Universitat de Girona

PROJECTE FI DE CARRERA

Estudi de la mobilitat en tres xarxes de transport públic

Titulació: Enginyeria Industrial (pla 2002)

Alumne: **Josep Barberillo Nualart**
(u1051158@correu.udg.edu)

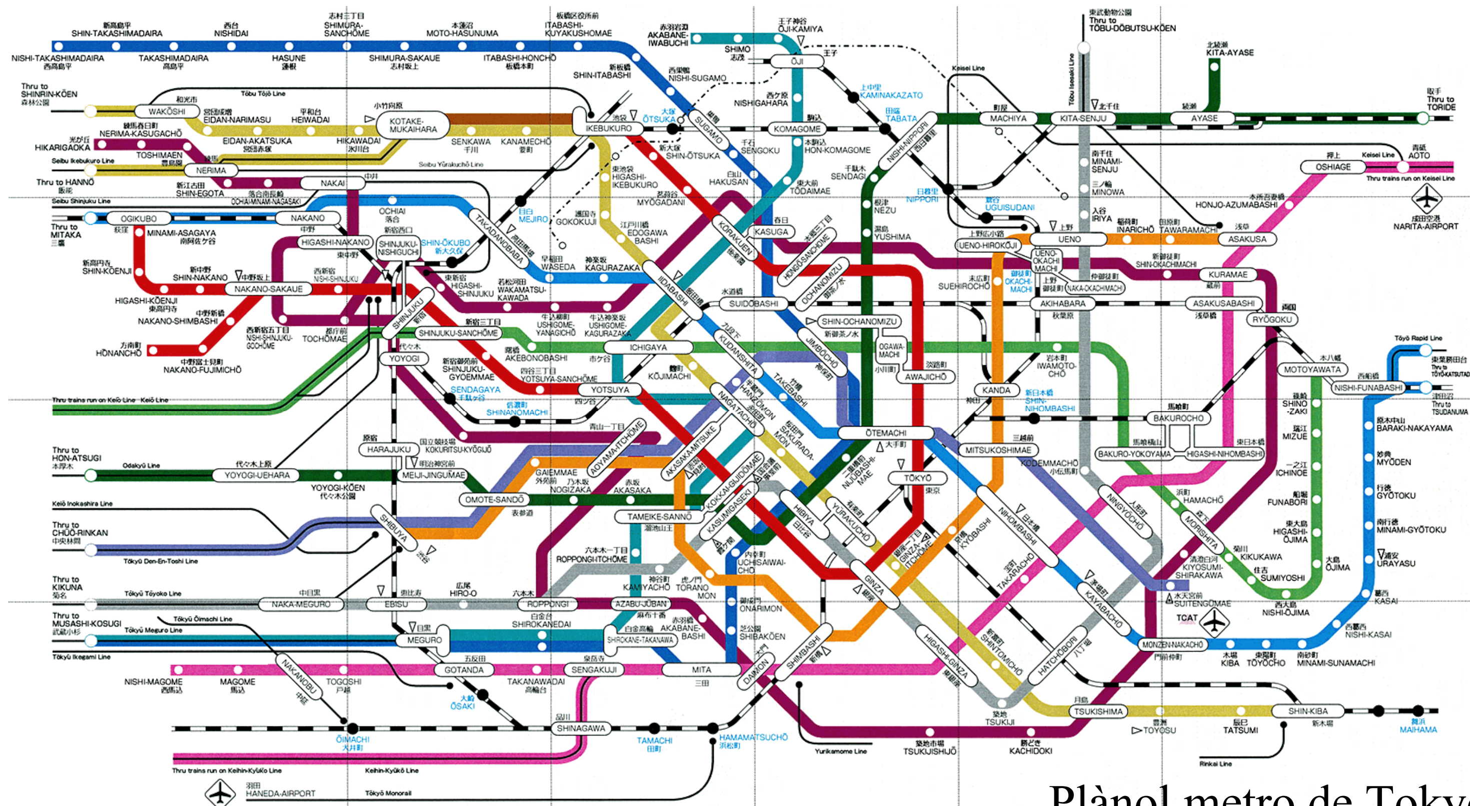
Director: **Joan Saldaña Meca**
(jsaldana@ima.udg.es)

Departament d'Informàtica i Matemàtica Aplicada

EPS Escola Politècnica Superior

QUÈ ÉS UNA XARXA?

•Des d'un punt de vista formal, una xarxa (o graf) és un conjunt de nodes connectats a través d'un seguit d'arestes.



Plànol metro de Tokyo

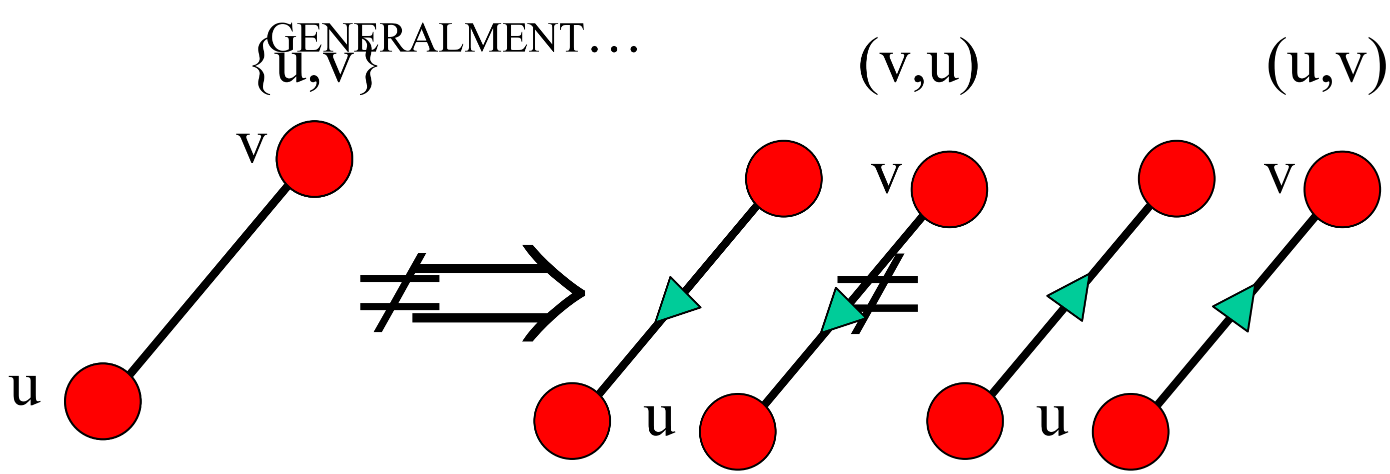
- Les arestes d'una xarxa poden ser dirigides o no dirigides:

  -En el cas d'aresta no dirigida que ens uneix els vèrtexs **u** i **v**, els camins $\mathbf{e_{uv}}$ i $\mathbf{e_{vu}}$ són equivalents ({u,v}) .

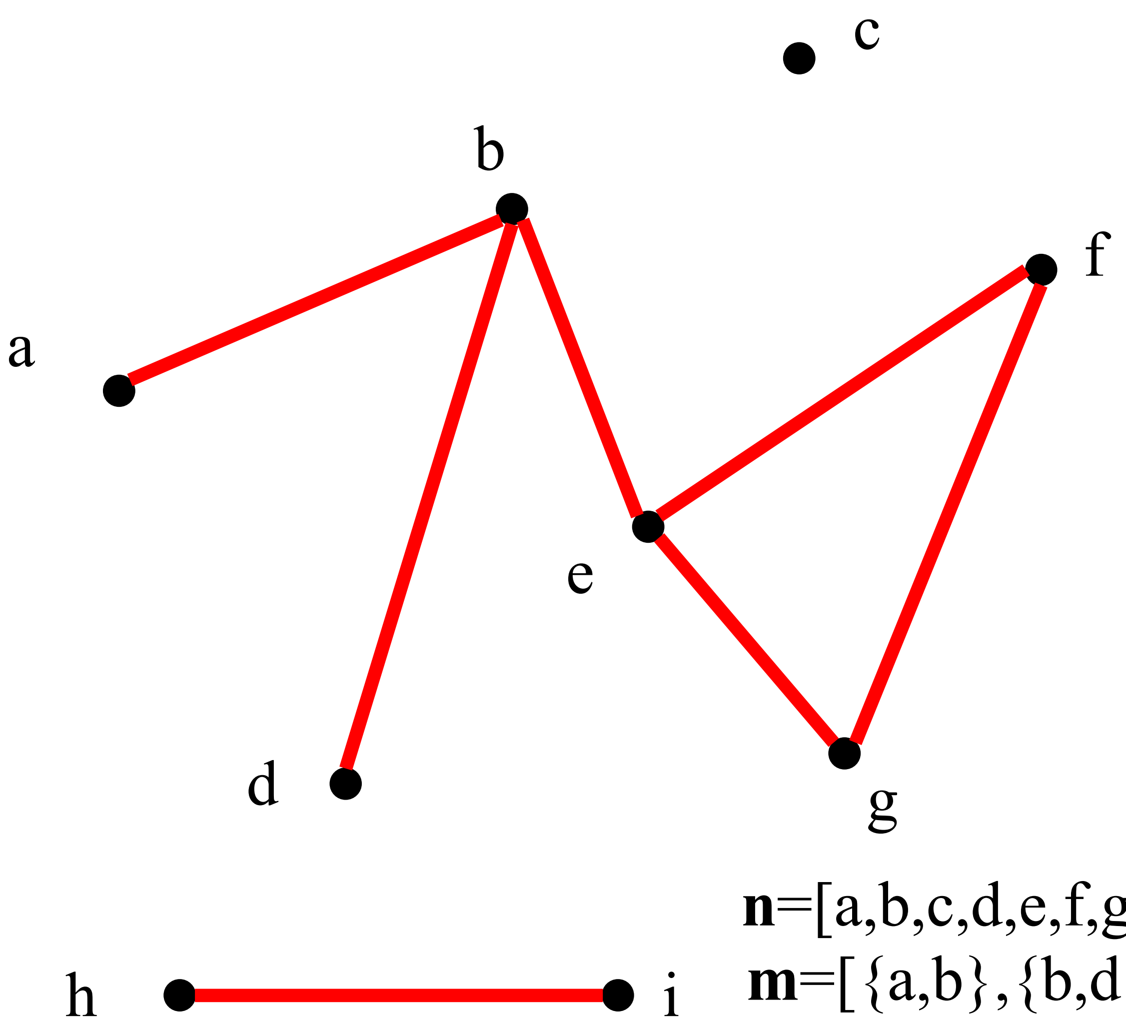
- Podem representar tot tipus de xarxes mitjançant **grafs**:

  -Un conjunt de **n** vèrtexs (nodes).

  -Un conjunt de **m** arestes (connexions entre nodes)



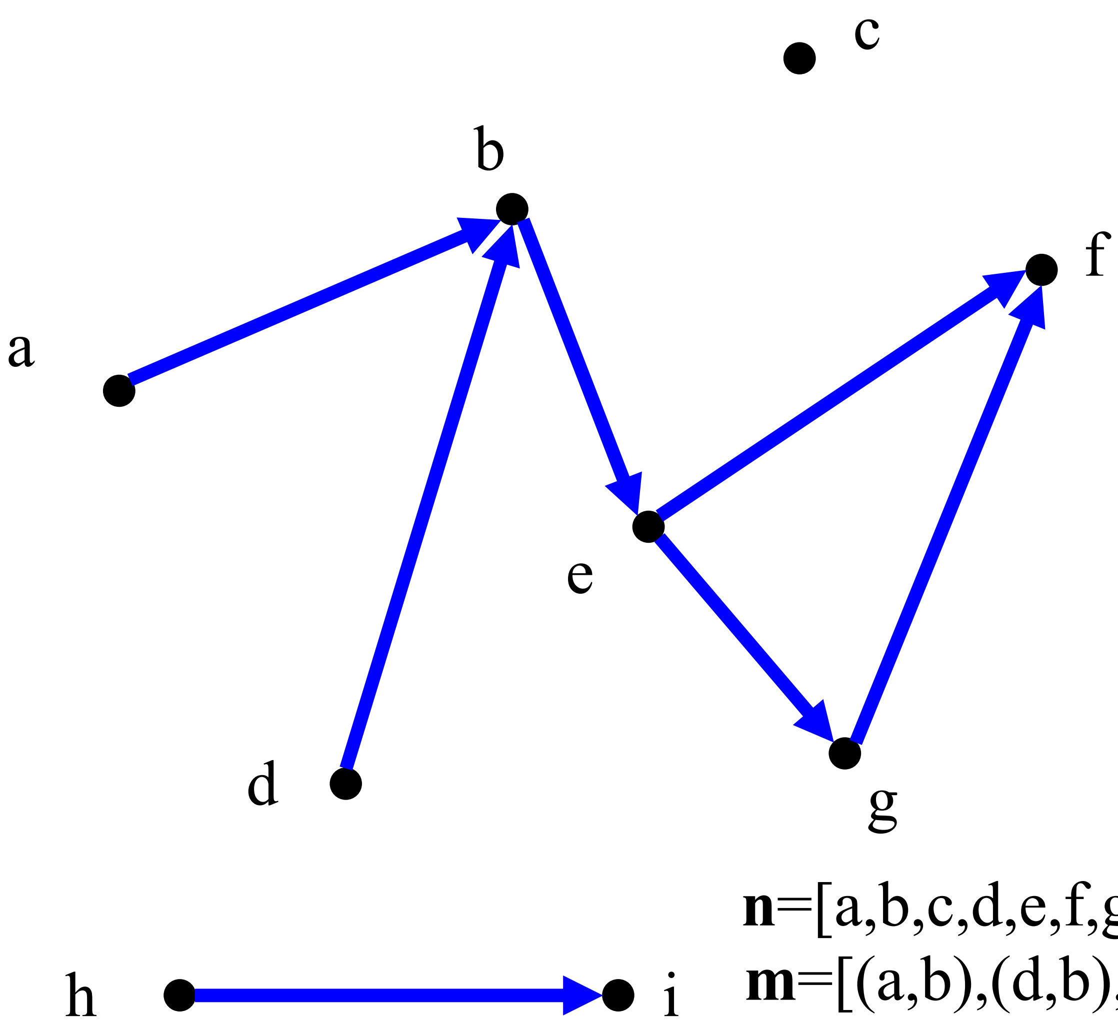
Exemple de graf **no dirigit**



n=[a,b,c,d,e,f,g,h,i]

m=[{a,b},{b,d},{b,e},{e,f},{e,g},{f,g},{h,i}]

Exemple de graf **dirigit**



n=[a,b,c,d,e,f,g,h,i]

m=[(a,b),(d,b),(b,e),(e,f),(e,g),(g,f),(h,i)]

CONCEPTE	ARESTES PARAL·LELES	BUCLES
GRAF	NO PERMÈS	PERMÈS
MULTIGRAF	PERMÈS	PERMÈS
GRAF SIMPLE	NO PERMÈS	NO PERMÈS

- Dues arestes són paral·leles quan són incidents amb els mateixos vèrtexs.

- Un arc amb el mateix vèrtex d'origen i de destí s'anomena un bucle.

• En un graf o multigraf G, es verifica que la suma dels graus de tots els vèrtexs és el doble que el nombre d'arestes:

$$\sum_{i=1}^{n} k_i = 2 \cdot m$$

essent k_i el grau del node i.

• El **nombre màxim d'arestes** que pot tenir un graf simple bé donat per la següent expressió:

$$m_{màx} = \frac{n(n-1)}{2} \quad \text{arestes}$$

- També podem incorporar pes a les arestes. Exemples:

  cost associat a transitar per l'aresta.

  capacitat de transport.

  flux d'informació...

Sigui G un graf de n vèrtexs:

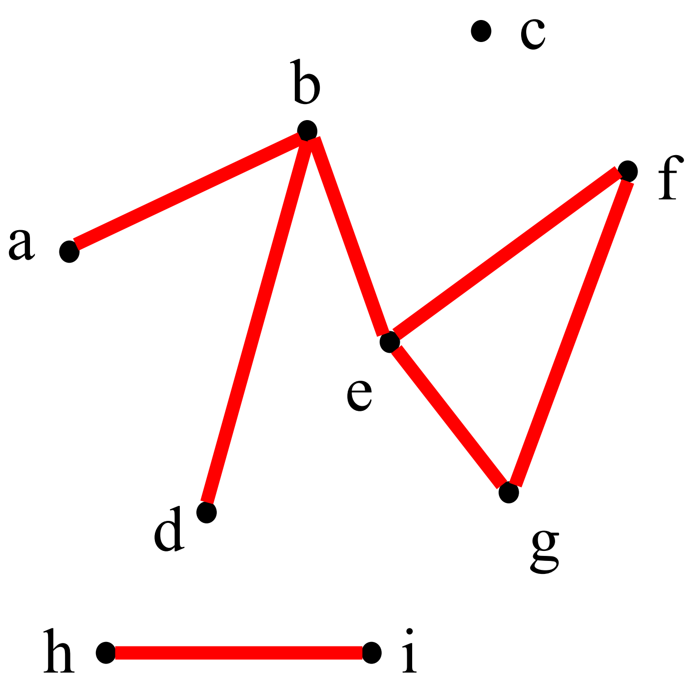
- La matriu d'ajacència **A** de G és una matriu booleana n x n de la forma **A**=$[a_{ij}]$, tal que:

  - l'element a_{ij} valdrà 1 si $\{v_i,v_j\}$ és una aresta de G

  - la matriu **A** serà **simètrica** si el graf és no dirigit.

Agafant l'exemple anterior, trobem la matriu d'adjacència **A** del graf:

$$A = \begin{array}{c} \begin{array}{ccccccccc} a & b & c & d & e & f & g & h & i \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \end{array}$$



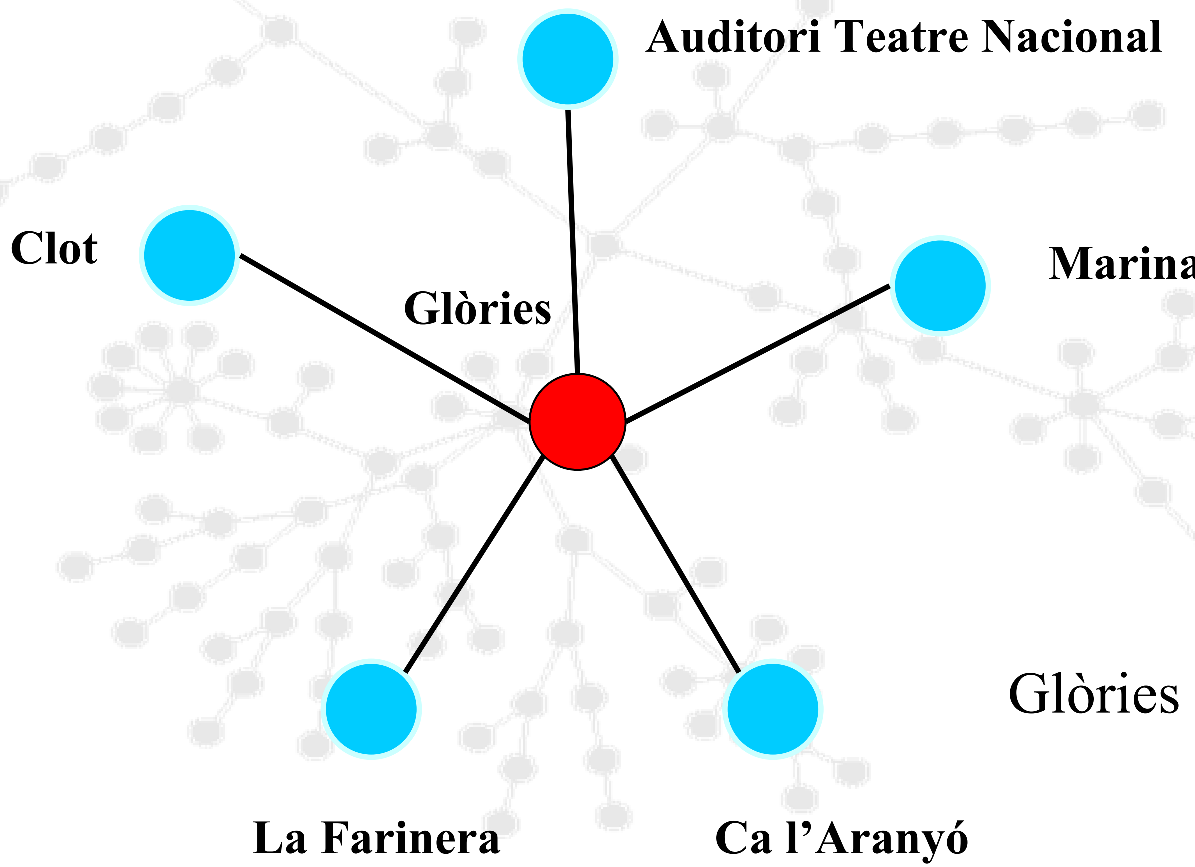
• Per tal d'estudiar la **topologia de la xarxa** definirem també conceptes com:

- grau → grau d'un node, grau mitjà primers veïns
- coeficient d'agrupament.
- camí més curt entre nodes.
- mesures de:
  - centralitat
  - d'accés a la xarxa i d'ocultació
  - internodalitat, page rank …

Conceptes que presentarem tot seguit.

GRAU VÈRTEX

- Anomenem grau d'un vèrtex al seu nombre total de connexions.



Glòries té grau: $k=5$

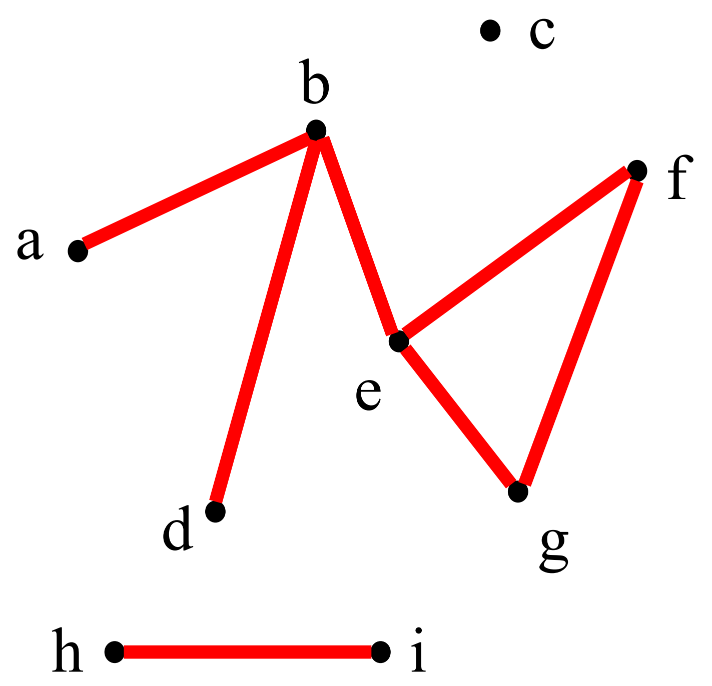
•Hem vist doncs que el grau d'un vèrtex és igual al nombre de primers veïns d'aquest.

•El grau d'un vèrtex $\boldsymbol{k_i}$ es pot calcular a partir de la matriu d'adjacència tal i com segueix:

$$k_i = (A^2)_{ii} = \sum_{j=1}^{n} A_{ij}$$

Exemple $k_i = (A^2)_{ii}$

$$A^2 = \begin{array}{c} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 3 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array} \begin{array}{c} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \end{array}$$



Agafant només els elements de la diagonal, comprovem que efectivament els seus valors corresponen al grau de cada node:

$k_a = k_h = k_i = k_d = 1$ $k_c = 0$

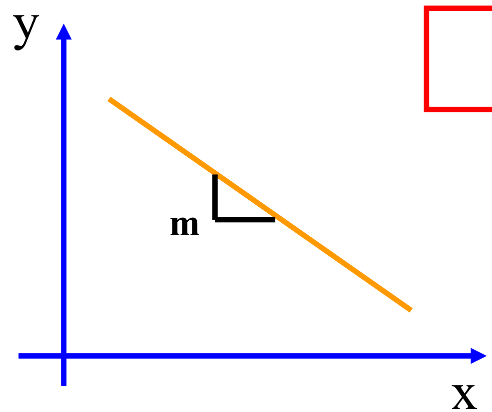
$k_b = k_e = 3$ $k_f = k_g = 2$

- Una de les propietats més importants que caracteritzen l'estructura d'una xarxa complexa és la distribució de graus P(k), que ens dóna la probabilitat que un node escollit a l'atzar tingui k connexions (o veïns).

- Els tipus de distribucions P(k) més importants són les següents:

Poisson	$P(k) = e^{-z} \cdot \frac{z^k}{k!}$
Exponencial	$P(k) = C \cdot e^{-\alpha \cdot k}$
Lliure d'Escala o potencial	$P(k) = C \cdot k^{-\gamma}$
Potencial truncada	$P(k) = C \cdot k^{-\gamma} \cdot e^{-\alpha \cdot k}$

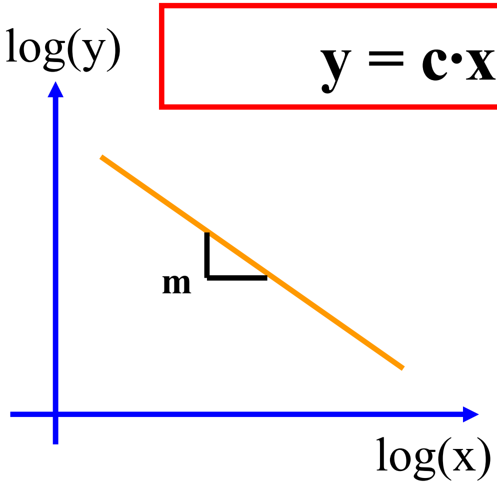
- En una correlació lineal:



$$\boxed{\mathbf{y = m \cdot x + b}}$$

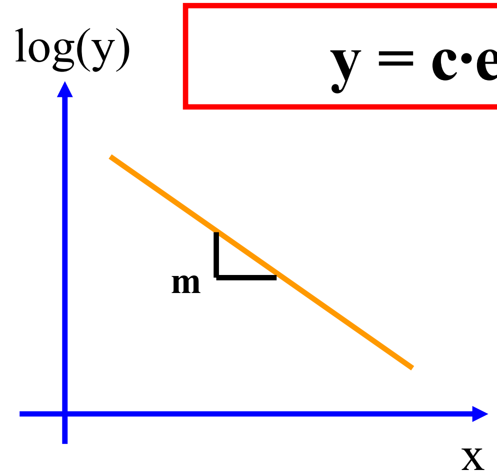
En una correlació potencial:

$$\boxed{\mathbf{y = c \cdot x^{m}}}$$



En una correlació exponencial:

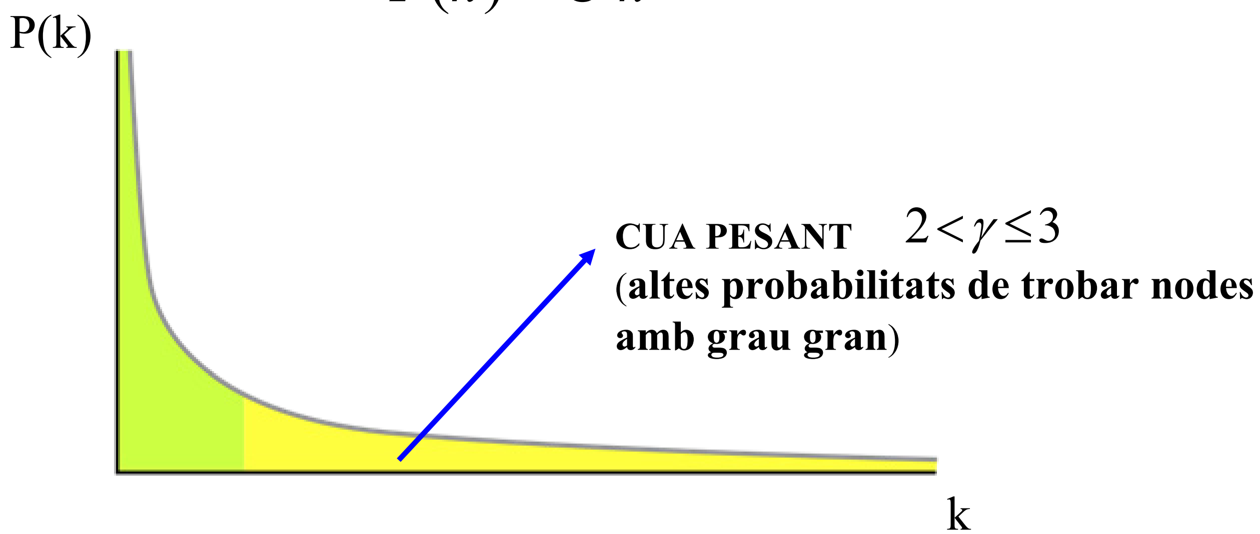
$$\boxed{\mathbf{y = c \cdot e^{m \cdot x}}}$$



•**La distribució potencial truncada** introdueix el terme exponencial $e^{-\alpha \cdot k}$ que té en compte els efectes que tenim pel fet de que la xarxa sigui de **tamany finit**.

Aquesta distribució és una alternativa respecte el comportament asimptòtic de la distribució potencial que en definitiva no és del tot realista en xarxes finites:

$$P(k) = C \cdot k^{-\gamma}$$



•En els casos que tinguem xarxes amb “cues primes” (baixa probabilitat de trobar nodes amb grau gran $\gamma>3$), llavors tant la **distribució exponencial** com la **distribució potencial** són bons candidats a descriure la fenomenologia de la xarxa.

•Aquest fet és degut a que costa distingir la diferència de bondat d’ajust entre una i altra corba quan tenim xarxes amb “cues primes”.

•Tot i això, l’exponencial sempre tendirà més ràpidament a zero, en canvi la potencial sempre es manté per sobre de la corba exponencial.

INTROD. A LES XARXES
GRAU VÈRTEX
GRAU MITJÀ PRIMERS VEÏNS
COEFICIENT DE AGRUPAMENT
CAMINS MÍNIMS
MESURA CENTRALITAT
MESURA PROXIMITAT
ACCÉS A LA XARXA
OCULTACIÓ NODAL
INTERNODALITAT
BETWEENNESS ARESTES
FUNNELING
PAGE RANK
CONCLUSIONS

- Demostrem-ho:

$$L=\lim_{k\to\infty}\left(\frac{e^{-\alpha k}}{k^{-\gamma}}\right)=\frac{0}{0} \longrightarrow L=\lim_{k\to\infty}\left(\frac{k^{\gamma}}{e^{\alpha k}}\right)=\frac{\infty}{\infty}$$

- Aplicant la regla de l'Hôpital per a solucionar la indeterminació i suposant que **γ=4**:

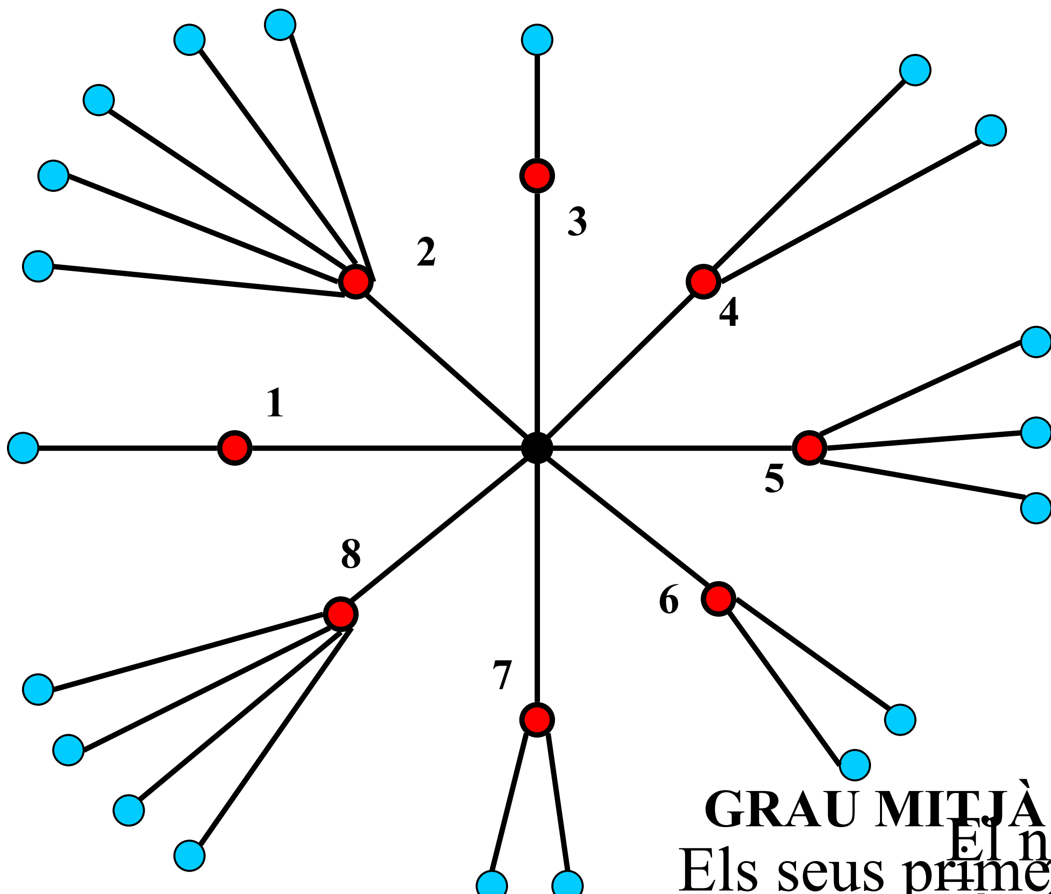
$$\boxed{\lim_{k\to\infty}\left(\frac{\gamma\cdot(\gamma-1)\cdot(\gamma-2)\cdot(\gamma-3)\cdot k^{\gamma-4}}{\alpha^{4}\cdot e^{\alpha k}}\right)}$$

Un cop realitzades les corresponents derivades, comprovem que al numerador obtenim un polinomi que depèn exclusivament de γ. Al fer el límit tindrem que:

$$\lim_{k\to\infty}\left(\frac{\gamma\cdot(\gamma-1)\cdot(\gamma-2)\cdot(\gamma-3)\cdot k^{\gamma-4}}{\alpha^{4}\cdot e^{\alpha k}}\right)=\frac{p(\gamma)}{\infty}=0$$

GRAU MITJÀ PRIMERS VEÏNS

- És el nombre mitjà de veïns que té un veí d'un vèrtex de grau k:



El node considerat té grau=8

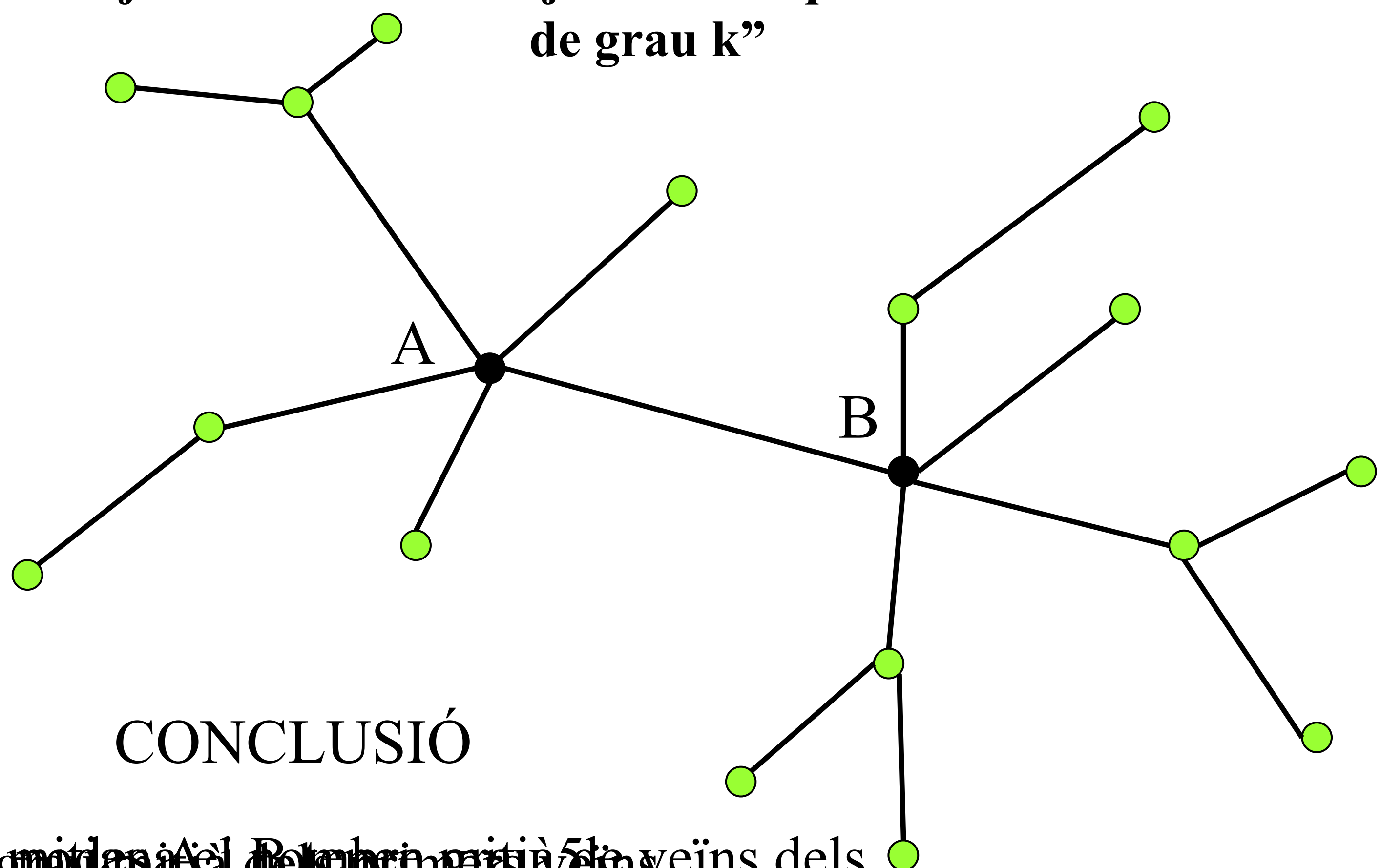
Els seus primers veïns tenen en total: 20 veïns

GRAU MITJÀ PRIMERS VEÏNS:

$$k = \frac{20}{8} = 2.5$$

•Estenent la definició anterior pel global d'una xarxa definim el grau mitjà dels primers veïns com:

"en mitjana el nombre mitjà de veïns que té un veí d'un vèrtex de grau k"



CONCLUSIÓ

Els nodes A i B tenen grau 5

El grau mitjà dels primers veïns de A val: 7/5=1.4

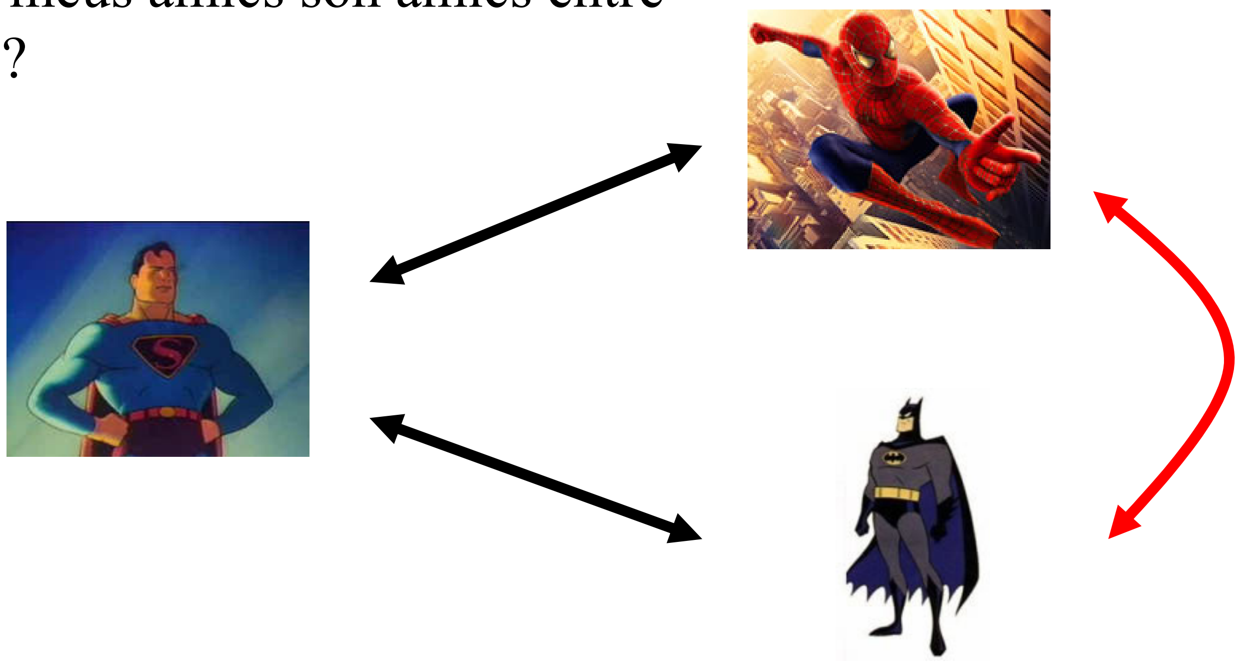
El grau mitjà dels primers veïns de B val: 9/5=1.8

En mitjana el nombre mitjà de veïns dels nodes de grau 5 val: (1.4+1.8)/2=1.6

COEFICIENT DE AGRUPAMENT

- El coeficient d'agrupament ens mesura la probabilitat que dos veïns d'un vèrtex, siguin també veïns entre ells.

Els meus amics són amics entre ells?



- Un valor elevat d'aquest coeficient reflecteix l'alta quantitat de triangles existents a la xarxa.

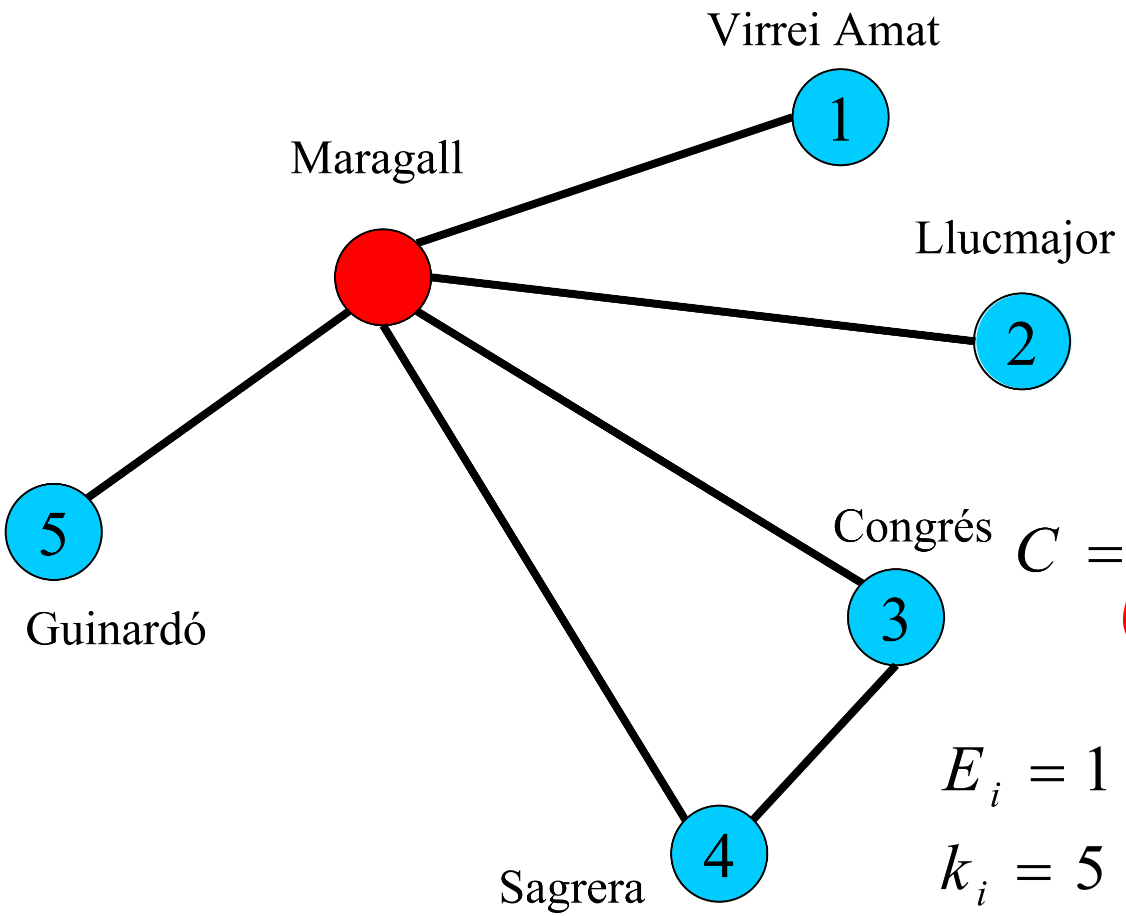
• El coeficient d’agrupament d’un node v bé donat per les expressions equivalents següents:

$$C_v(G) = \frac{nombre\ d'arestes\ entre\ veïns\ de\ v}{\binom{k_v}{2}}$$

$$C_v(G) = \frac{nombre\ de\ parelles\ ab, ac\ d'arestes\ adjacents\ per\ les\ quals\ bc\ es\ una\ aresta}{nombre\ de\ parelles\ ab, ac\ d'arestes\ adjacents}$$

• El coeficient d’agrupament mitjà de la xarxa el definim com:

$$C(G) = \sum_{v=1}^{n} C_v(G) / n$$



Nombre màxim d'arestes entre primers veïns

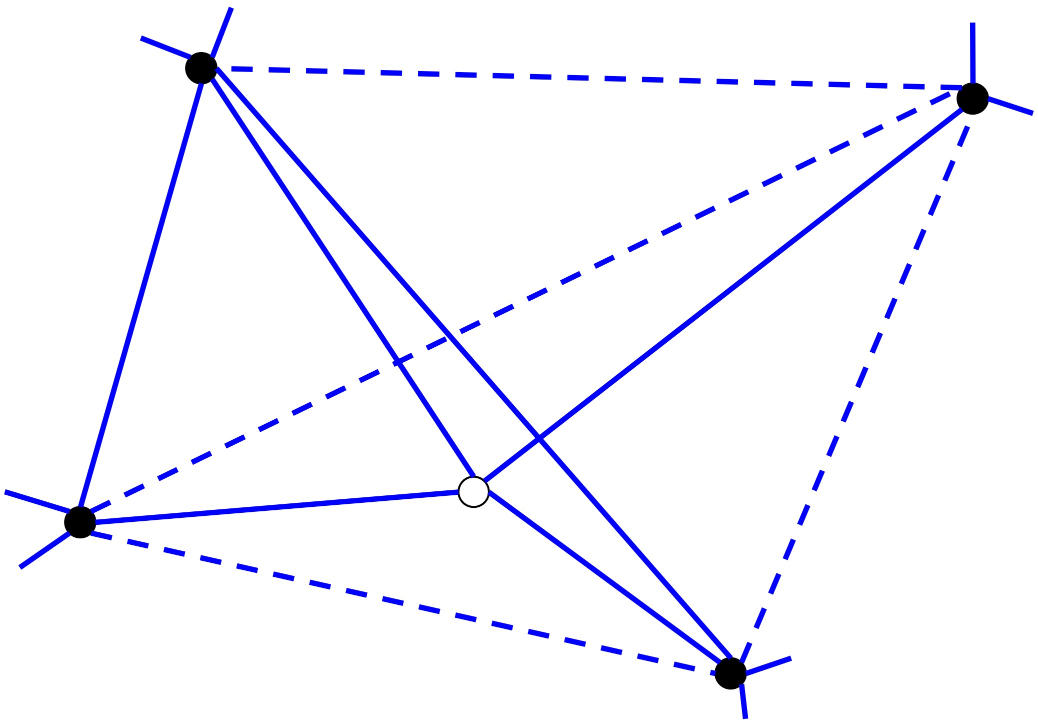
$$C = \frac{2 \cdot E_i}{k_i \cdot (k_i - 1)}$$

$$\begin{aligned} E_i &= 1 \\ k_i &= 5 \end{aligned} \Rightarrow C = 0.1$$

ki:nombre de primers veïns

Ei:nombre arestes existents entre primers veïns

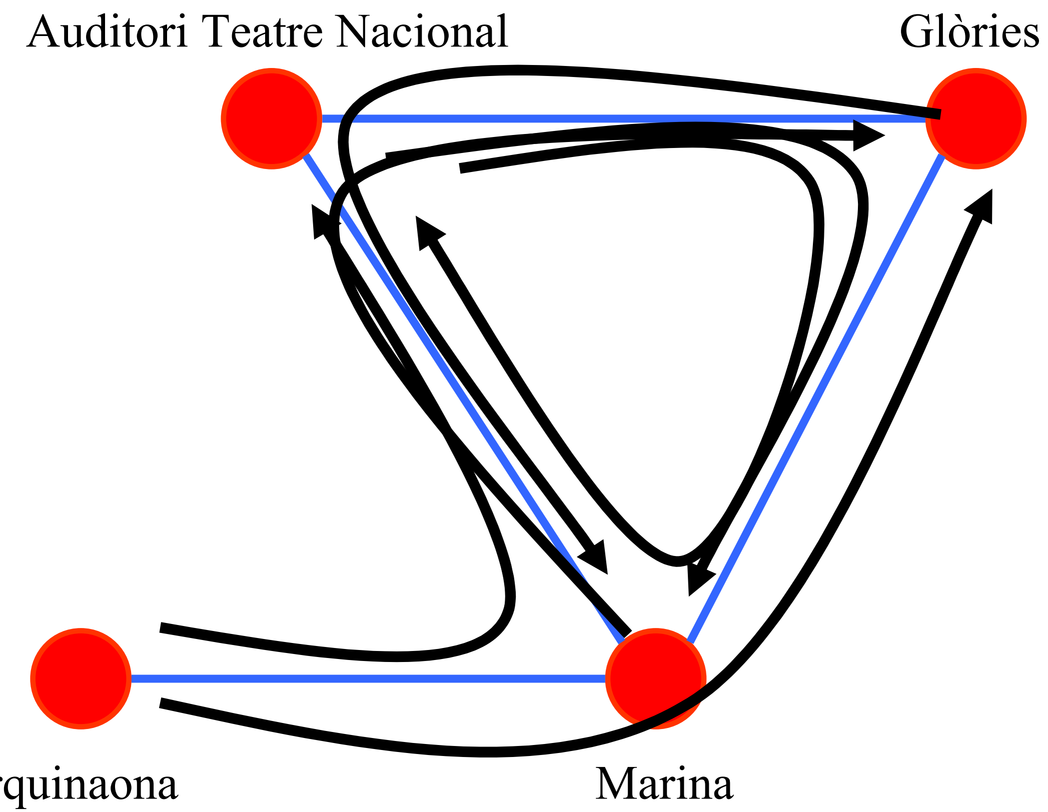
- El coeficient d'agrupament ens caracteritza la densitat de connexions en l'ambient més pròxim al node.



$$C = \frac{2y}{z(z-1)} = \frac{2\cdot 2}{4\cdot(4-1)} = \frac{1}{3}$$

n=4 vèrtexs m=5arestes

$$C = 3 \cdot \frac{N_\Delta}{N_U}$$



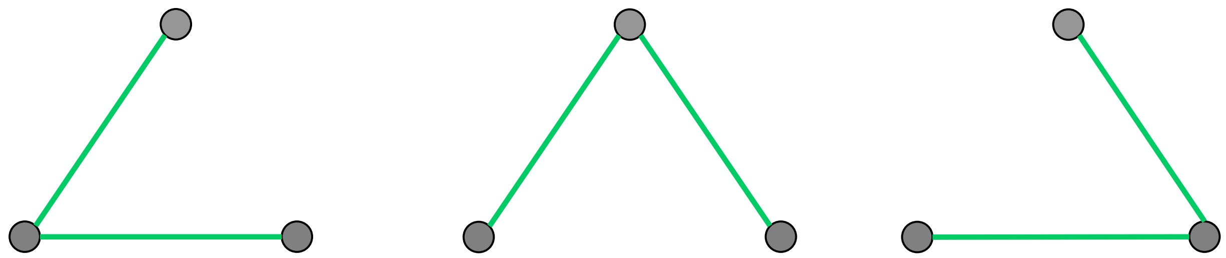
$$N_\Delta = 1$$

$$C = 3 \cdot \frac{N_\Delta}{N_U} = \frac{3}{5}$$

$$N_U = 5$$

$\mathbf{N_{\Delta}}$: nombre de triangles existents a la xarxa. Els triangles són trios de vèrtexs on cada vèrtex està connectat amb els altres dos.

$\mathbf{N_{U}}$: nombre de connexions triples de vèrtexs existents a la xarxa. Són tripletes de vèrtexs on un dels vèrtexs està connectat amb els altres dos (triangles oberts):



Nota: cada triangle està format per 3 tripletes de vèrtexs diferents.

• Alternativament podem calcular el **nombre de triangles** (NΔ) en una xarxa a partir de la matriu d'adjacència com:

$$\boxed{N\Delta = \frac{1}{6} trA^3}$$

tr: *traça*; suma dels elements de la diagonal de la matriu.

CAMINS MÍNIMS

- Definim una longitud ("1") igual per totes les arestes de la xarxa.

- Conseqüentment, la distància entre una parella de vèrtexs vindrà donada pel nombre d'arestes recorregudes per anar d'un vèrtex a l'altre.

- El camí mínim $\mathbf{l_{uv} = l_{vu}}$ entre els nodes **u** i **v** prendrà un valor donat pel mínim nombre d'arestes necessàries per anar de **u** a **v**.

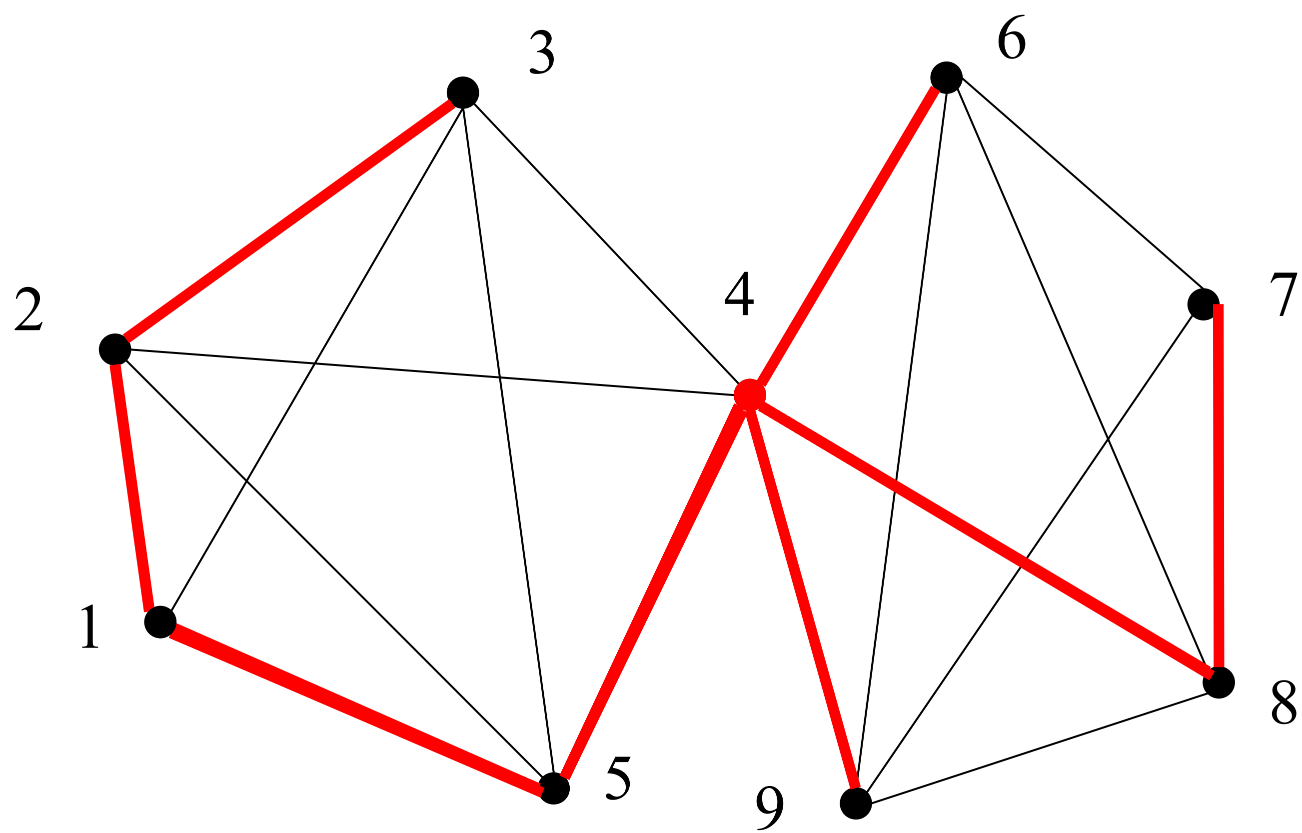
MESURA DE CENTRALITAT

• Aquesta mesura es defineix com la mitjana de les longituds dels camins mínims d'un node respecte tota la resta de nodes de la xarxa:

$$D_i = \frac{\sum_{i \neq j} d(v_i, v_j)}{(N-1)}$$

• És un indicador de "com de proper estic respecte la resta de nodes de la xarxa".

ESTUDI CENTRALITAT NODE 1



$L_{12}=1$ $L_{13}=2$ $L_{14}=2$ $L_{15}=1$

$L_{16}=3$ $L_{17}=4$ $L_{18}=3$ $L_{19}=3$ $D_1 = \frac{19}{(9-1)} = 2.375$

Repetint procediment pel node 4 obtenim un valor $D_4=10/8=1.25$ que ens indica una centralitat més elevada d'aquest.

MESURA DE PROXIMITAT

- Aquesta mesura la definim com el invers de la centralitat nodal:

$$T_i = \frac{N-1}{\sum_{j \neq i} d(v_i, v_j)}$$

On $d(v_i,v_j)$ és la longitud del camí més curt que va de v_i a v_j en G.

- La proximitat nodal ens permet entendre d'una forma més intuïtiva el concepte que estem tractant ja que nodes altament pròxims a la xarxa també tindran un valor alt d'aquesta mesura.

INFORMACIÓ DE ACCESSIBILITAT

- Ens mesura el nombre mitjà de preguntes que necessitem fer per localitzar una estació dins de la nostra xarxa partint d'un determinat node (**Nº BITS d'informació necessaris en mitjana**).

- És una mesura de com de bo és l'accés a la xarxa des del node considerat.

Atenció!: En tot moment estem considerant que ens volem moure per la xarxa emprant camins de cost mínim.

La probabilitat d'anar de **s** a **t** per un camí mínim bé donada per:

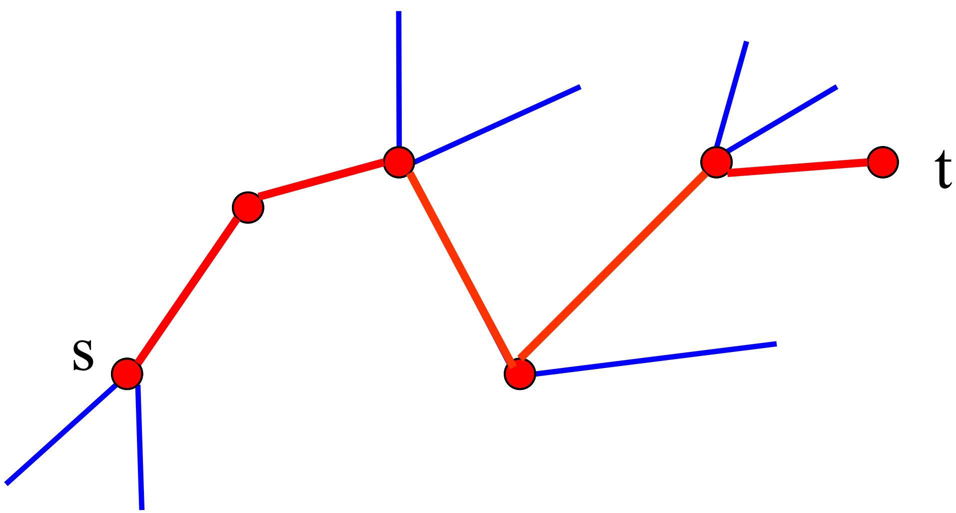
$$P[path(s,t)] = \frac{1}{k_s} \cdot \prod_{j \in path(s,t)} \frac{1}{k_j - 1}$$

EXEMPLE *Quina probabilitat tinc d'anar de s a t per un camí mínim?*

$$P(s \to t) = \frac{1}{3} \cdot \frac{1}{(2-1)} \cdot \frac{1}{(4-1)} \cdot \frac{1}{(3-1)} \cdot \frac{1}{(4-1)} = \frac{1}{3} \cdot \frac{1}{1} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{54}$$

Una vegada de cada 54 arribarem a **t** des de **s** a través d'un camí mínim.

Ens limita que no poguem tornar pel camí que venim.



La informació (en bits) que cal saber per anar de **s** a **t,** també anomenada "search information":

$$S(s \to t) = -\log_2 \sum_{\{path(s,t)\}} P[path(s,t)]$$

Accessibilitat a la xarxa des del node **s**:

$$\boxed{A_s = \frac{1}{(n-1)} \sum_{t \neq s} S(s,t)}$$

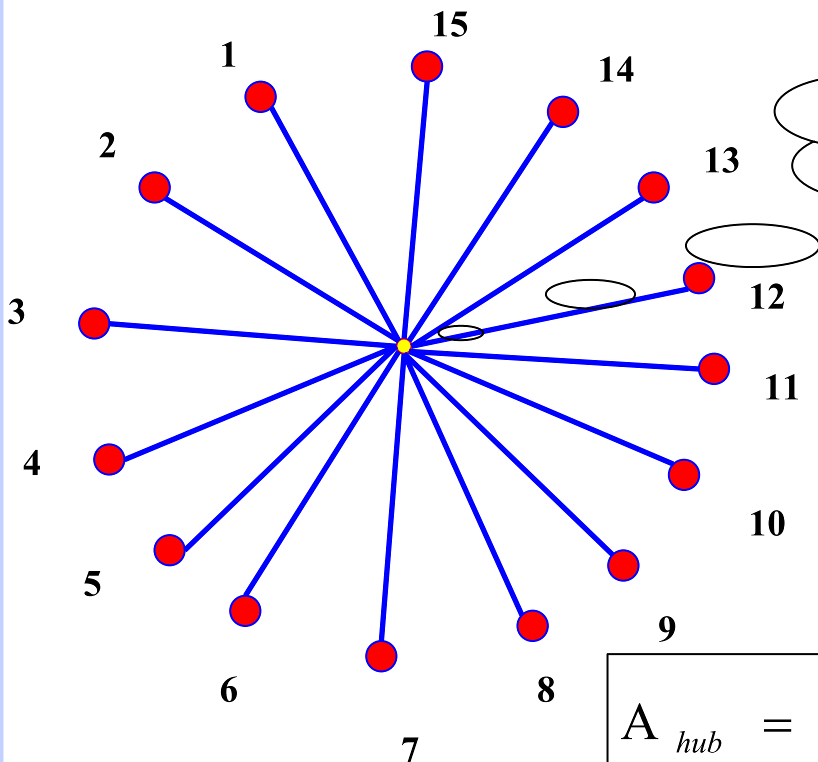
n: nombre de nodes

Per tal de caracteritzar la complexitat global de trobar estacions a la xarxa calculem la mitjana d'informació d'accés a aquesta des de qualsevol vèrtex com:

$$S = \frac{1}{n\cdot(n-1)}\sum_{s=1}^{n}\sum_{t\neq s} S(s,t)$$

Si volem comparar xarxes de diferent tamany podem escalar la mesura tal i com segueix:

$$\sigma = \frac{S}{\log_2(n)}$$



L'elecció és difícil…

Bits d'informació requerits: 3,91

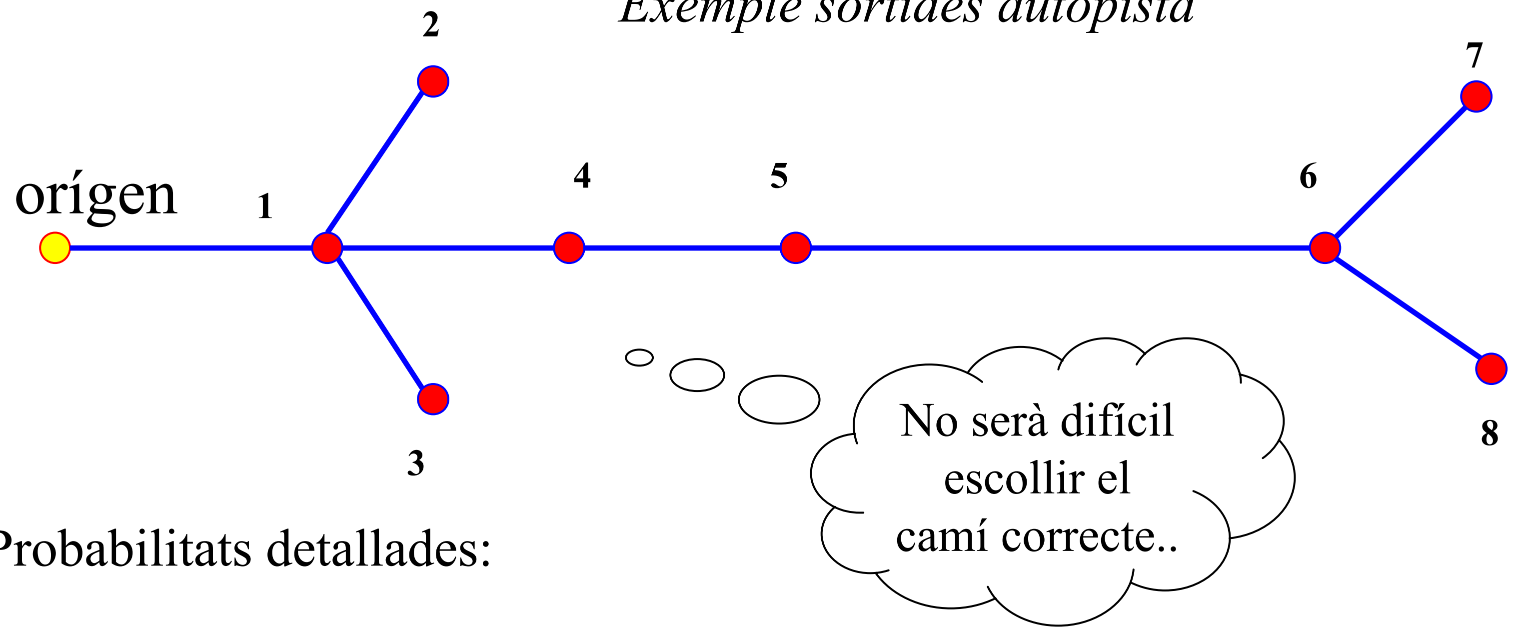
$$A_{hub} = \frac{1}{16-1}\cdot\left[15\cdot(3,91)\right] = 3,91\ bits$$

Recull possibles eleccions:

1: 0000	4: 0011	7: 0110	10: 1001	13: 1100
2: 0001	5: 0100	8: 0111	11: 1010	14: 1101
3: 0010	6: 0101	9: 1000	12: 1011	15: 1110

En un 'hub' és fàcil arribar-hi des d'altres veïns (no s'amaga), no obstant, sortint des del 'hub' és difícil arribar a un veí específic (té moltes opcions per escollir).

Exemple sortides autopista



Probabilitats detallades:

p(camí(0,1))=1 → *0 bits*

p(camí(0,2),(0,3),(0,4),(0,5)(0,6))=1/3 → *1,58 bits*

p(camí(0,7),(0,8))=1/6 → *2,58 bits*

L'accessibilitat a la xarxa des de l'orígen té un valor de 1,63bits:

$$A_0 = \frac{1}{8} \cdot (5 \cdot 1{,}58 + 2 \cdot 2{,}58) = 1{,}63\ bits$$

INFORMACIÓ DE LOCALITZACIÓ

• Ens mesura el nombre mitjà de preguntes que necessitem fer des d'un node qualsevol de la xarxa per a localitzar i arribar a un node concret mitjançant el corresponent camí més curt (mínim).

• És una mesura de com de bé "s'amaga" el node de la resta de la xarxa, és a dir, és una mesura de com de difícil és trobar el node.

• Quants bits d'informació necessiten saber (en mitjana) els altres nodes per a trobar-me?

La probabilitat d'anar de **t** a **s** per un camí mínim bé donada per:

$$P[path(t,s)] = \frac{1}{k_t} \cdot \prod_{j \in path(t,s)} \frac{1}{k_j - 1}$$

La informació que cal per anar de **t** a **s** en bits:

$$S(t \to s) = -\log_2 \sum_{\{path(t,s)\}} P[path(t,s)]$$

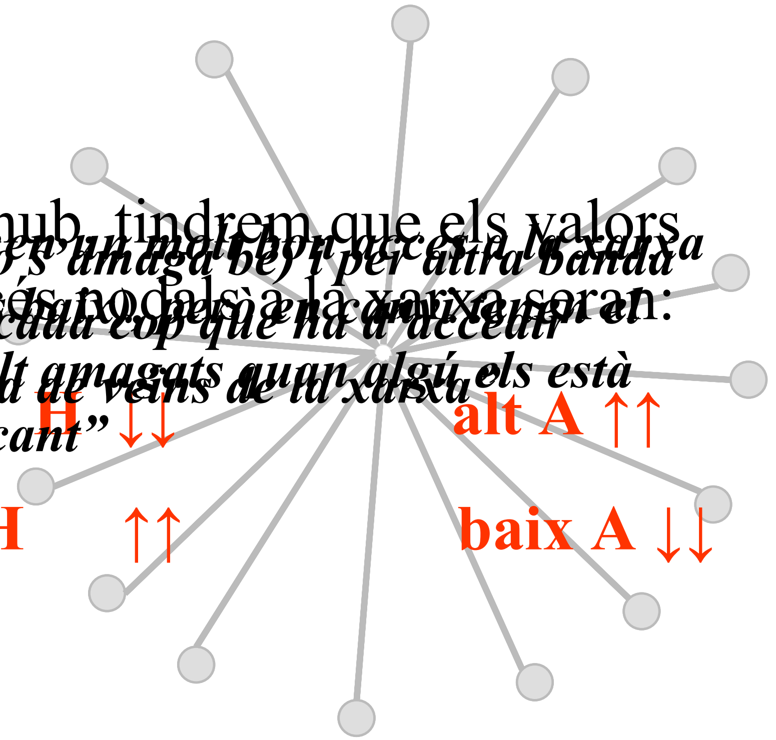
Informació de localització del node **s**:

$$\boxed{H_s = \frac{1}{n-1} \sum_{t \neq s} S(t,s)}$$

n: nombre de nodes

•Cal remarcar que la diferència amb el valor informació d'accés a la xarxa recau amb l'asimetria de quin és el node final de camí:

$$H_t = \frac{1}{n-1}\sum_{s\neq t} S(s,t) \quad \neq \quad A_t = \frac{1}{n-1}\sum_{s\neq t} S(t,s)$$



Retornant a l'exemple del hub, tindrem que els valors esperats d'ocultació i d'accés nodals a la xarxa seran:

“Els veïns del hub en canvi tenen un molt bon accés a la xarxa (poca informació necessària) però en canvi tenen el handicap de que queden molt amagats quan algú els està buscant”

“És molt fàcil trobar el hub (no s'amaga bé) i per altra banda personalment no té difícil cada cop que ha d'accedir adequadament a la resta de veïns de la xarxa”

HUB	**baix H** ↓↓	**alt A** ↑↑
VEÏNS	**alt H** ↑↑	**baix A** ↓↓

INTROD. A LES XARXES
GRAU VÈRTEX
GRAU MITJÀ PRIMERS VEÏNS
COEFICIENT DE AGRUPAMENT
CAMINS MÍNIMS
MESURA CENTRALITAT
MESURA PROXIMITAT
ACCÉS A LA XARXA
OCULTACIÓ NODAL
INTERNODALITAT
BETWEENNESS ARESTES
FUNNELING
PAGE RANK
CONCLUSIONS

INTERNODALITAT

•La Internodalitat d'un node és el nombre total de camins mínims entre totes les possibles combinacions de parelles de vèrtexs que passen a través d'aquest com a vèrtex intermedi.

$$\sigma(m) = \frac{1}{B} \sum_{i \neq j} B(i, m, j)$$

B(i,m,j): camins mínims entre i & j que passen per m.

B: total de possibles camins mínims entre parelles de nodes qualsevol i que són candidats a tenir el vèrtex *m* com a vèrtex intermedi:

$$B = \frac{(n-1)\cdot(n-2)}{2}$$

$$B = \frac{(n-1)\cdot(n-2)}{2}$$

A priori podríem pensar que l'expressió hauria de ser:

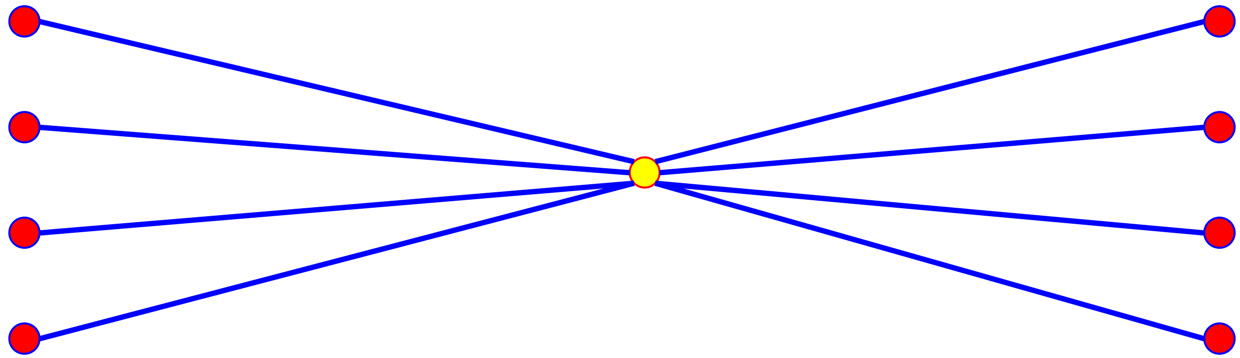
$$B = \frac{n\cdot(n-1)}{2}$$

Però per tal de no comptabilitzar els camins **(i,m)** o **(m,j)** on el vèrtex **m** ja no tindria opció de ser intermedi, cal que <u>restem 1 a cada factor del numerador</u>, obtenint l'expressió proposada segons el nostre criteri de normalització.

•Aquesta mesura ens indica la importància del vèrtex en el tràfic global de la xarxa.

•Podem intuir el fet que nodes amb un grau elevat, tinguin també una internodalitat elevada.

EXEMPLE COLL D'AMPOLLA



Tots els camins mínims entre nodes passen pel node central! Aquest node tindrà **Beetweenness igual a 1**!!

•La Internodalitat d'una aresta és el nombre total de camins mínims entre totes les possibles combinacions de parelles de vèrtexs que passen a través d'aquesta com a aresta intermèdia.

$$\sigma(\{a,b\}) = \frac{1}{B} \cdot \sum_{i \neq j} B(i,\{a,b\},j)$$

B(i,{a,b},j): camins curts entre i & j que passen a través de l'aresta {a,b} com aresta intermèdia.

B: total de possibles camins mínims entre parelles de nodes qualsevol.

En aquest cas tindrem que: $$B = \frac{n \cdot (n-1)}{2}$$

- Una aresta amb internodalitat elevada ens indicarà que canalitza la circulació d’un gran nombre de camins mínims entre vèrtexs a través de ella.

- Podem pensar que una aresta amb internodalitat elevada anirà a connectar possiblement a un vèrtex que a priori també tindrà una mesura de la internodalitat gran.

FUNNELING

- L'efecte '**funneling**' es dóna quan els diferents veïns d'un node no canalitzen els camins cap a la resta de nodes de la xarxa de la mateixa forma.

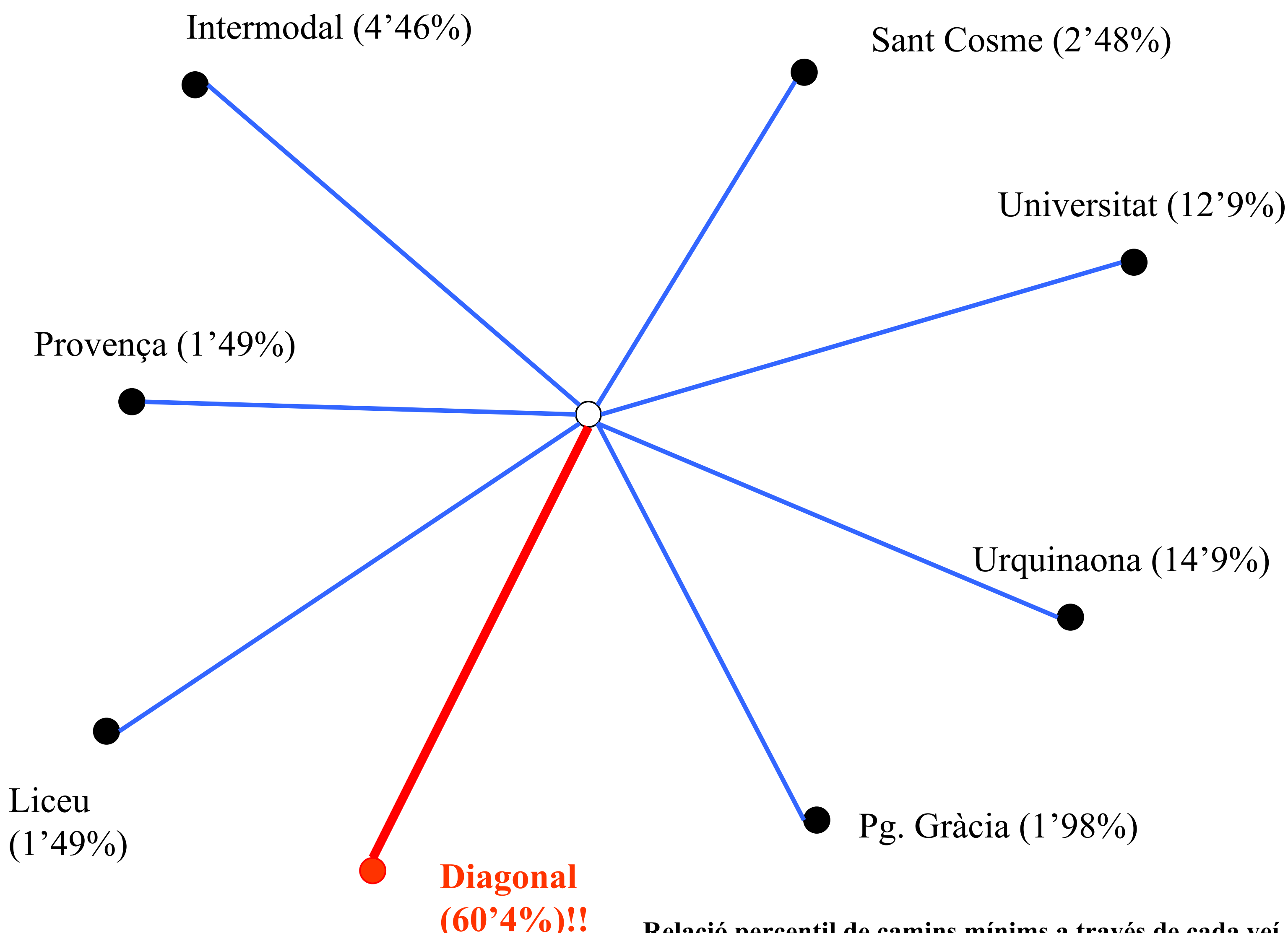
- És a dir, són un o dos dels primers veïns del node els que realment ens estan connectant amb la resta de la xarxa.

- En gran part, la majoria dels camins mínims que ens connecten amb la resta de la xarxa passen només a través d'aquests nodes que podem considerar "més famosos".

- M.E.J. *Newman, Scientific collaboration networks. Shortest paths, weighted networks, and centrality,* Phys. Rev. E64, 016132 (2001)

EXEMPLE

Plaça Catalunya



Relació percentil de camins mínims a través de cada veí

EXEMPLE **Plaça Catalunya**

- Observem que l'estació de **Plaça Catalunya** té un efecte funneling elevat ja que gairebé dues terceres parts dels camins mínims passen pel seu veí **Diagonal**.

- A una distància notable en quan a importància segueixen les estacions de **Urquinaona** (14'9% de camins mínims) i **Universitat** (12'9% de camins mínims).

PAGE RANK

•Els algorismes de cerca d'informació a Internet, com el *Page Rank* de Google, ens serveixen en aquest cas per establir un nou criteri alhora de classificar segons la seva importància dins de la xarxa les estacions estudiades.

•Aquesta mesura ens indica el percentatge de vegades que passarem pel node considerat si fem un viatge aleatori per la xarxa. Així doncs, cal esperar que una estació important dins la xarxa tingui un *Page Rank* elevat.

•D'aquesta forma, si naveguem aleatòriament per internet un temps suficientment gran, llavors tindrem una gran probabilitat de trobar-nos amb les pàgines de major *Page Rank*.

$$PR(v_i) = \frac{(1-d)}{n} + d \cdot \left(\frac{PR(v_1)}{k_1} + \frac{PR(v_2)}{k_2} + \ldots + \frac{PR(v_n)}{k_n} \right)$$

$PR(v_i)$: Page Rank node i-èsim

n: nombre total nodes xarxa

k_i: grau node i-èsim

d: assigna una probabilitat de salt no nul·la.

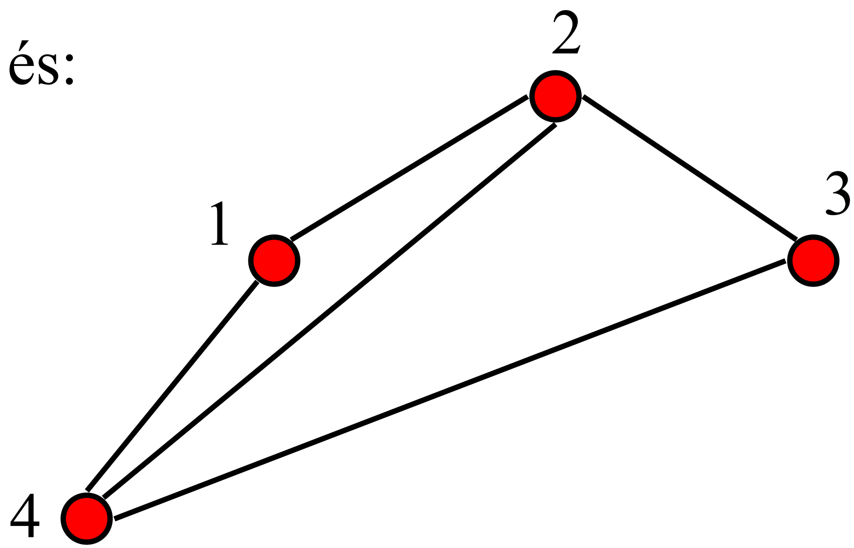
Sumatori pels veïns del node i-èsim en la iteració anterior

•Durant el nostre estudi, intentarem esbrinar quines relacions existeixen entre el Page Rank d'un node i alguna de les mesures anteriorment definides com la internodalitat nodal... .

EXPLICACIÓ DETALLADA CÀLCUL VECTOR PAGE RANK

La matriu d'adjacència és:

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



Suposant que un cop estem situats a un determinat vèrtex tenim igual probabilitat de seleccionar qualsevol enllaç sortint podem construir la matriu de probabilitats P:

$$P = \begin{pmatrix} 0 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \\ 1/2 & 1/3 & 1/2 & 0 \end{pmatrix}$$

La construcció de la matriu P, la podem formalitzar de forma generalitzada tal i com segueix:

$$c_j = \sum_{i=1}^{n} a_{ij}$$

$$p_{ij} = \begin{cases} a_{ij} / c_j & si \;\; c_j \neq 0 \\ 0 \;\; en \;\; altre \;\; cas \end{cases}$$

La matriu P és una matriu estocàstica per columnes, és a dir, la suma de cada columna val 1:

$$P = \begin{pmatrix} 0 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \\ 1/2 & 1/3 & 1/2 & 0 \end{pmatrix}$$

$C_1 = \sum a_{i1} = 1/2 + 1/2 = 1$

$C_2 = \sum a_{i2} = 1/3 + 1/3 + 1/3 = 1$

$C_3 = \sum a_{i3} = 1/2 + 1/2 = 1$

$C_4 = \sum a_{i4} = 1/3 + 1/3 + 1/3 = 1$

Una matriu **primitiva** es pot caracteritzar com una matriu irreductible i **no negativa**, amb algun element de la diagonal no nul.

Això però, no ens assegura que tingui un valor propi dominant ja que no tots els elements de la matriu són diferents de zero i per tant la matriu **no és primitiva**.

Una matriu A és **primitiva** quan al realitzar la potència (k) de la matriu tots els elements (a_{ij}) prenen valors positius.

D'aquesta forma ens assegurem que la matriu sigui irreductible.

$$A^k \Rightarrow \forall\, a_{ij} > 0$$

És per això que definim una nova matriu estocàstica de probabilitats que ens determina el fet que podem fer un salt cap a qualsevol altra pàgina de la xarxa amb la mateixa probabilitat :

e_{ij}=1/n

$$E=\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \Rightarrow graf\ \ complet$$

El coeficient de salt **d,** també anomenat vector de personalització ens permetrà trobar una combinació entre la connectivitat real de la xarxa (d=1) i una connectivitat totalment aleatòria (d=0):

$$(1-d)\cdot E=\begin{pmatrix} (1-d)/4 & (1-d)/4 & (1-d)/4 & (1-d)/4 \\ (1-d)/4 & (1-d)/4 & (1-d)/4 & (1-d)/4 \\ (1-d)/4 & (1-d)/4 & (1-d)/4 & (1-d)/4 \\ (1-d)/4 & (1-d)/4 & (1-d)/4 & (1-d)/4 \end{pmatrix}$$

$$\overline{1-d} \quad \overline{1-d} \quad \overline{1-d} \quad \overline{1-d}$$

$$d\cdot P=\begin{pmatrix} 0 & d/3 & 0 & d/3 \\ d/2 & 0 & d/2 & d/3 \\ 0 & d/3 & 0 & d/3 \\ d/2 & d/3 & d/2 & 0 \end{pmatrix}$$

$$\overline{d} \quad \overline{d} \quad \overline{d} \quad \overline{d}$$

Definim la matriu P' com:

$$\mathbf{P'=P\cdot d+(1-d)\cdot E}$$

El terme **(1-d)·E** dóna lloc a que tots els elements de **P'** siguin no nuls, per tant **P'** serà una matriu irreductible. L'efecte d'aquest terme és introduir salts aleatoris que no depenen de les propietats d'enllaç de la xarxa en sí.

•Valors de **d** pròxims a 1 ofereixen comportaments més realistes però poden arruinar la irreductibilitat i augmentar el nombre d'iteracions necessàries del mètode de la potència.

P':Matriu estocàstica i primitiva → ens garanteix l'existència d'un valor propi estrictament dominant:

(λ=1) de mòdul el radi espectral ρ(P')=1.

nota: si no tenim nodes sense sortida no cal fer cap més transformació (el graf és no dirigit i connex per tant ja està)

INTROD. A LES XARXES
GRAU VÈRTEX
GRAU MITJÀ PRIMERS VEÏNS
COEFICIENT DE AGRUPAMENT
CAMINS MÍNIMS
MESURA CENTRALITAT
MESURA PROXIMITAT
ACCÉS A LA XARXA
OCULTACIÓ NODAL
INTERNODALITAT
BETWEENNESS ARESTES
FUNNELING
PAGE RANK
CONCLUSIONS

$$P' = \begin{pmatrix} (1-d)/4 & [(1-d)/4]+d/3 & (1-d)/4 & [(1-d)/4]+d/3 \\ [(1-d)/4]+d/2 & (1-d)/4 & [(1-d)/4]+d/2 & [(1-d)/4]+d/3 \\ (1-d)/4 & [(1-d)/4]+d/3 & (1-d)/4 & [(1-d)/4]+d/3 \\ [(1-d)/4]+d/2 & [(1-d)/4]+d/3 & [(1-d)/4]+d/2 & (1-d)/4 \end{pmatrix}$$

$$\overline{1} \qquad \overline{1} \qquad \overline{1} \qquad \overline{1}$$

$$\vec{X}_{t+1} = P' \cdot \vec{X}_t = P' \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}_t$$

Aquest vector columna inicialment per (t=0) ha de ser estocàstic també.

Quan anem iterant i els valors de x_1,x_2,x_3 i x_4 convergeixin cap a un valor ja haurem trobat els valors de *Page Rank* PR1, PR2, PR3, PR4.

El procés iteratiu en aquest exemple en concret serà tal i com segueix:

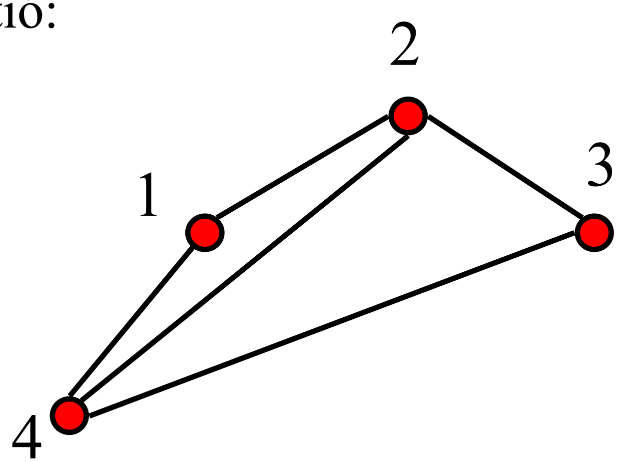
$$(x_1)_{t+1} = \frac{(1-d)}{4} + d\cdot\left(\frac{(x_2)_t + (x_4)_t}{3}\right)$$

$$(x_2)_{t+1} = \frac{(1-d)}{4} + d\cdot\left(\frac{(x_1)_t + (x_3)_t}{2} + \frac{(x_4)_t}{3}\right)$$

$$(x_3)_{t+1} = \frac{(1-d)}{4} + d\cdot\left(\frac{(x_2)_t + (x_4)_t}{3}\right)$$

$$(x_4)_{t+1} = \frac{(1-d)}{4} + d\cdot\left(\frac{(x_1)_t + (x_3)_t}{2} + \frac{(x_2)_t}{3}\right)$$

Els resultats de Page Rank obtinguts de l'aplicació del mètode de la potència al problema amb qüestió:



$$\mathbf{PR(v_1) = PR(v_3) = 20{,}48\%}$$

$$\mathbf{PR(v_2) = PR(v_4) = 29{,}52\%}$$

Altres resultats que es deriven de l'estudi de la xarxa:

centralitat nodes (v_1 i v_3)= 4/3
centralitat nodes (v_2 i v_4)= 1

clustering nodes (v_1 i v_3)= 1
clustering nodes (v_2 i v_4)= 2/3

informació d'accés nodes (v_1 i v_3)= 1,33 bits
informació d'accés nodes (v_2 i v_4)= 1,58 bits

informació localitz. nodes (v_1 i v_3)= 1,72 bits
informació localitz. nodes (v_2 i v_4)= 1,19 bits

grau mitjà primers veïns nodes (v_1 i v_3)= 2
grau mitjà primers veïns nodes (v_2 i v_4)= 4/3

RECABLEJAT per aconseguir xarxes aleatoritzades

•En la xarxa aleatoritzada també anomenada "xarxa nul·la o model nul", tots els vèrtexs tenen el mateix grau que en la xarxa original, però les connexions entre ells estan aleatoritzades.

APLICACIONS

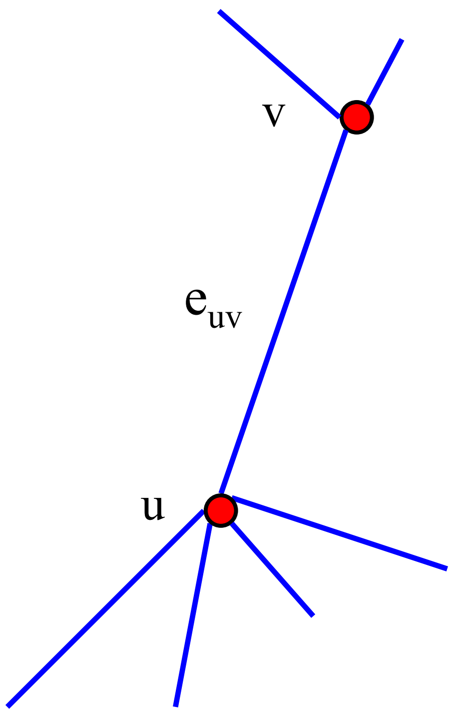
• La probabilitat $P(k_0,k_1)$ que dos vèrtexs amb grau k_0 i k_1 estiguin connectats amb una aresta, es compara amb la probabilitat $P_r(k_0,k_1)$ del model aleatoritzat de la mateixa xarxa.

•Maslov,S. & Sneppen,K. *Specificity and stability in topology of protein networks,* Science, **296**,910-913 (2002a)

•L'algoritme de recablejat conserva doncs la distribució de graus de la xarxa (conserva el grau de cada vèrtex individual).

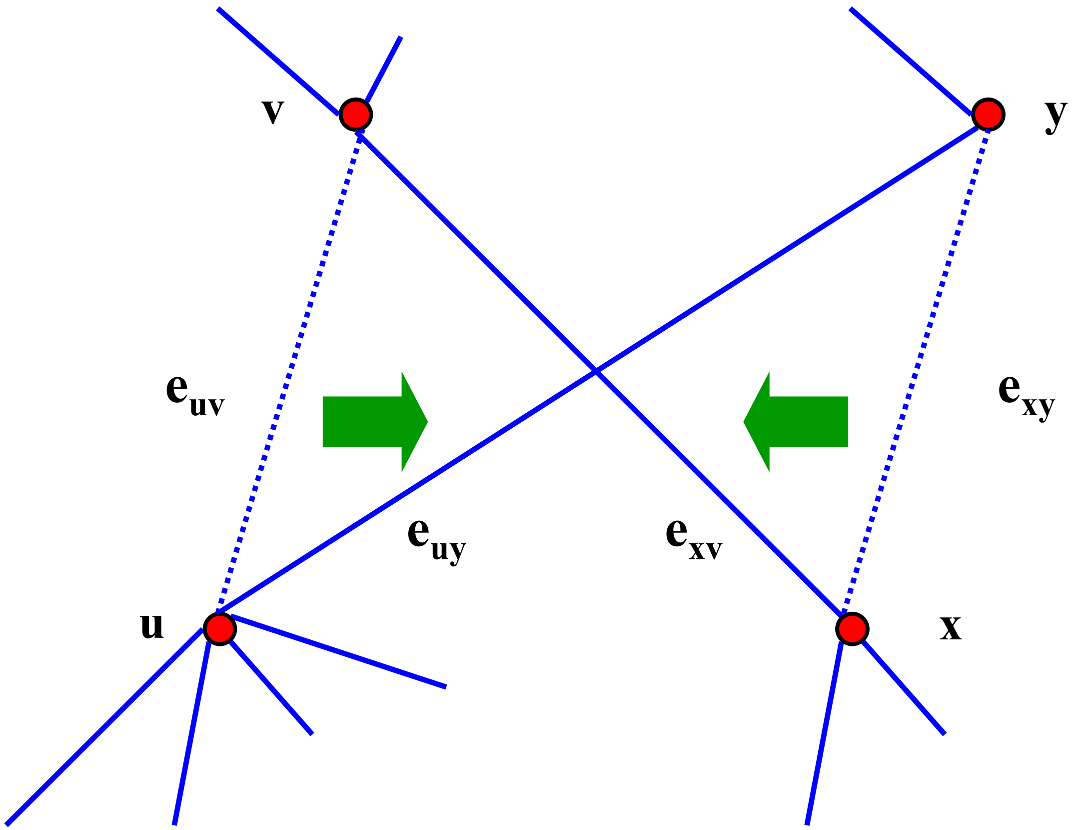
PROCEDIMENT OPERATIU

1.- Seleccionem de forma aleatòria una parella d'arestes:



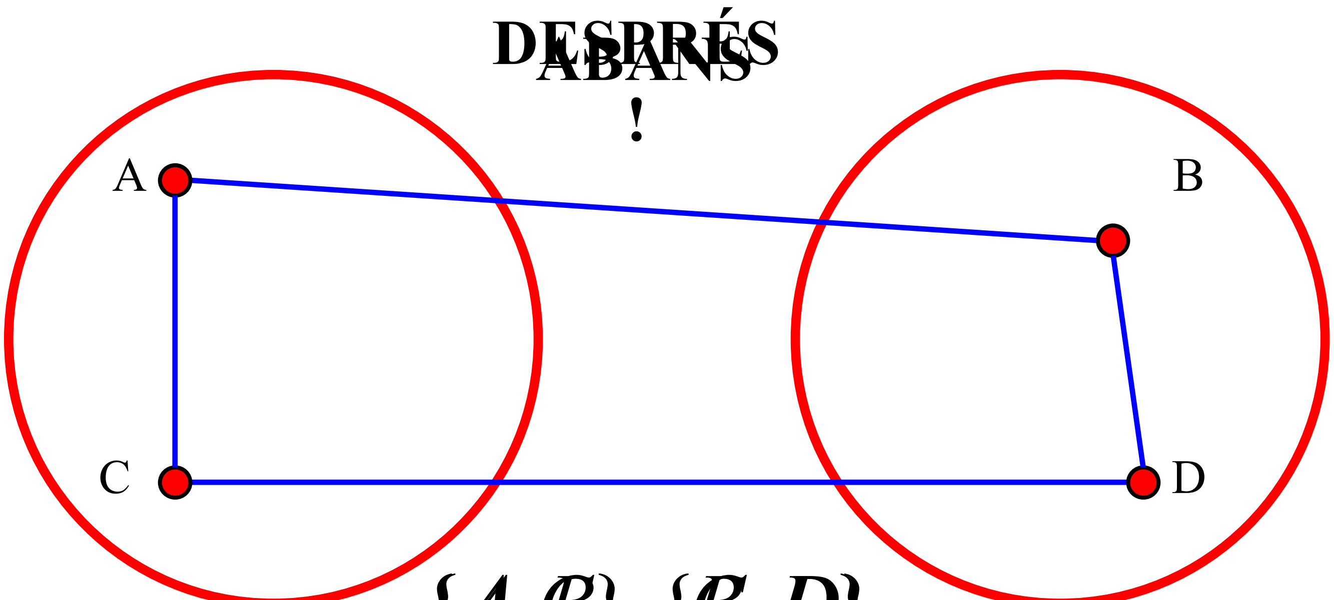
atenció: no seleccionar arestes que tinguin nodes en comú.

2.- Fem la reconnexió tot evitant la creació de links ja existents a la xarxa. D'aquesta forma evitarem l'aparició d'arestes múltiples connectant la mateixa parella de nodes. En el cas de que això passi reinicialitzarem el procés seleccionant dues noves arestes.



3.- Anem iterant el procés exposat fins que la informació inicial de la xarxa s'hagi perdut.

4.- Un cop realitzat el recablejat, pot passar que les modificacions introduïdes hagin generat l'aparició de vàries components connexes independents (mòduls). Així doncs, a posteriori escollirem la xarxa aleatòria com a vàlida si només si la xarxa resultant té el mateix nombre de components que la xarxa inicial.



$$\{A,B\},\{B,D\}$$

•**L'algoritme de recablejat** proposat té el següent problema que tot seguit comentem:

Es pot establir una tendència a que nodes de grau elevat estiguin connectats alhora amb nodes de grau elevat.

Això fa que la petjada de la correlació de graus no s'acabi de trencar. Això és degut al fet que un node de grau elevat té més arestes i per tant té també una probabilitat més elevada de ser escollit. Aquest fet provocarà que nodes de grau elevat tinguin la tendència de ser recablejats entre ells.

DIÀMETRE DE LA XARXA (ø)

•En un graf el diàmetre és la major distància entre dos punts del mateix.

$$\phi = \max(d\{v_i, v_j\})$$

Essent i≠j i tenint un total de [n·(n-1)/2] combinacions possibles de parelles de vèrtexs a considerar.

•Si obtenim diàmetre de la xarxa infinit poden estar passant dues coses: o bé el graf no és connex o bé aquest té infinitat de vèrtexs.

CONCLUSIONS

L'estudi que es durà a terme en aquest PFC pretén ser:

-Una anàlisi **comparativa** entre diferents xarxes de transport que alhora presenten topologies diverses en quan a mobilitat.

-Un estudi de les propietats i singularitats de les diferents xarxes estudiades.

-Un punt de partida per a l'estudi de les **lleis d'organització** d'aquestes xarxes, fins a dia d'avui unes lleis molt desconegudes i que van més enllà de les lleis d'interacció.

-Aprendre en definitiva les **eines de treball** d'un nou àmbit d'estudi (les xarxes complexes i l'estudi dels anomenats 'small worlds') que abarca molts camps i disciplines diferents.

Universitat de Girona

PROJECTE FI DE CARRERA

Estudi de la mobilitat en tres xarxes de transport públic

Titulació: Enginyeria Industrial (pla 2002)

Alumne: **Josep Barberillo Nualart**
(u1051158@correu.udg.edu)

Director: **Joan Saldaña Meca**
(jsaldana@ima.udg.es)

Departament d'Informàtica i Matemàtica Aplicada

EPS
Escola Politècnica Superior