ESPIANT

les XARXES SOCIALS amb TEORIA DE GRAFS

Pau Ventura i Alsina
Tutor: Juli Pérez
IES Pius Font i Quer
Manresa, octubre 2012

La gent que no es creu que les matemàtiques són senzilles,
és només que no s'ha adonat de com és de complicada la vida.

J. L. Von Neumann

Índex

Introducció

En aquest treball de recerca tractaré d'espiar xarxes socials, utilitzant una lupa molt potent: **la teoria de grafs.**

La teoria de grafs és una branca de les matemàtiques que s'aplica en camps tant diversos com ara les ciències socials, la lingüística, la física o l'enginyeria de la comunicació. Per exemple, només a nivell d'informàtica, té a veure amb intel·ligència artificial, llenguatges formals, sistemes operatius, escriptura de compiladors, organització i recuperació d'informació, etc. No es pot posar doncs en dubte que és útil desenvolupar i estudiar els grafs en general com a concepte abstracte, com a model matemàtic, i relacionar-lo amb les seves aplicacions.

Per altra banda, les persones ens relacionem entre nosaltres i constituïm grups socials. Avui en dia, en aquest món digitalitzat ningú pot negar que les xarxes socials en línia constitueixen grups socials a gran escala i es troben a tot arreu. A través d'elles circulen moltes dades, tant de contingut com de relacions, que poden ser extretes i analitzades, i justament la teoria de grafs és una bona eina per visualitzar-les.

Motivació de la recerca i objectius

A l'hora de triar el tema del treball de recerca vaig pensar que m'agradaria fer-lo de matemàtiques, però no només a nivell abstracte, sino en relació amb alguna aplicació que tingués a veure amb els meus interessos. Un dia el meu oncle em va ensenyar una aplicació del Facebook que dibuixava un graf com el de la figura 1 sobre les amistats de l'usuari.

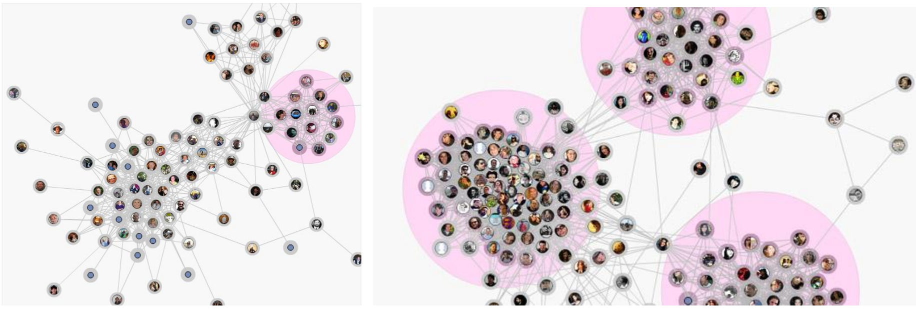


Figura 1: Exemples de grafs socials del facebook.

Em va sorprendre que poguessin saber tantes coses de mi i dels meus amics. Llavors se'm van acudir aquests interrogants:

- Com ho fan els propietaris de les xarxes socials en línia per saber tantes coses de nosaltres?
- Què en fanells d'aquesta informació?
- Qui espia les xarxes socials?
- Qui ens robarà la nostra privacitat?

A partir d'aquests interrogants va néixer la idea de fer un treball relacionant els grafs amb les xarxes socials.

El fet d'haver triat aquest tema també està motivat pel meu interès per les matemàtiques, perquè està relacionat amb els interessos del jovent, i perquè és una forma d'explicitar com les matemàtiques ens ajuden a saber més coses de la gent amb qui et relaciones. També m'ha servit per aprendre coses noves. Però calia concretar i fixar una línia, uns objectius, unes hipòtesis com a projecte del treball, que tingués coherència i em permetés abastar-ho a partir del nivell de 1r de bat. A grans trets el treball consisteix, per una banda, en entendre què són els grafs i les seves propietats, quina informació poden tenir, com extreure-la, etc. i, per altra, en explorar les xarxes socials en línia, veure com funcionen, fer-ne un anàlisi a la meva mida utilitzant els grafs i poder-ne extreure les meves conclusions.

Concretament els objectius d'aquest treball han estat:

- Demostrar que les matemàtiques són útils i aplicables.
- Mostrar com les matemàtiques es poden aplicar a les xarxes socials.
- Conscienciar la gent de la gran informació que es pot extreure de les xarxes socials utilitzant, per exemple, els grafs.
- Convèncer (amb exemples) de la poca privadesa de les xarxes socials.

Desenvolupament del treball

En primer lloc, em vaig documentar buscant informació per entendre la teoria de grafs. Per tal d'anar fixant idees feia un resum sobre els conceptes bàsics i em dibuixava grafs com a exemples. Em vaig anar acostumant als grafs i vaig descobrir que en tenia molts al meu voltant, fins i tot a l'institut. Això ha donat lloc al capítol 1.

Ara bé, el que és comú a tots els espais de l'institut són les persones i les relacions que hi ha entre elles, és a dir xarxes socials. Això incrementava el meu interès en relacionar els conceptes de grafs amb les xarxes socials, i em va sorprendre que, tot i que últimament les xarxes socials estan molt de moda, no trobava llibres ni pàgines web que en parlessin. Va ser en buscar-ho en anglès, que vaig començar a trobar informació sobre Social Network Analisis (SNA), però les referències corresponien a articles d'investigació. Això em va portar fins al Jordi Herrera i la Cristina Pérez de la Universitat Autònoma de Barcelona i em va obrir les portes del programa Argó. També vaig inspirar-me en la pel·lícula "Redes socialesï en una conferència organitzada per la Societat Catalana de Matemàtiques. En el capítol 2 explico com els grafs són un bon model per les xarxes socials.

El capítol 3 respon a la necessitat de posar en pràctica tots els coneixements que havia après sobre grafs i xarxes socials. Volia experimentar en una xarxa social a mida en la que, en primera persona, pogués realitzar el treball de camp, treure'n conclusions i comparar amb la realitat.

La inspiració inicial venia de la visió d'un graf del Facebook, per tant volia tornar a les xarxes socials en línia actuals, per espiar-les. La necessitat de treballar en xarxes obertes per fer l'extracció de dades em va portar a analitzar el Twitter i el Flickr al capítol 4, a partir de dades proporcionades per dos investigadors de la UAB.

Finalment, apart de les fonts d'informació consultades, he volgut incloure un apartat sobre eines TIC, per remarcar la descoberta del TeX, per tal de fer una bona presentació del treball escrit, i del Visone, per poder dibuixar i manipular els grafs. Ambdós són programes no trivials que he hagut d'aprendre, però que m'han estat molt útils.

Agraïments

No puc finalitzar aquesta introducció sense agrair el suport que he rebut de la gent del meu voltant. En primer lloc vull donar les gràcies al meu tutor Juli Pérez, professor de matemàtiques, que ha fet possible la realització d'aquest treball. En especial, també vull agrair a en Jordi Herrera i la Cristina Pérez, del Departament d'Enginyeria de la Informació i les Comunicacions de la Universitat Autònoma de Barcelona, per fer-me conèixer l'existència del programa Visone i proporcionar-me dades sobre xarxes socials en línia. A través d'ells aquest treball ha entrat en el programa Argó de la UAB.

Capítol 1

On s'amaguen els grafs de l'institut?

Introducció a la teoria de grafs

No hi ha cap branca de les matemàtiques, per abstracta que sigui,
que un dia no pugui ser aplicada al món real.
N. I. Lobatxevski

En aquest capítol explicaré els conceptes bàsics de la teoria de grafs, acompanyat d'exemples gràfics propis. Tot i ser conceptes teòrics estan molt relacionats amb la vida quotidiana. És per això, que he inclòs una secció per descobrir on s'amaguen els grafs en un entorn molt proper, l'institut de secundària. D'aquesta manera es veu com les matemàtiques no estan tant allunyades de la realitat i són més comprensibles.

A la primera secció comento l'origen dels grafs posant de manifest la seva implicació en àmbits molt diferents des de la geografia a l'electrònica, passant per exemple per la química. A la segona secció es revisen els conceptes bàsics de manera breu, amb exemples gràfics. Les seccions 3 i 4 aprofundeixen en alguns aspectes de la teoria matemàtica de grafs. Finalment, en l'última secció busco alguns dels grafs amagats en un institut de secundària, com si estiguéssim jugant a fet i amagar amb els grafs.

1.1 Origen dels grafs

La teoria de grafs té el seu origen en el famós problema dels set ponts de Königsberg, que va ser resolt per Leonhard Euler, l'any 1736, i va ser considerat el primer resultat de la teoria de grafs. Aquest problema consisteix en trobar un camí que passi per tots els set ponts que travessen el riu Pregel a la ciutat de Königsberg (actualment anomenada Kaliningrad) sense repetir-ne cap.

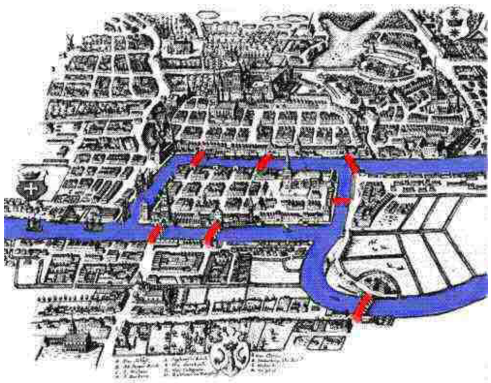


Figura 1.1: Recreació dels set ponts de Köningsberg.

Una altra fita històrica va ser l'any 1852, quan Francis Guthrie va plantejar el problema dels quatre colors: és possible pintar qualsevol mapa de països sense que dos països veïns tinguin el mateix color, utilitzant només quatre colors? No va ser resolt fins l'any 1976 per Kenneth Appel y Wolfgang Haken. Per resoldre aquest problema, els matemàtics van haver de definir termes i conceptes que són fonamentals en la teoria de grafs actual.

Posteriorment, en un àmbit ben diferent com és la química, també va ser utilitzada la teoria de grafs per Arthur Cayley, l'any 1857, per tal de distingir i enumerar els isòmers: compostos químics amb la mateixa fórmula (o composició) però diferent estructura molecular. Pocs anys més tard, el 1874, la teoria de grafs va ser utilitzada en l'àmbit de l'electricitat per Gustav Kirchhoff, per tal de formular les conegudes lleis que porten el seu nom. En els seus treballs sobre xarxes elèctriques, va ser el primer d'utilitzar els grafs de tipus arbre.

Durant el segle XX, la teoria de grafs va anar agafant consistència. El primer llibre sobre teoria de grafs va ser escrit per Dénes Konig i publicat l'any 1936. El progrés de la informàtica i de les tècniques de computació li va donar un bon impuls i l'ha convertit en una de les branques de la matemàtica aplicada amb més vitalitat.

Actualment la teoria de grafs és una branca de la matemàtica discreta molt important i activa a nivell d'investigació. Ha arribat a un desenvolupament tècnic força elevat amb parts reservades estrictament a especialistes i clarament fora de l'abast d'aquest treball. Està relacionada amb moltes altres branques de la matemàtica moderna com ara la teoria de grups, la topologia, la geometria algebraica, etc.

Al llarg de la història, la teoria de grafs ha servit per resoldre problemes de tipus molt diversos, o també per poder-los representar gràficament per tal d'interpretar-los més ràpidament i clarament. Tal com es pot suposar a partir del seu origen, té múltiples aplicacions en diferents àmbits com ciències de la computació, telecomunicacions, física, química i diferents enginyeries, com ara circuits elèctrics i xarxes d'ordinadors. Ara bé, també és utilitzada en la biologia, per la classificació i l'evolució de les espècies, i fins i tot per representar gràficament les migracions d'animals, de manera que es vegi clarament quins animals emigren a cada lloc, quan ho fan, etc. Per altra banda, també destaca el seu ús en l'àmbit empresarial i econòmic amb algoritmes que ajuden a reduir costos i planificar tasques.

Avui en dia utilitzem les seves aplicacions a la vida quotidiana sense ser-ne conscients. Per exemple, quan cerquem una ruta, sigui via googlemaps o utilitzant GPS, també intervenen els grafs. Els mapes es transformen en grafs, de manera que cada carrer és una aresta i cada cruïlla un vèrtex, i s'utilitza la teoria de grafs i alguns algoritmes per tal de poder calcular quin recorregut és el millor per arribar a la destinació donada (tenint en compte els paràmetres triats, ja sigui distància, peatges, qualitat de la carretera...).

En aquest treball relacionaré la teoria de grafs amb les xarxes socials com es veurà en el capítol següent.

1.2 Conceptes bàsics amb exemples

En aquesta secció s'inclouen les definicions bàsiques de grafs i alguns resultats ben coneguts. Al llarg de l'explicació presento exemples per tal que sigui més entenedor. De fet en el procés que he fet, la construcció dels exemples m'ha ajudat a entendre-ho millor.

Un **graf** G és un conjunt de **vèrtexs** V, un conjunt d'**arestes** E (de l'anglès *edge*), i dues **funcions d'incidència** $\iota, \tau : E \to V$ que indiquen quin és el vèrtex inicial i final de cada aresta. Els vèrtexs també s'anomenen **nodes** i sempre suposem que com a mínim n'hi ha un.

Un graf es dibuixa pintant un punt per cada vèrtex $v \in V$, i una fletxa per cada aresta $e \in E$ que comença a $\iota(e)$ i que acaba a $\tau(e)$. Si dues arestes diferents, $e_1, e_2 \in E$ comencen al mateix vèrtex i acaben al mateix vèrtex s'anomenen **arestes paral·leles**. Si una aresta comença i acaba en el mateix vèrtex, s'anomena **llaç**. Si un graf té arestes paral·leles o llaços, s'anomena **multigraf**.

En un graf, si una aresta e comença en el vèrtex u, i acaba en el vèrtex v, l'anomenem també (u, v) (es pot fer d'aquesta manera ja que no és un multigraf i, per tant, no hi hauran arestes diferents amb el mateix nom). En aquest cas, es diu que u i v són vèrtexs **adjacents**, i que e és **incident** als vèrtexs u i v.

Per dibuixar els grafs, utilitzo el programa Visone, que no permet etiquetar ni vèrtexs ni arestes amb subíndexs. Per tant, etiquetaré els vèrtexs directament amb números de manera que $V \subset \{1, 2, 3, \ldots\}$ i les arestes amb lletres, per tant $E \subset \{a, b, c, \ldots\}$. A la figura 1.2 es mostra un exemple d'un graf senzill, amb conjunt de vèrtexs $V = \{1, 2, 3, 4\}$, conjunt d'arestes $E = \{a, b, c, d, e\}$, i les funcions d'incidència ι i τ donades a continuació.

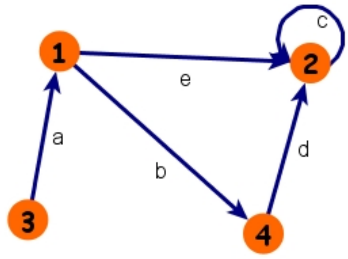


Figura 1.2: Exemple de graf indicant les seves funcions d'incidència.

$$\begin{array}{rcl} \iota\colon E & \longrightarrow & V \\ a & \mapsto & 3 \\ b & \mapsto & 1 \\ c & \mapsto & 2 \\ d & \mapsto & 4 \\ e & \mapsto & 1 \end{array} \qquad \begin{array}{rcl} \tau\colon E & \longrightarrow & V \\ a & \mapsto & 1 \\ b & \mapsto & 4 \\ c & \mapsto & 2 \\ d & \mapsto & 2 \\ e & \mapsto & 2 \end{array}$$

A la definició de graf, les arestes tenen una direcció (un inici i un final); en alguns llibres això ho anomenen **digraf** o **graf dirigit**. A la figura 1.3 es mostra un graf, un multigraf, i un graf dirigit.

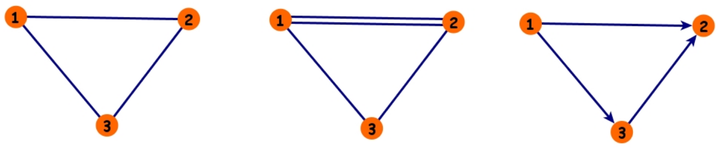


Figura 1.3: Exemple d'un graf, un multigraf i un graf dirigit, respectivament.

Si no diem el contrari, quan parlem de graf estarem excloent els multigrafs i els grafs dirigits.

L'**ordre** d'un graf $G = (V, E)$ és el nombre de vèrtexs que hi ha a G, és a dir, el cardinal del conjunt V, $|V|$. La **mida** de G és el nombre d'arestes que hi ha a G, és a dir el cardinal de E, $|E|$.

Per exemple, si fixem que l'ordre és 1, és a dir només tenim un vèrtex, no podem tenir cap aresta, a menys que permetem els llaços, com es mostra a la figura 1.4.

Figura 1.4: Exemples de multigrafs d'ordre 1.

D'ordre 2, és a dir amb 2 vèrtexs, només hi ha 2 grafs possibles, tal i com es veu a la figura 1.5. Notem que el primer té 2 vèrtexs desconnectats i és de mida 0, mentre que en el segon els vèrtexs estan connectats i el graf és de mida 1.

Figura 1.5: Tots els grafs d'ordre 2.

Per simplificar els gràfics, en els següents exemples no es té en compte la numeració dels vèrtexs.

D'ordre 3, els grafs possibles són els que es mostren a la figura 1.6 i tenen mides 0, 1, 2, 3 respectivament.

Figura 1.6: Tots els grafs d'ordre 3.

A continuació a les figures ?? i ??es mostren tots els graf possibles d'ordre 4. La figura 1.7 conté els de mida 0, 1 i 2, que tenen parts que no queden connectades. D'aquesta característica se'n diu ser connex o no connex i en parlaré a la següent secció. De fet és evident que, amb només dues arestes, no hi ha cap manera que tots els vèrtexs estiguin connectats, ja que per tal d'ajuntar 4 vèrtexs es necessiten com a mínim 3 arestes.

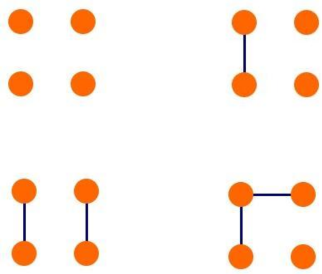


Figura 1.7: Tots els graf d'ordre 4 i mides 0, 1 i 2.

A la figura 1.8 es mostren els de mida 3, 4, 5 i 6. En aquest cas només n'hi ha un de mida 3 que té un vèrtex aïllat, la resta estan tots connectats.

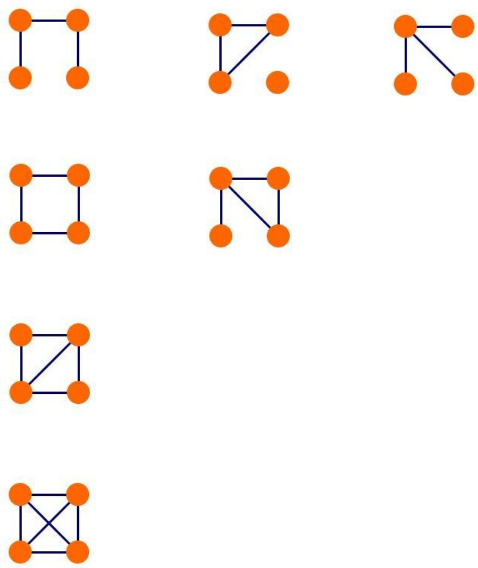
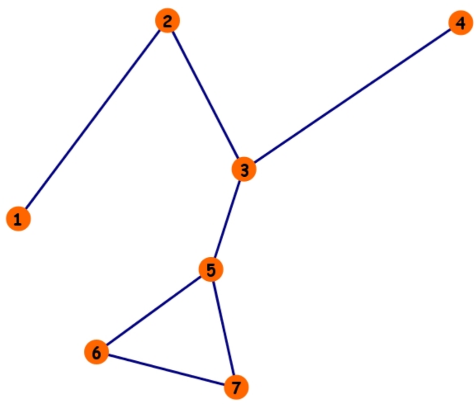


Figura 1.8: Tots els graf d'ordre 4 i mides 3, 4, 5 i 6.

El **grau** o **valència** d'un vèrtex és el número d'arestes que incideixen en ell. El denotarem $\deg v$. Si el graf és dirigit, enlloc de parlar del grau del vèrtex, parlarem del **grau d'entrada** i **grau de sortida**. El **grau mitjà** d'un graf no és res més que la mitjana aritmètica dels graus de cada vèrtex.

Apliquem-ho a un exemple. A partir del graf de la figura següent, fem la taula amb el grau de cada vèrtex.



Vèrtex	grau
v_1	1
v_2	2
v_3	3
v_4	1
v_5	3
v_6	2
v_7	2

Figura 1.9: Exemple de càlcul de grau dels vèrtexs.

Calculem ara el grau mitjà del graf de la figura 1.9:

$$\frac{1+2+3+1+3+2+2}{7} = \frac{14}{7} = 2.$$

Vèrtex	grau
v_1	6
v_2	2
v_3	3
v_4	0

Figura 1.10: Taula de vèrtexs i graus.

Problema.
És clar que donat un graf li podem calcular l'ordre, la mida, els graus, etc. Però també pot tenir sentit fixar uns valors i després buscar un graf que els satisfaci.

- *Qualsevol mida i ordre són possibles?*
- *Donada una taula qualsevol de vèrtexs i graus, sempre es podrà dibuixar un graf que la representi?*

Òbviament hi ha uns requisits mínims que s'han de complir, com els que es dedueixen a continuació, per tant la resposta és negativa per les dues preguntes.

Propietats. Suposem que tenim un graf $G = (V, E)$ d'ordre N i mida M. A partir dels conceptes introduïts podem deduir les propietats següents.

i) Cada aresta possible correspon a una parella de vèrtexs, per tant el nombre d'arestes és com a màxim el nombre de parelles de vèrtexs, que es compten com les convinacions de N elements agafats de dos en dos.

$$M \leq \binom{N}{2} = \frac{N!}{(N-2)!\,2!} = \frac{N(N-1)}{2}.$$

ii) El grau de qualsevol vèrtex ha de ser menor o igual que $N - 1$. Aquesta propietat és trivial ja que no es permeten arestes paral·leles ni llaços.

iii) Cada aresta fa augmentar el grau de dos vèrtexs en una unitat, per tant la suma dels graus de tots els vèrtexs és el doble del nombre d'arestes, és a dir,

$$\sum_{v \in V} \deg v = 2M.$$

De fet aquesta propietat sovint s'utilitza en grafs complicats per fer el recompte d'arestes, ja que és més fàcil fer automàtic el recompte de graus dels vèrtexs.

iv) Hi ha d'haver un nombre parell de vèrtexs amb grau senar. Això es dedueix de l'anterior perquè la suma dels graus és parell. Això ens permet afirmar, per exemple, que no existeix cap graf de 2013 vèrtexs tal que tots els vèrtexs tinguin grau 17.

Així és evident que no totes les taules de graus possibles corresponen a un graf, ja que només que no compleixin un dels requisits anteriors no tenen sentit. La taula de la figura 1.10 no correspon a cap graf, de les impossibles de dibuixar. Per exemple no pot ser que el vèrtex v_1 tingui grau superior a $4-1=3$. A més, hi ha un nombre senar de vèrtexs de grau imparell, la qual cosa fa que la suma de tots els graus doni senar, cosa que contradiu la propietat (iii).

Un graf s'anomena **complet** si cada vèrtex és adjacent a tots els altres vèrtexs. Aleshores, tal i com indica la propietat (i) es compleix que la mida del graf és $M = \dfrac{N(N-1)}{2}$.

Un graf s'anomena k-**regular** si tots els vèrtexs tenen el mateix grau (k).

És clar que un graf k-regular ha de tenir com a mínim $k+1$ vèrtexs. A partir de la propietat (iii) anterior també podem concretar la relació que hi haurà entre el nombre d'arestes i de vèrtexs. Primer calculem la suma dels graus,

$$\sum_{v \in V} \deg v = \sum_{v \in V} k = |V| \cdot k, \text{ apliquem ara la propietat (iii) } |V| \cdot k = 2|E|.$$

La figura 1.11 mostra exemples de grafs regulars amb $k = 1, 2, 3$. En efecte, tots compleixen l'afirmació anterior.

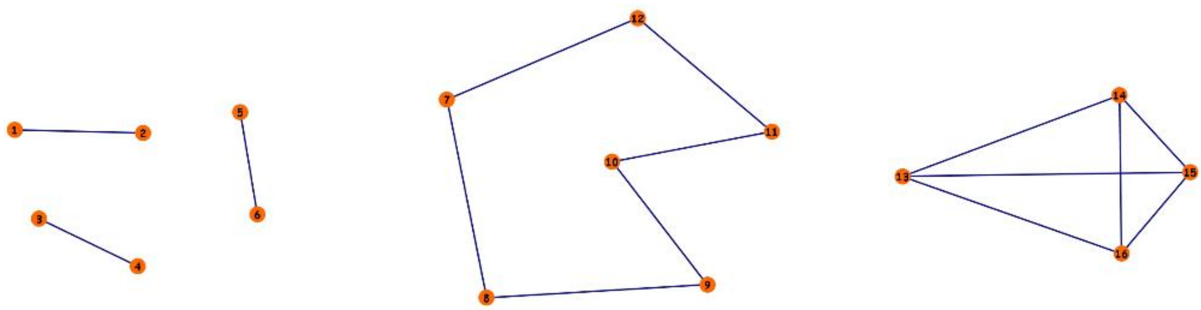


Figura 1.11: Exemples de grafs 1-regular, 2-regular i 3-regular, respectivament.

Diem que un graf és **bipartit** si el conjunt de vèrtexs es pot partir en dos grups disjunts de manera que cada aresta va d'un vèrtex d'un grup a un vèrtex de l'altre.

La figura 1.12 mostra un exemple de graf bipartit d'ordre 7 i mida 6. Els dos grups de vèrtexs són $\{1, 2, 3\}$ i $\{4, 5, 6, 7\}$

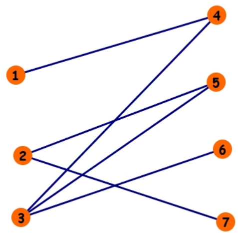


Figura 1.12: Exemple de graf bipartit.

No podem passar per alt la utilització de les matrius com a eines de la teoria de grafs. La matriu d'adjacència d'un graf es construeix posant a la casella (i, j) un 1 si els vèrtexs v_i, v_j són adjacents, i un 0 si no ho són. També és pot construir la matriu d'incidència posant un 1 a la casella (i, j) si l'aresta e_i és incident al vèrtex v_j.

Per exemple, la matriu d'adjacència del graf 3-regular de la figura 1.11 es mostra a la figura 1.13 i la seva matriu d'incidència a la figura 1.14

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Figura 1.13: Matriu d'adjacència del graf de la figura 1.11.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Figura 1.14: Matriu d'incidència del graf de la figura 1.11.

A partir de les propietats i operacions d'aquestes matrius, es dedueixen altres propietats del graf. Però en aquest treball no tractarem els detalls.

1.3 Camins, connexió i distància

Un **camí** p (en anglès *path*) és una seqüència finita de vèrtexs connectats consecutivament per arestes. La **longitud** $l(p)$ d'un camí $p = v_0, v_1, \ldots, v_n$, tal que $\tau v_i = \iota v_{i+1}$ és el nombre d'arestes n que travessa.

Sempre hi ha una aresta menys que el nombre de vèrtexs visitats, per aquest motiu comencem a numerar els vèrtexs des de 0.

Anomenem **vèrtex inicial** de p al primer vèrtex $\iota p = v_0$, i **vèrtex terminal** l'últim $\tau p = v_n$. Tots els altres vèrtexs s'anomenen **vèrtexs interns**. En el cas que els dos vèrtexs terminals coincideixin, és a dir, $\iota p = \tau p$, el camí comença i acaba en el mateix lloc i s'anomena **camí tancat** o **cicle**. L'**invers** d'un camí p és el mateix camí llegit en ordre contrari $p^{-1} = v_n, \ldots, v_0$; està clar que la longitud serà la mateixa, $l(p^{-1}) = l(p)$.

Observem que un camí pot passar més d'una vegada per un mateix vèrtex. Si això no succeeix direm que es tracta d'un camí **simple**. El mateix pels cicles, si no hi ha vèrtexs repetits (sense tenir en compte la repetició del primer amb l'últim, ja que sempre hi és) s'anomena cicle **simple**. A la figura 1.15 és mostra un graf amb un cicle de longitud 5 i un camí $p = 1, 5, 6, 7$ de longitud 3.

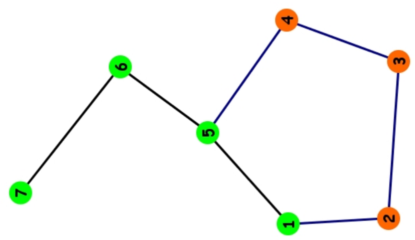
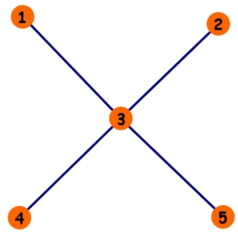


Figura 1.15: Graf amb un cicle.

La **distància** entre dos vèrtexs, u, v, d'un graf és la longitud mínima de tots els camins que els uneixen i la denotarem $d(u, v)$; Sí no hi ha cap camí que uneixi u i v direm que $d(u, v) = \infty$. Els camins de u a v que tenen aquesta distància mínima (n'hi pot haver més d'un) s'anomenen **geodèsiques**. Està clar que les geodèsiques són camins simples. També està clar que l'invers d'una geodèsica també és una geodèsica i per tant $d(u, v) = d(v, u)$.

A continuació, a la figura 1.16, veiem un exemple de graf amb totes les distàncies calculades.



$$\begin{array}{lllll} d(v_1,v_2)=2, & d(v_1,v_3)=1, & d(v_1,v_4)=2, & d(v_1,v_5)=2, & d(v_2,v_3)=1, \\ d(v_2,v_4)=2, & d(v_2,v_5)=2, & d(v_3,v_4)=1, & d(v_3,v_5)=1, & d(v_4,v_5)=2. \end{array}$$

Figura 1.16: Graf amb totes les distàncies calculades.

Un graf és **connex** si per cada dos vèrtexs qualsevols u, v existeix almenys un camí que els uneixi, és a dir que $d(u,v) \neq \infty$. En el cas que no sigui connex, el graf estarà format per dues o més **components connexes**. La figura 1.17 mostra un graf no connex, amb 2 components connexes, i la seva transformació en graf connex afegint 1 aresta.

Figura 1.17: Graf no connex, i el mateix graf afegint una aresta que el fa connex.

L'**excentricitat** d'un vèrtex u es denota $\text{exc}(u)$ i és la distància de u al vèrtex que estigui més lluny d'ell, és a dir,

$$\text{exc}(u) = \max_{v \in V} \{d(u,v)\}.$$

Per tant, des del vèrtex u amb camins de longitud com a màxim $\text{exc}(u)$ es pot arribar a qualsevol altre vèrtex del graf, mentre que amb camins de longitud inferior segur que no arribes a algun vèrtex.

El **radi** d'un graf G és l'excentricitat més petita dels seus vèrtexs, i el **diàmetre** l'excentricitat més gran. Matemàticament:

$$r(G) = \min_{v \in V} \text{exc}(v), \qquad \text{diam}(G) = \max_{v \in V} \text{exc}(v) = \max_{u,v \in V} d(u,v).$$

També s'anomena **centre** del graf els vèrtexs que tenen excentricitat igual al radi.

Així, el diàmetre és la longitud de la geodèsica més llarga possible. A partir dels vèrtexs del centre, el radi és la longitud de la geodèsica més gran per arribar als altres vèrtexs. Aquests conceptes tenen relació amb un cercle però observem que el diàmetre no sempre és el doble del radi, com es veu a l'exemple següent. La relació que compleixen és:

$$r(G) \leq \mathrm{diam}(G) \leq 2r(G).$$

Aquestes mesures només tenen gràcia si el graf és connex, ja que en el cas que no ho sigui les distàncies entre una component connexa i una altra seran infinit, la qual cosa fa que l'excentricitat de tots els vèrtexs sigui també infinita. Per tant, el radi i el diàmetre també seran infinit.

A la figura 1.18 hi ha un exemple d'un graf en el qual hi ha calculades les excentricitats dels vèrtexs (números vermells). Es dedueix que el seu radi és $r = 2$ i el diàmetre diam $= 3$. El centre del graf serien els vectors 2 i 3.

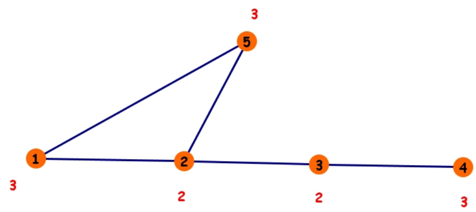


Figura 1.18: Graf amb les excentricitats, el radi i el diàmetre calculats.

Un camí (simple) que passi per totes les arestes una sola vegada es diu camí d'**Euler**. Tornant al famós problema dels set ponts de Köningsberg, té sentit estudiar l'existència de camins que recorrin totes les arestes.

Problema.
Donat un graf preguntar-se si existeix un camí d'Euler és el mateix que demanar si es pot dibuixar el graf sense aixecar el llapis del paper ni repetir arestes. Ho apliquem als grafs de la figura 1.19.

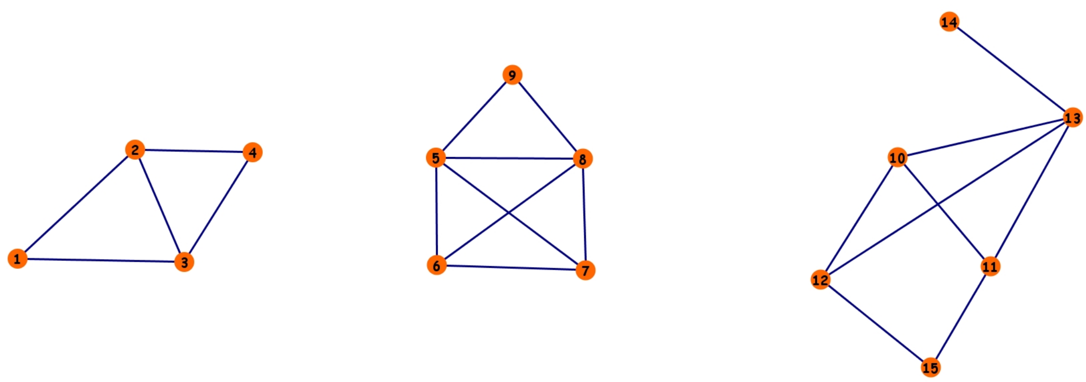


Figura 1.19: Es poden dibuixar sense aixecar el llapis del paper ni repetir línies?

En el primer dibuix, ja es pot veure a simple vista que la resposta és afirmativa. Per exemple el camí $p_1 = 2, 4, 3, 2, 1, 3$. També hi ha altres camins possibles començant amb el mateix vèrtex, $p_2 = 2, 1, 3, 2, 4, 3$, o en un altre vèrtexs, $p_3 = 3, 1, 2, 4, 3, 2$. No se'n troba cap que no comenci ni al vèrtex 1 ni al vèrtex 4. En el segon exemple, ja és més complicat trobar un camí d'Euler. Però després d'uns quants intents s'acaba trobant una solució: el camí $p = 6, 5, 8, 7, 5, 9, 8, 6, 7$. En el tercer cas, es prova però no se'n troba cap.

És perquè encara queden camins per provar o bé perquè no n'hi ha cap?

Podem respondre aquest problema utilitzant el famós teorema d'Euler.

Teorema.
Un graf G conté un camí Eulerià si i només si és connex (llevat de vèrtexs aïllats) i el nombre de vèrtexs de grau imparell és 0 o 2.

Observem que cada vegada que el camí passa per un vèrtex necessitem una aresta per entrar i una per sortir, per tant el seu grau ha de ser parell. Ara bé, el vèrtex inicial i el vèrtex final de qualsevol camí Eulerià no tancat han de ser de grau imparell. Per tant aquests dos són els vèrtexs de grau imparell que ens diu el teorema. En canvi en el cas d'un camí tancat, el vèrtex inicial i final coincideixen i per tant el seu grau és parell, això correspon al cas de 0 vèrtexs de grau imparell.

Ara ja podem afirmar que en el tercer exemple no hi ha cap camí d'Euler, ja que conté 4 vèrtexs de grau imparell. De fet, aquest teorema dóna resposta al problema dels set ponts de Königsberg. A la figura 1.20 es pot veure la seva representació en graf, prenent els ponts com a arestes i les zones de terra ferma com a vèrtexs. Es comprova que els quatre vèrtexs són de grau imparell i per tant, aplicant el teorema d'Euler, no és possible fer un recorregut que passi una única vegada per cadascun dels 7 ponts.

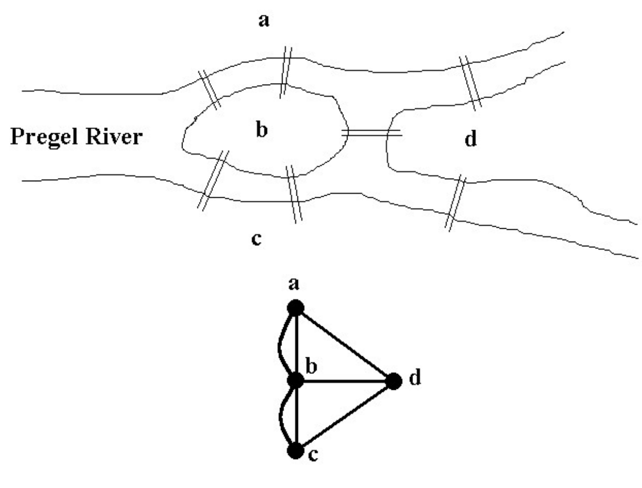


Figura 1.20: Els ponts de Königsberg i la seva representació en graf.

Un conjunt de vèrtexs d'un graf es diu **dominant** si tots els vèrtexs del graf són adjacents a algun vèrtex del conjunt. El **número dominant** d'un graf és el menor cardinal dels seus conjunts dominants, és a dir, dels conjunts dominants que tinguin el nombre mínim de vèrtexs.

1.4 Arbres

Els arbres són un tipus de grafs amb propietats especials. S'utilitzen per classificar i estructurar dades i s'apliquen a molts àmbits des de la teoria de la codificació, fins a ciències econòmiques o polítiques.

Un **arbre** T és un graf connex que no té cicles. Com que no hi ha cicles, per cada parell de vèrtexs u, v, hi ha una única geodèsica que els uneix. Això és perquè si tinguéssim dues geodèsiques, anant per una i tornant per l'altre tindríem un cicle. La geodèsica de u a v la denotem per $T[u, v]$. A la figura 1.21 veiem un exemple d'arbre i calculem la geodèsica del vèrtex 1 al 7: $T[1, 7] = 1, 2, 7$

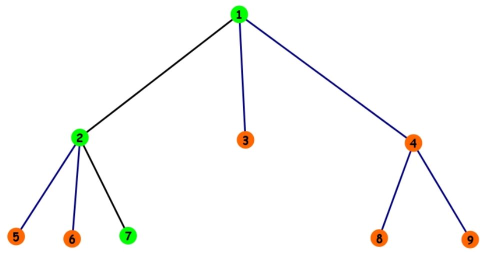


Figura 1.21: Exemple d'arbre amb una geodèsica.

Tots els arbres $T = (V, E)$ tenen sempre un vèrtex més que el nombre d'arestes, és a dir $|V| = |E| + 1$. Per justificar-ho comencem amb l'arbre més senzill: un sol vèrtex i cap aresta. A partir d'aquest arbre, cada vèrtex que s'afegeix estarà unit per una i només una aresta (ja que si estigués unit per més d'una formaria un cicle). Al revés no és cert, és a dir, hi ha grafs $G = (V, E)$ que compleixen $|V| = |E| + 1$ però no són arbres. La figura 1.22 en mostra un exemple.

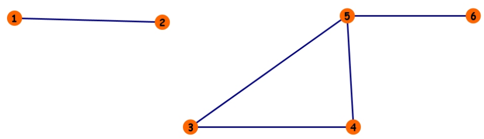


Figura 1.22: Graf amb 6 vèrtexs i 5 arestes, que no és arbre.

De fet, si un graf compleix la fórmula i és connex, llavors sí que és arbre; si compleix la fórmula i no té cicles, llavors també ha de ser arbre.

Hi ha un tipus d'arbres que serveix com a model de moltes classificacions, els **rooted trees**. La figura 1.21 n'és un exemple. En aquest cas es compleix que hi ha un vèrtex, l'1, de grau 3 que s'anomena arrel; els seus fills són $\{2, 3, 4\}$; tenen graus 4, 1 i 3, i a la vegada tenen 3, 0 i 2 fills, respectivament. Aquests últims són de grau 1, i s'anomenen vèrtexs terminals o també **fulles**, per similitud amb els arbres reals.

Sigui $G = (V, E)$ un graf. Un **arbre generador** de G és un subgraf de G que sigui arbre i que inclogui tots els vèrtexs de G.

Es pot veure que un graf connex té un o més arbres generadors. Una manera de construir-ne un pot ser la següent: es tria una aresta qualsevol i es mira si al suprimir-la queda un subgraf connex; en cas afirmatiu la suprimim. Repetim aquest procés fins que no en puguem suprimir cap. El resultat ha de ser un arbre ja és connex i no té cicles, ja que si en tingués podríem suprimir qualsevol aresta del cicle. És generador de G ja que conté tots els seus vèrtexs (perquè no n'hem tret cap).

Considerem per exemple el graf G de la figura 1.23. Observem que:

- L'ordre de G és 5 i la mida de G és 7, per tant està clar que no és un arbre.
- Per a trobar un arbre generador anem comprovant quines arestes podem suprimir. De fet com que el resultat ha de ser un arbre amb tots els vèrtexs, sabem que ha de tenir $5 - 1 = 4$ arestes, per tant n'hem de suprimir només 3.
- Si comencem a suprimir arestes per ordre, podem suprimir a. No podem suprimir b perquè deixaríem el vèrtex 1 aïllat. Suprimim c. No podem suprimir d perquè deixaríem el vèrtex 2 aïllat. Suprimim e, i parem el procés ja que portem eliminades 3 arestes. El resultat és l'arbre generador T que es veu a la figura 1.23.
- De fet també podríem haver suprimit b, d, g i obtindríem un altre arbre generador.

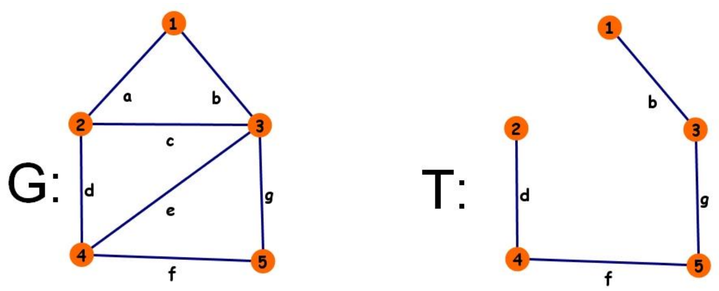


Figura 1.23: Graf G amb un possible arbre generador T.

Quan les arestes porten associat un pes o longitud determinada, es pot buscar un arbre generador tal que les seves arestes sumin el mínim possible. En el fons, el procés és similar a l'exemple anterior, però caldria treure les arestes de pes superior, si és possible. Aquest és un problema clàssic d'Investigació operativa conegut com *el problema del connector* que es resol amb els algoritmes de Prim o de Kruskal.

1.5 Grafs a l'institut

Les seccions anteriors poden semblar molt abstractes, però de fet hi ha molts grafs al nostre voltant. Per tal de posar uns quants exemples d'aquests, he buscat tot d'exemples senzills aplicats a l'institut. Així és com jugar a fet i amagar descobrint grafs en diferents racons de l'institut.

1.5.1 Grafs al laboratori de física i química

Mirant l'estructura de les molècules ens adonem que es poden interpretar com a grafs, on els àtoms són vèrtexs i els enllaços arestes. En la química inorgànica una fórmula representa generalment un sol compost. En els compostos orgànics, en canvi, existeixen substàncies formades pels mateixos àtoms però amb propietats diferents que es coneixen com a isòmers. La diferència esta en la geometria dels enllaços, per la qual cosa els grafs són un bon model de representació tal com va mostrar Arthur Cayley. Un exemple d'isòmers es mostra a la figura 1.24.

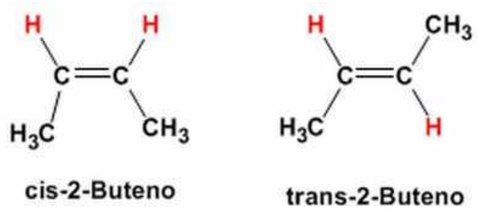


Figura 1.24: Exemple d'una parella d'isòmers.

Com podem veure en la figura 1.25, moltes molècules orgàniques formen una estructura amb un cicle format per 3, 4, 5, 6 o més vèrtexs.

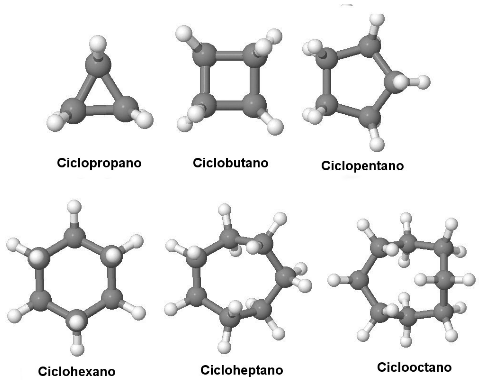


Figura 1.25: Exemple de molècules en forma de cicle.

També hi ha molècules amb estructura d'arbre. La molècula d'aigua és un exemple d'un arbre ben senzill, d'ordre 3 i mida 2. Però també hi ha arbres més grans, com ara qualsevol cadena d'àtoms de carboni. Els dos exemples els podem veure a la figura 1.26.

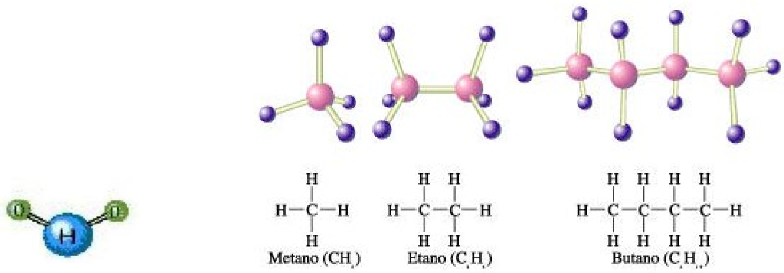


Figura 1.26: Molècula d'aigua i estructures amb àtoms de carboni.

La figura 1.27 mostra la molècula de Benzè i la de propanal, que són exemples de multigrafs, ja que tenen dobles enllaços que corresponen a arestes paral·leles.

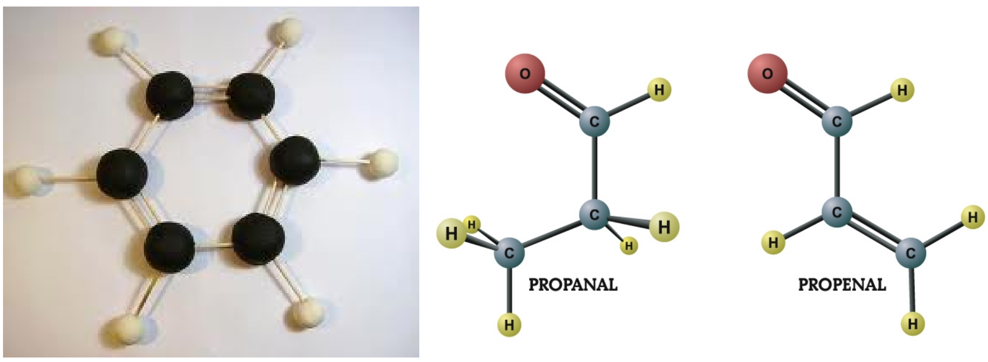


Figura 1.27: Molècula de benzè i molècula de propanal.

L'estructura de la glucosa també la podem representar amb un graf, figura 1.28. En aquest cas no és un arbre ja que clarament conté un cicle.

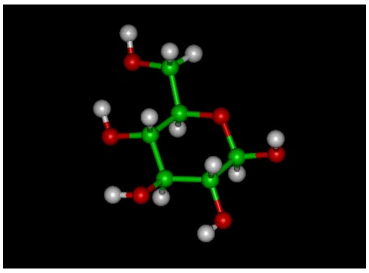


Figura 1.28: Molècula de glucosa.

Durant el procés de la glucòlisi, s'oxida la glucosa per tal d'obtenir energia per la cèl·lula en una sèrie de 10 reaccions enzimàtiques consecutives. Aquest procés es resumeix en la imatge gràfica de la figura 1.29 que a nivell de grafs es podria interpretar com que el graf es divideix en dos multigrafs i el cicle es parteix.

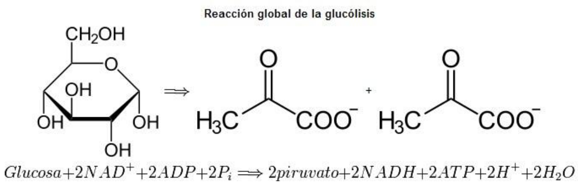


Figura 1.29: Esquema de la glucolisi.

Quan estem parlant d'un conjunt de molècules, els grafs també mostren les relacions entre elles. Per exemple, considerem un element bàsic com l'aigua, format per molècules com la de la figura 1.26. Considerem els ponts d'Hidrogen i els representem també en el graf com a arestes, de la mateixa forma que els enllaços, com mostra la figura 1.30. Com que aquests ponts depenen, en general, dels estats físics de l'aigua, els grafs obtinguts també. Així doncs, quan l'aigua es troba en estat gasós, no hi ha ponts entre les molècules i, per tant, el graf que formaria seria no connex. En el cas d'estat líquid, totes les molècules formen algun pont; però no són gaire estables i varien contínuament. En estat sòlid, els ponts formats són més estables i les molècules s'organitzen en forma d'hexàgon, una de les figures matemàtiques més estables.

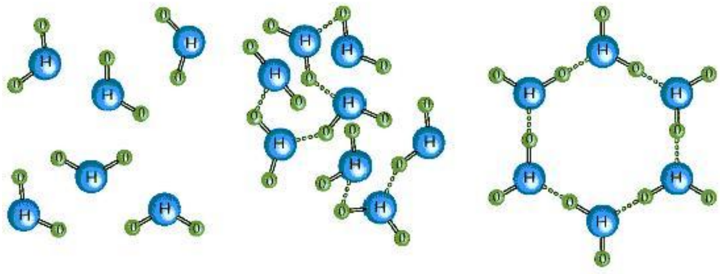


Figura 1.30: Enllaços segons l'estat físic de l'aigua: gasós, líquid i sòlid.

Utilitzar els grafs com a model de les estructures dels compostos químics també pot ser útil per a il·lustrar la diferència entre molècula i cristall iònic, en el que l'estructura geomètrica dels enllaços és essencial, si bé entra en joc la visió tridimensional. Com es veu a la figura 1.31.

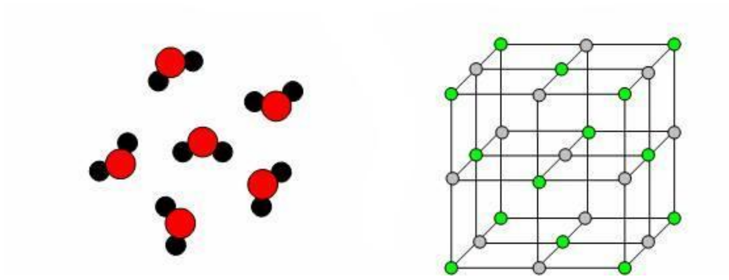


Figura 1.31: Diferència entre molècula i cristall iònic.

Com a exemple final destaquem l'ADN, la molècula que ens representa. A la figura 1.32 es mostren dues imatges de l'ADN. La primera és una imatge esquemàtica que ressalta la seva estructura com a graf. La segona dóna una visió tridimensional d'un fragment d'ADN. Va ser creada per la nostra classe dins de l'assignatura de química de 1r de Batxillerat amb el professor Jordi Marín. Vam utilitzar el programa *Imprudence* en el marc del projecte *Espurnik* (Departament d'Ensenyament de la Generalitat de Catalunya), que va donar lloc a una presentació en el congrés internacional de realitat virtual iED2012 de Boston a càrrec del professor del centre.

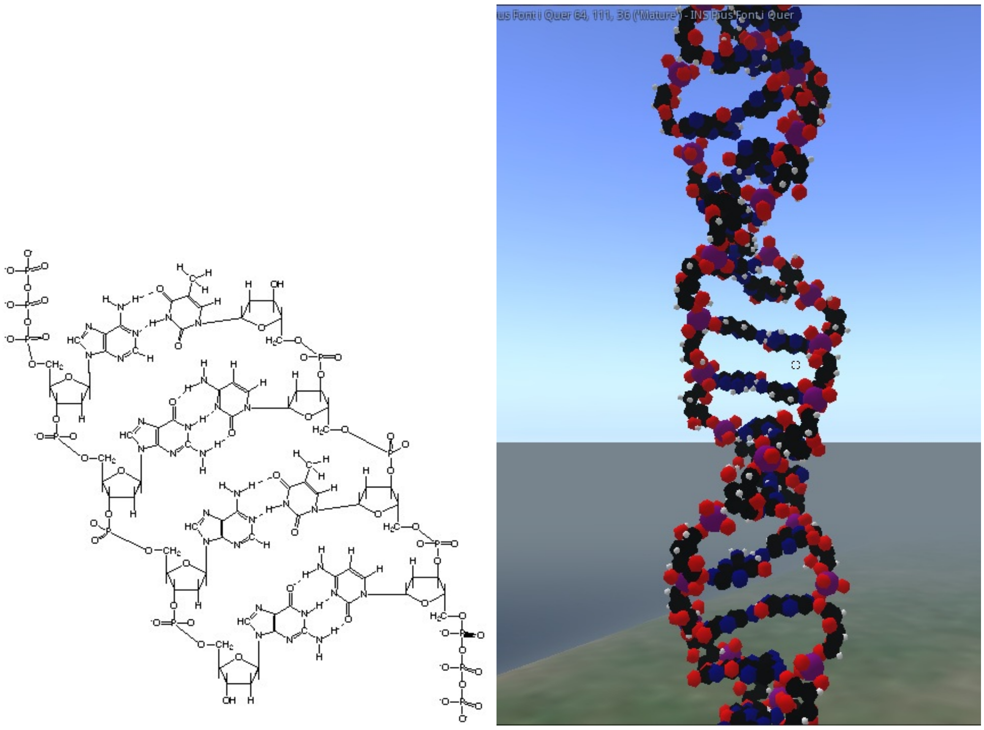


Figura 1.32: Representació de molècules d'ADN.

Així els grafs són un bon model per a representar l'estructura de les molècules o compostos químics en general, tenint en compte el paral·lelisme entre els seus elements que resumim en la taula següent.

Graf	$\longleftrightarrow$	Compost químic
vèrtex	$\longleftrightarrow$	àtoms
arestes	$\longleftrightarrow$	enllaços químics, ponts...

1.5.2 Grafs esportius

També es poden aplicar els grafs en moltes coses relacionades amb esports, el gimnàs, etc.

Per exemple, un torneig "round-robin", és a dir un torneig on cada equip juga contra tots els altres equips només una vegada. Aquest torneig es pot representar utilitzant un graf dirigit on cada equip representa un vèrtex. Les arestes anirien de cada equip vencedor al perdedor, de manera que quedaria un graf simple, sense arestes paral·leles ni llaços (ja que cada equip juga una sola vegada contra cadascun dels altres). Un cop creat aquest graf es veu molt més clarament i ràpidament quins equips són millors, calculant el grau de sortida (*Outdegree*) dels vèrtexs (equips).

Per tal de posar un exemple pràctic del nostre institut, represento gràficament de la manera esmentada un torneig de tenis taula.

Jugador 1 vs Jugador 2	7	11
Jugador 1 vs Jugador 3	10	12
Jugador 1 vs Jugador 4	11	4
Jugador 1 vs Jugador 5	5	11
Jugador 2 vs Jugador 3	9	11
Jugador 2 vs Jugador 4	11	2
Jugador 2 vs Jugador 5	11	13
Jugador 3 vs Jugador 4	11	8
Jugador 3 vs Jugador 5	7	11
Jugador 4 vs Jugador 5	3	11

A la taula costa veure quin dels 5 jugadors és millor. Per tal de millorar la visualització dels resultats, represento gràficament la informació de la taula mitjançant el graf corresponent. Cada jugador és un vèrtex, i cada partit una aresta del guanyador al perdedor. A aquest graf li calculo el grau d'entrada de cada vèrtex, i el pinto segons aquest, com mostro a la figura 1.33. A partir d'aquest graf es veu molt més clarament quins jugadors són millors que els altres, veient de color groc el que té el grau d'entrada més baix (és a dir, el millor jugador) i de color vermell el que té el grau d'entrada més elevat.

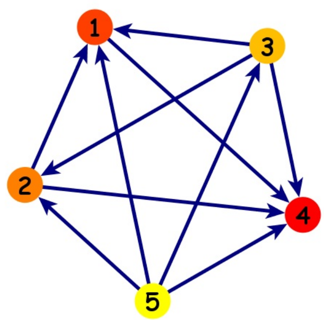


Figura 1.33: Graf del torneig de tenis taula.

1.5.3 On són els arbres de l'institut?

La resposta natural és al pati. En efecte, al pati podem trobar diferents tipus d'arbre com es mostra a la figura 1.34.

Figura 1.34: Arbres del pati de l'institut.

Però com que els arbres són un bon model de les estructures jeràrquiques i organitzatives, també trobem arbres en l'organització de l'institut (equip directiu, coordinadors, tutors...).

Com un bon exemple d'arbre fora del pati destaca la biblioteca com es veu a la figura (1.35). Hi ha molts llibres, revistes i documents per consultar i és molt important que estiguin ben organitzats. Això s'aconsegueix amb una bona classificació que té en compte el tipus de document, l'autor, l'any, etc. Aquesta classificació equival a un arbre.

Figura 1.35: Imatges de llibres classificats, a la biblioteca del centre i a la Biblioteca Nacional de Catalunya.

Per tal d'organitzar els llibres s'utilitza la **Classificació Decimal Universal (CDU)**. Aquesta manera de classificar la trobarem a totes les biblioteques.

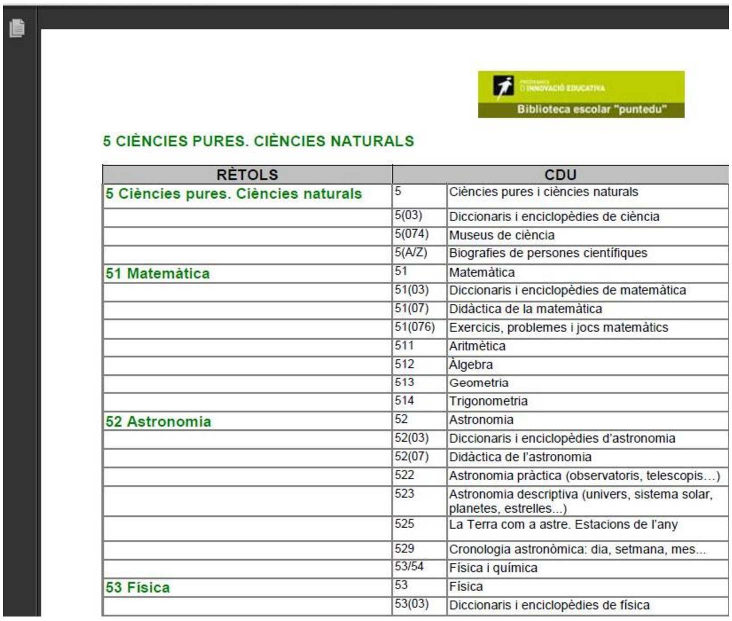
La CDU va ser desenvolupada a finals dels segle XIX pels juristes belgues Paul Otlet i Henri La Fontaine, fundadors de l'Institut Internacional de Bibliografia, a partir de l'adaptació de la Classificació Decimal de Melvil Dewey, d'Estats Units. Des del 1905 s'ha anat modificant i ampliant per actualitzar-se, i és un dels sistemes més utilitzats i traduïts en l'àmbit bibliogràfic. A Catalunya, Jordi Rubió en va realitzar l'any 1920 una primera adaptació.

La CDU és una classificació que divideix totes les àrees del coneixement humà en deu grans apartats que van del 0 al 9 i es van dividint seguint una estructura jeràrquica.

El primer nivell de classificació és el següent:

0. Generalitats : Diccionaris i Enciclopèdies.
1. Filosofia: Pensar i raonar.
2. Religió.
3. Ciències Socials: Política, educació, folklore...
4. Lingüística. Diccionaris.
5. Ciències exactes i naturals: Matemàtiques, zoologia...
6. Ciències aplicades: Oficis, producció...
7. Art, música, esports i oci.
8. Literatura: novela, poesia...
9. Història i Geografia. Biografia: Països, persones cèlebres...

Cada un d'aquests nivells es divideix en altres subnivells, que especifiquen el contingut del document. Així per exemple dins de la matèria 5, a les matemàtiques els correspon el 51, que es divideix en especialitats de les matemàtiques com es mostra a la figura 1.36, extreta del document sobre biblioteques escolars puntedu publicat per la Generalitat de Catalunya.


Biblioteca escolar "puntedu"

5 CIÈNCIES PURES. CIÈNCIES NATURALS

RÈTOLS	CDU	
5 Ciències pures. Ciències naturals	5	Ciències pures i ciències naturals
	5(03)	Diccionaris i enciclopèdies de ciència
	5(074)	Museus de ciència
	5(A/Z)	Biografies de persones científiques
51 Matemàtica	51	Matemàtica
	51(03)	Diccionaris i enciclopèdies de matemàtica
	51(07)	Didàctica de la matemàtica
	51(076)	Exercicis, problemes i jocs matemàtics
	511	Aritmètica
	512	Àlgebra
	513	Geometria
	514	Trigonometria
52 Astronomia	52	Astronomia
	52(03)	Diccionaris i enciclopèdies d'astronomia
	52(07)	Didàctica de l'astronomia
	522	Astronomia pràctica (observatoris, telescopis...)
	523	Astronomia descriptiva (univers, sistema solar, planetes, estrelles...)
	525	La Terra com a astre. Estacions de l'any
	529	Cronologia astronòmica: dia, setmana, mes...
	53/54	Física i química
53 Física	53	Física
	53(03)	Diccionaris i enciclopèdies de física

Figura 1.36: Classificació Decimal Universal de les ciències pures (adaptació puntedu).

Aquesta classificació, en el llenguatge de grafs, correspon a un arbre. L'esquema decimal està preparat per a que hi hagi 10 divisions cada vegada, però no tots els apartats estan assignats. L'esquema gràfic correspon doncs a un graf, amb les característiques següents:

- 1 vèrtex inicial, del que surten 10 arestes, per tant té grau 10.
- 10 vèrtexs de 1r nivell, als que arriba 1 aresta i en surten 10 més, per tant tenen grau 11.
- 10^2 vèrtexs de 2n nivell, també de grau 11.

  ...

- 10^r vèrtexs de nivell r que són els vèrtex finals als que arriba 1 aresta i no en surt cap més, per tant tenen grau 1

Òbviament el graf obtingut és un arbre del tipus *rooted-tree* explicat al capítol 1, i tots els vèrtexs terminals s'anomenen *fulles*.

1.5.4 Grafs a la sala d'ordinadors

Una altra de les aplicacions dels grafs que es troba en un institut són les xarxes d'ordinadors, que connecten entre si ordinadors, impressores o altres dispositius. Les xarxes tenen l'avantatge d'oferir un entorn versàtil i adaptable que facilita la feina en cooperació, i en el qual cada ordinador és autònom.

Un dels tipus més comuns són les **xarxes LAN**. Són xarxes d'àrea local, generalment d'ús privat. El seu abast es restringeix a una organització petita, com una empresa o un centre d'ensenyament. Com que les seves dimensions són limitades, cada ordinador es comunica amb qualsevol altre de la xarxa a gran velocitat de transmissió i amb poques errades.

En canvi, les xarxes MAN són xarxes d'àrea metropolitana petita. Connecten diversos segments de xarxes LAN mitjançant línies d'alta velocitat. Les xarxes WAN són xarxes d'àrea extensa que connecten ciutats, països i continents.

Pel que fa a la disposició física i lògica dels ordinadors, és a dir la connexió entre ells, les xarxes LAN poden ser de tipus:

- **lineal o bus:** Els ordinadors es connecten l'un darrere l'altre a la línia de transmissió, de manera que la xarxa només es veu afectada per les fallades de la línia.
- **en anella:** Cada ordinador es conecta amb els dos ordinadors adjacents. La línia de transmissió es tanca a través de les connexions en els equips, de manera que la fallada d'un equip provoca la fallada de la xarxa.

- **en estrella:** Tots els ordinadors es connecten a un dispositiu central anomenat switch, de forma que aquest element gestiona totes les comunicacions entre ells. El sistema és vulnerable a la fallada d'aquest element central.

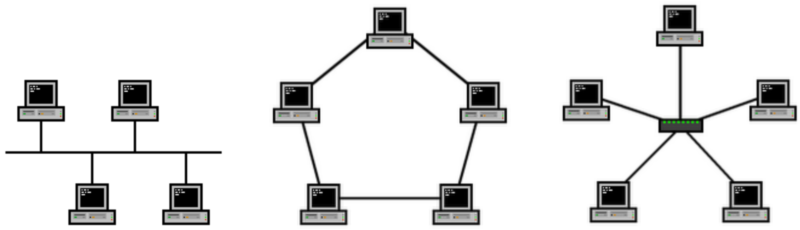


Figura 1.37: Xarxa en bus, xarxa en anell i xarxa en estrella.

A nivell de grafs, la connexió en anella correspon a un cicle i la connexió en estrella a un arbre. El graf d'una xarxa MAN vindria a ser la unió de diversos grafs de xarxes LAN.

Problema.
Donada una certa quantitat n de dispositius, ens podem preguntar quin tipus de connexió és millor. De fet això dependrà de la funcionalitat de la xarxa i per tant cal fixar les condicions exigides.

a) *Quina és la manera de que estiguin tots connectats, amb el mínim nombre de connexions?*

b) *De quina manera el nombre de dispositius intermitjos en connectar-ne dos qualsevols és mínim?*

Interpretem la xarxa com a graf. Tenim n vèrtexs. Per a que tots estiguin connectats entre ells ha de ser connex i necessitem almenys $n-1$ arestes. Per tant, amb aquest nombre mínim d'arestes tal i com s'ha justificat a la secció 1.4, ha de ser un arbre. Això respon la primera pregunta. Dels tipus de connexió anteriors, tant la connexió lineal com la d'estrella compleixen que són arbres. De fet, el tipus estrella té també altres avantatges. En un mitjà compartit per més de 2 equips és el que permet disposar de més ample de banda per una comunicació més econòmica i eficient.

En el cas del meu institut hi ha 3 sales d'ordinadors. He parlat amb els professors que ho gestionen per saber com estan connectats. Fa uns anys hi havia un switch a cada aula, d'on sortia un cable cap a un switch principal. Actualment, els cables de connexions van a parar a dos switchos que es comuniquen amb un cable amb el quadre principal. D'aquesta manera, s'estalvia un switch i es redueix el nombre de passos per connectar un ordinador amb un altre. En els dos casos l'estructura és bàsicament d'estrella.

En aquest context podem plantejar el problema següent, que podem resoldre aplicant les eines de teoria de grafs explicades a les seccions anteriors.

Problema.
Tenim 9 ordinadors connectats per cable segons la figura 1.38 on el valor indicat a cada aresta representa els metres de cable.

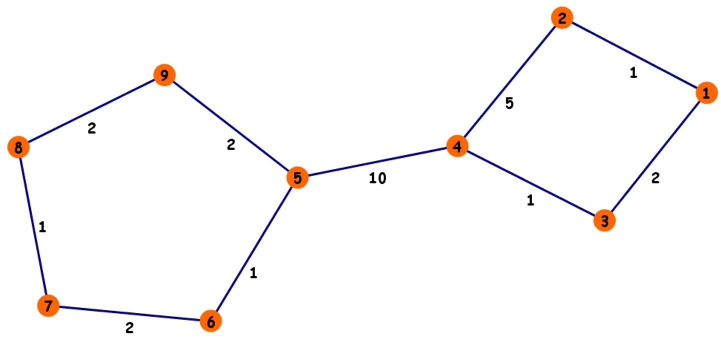


Figura 1.38: Graf dels 9 ordinadors i les seves connexions.

a) Volem comprovar l'estat de la xarxa enviant dades d'un ordinador a qualsevol altre de manera que les dades passin per tots els cables una sola vegada. *És possible fer-ho començant i acabant a l'ordinador 1? És possible fer-ho partint des d'un altre ordinador sense la necessitat d'acabar en el mateix?*

b) Dels 9 ordinadors se'n volen apagar 4, i deixar-ne 5 d'encesos, de manera que cada cable connecti un ordinador apagat amb un d'encès. *És possible? Com es pot fer?*

c) Es vol estalviar el màxim de metres de cable possibles mantenint els 10 ordinador connectats. *Quins cables puc suprimir? De quantes maneres es pot fer?*

Solució al problema.

a) No. És evident que per recórrer tot el graf començant i acabant al vèrtex 1, s'ha de passar per l'aresta $(4, 5)$ dues vegades. En canvi, la resposta a la segona pregunta és que sí, perquè el graf té 2 vèrtexs de grau senar (4 i 5), i la resta de grau parell. És a dir, és un graf eulerià. Per tant existeix un camí eulerià, però s'ha de començar i acabar en aquests dos vèrtexs. Per exemple, podriem passar pel camí $p = 5, 9, 8, 7, 6, 5, 4, 3, 1, 2$ o per $q = 4, 2, 1, 3, 4, 5, 9, 8, 7, 6$

b) No. El fet de poder repartir els ordinadors en apagats i encesos, de manera que 2 ordinadors apagats o dos encesos mai estiguin connectats per un cable, correspon a dir que el graf és bipartit. Per tant, la pregunta es podria plantejar com: el graf és bipartit? A partir de la definició és evident que qualsevol cicle d'un graf bipartit ha de tenir longitud parell. Com que el graf de la figura 1.38 té un cicle de longitud 5, $c = 5, 6, 7, 8, 9$, no pot ser bipartit.

c) Per tal de tenir els 9 ordinadors connectats amb les mínimes connexions possibles amb el mínim de metres de cable he de buscar un subarbre generador. Com que és arbre tindrà 8 arestes i, per tant, n'he d'eliminar 2 del graf inicial. A la secció 1.4 he explicat un mètode per a crear arbre generadors, que puc utilitzar per a resoldre aquest problema si tinc en compte d'eliminar sempre les arestes més llargues possibles: La més llarga és la $(4, 5)$, però no la puc treure perquè desconnectaria la xarxa; la següent més llarga és l'aresta $(2, 4)$ que sí que podem eliminar. Les següents més llargues són les de 2 metres, però la $(1, 3)$ no la podem eliminar perquè desconnectaria els ordinadors 1 i 2 de la resta; entre les altres de 2 metres, $(6, 7)$, $(5, 9)$ i $(8, 9)$, podem eliminar-ne una, la que vulguem. Un cop eliminada una qualsevol d'aquestes tres arestes, ja tenim un arbre i per tant no en podem eliminar cap més. Per tant hi ha tres arbres generadors amb el mínim de cable utilitzat.

Els grafs són doncs un bon model per a representar la topologia, és a dir la situació dels diferents elements, de les xarxes informàtiques.

Graf	$\longleftrightarrow$	Xarxa informàtica
vèrtex	$\longleftrightarrow$	processadors, impresores o altres dispositius
arestes	$\longleftrightarrow$	interconnexions entre ells

Capítol 2

Com són els grafs socials?

Aplicació de la teoria de grafs a les xarxes socials

Endinsa't prou en qualsevol cosa i hi trobaràs matemàtiques
D. Schlicter

Les xarxes socials, com a grups de persones que interaccionen, han existit sempre. Des de la família al grup de companys de classe, passant per l'equip esportiu o la colla de les vacances, les persones formem part de grups de persones que ens relacionem entre nosaltres. En general, aquestes xarxes de persones relacionades es formen a partir d'un espai, unes activitats i un temps compartit. Cada persona disposa d'una part d'informació de les altres persones de la xarxa i de les relacions entre elles, depenent del temps i l'espai en que coincideixen. Sovint en un lloc petit, com un poble, hi ha xarxes que abarquen tot el poble i per això es diu que en un poble tothom sap les xafarderies de tothom.

Amb les noves tecnologies, han aparegut les xarxes socials en línia, que tenen característiques diferents a les anteriors. Una diferència és que la condició sobre l'espai físic ha deixat de ser important, i les xarxes ja no depenen tant de la distància física. Una altra és que hi poden haver observadors externs a la xarxa que disposin d'informació de les persones de la xarxa i les relacions entre elles, sense que els membres de la xarxa en siguin conscients. Aquest és un dels riscos: la pèrdua de privacitat. L'objectiu d'aquest capítol és explicar què és l'anàlisi de xarxes socials a través del graf social corresponent, i plantejar el problema de la pèrdua de privacitat.

En aquest capítol en primer lloc explico què vol dir l'anàlisi de les xarxes socials, relacionat amb la Sociologia. Després presento breument les xarxes socials més comunes avui en dia. A continuació desenvolupo com es pot aplicar la teoria de grafs explicada en el capítol anterior a les xarxes socials en general. Aquesta anàlisi la posaré en acció en els capítols 3 i 4. A l'última secció d'aquest segon capítol comento els riscos que tenen les xarxes socials en línia.

2.1 Anàlisi de les xarxes socials

En les ciències socials, cap allà als anys 1930, Jacob Moreno, Kurt Lewin i Fritz Heider van ser els primers de representar en un diagrama les relacions entre un grup d'individus. Podríem dir que va ser l'inici de la utilització dels grafs com a model de les xarxes socials en la Sociologia.

En l'àmbit de l'antropologia, Radcliffe-Brown i Nadel van entendre que l'estructura social era el resultat de totes les relacions establertes entre individus: relacions de treball, amistat, negocis, etc. Posteriorment, a mitjans del segle XX, la investigació i l'anàlisi de l'estructura social va indicar que l'estructura de relacions establerta entre els individus afectava la societat com un tot. A causa d'aquest descobriment, als anys 60 i 70 ja hi havia investigadors a la Universitat de Harvard que buscaven maneres de mesurar i analitzar els fenòmens socials observats. A principis dels 70, Mark Granovetter va investigar el procés de difusió de la informació a través de xarxes socials, la qual cosa va marcar part de l'inici del que posteriorment ha estat anomenat Anàlisis de Xarxes Socials (Social Network Anàlisis, SNA).

En un món cada vegada més connectat i comunicat tecnològicament cal tenir en compte la capacitat de les xarxes socials de comunicació i la seva influència en diferents àmbits, i per tant la seva anàlisi és important.

Sabem que, a nivell de joves, l'estructura de relacions socials establertes és la base de la nostra comunicació i de l'intercanvi d'idees. Això influeix en l'acceptació o rebuig de propostes, estratègies i moltes altres coses. De fet, analitzar i comprendre l'estructura d'aquestes xarxes és un dels nous reptes que cal resoldre.

2.2 Descripció d'algunes xarxes socials en línia

A continuació faré un breu resum de les xarxes socials més conegudes. He agafat com a referència per una banda les que més utilitzem els joves, com el Facebook o el Twitter. Per altra banda, també he inclòs el LinkedIn, que a nivell d'adults és força utilitzat per temes de feina, per fer contactes professionals. També parlaré sobre el Flickr, utilitzat per totes les edats, per compartir imatges i videos.

Hi ha moltes altres xarxes socials com ara LastFM, Myspace... Però he comprovat que són poc conegudes a nivell d'institut i, per tant, no les he inclòs en aquest treball.

2.2.1 Facebook

El Facebook va ser fundat inicialment per Mark Zuckerberg, estudiant d'informàtica a la Universitat de Harvard (Estats Units), Dustin Moskovitz, Chris Hughes i Eduardo

Saverin, el 4 de febrer de 2004. En un principi la xarxa estava limitada als estudiants de la seva pròpia universitat, però aviat la van extendre a les universitats de l'àrea de Boston, i després de tot USA, i tot el món. Actualment és oberta a tota la gent, i la conformen més de 750 milions d'usuaris d'arreu del món.

El Facebook és una xarxa social gratuïta que pertany a una companyia privada: Facebook, Inc. El seu funcionament és similar al de qualsevol altre xarxa social. Els usuaris es registren i publiquen informació en el seu perfil. Aquesta xarxa permet agregar altres usuaris com a amistats. L'usuari pot compartir texts, videos, fotografies amb qualsevol altre usuari o només amb els que formen part dels seus contactes. A més, existeix el mur, un espai en el perfil de casa usuari que permet als seus amics publicar-hi missatges, visibles per tots ells. També ofereix aplicacions i jocs.

Figura 2.1: Pàgina d'inici del Facebook.

És una de les xarxes socials més conegudes actualment. Segons un estudi del 2009, era la xarxa més utilitzada en tot l'Internet. No he trobat dades explícites de Catalunya, però a Espanya hi ha aproximadament uns 14.500.000 usuaris i és un dels 20 països amb més usuaris. La pàgina amb més seguidors és la del Futbol Club Barcelona. Les edats mitjanes d'utilització del Facebook a nivell mundial són de 18 a 44 anys.

2.2.2 Twitter

El Twitter és un espai virtual creat l'any 2006. Inicialment servia com a via de comunicació dins d'una empresa i, mica en mica, es va anar extenent per tota la xarxa.

El Twitter permet publicar missatges de fins a 140 caràcters, amb l'objectiu principal de poder explicar a temps real el que t'està passant. Els missatges es diuen *Twitts*, i en català es parla de *piulades*. El fet que hi hagi un límit de 140 caràcters pot semblar un inconvenient, però serveix per resumir el que la gent vol penjar.

Els usuaris poden subscriure's a les actualitzacions realitzades per altres usuaris (s'anomenen seguidors) i, d'aquesta manera, estar al corrent de les actualitzacions que els interessin. Aquestes subscripcions esmentades són les que caracteritzen el Twitter com a xarxa social. A diferència de moltes altres xarxes, aquesta et crea un graf dirigit ja que el fet de "seguir"algun usuari no implica que aquest usuari et segueixi.

També cal destacar que, a diferència de Facebook, és una xarxa oberta que es pot explorar sense ser-ne usuari. Al capítol 4 n'utilitzaré dades.

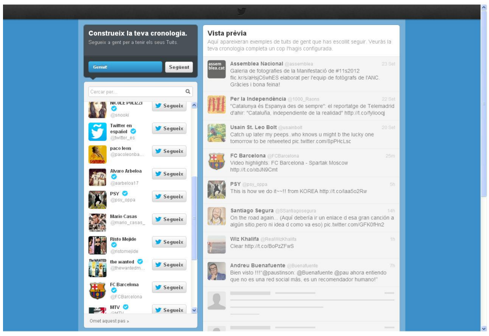


Figura 2.2: Vista prèvia del menú de twitts per seguir.

2.2.3 LinkedIn

LinkedIn és una xarxa social força utilitzada (el 2010 ja tenia més de 55 milions d'usuaris registrats). Aquesta xarxa social es diferencia de les altres xarxes socials com ara Facebook o Myspace, ja que el seu objectiu principal és facilitar les connexions a nivell professional. Serveix bàsicament per ajudar als professionals de tots els sectors a trobar altres professionals segons les seves necessitats, posar-se en contacte amb ells, crear negocis... Actualment LinkedIn s'ha convertit en una eina molt útil a la hora de crear connexions de treball gràcies a fòrums, grups i interconnexió amb altres xarxes socials. Però a nivell de joves no és gaire utilitzada.

Figura 2.3: Logo del linkedin.

2.2.4 Flickr

Flickr.com és un lloc web per compartir fotografies i videos, com ara el youtube.com. Va ser creat per Ludicorp l'any 2004 i comprada per Yahoo! l'any 2005.

Els principals avantatges són la gran qualitat en que poden ser penjades les imatges i videos, i la facilitat amb que es poden trobar i compartir. Com que no es necessita estar registrat per accedir als continguts de Flickr, molta gent l'utilitza per penjar fotografies de qualitat que després ensenyarà en altres xarxes socials mitjançant un enllaç. A part dels aproximadament 60.000.000 d'usuaris registrats, el Flickr és visitat per unes 80.000.000 de persones més. A l'agost del 2011 hi havia més de 6 bilions d'imatges penjades.

També es considera una forma indirecta de xarxa social. En aquest cas la interacció es correspon amb el fet de compartir contingut. A continuació incloc una imatge de la pàgina principal de Flickr, a la figura 2.4.

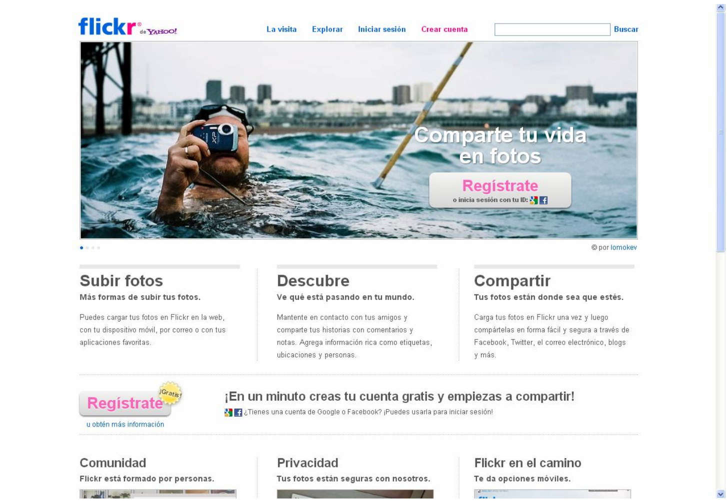


Figura 2.4: Vista prèvia de Flickr.

2.3 Els grafs socials com a model de xarxes socials

Per tal d'estudiar les xarxes socials s'utilitza la seva representació en forma de graf, obtenint el que s'anomena **graf social**. En aquesta secció explico com es pot aplicar la teoria de grafs per analitzar aquests grafs socials. Estudiaré el paral·lelisme entre els elements de les xarxes socials i els elements dels grafs i a continuació parlaré de les principals mesures utilitzades per l'anàlisi de xarxes socials, tant referents als usuaris com referents al graf en general.

2.3.1 Paral·lelisme d'elements

Per utilitzar els grafs com a model de les xarxes socials, hem d'establir una correspondència entre els seus elements de manera que així una xarxa social concreta queda identificada amb un graf G concret.

- **usuaris** $\leftrightarrow$ **vèrtex.** Cada persona o usuari de la xarxa és representat com un vèrtex o node.
- **relacions** $\leftrightarrow$ **arestes.** Cada relació entre dues persones es representa com una aresta que uneix els vèrtexs corresponents. Aquestes relacions poden ser de tipus diversos, ja siguin de vincles personals (amistat, respecte, preferència), transmissió

de recursos (operacions econòmiques i financieres) o comunicació (missatges en un i altre sentit).

- **mida de la relació ↔ pes de l'aresta.** No totes les relacions de les xarxes socials són igual de fortes o properes. Per reflectir això en el graf corresponent es poden utilitzar els pesos. Es tracta de donar un pes a cada aresta per valorar la força de la relació.

- **grup ↔ subgraf.** Si agafem un grup d'usuaris que ha estat delimitat per algun motiu concret que permet tractar-los com a subgrup concret, tenim un subgraf. Per exemple analitzant la xarxa social d'un institut, hom es pot restringir a una sola classe. De fet, a les xarxes socials en línia que funcionen a nivell mundial, el nombre d'usuaris és tan gran que la majoria d'estudis es limiten a subgrups.

2.3.2 Mesures referents a la centralitat dels usuaris

Un dels conceptes més utilitzats en l'anàlisi de xarxes és la centralitat de nodes, ja que permet una gran varietat d'aplicacions pràctiques. Un clar exemple d'una aplicació pràctica és la identificació de persones en una posició clau en una xarxa de deliqüència. És possible que la detenció de pocs individus amb un paper important dins una organització criminal la neutralitzi tota ja que talla la comunicació entre parts important de la xarxa de delinqüència.

En casos senzills la centralitat es veu d'una manera intuitiva a partir del gràfic, com en el graf de la figura 2.5 on és clar que el vèrtex 3 és el més central. Podem argumentar-ho dient que té un grau molt alt, que està el més proper possible de la resta de vèrtexs i que es troba en la majoria de camins que uneixen els altres vèrtexs. Les mesures bàsiques de centralitat no són res més que un intent de formalitzar aquestes idees per tal de poder calcular la centralitat en casos més complicats on no s'intueix tan fàcilment.

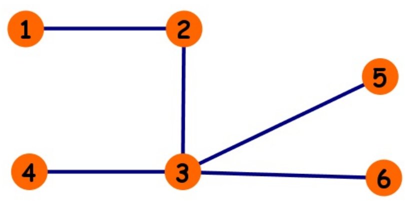


Figura 2.5: Exemple de graf social per valorar la centralitat.

Així, per calcular la centralitat dels nodes hi ha un seguit de mesures que s'apliquen als vèrtexs d'un graf. Per definir com n'és de central la posició d'un node es tenen en compte les tres mesures següents: la centralitat de grau $C_D(v)$ (*degree centrality*) la centralitat intermèdia $C_B(v)$ (*betweenness centrality*) i la centralitat de proximitat $C_C(v)$ (*closeness centrality*). Notem que també es poden considerar altres conceptes que hem vist a teoria de grafs com ara l'excentricitat.

- **Grau de centralitat.** El grau de centralitat C_D correspon al concepte de grau del vèrtex definit per grafs,
$$C_D(v_i) = \deg(v_i).$$
Aquesta mesura és la més senzilla i intuitiva de totes. En una xarxa social, un node amb una centralitat de grau elevada significa una persona amb moltes relacions i, per tant, cèntrica. Si el graf és dirigit cal tenir en compte per separat el grau d'entrada i el grau de sortida.

- **Centralitat intermèdia.** La centralitat intermèdia C_B d'un vèrtex és la suma de la proporció de geodèsiques, (camins de longitud mínima) que passen per ell, per cada parella de vèrtexs, és a dir
$$C_B(v_i) = \sum_{\substack{j,k \\ i \neq j \neq k}} \frac{\sigma_{v_j v_k}(v_i)}{\sigma_{v_j v_k}}$$
on n és l'ordre del graf, $\sigma_{v_j v_k}$ és el nombre de geodèsiques entre v_j i v_k, i $\sigma_{v_j v_k}(v_i)$ el nombre de geodèsiques entre v_j i v_k que passen per v_i.

  En una xarxa social una persona amb una centralitat intermèdia molt elevada significa que molta gent es relaciona a travès d'ella. Si es donés el cas que entre dues persones només hi hagués una sola geodèsica, les persones intermèdies serien claus ja que podrien ocultar o distorsionar la informació.

- **Centralitat de proximitat.** La centralitat de proximitat C_C d'un vèrtex té en compte la distància que hi ha entre aquest vèrtex i els altres, de manera que es calcula com l'invers de la suma de les distàncies entre aquest vèrtex i la resta.
$$C_C(v_i) = \frac{1}{\sum_j d(v_i, v_j)}.$$
Aquest valor sempre està entre 0 i 1. En una xarxa social, els valors més propers a 1 corresponen als usuaris que tenen relació més directe amb els altres usuaris.

- **Excentricitat.** L'excentricitat d'un vèrtex és la distància d'aquest vèrtex al que està més lluny d'ell.
$$\mathrm{exc}(v) = \max_{u \in V} d(v, u).$$
En una xarxa social, l'excentricitat d'un vèrtexs mesura la quantitat de persones que et separen de la persona més llunyana. Per tant les persones amb excentricitat més baixa són les més cèntriques.

Per exemple, si apliquem aquestes mesures al graf de la figura 2.5 obtenim

$$C_D(1) = 1 \quad C_D(2) = 2 \quad C_D(3) = 4\ldots$$

$$C_B(1) = 0 \quad C_B(2) = 4 \quad C_B(3) = 9\ldots$$

$$C_C(1) = \frac{1}{12} \quad C_C(2) = \frac{1}{8} \quad C_C(3) = \frac{1}{6}\ldots$$

$$\mathrm{exc}(1) = 3 \quad \mathrm{exc}(2) = 2 \quad \mathrm{exc}(3) = 2\ldots$$

Notem que, com que el graf és un arbre, el valor de $\sigma_{v_j v_k}$ que s'utilitza per al càlcul de la centralitat intermèdia és sempre 1.

2.3.3 Mesures referents al graf

Com hem vist al capítol 1, el **diàmetre** d'un graf es calcula com

$$\mathrm{diam}(G) = \max_{u,v \in V} d(u,v).$$

És l'excentricitat més alta, és a dir la longitud del camí més llarg entre tots els camins més curts que hi ha entre qualsevol parell de nodes. Dit en altres paraules, el diàmetre d'un graf és el major nombre de vèrtexs que has de travessar per anar d'un vèrtex del graf a qualsevol altre, pel camí més curt. Traduït al llenguatge de xarxes socials podem dir que el diàmetre és la quantitat màxima de gent que separa dues persones qualsevols.

L'experiment de Milgram del 1969 va ser un intent de mesurar el diàmetre del graf social dels Estats Units. Consistia en que una persona rebia una carta i l'havia de fer arribar a una altra persona que vivia a l'altra punta dels Estats Units, però només la podia enviar a algun conegut seu, i que aquest fes el mateix fins que algú que conegués a la persona establerta acabés la cadena. Milgram va acabar l'estudi dient que el diàmetre del graf social dels estats units era 6. De fet, tot i tenir molts vèrtexs, les xarxes socials són xarxes "small world", és a dir amb un diàmetre petit.

De manera anàloga, el **radi** d'un graf es calcula com

$$r(G) = \min_{v \in V} \mathrm{exc}(v).$$

El **centre del graf** està format pels vèrtexs que tenen excentricitat igual al radi. A nivell de xarxa social, els usuaris que estan al centre del graf arriben a tots els altre usuaris amb un nombre de passos igual al radi i, per tant, inferior als que necessitarien tota la resta d'usuaris.

També es poden considerar com a mesures del graf les **mitjanes de mesures** dels vèrtexs o de les arestes. Així es parla del grau mitjà del graf,que és la mitjana dels graus dels vèrtexs. Pero aplicant les propietats es pot calcular directament:

$$\text{grau mitjà de } G = \frac{\sum\limits_{v \in V} \deg v}{n} = \frac{2|E|}{n}.$$

2.4 Riscos en les xarxes socials en el tema de la privacitat

A les xarxes socials online actuals es fa pública una gran quantitat d'informació personal. Part d'aquestes dades personals han estat introduïdes pel mateix usuari al registrar-se o a través de comentaris i fotos. Però també hi ha informació de l'usuari introduïda per altres usuaris, per exemple en etiquetar fotos. La gran majoria dels usuaris veu aquest intercanvi d'informació de la mateixa manera que la comunicació que es viu en directe i no és conscient del que suposa per a la seva privacitat el fet que aquesta informació estigui escrita i registrada a la xarxa. No podem oblidar que els administradors hi tenen accés fins i tot encara que l'usuari l'esborri.

Però a més de les dades personals dels usuaris (sobre les quals els propis usuaris tenen bastant control) les xarxes socials contenen més informació: les relacions. La privacitat d'aquestes relacions és molt més complicada de controlar, ja que no depenen només de l'usuari en qüestió i, per tant, és impossible de controlar per un sol usuari. De fet, les relacions entre els usuaris donen molta informació sobre els líders i els nuclis dels grups, si hi ha sectors diferenciats, gent marginada, etc.

A partir del registre d'interaccions dels usuaris, siguin missatges directes, etiquetes en fotografies, comentaris de les publicacions o altres, es pot extrapolar informació sobre les relacions entre els usuaris. Utilitzant eines de la teoria de grafs, com les que s'han explicat en el capítol anterior, els administradors de la xarxa poden visualitzar i analitzar els usuaris i les seves relacions. Per exemple, poden saber quines persones són més populars, quines tenen més influència, quins són els temes de més interès, etc. Hi ha molts motius per voler recolectar aquesta informació de les xarxes socials online. A continuació s'expliquen algunes maneres de treure'n partit.

- **Publicitat:** Les empreses poden utilitzar la informació dels usuaris amb la finalitat de saber els seus gustos i interessos i, per tant, fer publicitat dirigida. També els serveix per crear anuncis personalitzats: adequats a l'edat de l'usuari, la localització geogràfica, sexe, interessos, professió...

  També es pot fer servir la informació per tal d'escollir certs punts estratègics i concentrar la publicitat allà, de manera que es redueixen costos però s'aconsegueix arribar a la mateixa quantitat de gent. Per tal de fer això, només cal escollir els punts estratègics segons la centralitat dels nodes. Així, per exemple, concentrant una campanya publicitària en uns quants nodes que estiguin ben comunicats es pot assegurar que tots els usuaris coneixen algú a qui la publicitat ha arribat.

- **Phishing:** La informació de cada perfil pot ser molt útil per un atacant que vol fer-se passar per un conegut de l'usuari o per una empresa en la qual l'usuari confia. Si es vol fer passar per un amic de l'usuari, disposa dels noms dels seus amics i també informació d'ells. En el cas que es vulgui fer passar per una empresa, disposa d'informació personal de l'usuari, cosa que pot contribuir a l'elaboració de missatges més creïbles.

- **Cossos policials:** Com ja hem dit, els grafs socials permeten analitzar les diverses interaccions entre un conjunt d'individus. Per posar un exemple pràctic, analitzant un graf d'un grup "delictiu"es pot trobar quin individu fa de connexió entre diverses cèl·lules del grup. Per tant la detenció d'aquest és probable que desestructuri el grup sense haver hagut de detenir gent en cada cèl·lula.

Potser molta gent pensa que aquests riscos no són reals, o estan molt exagerats. Però el millor argument per adonar-nos que realment aquesta informació queda registrada i perdem privacitat ens el donen les mateixes xarxes socials en la lletra petita de les condicions de privacitat. Potser no totes ho fan explícit, però algunes sí. Per exemple el Twitter en crear un usuari nou, ofereix un enllaç (https://twitter.com/privacy) en el qual explica les condicions de privacitat. Entre elles es pot llegir textualment:

You understand that through your use of the Services you consent to the collection and use (as set forth in the Privacy Policy) of this information, including the transfer of this information to the United States and/or other countries for storage, processing and use by Twitter.

Més concretament, a l'apartat de *Information Collection and Use* diu que això no es refereix només als missatges que hom publica, sinó també a l'estructura de seguidors, és a dir la xarxa social. També remarca que la informació queda arxivada (*United States Library of Congress*). A continuació s'inclou part del text:

Our Services are primarily designed to help you share information with the world. Most of the information you provide us is information you are asking us to make public. This includes not only the messages you Tweet and the metadata provided with Tweets, such as when you Tweeted, but also the lists you create, the people you follow, the Tweets you mark as favorites or Retweet, and many other bits of information that result from your use of the Services. Our default is almost always to make the information you provide public for as long as you do not delete it from Twitter, but we generally give you settings to make the information more private if you want. Your public information is broadly and instantly disseminated. For instance, your public user profile information and public Tweets may be searchable by search engines and are immediately delivered via SMS and our APIs to a wide range of users and services, with one example being the United States Library of Congress, which archives Tweets for historical purposes. When you *share information or content like photos, videos, and links via the Services, you should think carefully about what you are making public.*

Als capítols següents s'ha fet una simulació de com es pot obtenir informació de la xarxa social utilitzant els grafs, de manera que exemplifica l'anterior.

Capítol 3

Xarxes socials en primera persona

TREBALL DE CAMP

Una imatge val més que mil paraules
(Popular)

En aquest capítol, per tal de dur a la pràctica l'anàlisi de xarxes socials explicat al capítol anterior, he dissenyat un experiment amb l'objectiu de comprovar l'eficàcia de la teoria de grafs, i mostrar la quantitat d'informació que es pot extreure d'una simple xarxa, gràcies a l'aplicació de les matemàtiques.

Les xarxes socials són tan extenses que per fer un estudi cal reduir-se a una part. Amb una enquesta a la meva classe de 1r de Batxillerat, podia obtenir una xarxa de mida adequada, fer el graf social corresponent i analitzar-lo, i comparar-lo amb la meva percepció de la realitat. Aquest experiment doncs, ha consistit en:

- Detectar algunes relacions socials reals en el grup classe, a través d'una enquesta.
- Representar la informació en un graf com a model.
- Analitzar el graf social utilitzant la teoria de grafs i la seva relació amb les xarxes socials, i extreure'n conclusions.
- Contrastar les conclusions amb la meva percepció de la realitat.

En aquest capítol descric en primer lloc els objectius i disseny de l'enquesta. Després explico com he fet la recollida de respostes. A la tercera secció faig un parèntesis de les enquestes per explicar el projecte Visone, l'eina informàtica que faré servir per tractar els grafs. La secció quatre conté l'anàlisi dels grafs socials de les preguntes de l'enquesta. Utilitzo la primera pregunta per detallar el funcionament del Visone (com a manual del programa). Per la resta de preguntes comento directament els resultats.

3.1 Objectius i disseny de l'enquesta

Per portar a terme una anàlisi d'una xarxa, cal tenir en compte els següents conceptes, que concretem de cara al nostre experiment.

- **Conjunt a estudiar:** en el nostre cas, una classe de joves de 1r de Batxillerat, com a mostra del Batxillerat.
- **Relacions i qualitats que ens interessen:** percepció en el grup d'habilitats acadèmiques, esportives o relacionals. (Simplement d'amistat i confiança.)
- **Informació necessària:** caldrà obtenir l'opinió de les persones de la classe referent als aspectes mencionats.

Per tal d'aconseguir informació, s'han de definir els tipus de fonts. Normalment s'utilitzen fonts secundàries i primàries. **Les fonts primàries** (directes) proporcionen la informació més rellevant. Van des de l'observació directa fins a entrevistes o enquestes, que són les més utilitzades en processos d'investigació social. En aquest cas he utilitzat un qüestionari amb preguntes concretes.

Les fonts secundàries (indirectes) serien la documentació on es faci referència als conjunts socials que estem analitzant. En el cas que tracta aquest treball, no disposo d'informació documentada sobre enquestes similars a joves de la ciutat, ni comarca, ni tan sols a nivell de Catalunya. Per altra banda, el fet de conèixer el grup enquestat en primera persona podia aportar informació, però hi havia el risc que influís en l'anàlisi. És per aquest motiu que aquesta informació no s'ha contrastat fins al final.

Un **qüestionari** és precisament l'instrument que m'ha permès recaptar la informació necessària per aquest experiment. Ara bé, el qüestionari ha de complir certes normes internes de coherència, qualitat i objectivitat.

Els punts més importants en la preparació d'una enquesta són:

- fixar el tipus d'informació que es vol que donin les preguntes,
- concretar les preguntes que es faran,
- analitzar la interrelació de les preguntes,
- organitzar les preguntes en una seqüència senzilla i lògica.

Vull detectar diferents tipus de relacions entre els alumnes de la classe. Per una banda, vull que tinguin relació amb activitats acadèmiques (grups per fer treball) o activitats de lleure (equip esportiu), i per l'altra banda de relació de confiança i amistat entre la gent. Per tal d'obtenir aquesta informació, he redactat 5 preguntes concretes. La figura 3.1 mostra el qüestionari que vaig passar a la classe.

Respon les següents preguntes utilitzant el número de la llista enlloc del nom de la persona.

NÚMERO DE LA PERSONA ENQUESTADA:

1) Amb qui faries un treball de grup de català o castellà, en que la nota importa molt? (4 persones).
2) Amb qui faries un treball de grup de ciències, en que la nota importa molt? (4persones).
3) Amb qui aniries d'equip per fer un partit de futbol? (4 persones)
4) Gent amb qui pots parlar de tot, amb qui et potes be. (màxim 5)
5) Amb qui compartiries un secret o una cosa important per tu (màxim 3)

Figura 3.1: Enquesta sense omplir.

Per la redacció de les cinc preguntes de l'enquesta s'ha tingut en compte que les paraules han de ser clares, concretes i sense complicacions lingüístiques i/o coses per l'estil. La pregunta 1 i la 2 fan referència a la part d'activitats acadèmiques, la 3 tracta sobre una activitat esportiva, i la 4 i la 5 tracten les relacions d'amistat. També he tingut en compte l'ordre, començant amb preguntes menys personals i més fàcils de respondre, i després progressant cap als aspectes més personals. A més, les preguntes inicials busquen despertar l'interès de la persona enquestada ja que el grup correspon a la seva classe d'institut.

Per tal de protegir la identitat de les persones que responien l'enquesta, he etiquetat de manera aleatòria a cada persona. Aquesta etiqueta aleatòria l'he construït a partir de la llista alfabètica de la classe assignant a cada persona un número aleatori entre 1 i 26 (ja que he passat l'enquesta a 26 persones) sempre i quan no hagi estat ja utilitzat. A la figura 3.2 es mostra una de les possibles taules d'anonimització de l'enquesta. De fet, fixem-nos que per un total de 26 persones hi ha

$$26! = 4.032914611 \times 10^{26}$$

possibles taules diferents que corresponen a les possibles permutacions. Per tant la probabilitat que una persona externa relacionés les persones reals de la classe amb el número assignat (un cop destruïda la taula) és gairebé 0, ja que:

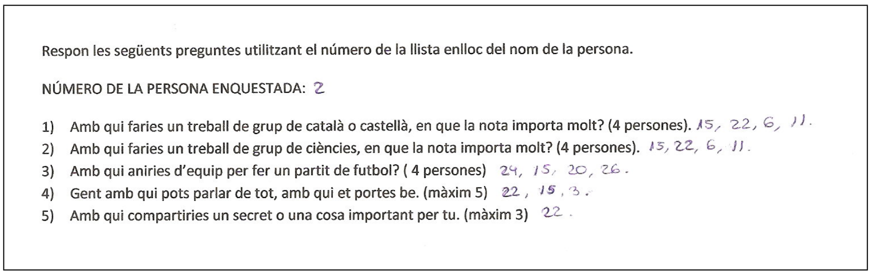
$$p = \frac{1}{4.032914611 \times 10^{26}} \leq 4.04 \times 10^{-26}.$$

Persones	Número aleatori
Nom 1	12
Nom 2	5
Nom 3	24
⋮	⋮
Nom 25	26
Nom 26	7

Figura 3.2: Exemple de taula aleatòria per anonimitzar l'enquesta.

3.2 Recollida de respostes

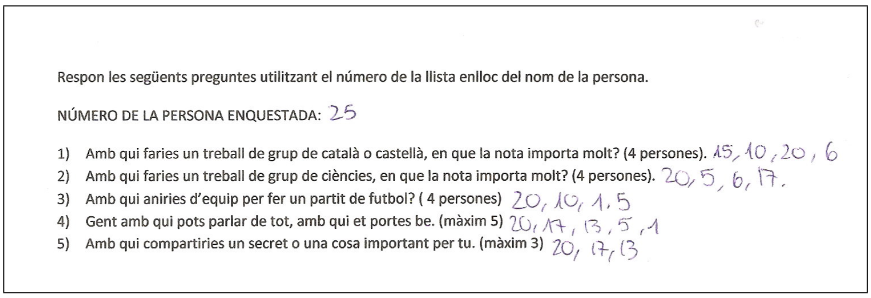
S'ha passat l'enquesta a una classe de 26 persones, numerant-les amb una permutació, com s'ha explicat abans, per conservar l'anonimat. Tot i això, alguna persona va posar el nom, cosa que vol dir que li importa poc la privacitat i que no sap seguir les instruccions. A les figures 3.3 i 3.4 es mostren dues enquestes concretes amb les seves respostes.


Respon les següents preguntes utilitzant el número de la llista enlloc del nom de la persona.

NÚMERO DE LA PERSONA ENQUESTADA: 2

1) Amb qui faries un treball de grup de català o castellà, en que la nota importa molt? (4 persones). 15, 22, 6, 11.
2) Amb qui faries un treball de grup de ciències, en que la nota importa molt? (4 persones). 15, 22, 6, 11.
3) Amb qui aniries d'equip per fer un partit de futbol? (4 persones) 24, 15, 20, 26.
4) Gent amb qui pots parlar de tot, amb qui et portes be. (màxim 5) 22, 15, 3.
5) Amb qui compartiries un secret o una cosa important per tu. (màxim 3) 22.

Figura 3.3: Respostes de la persona 2 a l'enquesta.


Respon les següents preguntes utilitzant el número de la llista enlloc del nom de la persona.

NÚMERO DE LA PERSONA ENQUESTADA: 25

1) Amb qui faries un treball de grup de català o castellà, en que la nota importa molt? (4 persones). 15, 10, 20, 6
2) Amb qui faries un treball de grup de ciències, en que la nota importa molt? (4 persones). 20, 5, 6, 17.
3) Amb qui aniries d'equip per fer un partit de futbol? (4 persones) 20, 10, 1, 5
4) Gent amb qui pots parlar de tot, amb qui et portes be. (màxim 5) 20, 17, 13, 5, 1
5) Amb qui compartiries un secret o una cosa important per tu. (màxim 3) 20, 17, 13

Figura 3.4: Respostes de la persona 25 a l'enquesta.

En general l'enquesta ha anat molt bé, amb una participació total de la classe. Era curta i ràpida de contestar, un dels objectius inicials. Les preguntes eren clares cosa que remarca la fiabilitat de les respostes. Tot i això, hi van haver petits malentesos en les 3 primeres preguntes, ja que alguns pensaven que els grups eren en total de 4 persones i que, per tant, n'havien de posar només 3. També hi va haver una persona que no va posar el seu nombre identificador a la part on és requerit. Tot i descobrir-ho a posteriori, com que només era una, va ser fàcil de solucionar.

Per poder analitzar bé les respostes amb l'ajuda dels grafs, cal prèviament fer el buidat de les enquestes en un format adequat. Aquest buidatge l'he fet en un fitxer d'Excel, de manera que he creat un full per cada pregunta.

Hi ha diferents formes d'organitzar les dades de les relacions entre un mateix conjunt de persones. Poso a la primera columna i la primera fila els noms de les persones (en aquest cas números) com a etiquetes de files i columnes. Aleshores he marcat amb un 1 les caselles en que la persona de la fila ha escollit el de la columna. De fet, és la idea de la matriu d'adjacència.

La figura 3.5 mostra la taula de la pregunta 1 com a exemple.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Pr. 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1						1									1			1		1						
2						1					1				1							1				
3						1	1	1							1											
4						1				1										1		1				
5															1	1				1			1			
6		1													1					1		1				
7		1	1						1						1											
8							1						1		1					1						
9															1					1		1	1			
10		1				1														1		1				
11						1				1								1		1						
12				1					1	1									1							
13																	1	1		1			1			
14						1									1					1		1				
15			1				1	1	1																	
16						1												1		1						1
17									1		1		1							1						
18	1									1										1						1
19		1				1										1				1						
20	1					1				1					1											
21						1				1					1					1						
22		1				1			1											1						
23					1	1									1					1						
24		1	1												1	1										
25						1				1					1					1						
26	1										1							1					1			

Figura 3.5: Buidatge de la pregunta 1 en una taula d'Excel.

Interpreto cada taula com a matriu d'adjacència del graf. Cada persona equival a un vèrtex, i les numerarem com $v_1, \ldots, v_n$. Cada aresta equival a la relació entre aquestes persones. No és una relació simètrica i per tant es tracta d'un graf dirigit. Així si la persona v_i ha triat a v_j, tindrem l'aresta $e_{ij} = \overrightarrow{v_i v_j}$, que no és igual a $e_{ji} = \overrightarrow{v_j v_i}$.

Per dibuixar el graf a partir de la matriu d'adjacència i calcular les mesures explicades al capítol anterior, utilitzaré el programa Visone.

3.3 El projecte Visone, al servei de l'anàlisi de xarxes socials

Visone és un projecte per desenvolupar models i algoritmes per integrar l'anàlisi i la visualització de les xarxes socials. De fet, el seu nom correspon a:

VI = *visual* **SO** = *social* **NE** = *Network*

Figura 3.6: Logo del programa Visone.

Una part important del projecte és el disseny i implementació d'eines informàtiques destinades a la recerca i la docència en anàlisi de xarxes socials. Està especialment dissenyat per permetre l'aplicació de mètodes visuals innovadors i avançats amb facilitat i precisió tant als que són especialistes, com als que no ho són.

3.3.1 Origen del programa

El projecte visone va néixer d'un contacte entre el Grup d'Algoritmes i Estructures de dades del Departament d'Informàtica i el Grup de Política domèstica i Administració pública del Departament de Política i Gestió de la Universitat de Konstanz, Alemanya. El 1996 es va constituir en aquella Universitat un grup de recerca interdisciplinari finançat pel projecte conjunt de Dorothea Wagner (doctora en ciències de la computació) i Wolfgang Seibel (doctor en ciències polítiques) que tenia també com a membres inicials Ulrik Brandes, Patrick Kenis, Jörg Raab, i Volker Schneide. Però el 1995, D. Wagner ja havia publicat l'article *VLSI Network Design: a Survey.* a la revista "Handbooks in Operations Research/Management Science" d'Holanda.

Així podem dir que la idea original del projecte es remunta a l'any 1995, i té doncs la mateixa història que els de la meva generació. El projecte s'ha desenvolupat al mateix temps que nosaltres creixíem.

3.3.2 Característiques principals del programari

El software desenvolupat pel grup de recerca dins del projecte també s'anomena Visone, i les seves característiques principals són:

- interfície gràfica interactiva, adaptada a les xarxes socials.
- visualitzacions de xarxa innovadores.
- possibilitat de representar relacions de diferents tipus.
- disponible amb Java per a Windows, Linux i MacOS.
- importació i exportació de formats estàndard per a les dades de xarxes socials.
- qualitat d'exportació en format JPEG, PDF, SVG, Metafile, i altres formats.

Es pot descarregar el Visone gratuïtament de la pàgina web: http://www.visone.info/html/download.html on hi ha totes les versions, com es mostra a la figura 3.7. El programa no necessita una instal·lació prèvia. Les instruccions són en anglès.

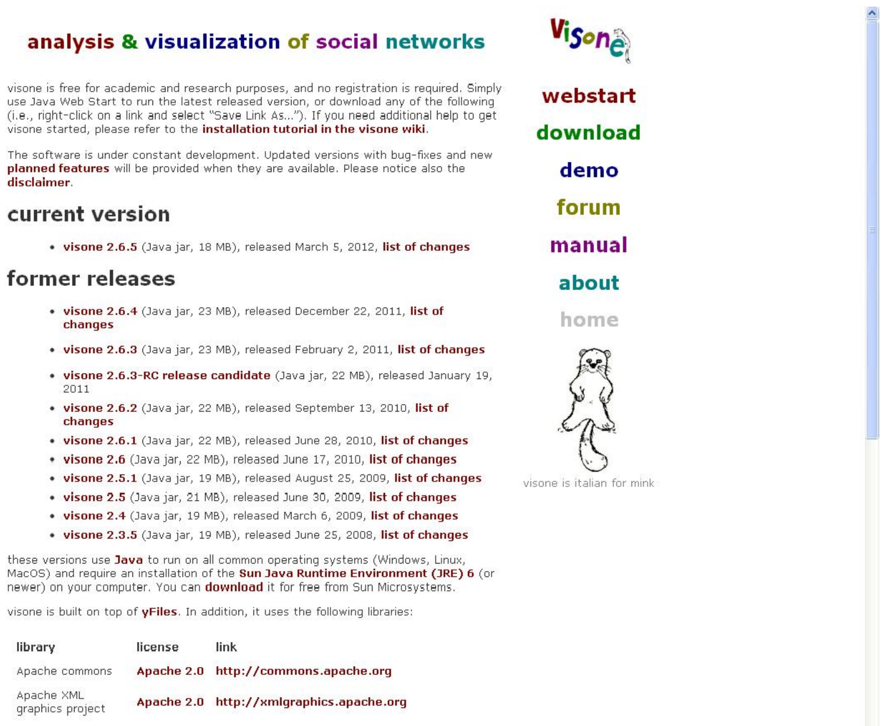


Figura 3.7: Pàgina web http://www.visone.info/html/download.html

A continuació descrivim els trets principals d'aquest software que utilitzarem a la resta del capítol. A l'executar el programa s'obre la pantalla principal, dividida en tres zones: la zona superior de menús, al marge esquerre la zona d'operacions predeterminades i a la dreta l'àrea per dibuixar els grafs.

La secció de menús és molt semblant a altres programes, amb un format similar, la qual cosa en facilita l'aprenentatge i l'ús. Les pestanyes són `File, Nodes, Links, Layout View` i `Help`.

En la pestanya de `File`, hi ha les opcions relacionades amb gestionar i administrar els arxius: `open, create, save`... Els desplegables `Nodes` i `Links` contenen les opcions que ens permeten canviar la visualització de vèrtexs i les seves connexions, com ara canviar color, grandària, forma...

La secció d'operacions predeterminades serveix un cop ja creat el graf. Es divideix en dues parts: `Analisis/Visualization` i `Selection/Geometry/Templates`.

L'apartat `Analisis` serveix per calcular certs paràmetres en el graf com ara mesures de centralitat, grau, etc. L'apartat `Visualization`, en canvi, serveix per poder triar com visualitzes els paràmetres obtinguts amb l'anàlisi, amb colors diferents, mides o ubicació. `Selection` serveix per veure totes les propietats de cada vèrtex. `Geometry`, en canvi, serveix per poder realitzar canvis en la distribució dels vèrtexs, la seva rotació, escalament, etc. L'última secció, `Templates`, permet controlar el número i tipus de vèrtexs i arestes que es van integrant al graf.

3.3.3 Els primers passos amb Visone

Les funcions bàsiques per tal de començar a utilitzar el Visone són les següents.

- Per crear un vèrtex, cal posar el ratolí on vulguis crear-lo, i clicar amb el botó dret. El programa donarà valors predeterminats (forma i color) que després es podran canviar a través de les opcions.
- Per crear arestes, cal clicar en el vèrtex inicial i després clicar en el vèrtex terminal. Altre cop, el programa donarà valors predeterminats a aquestes arestes. Si en comptes de clicar en un vèrtex terminal, es clica sobre l'espai en blanc automàticament es crea un nou vèrtex enllaçat a l'anterior.
- Per tal de moure un o més vèrtexs, s'ha de clicar a l'icona en forma de llapis que hi ha al menú de d'alt, al costat de les lupes, i després de clicar-la arrossegar el vèrtex.

Quan es tracta d'analitzar les xarxes socials, es tendeix a importar matrius de dades d'un altre fitxer. L'organització d'aquestes dades acostuma a ser de dos tipus: mode 1 (les relacions entre un conjunt i ell mateix) i mode 2 (un conjunt que es refereix a un conjunt totalment diferent). En el nostre cas, farem servir el mode 1, ja que analitzarem les relacions de les persones d'una classe entre elles. Per tal de recollir les dades, i després importar-les al Visone farem servir l'Excel.

3.3.4 Altres programes per a l'anàlisi de xarxes socials

Apart d'aquest programari, també existeixen altres programes que poden ser d'utilitat per a l'anàlisi de xarxes socials. Per exemple hi ha el programa del grup Sienna, localitzat a Seattle (USA), que és de compra. El programa UCINET és també de compra i està comercialitzat per Analytic Technologies localitzats a Lexington (USA), i es complementa amb el Netdraw com a mòdul de visualització dels grafs.

Un altre programa és el Pajek (aranya en Slovene), només per Windows. Es pot descarregar gratuïtament a la pàgina web http://pajek.imfm.si/doku.php?id=download. Està pensat per xarxes molt grans, i per tant fora de l'abast d'aquest treball. Va néixer també al 1995-1996 a Eslovènia, un país que pocs anys abans havia declarat la seva independència.

3.4 Anàlisi de les respostes

Per cada pregunta amb les seves respostes he d'obtenir el graf social corresponent, representant cada alumne de la classe com un vèrtex, i cada relació com una aresta. Així tindré 5 grafs diferents com a model de les relacions entre els alumnes de la meva classe, en diferents aspectes.

Per obtenir el graf transformo el fitxer Excel en .cvs, el visualitzo i en calculo els paràmetres amb el programari Visone.

A continuació comento els resultats de l'anàlisi de cada pregunta per separat. Finalment, en trec unes conclusions globals i les comparo amb la meva percepció de la realitat.

3.4.1 Anàlisi detallat de la pregunta 1

En aquest apartat descriuré amb detall com construeixo la representació visual del graf de les respostes a la pregunta 1 utilitzant el Visone. En certa manera aquest apartat es pot veure com a guia d'utilització d'aquest programa.

Primer hem d'importar les dades al programa Visone. Per tal de fer-ho, transformarem les dades emmagatzemades en un fitxer Excel per tal que tingui l'extensió .cvs:

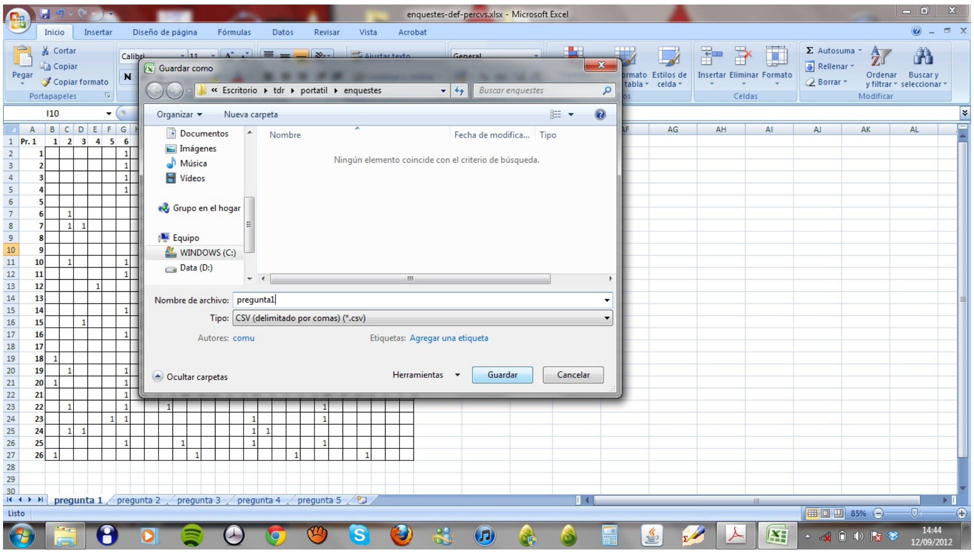


Figura 3.8: Base de dades en extensió .cvs

Aquest format és compatible amb el Visone i, per tant, no hi ha cap problema a l'obrir-lo. En les opcions d'importació cal seleccionar `Row labels` i `column labels`, per tal que importi les etiquetes que corresponen al format de les dades de l'Excel.

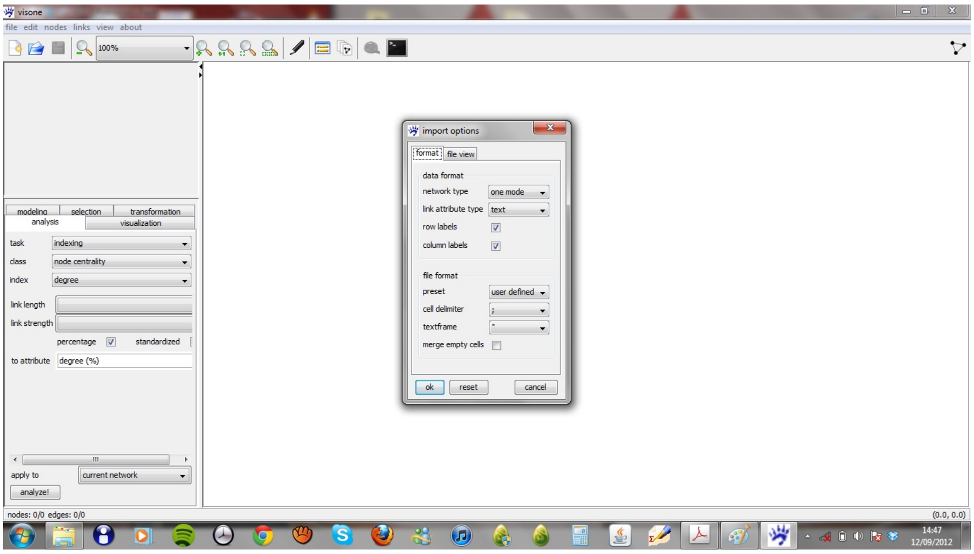


Figura 3.9: Importació de les dades .cvs al Visone.

Un cop importat el fitxer .cvs, el Visone dibuixa els vèrtexs i les arestes corresponents a les dades del fitxer i ofereix la visualització del graf de la figura 3.10, a partir de la qual ja es pot començar a treballar.

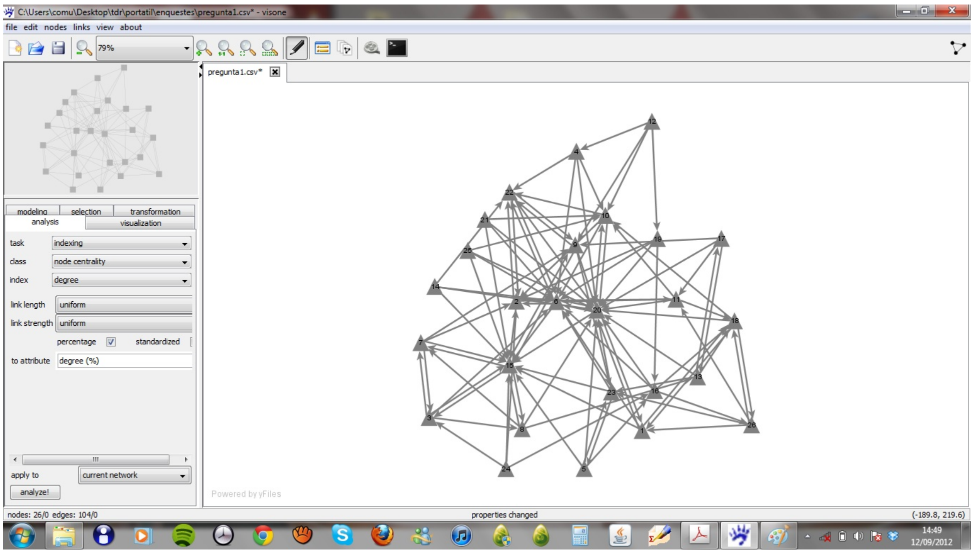


Figura 3.10: Graf creat amb les dades importades de les respostes a la pregunta 1.

Com podem veure en la imatge anterior, ja tenim un graf que representa tota la gent de la classe i les seves relacions, segons la primera pregunta, que deia:

Amb qui faries un treball de grup de català o castellà, en què la nota importa molt?

Però el graf obtingut resulta complicat i no es veu gaire clar. Per tal de solucionar això, canviarem manualment els colors i les grandàries dels vèrtexs, la mida i la posició de les etiquetes dels vèrtexs, i la forma, color i la mida de les arestes.

Per tal de canviar color, mida i forma dels vèrtexs, anem al submenú `Nodes` i cliquem `Select All`. Un cop seleccionats tornem a obrir el submenú `Nodes` i cliquem `Properties`, que obrirà una finestra amb les opcions: `General, Label, Attributes`. En el `General` podem canviar la forma, el color i mida dels vèrtexs. També podem canviar la ubicació introduint les coordenades que ens interessin.

De manera anàloga, si seleccionem la pestanya `Label`, podem canviar la forma, color, mida i ubicació de les etiquetes dels vèrtexs seleccionats.

La pestanya `Attributes` mostra les mesures, que es poden calcular sobre els vèrtexs, que calcularem posteriorment.

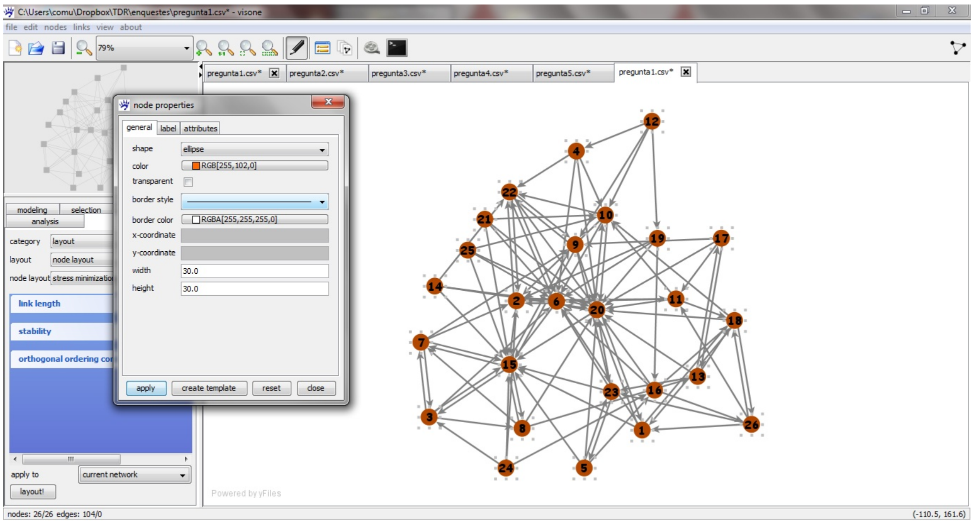


Figura 3.11: Propietats generals de tots els vèrtexs del graf social de la pregunta 1.

Per poder modificar les arestes, anem al submenú `Links` i cliquem `Select All`. Un cop seleccionades les arestes, tornem a obrir el submenú `Links` i cliquem `Properties`, que obrirà una finestra amb les opcions: `General, Label, Attributes`. En general podem canviar el color, gruix i estil de l'aresta, com es pot veure a la figura 3.12.

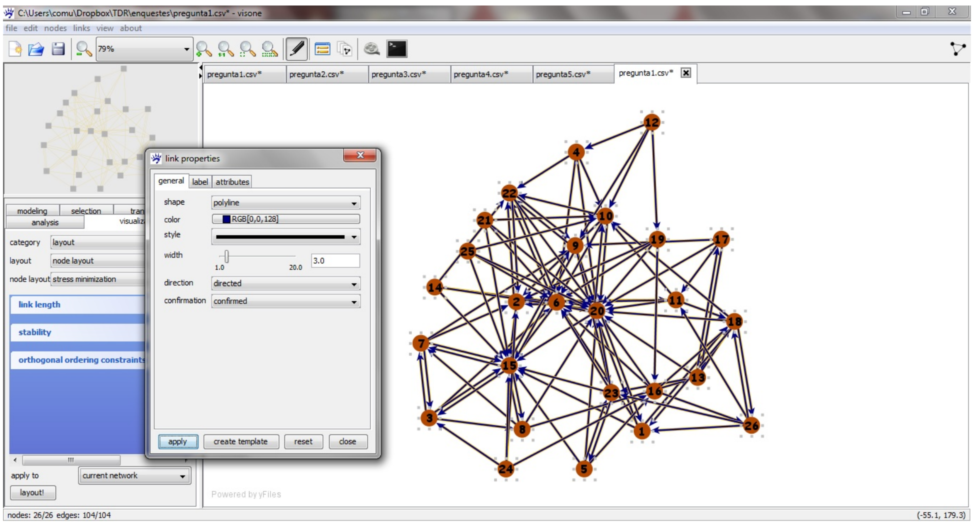


Figura 3.12: Propietats generals de les arestes del graf social.

Notem que amb aquests canvis, l'estructura general del graf no ha canviat, ja que no hem mogut cap vèrtex de lloc, però ja es veu més clar. Tot i això, encara no es veuen del tot bé les arestes, ja que moltes es superposen.

En la utilització del Visone per a l'anàlisi de xarxes socials, l'eina `Layout` és molt útil, especialment en un graf amb molts vèrtexs i arestes, ja que permet reorganitzar les estructures formades per tal de crear noves visualitzacions. Hi ha moltes visualitzacions possibles, però les més útils i que queden millor són les següents: `Circular Layout`, `Stress Minimization` i `Status Layout`, amb exemples a continuació.

El **Circular Layout** reorganitza l'estructura representant-la en forma circular,triant les distàncies entre els vèrtexs i el radi de la estructura circu-relacional.

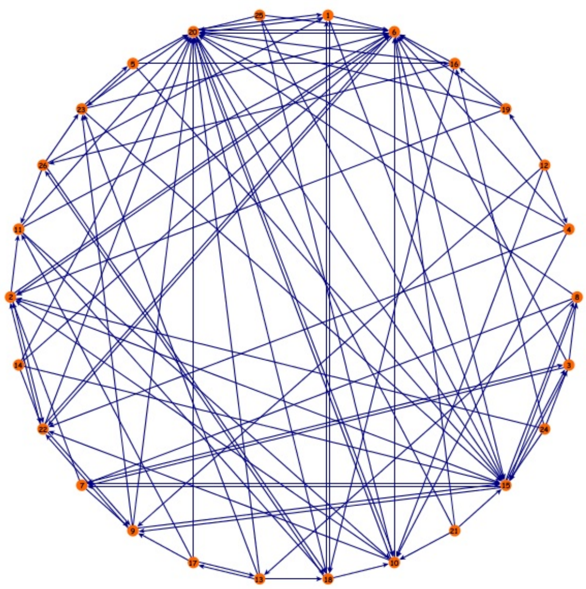


Figura 3.13: Graf amb estructura *circular*.

El **Status Layout** posa tots els vèrtexs alineats a la part inferior de la pantalla, i totes les arestes per sobre. Això permet veure molt ràpidament la quantitat d'arestes que van a cada vèrtex.

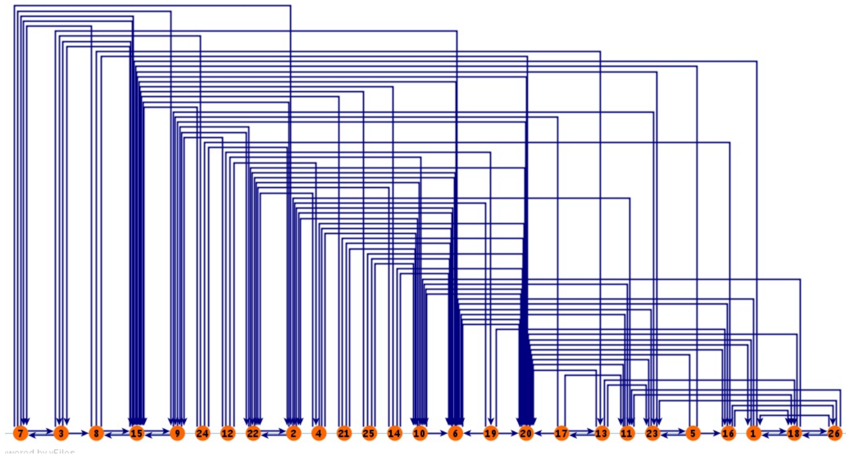


Figura 3.14: Graf amb estructura *status layout*.

El **Stress Minimization** crea el dibuix del graf intentant minimitzar la mida total de totes les arestes, evitant conglomerats de vèrtexs.

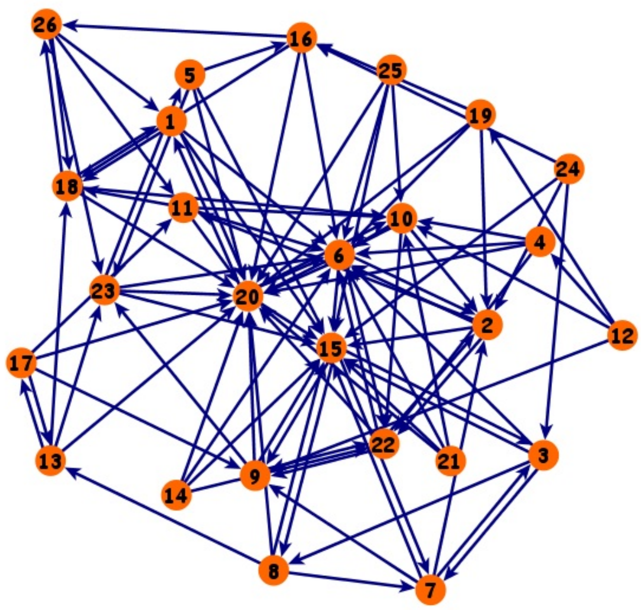


Figura 3.15: Graf amb estructura *stress minimization*.

Un cop utilitzada l'eina `Layout` el graf ja es veu molt més clarament i, per tant, ja no modifiquem més el seu aspecte. Ara ja podem passar a l'anàlisi.

Per analitzar una xarxa social general, ens fixarem en les mesures presentades al capítol anterior, especialment les que fan referència a la centralitat dels vèrtexs. Per tal de calcular la centralitat de cada vèrtex faré servir mesures diferents: centralitat de grau, centralitat intermèdia, centralitat de proximitat, excentricitat i page-rank (un patró que utilitza el google per les seves búsquedes, basat amb la importància dels vèrtexs relacionats). A continuació mostro la forma de calcular aquestes mesures amb el programa.

Ens situem en el menú de la part esquerra de la pantalla de Visone, en l'opció `Analisis`. Dins d'aquesta pestanya, a l'apartat `index` podem triar quines mesures volem aplicar.

El grau de cada vèrtex no és res més que la suma del grau d'entrada i del grau de sortida. En l'enquesta, no serveix de res calcular el grau de sortida ja que les preguntes el fixaven, i per tant calculo el grau d'entrada amb la instrucció `indegree`. Les altres mesures, calculades amb la instrucció corresponent, són: la centralitat intermèdia (`betweenness`), la centralitat de proximitat (`closeness`), l'excentricitat (`eccentricity`) i page-rank (`pagerank`). Els valors calculats es guarden com a atributs de cada vèrtex. A la imatge 3.16 mostro els valors calculats pels vèrtexs 15 i 4, amb característiques ben diferents.

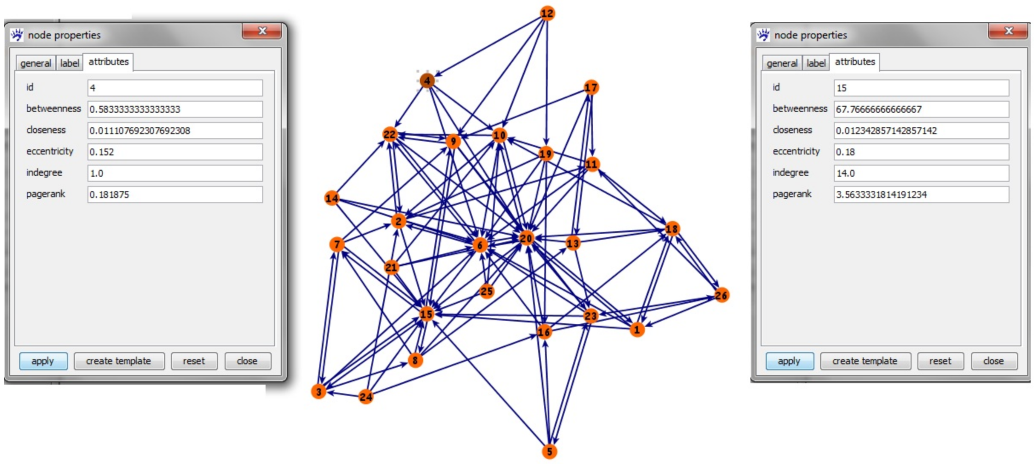


Figura 3.16: Atributs de dues de les persones enquestades, referents a la pregunta 1.

Un cop calculades totes les mesures, és bo treballar amb representacions gràfiques que permetin visualitzar la informació de tots els vèrtexs a la vegada. Notem que això té molt de sentit en els grafs socials ja que tenen molts vèrtexs i mirar els atributs de cadascun seria poc pràctic. Per fer-ho, anem a la opció `Visualization` i a l'apartat `category` posem `mapping`. Per exemplificar-ho, represento els atributs de maneres diverses, amb els menús `type` i `property`.

Per tal de representar el grau d'entrada, posaré `color` i `node color`, triaré dos colors ben diferents i clicaré `visualize`. El Visone ens pintarà els vèrtexs més d'un color o de l'altre, segons el seu grau d'entrada, com es mostra en la figura 3.17.

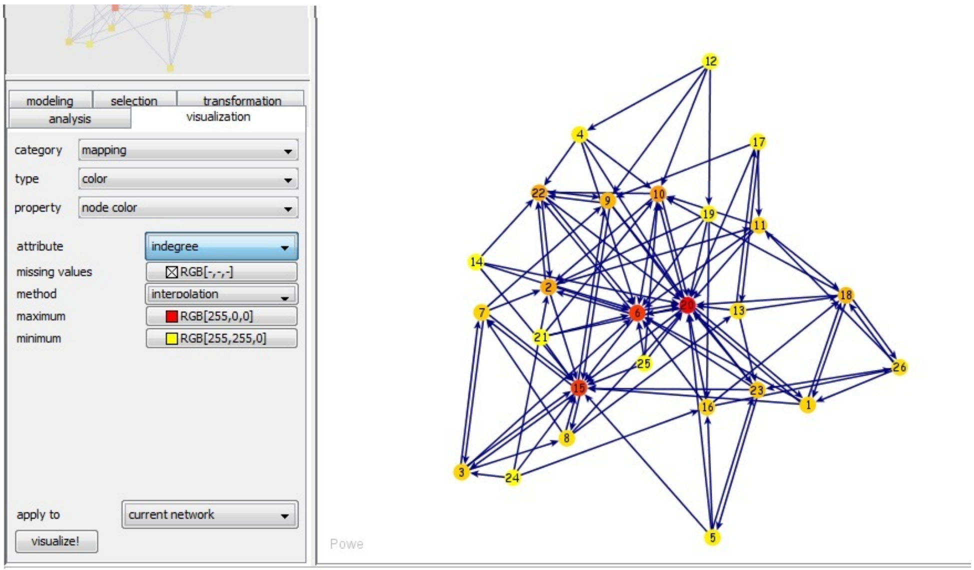


Figura 3.17: Graf amb el grau d'entrada representat amb colors.

La centralitat intermèdia la represento canviant la mida dels vèrtexs. Per tal de fer-ho, posaré l'opció `size` i `node area`. Tal com es mostra a la figura 3.18, notem que els vèrtex més grans són els que tenen una centralitat intermèdia superior. Ara bé aquests no tenen perquè coincidir amb els que tenen un grau d'entrada superior, destacats segons el color.

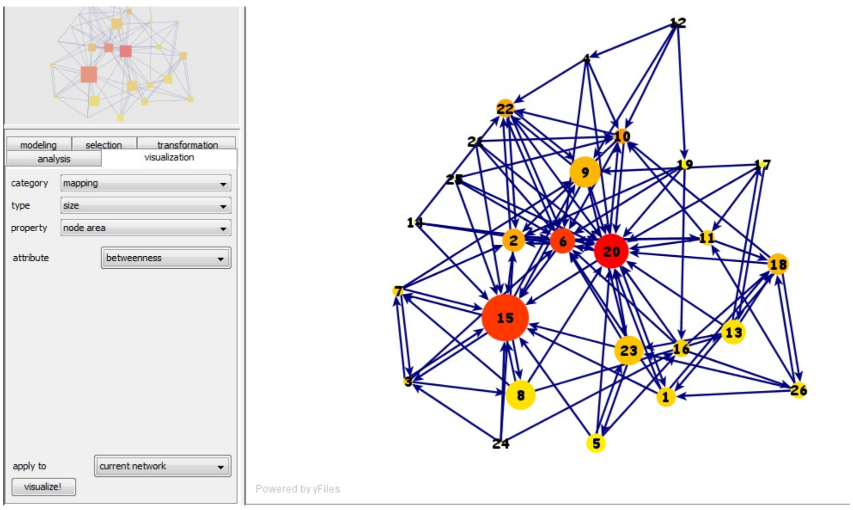


Figura 3.18: Graf amb la centralitat intermèdia representada amb la mida dels vèrtexs.

Apart de les mesures presentades a la secció anterior, n'existeixen altres, disponibles al Visone, que també són interessants. Per exemple en un graf dirigit té sentit mesurar el pagerank, com fa el Google. La idea és mesurar la importància d'un vèrtex tenint en compte la importància dels que li envien arestes i el nombre d'arestes que surten d'ells, en els que la seva importància queda repartida.

La mesura pagerank la represento canviant l'amplada dels vèrtexs. Per tal de fer-ho, posaré l'opció `size` i `node width`. A continuació es mostra el graf resultant a la figura 3.19.

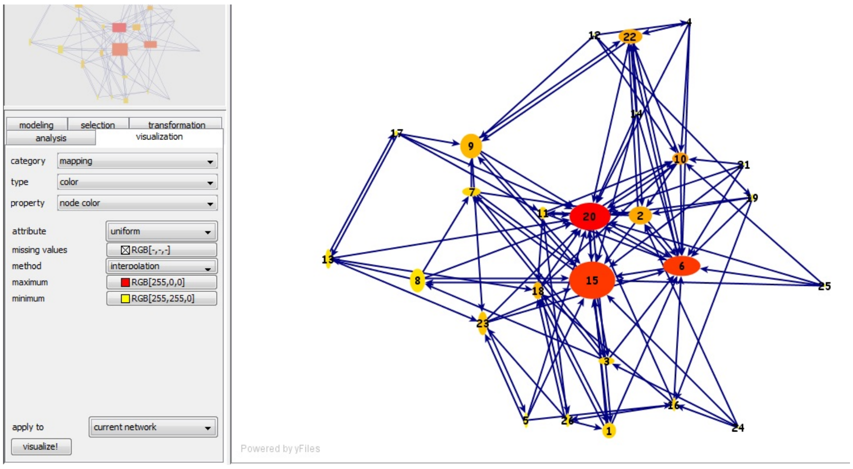


Figura 3.19: Graf amb el pagerank representat amb l'amplada dels vèrtexs.

A partir d'aquesta visualització conjunta de les 3 mesures (grau d'entrada, centralitat intermèdia i pagerank), puc extreure conclusions sobre la pregunta 1, que deia:

Amb qui faries un treball de grup de català o castellà, en què la nota importa molt?

Clarament, els vèrtexs que destaquen en aquesta pregunta són el 15, 20, 6, 9 i 22, és a dir, aquestes persones són les més valorades, en general, per fer un treball de grup de llengües. Ara bé, si mirem les mesures per separat hi ha variacions.

- Mirant el color veiem que la persona 20 ha estat la més triada, seguida de la 6 i la 15.
- En canvi, mirant l'alçada veiem que és la 15 la que enllaça amb les tries de més gent. La 20 també té una bona centralitat intermèdia, cosa d'esperar. També destaquen les persones 9, 8, i 23 que, encara que no fossin molt triades, enllacen amb les tries de molta gent.
- Mirant l'amplada, que representa el pagerank, veiem que destaquen de nou la persona 15, 20 i 6, que de fet també s'han triat entre elles.

De fet, només m'he fixat amb aquestes mesures perquè les altres no tenen gaire sentit en aquest graf social, ja que està format només per 26 persones i hi havia restriccions prèvies sobre el nombre d'arestes sortints de cada vèrtex. Així per exemple, la centralitat de proximitat és molt semblant a tots els vèrtexs perquè el radi és molt petit i tots estan fortament comunicats.

A la resta de preguntes transformaré el graf per a visualitzar també la centralitat de grau, la centralitat intermèdia i el pagerank, com en aquesta pregunta.

3.4.2 Anàlisi de la pregunta 2

De manera similar a l'apartat anterior, aplicant el potencial del programa Visone, obtinc la representació del graf associat a les respostes de la pregunta 2 per fer-ne l'anàlisi. La pregunta 2 deia:

Amb qui faries un treball de grup de ciències, en què la nota importa molt?

A continuació, la figura 3.20 mostra la taula d'Excel que recull les respostes. Notem que és difícil treure conclusions a partir de la taula. En canvi, a partir del graf és molt més fàcil.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Pr. 2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
1					1					1										1			1			
2						1					1				1							1				
3							1	1		1					1											
4											1				1					1			1			
5											1					1				1			1			
6					1				1	1										1						
7		1	1			1									1											
8							1						1							1			1			
9					1	1					1												1			
10		1				1														1		1				
11					1											1				1			1			
12				1					1	1									1							
13											1						1			1			1			
14	1																	1			1					1
15			1				1	1	1																	
16					1						1									1			1			
17									1		1		1							1						
18										1										1			1			1
19						1				1						1							1			
20						1					1						1						1			
21						1			1	1										1						
22		1				1					1												1			
23					1					1	1									1						
24					1				1											1		1				
25					1	1											1			1						
26	1										1							1					1			

Figura 3.20: Buidatge de la pregunta 2 de l'enquesta en una taula d'Excel.

Introduint directament les dades al Visone el graf obtingut és el de la figura 3.21, a partir del qual se'n pot extreure poca informació. He inclòs aquesta figura per tal de poder valorar la diferència amb la figura 3.22 que mostra ja el graf amb una visualització d'acord amb les mesures de centralitat que volem analitzar.

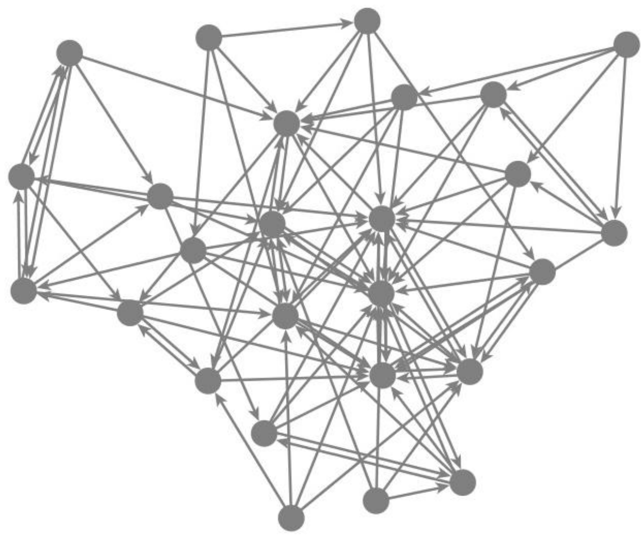


Figura 3.21: Graf social de la pregunta 2, creat directament a partir de les dades.

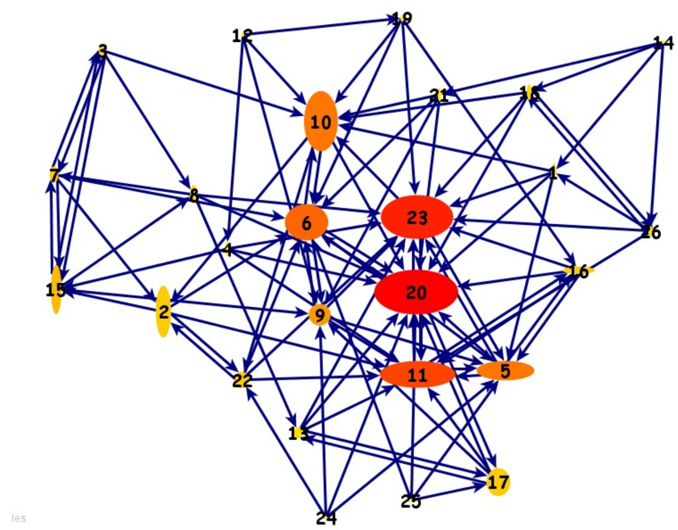


Figura 3.22: Graf social de la pregunta 2, amb interpretació gràfica de les mesures.

Observem que tot i tenir la mateixa forma, ja que no canvio el layout, el fet de marcar amb color, alçada i amplada les mesures analitzades permet fer l'anàlisi visual del graf social molt millor.

Així, a la vista de la figura 3.22, podem extreure les conclusions següents.

- Amb la centralitat de grau pintada de color, veiem que les persones més triades per a un treball de ciències són les 23, 20 i 11.
- Mirant la centralitat intermèdia com a alçada de cada vèrtex veiem que les persones 20 i 23 segueixen destacant, però són lleugerament superades per les persones 2 i 15, tot i que tenien una centralitat de grau baixa. Això vol dir que, tot i que la 20 i 23 són les més triades, les 2 i 15 es troben enmig de més tries i encaixarien en més grups.
- El vèrtex 10 és amb diferència el que té la centralitat intermèdia més alta, és a dir està enmig de les tries i seria ben valorat en la majoria de grups, tot i que no sigui tant triat.
- Sobre el pagerank, que es visualitza amb l'amplada, veiem que el 23, 20 i 11 tornen a destacar. També destaca el 5, ja que està molt ben connectat perquè ha estat triat pels tres vèrtexs que destaquen en global. Aquest vèrtex ens dóna un bon exemple del que mesura el pagerank, ja que els vèrtexs més triats, també s'han triat entre ells.

3.4.3 Anàlisi de la pregunta 3

La pregunta 3 deia:

Amb qui aniries d'equip per fer un partit de futbol?

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Pr. 3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1				1	1															1				1		
2															1					1				1		1
3					1			1				1			1											
4	1									1		1												1		
5				1								1								1				1		
6					1				1	1										1						
7				1											1						1			1		
8				1	1															1				1		
9	1				1						1												1			
10	1			1																1				1		
11	1			1	1																			1		
12				1						1										1				1		
13				1																	1		1			
14	1			1	1																1					
15	1		1					1																1		
16				1																1	1			1		
17					1						1		1											1		
18	1											1									1					1
19				1						1											1			1		
20	1			1	1																			1		
21	1			1																1				1		
22								1	1											1				1		
23	1				1						1													1		
24				1						1									1		1					
25	1				1					1										1						
26	1										1							1					1			

Figura 3.23: Buidatge de la pregunta 3 de l'enquesta en una taula d'Excel.

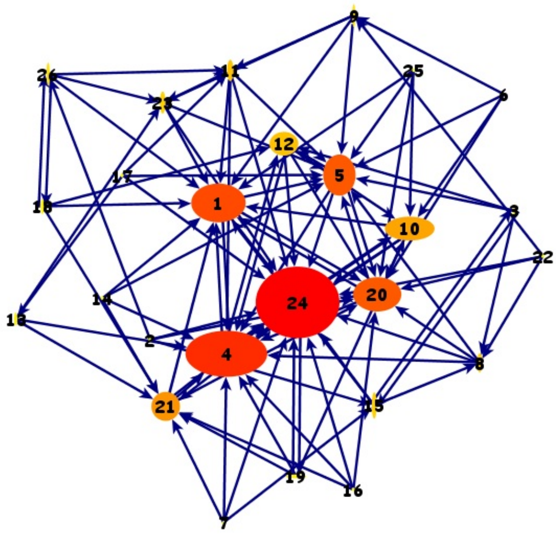


Figura 3.24: Graf social de la pregunta 3, amb interpretació gràfica de les mesures.

La pregunta 3 té la particularitat que les respostes estan fortament correlacionades amb el sexe. A la vista del graf és raonable deduir que els vèrtexs 24, 4, 1, 20, 5, 10, 12, 21 són nois, ja que són els vèrtexs destacats en una o altra mesura. Segurament la majoria de noies estarien en els vèrtexs de color més clar i més petits.

El graf destaca els que deuen ser millor jugadors de futbol segons el color (24, 4, 1, 5 i 20), que es correspon amb ser triat per moltes persones, siguin nois o noies. Notem que de fet hi ha vèrtexs que dirigeixen totes les seves tries a aquest conjunt, per exemple el vèrtex 14, que tant pot ser noi com noia.

Amb l'alçada es reflexa la centralitat intermèdia, que correspondria a ser capaç de jugar bé en equip. En canvi l'amplada dóna informació sobre si s'ha estat triat pels bons jugadors. Els vèrtexs 5 i 10 exemplifiquen les dues mesures. El 5 correspon a una persona que encaixaria bé en la majoria d'equips, tot i que no ha estat triada pels bons jugadors. En canvi el 10 ha estat triat per menys persones però els que l'han triat són considerats bons jugadors.

3.4.4 Anàlisi de la pregunta 4

La pregunta 4 deia:

Gent amb qui pots parlar de tot, amb qui et portes bé.

En aquesta pregunta el tipus de connexió entre els vèrtexs varia respecte les preguntes anteriors, ja que correspon a una relació d'amistat entre les persones. A diferència de les preguntes anteriors, en aquesta es demanava un màxim de 5 relacions possibles, enlloc d'obligar a citar-ne obligatòriament 4.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Pr. 4	1	2	3	4	5	6	7	8	9	10	11	12	13
1						1							
2			1										
3							1	1					
4							1		1	1			
5											1		1
6	1	1											
7			1	1				1					
8							1						
9				1									
10				1					1		1		
11			1		1								
12			1	1			1						
13											1		
14	1	1				1							1
15		1	1				1						
16		1									1		1
17					1				1				1
18												1	
19				1			1		1	1			
20	1				1						1		1
21								1				1	
22		1				1			1				
23					1						1		
24				1					1			1	
25	1				1								1
26	1					1							

A	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Pr. 4	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1							1		1			
2		1							1				
3	1		1										
4	1										1		
5			1	1			1						
6									1				
7	1		1		1								
8							1						
9													
10						1				1			
11				1			1						
12		1				1							
13				1			1			1			
14													
15													
16				1						1			
17							1		1				
18											1		1
19											1		
20				1									
21	1									1	1		
22	1												
23			1		1								
24						1		1					
25				1			1						
26					1								

Figura 3.25: Buidatge de la pregunta 4 de l'enquesta en una taula d'Excel.

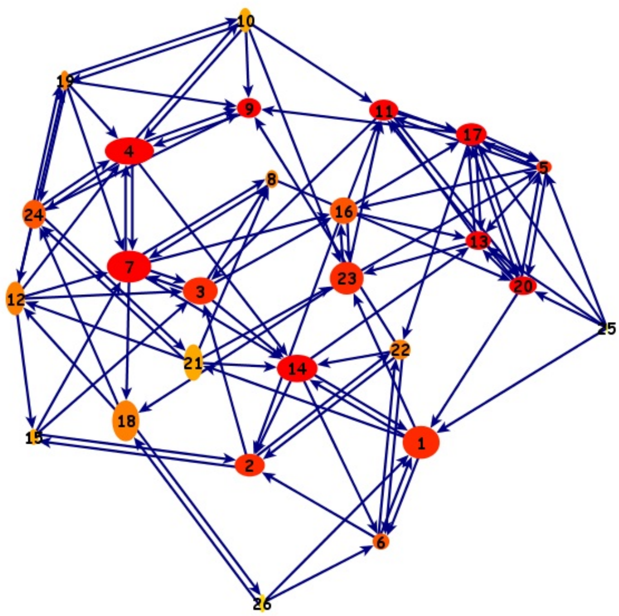


Figura 3.26: Graf social de la pregunta 4, amb interpretació gràfica de les mesures.

El graf de la figura 3.26 mostra ja el resultat del càlcul de les mesures interpretat gràficament i permet fer l'anàlisi de manera visual. Observem que gairebé tots els vèrtexs tenen bona centralitat, la qual cosa ens diu que és un grup de gent sociable, on gairebé tothom es relaciona amb la resta i té bastants amics. Pel que fa a les altres mesures, centralitat intermèdia i pagerank, els resultats de la majoria també són similars.

En sentit contrari destaca, però, que hi ha algunes persones amb més dificultats de relació o que es restringeixen a un cercle més reduït, com serien els vèrtexs 25, i 26.

Al graf també hi podem veure algun subgrup més relacionat entre ells, com per exemple el format pels vèrtexs 20, 13, 5, 17 i 11. Això entraria dins de la detecció de comunitats, una de les línies de recerca més dinàmiques actualment a nivell de grans xarxes socials.

3.4.5 Anàlisi de la pregunta 5

Finalment, la pregunta 5 era la més personal, per analitzar més a fons la força de les relacions d'amistat, ja que demanava:

Amb qui compartiries un secret o una cosa important per a tu?

En aquest cas es demanava un màxim de tres persones ja que s'entén que estem parlant de relacions personals més profundes que les de la pregunta anterior, en que el màxim era 5. Això farà que surti un graf amb menys arestes que el de la pregunta quatre, ja que ens estem restringint a les relacions més fortes. Aquest graf doncs, ha de ser un subgraf de l'anterior i el seu aspecte ha de ser clarament diferent dels anteriors.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
Pr. 5	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
1														1												
2																						1				
3							1	1							1											
4							1		1	1																
5																	1			1			1			
6		1																				1				
7			1					1							1											
8							1								1											
9				1																						
10				1					1										1							
11					1											1							1			
12			1	1											1											
13																	1									
14	1	1										1														
15			1			1																				
16											1												1			
17					1								1							1						
18																										1
19				1					1	1																
20					1								1				1									
21														1										1		
22		1				1																				
23											1		1			1										
24				1																	1					
25													1				1			1						
26																		1								

Figura 3.27: Buidatge de la pregunta 5 de l'enquesta en una taula d'Excel.

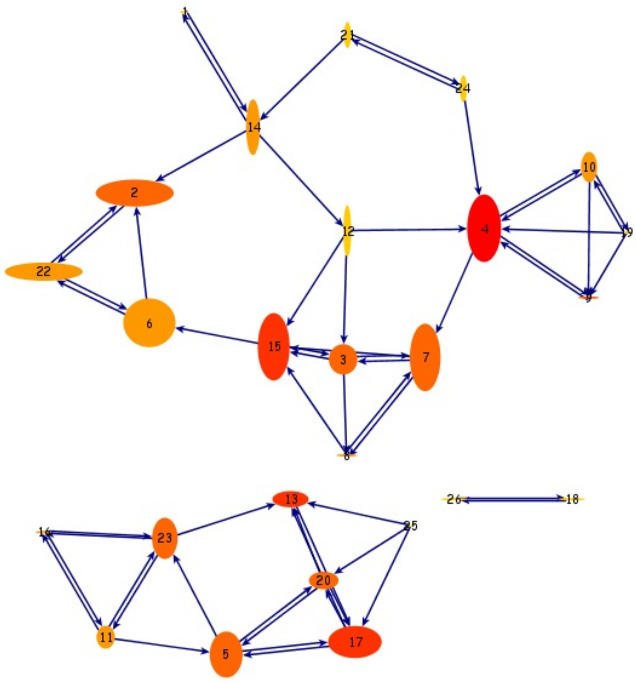


Figura 3.28: Graf social de la pregunta 5, amb interpretació gràfica de les mesures.

A continuació, mirant el graf de la figura 3.28 faig llista d'alguns comentaris.

- Reparteix clarament els vèrtexs en grups, i queden tres comunitats no connexes. Per exemple, queda ben diferenciat el subgrup format per les persones 16, 11, 23, 5, 13, 20, 17 i 25.
- També es dóna el cas d'una parella de persones connectades entre elles que queda desconnectada de la resta, 26 i 18.
- En els dos grups grans hi ha persones amb valors de centralitat de grau, centralitat intermèdia i pagerank elevats.
- No hi ha cap vèrtex que domini en tots els aspectes; comparem per exemple el vèrtex 22 i el vèrtex 4.
- La majoria de relacions són recíproques, la qual cosa té molt sentit ja que la pregunta es refereix a un gran nivell de confiança.

3.4.6 Anàlisi global

A partir de les mesures que hem visualitzat en els grafs de les respostes al qüestionari, podem fer una valoració dels resultats. En certa manera, a través de la teoria de grafs, estem fent una radiografia del grup de persones de la xarxa social analitzada. Si aquesta xarxa correspon a un grup classe, aquesta radiografia pot ser d'utilitat per al professorat, i fins i tot pot ser una eina per tutoria. Tal com diu la dita popular de l'inici del capítol *una imatge val més que mil paraules.*

A partir de la visualització conjunta de les tres mesures esmentades a la pregunta 1 podem deduir que les persones més bones fent treballs de català o castellà són el 15, 20, 6, 9 i 22. La pregunta 2, en canvi, fa referència a la gent bona en treballs de ciències que són la 20, 23, 11, 10, 6 i 5. En conjunt, la percepció és que aquestes persones són bones en l'assignatura corresponent, o són treballadores i saben treballar en grup. Notem que només dues persones destaquen tant per ciències com per lletres (20 i 6). Considero que és raonable, ja que la majoria de gent tendeix a ser més bo en una o altra cosa, i no pas en les dues alhora. La poca coincidència deu ser deguda a que en aquest grup de batxillerat justament hi ha persones d'una modalitat de ciències i una de lletres. També pot ser que hagin destacat en les dues matèries perquè són bones treballant en grup.

Contrastant amb la meva pròpia percepció de la realitat, ja que conec de quines persones estem parlant, puc confirmar alguna de les hipòtesis anteriors. Les dues persones que destacaven són treballadores i treuen bons resultats de les dues branques. Una altra cosa que veig és que en conjunt totes les persones que destaquen menys una són del Batxillerat tecnològic.

Comparant els resultats de la pregunta 3 amb la realitat, comprovo que efectivament totes les persones destacades, 24, 4, 1, 20, 5, 10, 12, 21 són nois. Però també m'adono que la gent tria més aviat a gent amb qui acostuma a jugar, que no pas altres tot i saber que juguen millor. Les excepcions són el 24 i 4, que els tria molta gent ja que són, amb diferència, els més bons.

La pregunta 4 i la 5 estan molt relacionades, amb la diferència que la 5 aprofundeix molt més en el tema confiança, mentre que la 4 només es fixa en l'amistat. És per això que en la 5 es formen més subgrups tancats, que es veuen com a tres components connexes. De fet, cal tenir en compte que la classe estava integrada en aquell moment per alumnes de 2 modalitats ben diferents (tecnològic i econòmic), la qual cosa pot explicar el perquè d'aquests subgrups (ja que hi havien algunes assignatures que les fèiem per separat). Per altra banda, no hi ha cap vèrtex que destaqui descaradament respecte els altres, cosa que correspon a la realitat ja que no hi ha cap persona que sigui la més popular de la classe.

Capítol 4

Amb la lupa sobre xarxes socials en línia

ANÀLISI DE XARXES SOCIALS EN LÍNIA

Digues-me amb qui vas i et diré qui ets
(Popular)

En aquest capítol l'objectiu és analitzar un fragment d'una xarxa social en línia real, utilitzant el graf social corresponent, per demostrar la poca privacitat que tenen.

Les xarxes socials online són tremendament grans i s'estudien reduint-se a una part. En aquest capítol estudio dos fragments reals de dues xarxes socials online.

Per tal d'aconseguir informació de les relacions d'una xarxa social en línia podria anar visitant amb un navegador web tots els perfils dels usuaris d'aquesta xarxa. Però aquest procés és molt lent i si vull visitar molts usuaris resulta impossible fer-ho amb aquest mètode. Per tant cal un programa que permeti recol·lectar aquestes dades de forma automàtica. Aquests programes s'anomenen *spiders* (aranyes), o *crawlers* (rastrejadors). Les aranyes poden fer aquesta recol·lecció d'informació de maneres diferents. Comencen en una o vàries pàgines o usuaris on busquen referències a altres pàgines o usuaris, que seguidament analitzaran buscant noves referències. Aquest procés pot anar-se repetint infinitament i, per tant, en algun moment s'ha de parar, i hi ha diferents algoritmes per fer-ho.

Les *aranyes* són molt complicades de programar i necessiten un ordinador potent per funcionar, per tant, he hagut de recórrer a llocs especialitzats, que em poguessin proporcionar les dades d'algun fragment d'una xarxa social en línia. Gràcies al programa ARGÓ, m'he posat en contacte amb en Jordi Herrera i la Cristina Pérez, del Departament d'Enginyeria de la Informació i les Comunicacions de la Universitat Autònoma de Barcelona, que m'han proporcionat dades de dues xarxes diferents: Twitter i Flikr.

Aquestes dades han estat obtingudes per una aranya, començant pel perfil d'un usuari concret de cada xarxa, ubicat a Catalunya. Aquesta aranya ha utilitzat l'algoritme FIFO (First In First Out) que funciona de la manera següent. Donada una pàgina inicial (representada com a vèrtex), busca totes les referències a altres pàgines (representades com a vèrtexs veïns) i aquestes noves pàgines es posen a la cua de pàgines per explorar. Llavors explora els vèrtexs d'aquesta cua, començant pel primer, és a dir, la primera referència trobada a la pàgina inicial. Com que les noves referències que troba són afegides a la cua, no seran explorades fins que tots els veïns del primer vèrtex hagin estat completament explorats. Per tant, hi ha dos tipus de vèrtexs: els **totalment explorats** i els **descoberts** (veïns dels explorats) que encara no han estat explorats per l'aranya.

Per tal d'analitzar les dades, utilitzaré les mateixes mesures aplicades en el capítol anterior. A continuació, en el primer apartat transformaré les dades referents al Twitter en un graf, i les analitzaré. En el segon apartat representaré i comentaré les dades referents al Flickr.

4.1 Espiant a la xarxa Twitter

En aquest apartat, utilitzarem la informació extreta de la xarxa Twitter a partir d'un usuari, explorant tots els usuaris amb qui es relaciona, de manera que s'arriba fins a un total de 106 usuaris totalment explorats. A partir d'aquests 106 usuaris tenim 223329 nodes descoberts i 245307 relacions descobertes. Ara bé, aquests últims nodes no estan del tot explorats, i només estan vinculats als usuaris anteriors per poques arestes. En aquestes condicions, és habitual eliminar aquests vèrtexs i les arestes corresponents per no distorsionar l'anàlisi i per simplificar-lo. Així doncs el graf conté 1 usuari inicial, 105 dels usuaris que es relacionen directament amb ell i les relacions entre ells, que fan un total de 377 arestes.

Encara que les relacions de Twitter són dirigides, en aquest cas ens limitarem a l'estudi del graf subjacent, sense tenir en compte la direcció.

El fitxer de dades inicial provinent de l'exploració a través de la xarxa pels *crawlers* té extensió `.gml`. Introduint aquest fitxer en el Visone obtinc el graf de la figura 4.1.

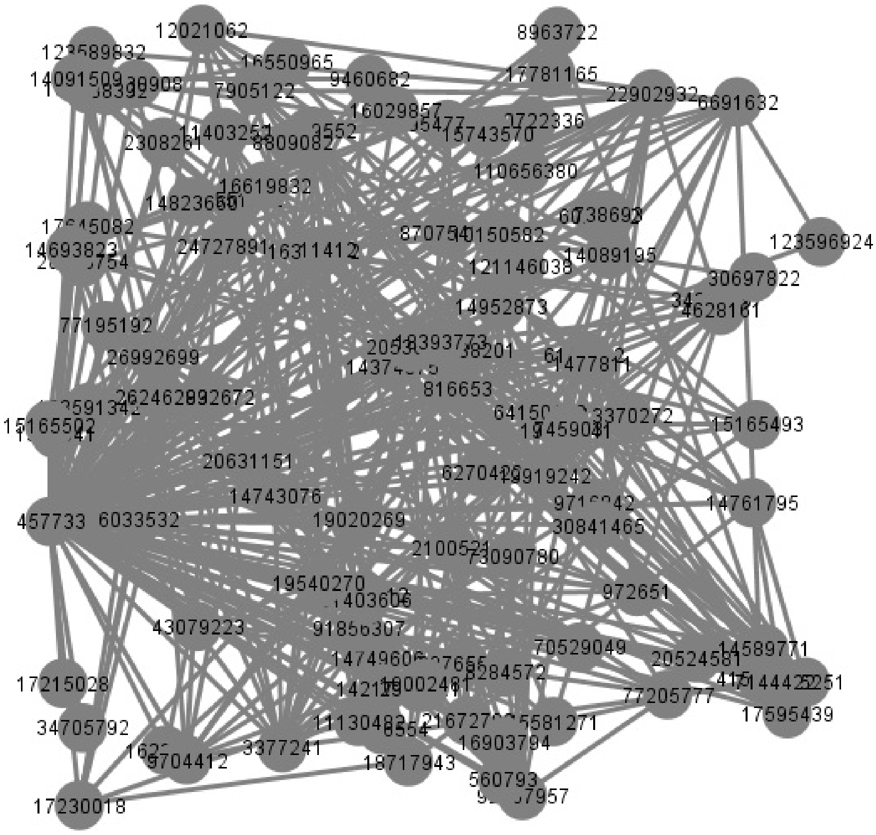


Figura 4.1: Visualització `random` del graf de twitter explorat.

Com en el cas del capítol anterior, aquest graf dóna poca informació a primer cop d'ull. És gràcies a la teoria de grafs, que a través del càlcul de les mesures de centralitat, es poden aplicar algoritmes per representar-ho millor. Per millorar la visualització hem explorat diferents tipus de layout, que es mostren a la figura 4.2.

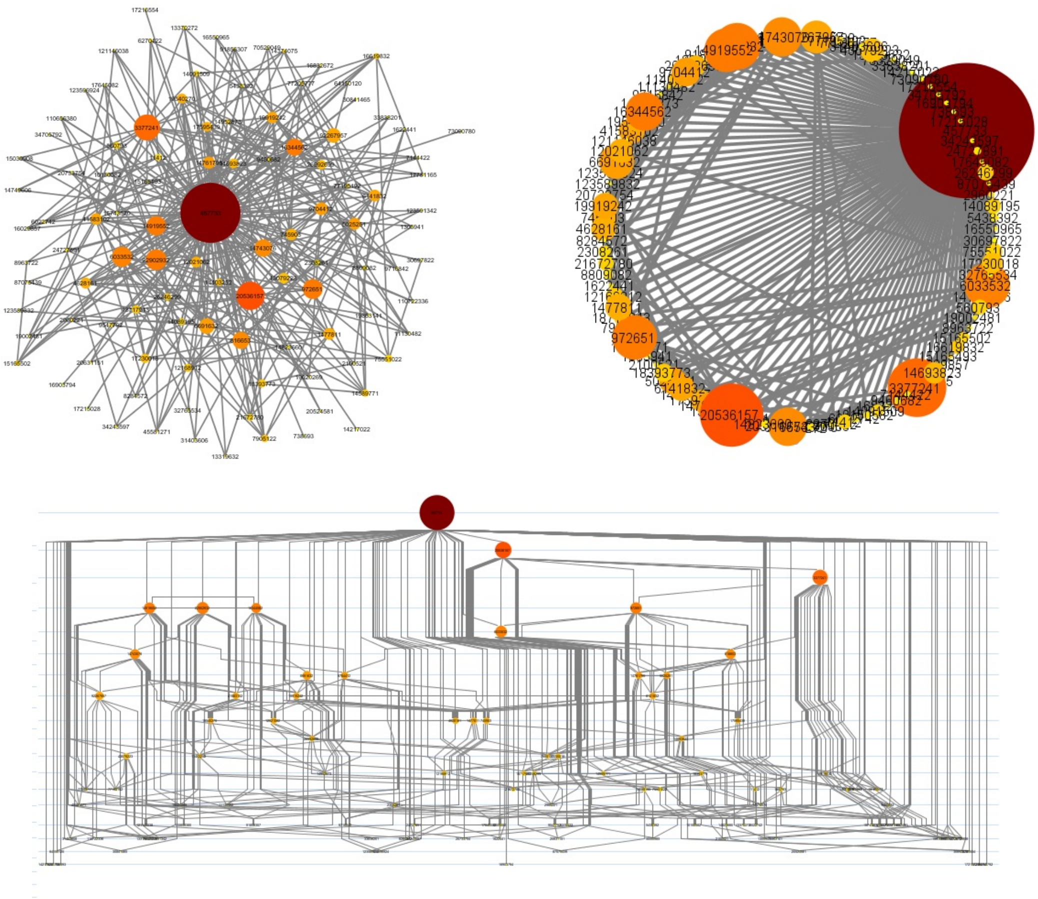


Figura 4.2: Mostres de visualitzacions del graf de twitter explorat: `stress minimization`, `centrality layout` i `status layout` respectivament.

Notem que cadascuna reflexa diferents aspectes, i ja se'n pot extreure alguna informació més útil. A totes les visualitzacions he representat amb el color del vèrtex el seu grau i amb la mida del vèrtex la seva centralitat intermèdia. En tots els casos queda reflectit el vèrtex corresponent a l'usuari des del que es va iniciar l'exploració com el vèrtex amb més grau (color més fosc) i més centralitat intermèdia (grandària). Això és lògic perquè està directament comunicat amb tots els altres. D'entre els relacionats amb ell en destaquen també uns quants, segons mida i color, en funció del seu grau i del nombre de camins en els que intervenen.

El cas de `stress minimization` reflexa la centralitat del node inicial posant-lo al mig, i tota la resta al seu voltant. D'entre aquests, els vèrtexs amb grau inferior i més mal comunicats queden a la perifèria, en els punts més allunyats del centre.

La visualització que ens dóna el `centrality layout` col·loca tots els vèrtexs en una circumferència, de manera que totes les relacions passen dins del cercle i es representen com a cordes. Els vèrtexs amb més grau queden repartits per la circumferència, evitant així que quedin acumulats. Dit d'una altra manera, reparteix les cordes. L'únic punt d'acumulació de cordes és el punt inicial, cosa inevitable. Els vèrtexs amb grau petit connectats entre si queden a prop per tal d'evitar que les cordes que generen es vegin massa. Per tant, les cordes que destaquen perquè travessen el cercle son les que connecten vèrtexs amb grau elevat.

Finalment, el `status layout` ens mostra una ordenació dels vèrtexs de més a menys grau, per capes horitzontals (de dalt a baix). Als extrems dret i esquerre col·loca a baix de tot els vèrtexs que només estan connectats amb l'inicial, per tal d'evitar encreuaments d'arestes. L'estructura recorda un arbre, ja que posa cada vèrtex important com a punt de ramificació de l'arbre, és a dir, amb tots els vèrtexs que influeix a sota seu. En aquesta visualització es veuen clarament els fluxes d'influència.

4.2 Espiant a la xarxa Flickr

En aquest apartat utilitzarem també les dades cedides per J. Herrera i C. Pérez sobre un fragment de la xarxa Flickr. A diferència de l'anterior, en aquest cas analitzarem les dades d'un graf molt més gran ja que no s'han suprimit els vèrtexs terminals. En aquest graf hi ha un total de 25255 vèrtexs descoberts, dels quals 102 han estat totalment explorats amb el web-crawling. El nombre de relacions descobertes és 34085.

Introduint el fitxer de dades .gml directament al Visone, obtinc un graf molt estrany, figura 4.3, en el qual tots els 25255 nodes estan comprimits de manera que no es veu res. Per tant, cal millorar la visualització.

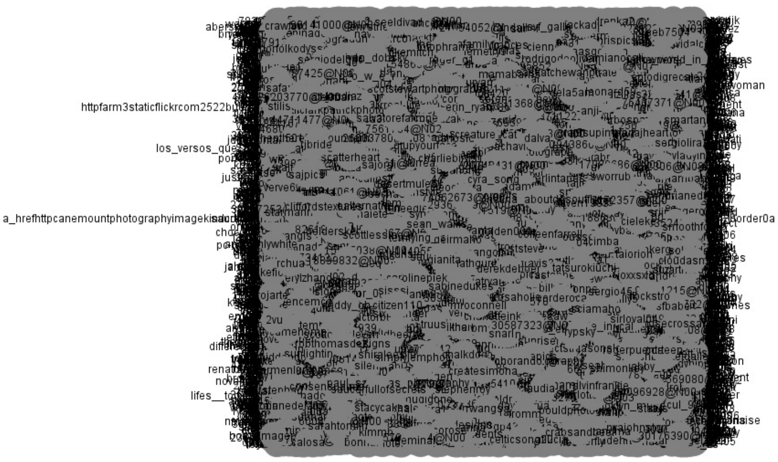


Figura 4.3: Graf obtingut directament de l'importació de dades del Flickr.

En primer lloc, he utilitzat el `spring embedder` per tal de representar el graf sense que quedin tots els vèrtexs tan acumulats. En segon lloc, he calculat els graus de tots els vèrtexs. Després els he pintat segons el seu grau, posant vermell fort els de més grau i blau cel els de menys, tal com es pot veure a la imatge 4.4.

Mirant la figura 4.4 veiem tots els vèrtexs pintats de blau cel, amb tonalitats pràcticament iguals, i cap de color vermell fort. Això és perquè només hi ha 102 vèrtexs totalment explorats, que són els que tenen grau més elevat, i queden tapats entre totes les arestes.

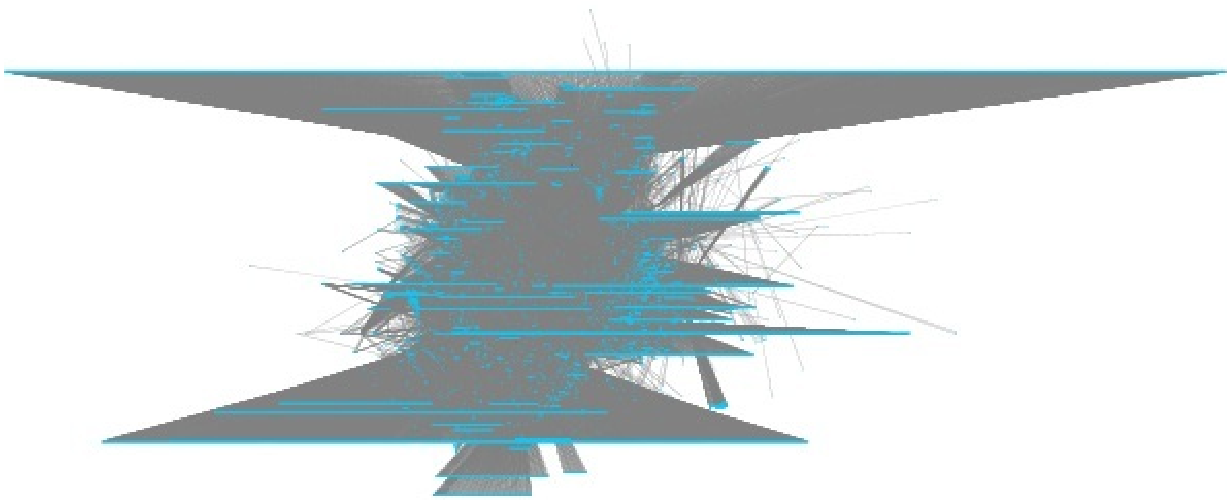


Figura 4.4: Graf obtingut directament de l'importació de dades del Flickr.

Com que és un graf tan gran, al visualitzar-lo tot complet, no s'aprecien els vèrtexs ni les seves arestes. És per això que incloc uns quants zooms fets a diferents parts del graf a les figures 4.5 i 4.6.

A la primera figura es veu un fragment aleatori del graf, amb vèrtexs amb grau molt més elevat que d'altres, i tot d'arestes. És una zona amb moltes arestes, és a dir molt densa.

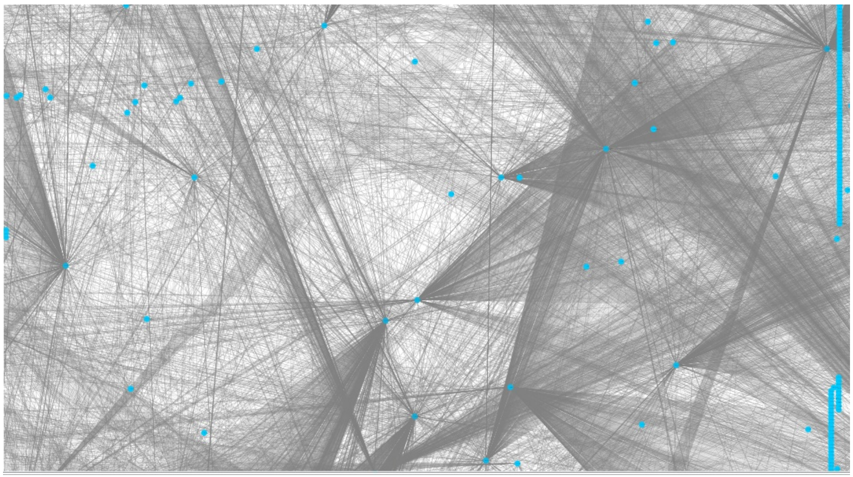


Figura 4.5: Zoom d'una part aleatòria del graf.

A la segona figura, es mostren dues imatges amb característiques diferents. A l'esquerra es veu una zona poc densa, en la que destaquen alguns vèrtexs. Notem que la quantitat d'arestes que surten d'un o altre vèrtex és ben diferent, és a dir, hi ha usuaris molt més relacionats que altres, dins d'aquest fragment de xarxa. També es veu que alguns vèrtexs estan dibuixats molt propers i amb poques relacions cap a altres. A la imatge de la dreta es veu un dels 102 vèrtexs totalment explorats. Aquest vèrtex està marcat amb vermell fort i es veu clarament que el seu grau és molt superior a tota la resta, ja que casi totes les arestes de la zona són incidents a ell.

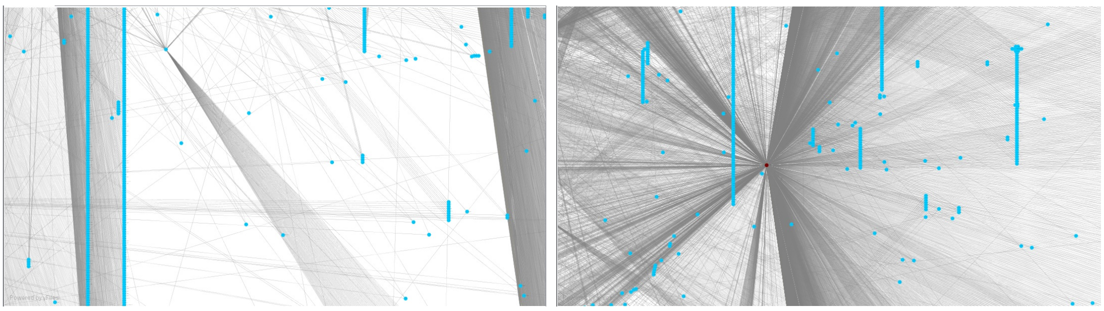


Figura 4.6: Zoom d'una part poc densa del graf i zoom d'un dels vèrtexs totalment explorat, respectivament.

Com que el nombre de vèrtexs i arestes és tant elevat, no m'ha estat possible calcular gaires paràmetres ja que necessitaria un ordinador molt més potent. Per tant, només he calculat el grau d'entrada i l'he pintat de color. Cal remarcar també que no és possible utilitzar qualsevol layout, pel mateix problema. Per exemple em resulta impossible utilitzar `stress minimization`, ja que l'ordinador es penja.

Capítol 5

Conclusions

Un cop acabat el treball, puc afirmar que els objectius principals plantejats s'han assolit. He après un tema nou de matemàtiques, la teoria de grafs, i l'he aplicat a les xarxes socials. Això complia els dos primers objectius: demostrar que les matemàtiques són útils i aplicables a la vida real, i estudiar les xarxes socials aplicant una part de les matemàtiques.

A través de les respostes obtingudes en les enquestes i de l'anàlisi conseqüent, com també de l'anàlisi dels fragments de les dues xarxes socials en línia que he estudiat (Twitter i Flickr), he comprovat la gran quantitat d'informació que se'n pot extreure. De la mateixa manera que jo, amb 16 anys, he pogut obtenir aquesta informació, qualsevol altra persona podria obtenir-la. Això demostra la poca privadesa d'aquestes xarxes socials. Cal remarcar que encara que algunes xarxes tinguin una privacitat més alta, com ara el Facebook, tenen també un cert risc ja que els propietaris d'aquestes si que accedeixen a tota la informació. Això queda confirmat, per exemple, pel fet de que els anuncis que un usuari pot veure són personalitzats als seus gustos.

De fet, posteriorment a l'entrega d'aquest treball (octubre 2012), Facebook va fer públics certs canvis en les seves polítiques de privacitat que van confirmar les meves sospites.

És increïble el nombre i la variació de preguntes que m'he trobat a mesura que anava elaborant el treball. Moltes d'aquestes semblaven interessants, però aquest treball no podia abarcar-ho tot, de manera que vaig haver d'escollir.

Per exemple, un dels objectius que m'havia plantejat inicialment era crear un programa per dibuixar un graf a partir de la llista dels seus elements. Ara bé, en descobrir l'existència de programari ja creat no comercial i de lliure accés, vaig pensar que seria més útil aprendre a utilitzar aquest programari i fer-lo servir per analitzar dades. Una altra cosa que em vaig plantejar era crear una aplicació del Facebook que creés un graf amb les teves amistats i l'analitzés. No obstant, era molt complicat de programar atès que es necessitaven molts permisos del Facebook.

També, arran d'una conferència de J. Domingo-Ferrer, em va interessar el problema de l'anonimització. És habitual rebre publicitat relacionada amb les consultes fetes a Google. Això vol dir que a través de Google fan un perfil de la gent per fer pubicitat dirigida. Té sentit aleshores, establir protocols per redirigir consultes per tal d'emmascarar el propi perfil. Em vaig plantejar com es podria traslladar això al Facebook o a una altra xarxa social. No era gens fàcil ja que es tracta de falsejar la informació de cara als "espies"però no de cara als amics, però podria ser molt interessant trobar algun patró a seguir.

En general crec que he après moltes coses fent aquest treball, tant de contingut com a nivell d'eines informàtiques i passos per fer un treball. Això m'ha portat molta feina i temps, i he hagut de fer un esforç suplementari però, mirat en global, ha valgut la pena.

Com a idea final vull remarcar de nou que queda demostrat la poca privadesa de les xarxes socials, i m'agradaria aconseguir l'objectiu de conscienciar i convèncer la gent del meu voltant.

Epíleg

Dos mesos després d'haver entregat aquest treball, el Facebook va fer públic un canvi de normativa que li permet fer un cert ús comercial de la informació dels usuaris. De fet, era un dels riscos dels que alertava en aquest treball.

Fonts d'informació

Llibres i articles consultats

1. Aggarwal, C. C. *Social Netxork Data Analytics*, Editorial Springer, 2011.
2. Caballero, R. et al. *Matemàtica discreta para informáticos*, Editorial Pearson, 2007.
3. Comellas, F., Fàbrega,J., Sànchez, A., i Serra, O. *Matemàtica discreta*, Edicions UPC, 1994.
4. García, F. *Matemática discreta*, Editorial Thomson, 2005.
5. Gimbert, J. *Les matemàtiques de Google: l'algoritme PageRank*, Butlletí Societat Catalana de Matemàtiques, pàg. 29-55, 2011.
6. Lamport, L. *Latex- A Document Preparation System*, Editorial Addison-Wesley, 1985.
7. Pérez, C. *La privacitat de les connexions dels usuaris d'una xarxa social*, Projecte de final de carrera d'Enginyeria Superior en Informàtica, UAB 2010.
8. Scott, J. *Social Network Anélisis. A Handbook*, Sage Publications, 1996.

Pàgines web consultades

1. Abeyà, M. Gabarró, M. Gasol A. i Rebordosa A. (Dept. Ensenyament, Generalitat de Catalunya), *Adaptació de la CDU per a les biblioteques escolars infantil, primària i secundària* disponible a http://bd.ub.edu/bescolar/IMG/pdf/adaptacio_cdu.pdf (consulta agost 2012)
2. Butler, P. *Visualizing Friendships*, disponible a https://www.facebook.com/note.php?note_id=469716398919 (consulta desembre 2011)
3. *Comprehensive TeX Archive Network*, http://www.ctan.org/ (consulta gener 2012)

4. Fundació CTIC, *Redes Sociales*, disponible a https://sites.google.com/a/xtec.cat/calaix-de-sastre-6a-2011/internet-segura/ /xarxessocials (consulta febrer 2012)

5. http://pajek.imfm.si/doku.php?id=download (consulta març 2012)

6. López, F. *Las estadísticas de Facebook España 2011*, disponible a http://www.ticbeat.com/socialmedia/estadisticas-facebook-espana-2011-slideshare/ (consulta juny 2012)

7. *Social Network Analysis*, disponible a http://www.orgnet.com/index.html (consulta juny 2012)

8. Universität Konstanz, *Projecte Visone*, disponible a http://visone.info/html/about.html (consulta març 2012)

9. Twitter, https://twitter.com/privacy (consulta maig 2012)

10. Viquipèdia, *Classificació Decimal Universal*, disponible a http://en.wikipedia.org/wiki/Universal_Decimal_Classification (consulta agost 2012)

11. Viquipèdia, *Social Network Analysis*, disponible a http://en.wikipedia.org/wiki/Social_network_analysis

Altres recursos utilitzats

1. Conferència a càrrec de J. Domingo-Ferrer (Universitat Rovira i Virgili), TITOL, en el marc de la *15a Trobada matemàtica* organitzada per la Societat Catalana de Matemàtiques el maig 2012.

2. Pel·lícula: *Redes Sociales*, dirigida per D. Fincher, estrenada l'octubre de 2010.

3. Programa de ràdio "El Mirall", *Ús i abús de les xarxes socials*, emès el 6 de setembre disponible també a http://www.catradio.cat/audio/661566/Us-i-abus-de-les-xarxes-socials.

Índex de figures

ANNEX: Programari utilitzat

A continuació comento alguns dels recursos informàtics utilitzats per a l'elaboració del treball.

LaTeX

Tot el treball està escrit utilitzant LaTeX com a processador de textos. És la primera vegada que utilitzo aquest programa i ha requerit un esforç inicial extra per aprendre'l. En global crec que ha valgut la pena fer aquest esforç, ja que la presentació final queda molt millor que si estigués fet amb els programes habituals per processar textos, com el Word. A part, aquest esforç es veurà recompensat, ja que en un futur podré utilitzar el LaTeX per altres treballs, ja que presenta molt avantatges, en especial pel que fa al format de fórmules i la gestió de tot el contingut.

El LaTeX és programari lliure i és molt utilitzat a nivell acadèmic, especialment a les universitats en els àmbits relacionats amb les ciències (matemàtiques, informàtica, física,...). Va ser elaborat per L. Lamport a partir del llenguatge TeX inventat per D. Knuth.

Per utilitzar-lo es necessita editar un fitxer amb extensió `.tex`, després compilar-lo amb el LaTeX o algun procés compost, que crea els fitxers amb extensió `.dvi`, o `.pdf` que es poden visualitzar amb altres programes com el *Yap* o l'*Adobe*. Per a editar el fitxer hi ha opcions molt diverses, segons es treballi en Windows, o Linux. La que jo he utilitzat és el WinEdt, que ja porta preparats una sèrie de menús que faciliten el procés de compilar i també la inserció de fórmules i instruccions especials. A continuació, la figura 5.1 mostra una captura de pantalla durant l'edició d'aquest treball.

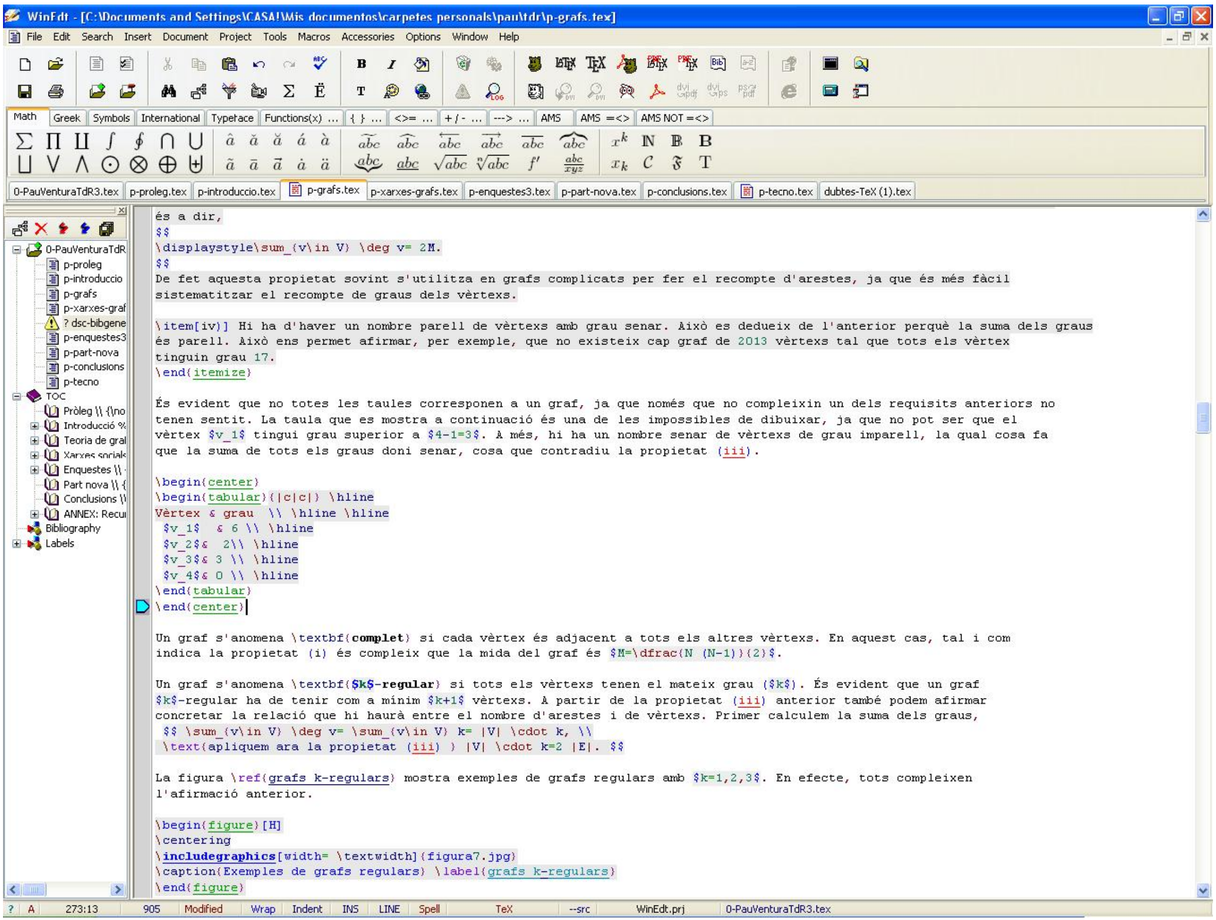


Figura 5.1: Captura de pantalla d'una part del treball amb l'editor WinEdt.

El procés de compilació es realitza en el llenguatge de l'ordinador i ens apareix la pantalla negra bàsica. A la captura de pantalla de la figura 5.2 es mostra el procés de compilació amb LaTeX, que crea un fitxer amb extensió `.dvi`, que vol dir device independent, que és exportable i vàlida arreu.

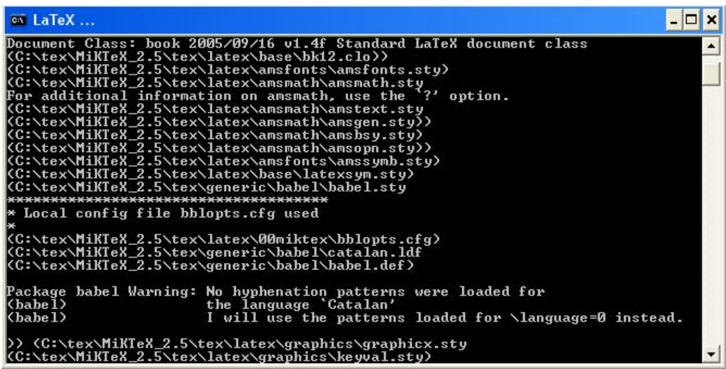


Figura 5.2: Captura de pantalla d'una part del treball compilant amb LaTeX.

Per a incloure figures, cal compilar amb el PDF TeXify, com es mostra a la figura 5.3. En aquests processos es generen també fitxers auxiliars, corresponents a l'índex, les referències, la llista de figures, etc. De fet en aquest treball, sense comptar figures, intervenen més de 50 fitxers.

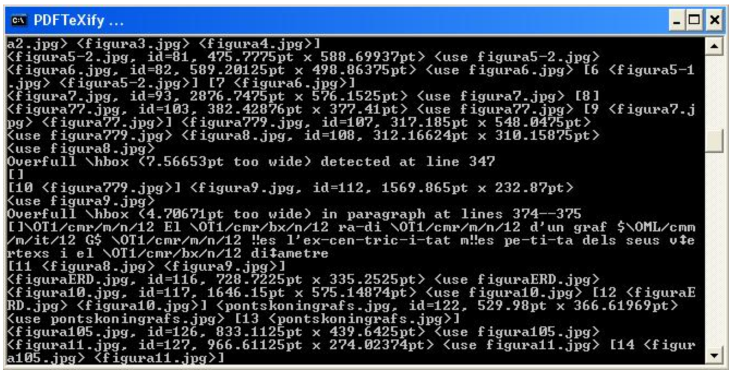


Figura 5.3: Captura de pantalla compilant una part del treball per crear el pdf amb PDF TeXify.

Dins dels avantatges d'aquest programari vull destacar els següents.

- **Format general.** Permet centrar-se exclusivament en el contingut sense preocupar-se en els detalls de format, que es gestionen de manera general i es poden modificar en qualsevol moment de manera ràpida.

- **Estructura general** Permet estructurar fàcilment el document en capítols, seccions, bibliografia, índex..., amb comptadors i numeració automàtica.

- **Projecte-arbre** Es pot crear un fitxer general que crida a altres fitxers, de manera que crea un arbre amb l'estructura dels diferents capítols. Així es poden organitzar els fitxers per capítols, i només et cal obrir el capítol en el qual estàs treballant. Això va bé perquè no cal compilar cada cop tot el treball.

- **Automatització** de l'índex general i de l'índex de figures.

- **Fórmules.** Aquest programa ja està pensat per tal de poder-hi introduir fórmules matemàtiques. Això és un gran avantatge, ja que es poden escriure tot tipus de fórmules per ser reproduïdes amb una tipografia professional.

- **Taules i figures.** Permet crear taules i afileraments amb molts paràmetres possibles. També permet incloure figures de diferents formats. Tant les taules com les figures poden ser definits com a elements flotants per facilitar la seva distribució a les pàgines, la seva enumeració automàtica, el text de peu de figura associat i el llistat final de l'índex de figures.

Tot i que com es veu és un programa amb un gran potencial, té també alguns inconvenients. El principal és que cal programar instruccions i, per tant, requereix un temps i un esforç d'aprenentatge. En l'elaboració concreta, les dificultats amb què m'he trobat són les següents.

- Tot i que està preparat a nivell mundial, he tingut alguns problemes amb els accents.
- La inclusió d'imatges no és tan directa com en altres programes, i ha requerit instal-lar algun paquet addicional.
- Cal anar compilant per veure com queda, ja que no és WYSIWYG (What You See Is What You Get)
- Com que és un programa, la sintaxi ha de ser estrictament correcta, sovint dóna errors i no es pot compilar. No sempre és fàcil veure l'origen de l'error per poder corregir-lo.

Visone

Per dibuixar i manipular els grafs he utilitzat el programa Visone, seguint la recomanació d'en J. Herrera i la C. Pérez, als que he consultat alguns dubtes sobre el programa. És també un programa que he hagut d'aprendre de nou, i això requereix temps i paciència. Potser per això m'ha semblat útil incloure en el treball una part de descripció del programari. Era una forma constructiva d'explicar la part de treball de camp de visualització i anàlisis, a la vegada que funcionava com a manual exemplificat.

Tot i així, aprofito per incloure aquí l'opinió personal sobre aquest programa, valorant avantatges i inconvenients.

Està molt bé que sigui compatible amb Excel, la qual cosa permet importar dades d'una manera fàcil i pràctica. De fet també permet obrir fitxers `.gml` construïts per dades proporcionades per rastrejadors.

Ara bé, a l'hora de dibuixar els grafs exemple de la secció 1.2, he comprovat que no funciona bé per a representar llaços. De fet és natural perquè està pensat per a analitzar xarxes socials. Els llaços representarien les relacions d'una persona amb ella mateix i això no té sentit.

Una altra de les dificultats és que no es poden etiquetar els vèrtex o les arestes amb subíndex. A l'hora d'etiquetar-los, doncs, he utilitzat números pels vèrtexs i lletres per les arestes.

Altres programes

He utilitzat també altres programes al llarg del treball. De fet, les eines informàtiques estan tan integrades en la vida quotidiana que les utilitzem contínuament.

Per exemple, s'ha utilitzat el Word de manera puntual pel redactat de l'enquesta. Per tal de tabular les respostes de l'enquesta, s'ha fet servir l'Excel. Tal com s'ha dit al capítol 3, per poder importar aquestes dades al Visone ha calgut exportar-les a l'extensió .cvs.

Per retallar i redimensionar algunes figures, captures de pantalla, imatges i fotografies he utilitzat el Paint.