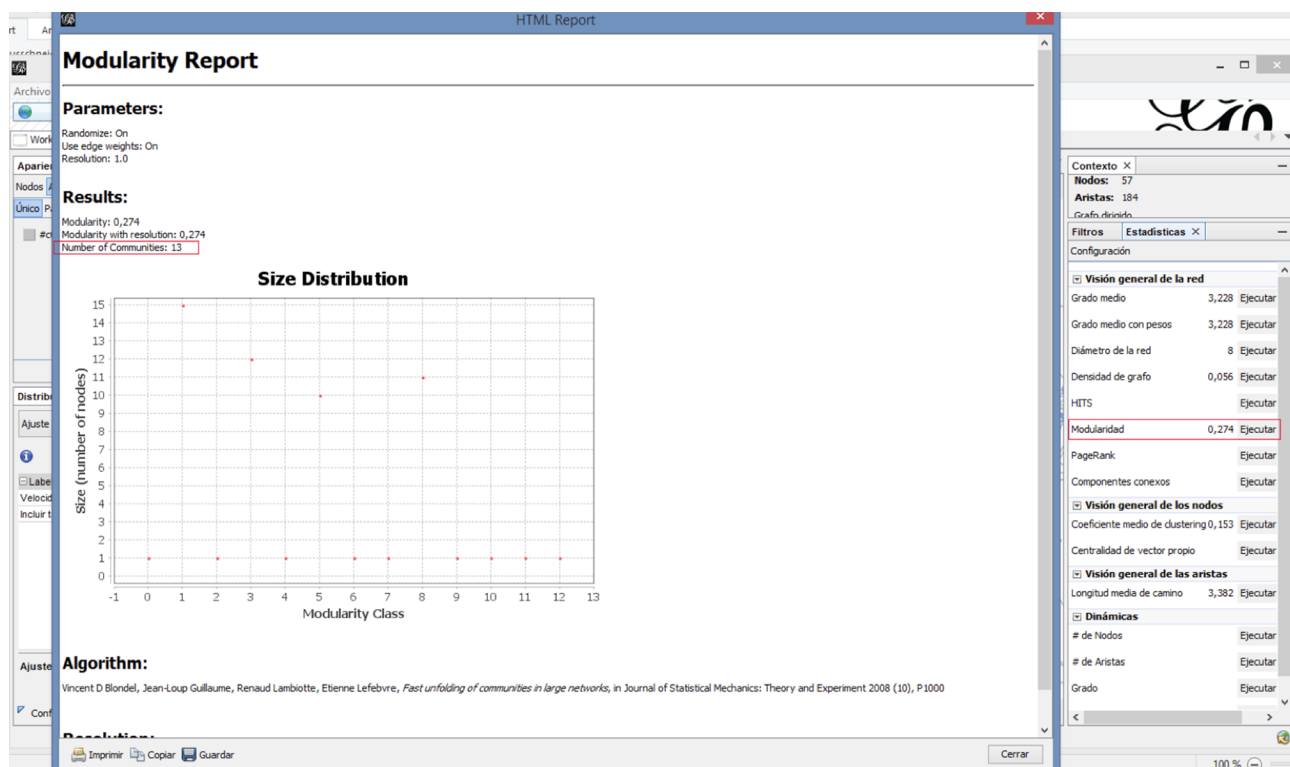
Per analitzar les xarxes és important detectar i estudiar les comunitats. Per aconseguir-ho, hem de visualitzar els clústers de la xarxa. Les comunitats les podem detectar amb diferents algorismes. Gephi implementa el mètode Louvain.

Per a detectar les comunitats o clústers, hem d'anar al panell d'«Estadístiques» i seleccionar «modularitat». S'obre un menú en què hem de marcar «Aleatori» i acceptar.

Vegeu també

Al subapartat 2.2 veurem un exemple d'identificació de comunitats en una xarxa social, mitjançant Gephi.

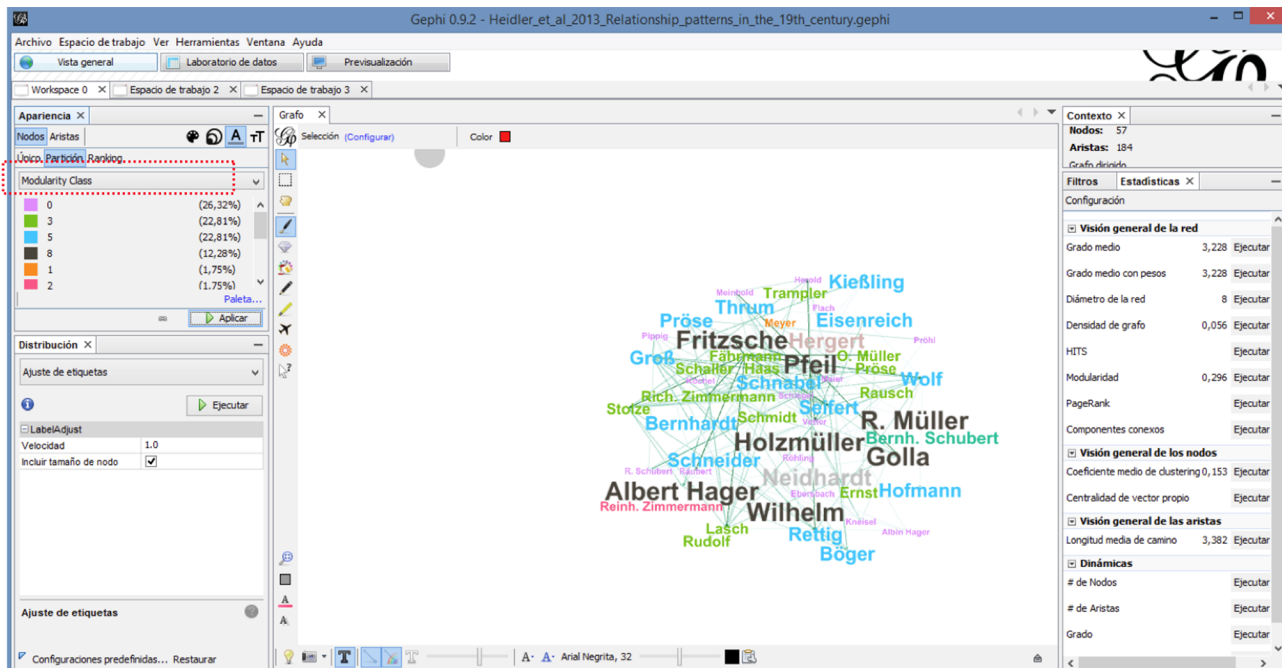
Figura 23



7) Partició

L'algorisme de detecció de la comunitat creat en **Modularity Class** dona valor a cada node. La pantalla de partició és a la part esquerra de la pantalla. Pot usar aquesta nova dada per pintar comunitats. Un cop seleccionat es prem «aplicar». A continuació veiem la llista de partició i les comunitats que trobem diferenciades per colors.

Figura 24



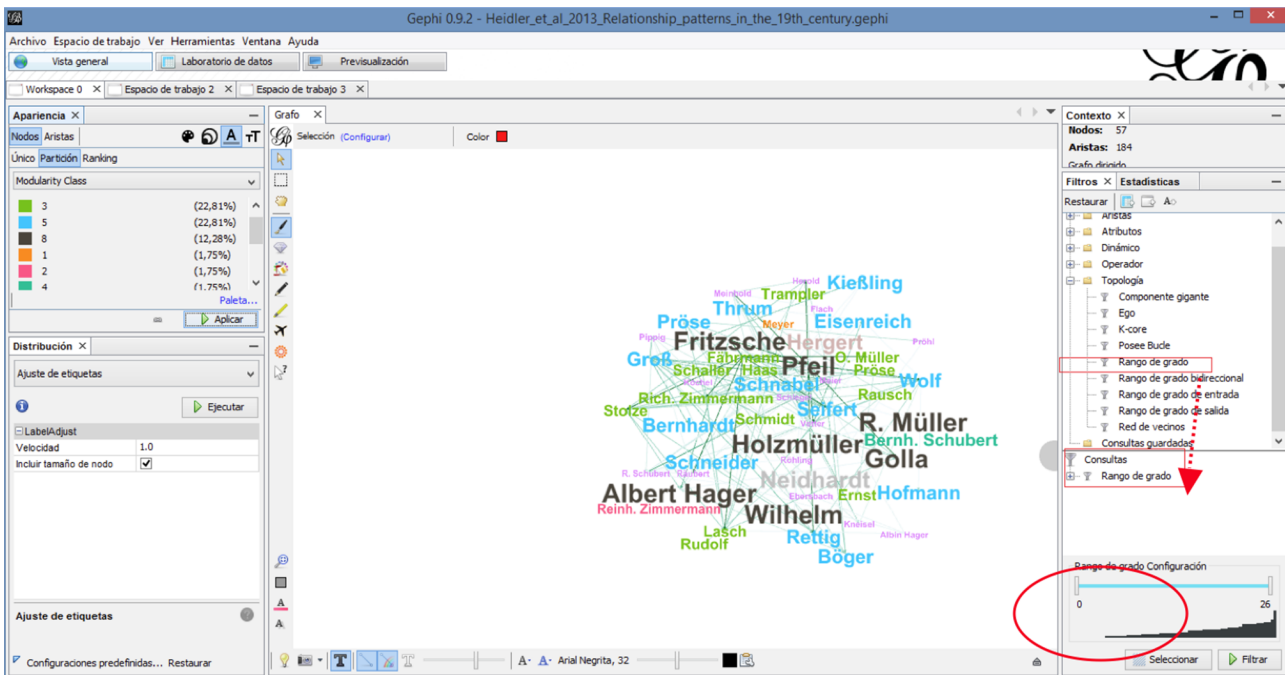
Si la xarxa es veu molt densa, perquè hi ha molts nodes i arestes, podem emprar els filtres que trobem a la part dreta de la pantalla, seleccionant «topologia». Ho veurem a continuació.

8) Filtració

Els filtres per amagar nodes o arestes els seleccionem mitjançant la funció «filtres» al costat de la pestanya d'estadístiques. Trobarem diferents categories en el directori de «topologia». Hem de moure el filtre que seleccionem cap a l'espai de baix. El movem amb el ratolí.

Seleccioneu «Rang de grau» a la categoria de topologia i arrossegueu-lo amb el ratolí a les consultes.

Figura 25



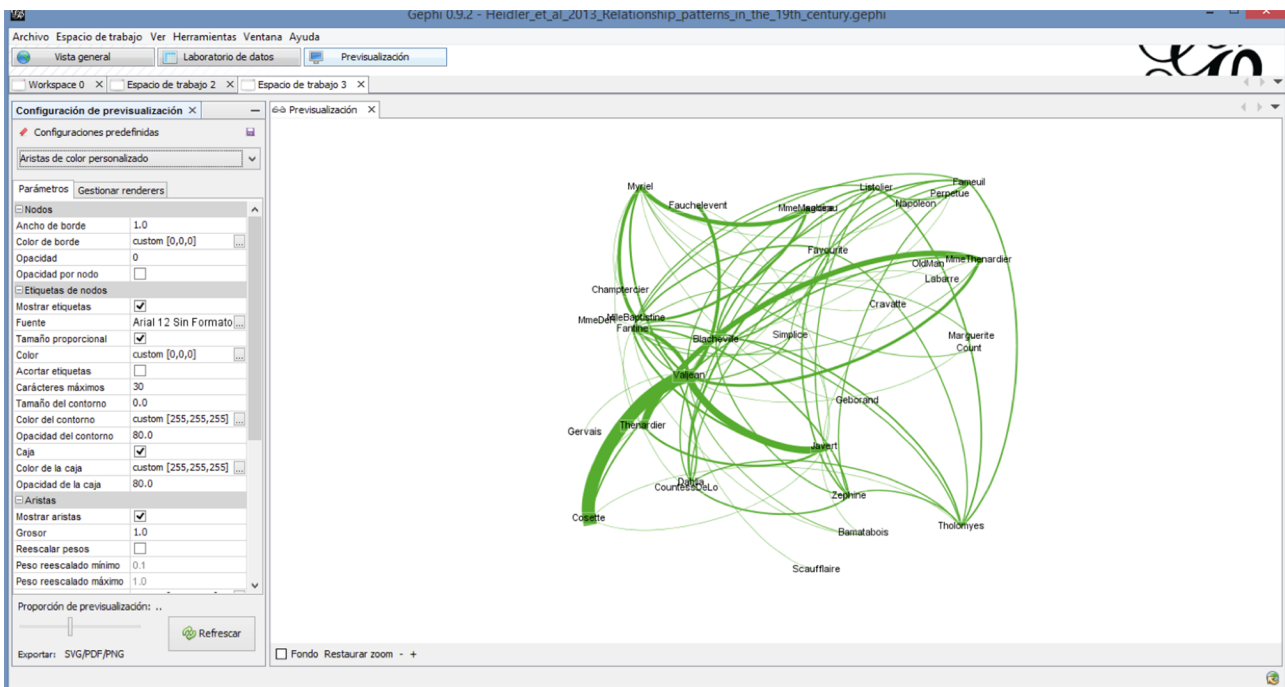
En el rang podem indicar el grau per sota del qual no volem que es mostrin els nodes. Amb el ratolí arrosseguem el límit, fins arribar al rang que volem definir.

Un cop indicats els valors, els colors i les etiquetes podem previsualitzar com queda la nostra xarxa. Finalment l'exportarem, com veurem a continuació.

2.1.5. Pas 5: previsualització

La pantalla «Previsualització» la trobem clicant la pestanya de la part superior esquerra. Allà seleccionem els paràmetres que considerem (proporció de la visualització, mostrar etiquetes, retallar etiquetes..., i podem fer els últims retocs al gràfic final. Quan hàgim acabat premem «refrescar» per a visualitzar la versió final.

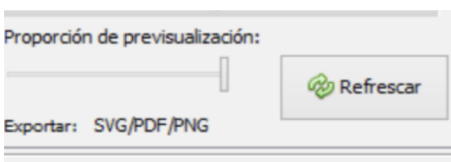
Figura 26



2.1.6. Pas 6: exportació

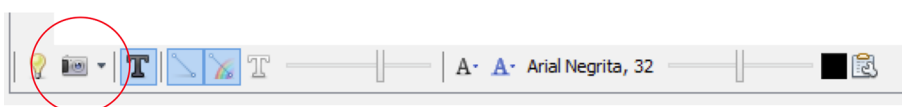
Podem exportar el gràfic resultant de la nostra xarxa en un fitxer SVG/PDF/PNG. Ho hem de fer des de la pantalla de previsualització. A la part inferior veiem «Exportar», hi cliquem i ja només ens falta posar-hi un nom i seleccionar el tipus de fitxer.

Figura 27



També podem fer captures de pantalla amb la icona d'un aparell de foto que hi ha a la barra inferior de la pantalla de Gephi.

Figura 28



2.2. Exemple d'anàlisi dels amics de Facebook amb Gephi.

Detecció de comunitats

Aquest exemple està basat en el model il·lustrat per Sherlock (2014). Serveix per veure la seqüencialitat de passos a fer, eines a emprar, i aspectes que es poden visualitzar quan som davant de la pantalla on tenim la representació d'una xarxa social.

2.2.1. Descàrrega i preparació de dades

Aquest exemple fa servir com a model la xarxa d'amics d'un usuari al Facebook. Per fer l'anàlisi, s'utilitza un fitxer GDF (Graph Data Format File) que conté una xarxa de relacions entre nodes. Un fitxer GDF és un com un fitxer CSV, però amb un format adequat per representar grafs. Per a crear un fitxer GDF de manera ràpida i senzilla es pot utilitzar l'aplicació de Netvizz de Facebook.

2.2.2. Visualització i anàlisi

Per visualitzar les dades, primer hem de carregar-les a Gephi, tal com hem vist anteriorment.

Quan visualitzem les xarxes podem trobar-hi diversitat de formes, i tots els seus components. A partir de la visualització podem fer una sèrie d'anàlisis. Aquest és el tema de l'assignatura Respondre preguntes de recerca analitzant xarxes. Tot i així, en aquest exemple avancem aspectes que la visualització ens indica. Aquí veurem els següents tres aspectes:

- com detectar comunitats,
- detalls dels membres i
- influents.

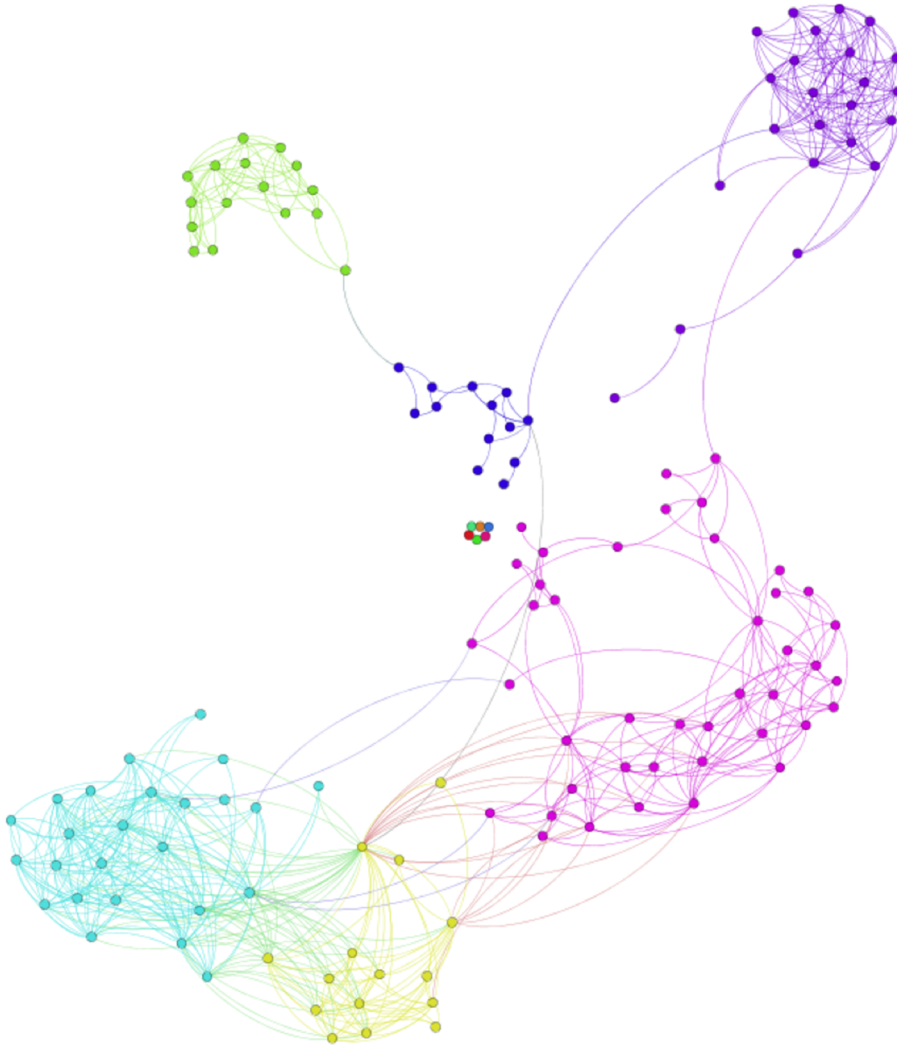
1) Detecció de comunitats

Per detectar comunitats d'amics, podem utilitzar diferents mètodes. En aquest exemple, utilitzarem el mètode de Louvain. Recordem que aquest mètode descompon la comunitat d'amics en comunitats secundàries més petites. La figura 29 mostra el gràfic de la xarxa d'amics de Facebook després d'executar l'algorisme. Cada petit punt del gràfic representa una persona i cada línia o aresta indica que aquestes persones són amigues entre elles. L'algorisme ha agrupat les persones en comunitats i s'ha codificat cada comunitat en el gràfic en un color depenent de la comunitat en què es troba la persona.

Vegeu també

La càrrega de dades a Gephi s'ha tractat al subapartat 2.1.

Figura 29. Detecció de la comunitat mitjançant l'aplicació de Gephi del mètode Louvain



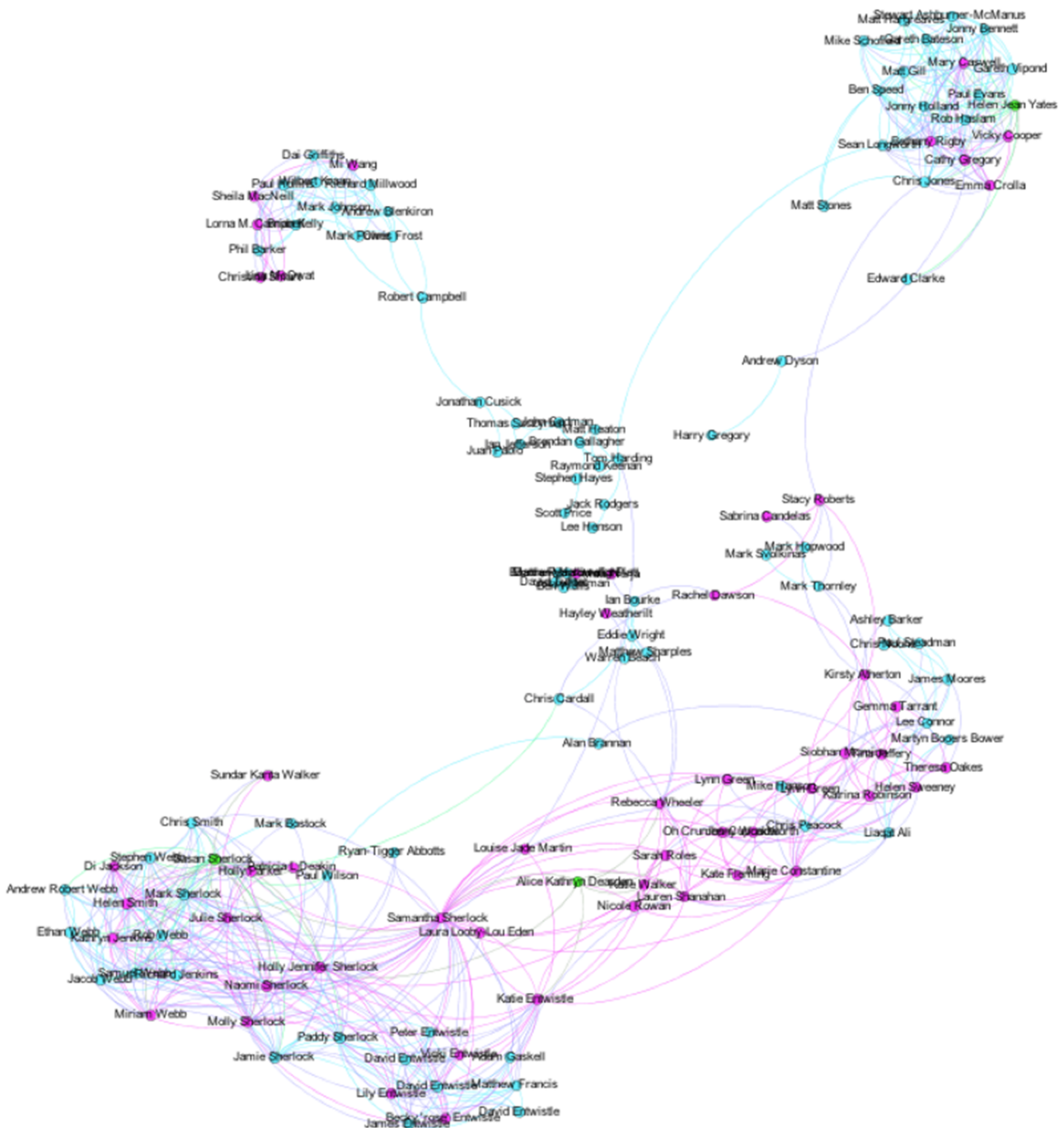
Font: Sherlock (2014).

En aquest exemple, sembla que hi ha quatre comunitats diferents amb diverses comunitats més petites, que semblen vincular les comunitats més grans. Podem veure de prop la pertinença de cada membre a un grup diferent. En aquest cas, els quatre grups principals són: amics de secundària, familiars, amics del treball actual i amics d'un treball anterior. El petit grup al mig que els uneix està majoritàriament format per amics de l'escola primària, d'una àrea local o universitat. També podem explorar aquestes persones examinant-ne els noms i el sexe.

2) Detall dels components

Podem visualitzar també els noms i el sexe dels components. En el gràfic, s'hi poden afegir els noms dels amics i també es pot codificar el color per gènere (rosa per nena, blau per nen i verd per a persones que no tenen el gènere indicat al perfil de Facebook) (figura 30).

Figura 30. Noies i nois a la xarxa d'amics d'un usuari al grup de Facebook, amb noms



Font: Sherlock (2014).

Respecte a aquesta figura 30, és interessant de ressenyar que en el centre hi ha una àrea que inclou només nois. Això pot ser per diverses raons. Una d'elles pot ser que es tracti d'amics de la universitat, on l'usuari de Facebook va fer uns estudis, relacionats amb una disciplina molt dominada per homes.

3) Influència

Un altre aspecte que es pot veure quan analitzem la xarxa és el de la **influència**. Podem descobrir qui és influent, i trobar l'influent entre diferents comunitats i amics. Per fer-ho, hem d'aplicar les **mesures de centralitat** entre

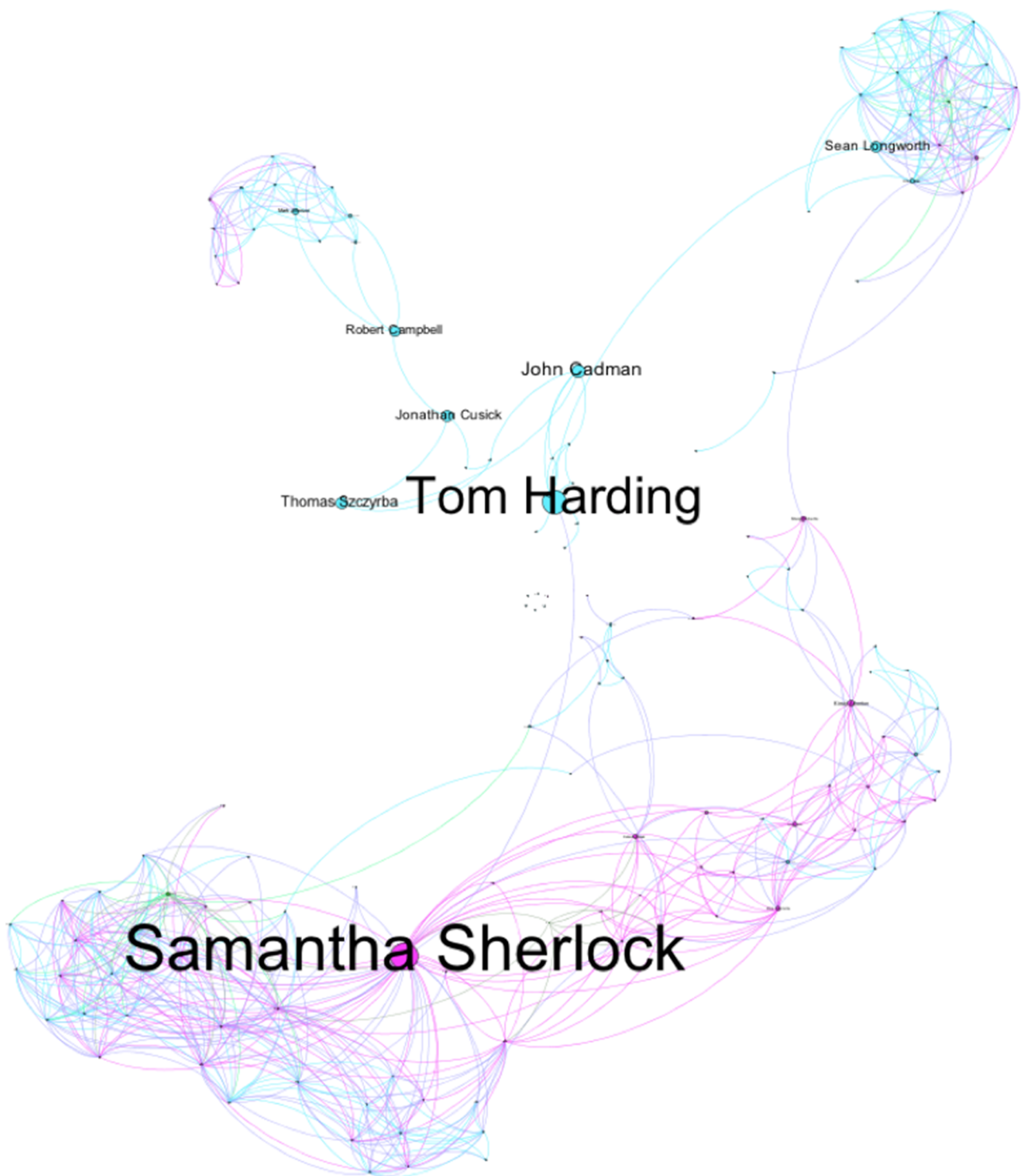
totes les persones de la xarxa. Les mesures de centralitat es refereixen a un node, al lloc que ocupa en la xarxa, i per tant, té un valor estructural. Hi ha diferents mètriques de centralitat, unes més senzilles com el **grau** que representa el nombre de connexions d'un node (com hem vist abans) i d'altres més complexes, com la *betweenness centrality*).

La *betweenness centrality* (o intermediació) mesura el nombre de **camins** més curts entre dos nodes que passen per un node concret. Els nodes amb alta centralitat són influents perquè connecten àrees diferents de la xarxa, i es troben al centre de molts grups de relacions. La figura 31 mostra les etiquetes i els nodes ponderats de manera que es mostren els noms dels portaveus més rellevants.

Vídeo recomanat

Podem veure en la següent referència un vídeo d'exploració dels amics de Facebook que ofereix Sherlock (2014): <https://youtu.be/OgaoS7sUGZI>.

Figura 31. Mesures de centralitat entre interiors



Font: Sherlock (2014).

2.3. Exemple: visualització de la xarxa d'interaccions al voltant d'un *hashtag* a Twitter

Aquest exemple està basat en un *hashtag* de Twitter, «#LasInvisibles» llançat durant un programa de televisió (*Salvados* de l'11 de març de 2018). Veurem les interaccions que s'han produït al voltant del *hashtag*. Ens servirà per estudiar-ne la centralitat, els detalls dels nodes, les comunitats d'usuaris, com podem millorar la visualització i l'exportació de la imatge.

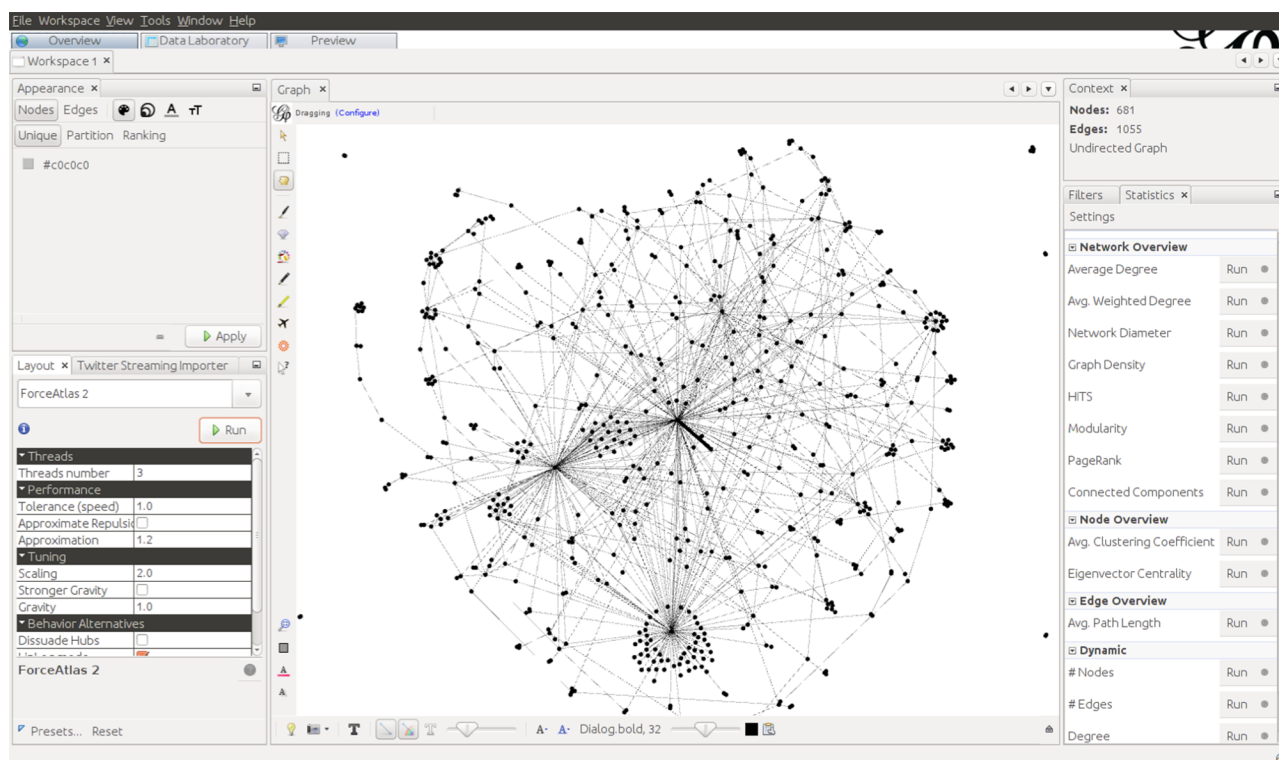
2.3.1. Càrrega d'arxius

Ara mirarem un altre exemple basat en dades de Twitter. Farem servir l'arxiu generat a l'assignatura Big data i anàlisi de xarxes socials: conceptes i eines amb l'eina Flocker, en aquest cas la xarxa d'interaccions al voltant del *hashtag* «#LasInvisibles».

Quan obrim el fitxer, veiem nodes i arestes acumulades i solapades de manera desordenada. Per a canviar la disposició dels nodes a l'espai fem servir un algorisme de *layout*, en aquest cas seleccionem ForceAtlas2 (trobem una llista d'algorismes de *layout* al panell de sota a l'esquerra).

L'algorisme ForceAtlas2 simula un sistema de forces físiques, en què els nodes que estan connectats s'atrauen entre ells, i els que no estan connectats es repel·leixen. A més, hi ha una força de gravetat que atreu tots els nodes cap al centre. Així, l'algorisme va disposant els nodes de tal manera que els que estan connectats entre ells són a prop, i es pot observar de manera visual com emergeixen diferents comunitats d'usuaris. Marquem l'opció «LinLog mode» amb la qual la força de repulsió és logarítmica, i per conseqüència els nodes es distribueixen més per l'espai.

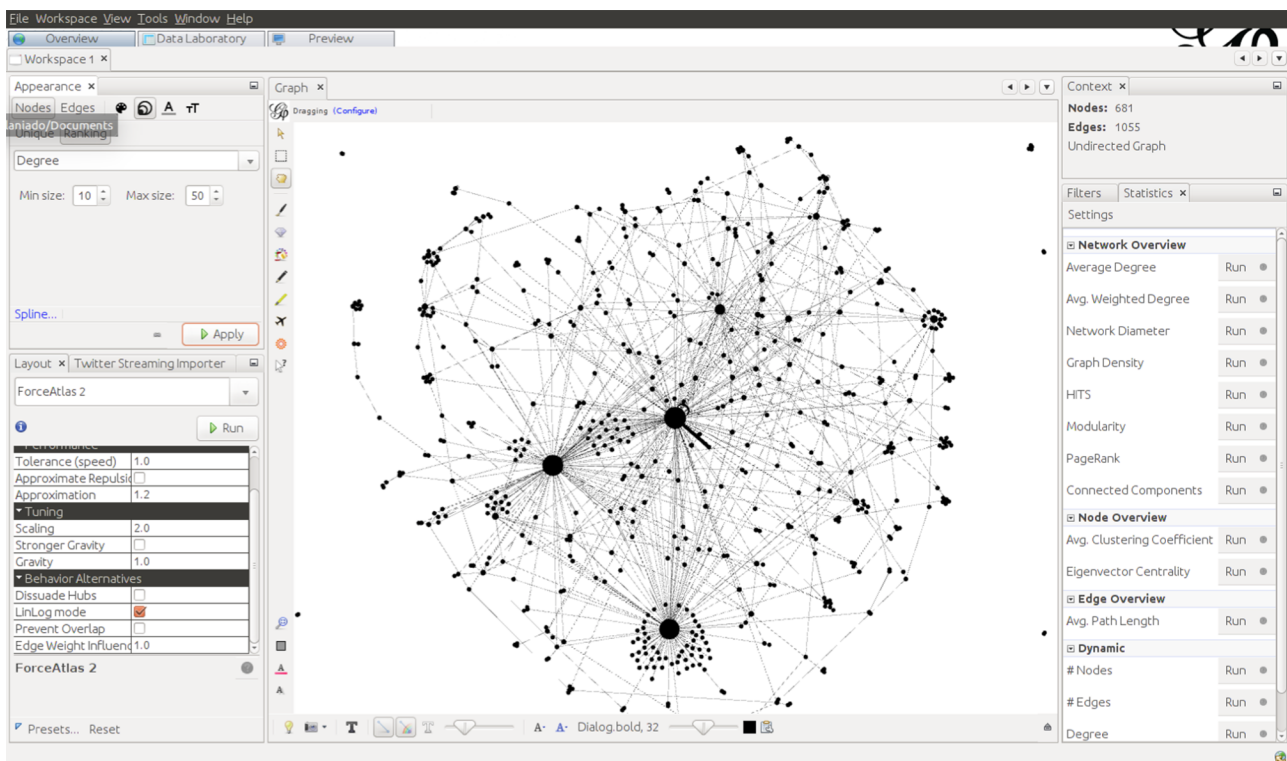
Figura 32



2.3.2. Centralitat

Ara canviarem la mida dels nodes. Al panell «Appearance», a dalt a l'esquerra, seleccionem «Nodes» i després «Size» (té la icona amb cercles de diferents mides). Ara està seleccionada la opció «Unique», que fa que tots els nodes tinguin la mateixa mida. Seleccionem «Ranking» i aquí podem escollir una mida pels nodes, en un menú desplegable, segons quin atribut vulguem assignar. Seleccionem «Grau» per a destacar els nodes amb més connexions. Juguem amb el valor mínim i màxim i escollim, en aquest cas, 10 i 50. A baix se'n pot veure el resultat, on s'observa que la xarxa es desenvolupa al voltant d'alguns nodes molt centrals.

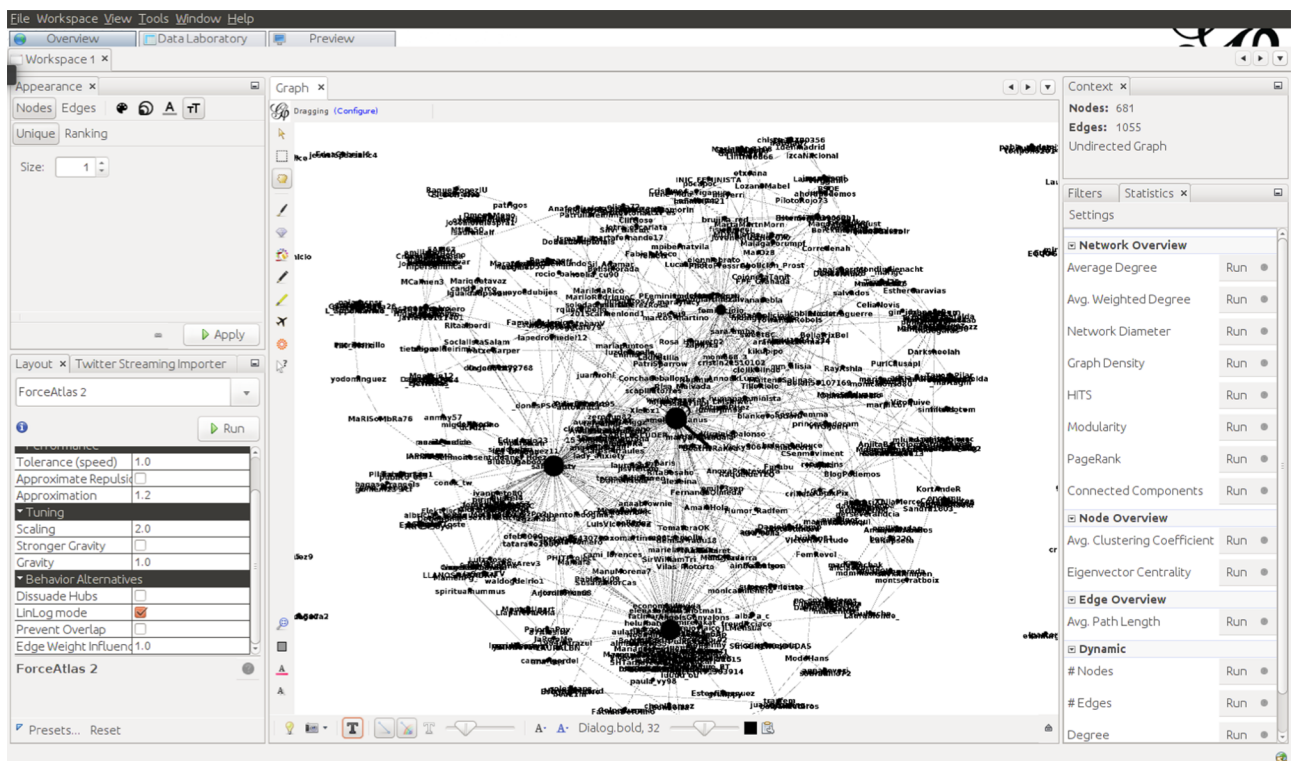
Figura 33



2.3.3. Detalls dels nodes

Ara, per a veure quins són aquests nodes, seleccionem l'opció «Show node labels» (la T negra a baix a la esquerra).

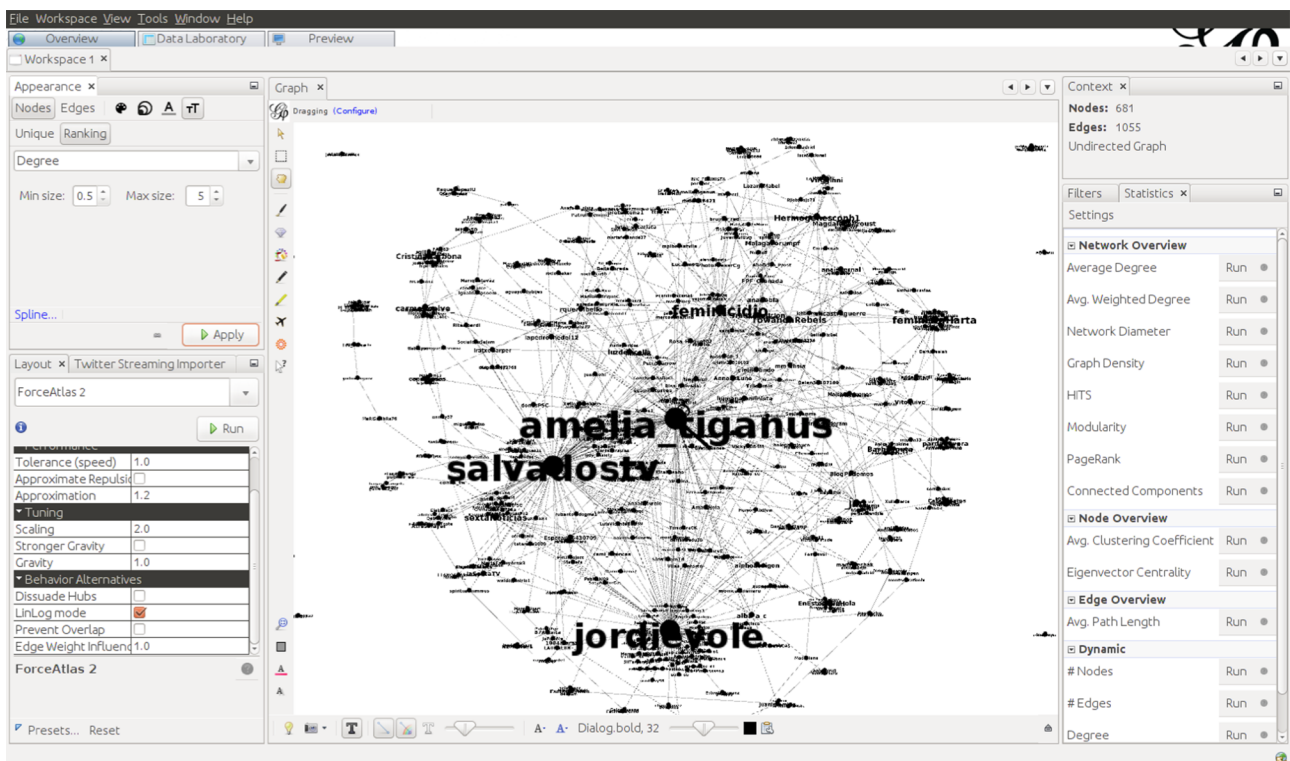
Figura 34



Com es veu a dalt, hi ha moltes etiquetes solapades. Per tal que destaquin les etiquetes dels nodes més centrals, anem de nou al panell «Appearance» - «Nodes» que trobarem a dalt a l'esquerra, i seleccionem l'última opció de la dreta «Label Size» (la icona amb una T gran i una petita). Com en el cas de la mida dels nodes, escollim que la mida depengui d'algun «Ranking» dels nodes, i seleccionem com a atribut pel rànkung el «Degree». Fixem de nou un valor mínim i un de màxim, en aquest cas 0.5 i 5.

Com es pot veure a baix, d'aquesta manera podem veure els nodes principals de la conversa: @salvadostv (el programa que va llançar l'hashtag), @jordievole (el conductor del programa) i @amelia_tiganus (una activista feminista).

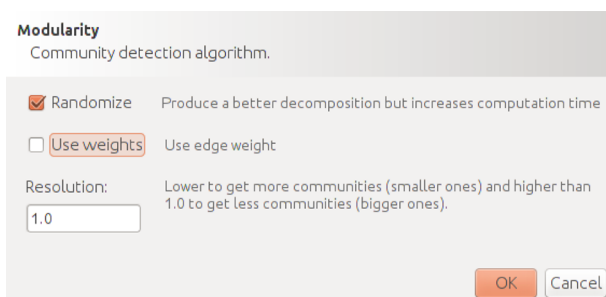
Figura 35



2.3.4. Comunitats d'usuari

Ara, per a detectar comunitats d'usuari, anem al panell de la dreta, a «Estatística» i «Network overview» i hi llancem l'algorisme «Modularity». És una implementació del mètode Louvain per a la detecció de comunitats a través de la maximització de la modularitat, com hem vist a dalt. Deixem els valors per defecte i llancem l'algorisme.

Figura 36

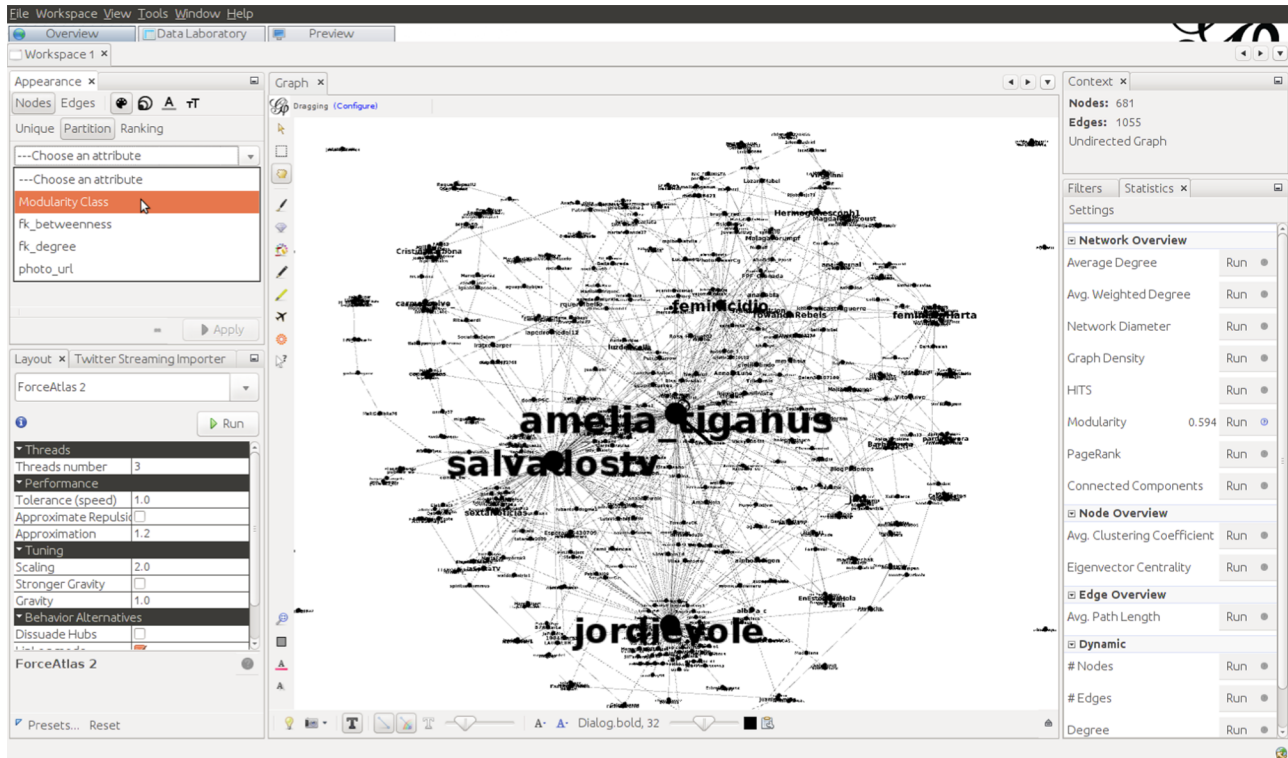


Ara que hem fet córrer aquest algorisme, ens apareixen unes estadístiques sobre els resultats, i el programa afegeix aquests resultats a les mètriques que podem fer servir a la visualització.

Concretament, volem pintar els nodes de colors diferents segons la comunitat a la qual pertanyen. Anem de nou a «Appearance» - «Nodes», i en seleccionem la primera opció «Color» (amb la icona d'una paleta). Ara escollim «Partition»

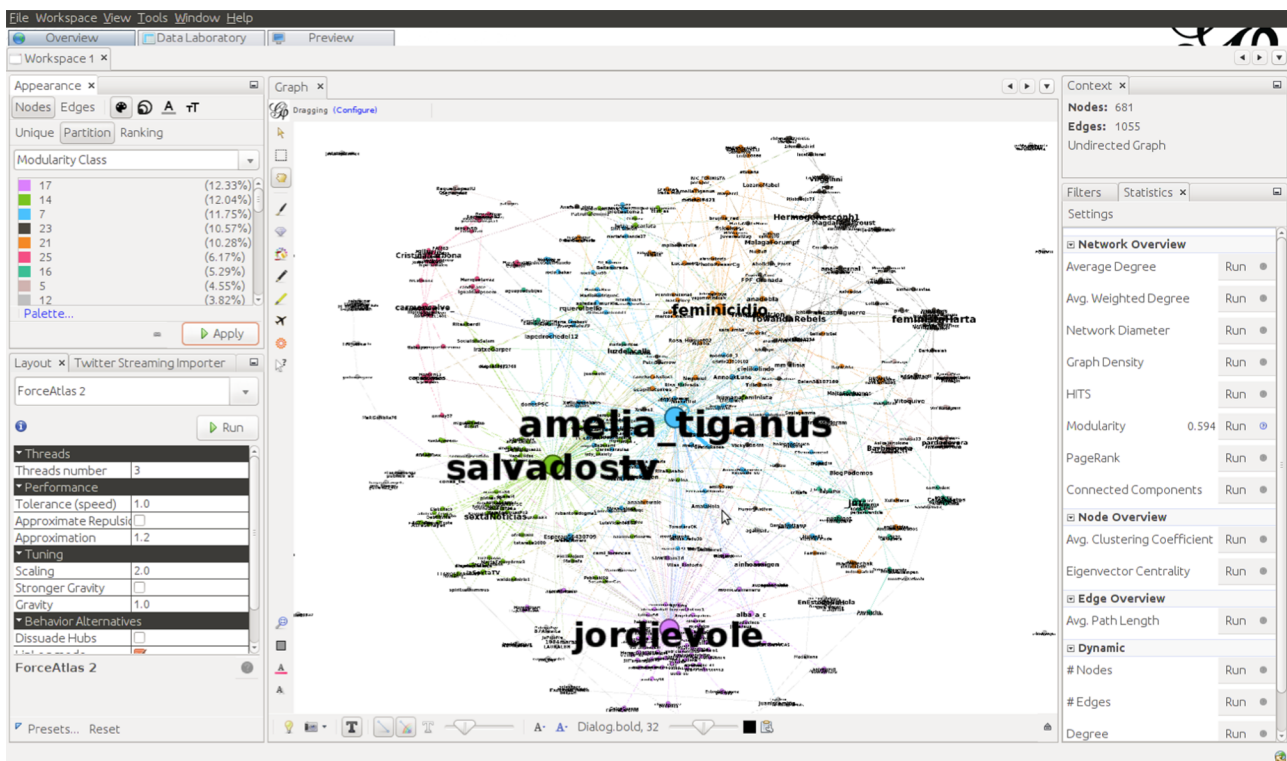
per a assignar un color a cada partició de nodes, i al menú desplegable seleccionem com a atribut per a definir les particions «Modularity Class», com es veu a la imatge següent.

Figura 37



Obtenim el resultat que es pot veure a continuació: els colors ensenyen les particions en què l'algorisme Louvain ha dividit els nodes. S'observen tres particions al voltant dels tres nodes més centrals, i algunes altres particions. Els nodes del mateix color tendeixen a ocupar la mateixa àrea, ja que l'algorisme ForceAtlas2 també tendeix a posar junts els grups de nodes que estan connectats entre ells. En altres paraules, tenim ara dos criteris diferents per a observar els grups (o comunitats) de nodes a la conversa: la disposició a l'espai, segons l'algorisme de Layout ForceAtlas2, i el color dels nodes, segons el mètode de Louvain per a detectar comunitats.

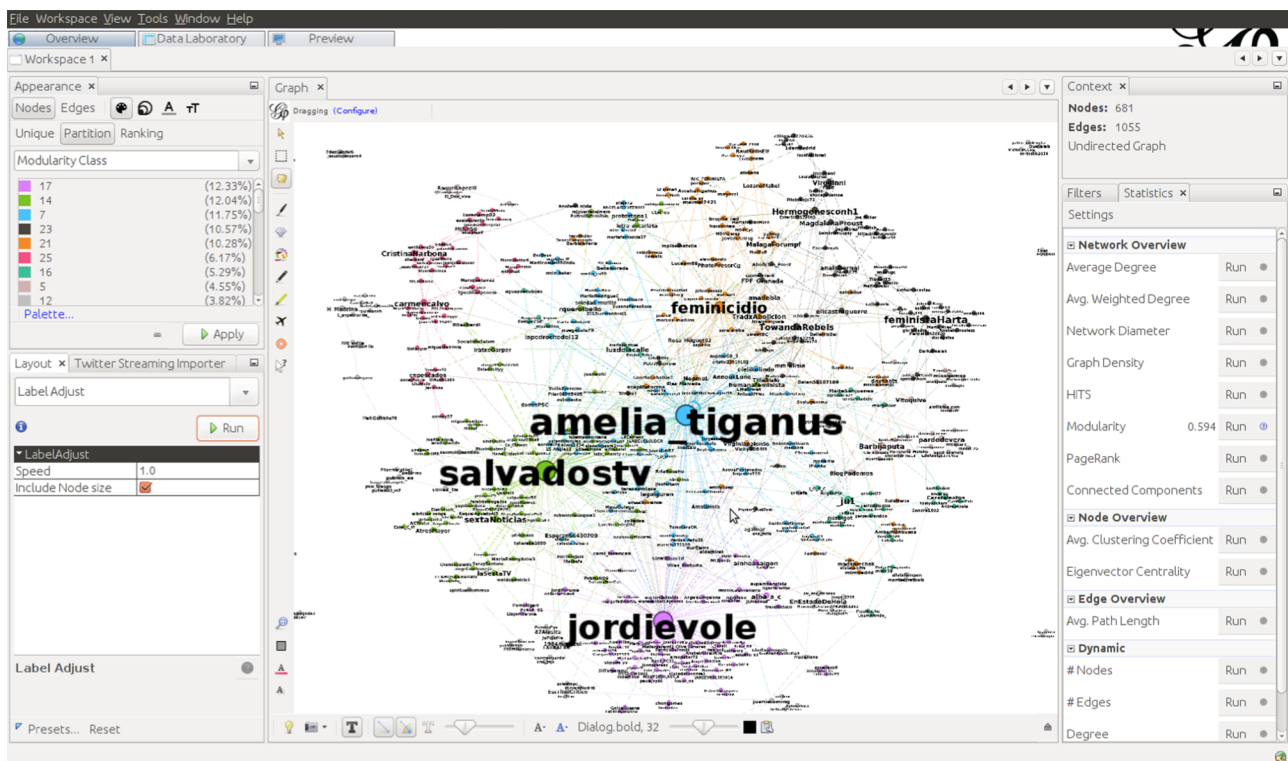
Figura 38



2.3.5. Millora de la visualització

Encara podem millorar una mica més la visualització fent servir un altre algorisme de *Layout*: «LabelAdjust», que efectua petits moviments per a evitar que les etiquetes dels nodes es solapin. Com es veu a la imatge següent, la posició dels nodes canvia molt poc, però permet que no hi hagin solapaments d'etiquetes i es puguin llegir millor els noms dels nodes.

Figura 39



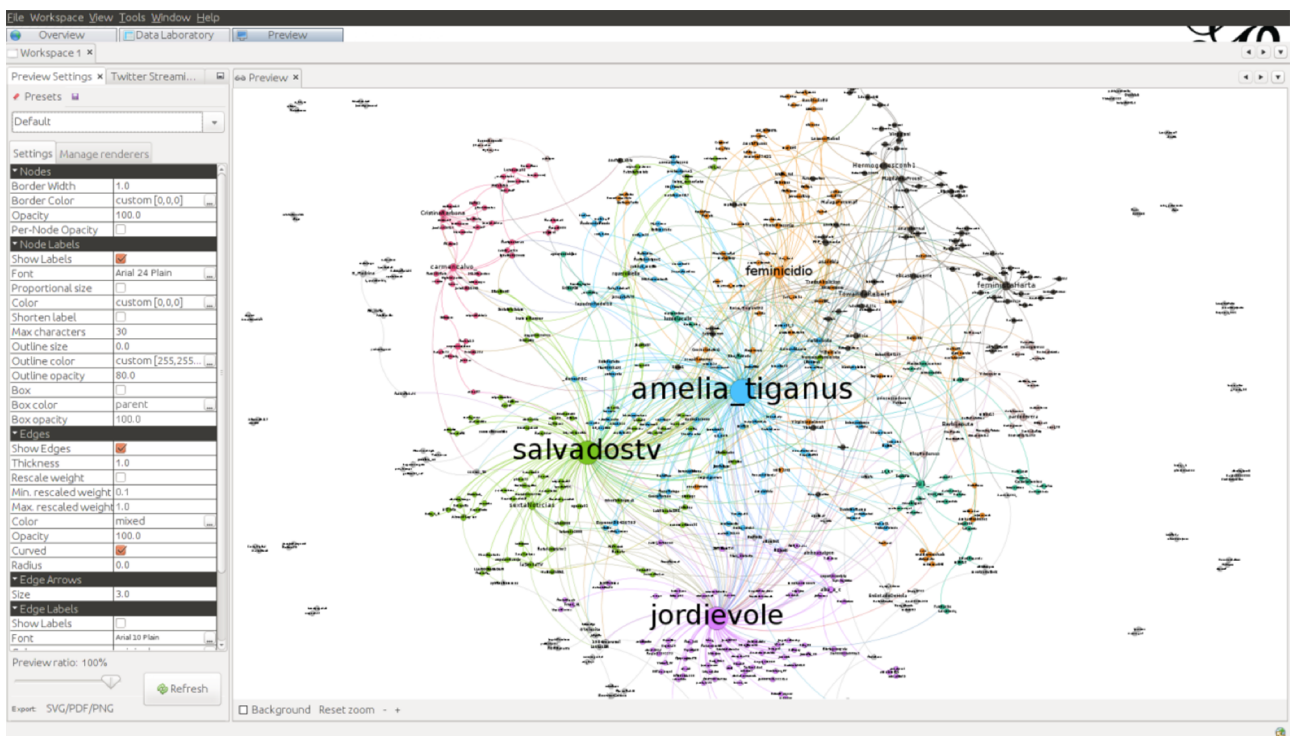
2.3.6. Exportació de la imatge

Ara que hem finalitzat la visualització de la xarxa, anem al tab «Preview», per a exportar-la com a imatge. Aquí podem tocar varies opcions per a millorar l'aspecte de la visualització, fent «Refresh» cada vegada per a veure el resultat dels canvis.

Com que és una xarxa dirigida, posem les arestes corbades («edges» - «curved»). Aquesta és una convenció bastant establerta per a visualitzar xarxes: en el cas d'una xarxa no dirigida hi ha suficients arestes rectes, mentre que en el cas d'una xarxa dirigida, les arestes corbades permeten indicar la direcció d'un enllaç. Per convenció es segueixen les arestes sempre en sentit horari. Així, si entre dos nodes hi ha un cicle, arestes corbades en els dos sentits, es tracta d'una connexió recíproca, si no la corba indica el sentit de la connexió unidireccional.

A «Node_labels» marquem «Show_labels» per a ensenyar les etiquetes. Es poden posar les etiquetes amb mida proporcional a la mida dels nodes, però en el nostre cas ja havíem posat les etiquetes com a proporcionals al grau, així que per a evitar que s'apliqui dues vegades la mateixa transformació, especifiquem que les etiquetes no siguin proporcionals (desseleccionem l'opció «Proportional size»).

Figura 40



Ara podem exportar el graf com a imatge png o svg, o com a pdf.

3. Visualització de la difusió d'informació: viralitat

Una de les aplicacions de la visualització és identificar el recorregut d'un contingut en línia per veure'n el potencial viral i el recorregut.

La informació a les xarxes socials es difon entre una audiència de volum i perfils diversos. Per tal que el contingut que volem difondre arribi a un gran nombre de persones, ha de produir-se una disseminació ràpida i àmplia, com la propagació dels virus. La comunicació esdevé viral quan s'activa el procés de «contagi» (Sivera, 2014) i per això cal la confluència d'una sèrie de factors. Peralta (2016) ens recorda que el contingut viral en xarxes socials es caracteritza per propagar-se gairebé sense control, gràcies als usuaris que ho comparteixen moguts per algun tipus d'emoció. Per tant, perquè els continguts en les xarxes socials es facin virals calen un seguit de circumstàncies que afavoreixin la difusió dels continguts fins que aquests es facin virals.

A un professional de màrqueting en línia, no només li interessa crear contingut viral en xarxes socials sinó trobar la fórmula mitjançant la qual es pot fer arribar un missatge al major nombre possible de persones i a ser possible de manera orgànica, és a dir, sense recórrer al tràfic de pagament de les xarxes socials (Peralta, 2016).

Tsvetovat i Kouznetsov (2011) senyalen que no hi ha cap recepta preparada per impulsar la difusió d'un contingut. Tanmateix, podem comprendre el que n'impulsa la difusió i com les nostres decisions poden impulsar-ho. Aquests autors fan recomanacions des del punt de vista dels elements que integren les estructures de les xarxes socials.

De la mateixa manera, la investigadora Sivera (Bigas, 2015) fruit de la seva recerca apunta algunes condicions per a tenir èxit en les campanyes publicitàries en la xarxa, i fer-ne créixer la capacitat viral. Aquesta autora es refereix als components que han de tenir els continguts.

En definitiva, part dels ingredients per aconseguir un contingut viral està constituït per elements estructurals de la teoria de les xarxes socials i components dels missatges espontanis, privats o publicitaris.

Tot i que no hi ha receptes màgiques, en aquest apartat en primer lloc veurem alguns dels aspectes de les estructures i mètriques de les xarxes que contribueixen que un contingut sigui viral: centralitat, mencions, mems, massa crítica, i saturació són alguns dels conceptes que hem de tenir en compte si volem visualitzar la difusió de la informació a les xarxes socials. En segon lloc, veurem exemples de contingut viral de tres tipus.

3.1. Divulgació per xarxes

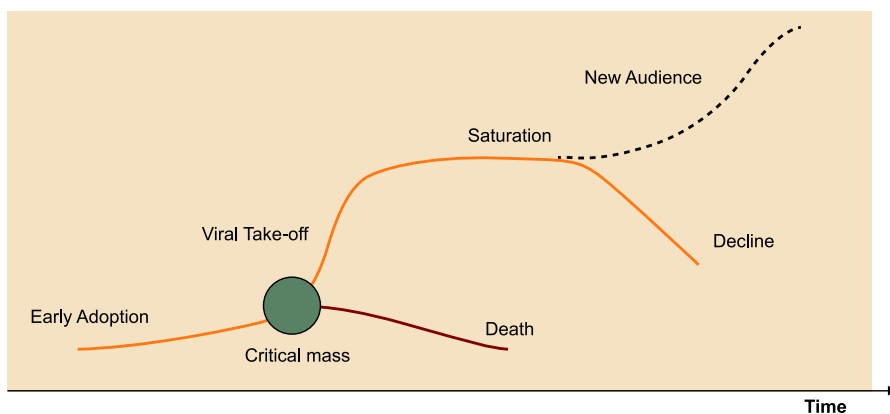
En aquest apartat ens centrem en aspectes de xarxes socials, veient el que ens diuen Tsvetovat i Kouznetsov (2011). Aquests autors il·lustren la viralitat amb un exemple: la publicació d'un vídeo a YouTube i la seva difusió per la xarxa personal. Aquest vídeo té un potencial per fer-se viral, i es farà viral depenent en gran mesura de qui el veu i què fan amb ell després. Al principi, les úniques persones que el veuen són a la xarxa d'ego immediata. Per tant, el creixement del nombre de punts de vista és molt lent i lineal en el temps. Al final, el nombre total de visualitzacions que obtindrà el vídeo estarà relacionat matemàticament amb el nombre de seguidors que té la xarxa pròpia on s'ha difós: el **grau de centralitat**.

A més de la centralitat, hi ha altres elements que impulsen la divulgació. Una d'elles és que pugui passar a una altra xarxa. Els amics de la meua xarxa comparteixen el vídeo als seus amics (i així successivament, exposant-hi més i més persones), el vídeo serà el més popular de YouTube.

Una altra cosa que pot passar és el de les **mencions** i la divulgació de comentaris que reforcin la idea de mencions, i fins i tot, que es converteixi en un **mem**. Per exemple, en algun lloc del centre del procés, un dels meus amics diu «has vist aquest vídeo?» I l'altre respon «sí, ja el tinc», i això, fins i tot, pot reforçar la idea que el vídeo constitueix un mem, un nou artefacte cultural que s'aixeca. De sobte, exposar els altres al mem es converteix en una segona naturalesa o fins i tot en una acció obligatòria.

Perquè el vídeo acabi sent viral, el mem ha d'aconseguir arribar a una **massa crítica** en un cert temps, si no la seva taxa de compartició comença a caure i, finalment, mor. En canvi si s'aconsegueix una massa crítica, creixerà exponencialment fins que arribi al punt de saturació, és a dir, quasi tots els que podrien haver adquirit el mem ja ho han fet. A partir d'aquest moment, el mem disminuirà. Tsvetovat i Kouznetsov (2011) ho il·lustren a la figura 41.

Figura 41. Corba de la difusió: massa crítica, boom i davallada



Font: Tsvetovat i Kouznetsov (2011).

Una nota important és que les nocions de **massa crítica** i **saturació** s'apliquen a cada xarxa o comunitat: és possible que un mem arribi al punt de saturació d'una comunitat i que es desconegui pràcticament en un altra comunitat. No obstant això, atès que les comunitats estan interrelacionades, és possible que un mem surti a una nova comunitat i trobi una nova massa crítica i un punt de saturació encara més gran (Tsvetovat i Kouznetsov, 2011).

Tsvetovat i Kouznetsov (2011) ens recorden la importància de la saturació i massa crítica en l'estratègia que varen seguir els seus fundadors en l'origen i extensió de Facebook. Quan el 2003 es va llançar Facebook (sota el nom original de «Facemash»), aquest es va limitar a una petita comunitat densa -graduats de Harvard. Llançar-ho en un espai reduït els va permetre arribar ràpidament a la massa crítica. De fet, en aproximadament quatre hores des del seu llançament, el lloc va atreure quatre-cents cinquanta visitants, o al voltant del 6% de la població de pregrau total. Recordeu la dada 6%, és important. Dins de Harvard, a principis de Facebook van passar-hi una sèrie d'iteracions fins arribar a un disseny bastant escalable i, finalment, va assolir la saturació al voltant del 50% de les infermeres de Harvard. A partir d'aquí, el seu abast es va ampliar a altres escoles de la Ivy, una per una, i finalment a totes les universitats dels EUA. Finalment, el 2005, es va obrir als estudiants de secundària i es va establir l'escenari de la dominació mundial.

La regla que Facebook va seguir era aconseguir una saturació dins d'una comunitat abans de passar a una comunitat més gran. D'aquesta manera, mai no es va perdre la massa crítica i els nous membres es van adaptar perfectament a l'estructura social creada per companys d'altres escoles. Aquesta regla va contribuir a fer que Facebook fos diferent d'altres xarxes socials anteriors com SixDegrees i Friendster.

Com ens recorden Tsvetovat i Kouznetsov (2011), la **massa crítica** es pot estimar mesurant la probabilitat que un enllaç creat aleatòriament d'A a B formarà una o més tríades obertes amb altres nodes. Això és proporcional al nombre de nodes ja connectats (es duplica el nombre de nodes):

$$P(\text{tríada oberta}) \sim \# \text{ ties} / (2 * \# \text{ nodes})$$

D'aquesta manera, la probabilitat d'una tríada oberta arriba al 50% quan es connecten una quarta part dels nodes, i cada nova connexió fa que les regles de tancament triàdiques generin encara més connexions, augmentant encara més la probabilitat d'una cascada (figura 41).

En els experiments fets per Tsvetovat i Kouznetsov (2011), aquests autors han detectat que la transició del creixement lineal (on les connexions s'afegeixen una per una) per esdevenir viral ocorre en algun lloc al voltant del 7% de den-

sitat –i, si el 7% de les persones que hi assisteixen admeten un mem, repiulen un vídeo, s'uneixen a un lloc, etc., la resta també ho farà ràpidament en una ona viral.

En definitiva, quan visualitzem una xarxa social hem de tenir en compte i intentar visualitzar les xarxes de components com «m'agrada», mencions, re-enviaments, etc.

3.2. Tipus de contingut viral

Peralta (2016), per la seva experiència, classifica el contingut viral en xarxes socials en tres tipus: espontani, estratègic i circumstancial. Vegem-ne aquests tres tipus, a partir dels exemples que ens descriu.

3.2.1. Contingut viral espontani

El contingut viral espontani és aquell que es difon de forma orgànica sense que hi hagi hagut una intenció prèvia de fer-lo arribar a tantes persones. Compartim alguna cosa en les nostres xarxes socials perquè ho puguin veure els nostres contactes. Pot passar que aquests ho difonguin entre els seus contactes, i per tant la difusió s'amplia, sense que ho haguem planificat.

Ens agrada rebre «m'agrada», favorits o *retweets* com a reconeixement de la nostra comunitat, ja que d'aquesta manera podem pensar que el contingut és interessant, divertit i/o útil per a ells.

3.2.2. Contingut viral estratègic

Quan el contingut que difonem ha estat pensat i planificat perquè arribi el més lluny possible, el contingut serà estratègic. Per tant, la difusió està més o menys calculada. Serà viral depenent del volum d'audiència al qual arribi.

Peralta (2016) ens explica un exemple de contingut i viralitat planificada, per quan vulguem denunciar alguna cosa i busquem alhora arribar a molta gent, perquè la companyia reaccionï o per alertar altres usuaris de les males pràctiques. Això va ser precisament el que va succeir el desembre de 2013 quan l'empresa Vueling va fer pagar a un viatger un seient per al seu violí, si no el volia facturar. La companyia li va fer pagar una quantitat que considerava abusiva per portar el violí «assegut» al seu costat si no volia facturar-lo.

Figura 42



El viatger va dissenyar una estratègia per denunciar el seu cas intentant arribar al major nombre de persones possibles. L'estratègia era més complexa que publicar un senzill *post* a Facebook. En aquesta ocasió, el seu blog va ser fonamental i les xarxes socials van complir un paper primordial. El sentiment generat en els lectors del seu cas va ser el detonant d'aquesta viralitat estratègica. En menys de vint-i-quatre hores, l'article va rebre més de 25.000 visites en el blog i només a Facebook va ser compartit més de 6.900 vegades.

3.2.3. Contingut viral circumstancial

Aquest tipus de contingut estaria classificat entre els dos anteriors. En aquest cas es busca d'una manera calculada, però no premeditada, aprofitar una circumstància concreta per intentar arribar a tanta gent com sigui possible.

El contingut viral circumstancial acostuma a ser producte d'un moment concret en què es donen unes circumstàncies impredecibles, però idònies, que es poden utilitzar de forma conscient per aconseguir el major nombre d'usuaris.

Pot tractar-se d'un esdeveniment amb gran difusió que tracti el tema del contingut que volem viralitzar, un *hashtag* que és *trending topic*, una persona influent que es fixa en el seu contingut i ho difon, etc.

Peralta (2016) ens dona l'exemple d'una piulada que va enviar una tarda qualsevol mentre estava ordenant uns llibres al seu estudi. Per casualitat, consultant les seves llistes de Twitter, l'autor va veure que Arturo Pérez-Reverte estava

tuitejant, i va aprofitar per fer una piulada amb una foto dels seus llibres, entre els quals n’hi havia un de Arturo Pérez-Reverte. L’autor era un lector habitual de les seves obres i de vegades seguia també les seves aparicions explosives a Twitter tractant tot tipus de temes. És per això que va intentar experimentar aprofitant l’ocasió. Va fer una foto dels seus llibres i va enviar la següent piulada:

Figura 43



En poc menys de trenta minuts, i gràcies al fet que el mateix Pérez-Reverte va retuitar la foto. Els resultats de la piulada van ser els que es veuen a continuació:

Figura 44

Actividad de Tweets

David Peralta @peraltadavid
 Terminado el último libro de @perezreverte. Ocupa ya su sitio junto a los demás #leer #recomendado
pic.twitter.com/NtK0cPbUsV

Promocione su Tweet
 Su Tweet tiene 4.238 interacciones en total hasta ahora.
 ¡Consiga más interacciones con este Tweet!

Promocione su Tweet

Impresiones	67.597
Interacciones totales	4.238
Interacciones con el contenido multimedia	2.819
Abrir el detalle	704
Clics en el enlace	289
Clics en el perfil	249
Me gusta	101
Retweets	26
Clics en la etiqueta	24
Respuestas	16
Seguimientos	10

4. Altres eines per visualitzar dades de Facebook

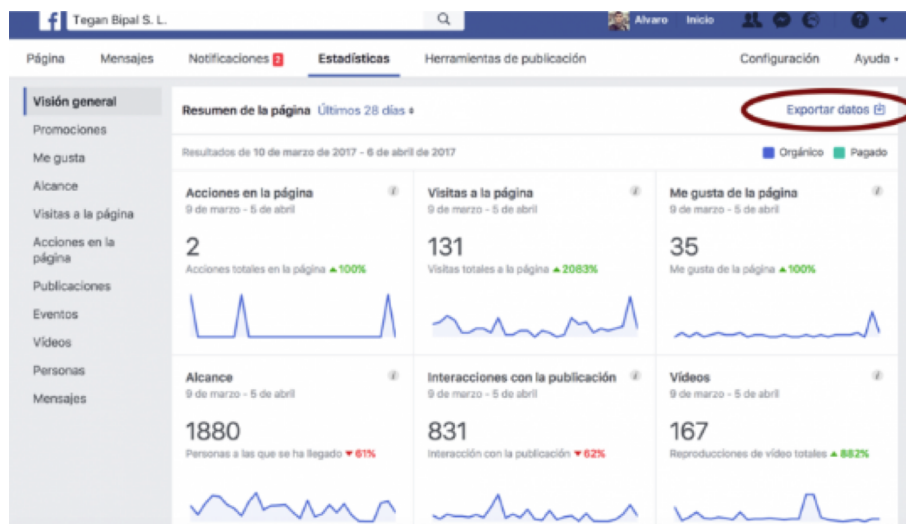
En aquest apartat presentem diverses eines per visualitzar dades de Facebook. En primer lloc, estudiarem **Facebook Insights** que presenta més facilitat per a l'usuari sense coneixement informàtics ni de programació. En segon lloc, descrivim de forma ràpida les prestacions que ofereix **RFacebook**. Aquest últim és més complex d'emprar, atès que requereix coneixements previs de programació.

4.1. Facebook Insights

És una eina que aquesta xarxa social posa a disposició dels administradors de perfils per fer un seguiment de la interacció dels usuaris amb un compte o perfil de Facebook, i analitzar les estadístiques. Això ajuda a comprendre millor el rendiment de la pàgina.

Veiem les estadístiques, clicant a l'opció d'estadístiques des de la nostra pàgina de Facebook. Podem veure els totals de continguts, o seleccionar per períodes, increment d'usuaris, perfils usuaris, etc. (figura 45).

Figura 45. Finestra d'estadístiques



Font: Mariscal (2017).

Des d'aquest menú podem descarregar les dades de tres tipus: pàgines (accés a fonts i públics de la pàgina de fans), publicacions que hem fet (abast, orgànic, impressions i comentaris), i dades de vídeos concrets i reproduccions totals (Mariscal, 2017).

Per a més informació sobre totes les prestacions i usos de Facebook Insights podem emprar algun dels tutorials que trobem a la xarxa:

- A. Mariscal (2017). *Cómo descargar las estadísticas de tu página de Facebook*: <<http://alvaromariscal.com/como-descargar-las-estadisticas-de-tu-pagina-de-facebook/>>.
- *What You Need to Know About Facebook's New Insights and Analytics*: <<https://blog.kissmetrics.com/facebook-insights-and-analytics/>>.
- C. Miñana (2015). *Facebook Insights, súper tutorial en español*: <<https://www.publicidadenlanube.es/facebook-insights-tutorial-espanol/>>.
- A. Schou (2014). *¿Cómo instalar y usar Facebook Audience Insights?*: <<http://andreasschou.es/2014/11/como-instalar-y-usar-facebook-audience-insights/>>.
- Chema Espejo (2016). *Estadísticas de Facebook Insights ¡Guía básica de analítica para dummies!*: <<https://josefacchin.com/estadisticas-facebook-insights-guia/>>.

A continuació pararem atenció a alguns aspectes que treballem amb les estadístiques.

4.1.1. Viralitat de publicacions

Un dels aspectes que ens permet treballar les estadístiques és la viralitat de les publicacions.

Podem veure les publicacions més recents i quines publicacions han tingut més èxit. Es pot veure el detall pels tipus següents: tots els tipus de publicacions, publicacions, fotografies, enllaços, vídeos, publicacions de plataformes, preguntes.

Les dades que en podem obtenir són: data en què es va fer la publicació, títol, abast, usuaris que la van mirar, usuaris que la van difondre, i el percentatge de la seva viralitat (figura 46).

Figura 46. Detall d'estadístiques per posts publicats

All Posts Published August 08, 2013 to November 05, 2013							
Reach: Organic / Paid Post Clicks Likes, Comments & Shares							
Published	Post	Type	Targeting	Reach	Engagement	Promote	
10/15/2013 9:52 am	Here are 101 tips, tools, apps, and resources to help you increase revenue with Facebook. http://bit.ly/1ENMADU	📧	🌐	37.7K	1.1K 488	\$53.00 Spent	
10/02/2013 12:35 pm	Need an ecommerce solution that can satisfy the needs of small businesses, enterprises, and developers? Stay	🔗	🌐	452	13 8	Boost	
09/30/2013 11:33 am	Still want keyword data for your website or blog? Here are five ways to get it after Google encrypts search and	🔗	🌐	912	56 14	Boost	
09/19/2013 2:21 pm	Planning to sell products or services online? Here's a great resource guide for all things Shopify for beginners!	🔗	🌐	35.5K	120 174	\$40.00 Complete	
09/17/2013 10:45 am	Want to learn more about promoting your content? These 32 experts share their best tips!	📧	🌐	36.6K	496 249	\$40.00 Complete	

Font: Hines (2014).

Conèixer la viralitat de la publicació ens dona coneixements per crear més contingut que pot agradar als nostres usuaris, per assolir més fidelitzacions a la nostra marca, arribar a més públic i, en definitiva, augmentar el trànsit a la vostra pàgina.

4.1.2. M'agrada

Podem mirar l'apartat «Gustos» per veure les dades demogràfiques i les ubicacions dels fans. També es poden veure els grups d'edat. Aquesta informació ens facilita personalitzar les nostres publicacions (figura 47).

Figura 47. Informació demogràfica del públic de Facebook que explorem



Font: Schou (2014).

Amb la informació demogràfica sobre els nostres fans, podem crear persones del nostre públic objectiu. Això pot ser molt beneficiós a l'hora de crear pàgines de destinació personalitzades, pàgina d'inici personalitzada i missatges de màrqueting globals personalitzats al públic.

També es pot cercar per nous gustos i fins i tot per «no m'agrada». Aquesta informació la podem emprar per analitzar la nostra pàgina i veure les publicacions que vàrem escriure aquells dies, per tal de tenir una indicació sobre el que pot haver sortit malament. Per exemple, ens podem preguntar sobre l'atractiu del contingut, o la manera de publicar el contingut (massa articles i poques imatges).

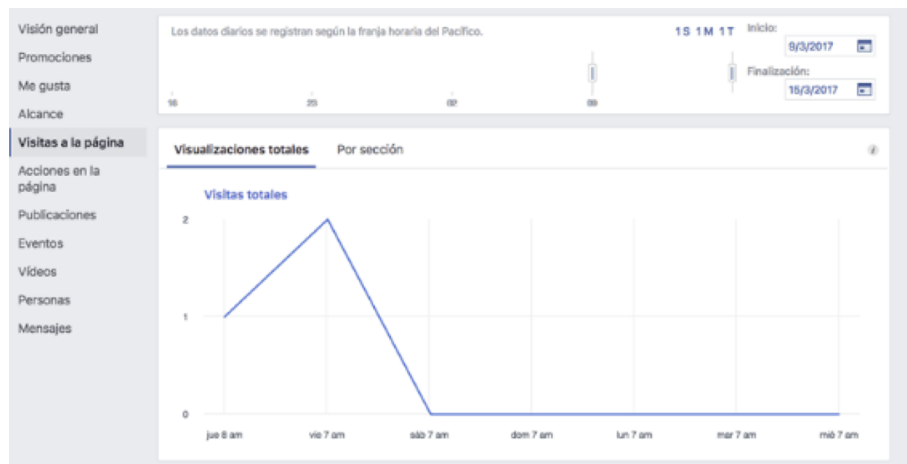
4.1.3. Abast

Quan visualitzem l'«abast» podem veure l'abast de les publicacions (el nombre de vegades que la gent entra al nostra *post*, el nombre de recomanacions que ha fet la gent, els «m'agrada», comentaris i compartició de continguts. Els aspectes que aquesta visualització ens permet analitzar són: quina part del nostre contingut està destinat a arribar als nostres fans, qui parla de la nostra pàgina, o si alguns dels nostres seguidors fan referència a la nostra pàgina a altres usuaris o comparteixen les nostres imatges amb els amics, entre d'altres aspectes.

4.1.4. Visites a la pàgina

Aquesta secció ens ajuda a veure les visualitzacions de la nostra pàgina i visites úniques de visitants (aquells que ens han cercat en lloc d'haver fet clic a l'anunci de Facebook). També podem veure el nostre abast en termes de cerques orgàniques, cerques pagades i virals (figura 48).

Figura 48. Gràfic de visites a la pàgina



Font: Farucci (2018).

4.1.5. Extracció de dades

Podem exportar les estadístiques amb el botó dret superior de la pàgina Insights (figura 49).

Figura 49. Menú per exportar dades de Facebook

Font: Mariscal (2017).

Des del menú es pot seleccionar el tipus de dades (dades de pàgines, publicacions o vídeos), l'interval de dates del qual es volen les dades, i el tipus d'arxiu que es vol; excel (.xls) o valors separats per comes (.csv).

4.2. Rfacebook: Accés a l'API de Facebook via R (opcional)

A continuació presentem un seguit d'eines més complexes que requereixen coneixements de programació. Per tant, la seva lectura és opcional.

En l'assignatura Big data i anàlisi de xarxes socials: conceptes i eines ja hem vist algunes de les funcionalitats d'aquest paquet d'R. En aquesta assignatura veurem com obtenir diferents tipus de dades: les d'una pàgina de Facebook, les d'un grup de Facebook, les informacions personals, les de pàgines de fans, i les de «m'agrada». Utilitzarem els paquets de codis creats per Pablo Barberá (Barberá, 2013, 2017).

4.2.1. Analitzar les dades d'una pàgina de Facebook

Per a Barberá (2013) les pàgines de Facebook són probablement la millor font d'informació per saber com les persones utilitzen aquest lloc de xarxes socials, ja que se'n poden recollir totes les publicacions i comentaris combinant les funcions *getPage* i *getPost*. Aquest autor ens ofereix com exemple com s'ha popularitzat la pàgina de Facebook «Humans of New York» i quin tipus d'audiència té.

El primer pas seria recuperar un marc de dades amb informació sobre totes les seves publicacions utilitzant el codi següent. Per assegurar-nos que recopilem totes les publicacions, hem de configurar *n* en un nombre molt alt i la funció s'aturarà automàticament quan arribi al total de publicacions disponibles (3.674) (figura 50).

Figura 50. Anàlisi de dades d'una pàgina de Facebook amb R

```

page <- getPage("humansofnewyork", token, n = 5000)

## 100 posts (...) 3674 posts

page[which.max(page$likes_count), ]

##           from_id           from_name
## 1915 102099916530784 Humans of New York
##           message
## 1915 Today I met an NYU student named Stella. I took a photo of her. (...)
##           created_time type
## 1915 2012-10-19T00:27:36+0000 photo
##
##           link
## 1915 https://www.facebook.com/photo.php?fbid=375691212504985&set=a.102107073196735.4429.102099916530784&type=1&r
##           relevant_count=1
##           id likes_count comments_count
## 1915 102099916530784_375691225838317      894583      117337
##           shares_count
## 1915      60528

```

Font: Barberá (2013).

La publicació més popular ha rebut gairebé 900.000 lectures i 120.000 comentaris, i es va compartir més de 60.000 vegades. Com podem veure, les variables retornades per cada publicació són les mateixes que quan busquem publicacions de Facebook: informació sobre el contingut de la publicació, el seu autor i la seva popularitat i abast. Amb aquest marc de dades, és relativament senzill visualitzar com ha crescut la popularitat de «Humans of New York» exponencialment amb el pas del temps.

El codi següent mostra com agregar les mètriques cada mes per calcular el recompte mitjà de «m'agrada / comentaris / accions per publicació» (figura 51).

Figura 51. Càlcul del recompte per mes

```

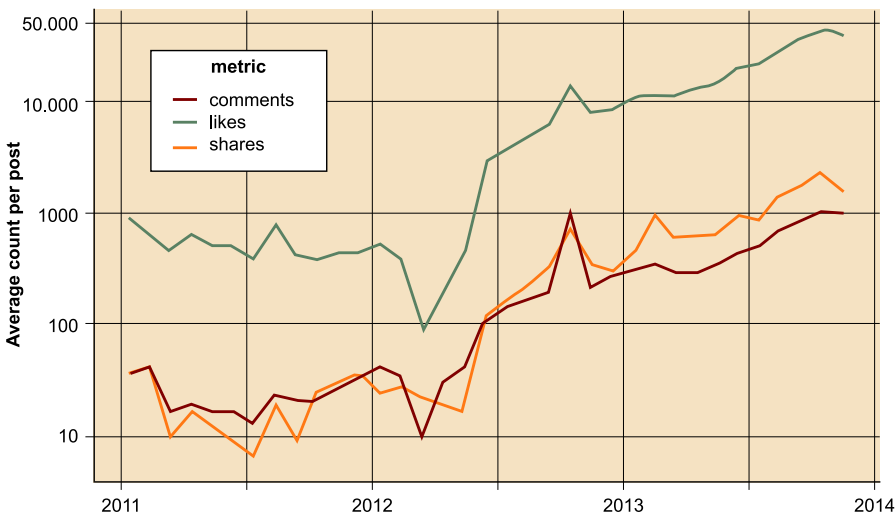
## convert Facebook date format to R date format
format.facebook.date <- function(datestring) {
  date <- as.POSIXct(datestring, format = "%Y-%m-%dT%H:%M:%S+0000", tz = "GMT")
}
## aggregate metric counts over month
aggregate.metric <- function(metric) {
  m <- aggregate(page[[paste0(metric, "_count")]], list(month = page$month),
    mean)
  m$month <- as.Date(paste0(m$month, "-15"))
  m$metric <- metric
  return(m)
}
# create data frame with average metric counts per month
page$datetime <- format.facebook.date(page$created_time)
page$month <- format(page$datetime, "%Y-%m")
df.list <- lapply(c("likes", "comments", "shares"), aggregate.metric)
df <- do.call(rbind, df.list)
# visualize evolution in metric
library(ggplot2)
library(scales)
ggplot(df, aes(x = month, y = x, group = metric)) + geom_line(aes(color = metric)) +
  scale_x_date(date_breaks = "years", labels = date_format("%Y")) + scale_y_log10("Average count per post",
  breaks = c(10, 100, 1000, 10000, 50000)) + theme_bw() + theme(axis.title.x = element_blank())

```

Font: Barberá (2017).

De manera gràfica veiem l'evolució al llarg dels anys. Per exemple, al novembre de 2013, la publicació mitjana rebia al voltant de 40.000 m'agrada (figura 52).

Figura 52. Visualització de les mètriques



Font: Barberá (2017).

També és possible seleccionar publicacions per data. Per exemple, imagineu que volem rebre només publicacions publicades durant el 2015, el codi és (figura 53):

Figura 53. Codi per seleccionar publicacions per any

```

page <- getPage("humansofnewyork", token, n = 5000, since='2015/01/01', until='2015/12/31')

```

Font: Barberá (2017).

Per obtenir més informació sobre cada publicació individual, podem utilitzar la funció *getPost*, que retornarà les mateixes variables que les anteriors, així com una llista de comentaris i m'agrada. Seguint amb l'exemple, el codi següent mostra com recollir una llista de mil usuaris que indiquen que els agrada la publicació més recent, per la qual també recollirem informació per tal d'analitzar l'audiència d'aquesta pàgina en termes de gènere, idioma i país (figura 54).

Figura 54. Codi per obtenir informació individualitzada per publicació

```
post_id <- head(page$id, n = 1) ## ID of most recent post
post <- getPost(post_id, token, n = 1000, likes = TRUE, comments = FALSE)
users <- getUsers(post$likes$from_id, token)

## 500 users -- 1000 users --

table(users$gender) # gender

## female  male
##   784   209

table(substr(users$locale, 4, 5)) # country

## AZ BE BG BR CA CZ DE DK EE ES FR GB GR HR HU IL IR IT
##  1  1  1  8  1  2  8  8  1  4  5 141  1  3  1  1  1  9
## LA LT NL PI PL PT RO RS RU SE SI SK TH US VA
##  8  3  5  5  5  2  1  1  8  2  1  1  1 758  1

table(substr(users$locale, 1, 2)) # language

## az bg cs da de el en es et fa fr he hr hu it la lt nl
##  1  1  2  8  8  1 904 12  1  1  6  1  3  1  9  1  3  6
## pl pt ro ru sk sl sr sv th
##  5 10  1  8  1  1  1  2  1
```

Font: Barberá (2013).

4.2.2. Analitzar les dades d'un grup de Facebook

Igual que les pàgines públiques de Facebook, les dades de grups públics també es poden descarregar fàcilment amb la funció *getGroup*. Tingueu en compte que això només funcionarà per a grups dels quals l'usuari autenticat és membre (figura 55).

Figura 55. Codi per analitzar dades de grups de Facebook

```
group <- getGroup("150048245063649", token, n=50)
```

Font: Barberá (2017).

4.2.3. Extracció d'informació personal

Rfacebook també permet llegir informació personal sobre l'usuari autenticat, com ara la llista de m'agrada i el contingut de la Newsfeed. La informació privada sobre els amics de l'usuari ja no està disponible a través de l'API (figura 56).

Figura 56. Codi per extreure informació personal

```
getLikes('me', token = token)[1,]
## id          names
## 1 687958677914631 FiveThirtyEight
## website
## 1 http://www.fivethirtyeight.com/
```

```
getNewsfeed(token, n=1)
## 1 51191684997 Rob DenBleyker <NA> <NA>
##
##                                     message
## 1 Sorry for the late comic, it's up now!\n\nhttp://explosm.net/comics/3512/
##           created_time type          link
## 1 2014-04-02T12:38:46+0000 link http://explosm.net/comics/3512/
##
##           id likes_count comments_count shares_count
## 1 51191684997_10152084439949998          6942          110          497
```

Font: Barberá (2017).

4.2.4. Informació de la pàgina de Fans

Rfacebook proporciona un accés limitat a les funcions d'informació, comptes d'administració disponibles de pàgines públiques. Vegeu-ne la documentació (<https://github.com/pablobarbera/Rfacebook/blob/master/Rfacebook/R/getInsights.R>) de *getInsights* per obtenir més informació (Barberá, 2017).

4.2.5. Anàlisi de «m'agrada»

La visualització de les dades dels «m'agrada» ens permet també fer una anàlisi dels continguts. Veurem un exemple facilitat per Julian Hillebrand (2014) que mostra com amb els «m'agrada» de Facebook podem fer una anàlisi de gènere. Tots els codis els podem copiar de la pàgina <http://thinktostart.com/gender-analysis-facebook-post-likes/> i de la pàgina <https://www.r-bloggers.com/gender-analysis-of-facebook-post-likes/>.

De la mà de Hillebrand (2014) farem els següents passos:

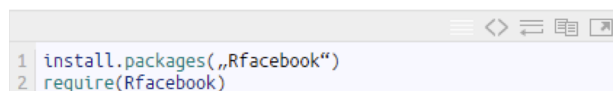
1) Instal·lació del paquet i obtenció del *token* d'autenticació

Primer instal·lem el paquet *devtools* i ens connectem a l'API de Facebook via App. Podem obtenir la informació i els codis a: <http://thinktostart.com/analyzing-facebook-with-r/>.

2) Càrrega del paquet Rfacebook

A continuació carreguem el paquet Rfacebook.

Figura 57. Paquet Rfacebook



```
1 install.packages(„Rfacebook“)
2 require(Rfacebook)
```

Font: Hillebrand (2014).

3) Anàlisi de gènere de missatges publicats a Facebook

Primer definirem el **nombre de publicacions** de la pàgina que volem analitzar. Aquestes són sempre les més recents. Però poden tenir molts m'agrada i, ja que hem de processar diferents accions en tots els casos, pot durar molt de temps si s'escull una gran quantitat de publicacions.

```
number_posts <- 2
```

Definim el **nom de la pàgina** que volem analitzar. En aquest cas escollim la pàgina 'Forbes'

```
page_name <- “forbes”
```

Descarreguem els comentaris de la pàgina amb:

```
page <- getPage(page_name, token, n = number_posts, feed = FALSE)
```

Això ens retorna un *data frame* amb el nombre de *posts* que hem sol·licitat, en cas que estigui disponible. Les publicacions tenen els següents atributs:

```
from_id, from_name, message, created_time, type, link, id, likes_count,
comments_count, shares_count
```

Per al nostre anàlisi només necessitem l'identificador de columna que conté un identificador únic per a totes les publicacions, també anomenat id de publicació.

```
posts <- page$id
```

4) Obtenció de detalls dels «m'agrada»/ agradaments d'un post

En els propers passos, es produeixen bàsicament dos processos. En primer lloc, creem una nova entrada en el nostre marc de dades final des de la publicació que estem analitzant en aquest moment. A continuació, utilitzem la seva id per obtenir més informació sobre aquesta publicació amb la funció *getPost ()*.

El quadre de dades retornat, bàsicament, conté tres valors: publicació, m'agrada, comentaris.

Aquestes categories contenen diverses llistes amb més dades, però només necessitem les dades emmagatzemades a la secció «m'agrada». Allà podem trobar els camps *from_name* i *from_id* per a tots els gustos de la publicació.

D'aquesta manera, extraurem *user_id* que és el camp *from_id* i obtenim informació sobre l'usuari amb la funció *getUsers ()*. A partir de les dades de l'usuari, extraïem el sexe i el desem en un marc de seqüència temporal.

Després de processar tots els «m'agrada» de la publicació i emmagatzemar el gènere de cadascun d'ells en el marc de sexe, es divideix en tres categories: masculí, femení i altres. Per tant, comptem quantes persones van dir que són «homes», «dones» o alguna cosa diferent.

A continuació, guardem els resultats a la nostra *data_frame_gender* i processem les properes publicacions de la mateixa manera.

Figura 58. Detall dels «m'agrada»

```

1 data_frame_gender <- data.frame(post=character(),male=numeric(),female=numeric(),etc=numeric(),likes=numeric(),type=character(),stringsAsFactors=FALSE)
2
3 for(i in 1:length(posts))
4 {
5   temp <- posts[i]
6   #dataframe values:
7   #post id
8   #likes count
9   #
10  post <- getPost(temp,token)
11
12  data_frame_gender[i,1] <- post$post$message
13  data_frame_gender[i,5] <- post$post$likes
14  data_frame_gender[i,6] <- post$post$type
15
16  gender_frame <- data.frame(gender=character(),stringsAsFactors=FALSE)
17
18  for(j in 1:length(post$likes$from_id))
19  {
20    likes <- post$likes$from_id
21    user_id <- likes[j]
22
23    user <- getUsers(user_id,token=token)
24
25    gender <- user$gender
26
27    gender_frame[nrow(gender_frame)+1,] <- gender
28  }
29
30
31  number_males <- nrow(subset(gender_frame, gender=="male"))
32  number_females <- nrow(subset(gender_frame, gender=="female"))
33  number_etc <- data_frame_gender[i,5] - (number_males+number_females)
34
35  data_frame_gender[i,2] <- number_males
36  data_frame_gender[i,3] <- number_females
37  data_frame_gender[i,4] <- number_etc
38
39 }

```

Font: Hillebrand (2014).

5) Visualització de les dades

Per visualitzar les dades hem de definir el fragment i afegir-hi els noms, seguint aquest codi:

Figura 59. Visualització de dades

```
1 slices <- c(sum(data_frame_gender$male),sum(data_frame_gender$female),sum(data_frame_gender$etc))
2
3 pct <- round(slices/sum(slices)*100)
4 lbls <- names(data_frame_gender[2:4])
5 lbls <- paste(lbls, pct) # add percents to labels
6 lbls <- paste(lbls,"%",sep="") # ad % to labels
7
8 pie(slices, labels = lbls, main="Gender Distribution of all analyzed posts")
```

Font: Hillebrand (2014).

Bibliografia

Aragón, P.; Kappler, K. E.; Kaltenbrunner, A.; Laniado, D.; Volkovich, Y. (2013). «Communication dynamics in twitter during political campaigns: The case of the 2011 Spanish national election». *Policy & Internet* (vol. 5, núm. 2, pàg. 183-206).

Barberá, P. (2013). *Introducing the Rfacebook package*. <http://pablobarbera.com/blog/archives/3.html?s_tact=C43202QW>.

Barberá, P. (2015). *Rfacebook: Access to facebook API via R*. <<https://github.com/pablobarbera/Rfacebook/blob/master/README.md>>.

Bigas, N. (2015). Una fórmula para conseguir viralidad publicitaria en internet. Entrevista a Silvia Sivera. <<https://www.uoc.edu/portal/es/news/actualitat/2015/025-viralidad-internet.html>>.

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

Farucci, C. (2018). *Métricas de Facebook: qué estadísticas medir para analizar tu actividad*. <<https://www.borjagiron.com/facebook/metricas-facebook-estadisticas-medir-analizar/>>.

Grandjean, M. (2016). *Una (genial) introducción y tutorial al Gephi*. <<http://ars-uns.blogspot.com.es/2016/05/una-genial-introduccion-y-tutorial-al.html>>.

Hillebrand, J. (2014). *Gender Analysis of Facebook Post Likes*. <<http://thinktostart.com/gender-analysis-facebook-post-likes/>>.

Laniado, D.; Volkovich, Y.; Scellato, S.; Mascolo, C.; Kaltenbrunner, A. (2017, 15 d'agost). The Impact of Geographic Distance on Online Social Interactions. *Information Systems Frontiers* (pàg. 1-16). <https://doi.org/10.1007/s10796-017-9784-9>.

Mariscal, A. (2017). *Cómo descargar las estadísticas de tu página de Facebook*. <<http://alvaromariscal.com/como-descargar-las-estadisticas-de-tu-pagina-de-facebook/>>.

Newman, M. E. (2006). «Modularity and community structure in networks». *Proceedings of the national academy of sciences* (vol. 103, núm. 23, pàg. 8577-8582).

Peralta, D. (2016). *Los 3 tipos de contenido viral en redes sociales*. <<https://davidperalta.es/los-3-tipos-de-contenido-viral-en-redes-sociales/>>.

Sherlock, David (2014). Example of Facebook Friends Analysis [en línia]. A: *David Sherlock's blog*. [Consulta: abril, 2018]. <<https://davidsherlock.co.uk/example-facebook-friends-analysis/>>.

Sivera, S. (2014). *Marketing viral: claves creativas de la viralidad publicitaria*. Barcelona: Universitat Ramon Llull [tesis doctoral].

Tsvetovat, M.; Kouznetsov, A. (2011). *Social Network Analysis for Startups*. Sebastopol, CA.: O'Reilly.

