

---

# ***Big data* i anàlisi de xarxes socials: conceptes i eines**

---

PID\_00257251

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 pel professor: Montserrat Garcia Alsina (2018)

# Índex

<b>Introducció</b> .....	5
<b>Objectius</b> .....	7
<b>1. Big data: gestió de grans volums de dades</b> .....	9
1.1. Concepte .....	9
1.2. Fases de la cadena de valor .....	12
1.2.1. Generació .....	12
1.2.2. Adquisició .....	13
1.2.3. Emmagatzematge .....	13
1.2.4. Anàlisi .....	14
1.3. Eines .....	14
1.3.1. Maquinari .....	15
1.3.2. Programari .....	15
<b>2. Social media: cadena de valor i eines per a gestionar i analitzar la informació</b> .....	19
2.1. Generació de dades en <i>social media</i> .....	19
2.1.1. On es generen .....	19
2.1.2. Quines dades es generen .....	21
2.2. Adquisició de les dades: d'on, com i sobre què extraïem dades .....	28
2.2.1. Accions .....	29
2.2.2. Eines .....	29
2.3. Emmagatzematge de les dades .....	45
2.4. Anàlisi de xarxes socials. Gephi, Pajek, R, CiteSpace, TouchGraph .....	46
2.4.1. Pajek .....	48
2.4.2. Gephi .....	49
<b>3. Projecte de recerca i anàlisi de xarxes</b> .....	51
3.1. Pla de <i>social media</i> .....	52
3.1.1. Identificar la missió: preguntes .....	52
3.1.2. Visió .....	53
3.1.3. Anàlisi DAFO: preguntes .....	53
3.1.4. Estructurar el pla de social media .....	55
3.2. Planificar la recerca .....	63
<b>Bibliografia</b> .....	65





## Introducció

Avui en dia disposem d'un gran volum de dades, generades per diferents dispositius: telèfons mòbils, sensors, navegació per webs, *social media*, entre d'altres fonts de dades. L'adequada gestió d'aquests grans volums de dades pot generar valor i és origen de creixement econòmic. Això succeeix si les dades són acurades i responen a paràmetres de qualitat que garanteixen que el resultat de l'anàlisi és fiable. Aquesta gestió de dades, les tècniques d'anàlisi i la tecnologia emprada és el que es coneix pel terme anglès *big data*.

El *big data* obre múltiples oportunitats al màrqueting digital i permet oferir nous models d'anàlisi d'usuaris que ens aporten una visió 360°. Aquests models permeten realitzar, per exemple, una segmentació del públic, i d'aquesta manera es poden fer continguts i serveis personalitzats i ser capaços d'invertir els recursos d'una manera més intel·ligent i, per tant, optimitzar-ne els resultats.

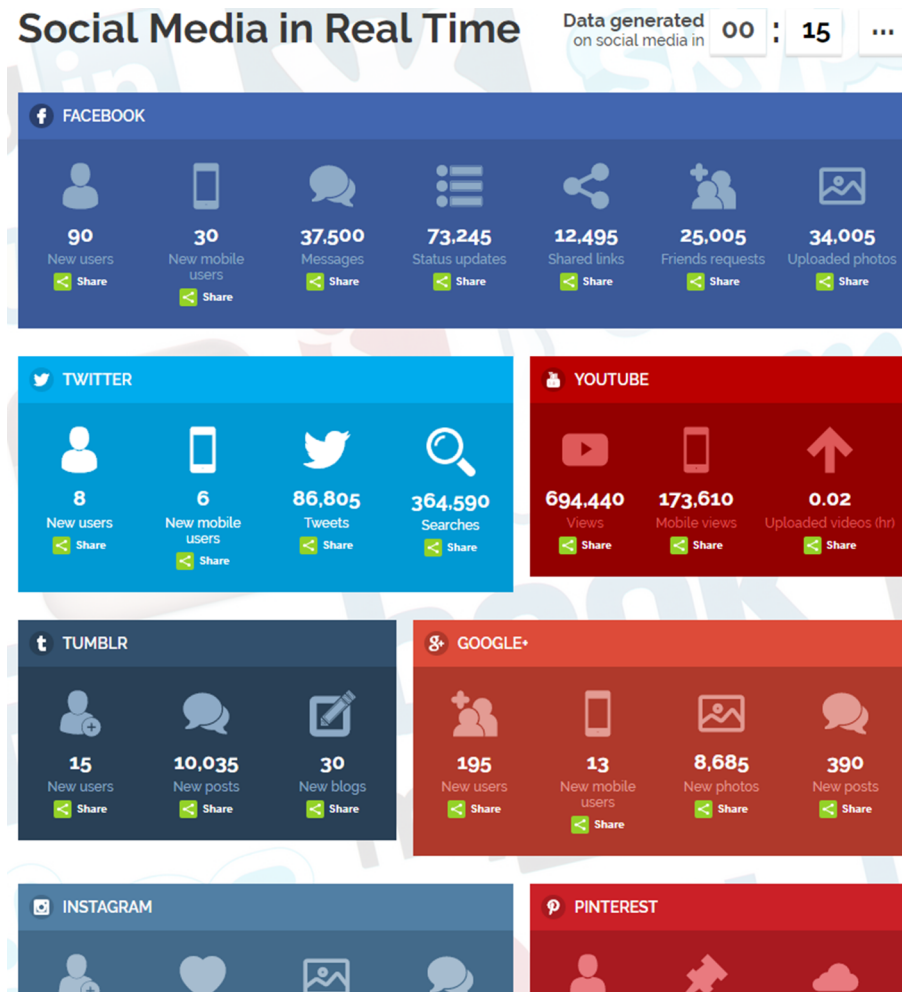
Les xarxes socials generen un gran volum de dades si considerem tant els continguts de les notícies, notes o comunicacions unilaterals, bilaterals o multilaterals que tenen lloc als *social media*, i la pròpia informació que s'origina quan circulen els continguts i es produeixen interaccions en el si de les xarxes. Les xifres d'usuaris de xarxes socials augmenten any rere any i les noves generacions hiperconnectades se senten còmodes navegant i comunicant-se mitjançant les xarxes socials.

Les xarxes socials van començar sent exclusivament per a ús personal, però també han demostrat que són útils per a les empreses. És econòmic i fàcil de promocionar productes i serveis independentment del mercat. Amb l'ajut de totes les dades que els usuaris generen diàriament, les empreses/entitats poden usar segments de mercat per captar nous usuaris i així augmentar les vendes i la consciència de marca o bé el nombre de clients interessats en els seus serveis.

L'àmplia penetració de les xarxes socials a les nostres vides ha provocat fins i tot l'aparició de nous treballs, entre ells els *influencers*, una persona amb un alt volum de seguidors que escolten els seus missatges i els difonen. Són un nou model de líders d'opinió i YouTube i Instagram són els *social media* per excel·lència per a l'*influencer* de màrqueting.

Ens podem fer una idea dels grans volums de dades si mirem Coupofy. Aquesta eina presenta una infografia interactiva que mostra quantes dades es generen cada segon a les xarxes socials (figura 1).

Figura 1. Captura de pantalla de la generació de dades per segon



Font: imatge facilitada per Coupofy, <http://www.coupofy.com/social-media-in-realtime/>.

Les publicacions de Facebook obtenen *likes* o m'agrada a un ritme de més de quatre milions per minut. Els usuaris envien piulades a Twitter a una velocitat d'aproximadament 350.000 per minut. I més de tres-cents hores de vídeo es carreguen a YouTube cada seixanta segons.

Aquesta gran generació de dades és un dels grans reptes als quals s'enfronten avui institucions públiques i privades, però l'important és saber què fer amb aquestes dades i com. Aquest és el veritable èxit.

En aquest mòdul estudiarem, en primer lloc, el concepte de *big data*, i les eines emprades per a gestionar el gran volum de dades. Això ens servirà de base per a treballar, en segon lloc, com explotarem i gestionarem les dades dels *social media*, prenent com a base la cadena de valor. Per últim, treballarem com plantejar preguntes a les dades.

## 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 la cadena de valor del *big data*.
2. Obtenir una panoràmica de les eines de gestió i anàlisi de les xarxes socials.
3. Saber plantejar un pla o projecte d'anàlisi de xarxes socials i gestionar-lo.
4. Saber formular preguntes de recerca per orientar la cerca i l'anàlisi d'informació.
5. Identificar fonts de dades rellevants.
6. Dissenyar un pla/projecte i gestionar-lo en un entorn professional o de recerca.
7. Aplicar l'ús de les TIC.



## 1. *Big data*: gestió de grans volums de dades

El *Big data* o dades massives és un concepte abstracte que ha esdevingut una paraula de moda en diversos àmbits: negocis, enginyeria informàtica, informació i documentació, o sistemes d'informació, entre altres.

L'increment exponencial de la quantitat de dades accessibles, processables i analitzables a gairebé tots els camps de l'activitat humana ha suposat una revolució per al coneixement. Quan acumulem grans volums de dades, apareix la necessitat de veure què podem fer-ne i quina informació podem extreure'n. Aquest fet crea la necessitat de gestionar-les amb una finalitat organitzativa, i disposar de tecnologia i metodologies específiques. La mateixa gestió de dades ens porta a generar informació que sigui rellevant en el context de l'organització, és a dir, a generar coneixement per a l'acció, que sigui aplicable a diferents àmbits, per exemple, per a la presa de decisions, per al disseny d'accions o per a l'elaboració de plans estratègics.

Per tant, quan parlem de dades massives, estem parlant també de gestió de la informació i de generació de coneixement per a l'acció. En el cas dels *social media*, veiem que també estan circulant grans volums de dades. Això ho estudiarem en el següent apartat.

Aquest camp científic és el que ens dona les pautes metodològiques per a gestionar grans volums de dades amb la finalitat de crear valor mitjançant una sèrie de processos i procediments. Però, a més, hem de tenir en compte que hem de disposar de la tecnologia adequada per a la captura de dades, processar-les, analitzar-les i interpretar-les de manera ràpida i eficient (Gómez i Conesa, 2015).

Així doncs, quan definim dades massives hem de respondre les següents qüestions: Per a què serveixen? Com les podem gestionar? Amb què?

Estudiem ara en aquest apartat què és el *big data* des de la perspectiva de les seves característiques, la seva cadena de valor i les eines tecnològiques (programari i maquinari) que ajuden a la gestió dels grans volums de dades.

### 1.1. Concepte

El *big data* sovint es defineix a partir de les característiques de les dades, que comencen per V. El volum és la primera característica de la definició de dades massives, que marca indirectament una de les novetats de les dades massives, i una de les limitacions de les aplicacions informàtiques que durant les últimes dècades gestionaven dades. A més del volum, en la definició s'hi han considerat aspectes com la varietat de les dades i la velocitat del processament.

Posteriorment, altres autors, quan n'han aprofundit l'estudi, han afegit més V a les tres V inicials (volum, varietat i velocitat). Aquestes noves V estan vinculades amb la qualitat de les dades, el tractament i la gestió, o prestacions de tecnologia i programari. Aquestes característiques són: la veracitat, el valor, la visualització, la verificació, la variabilitat i la viabilitat (Chen i altres, 2014; Padgavankar i Gupta, 2014).

A continuació expliquem algunes de les V associades a les dades massives:

1) **Volum:** Fa referència a dos aspectes. El primer és l'increment exponencial de les dades, fruit de l'ús de noves tecnologies i la facilitat de generar dades digitals. El segon aspecte és el repte que suposa recollir, emmagatzemar, processar i integrar grans volums de dades procedents de fonts molt variades i distribuïdes.

La gestió de grans volums de dades és determinant per a impulsar el desenvolupament de plataformes capaces de capturar, emmagatzemar, analitzar i visualitzar aquestes dades.

2) **Varietat:** Les dades són molt variades en quant a tipologies, format i estructures usades per a la seva organització i presentació. Això es deu a que procedeixen de fonts molt diverses i estan generades per dispositius diversos (mòbils, àudio, vídeo, sistemes GPS, sensors de temperatura, mesuradors de vent, sensors RFID, missatges de Twitter, webs, internet de les coses, etc.).

Tenint en compte les estructures amb les quals s'organitzen les dades, i la manera com s'emmagatzemen, podem classificar les dades segons el nivell d'estructuració: estructurades, no estructurades i semiestructurades (Gómez i Conesa, 2015).

- Les **dades estructurades** són les que s'emmagatzemen en camps, la longitud dels quals, la denominació i el format han estat predefinits, i s'emmagatzemen en taules de bases de dades. Per tant, coneixem per avançat l'organització de les dades, l'estructura i el tipus de cada dada, la seva posició i les possibles relacions entre les dades. La informació ve representada per dades elementals, no compostes per d'altres estructures.
- Les **dades no estructurades** són documents en el format que han estat creats, i per tant, no tenen un format específic, ni estan emmagatzemats en una taula. La informació no està representada per dades elementals, sinó per una composició cohesionada d'unitats estructurals de nivell superior. La interpretació i manipulació d'aquestes dades és més complexa (Gómez i Conesa, 2015).
- Les **dades semiestructurades** tenen dades elementals, però no tenen una estructura fixa, tot i que tenen algun tipus d'estructura implícita o autodefinida. És el cas de documents escrits amb llenguatge html o xml. Aquest

tipus de dades presenten complexitat en el procés de carrega, perquè s'hi han d'afegir tantes excepcions com possibilitats de variació tinguin les dades. Per exemple, en el cas d'un document en xml, l'ordre de les dades no és important, per tant, quan la base de dades carregui un document, primer l'haurà d'interpretar. A més, a l'hora de carregar-lo, si algun dels camps no està estructurat es produeix pèrdua d'informació.

Per tant, la gestió de dades no estructurades i semiestructurades és un dels reptes de les dades massives.

**3) Velocitat:** El processament de les dades massives s'ha de fer en el mínim temps possible, i en alguns casos en temps real.

La velocitat de gestió i anàlisi de les dades permet veure amb suficient antelació fets i accions que ens permeten reaccionar amb temps: clients que es donen de baixa, mal funcionament d'un servei o d'un producte, etc. Aquesta característica està vinculada a les eines que permeten gestionar dades massives i fa referència a la velocitat de càrrega, la velocitat de processament i la velocitat de visualització de les dades per a facilitar l'anàlisi i l'extracció de patrons.

**4) Veracitat:** Donat que les dades massives ens permeten extreure informació per a prendre decisions i crear coneixement per a l'acció, és important que siguin fiables i siguin fidels a la realitat. Per aquest motiu, la veracitat està relacionada amb la certesa o incertesa i l'exactitud o inexactitud de les dades.

Per assegurar-ne la veracitat, és precís aplicar una sèrie de mesures, com ara protegir les dades d'atacs malignes, i actualitzar les dades, tenint-ne en compte la variabilitat. Aquests aspectes estan vinculats amb el tractament de les dades, més concretament amb la neteja de les dades i actualització.

**5) Valor:** L'anàlisi de les dades i l'extracció d'informació creen coneixement que és font d'innovació, competitivitat, productivitat i ocupació. Per això, la gestió i l'anàlisi de les dades massives han d'estar orientades a la creació de valor, com a mitjà per a justificar els esforços invertits.

**6) Visualització:** Les prestacions del programari que gestiona les dades han de facilitar que volums grans de dades es presentin visualment de manera pràctica, dinàmica, interactiva i comprensible. A més, la visualització ha d'anar acompanyada del context per facilitar-ne l'anàlisi.

**7) Verificació:** La verificació permet ratificar la seguretat de dades i la seva integritat mitjançant missatges d'autenticitat, signatures digitals, certificats de tercers i altres sistemes (Liu i altres, 2015).

**8) Variabilitat:** Aquesta característica està relacionada amb la velocitat amb què es produeixen les dades, ja que la generació de dades pot produir l'obsolescència d'altres. Per tant, cal introduir procediments específics per a la seva gestió en termes d'eficiència, al mateix nivell que la veracitat.

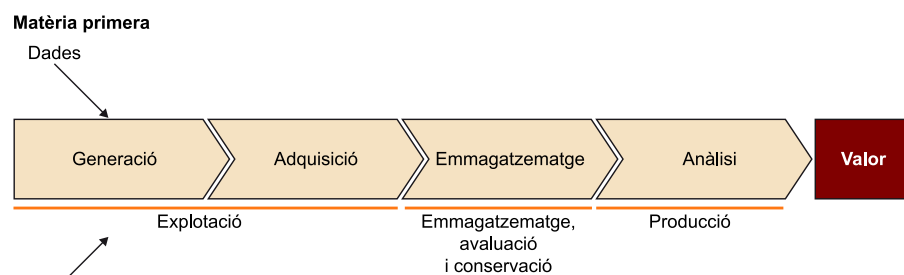
**9) Viabilitat:** Les plataformes i el programari per a gestionar dades tenen capacitats i complexitats diferents, i, per tant, els pressupostos resultants per comprar infraestructures tecnològiques necessàries poden ser molt elevats. Cal observar el cost en quant a infraestructures i eines necessàries per emmagatzemar i analitzar les dades massives, garantint-ne la velocitat, veracitat, verificació i variabilitat. En qualsevol cas, la característica de viabilitat ens recorda que la inversió ha d'estar justificada pel valor que s'extregui de les dades i que aquestes han d'estar orientades als objectius que l'organització tingui fixats.

## 1.2. Fases de la cadena de valor

Les dades són la matèria primera que cal processar i transformar per a generar valor, mitjançant una sèrie de processos. Aquests es presenten visualment en la següent cadena de valor i es poden englobar en tres: explotació, emmagatzematge i producció (Chen i altres, 2014) (figura 2). El procés d'emmagatzematge cal que vagi acompanyat de l'avaluació i el manteniment per garantir la qualitat de les dades emmagatzemades per a la seva recuperació. Les dades que es generen a les xarxes socials també formen part de la cadena de valor dels grans volums de dades.

Mirem ara les fases de la cadena de valor de les dades: generació, adquisició, emmagatzematge i anàlisi (figura 2).

Figura 2. Cadena de valor de les dades



Font: Garcia-Alsina (2017).

### 1.2.1. Generació

La cadena de valor de les dades massives comença amb la generació de dades procedents d'una gran varietat de fonts. Hi ha diferents entorns on es generen grans quantitats de dades, i en molts casos de manera automàtica: rastres que deixem de les cerques que fem per internet, dades que generen els sensors i



càmeres de vigilància que s'instal·len en els edificis o naus industrials, sistemes d'informació existents a les organitzacions, dades i informació que generen les administracions públiques o la investigació, internet de les coses, etc.

Recordem que aquestes dades tenen com a una de les seves característiques la varietat. Aquesta s'engloba en tres tipus de dades, atenent a la seva organització: estructurades, desestructurades i semiestructurades.

### **1.2.2. Adquisició**

El segon esglaió de la cadena de valor és l'obtenció de dades, generades per diferents tipus d'objectes. Per obtenir dades i analitzar-les, abans les hem de descarregar dels llocs on es generen. Tanmateix, no podem descarregar totes les dades, per tant, hem de seleccionar tan sols les dades que satisfan les nostres necessitats d'informació. Un cop tenim descarregades les dades seleccionades, aquestes requeriran un tractament per evitar redundàncies, i inconsistències. Per això podem dir que en aquest esglaió de la cadena trobem tres subfases: recollida, transmissió i preprocessament.

En conseqüència, en aquesta fase efectuem les següents accions:

- Definir els objectius que es volen assolir i les preguntes d'acord a les necessitats d'informació manifestades.
- Identificar les fonts de dades adients.
- Especificar quina metodologia emprem per extreure dades.
- Tractar dades: integració i federació de dades, neteja, eliminació de redundàncies.

### **1.2.3. Emmagatzematge**

Un cop hem obtingut les dades, aquestes han d'estar emmagatzemades en una plataforma, en un format adequat que n'asseguri una recuperació ràpida i posterior anàlisi i extracció de valor. El sistema d'emmagatzematge ha d'assegurar la consistència de les dades, en el cas que hi hagi diverses còpies de les mateixes dades (en previsió de caigudes de servidors), i reproduir-les en diversos servidors.

En l'emmagatzematge juga un paper important definir com guardar les dades de cara a l'eficiència en la recuperació dels conjunts de dades, com plantejar les cerques i consultes de les dades emmagatzemades, i com presentar-les als usuaris. En aquest sentit, s'ha d'explorar quin servei ens poden donar diversos àmbits com:

- a) Llenguatges controlats, per treballar la precisió de la classificació i indexació de tots els conjunts de dades.
- b) Usabilitat dels magatzems de dades.
- c) Visualització de les dades.

En la segona part d'aquest mòdul veurem aquestes fases de la cadena de valor, centrades en com treballem les dades dels *social media*.

#### 1.2.4. Anàlisi

La cadena de valor de les dades massives té com a fase final l'anàlisi de les dades. Aquesta fase és clau per extreure el valor que les dades contenen amb la finalitat de crear coneixement per a l'acció. Des de sempre s'ha donat valor a les dades i a la informació per a la presa de decisions. Les organitzacions sempre han disposat de dades, encara que fossin en format paper. La disciplina de la intel·ligència competitiva és un exemple dels processos conduents a extreure informació estratègica i coneixement per a l'acció, mitjançant tècniques d'anàlisi. Els avenços en la tecnologia han permès que l'anàlisi compti amb el suport d'eines cada vegada més potents. Els magatzems de dades (*data lake*) tenen com un dels seus objectius facilitar l'anàlisi de les dades.

L'anàlisi engloba un conjunt de procediments i models estadístics per extreure informació d'un ampli conjunt de dades (Kune i altres, 2016). En aquest sentit, hi ha una sèrie de mètodes desenvolupats, com són la mineria de dades, els factors, les correlacions, les regressions, el test A / B, l'estadística, etc. (Chen i altres, 2014).

#### 1.3. Eines

Per a la gestió de grans volums de dades són necessaris maquinari i programari específics. Calen infraestructures i eines que permetin la rapidesa de processament, emmagatzematge a gran escala i xarxes potents i ràpides, a preus més baixos.

En entorns d'indústria, els equipaments utilitzats per dur a terme l'anàlisi i la mineria de grans conjunts de dades solen comprendre grans sistemes distribuïts, amb centenars de nuclis de còmput i gran quantitat de memòria, tant en forma física (*in situ* a l'empresa/entitat o institució corresponent) com en la xarxa (*Cloud Computing*). En tots els casos, l'estructura lògica del sistema de mineria sol estar composta per dos elements principals:

a) Un sistema d'emmagatzematge massiu comú on es guarden les dades. Aquest emmagatzematge pot adoptar diferents noms, segons la mida del sistema, i l'origen i la naturalesa de les dades emmagatzemades. Aquests noms són: *data lake*, *data warehouse*, *data mart*.

b) Un sistema de còmput i processat que treballa sobre les dades de la part anterior.

### 1.3.1. Maquinari

Avui en dia tenim al nostre abast el fet de poder consumir aquesta informació sense necessitat de comptar amb un caríssim supercomputador. Podem arribar a fer proves amb sistemes d'uns 8-16 GB de RAM i uns quants GB de disc dur. En un entorn de producció es fa servir *comodity hardware*, és a dir, màquines que no necessiten tenir un alt grau de fiabilitat i sofisticació, com per exemple sistemes RAID (Redundant Array of Inexpensive Disks) que combinen diferents discs magnètics o d'estat sòlid en una sola unitat lògica per redundar dades per temes de seguretat, o augment de transferència (velocitat), discs durs *enterprise*, components redundants, entre d'altres. Màquines d'aquest tipus són molt més barates i, si es trenquen, són fàcilment substituïbles. El pes de la fiabilitat recau sobre el programari. Haurem, això sí, de dimensionar les característiques segons els requeriments del nostre escenari. Una configuració típica d'un node d'un entorn productiu podria ser una màquina amb 8-24 cores, 32-256 GB de ram i uns 8-12 discos HDD.

Per introduir-nos en l'ús de programari necessari per a tractar grans volums de dades no cal tenir un clúster de màquines potents, podem trastejar unes simples màquines virtuals en el nostre PC (VM, sigles de la terminologia anglosaxona *Virtual Machines*) o amb alguna VM com la de Cloudera QuickStart VM que ja ho té tot muntat, i que serveix justament per aprendre.

### 1.3.2. Programari

Mirem ara les diferents opcions d'emmagatzematge des del punt de vista del programari: *data lake*, Hadoop, Hbase, Hive i Apache Spark.

#### **El *data lake***

Actualment les dades s'emmagatzemen als *data lake*, que se solen construir mitjançant el programari Hadoop que veurem a continuació.

Un *data lake* és un repositori que conté una gran quantitat de dades en brut en el seu format natiu. Aquestes dades es mantenen allí fins que es necessiten. Fins fa poc, les bases de dades que gestionaven de manera intel·ligent grans volums de dades eren els *data warehouse*, que posteriorment també se'ls ha

nomenat *business intelligence*. Entre aquests dos programaris hi ha diferències. Vegem a continuació algunes de les característiques dels *data lake*, i les diferències amb els *data warehouse*.

Un *data lake* utilitza una **arquitectura plana** per emmagatzemar les dades en brut. A cada element del *data lake* se li assigna un identificador únic i se l'etiqueta amb un ampli conjunt de metadades. Quan alguna qüestió de negoci ho requereix, podem sol·licitar les dades relacionades del *data lake*, i amb aquest conjunt més petit de dades podem fer l'anàlisi que ens permeti respondre preguntes. Es diferencien dels *data warehouse* perquè aquests emmagatzemen les dades en arxius o carpetes de forma jeràrquica.

Una altra característica del *data lake* és la **distribució de dades** per múltiples nodes que poden estar en el núvol en lloc de romandre en un entorn fix i estructurat basat en esquemes. Això és un altre tret diferencial del *data lake* vers el *data warehouse*.

Per últim, destaca com a característica dels *data lake* l'autonomia de l'usuari, perquè primer podem explorar i esbrinar què hi ha i després decidir quin és l'esquema que organitza millor les dades d'acord a les nostres necessitats d'informació. Això dona més llibertat a l'usuari per a seleccionar i combinar les dades, per extreure el valor d'acord a les seves necessitats d'informació. Aquesta característica és un altre tret diferencial dels *data lake* vers els *data warehouse*. En general, en un *data warehouse* hi ha professionals d'IT que creen models de dades segons els seus criteris, i que si exclouen les necessitats dels decisors de l'organització, pot obstaculitzar la productivitat i el valor de negoci que han aportat per les dades.

## **Hadoop**

És un programari obert per a processament de grans conjunts de dades distribuïdes. Per tant, forma part d'un sistema distribuït i està disponible sota llicència Apache. Els sistemes distribuïts són un conjunt d'ordinadors ubicats en diferents llocs, però connectats entre si per una xarxa de comunicacions. Cada ordinador té instal·lat el seu programari i els seus perifèrics que l'usuari percep com un sol sistema.

Aquest és un dels programaris més utilitzats en l'àmbit de sistemes distribuïts, incloent-hi moltes de les grans empreses del sector de Tecnologies de la Informació.

Hadoop consta de tres grans parts:

- **Hadoop Distributed File System** (HDFS, «Sistema de Fitxers Distribuït Hadoop»).
- **Hadoop MapReduce**, un *framework* de processat paral·lel.

- **Hadoop YARN**, és un *framework* afegit en versions recents, per a planificar treballs i gestionar recursos.

Vegem aquests elements:

### 1) Hadoop Distributed File System (HDFS)

És un sistema d'arxius basat en una arquitectura mestre-esclau. És escalable, tolera els errors (mal funcionament del maquinari o del programari), i compta amb una arquitectura distribuïda. Els arxius estan distribuïts en diverses màquines per al seu processament, tot i que sembla que treballi en un sol arxiu.

### 2) MapReduce

És un model de programació creat per aplicacions que han de processar en un maquinari grans quantitats de dades de forma paral·lela, i pot ser executat en diversos llenguatges de programació, com Java, Ruby, Python i C ++. Quan les dades entren per ser processades, aquestes es divideixen en grups per al seu processament de manera distribuïda, en paral·lel, en diferent maquinari per després combinar el resultat.

### 3) Yarn-MapReduce

És l'evolució de MapReduce per a Hadoop 2.0. La funció de YARN en Hadoop és proporcionar un entorn que gestioni els recursos per a realitzar treballs de computació. YARN s'encarrega de distribuir la feina a fer tenint en compte on són les dades a processar, a més de gestionar les pròpies execucions dels programes. Amb YARN podem separar el sistema de fitxers del sistema d'execució. Per tant, podem fer servir el sistema de fitxers distribuïts de Hadoop (HDFS) sense passar per YARN o bé utilitzar aquest gestor de recursos per a accedir a aplicacions (com és el cas de la majoria de les eines que treballen en Hadoop).

## Hbase

Aquest programari es pot considerar «la base de dades» de Hadoop. Basat en BigTable de Google, proporciona la capacitat de crear taules amb milions d'entrades i permet fer accessos de lectura/escriptura ràpida i consistent. També és versionada (control de versions) i no-relacional, de manera que ofereix flexibilitat. A més, és fàcilment connectable, ja sigui per la seva API (Application Programmable Interface) java o mitjançant *web services*. Podem fer servir HBase per donar forma a les nostres dades.

## Hive

És una base de dades que facilita llegir, escriure i gestionar conjunts de dades, emmagatzemades de manera distribuïda. Estructura les dades en taules i vistes i, com totes les bases de dades, permet fer consultes i extreure informació usant el seu propi llenguatge d'interrogació anomenat HiveQL, un tipus de SQL.

## Apache Spark

Apache Spark és un *framework* de còmput distribuït, ràpid i de propòsit general. És capaç de ser executat per sobre de diversos sistemes distribuïts (Apache Hadoop, Apache Mesos), en el núvol (Amazon EC2), i en solitari (standalone), adaptant-se a pràcticament qualsevol entorn d'ús. A més, disposa de llibreries d'aprenentatge computacional (MLLib), operacions amb llenguatge d'interrogació SQL (Spark SQL), tractament de gràfics (GraphX) i tractament de fluxos de dades (Spark Streaming). Aquestes llibreries es troben disponibles, en la seva majoria, per als llenguatges de programació Java, Scala i Python. Recentment, s'ha començat a afegir suport per R, un llenguatge àmpliament utilitzat en estadística i anàlisi de dades. A més, com en el cas de Hadoop, Spark és un programari amb un ampli grau d'utilització en les grans empreses del sector.

## 2. *Social media*: cadena de valor i eines per a gestionar i analitzar la informació

En aquest apartat estudiarem la cadena de valor de les dades, aplicada als *social media*. Per tant, veurem en primer lloc on es generen dades, i quin tipus de dades. En segon lloc, treballarem com descarregar les dades i amb quines eines. A continuació, on i com les emmagatzemarem. Finalment, coneixerem les eines per analitzar les dades.

### 2.1. Generació de dades en *social media*

En aquest apartat veurem en quines *social media* es generen dades i quin tipus de dades trobem.

#### 2.1.1. On es generen

Tots coneixem les anomenades xarxes socials, i molts en som usuaris privats o professionals. Tot i així, veurem algunes de les xarxes socials on es generen dades que ens permeten extreure informació en l'actualitat. Les descriurem a continuació.

#### **Twitter**

És un servei de *microblogging* que permet als usuaris enviar i llegir missatges de text d'una llargada màxima de dos-cents vuitanta caràcters anomenats *tweets* o piulades. És el canal de comunicació per excel·lència i permet:

- Monitoritzar el nínxol de mercat que és objecte del nostre interès.
- Tenir vigilada la competència.
- Dirigir l'audiència cap a una determinada pàgina web, per exemple la de la nostra empresa.
- Difondre el contingut durant les hores de més activitat.
- Identificar una llista d'influenciadors dins d'un sector concret, i guanyar-nos-en la confiança.
- Compartir continguts interessants per a la nostra audiència, que no tenen per què ser exclusivament nostres.

#### **Facebook**

És una xarxa social llançada el 2004 i que pertany a la companyia privada Facebook, Inc. La xarxa permet d'afegir-hi gent com a amics, i enviar-los missatges i compartir enllaços, fotografies i vídeos, entre d'altres coses. És obert a tothom qui tingui més de tretze anys, i només cal una adreça de correu electrònic vàlida per a registrar-s'hi. És una de les xarxes socials més conegudes

actualment. Té com a principi un mur particular per a cada usuari en el qual es van succeint les diverses publicacions de les amistats (membres que hom ha acceptat com a amics) o dels grups o pàgines als quals hom ha clicat a «M'agrada».

Permet les següents accions:

- Plantejar els continguts de Facebook com una forma de fer participar la nostra audiència.
- Proporcionar jocs, enquestes, sortejos, promocions exclusives per als fans i dissenyar, sobretot, contingut que sigui fàcilment compartible.
- Fer una planificació cronològica i escalada perquè la nostra pàgina no sembli una ciutat en construcció.

Hem de ser conscients que sense pagar pels anuncis de Facebook és difícil que les nostres publicacions arribin a més d'un 5% de l'audiència.

Hem de comprovar si Facebook per al nostre tipus de negoci/entitat és el canal que més conversions ens genera.

## **Instagram**

És una xarxa social especialitzada, una aplicació mòbil i web gratuïta desenvolupada per Kevin Systrom i Mike Krieger l'any 2010 per compartir imatges i vídeos. L'aplicació permet als usuaris fer fotografies, aplicar-hi filtres i marcs si es desitja i, finalment, mostrar-les a les seves amistats o seguidors, ja sigui a la mateixa plataforma o a diverses xarxes socials, incloent-hi Facebook, Twitter, Quadrangular, Tumblr, Flickr i Posterous.

És una de les xarxes que més creix i la que més possibilitats ofereix per donar a conèixer la nostra empresa/entitat des de dins.

Aquesta eina permet fer les següents accions:

- Atraure trànsit al vostre web.
- Humanitzar la nostra marca mostrant el nostre treball i la nostra gent i guanyar-nos la seva audiència.

## **LinkedIn**

És una xarxa social orientada als professionals i als negocis. Fundada als EUA el desembre de 2002, va ser presentada oficialment el maig de 2003. El principal propòsit d'aquesta xarxa és posar en contacte i establir col·laboracions entre professionals qualificats.



Aquesta eina permet fer les següents accions:

- Aconseguir el major nombre de seguidors a la nostra pàgina d'empresa/entitat, i que aquests recomanin el nostre contingut professional.
- Utilitzar un llenguatge més sobri i adequat per al sector.
- Difondre el nostre contingut als nostres treballadors i als respectius contactes.
- Usar-la de manera estratègica.

### 2.1.2. Quines dades es generen

Cada vegada que un usuari ofereix les seves dades mitjançant un canal, com pot ser una pàgina de Facebook, blog, Twitter, LinkedIn o una pàgina web entre d'altres, genera un *lead*. Un *lead* és un anglicisme que fa referència a l'usuari que ha demostrat interès en l'oferta de la marca, amb una sol·licitud d'informació addicional o oferint les seves dades al canal. L'usuari obté a canvi una participació en un concurs, l'obtenció d'un servei o producte, l'accés a un contingut com pot ser un *e-book* o un tutorial i fins a una subscripció.

Però quin tipus de dades generen els usuaris de les xarxes socials? Quines dades generen les mateixes xarxes socials? Cada usuari genera i comparteix una gran quantitat d'informació a la xarxa que es pot agrupar en les categories següents: perfil d'usuari, amistats, seguidors, a qui seguir, fotos, interaccions, grups i dades de control de les interaccions entre els components de les mateixes xarxes. A més, les mateixes xarxes socials generen una sèrie de mètriques, que són dades també susceptibles de ser analitzades per generar valor. Tot seguit, veurem amb més detall aquests tipus de dades, les que generen els usuaris i les que registren i ofereixen les mateixes xarxes.

### Dades generades per l'usuari

Les dades que genera l'usuari són tant personals com els gustos i preferències en relació als continguts que circulen en les xarxes socials.

#### 1) Perfil d'usuari

L'usuari pot compartir el seu gènere, l'aniversari, l'adreça de correu electrònic, l'adreça IP d'origen, la data en què es va unir a la xarxa social, el telèfon, la seva formació, el lloc on treballa i la població on viu. Un usuari també pot compartir el seu estat civil actual, incloent-hi «Solter/a», «En una relació», «Promès/a» o «Casat/da». Per als tres últims estats, hi ha un altre compte d'usuari associat.

També pot especificar alguns dels seus *hobbies* o interessos en la informació en el seu perfil, p. ex., «com la música», «com Hello-Kitty». Els interessos selectius poden tenir un paper important en l'anàlisi de les característiques d'un usuari.

## 2) Amistats, seguidors i a qui seguir

Els usuaris del graf social poden estar connectats entre si, si estableixen amistat (en el cas de Facebook) i si es segueixen mútuament a Twitter. L'iniciador (és a dir, un usuari) d'una amistat pot revocar la sol·licitud d'amistat o deixar de seguir a un altre usuari en qualsevol moment. Després de rebre la sol·licitud d'amistat o veure que l'està seguint, aquest usuari pot negar o aprovar l'amistat o bloquejar-ne el seguiment. En el cas de Facebook, l'usuari pot tornar a aprovar l'amistat, tot i que la hi hagi negat abans. Això passa sovint després que un usuari comprova i verifica la informació de l'iniciador.

## 3) Fotos pujades a les xarxes socials

Per a cada foto, l'usuari pot etiquetar amistats i/o afegir-hi la ubicació i *hashtags* que poden descriure el context del micromoment de l'usuari.

## 4) Interaccions amb els *posts* de Facebook o tweets

L'usuari es converteix en moderador i, òbviament, en subscriptor. Els amics o seguidors dels usuaris responen a la publicació, escrivint-hi comentaris i mostrant el seu interès en una publicació o comentari, repiulant, compartint la publicació o afegint-hi un m'agrada o altres opcions, en el cas de Facebook. Per tant, una publicació o un comentari té una llista d'usuaris a qui «els agrada» o que han compartit aquella publicació.

## 5) Grup de Facebook

Un grup de Facebook és creat per un usuari per a persones que comparteixen interessos similars o que viuen al mateix lloc. Cada grup també té moltes discussions/publicacions creades pels seus membres.

## 6) Característiques dels dispositius dels usuaris

Es poden obtenir dades sobre els dispositius (marca, sistema operatiu, etc.) i es poden usar per a conèixer més a l'usuari.

## 7) *Fan Page* de Facebook:

Una *Fan Page* de Facebook és una funcionalitat més de Facebook creada sobretot per ser un canal de comunicació entre persones unides per un interès (fans) dins de la xarxa social. Per tant, aquest tipus de pàgina té per objectiu

unir persones o empreses interessades en un tema en comú, sense necessitat que ningú n'hagi d'aprovar l'amistat. És el fan el que decideix si segueix o no les actualitzacions d'una pàgina.

És una eina de comunicació per a les empreses, ja que poden estar en contacte amb els seus clients actuals o potencials.

Una *Fan Page* conté una sèrie de seccions que generen un seguit de mètriques. Vegem-les en l'apartat següent.

### **Dades generades per les mètriques de les xarxes socials**

Vegem principalment les mètriques de dues xarxes socials: *Fan Page* de Facebook i Twitter. Més endavant oferirem més detalls sobre les mètriques i les eines útils per avaluar les xarxes socials, que ens subministren un elevat nombre de dades.

#### **Vegeu també**

En l'apartat 3 d'aquest mòdul s'ofereixen més detalls sobre les mètriques i les eines útils per avaluar les xarxes socials.

#### **1) *Fan Page* de Facebook**

Les seccions i mètriques de *Fan Page* són les següents: resum, m'agrada, abast, visites, *posts*, vídeos i gent. A continuació, vegem-ne una petita descripció.

##### **a) Resum (*Overview*)**

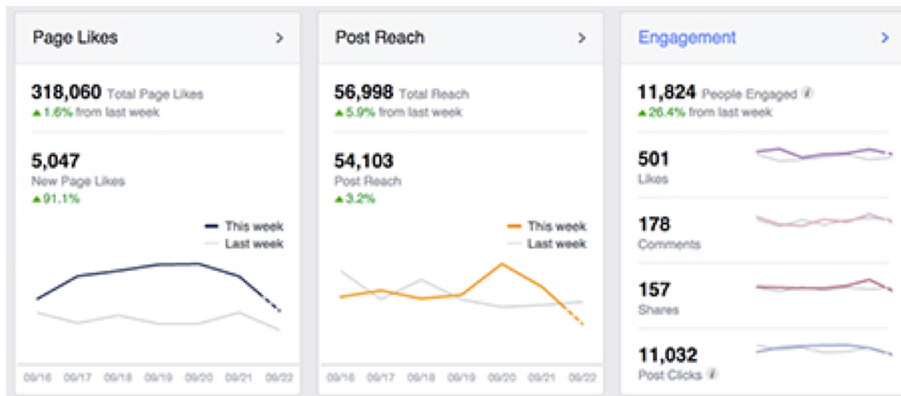
Aquesta secció proporciona una instantània dels últims set dies del rendiment de la *Fan Page*. Se centra en tres àrees bàsiques:

- *Page likes*: total i nous *likes* per a la pàgina.
- Publicació d'abast (*Reach*): nombre total de persones úniques a qui s'ha mostrat la pàgina i les publicacions.
- Compromís (*Engagement*): nombre total de persones úniques que van participar amb la pàgina, així com diferents tipus d'interacció.

També podem veure les cinc publicacions més recents i fer servir més eines:

- Pàgines per veure (*Pages to Watch*): veure com funcionen les empreses/entitats similars a la nostra. Per exemple, si som un detallista de joies, podem publicar idees basant-nos en entrades atractives d'altres minoristes.
- Aplicació del gestor de pàgines (*Pages Manager App*): Es pot descarregar l'aplicació gratuïta de la botiga iTunes Store o Google Play.

Figura 3. Secció Resum



## b) Agradaments (*likes*)

A la secció *likes* veurem tres mètriques bàsiques:

- *Page likes*: el total de *likes* de la pàgina per dia, durant un període de vint-i-vuit dies.
- *Net likes*: el nombre de nous *likes* menys el nombre de clics no relacionats.
- On es van produir els *likes* de la pàgina (*Where Your Page Likes Happened*): el nombre de vegades que li ha agradat la pàgina, desglossat per on va passar.

Figura 4. Secció Agradaments



## c) Abast (*Reach*)

La secció *Reach* mostra:

- Abast de la publicació (*Post Reach*): la quantitat de persones a les quals s'ha publicat la nostra publicació, desglossada per abast pagat i orgànic.
- Gustos, comentaris i accions (*likes, comments and shares*): el compromís positiu que ens ajuda a arribar a més persones.

- Amaga, informa com a correu brossa i «no m'agrada» (*Hide, Report as Spam and Unlikes*): el compromís negatiu disminuirà l'abast del nombre de persones a les quals arribem.
- Abast total (*Total Reach*): la quantitat de persones a les quals li ha servit qualsevol activitat de la vostra pàgina.

Figura 5. Secció Abast



#### d) Visites

Sota visites, s'hi veuen els següents aspectes:

- Visites de pàgines i pestanyes (*Page and Tab Visits*): el nombre de vegades que es va veure cada pestanya de la pàgina.
- Referències externes (*External Referrers*): el nombre de vegades que la gent va arribar a la pàgina des d'un lloc web de Facebook.

Cal omplir la pàgina amb informació bàsica perquè la puguin trobar les persones que més ens importen.

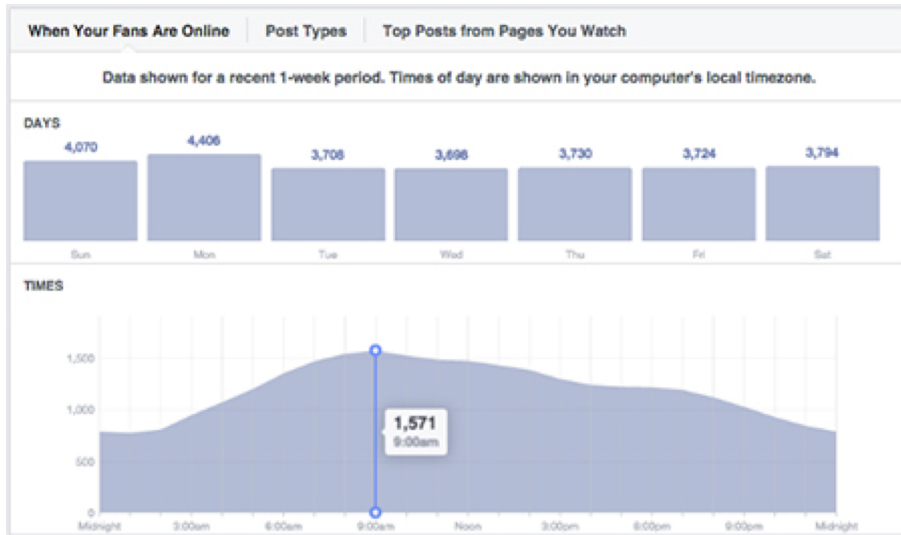
#### e) Posts

La pestanya «Publicacions» es divideix en les tres seccions següents:

- Quan els teus fans estiguin en línia (*When Your Fans Are Online*): us mostra quan les persones a qui els agrada la pàgina són a Facebook.
- Tipus de missatges (*Posts Type*): mostra l'èxit dels diferents tipus de publicació basats en l'abast mitjà i el compromís.
- Publicacions principals de les pàgines que visualitzeu (*Top Posts from Pages You Watch*): ofereix la participació de les publicacions de les pàgines que estem observant.

És possible programar les publicacions perquè apareguin quan el públic de la pàgina acostuma a estar en línia. També es pot revisar els tipus de publicacions que agraden més al públic, com ara publicacions de vincles o publicacions de fotografies.

Figura 6. Secció *Posts*



## f) Vídeo

La secció de vídeos mostra els aspectes següents:

- Vistes de vídeo (*Video Views*): la quantitat de vegades que els vídeos de la nostra pàgina s'han vist durant tres segons o més.
- Vistes de trenta segons (*30-Second Views*): la quantitat de vegades que els vídeos de la nostra pàgina s'han vist durant trenta segons o més. Si un vídeo té menys de trenta segons, es calcula quan la gent en veu el 97%.
- Vídeos principals (*Top Videos*): els vídeos més vistos a la pàgina que hem mirat durant tres segons o més.

## g) Gent (*People*)

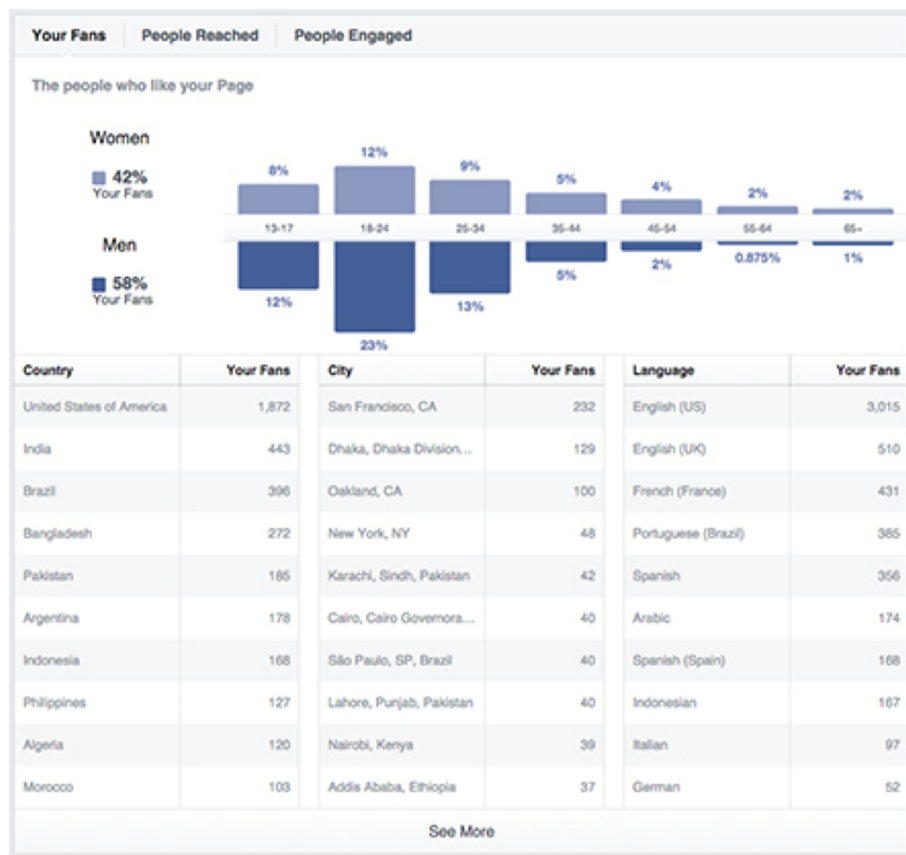
Aquesta pestanya divideix la gent entre: fans o seguidors, gent a la que arribem i gent compromesa amb la marca/entitat.

- **Seguidors** (*Your Fans*): es visualitza el gènere, l'edat, la ubicació i l'idioma de les persones a qui els agrada la pàgina.
- **Gent a la que arribem** (*People Reached*): consulteu les persones a les quals els ha aparegut la nostra publicació durant els últims vint-i-vuit dies.

- **Gent compromesa amb la marca (People Engaged):** és possible esbrinar a qui ha agradat, qui ha comentat, qui ha compartit les publicacions o ha participat amb la pàgina durant els últims vint-i-vuit dies.

Un cop hem conegut qui és el nostre públic més compromès, s'hi poden adaptar millor les publicacions avançades. Les vostres publicacions avançades s'haurien d'orientar al públic més compromès per mantenir-los-hi interessats. Per exemple, Allison Park Store va canviar les seves publicacions avançades per estar menys enfocades a interessos masculins com el golf, quan va saber que la majoria dels seus seguidors eren dones.

Figura 7. Secció Gent



## 2) Panells de control Twitter

A continuació veurem les mètriques que ofereix Twitter i les dades que en podem obtenir per analitzar la nostra xarxa social.

- **Activitat del tweet:** en mesura la interacció i així es pot saber com podem fer que els tweets tinguin més èxit.
- **Seguidors:** explora els interessos, ubicacions i demografia dels seguidors.
- **Cards de Twitter:** vigila la forma en què les Cards de Twitter motiven clics, descàrregues d'app i retweets.

## 2.2. Adquisició de les dades: d'on, com i sobre què extraïem dades

El segon esglaió de la cadena de valor és l'obtenció de dades, generades per diferents tipus d'objectes.

Alguns dels reptes clau en l'adquisició de dades de les xarxes socials són els següents:

a) **Dades no estructurades:** les dades de les xarxes socials estan molt desestructurades, ja que en contenen de diversos tipus, en lloc d'un text estructurat senzill. Depenent de la xarxa social, podem tenir dades de perfil, connexions, etiquetes, missatges publicats, recomanacions, comunitats i l'historial de navegació. Però també contenen informació multimèdia, com ara àudio o vídeo sobre text.

b) **Fonts d'informació heterogènies:** les xarxes socials són molt heterogènies amb diferents tipus de xarxes d'informació. Contenen múltiples tipus de nodes i enllaços. Els formats de dades en diferents serveis de les xarxes socials sovint són variats i incompatibles. Aquesta heterogeneïtat és un punt important en la recollida de dades.

c) **Recollida dinàmica de dades:** les xarxes socials són altament dinàmiques i evolucionen constantment. L'anàlisi dinàmica de les xarxes socials ha tingut recentment molta importància i requereix, per tant, també una recollida dinàmica de dades. Aquesta recollida dinàmica i incremental de les dades podria contribuir a una millor extracció de coneixement a partir de les dades. Tanmateix, aquesta adquisició dinàmica de dades necessita considerar diverses característiques temporals, com ara la durada de la recollida, els diferents intervals de temps en què es recolliran les dades i la taxa de recaptació. Aquests aspectes s'han de tenir en compte en el pla del *social media*, i en el projecte de recerca i anàlisi de xarxes.

d) **Potència de processament i espai d'emmagatzematge disponible:** la proliferació de les xarxes socials, pel que fa al nombre d'usuaris i publicacions, ha provocat el creixement exponencial de les dades de xarxes socials. Per a la manipulació d'aquestes dades és necessari un sistema robust que tingui un alt poder de processament i un bon espai d'emmagatzematge.

L'obtenció de dades de les xarxes socials es fa mitjançant diversos mètodes, tècniques i eines. En aquest apartat en veurem dos aspectes:

- Les accions que s'han de fer per extreure i gestionar dades i les eines que s'hi utilitzen.
- Les tecnologies que donen suport a aquesta fase.



### 2.2.1. Accions

Per obtenir dades, hem d'efectuar un conjunt d'accions. Vegem-les amb més detall, tot seguit:

1) Definir els objectius que es volen cobrir i les preguntes que volem respondre analitzant les dades. Haurem de definir el tema i les preguntes de recerca que volem respondre durant l'anàlisi de les xarxes socials. Les preguntes que farem són: què, quins, com, on, etc.

2) Identificar les fonts on es generen les dades que es necessiten. Per tant, concretar quines xarxes socials emprarem.

3) Extracció de dades mitjançant les API (Application Programmable Interface) pròpies de les xarxes socials (Facebook Graph API, Twitter API, Facebook Insights) o amb l'ajuda de programes lliures com R i el llenguatge de programació Python. Aquests programaris i llenguatge ens permeten interaccionar amb aquestes API d'una forma més senzilla gràcies als paquets que alguns usuaris han elaborat i han compartit. També podem trobar eines de pagament que a més d'extreure dades ens permeten analitzar-les i visualitzar-les.

### 2.2.2. Eines

Les eines que emprarem per descarregar les dades de les xarxes són diverses, i varien en la complexitat i prestacions. En aquest curs treballarem amb les eines més bàsiques, per tal de familiaritzar-nos amb la gestió de dades. Tanmateix, veurem a nivell descriptiu eines més complexes, per tal que en un futur estigüeu en disposició d'aprofundir-hi.

Eines **bàsiques** són:

- Netvizz per a descarregar dades de Facebook.
- Flocker per a descarregar dades de Twitter.

Les eines més **complexes** són les que descarreguen dades mitjançant API (Application Programmable Interface) i requereixen coneixements de llenguatge de programació, com Python. Es tracta de Twitter API, R, RStudio, R-commander, Facepy i Rfacebook.

Aquestes eines les veurem en els següents apartats.

#### **Adquisició mitjançant eines bàsiques**

Les eines bàsiques amb les que treballarem són Netvizz per a descarregar dades de Facebook, i Flocker per descarregar dades de Twitter. Vegem-les tot seguit.

## 1) Netvizz per a Facebook

Les polítiques de Facebook en material d'accés a les dades són bastant restrictives. Les dades privades no són accessibles, mentre que part de les dades públiques són accessibles mitjançant l'API. Facebook introdueix sovint canvis i noves restriccions a l'API, així que sempre haurem de cercar informació i eines actualitzades per a evitar incórrer en problemes.

És possible desenvolupar codi amb algun llenguatge de programació, com ara Python o R, per a accedir directament a les API, però això requereix un cert coneixement tècnic. Qui estigui interessat en desenvolupar el seu propi codi per a accedir a les API de Facebook trobarà una explicació detallada a la propera secció (opcional). Aquí, veurem com obtenir dades sense necessitat de programar.

Per tal de recopilar dades de Facebook, la manera més senzilla és amb una eina que es connecti a l'API. L'eina actualment més completa i documentada és Netvizz, desenvolupada per Bernhard Rieder a la Universitat d'Amsterdam - Digital Methods Initiative.

A continuació, veurem els passos necessaris per a recopilar dades de Facebook amb Netvizz.

### a) Descàrrega de dades d'activitat en un grup o pàgina pública

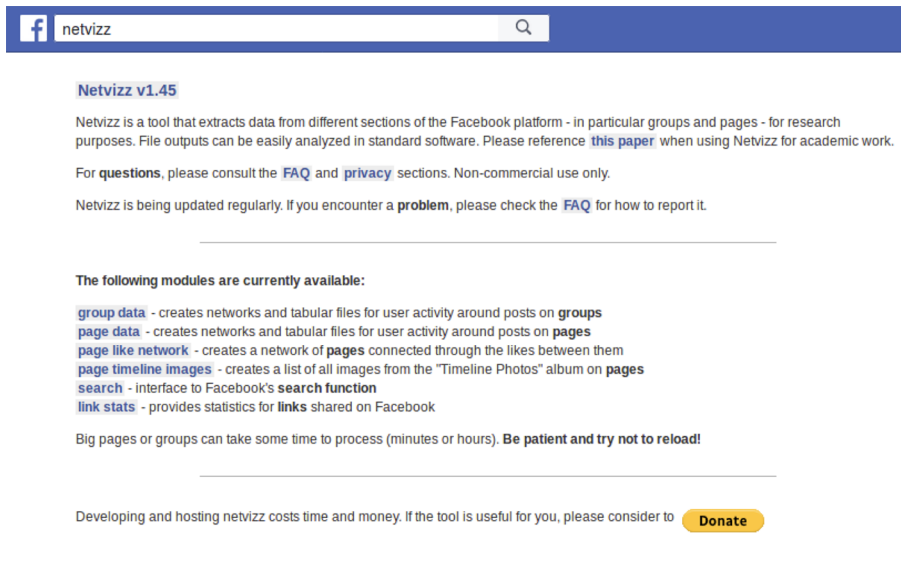
Ara veurem els passos necessaris per a obtenir dades sobre l'activitat d'un grup o pàgina de Facebook.

#### **Pas 1: instal·lar NetVizz**

Per a recopilar dades es necessita un compte de Facebook. Un cop hem entrat al nostre compte, escrivim Netvizz a la barra de recerca. Instal·lem aquesta aplicació i n'acceptem les condicions d'us.

Apareixerà una pantalla semblant a aquesta:

Figura 8



## Pas 2: seleccionar el tipus de dades

En l'aplicació s'hi presenten sis opcions:

- *group data*: per a recopilar dades d'activitat a un grup públic;
- *page data*: per a recopilar dades d'activitat a una pàgina pública;
- *page like network*, per a recopilar la xarxa de les connexions de *like* entre pàgines, a partir d'una pàgina donada;
- *page timeline images*, per a recopilar les imatges d'una pàgina pública;
- *search*, per a accedir a la funció de cerca de Facebook;
- *link stats*, per a recopilar estadístiques sobre enllaços compartits al Facebook.

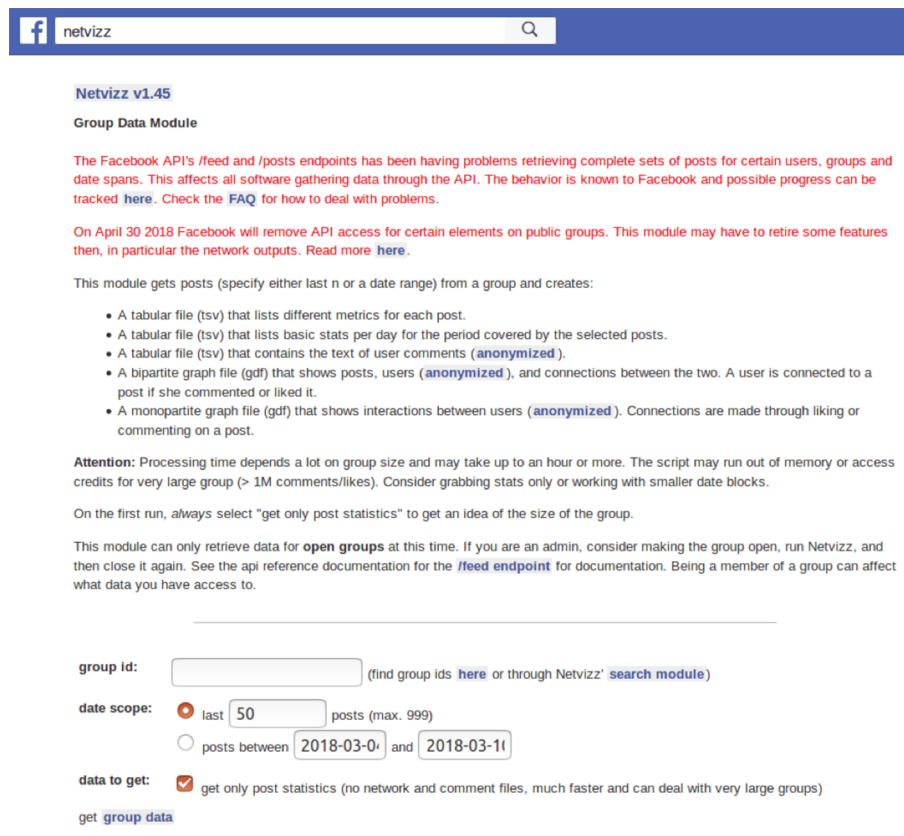
Per a aquest exercici escollirem entre *group data* i *page data*.

En els dos casos, la mateixa aplicació de Netvizz anonimitza els usuaris per raons de privacitat, de manera que veureu etiquetes amb números i no els noms dels usuaris.

## Pas 3: escollir el grup o la pàgina

Escollint l'opció *group data* ens trobem amb aquesta pàgina, que explica el seu funcionament i el format d'output, alhora que permet especificar-ne els paràmetres.

Figura 9



**Netvizz v1.45**  
**Group Data Module**

The Facebook API's /feed and /posts endpoints has been having problems retrieving complete sets of posts for certain users, groups and date spans. This affects all software gathering data through the API. The behavior is known to Facebook and possible progress can be tracked [here](#). Check the [FAQ](#) for how to deal with problems.

On April 30 2018 Facebook will remove API access for certain elements on public groups. This module may have to retire some features then, in particular the network outputs. Read more [here](#).

This module gets posts (specify either last n or a date range) from a group and creates:

- A tabular file (tsv) that lists different metrics for each post.
- A tabular file (tsv) that lists basic stats per day for the period covered by the selected posts.
- A tabular file (tsv) that contains the text of user comments ([anonymized](#)).
- A bipartite graph file (gdf) that shows posts, users ([anonymized](#)), and connections between the two. A user is connected to a post if she commented or liked it.
- A monopartite graph file (gdf) that shows interactions between users ([anonymized](#)). Connections are made through liking or commenting on a post.

**Attention:** Processing time depends a lot on group size and may take up to an hour or more. The script may run out of memory or access credits for very large group (> 1M comments/likes). Consider grabbing stats only or working with smaller date blocks.

On the first run, always select "get only post statistics" to get an idea of the size of the group.

This module can only retrieve data for **open groups** at this time. If you are an admin, consider making the group open, run Netvizz, and then close it again. See the api reference documentation for the [/feed endpoint](#) for documentation. Being a member of a group can affect what data you have access to.

---

**group id:**  (find group ids [here](#) or through Netvizz' [search module](#))

**date scope:**  last  posts (max. 999)  
 posts between  and

**data to get:**  get only post statistics (no network and comment files, much faster and can deal with very large groups)

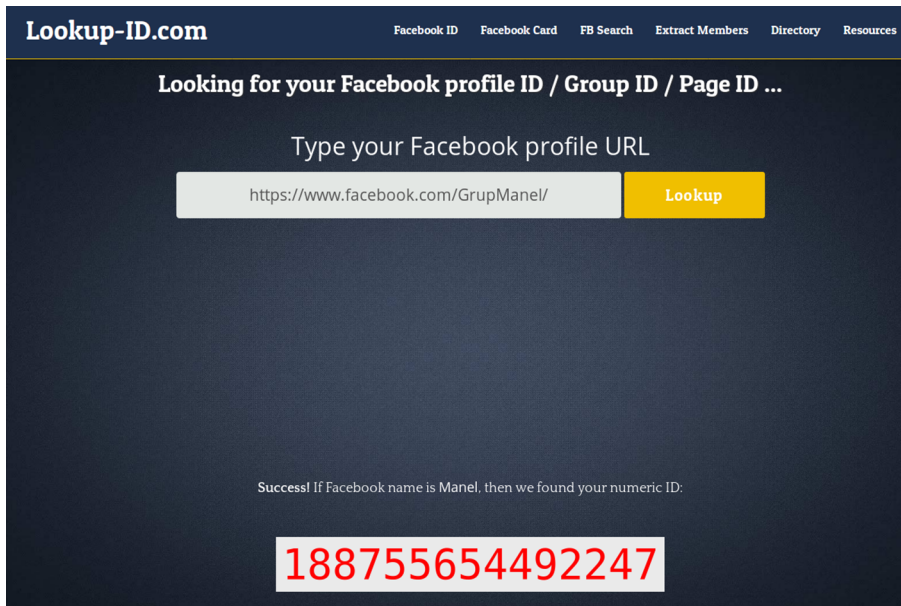
[get group data](#)

En aquest exemple farem servir la pàgina del grup musical Manel.

#### Pas 4: obtenir l'ID del grup (o pàgina) escollit

Com s'explica a la pàgina, per a poder obtenir les dades, cal introduir l'ID numèric del grup (o pàgina) escollit. Per tal d'obtenir l'ID corresponent a un grup o pàgina, es pot fer servir el servei Lookup-ID.com. Aquí, posant-hi l'URL d'un grup o pàgina de Facebook, es pot trobar l'ID corresponent, tal com s'ensenya a la figura següent.

Figura 10. ID del grup Manel



Un cop obtingut l'ID, podem copiar-lo i passar a realitzar la consulta.

#### Pas 5: realitzar la consulta

Ara ja tenim tots els elements per a poder realitzar la consulta indicant els valors dels diversos paràmetres. Per exemple, per a obtenir totes les dades d'activitat a la pàgina de Manel de l'última setmana, considerant tant *posts* de la pàgina com dels seus fans, realitzem la consulta que us ensenyem a la figura següent.

Figura 11



**Netvizz v1.45**

**Page Data Module**

On February 5 2018 Facebook has removed API access for a number of elements on public pages. This includes fans per country and all user information, which means user-post bipartite graphs can no longer be generated and users can no longer be distinguished in comment files.

This module gets posts (specify either last n or a date range) on a page and creates a number of files:

- A tabular file (tsv) that lists a series of metrics for each post.
- A tabular file (tsv) that lists basic stats per day for the period covered by the selected posts, including reactions per post.
- A tabular file (tsv) that contains the text of user comments (no user information).

**Attention:** Processing time depends a lot on page size and may take up to an hour or more. The script may run out of memory or access credits for very large pages (> 1M comments/likes). Consider grabbing stats only or working with smaller date blocks. On the first run, always select "post statistics only" to get an idea of the size of the page.

Which posts are retrieved may depend on whether you like the page or not.

The Facebook API's /feed and /post endpoints have stopped retrieving complete sets of posts for certain users, pages and date spans. This can affect all software gathering data through the API. [According to Facebook](#), these endpoints now show the same posts the logged user would see on the page surface.

See the api reference documentation for the [page/feed endpoint](#).

Check the [FAQ](#) for how to deal with problems.

---

**page id:**  (find page ids [here](#) or through Netvizz' [search module](#))

**date scope:**

last  posts (max. 999)

posts between  and

**data to get:**

post statistics only (post metrics, stats per day ~~and fans per country~~)

post statistics and 200 top ranked comments per post

full data (full ~~network~~ and comment files, can fail for larger pages)

get [post by page only](#) or [posts by page and users](#)

---

## Pas 6: desar l'arxiu

Ara podem guardar l'arxiu zip que conté els resultats:

Figura 12



**Netvizz v1.45**

Trying to get 50 posts.

pid: 188755654492247 / until: 2017-09-07T17:01:34+0000 - 50 of 50 (50)

Retrieved data for 50 posts.

comments retrieved.

**Download**

Compressing files...

page\_188755654492247\_2018\_03\_11\_22\_08\_25.tab  
 page\_188755654492247\_2018\_03\_11\_22\_08\_25\_statsperday.tab  
 page\_188755654492247\_2018\_03\_11\_22\_08\_25\_fanspercountry.tab

Your files have been generated. 2 files were zipped. Download the [zip archive](#).

For file descriptions, refer to the main module page and for any problems check the [FAQ](#).

Fixeu-vos bé en la carpeta on guardeu aquest arxiu i anomeu-lo. Preferentment, poseu-hi un nom que identifiqueu amb facilitat.

Cal comentar que, com està indicat en vermell a la interfície del Netvizz, uns canvis recents a l'API de Facebook han posat una ulterior restricció, així que ja no és possible obtenir la xarxa de les interaccions a un grup o pàgina de Facebook, sinó només les estadístiques sobre l'activitat al voltant de cada *post*.

L'arxiu comprimit que hem desat conté dos fitxers csv, en què cada línia correspon a un *post*, i en mostra diverses característiques i estadístiques, com el que s'indica a la pàgina de l'aplicació, i el que reportem a continuació.

### **post stat file - tsv format - rows are posts**

*type*: Facebook's post classification (e.g. photo, status, etc.)

*by*: either "post\_page\_pageid" (post by page) or "post\_user\_pageid" (post by user);

*post\_link*: direct link to the post;

*post\_message*: text of the post;

*picture*: the picture scraped from any link included with the post;

*full\_picture*: the picture scraped from any link included with the post (full size);

*link*: link URL (if the post points to external content);

*link\_domain*: domain name of link;

*post\_published*: publishing date

*post\_published\_unix*: publishing date as Unix timestamp (for easy conversion and ranking);

*post\_published\_sql*: publishing date in SQL format (some analysis tools prefer this);

*likes*: number of actually retrieved likes a post received or a user made;

*likes\_count\_fb*: Facebook provided like count for posts;

*comments\_count\_fb*: Facebook provided comment count for posts;

*reactions\_count\_fb*: Facebook provided reactions count for posts (includes likes);

*shares\_count\_fb*: Facebook provided share count for posts;

*engagement*: sum of comment, reaction, and share counts;

*comments\_retrieved* (not in "stats\_only" mode): actually retrieved comments (may be lower than *comments\_count\_fb* due to privacy or deletion);

*comments\_base* (not in "stats\_only" mode): number of base level comments (in threaded conversations);

*comments\_replies* (not in "stats\_only" mode): number of reply level comments (in threaded conversations);

*comment\_likes\_count* (not in "stats\_only" mode): number of likes on comments for this post;

*rea\_X*: these columns give counts for the retrieved reactions (NONE, LIKE, LOVE, WOW, HAHA, SAD, ANGRY, THANKFUL);

#### **b) Xarxa de *likes* entre pàgines**

De manera anàloga, en lloc de *group data* o *page data* podem escollir l'opció *page like network*, amb la qual obtindrem la xarxa de *likes* entre pàgines, a partir d'una pàgina donada.



Figura 13

**Netvizz v1.45**

**Page Like Network Module**

This module starts with a selected page (the "seed") and retrieves all the pages that page likes. It will continue until the specified crawl depth is reached (currently limited to 2). The output is a network file (gdf format) containing a (directed) network of pages. Because node ids are unique, you can combine several networks in gephi.

You can now also use several seed ids, separated by a comma. If several ids are used, crawl depth is reduced to a max of 1. Use a crawl depth of 0 to get only the relations between your seeds.

188755654492247 page id (find page ids [here](#) or through Netvizz' [search module](#))

1 depth (max 2)

[start](#)

**page network (directed) - gdf format - nodes are pages**  
 name: the page id  
 label: the page name  
 username: the page username  
 category: page category according to Facebook's ontology  
 post\_activity: posts per hour, based on the last 50 posts  
 fan\_count: number of likes a page has received  
 talking\_about\_count: current attention metric provided by Facebook  
 users\_can\_post: whether page users are allowed to publish posts on the page  
 link: a direct link to the page

L'app comprovarà els *likes* de la pàgina indicada a altres pàgines, i al seu torn els *likes* de totes aquestes pàgines a altres. Si es canvia el valor del paràmetre de la profunditat d'exploració (paràmetre *depth*) d'1 a 2 es realitzarà una altra iteració a partir de les pàgines trobades.

El resultat en aquest cas és un arxiu de xarxa en format gdf, que es pot obrir amb el programa Gephi, com veurem a l'assignatura Visualitzar i entendre interaccions a les xarxes socials.

## 2) Flocker per a Twitter

Si amb Facebook hem vist que hi ha moltes restriccions a l'hora de recopilar dades de manera automàtica, les polítiques de Twitter són més obertes. En particular, és possible obtenir lliurement fins a l'1% del volum total de piulades en un moment determinat, sempre que es recopilin en temps real.

Per a obtenir dades de Twitter de manera senzilla, farem servir el servei web Flocker desenvolupat pel col·lectiu Outlier.

Com en el cas de Facebook, per a poder recopilar dades cal tenir un compte, i accedir-hi amb aquest.

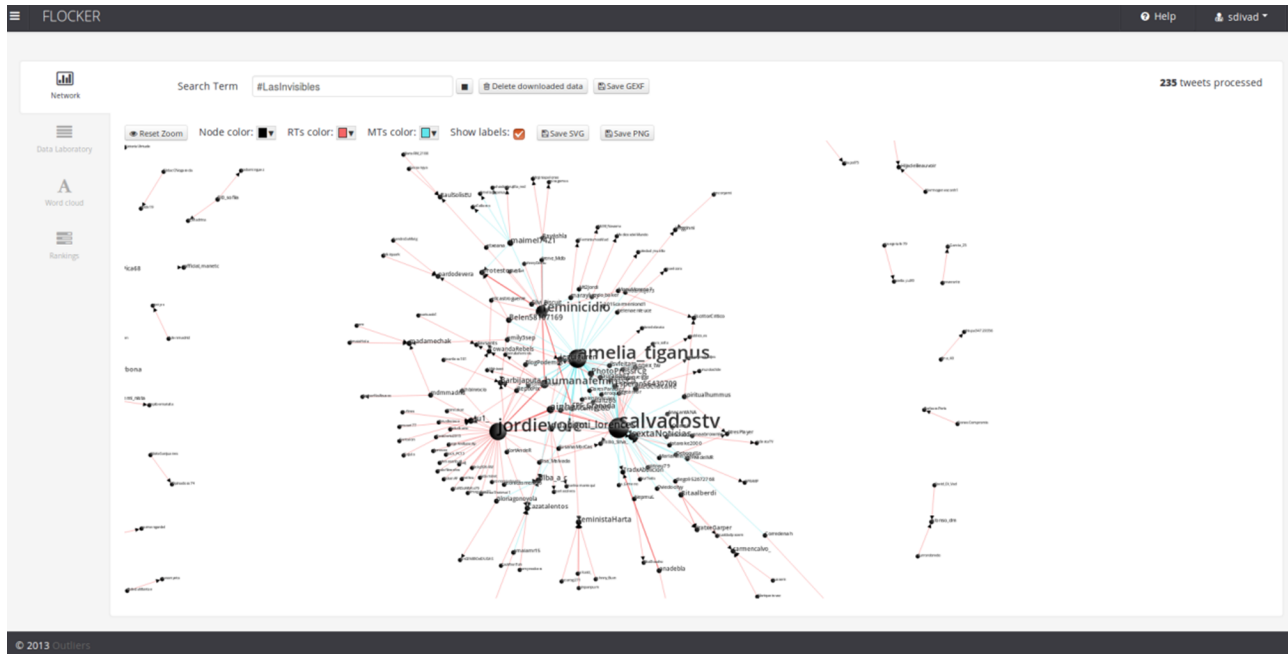
Un cop introduït el *log in* amb el nostre compte de Twitter, podem inserir un terme de cerca, com ara un *hashtag*, per a començar a monitorar tots els continguts al voltant d'aquesta paraula clau en temps real.

### Vegeu també

El programa Gephi es tracta a l'assignatura Visualitzar i entendre interaccions a les xarxes socials.

L'eina dibuixa una xarxa en què cada node és una usuària o un usuari de Twitter, i cada enllaç una interacció entre dos usuaris, com ara una repulada o una menció. Els colors dels enllaços representen el tipus d'interacció, i la mida d'un node depèn del seu número de connexions.

Figura 14



La figura ensenya un exemple amb el *hashtag* #LasInvisibles, relacionat amb un programa de *Salvados*.

En qualsevol moment es pot descarregar un fitxer en format gexf, que conté la xarxa d'interaccions generada fins a aquell punt. També es pot aturar el monitoratge amb el botó a la dreta del terme de cerca.

A l'assignatura Visualitzar i entendre interaccions a les xarxes socials aprendrem a obrir el fitxer generat amb el programa Gephi. Això ens ajudarà a analitzar i visualitzar la xarxa.

### **Adquisició mitjançant les API (opcional)**

Hi ha un seguit d'eines més complexes que requereixen més coneixement de programació. Els exposem a continuació, a títol informatiu, però sense l'objectiu de treballar-los de manera pràctica durant l'assignatura.

#### **1) Twitter API**

Per poder usar l'API de Twitter serà necessari crear l'objecte de TwitterAPI amb les *key* generades per l'API de Twitter.

Un cop generat l'objecte, podrem fer trucades a les diferents funcions que ens ofereix l'API de Twitter.

Realitzant la consulta rebrem un JSON (JavaScript Object Notation), un format de text lleuger per a l'intercanvi de dades. Per a poder generar el graf a partir de les dades recuperades d'aquest json, i aplicar-hi les diferents mesures que escollim, es pot usar per exemple el *framework* de Python o el paquet de R per al tractament de grafs anomenat Igraph. Al mòdul 3 ho veurem detalladament.

Python és un llenguatge de programació interpretat, que és molt adequat per a començar a programar, ja que es tracta d'un llenguatge molt potent però amb una sintaxi simple i clara. Va ser creat per Guido van Rossum l'any 1991.

## 2) Instal·lar Python al nostre ordinador

La versió més actual és la versió 3 de Python, encara que la versió 2 encara es manté i es desenvolupa. Les instruccions per instal·lar Python en el nostre ordinador són les següents:

- Ves a la pàgina web <https://www.python.org/downloads/windows/> i prem el botó «Latest Python 3 Release». Comprova que sigui la darrera versió de Python 3 i no la Python 2.
- Ubuntu o altres distribucions derivades de Debian: obre un terminal i executa-hi l'ordre següent: `sudo apt install python3 idle3`.

## 3) Instal·lació R, RStudio i R-commander

R és un programa estadístic de codi obert. Pot executar-se en qualsevol sistema operatiu (Window, MAC o Linux).

RStudio és un entorn integrat de desenvolupament per a R. És un entorn agradable on executar R.

Rcommander és una interfície gràfica d'usuari per a R. Ofereix, mitjançant els menús, accés a moltes de les funcionalitats d'R.

Vegem com instal·lar aquests programes.

### a) Instal·lació de R:

- Ves a <http://cran.at.rproject.org/>.
- Selecciona el sistema que tinguis instal·lat en el teu ordinador (Windows, MAC o Linux).
- Instal·la'l utilitzant les opcions que ofereix per defecte.

**b) Instal·lació RStudio:**

- Ves a <http://www.rstudio.com/products/rstudio/download/>.
- Selecciona el sistema que tinguis instal·lat en el teu ordinador (Windows, MAC o Linux).
- Instal·la'l utilitzant les opcions que ofereix per defecte.

**c) Instal·lació R-commander:**

- (Només si treballes amb MAC). Descarrega i instal·la la llibreria `tcltk8.xx1.dmg` de <http://cran.rproject.org/bin/macosx/tools/>.
- Obre l'RStudio.
- Ves a la pestanya «Packages» i clica a «Install Packages».
- Comença a escriure «Rcmdr» a la finestra fins que vegis que apareix a la llista i instal·la el paquet.

**4) Facebook - Graph API**

La Graph API (Application Programmable Interface) de Facebook és la manera principal d'obtenir el graf social de Facebook. És una API que s'utilitza per consultar dades, publicar notícies noves, pujar fotos, extreure el nombre de m'agrada o *likes* i una gran varietat de tasques més.

Per poder extreure dades de Facebook utilitzant Graph API, un codi Python o R, cal registrar-se com a desenvolupador a Facebook i després obtenir un *token* d'accés. Aquests són els passos que cal seguir:

- Ves a l'enllaç [developers.facebook.com](https://developers.facebook.com) i crea-hi un compte.
- Ves a l'enllaç [developers.facebook.com/tools/explorer](https://developers.facebook.com/tools/explorer).
- Ves a «Les meves aplicacions» («My apps») a l'extrem superior dret i selecciona «Afegeix una nova aplicació» («add a new app»). Tria un nom de visualització i una categoria i, a continuació, «Crea un identificador d'aplicació» («Create App ID»).
- Torna a l'enllaç [developers.facebook.com/tools/explorer](https://developers.facebook.com/tools/explorer). Veuràs «Graph API Explorer» a sota de «Les meves aplicacions», a l'extrem superior dret. Des del desplegable de «Graph API Explorer», selecciona la teva aplicació.
- Selecciona «Obtenir *token*» («Get Token»). A partir d'aquest menú desplegable, selecciona «Obtenir *token* d'accés d'usuari» («Get User Access Token»). Selecciona els permisos del menú que hi apareix i selecciona «Obtenir *token* d'accés» («Get Access Token»).
- Ves a l'enllaç [developers.facebook.com/tools/accesstoken](https://developers.facebook.com/tools/accesstoken). Selecciona-hi «Depura» («Debug»), corresponent al «User Token». Ves a «Ampliar l'accés

a *token*» («Extend Token Access»). Això garantirà que el *token* no caduqui cada dues hores.

### a) Llegir

Amb una sol·licitud a Graph API, HTTP GET es poden obtenir tots els nodes i arestes (*edges*). Per exemple, si voleu recuperar informació sobre l'usuari actual, es realitzarà una sol·licitud HTTP GET de la manera següent:

```
GET /v2.5/me HTTP/1.1
Host: graph.facebook.com
```

La majoria de les sol·licituds a l'API s'han de signar amb un identificador d'accés (*access token*). Per determinar quins permisos són necessaris en aquest identificador d'accés, es pot consultar la referència en aquest enllaç de la Graph API corresponent al node o aresta que desitgeu llegir. També podeu utilitzar l'explorador de la Graph API per generar ràpidament identificadors, amb l'objectiu d'explorar les diferents opcions de l'API i familiaritzar-vos amb el funcionament.

El node/me és un extrem especial que es tradueix en l'objecte *user\_id* de la persona (o l'objecte *page\_id* de la pàgina de Facebook) que l'identificador d'accés farà servir per fer sol·licituds a l'API. Si tinguessis un identificador d'accés d'usuari, podries recuperar totes les fotos d'un usuari mitjançant la següent sol·licitud:

```
GET graph.facebook.com/me?fields=id,name,picture
```

### b) Seleccionar camps

De forma predeterminada, quan es realitza una consulta no es retornen tots els camps d'un node o aresta. Podeu triar els camps o arestes que voleu que es retornin mitjançant els *fields* de consulta de paràmetre. Aquest paràmetre resulta molt útil per realitzar trucades més eficients i ràpides a l'API.

Per exemple, en una trucada a l'API Gràfic en què s'utilitza el vostre identificador d'accés d'usuari (*access token*): <https://graph.facebook.com/me?fields=id,name,picture> només es retornarà l'identificador, el nom i la foto del teu perfil:

```
GET graph.facebook.com
/ me / fotos
```

Podeu trobar més informació en el següent enllaç.

## 5) Facepy: accés a la Graph API de Facebook via Python

Facepy fa que sigui més senzill d'interactuar amb les API de Facebook. És una classe de Python. Tingueu en compte que Facepy no fa l'autenticació amb Facebook; només consumeix la seva API. Per obtenir un *token* d'accés per consumir l'API en nom d'un usuari, utilitzeu una biblioteca OAuth adequada per a la plataforma o el procediment que hem descrit a l'apartat anterior.

Per instal·lar Facepy cal tenir instal·lat Python i pip. Usant la Terminal o Consola i mitjançant pip, escriviu la següent ordre:

```
$ pip install facepy
```

Aquest seria l'script de Python per usar Facepy:



## 6) Rfacebook: accés a la API de Facebook via R

Aquest paquet d'R proporciona una sèrie de funcions que permeten als usuaris d'R accedir a l'API de Facebook per obtenir informació sobre pàgines, grups i publicacions públiques, així com algunes de les dades privades de l'usuari autenticat.

### a) Instal·lació i autenticació

Rfacebook es pot instal·lar directament des de CRAN (xarxa de ftp i servidors web distribuïts a tot el món que emmagatzemen versions idèntiques actualitzades de codi i documentació per a R. Tanmateix la versió més actualitzada sempre serà a GitHub. El codi següent mostra com instal·lar-se'l a partir de CRAN:

```
install.packages("Rfacebook") # from CRAN
```

La majoria de les sol·licituds de l'API requereixen l'ús d'un *token* d'accés. Hi ha dues maneres de fer sol·licituds autenticades amb Rfacebook. Una opció és generar un *token* temporal a la Graph API Explorer. A continuació, només copieu i enganxeu el codi a la consola R i deseu-lo com a vector de cadena per passar-ho com a argument a qualsevol funció de Rfacebook, com es mostra a continuació. Tanmateix, tingueu en compte que aquest *token* d'accés només serà vàlid durant dues hores.

```
# token generated here:
# https://developers.facebook.com/tools/explorer
token <- 'XXXXXXXX'
me <- getUsers("me", token, private_info=TRUE)
me$name # my name
# [1] "Pablo Barberá"
me$hometown # my hometown
# [1] "Cáceres, Spain"
```

És possible generar un *token* «de llarga durada» (vàlid durant dos mesos) usant la funció fbOAuth, però el procés és una mica més complicat. Per obtenir un tutorial pas a pas, consulteu aquesta publicació de blog de JulianHi.

L'exemple anterior mostra com recuperar informació sobre un usuari de Facebook. Tingueu en compte que això es pot aplicar per obtenir informació sobre qualsevol usuari (o vector d'usuaris) amb perfils públics, és a dir, pàgines. Tots dos noms de pantalla d'usuari o identificadors es poden utilitzar com a arguments, per exemple:

```
getUsers(c("barackobama", "donaldtrump"), token)
```

Aquesta funció no es pot utilitzar per obtenir informació del perfil dels usuaris amb perfils privats, a excepció dels perfils que utilitzen l'aplicació creada per l'usuari autenticat. Quan s'utilitza un *token* temporal, només es retorna la informació dels perfils que hagin creat un *token* temporal a la xarxa d'amics de l'usuari autenticat.

#### b) Analitzant la vostra xarxa d'amics

D'acord amb Barberá (2013), la funció getFriends permet que l'usuari capturi informació sobre els seus amics de Facebook, sempre que utilitzi l'aplicació creada per autenticar-se amb l'API. Atès que els identificadors d'usuari s'assignen en ordre consecutiu, és possible saber quins dels nostres amics van ser els primers a obrir un compte de Facebook.

```
my_friends <- getFriends(token, simplify = TRUE)
head(my_friends$id, n = 1) # get lowest user ID
```

Per accedir a informació addicional sobre una llista d'amics mitjançant l'aplicació *token*, podeu utilitzar la funció `getUsers`, que retornarà un marc de dades amb les dades de Facebook dels usuaris. Algunes de les variables disponibles per a tots els usuaris són: sexe, idioma i país. També és possible obtenir l'estat de relació, la ciutat natal, l'aniversari i la ubicació dels nostres amics si establim `private_info = TRUE`. (Tingueu en compte que el llenguatge i el país s'extreuen dels codis locals).

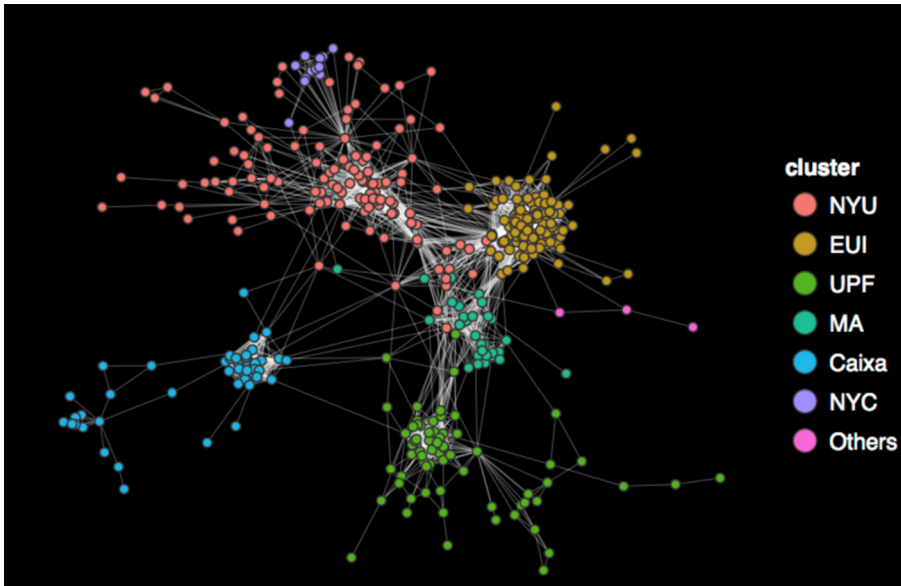
```
my_friends_info <- getUsers(my_friends$id, token, private_info=TRUE)
table(my_friends_info$gender) # gender
table(substr(my_friends_info$locale, 1, 2)) # language
table(substr(my_friends_info$locale, 4, 5)) # country
table(my_friends_info$relationship_status) # relationship status
```

Finalment, la funció `getNetwork` extreu una llista de totes les amistats recíproques entre els amics mitjançant l'aplicació amb la qual es va crear el testimoni. Aquesta llista es pot utilitzar per analitzar i visualitzar una xarxa d'ego de Facebook. El primer pas és utilitzar la funció `getNetwork`. Si l'opció de format està establerta igual a `Edgelist`, retornarà una llista de totes les vores d'aquesta xarxa. Si `format = adj.matrix`, retornarà una matriu adjacent de dimensions (nxn), amb n el nombre d'amics i 0 o 1 que indica si l'usuari a la fila 'i' també és amic de l'usuari a la columna 'j'.

```
mat <- getNetwork(token, format = "adj.matrix")
dim(mat) # dimension of matrix
```

Aquesta matriu de l'adjacència es pot convertir en un objecte `igraph`, que facilita la tasca de computar mesures de centralitat, descobrir comunitats o visualitzar l'estructura de la xarxa. Com a il·lustració, la trama que hi ha a continuació mostra una xarxa d'ego d'un usuari de Facebook, on els colors representen els clústers descoberts amb un algoritme de detecció de la comunitat, que es superposa amb la pertinença a comunitats fora de línia.



Figura 15. *Introducing the Rfacebook package*

P. Barberá (2013); <http://pablobarbera.com/blog/archives/3.html>.

La visualització i anàlisi de xarxes les veiem en altres assignatures.

### c) Cercar publicacions públiques a Facebook

Abans de la versió 2.0 de l'API de gràfics, es podia recopilar publicacions a Facebook que esmentaven paraules clau específiques, utilitzant la funció search-Facebook. Aquesta opció ja no està disponible mitjançant l'API pública.

En l'assignatura Visualitzar i entendre interaccions a les xarxes socials veurem com analitzar dades d'una pàgina de Facebook mitjançant aquesta eina.

## 2.3. Emmagatzematge de les dades

En aquest subapartat veiem on podem guardar el conjunt de dades i com fer-ho per tal de poder-les recuperar quan ho necessitem. Cal emmagatzemar les dades amb l'estructura més adequada per a la seva tipologia i també depenent de l'ús que en farem posteriorment.

Les dades es poden emmagatzemar al núvol si tenim una gran quantitat de dades, com per exemple: S3 (Amazon Web Services), Google Cloud Storage (Google) o Azure Storage (Microsoft).

Tenint en compte la tipologia de les dades de xarxes socials, que acostumen a ser no estructurades, podríem guardar-les usant bases de dades orientades a grafs (BDOG) com per exemple Neo4J. Aquest tipus de bases de dades representen la informació com a nodes d'un graf, i les seves relacions com a arestes d'aquest graf, de manera que es pot fer ús de la teoria de grafs per tal de recórrer la base de dades. Aquestes bases de dades permeten:

- Consultes més àmplies i no delimitades per taules, per exemple «Mostri totes les dades que continguin el nom *Anna*».
- No cal definir un nombre determinat d'atributs, es podria donar el cas que una persona tingui quatre noms relacionats i en canvi una altra només dos, sense malgastar espai.
- No cal definir una mida pels registres, fet que evita també possibles fallades a la base de dades.
- Permet molts tipus de recorreguts per les dades, per exemple, el jeràrquic, que obté el node avi i viceversa.

## **2.4. Anàlisi de xarxes socials. Gephi, Pajek, R, CiteSpace, TouchGraph**

L'anàlisi de les dades que obtenim de les xarxes socials es fa mitjançant la teoria de xarxes socials i aplicant una sèrie de tècniques i eines tecnològiques. La teoria de les xarxes socials la veurem a l'assignatura Respondre preguntes de recerca analitzant xarxes. En aquest apartat presentem un ventall d'eines que permeten veure i analitzar les xarxes socials. Tot i que aquest curs treballarem amb Gephi, volem que conegueu l'existència d'altres.

La necessitat creixent de visualització de xarxes complexes ha propiciat el desenvolupament d'una gran col·lecció d'eines programari, aptes per a tots els públics.

Algunes estan dirigides a perfils més tècnics, però també en trobem d'intuïtives, adequades per a usuaris que no tenen coneixements de programació.

La majoria permeten el càlcul de mesures locals i globals de la xarxa, la seva visualització gràfica i la detecció de comunitats i, en general:

- Creació de xarxes.
- Visualització i manipulació de xarxes.
- Anàlisi qualitatiu i quantitatiu/estadístic de xarxes.
- Detecció de comunitats.
- Anàlisi predictiu (models de contagi, predicció d'enllaços, etc.).

Les eines més populars són Gephi, Pajek, CiteSpace, TouchGraph i les diferents biblioteques desenvolupades en R (igraph, sna, TNET, statnet i NetworkX).

A continuació teniu una taula comparativa de les eines que hem comentat:

Taula 1. Comparativa d'eines per analitzar xarxes socials

<b>Programari Versió</b>	<b>Gephi 0.7 alpha</b>	<b>Pajek 1.26</b>	<b>CiteSpace</b>	<b>TouchGraph</b>	<b>igraph</b>	<b>NetworkX 0.6</b>
<b>Tipus</b>	Programari stand-alone	Programari stand-alone			Llibreria	Llibreria
<b>Plataforma</b>	Java	Windows	Java	Java	R	Python
<b>Llicència</b>	GNU GPL	Lliure per a ús no comercial	Lliure	Lliure	Lliure	BSD Llicència
<b>Temps computació esperat</b>	Mitjà (Java)	Ràpid (C)			Ràpid (R)	Ràpid (Python)
<b>Núm. nodes tractables</b>	150.000	500.000				1.000.000
<b>Temps per carregar 10<sup>5</sup> nodes i 10<sup>6</sup> arestes</b>	40 segons	24 segons				137 segons
<b>Formats</b>						
GML	Sí	No		No	Sí	Sí
Pajek (.net)	Només importar	Sí		Sí	Sí	Sí
GraphML	Sí	Només exportar			Sí	Sí
DL	Sí	Sí			Sí	No
GEXF	Sí	Sí			Sí	No
Txt			Si		No	
CSV				Sí	No	
<b>Tipus grafs</b>						
2-mode grafs	No	Sí			Sí	Sí
Multirelacional grafs	No	Sí			No	No
Temporalitat	No	Sí			Sí	Sí
<b>Indicadors</b>						
Grau centralitat	Sí	Sí				Sí
Centralitat intermèdia	Sí	Sí				Sí
Centralitat de proximitat	Sí	Sí				Sí
HITS	Sí	No				Sí
Page Rank	Sí	No				Sí
<b>Algoritmes Clustering</b>						
Dendograma	Sí	Sí				No

Programari Versió	Gephi 0.7 alpha	Pajek 1.26	CiteSpace	TouchGraph	igraph	NetworkX 0.6
Valoracions						
Facilitat d'ús	Sí					
Aportació de coneixement	Sí					
Estadístiques i gràfiques						
Visió de l'estructura	Sí	Sí				
Modificació de la informació	Sí	Sí				

Font: Combe i altres (2010)

A continuació comentem els avantatges i desavantatges de Gephi i Pajek, dues de les eines més ben valorades per estudiar xarxes socials.

### 2.4.1. Pajek

Entre els **avantatges** d'aquesta eina destaquem les següents:

- És un *software* gratuït.
- És una de les eines amb més potencial pel fet que el graf es pot crear manualment i retocar, canviar, modificar o afegir el que necessitis al graf per poder veure la informació que es necessiti en cada moment. Es poden afegir diferents patrons o estadístiques per als nodes que facin diferències i que es necessitin veure dins de la xarxa que s'utilitza.
- Porta molts anys en el mercat tecnològic i està constantment rebent actualitzacions de millora que fan que cada dia l'eina sigui més completa per treure el major rendiment als grafs que s'estudiïn amb ella.

Tot i que és una eina potent, també presenta **desavantatges**. Alguns d'aquests els veiem tot seguit.

- Tot es realitza manualment. La forma d'utilitzar l'eina és creant nosaltres mateixos el graf que volem estudiar. Això fa que el temps de realització del graf en comparació amb la resta d'eines que ens ho fan de forma automàtica sigui molt més gran.
- Per a poder realitzar qualsevol tipus de graf professional o un graf de grans mesures, es requereixen coneixements avançats sobre l'eina per a poder realitzar la xarxa de forma correcta i adequada. Això fa que aquesta eina

tingui molt potencial però per a les persones que sàpiguen realment utilitzar-la.

- No ofereix cap tipus d'estadística ni dades que puguin ser rellevants per a un estudi seriós d'una xarxa social.

### 2.4.2. Gephi

D'entre els **avantatges** d'aquesta eina presentem els següents:

- És un *software* gratuït amb tot tipus de complements.
- Té infinitat de modificacions que pots aplicar al graf de la xarxa social que estiguis utilitzant. Aquesta eina és coneguda com el Photoshop dels grafs. Podem modificar-ho al nostre gust, depenent de l'estudi o la informació que necessitem treure de la nostra xarxa per veure-ho de forma més clara.
- Facilitat d'ús de la seva interfície. Totes les opcions que ens proporciona Gephi són còmodes, intuïtives i fàcils d'aplicar, fet que suposa un gran avantatge per a qualsevol usuari i fa possible el seu ús a nivell personal com a nivell professional en el cas que es necessités.
- És bastant ràpid, es pot importar qualsevol graf de forma ràpida i còmoda i analitzar-lo, modificar-lo o canviar-lo amb la mateixa rapidesa.
- S'hi poden analitzar diverses xarxes socials, per exemple: Facebook i Twitter.

Malgrat aquests avantatges, també trobem alguns **desavantatges**, que presentem tot seguit.

- Aquesta eina només es basa en la part estructural de qualsevol xarxa social. Això fa que no es pugui analitzar la base de dades de les persones que componen el graf, ni realitzar estadístiques ni treure gràfiques per tenir més informació pel que fa a les dades directament. És cert que aquesta eina mostra la base de dades i es pot treballar sobre ella. Tanmateix, només ens mostra els seus resultats sobre la base del graf creat i mai realitza estadístiques directes i gràfiques sobre les dades de la taula que crea el graf. Això és un desavantatge bastant gran, ja que la gran majoria de la informació que es pot extreure del graf serà pel que fa a la seva estructura i no a la informació directa. Per tant, cal tenir-ho en compte en les decisions que es prenguin o les conclusions que se n'extreguin.
- Podem dir que no és una eina independent. El problema és que si volem analitzar qualsevol graf s'ha d'importar des de fora del programa. Necessitem una aplicació o algun altre tipus de programa que generi el graf des-

prés, per poder modificar-lo des de l'eina i extreure'n la informació. Tot i que això és un inconvenient, hem d'observar que Gephi llegeix com a input molts tipus de format, també un senzill CSV... per tant, aquest desavantatge té una part de punt a favor.

- No totes les aplicacions que extreuen les dades de les xarxes socials són gratuïtes. Per exemple, si volguéssim extreure el graf de la nostra xarxa social de Twitter necessitariem una aplicació que fos de mineria de dades per demanar la informació de Twitter, i un cop aconseguida importar-la a Gephi per a la seva anàlisi. L'única aplicació que hem trobat gratuïta per utilitzar amb Gephi és una eina que s'utilitza a Facebook anomenada Netvizz. Hem de tenir en compte que aquest aspecte no és una limitació de Gephi sinó general. Facebook està supertancat (de fet sembla que Netvizz mateix probablement no durarà molt) però per a Twitter hi ha moltes eines.
- Té moltes opcions i ens podem perdre en elles i al final no trobar l'opció que es busca per a modificar una cosa específica del graf que s'està estudiant.

En els altres assignatures treballarem amb Gephi, i veurem com instal·lar-lo.

### 3. Projecte de recerca i anàlisi de xarxes

El potencial de les xarxes socials és immens, i des del punt de vista del màrqueting s'aprofita. Per exemple, mitjançant el telèfon intel·ligent consultem la cita amb el metge, una recepta per sopar, els millors restaurants japonesos d'una ciutat, una empresa que ofereix un servei determinat o quina és la millor ruta per arribar al meu destí. Tots aquests instants Google els anomena micromoments.

En termes de màrqueting, això significa que l'itinerari del client (*customer journey*) està fragmentat en milers de micromoments que ocorren en temps real i cada un d'ells és una oportunitat única per a les marques per impactar en els potencials clients. A l'hora d'elaborar una estratègia de continguts, les entitats/marques han d'identificar quines són les diferents necessitats dels usuaris durant aquests micromoments i quins canals són els més adients per comunicar-se amb els usuaris.

La societat ha canviat la manera de consumir, de comunicar-se entre sí i amb les marques/entitats. Això ha fet necessari canviar la forma de comunicar-se amb els usuaris d'una entitat/marca, estar presents en tots els canals en què hi ha els nostres clients i usuaris, estar *always on* i connectats.

Una entitat/marca socialment connectada acompanya els clients en el procés de coneixement del servei/producte, d'acostament a l'entitat/marca. En definitiva, està present en els moments en què sigui útil el servei o producte per a l'usuari.

Acompanyar no vol dir vendre. Significa aconsellar, oferir informació de valor, crear una relació de confiança. L'*engagement* del qual tant es parla.

L'òptim és que aquesta relació amb l'usuari comenci abans que aquest en tingui la necessitat. Que l'usuari s'acosti a l'entitat/marca perquè aquesta li ofereix alguna cosa interessant, alguna cosa que li aporta valor, no perquè aquesta li faci una oferta. És a dir, interessa que a poc a poc es vagi acostant a la marca fins a convertir-se en fan i finalment en recomanador de l'entitat/marca (*advocate*).

Evidentment, això només passa en una situació ideal. La realitat és que el procés no sigui lineal, sinó que es produeixen salts entre les diferents etapes, ja que intervenen factors aliens al servei/marca (recomanacions d'amics, competència, etc.).

Hi ha molts negocis/entitats que, encara que tenen una presència relativament interessant en xarxes socials, no acaben de tenir gaire clares les seves estratègies a curt i a llarg termini, ni com poden extreure valor de les xarxes. Això es deu principalment a que no disposen d'un pla bàsic de *social media* en les seves estratègies comercials o de comunicació.

Recordem que gestionem les dades per obtenir valor per a la nostra organització. Ho hem de fer de manera eficient. Per això, hem de planificar molt bé la nostra tasca, tant des del punt de vista del pla de *social media* per al màrqueting de l'organització, amb la finalitat de crear la xarxa de la qual extreure dades, com des del punt de vista de les preguntes que hem de fer per obtenir respostes de les xarxes.

En aquest últim apartat del mòdul, treballarem com orientar un projecte de recerca per saber quines dades necessitem descarregar a fi d'extreure la informació que doni respostes a les nostres preguntes inicials. Primer veurem com planificar i dissenyar un pla de *social media*, i després com plantejar les preguntes de recerca.

### **3.1. Pla de *social media***

Un pla de *social media* en màrqueting o en xarxes socials és l'organització, planificació i adequació dels diferents mitjans socials que tenim al nostre abast per a establir una via de comunicació en línia amb els nostres usuaris amb la finalitat d'aconseguir els nostres objectius i on el nostre usuari sempre es situa al centre.

Les xarxes socials es poden emprar com un canal dins de les nostres estratègies de comunicació que, ben usat i gestionat per un *community manager*, pot servir per a tenir una connexió directa tant amb els clients actuals com amb els clients potencials.

Abans de dissenyar un pla de *social media* cal disposar de la màxima informació del negoci i/o entitat per tal de plantejar una estratègia adequada. Així doncs, és necessari un pas preliminar de *briefing* i anàlisi amb preguntes essencials que cal respondre. Després, s'ha de planificar i estructurar el pla de *social media*. Per això, serà important conèixer quina és la missió i visió de l'organització, quines són les seves debilitats, fortaleses, amenaces i oportunitats. A continuació veurem com fer-ho.

#### **3.1.1. Identificar la missió: preguntes**

La missió d'una organització fa referència al motiu, propòsit i raó per les quals existeix. En concret explica els següents aspectes:

- Què vol complir a l'entorn on actua.
- Què vol fer, és a dir quins són o haurien de ser els seus productes i serveis.



- Per a qui treballa, és a dir, qui són els clients.
- Quins factors poden influir en la seva activitat (entorn intern i extern, recursos disponibles, capacitats distintives...).
- En quin àrea geogràfica competeix.

Moltes empreses en el document fundacional detallen la missió de l'organització, o en els certificats de gestió de la qualitat. Tanmateix, aquesta missió pot evolucionar en el temps o no estar prou detallat. Per tant, si volem conèixer amb detall la missió, hem d'extreure dades a partir de preguntes, com les següents:

- Qui són els nostres clients?
- Quins són els nostres majors productes o serveis?
- En quines àrees geogràfiques competim?
- Quina és la tecnologia bàsica?
- Quines són les nostres creences bàsiques, valors, aspiracions i prioritats filosòfiques?
- Quines són les nostres majors fortaleces i avantatges competitiu?
- Quines són les nostres responsabilitats públiques i quina imatge volem o hem de projectar?

### 3.1.2. Visió

Un altre element que defineix i influeix les organitzacions és la seva visió. Aquesta es refereix a cap a on vol anar l'organització a llarg termini i en què s'ha de convertir, tenint en compte el seu entorn intern i extern.

El detall de la visió l'assolim fent preguntes com les següents:

- Clarifica la direcció i el propòsit de l'organització?
- La direcció està articulada i s'entén?
- La direcció és ambiciosa?

### 3.1.3. Anàlisi DAFO: preguntes

Un altre tipus d'informació rellevant per conèixer l'empresa és l'anàlisi intern i extern de l'organització. Això es fa amb l'anomenat **DAFO**, que és l'acrònim de **Debilitats, Amenaces, Fortaleces i Oportunitats**. Recollir dades sobre aquests aspectes, ens dona l'oportunitat de conèixer quines són les debilitats i les fortaleces de l'organització, i quines són les amenaces i les oportunitats de l'entorn. Amb aquesta informació podem definir accions per mantenir i potenciar les fortaleces, contrarestar les debilitats, aprofitar les oportunitats de l'entorn, i preparar-nos per afrontar les amenaces de l'entorn.

Les xarxes socials són una bona font d'on extreure les dades i construir aquesta informació. Per tal de conèixer les debilitats, les amenaces, les fortaleces i les oportunitats referents a la nostra organització, hem de preparar un llistat de

preguntes prèvies al disseny d'un pla de *social media* per a màrqueting. Atès que les debilitats i les fortaleses es refereixen a la part interna de l'organització, primer hem de mirar internament l'empresa, i recollir dades sobre aspectes com, per exemple, la seva organització, els recursos, els productes o els serveis que ofereix, entre d'altres. L'observació de l'entorn ens permet veure aspectes molt diferents: comercials, com ara la competència; demogràfics; socials, com ara els gustos de la població o les preferències; polítics, com els canvis legislatius; entre d'altres.

A continuació trobareu un exemple de les preguntes que podem fer servir per extreure informació i efectuar el DAFO de la nostra organització.

#### 1) Mirant cap al producte o servei:

- Què venem/ofereix? Descripció dels productes o serveis que disposem en el nostre negoci/entitat, especificant-hi el principal i els complementaris.
- On el venem/ofereix? Quin és el nostre àmbit geogràfic.
- Té lloc web? Ven/ofereix serveis a internet? Què s'ha fet en l'àmbit digital?
- És un servei estacional? Hi ha èpoques de major demanda? Quines?
- Quina és la nostra situació actual en el mercat?
- Quina és la nostra proposta de valor: en què ens diferenciem de la competència?

#### 2) Mirant cap al client o públic objectiu:

- A qui venem/ofereix serveis? A quin(s) *target(s)* ens estarem dirigint?
- Quines característiques tenen els *targets* als quals ens dirigim?
- Quin és el nostre públic objectiu en xarxes socials?
- Quina necessitat estem cobrint amb el nostre servei o producte? Tenim diferents serveis específics per a cada públic objectiu?
- Com es comporta el nostre client en línia? Són usuaris d'internet? Fan servir les xarxes?
- Quines són les pors o barreres que pot tenir aquest client? Per a això podem mirar fòrums, comentaris de blocs, en webs especialitzades... per descobrir què pensen dels tipus de serveis que estem oferint.

- Quins són els idiomes dels nostres clients?
- D'on procedeixen els nostres clients? Són locals, són nacionals, o són estrangers?

### 3) Mirant cap a l'empresa, negoci o entitat:

- Quina és la mida de l'empresa/entitat?
- Quins recursos té disponibles?
- Què s'ha fet fins ara en línia?
- Tenim presència en els *social media*? Com hi estem presents?
- S'han realitzat accions en els *social media* d'algun tipus?
- Es parla de nosaltres en els mitjans socials: blogs, fòrums, xarxes...?
- Què es diu de nosaltres? És positiu o negatiu?

### 4) Mirant cap a la competència:

- Qui és la competència en línia? S'ha d'estudiar què fa i què ofereix. Tots usem Google o similar per buscar informació sobre un tema que ens interessa. Hem d'investigar quines són les paraules claus que el nostre potencial client utilitzaria per trobar aquesta informació i mirar quines empreses surten a les primeres posicions, ja que aquestes seran les nostres competidores.
- Com ofereixen els seus serveis a la web i quin és el seu argument principal?
- Quins canals socials utilitzen els competidors i com hi actuen? Existeix interacció? Són actius? Fan campanyes?

Un cop tenim la resposta a totes aquestes preguntes podem omplir el requadre DAFO (debilitats, amenaces, fortaleces i oportunitats) (taula 2), Aquest serà de gran utilitat per tenir la informació necessària sobre l'empresa/entitat. A partir d'aquesta informació es podran definir accions.

Taula 2. Marc d'anàlisi DAFO

Factors interns	Factors externs
Debilitats	Amenaces
Fortaleces	Oportunitats

#### 3.1.4. Estructurar el pla de social media

A continuació proposem uns deu passos per tal d'estructurar de manera eficaç el pla de *social media*:

##### 1) Pas 1. Objectius: què volem aconseguir?

Cal afegir als objectius generals de l'empresa/entitat aquells que siguin essencials per encaminar de la manera més òptima les nostres metes específiques en l'àmbit digital.

Alguns exemples podrien ser: augmentar vendes, captar nous usuaris, augmentar la nostra comunitat, aconseguir la viralització del nostre contingut, augmentar la visibilitat de la marca, reduir costos, augmentar marges, entre d'altres.

## 2) Pas 2. Estratègies: Quin camí escollirem?

Quan tenim clar què volem aconseguir, cal estudiar les possibilitats que tenim per aconseguir aquestes metes que ens estem proposant.

Les estratègies es fixen com el camí del qual no hem de desviar-nos i són la manera com aconseguirem el que ens estem proposant.

Cada objectiu tindrà diverses estratègies que es complementaran per assolir-lo.

Per exemple, si volem aconseguir nous clients, les estratègies podrien ser:

- Estratègies de promoció i vendes.
- Estratègies d'atracció.
- Estratègies de captació de *leads*.

## 3) Pas 3. Accions: com ho aconseguirem?

En aquest pas, concretarem cadascuna de les accions que necessitem dur a terme per posar en marxa les estratègies que hem seleccionat.

És en aquest moment quan donem els detalls sobre les accions i/o campanyes concretes i les organitzem temporalment en el nostre calendari.

Seguint amb l'exemple anterior, tenim les següents accions:

### a) Accions a realitzar per a l'estratègia de **promoció i vendes**:

- Ofertes puntuals en períodes clau en la nostra empresa/entitat.
- Concursos amb regals d'entrades, de menús, caps de setmana gratis...
- Campanyes publicitàries segmentades a Facebook/Twitter...

### b) Accions per a les estratègies de **màrqueting d'atracció**:

- Ús de la tècnica de màrqueting de continguts.
- Crear i potenciar la creació de contingut d'un bloc.
- Maximitzar el SEO dels articles.

- Compartir i comentar aquests *posts* en les diferents xarxes socials.

#### c) Accions per a l'estratègia de **captació de leads**:

- Creació de formularis de subscripció adequats a objectius i visibles.
- Incentius per subscripció.
- Creació de *landing pages* concretes.

#### 4) Pas 4. Elecció dels canals: on actuarem?

Ja sabem què volem aconseguir i com ho farem, per tant és hora d'establir quines són aquelles plataformes socials en què invertirem els nostres esforços.

Per fer aquesta elecció cal tenir en compte tres factors que són determinants per no malbaratar el nostre pressupost:

- Qui és el nostre client i en quines xarxes s'està movent.
- Els objectius i accions que hem planificat.
- Les característiques intrínseques del nostre servei o producte.

#### 5) Pas 5. Planificació específica per a cada canal social

Un cop tenim seleccionats els canals socials, cal marcar el pla de treball específic per a cada canal, tenint sempre en compte l'objectiu concret que hem d'assolir en aquesta plataforma: campanyes, freqüència de publicacions, i to. És en aquest punt on hem d'incloure el que ens ajudi en la nostra tasca diària.

#### 6) Pas 6. Establiment i medició dels indicadors clau de compliment (KPI)

Una vegada que està tot en marxa, necessitem conèixer el grau d'abast que tenen les nostres accions per poder saber quina repercussió i quin percentatge d'èxit estan tenint en cadascun dels objectius proposats.

Per aquest motiu cal definir els indicadors clau d'acompliment, també coneguts per les seves sigles en anglès KPI (*Key Performance Indicators*). Aquests són útils per a mesurar les nostres accions.

Recordem que els KPI són indicadors que ens ajuden a mesurar i quantificar el nostre treball, havent prèviament definit uns objectius i unes activitats que ens recolzen la consecució d'aquests objectius.

Els KPI s'utilitzen en el món de l'economia i les finances des de fa molts anys amb un únic objectiu: millorar l'eficiència i la productivitat d'un o un conjunt de serveis, aconseguint d'aquesta manera que el nostre equip de treball funcioni de manera coordinada.

## 7) Pas 7. Elecció d'eines que s'usaran per a la gestió i la medicació

Cal triar i seleccionar les eines que resultin més còmodes i funcionin millor per mantenir ordenades les nostres dades, programar la publicació i extreure dades sobre el funcionament de les nostres xarxes socials.

En el mercat existeixen moltes eines i haurem de conèixer-les una mica més per poder triar aquelles que s'ajustin millor a les nostres necessitats per a la gestió d'informació.

### A. Eines de gestió general

#### a) Klout

Analitza la influència en les xarxes socials d'una persona o marca. És una de les eines més utilitzades en el món. Per saber la influència realitza un mesurament tenint en compte diversos factors com:

- La valoració numèrica de la nostra influència.
- Calcula l'abast efectiu del nostre compte.
- Mesura la influència que tenim sobre comptes que interaccionen amb el teu contingut.
- Comprova el tipus de seguidors que té un usuari i el nivell d'influència que aquest posseeix.
- Mesura la probabilitat que tenim que difonguin el nostre contingut a les xarxes socials.
- Permet gestionar i programar els nostres continguts.

#### b) Buffer

És una eina molt útil per programar el moment de millor impacte per compartir els nostres continguts en xarxes socials (Facebook, Twitter, Google+ o LinkedIn). Una altra de les seves funcionalitats és que suggereix continguts relacionats amb els nostres interessos perquè els puguem publicar.

#### c) Hootsuite

L'eina de gestió de Xarxes Socials més usada al món. Permet administrar de forma senzilla diversos perfils socials (Twitter, Facebook Google+ o LinkedIn, etc.). Entre les seves funcionalitats destaquen:

- Programació dels nostres continguts.
- Agregació il·limitada de *feeds* de notícies.
- Escurçador d'URL propi.
- Integració amb Google Analytics.
- Permet crear informes personalitzats.

#### d) Everypost

És una bona eina per difondre continguts en les plataformes de *social media* com (Facebook, Twitter, Google+, LinkedIn, Tumblr, etc.). Algunes de les seves característiques són:

- Compartir textos, fotos i vídeos.
- Ordenar els teus canals favorits.
- Eскурçar les URL.
- Eскурçar els missatges sense limitar-nos en cent-quaranta caràcters.

#### e) bit.ly

És la millor eina de *social media* per escurçar enllaços. Una de les eines essencials per a un *community manager*. Ens permet analitzar la quantitat de clics i el lloc de procedència de les visites als nostres continguts.

#### f) Feedly

Una molt bona eina per gestionar les alimentacions de continguts (*feeds*) dels nostres blocs preferits. Molt útil per organitzar i compartir els continguts que ens interessin.

### B. Eines específiques de Facebook

#### a) Agorapulse

Es tracta d'una eina d'anàlisi de pàgines de Facebook que ens permet utilitzar estadístiques avançades en temps real per mesurar el rendiment de les nostres pàgines i comparar les nostres estadístiques amb les de la nostra competència. Les seves principals funcionalitats són les següents:

- Aplicacions per a pàgines de Facebook, amb les que podrem crear concursos, sortejos, etc.
- Gestionar el nostre mur de Facebook des de la pròpia eina, fet que permet disposar de tota la informació en un sol lloc.
- Enriquir la nostra base de dades, amb informació extreta dels nostres fans.

#### b) Facebook Insights

Eina d'estadístiques de Facebook, que permet analitzar l'èxit de la *Fan Page*, així com de les campanyes realitzades. Ens permet realitzar les següents accions:

- Planificar segons popularitat i *target*.
- Monitoritzar la competència.
- Crear informes automatitzats.

- Saber quins són els articles més visitats i així utilitzar-los en campanyes.

### c) Easy Promos

Aquesta eina permet crear promocions i concursos a la nostre *Fan Page*. Amb ella també podem gestionar les nostres promocions. Algunes de les seves funcionalitats són:

- Visualitzar i gestionar els usuaris que han participat en la promoció al mur.
- Filtrar els usuaris i les seves participacions per així determinar qui entra en el sorteig.
- Fer un sorteig aleatori entre els participants seleccionats.
- Publicar el resultat del concurs, amb els guanyadors.
- Exportar els participants, els comentaris i m'agrafes en un fitxer Excel o CSV.

### d) LikeAlyzer

Es tracta d'una eina gratuïta molt còmoda i senzilla per analitzar la nostra *Fan Page*. Proporciona un completíssim informe en què destaquen aquells aspectes en els quals hem de millorar. Entre les seves funcionalitats destaquen:

- Analitzar la nostra competència.
- Extreure informació sobre la freqüència de les nostres publicacions.
- Veure els tipus de publicacions.
- Conèixer els horaris en què es rep millor ràtio de resposta dels usuaris.

### e) Fanpage Karma

És una bona eina per analitzar una pàgina d'empresa/entitat a Facebook. Ens permet analitzar i millorar perfils en diverses xarxes socials. Algunes altres de les seves característiques són:

- Saber quins són els nostres continguts més reeixits (publicacions i *hashtags*).
- Cercar els millors temes i continguts per a la nostra pàgina i de les pàgines dels nostres competidors.
- Analitzar els nostres perfils i els dels nostres competidors amb indicadors, avaluacions i informes.
- Analitzar i millorar les nostres publicacions;
- Identificar els millors moments per a publicar.
- Saber quan es produeix la major activitat dels nostres fans.

## C. Eines específiques de Twitter

### a) SocialBro



És una de les eines més potents per obtenir una anàlisi detallada del nostre compte a Twitter. Permet saber l'activitat de la nostra comunitat i dels principals influenciadors. Altres funcionalitats que proporciona són:

- Gestionar les nostres llistes.
- Realitzar un seguiment dels nostres seguidors.
- Gestionar i millorar la nostra comunitat.
- Analitzar la nostra competència. Descobrir com es comporten els nostres competidors, què publiquen i qui els segueix.
- Descobrir nous usuaris per idioma, ubicació geogràfica, antiguitat del compte, ràtio seguidors/seguït, paraules clau.
- Saber quina és la millor hora per piular i, per tant, per millorar la visibilitat de cada piulada.

#### b) Topsy

És una eina de monitorització de tendències a Twitter. Amb ella podrem veure quins són els continguts més esmentats de la nostra competència. També ens permet realitzar les següents accions:

- Trobar les piulades més populars de qualsevol temàtica.
- Identificar els millors i més importants influenciadors a Twitter.
- Cercar i filtrar per contingut, enllaços, piulades, fotos, vídeos o influenciadors.

#### c) TweetDeck

El gran avantatge respecte a les dues anteriors és que és gratuïta. A més inclou:

- Administració d'un nombre il·limitat de comptes.
- Possibilitat de programar les publicacions.
- Integració amb <https://bitly.com/>.

#### d) Twitter Analytics

És l'eina d'estadístiques de Twitter, que permet analitzar l'activitat de les nostres piulades en temps real, així com l'*engagement rate*, és a dir, totes les persones que fan clics als enllaços i la quantitat de repiulades que obtenen. Altres característiques que proporciona són:

- Saber el nombre de favorits i la quantitat de respostes que té cada publicació.
- El volum total d'impressions en els últims vint-i-vuit dies.
- La mitjana d'impressions per dia i una comparativa amb el període anterior.
- Una gràfica amb l'estadística i evolució del nombre de seguidors.
- Saber els interessos dels nostres seguidors.

- La ubicació geogràfica dels nostres seguidors: per països i ciutats.

#### e) ManageFlitter

Ens permet visualitzar a la nostra comunitat de Twitter i filtrar-la per saber els usuaris que ens segueixen, usuaris amb poca activitat o que estiguin inactius. També ens proporciona altres funcionalitats com les següents:

- Executar *unfollows* de forma massiva (fins a un límit de cent per dia en la seva versió gratuïta).
- Saber les estadístiques més rellevants dels nostres seguidors.
- Analitzar comptes falsos i inactius per deixar de seguir-los.
- Cercar influenciadors de la temàtica que ens interessa.
- Programar els nostres continguts perquè es publiquin a Twitter a l'hora de més impacte.

#### f) MetricSpot

És una eina gratuïta que permet realitzar una analítica molt completa en qualsevol compte de Twitter. Entre altres funcionalitats destaquen les següents:

- Anàlisi de la nostra competència, identificant als usuaris més actius per intentar captar-los com a seguidors.
- Saber quines són les piulades més populars.
- Mesurar la influència a Twitter.
- Saber quins són els *hashtags* més utilitzats.
- Analitzar la nostra activitat diària i setmanal.
- Proporció de piulades, RT i respostes publicades.

### 8) Pas 8. Elecció del personal necessari

Quan ja es té tot el pla de *social media* establert, cal escollir el personal adient per dur a terme les accions. Per tant, cal veure a qui necessitem i quines seran les seves funcions.

### 9) Pas 9. Protocol d'actuació per a casos de crisi de reputació

Actuar a les xarxes ofereix molts beneficis, però també és veritat que es poden produir situacions difícils que s'hauran de gestionar adequadament.

Establir dins del pla de *social media* per a màrqueting un protocol d'actuació que sigui conegut per tot el personal implicat en la gestió dels nostres mitjans socials és una arma de prevenció que pot ser molt útil en casos de crisi de reputació.

### 10) Pas 10. Pressupost: quant em gastaré?

Aquest últim pas és tan important com el primer i vindrà determinat tant pels recursos econòmics dels quals disposa l'empresa/entitat, com per la necessitat d'inversió que requereix el nostre pla de *social media* per a màrqueting.

Sense una inversió correcta, és gairebé impossible pretendre obtenir algun guany significatiu.

### 3.2. Planificar la recerca

Tenint en compte tot el que hem dit, resumim a continuació els passos que hem de seguir quan vulguem iniciar una recerca dins de les nostres dades. Recordem que analitzem dades per satisfer les necessitats d'informació detectades.

En primer lloc hem de seleccionar un tema en concret i definir els objectius pel quals iniciem la recerca. Podem expressar un objectiu general i després concretar-lo en d'altres més concrets. El tema escollit, l'objectiu general i els específics han d'estar alineats amb l'estratègia de l'organització. Poden estar vinculats a aspectes identificats en el DAFO o en el pla de *social media*.

En segon lloc, a partir dels objectius hem d'elaborar les preguntes concretes que volem respondre amb les dades. Són les preguntes de recerca, que volem respondre amb l'anàlisi de les xarxes socials. Preguntes que farem són: què, qui, quins, com, on, quants?

En tercer lloc, identificarem les fonts on es generen les dades que es necessiten. Per tant, haurem de decidir, d'acord al pla de *social media*, quines xarxes socials emprarem, quins continguts i quines mètriques, en concret. Hem de tenir en compte els tipus de dades que ofereix cadascuna de les mètriques que hem vist abans.

En quart lloc, extraurem les dades, les tractarem, les organitzarem, emmagatzemarem i classificarem.

Per últim, les analitzarem tenint en compte les preguntes de recerca efectuades.



## Bibliografia

**Chen, M.; Mao, S.; Liu, Y.** (2014). *Big Data: A Survey. Mobile New Application* (vol. 9, pàg. 171-209).

**Combe, D.; Largeron, C.; Egyed-Zsigmond, E.; Géry, M.** (2010). «A comparative study of social network analysis tools» [en línia]. *Web Intelligence & Virtual Enterprises* (oct.). Saint Etienne. <<https://hal.archives-ouvertes.fr/hal-00531447/document>>

**Garcia-Alsina, M.** (2017). *Big data: gestión y explotación de grandes volúmenes de datos*. Barcelona: Editoria UOC-El Profesional de la Información.

**Gomez García, J. L.; Conesa i Caralt, J.** (2015). *Introducción al big data*. Barcelona: Oberta UOC Publishing.

**Kune, R.; Konugurthi, P. K.; Agarwal, A.; Chillarige, R.; Buyya, R.** (2016). «The anatomy of big data computing». *Software Practice and Experience* (núm. 46, pàg. 79-105).

**Liu, C.; Yang, C.; Zhang, X.; Chen, J.** (2015). «External integrity verification for outsourced big data in cloud and IoT: A big picture». *Future Generation Computer Systems* (núm. 49, pàg. 58-67).

**Padgavankar, M. H.; Gupta, S. R.** (2014). «Big Data Storage and Challenges». *International Journal of Computer Science and Information Technologies* (vol. 5, núm. 2, pàg. 2.218-2.223).

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

