

## RECUPERACIÓ DE LA INFORMACIÓ

Data: 26 de gener de 2017

Control 2

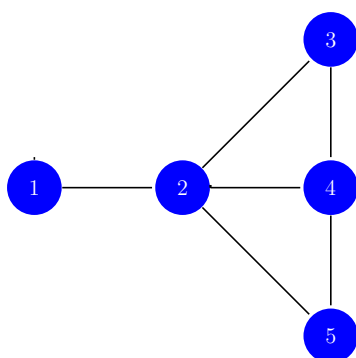
Temps: 2 hores

### Problema 1 [3 punts]

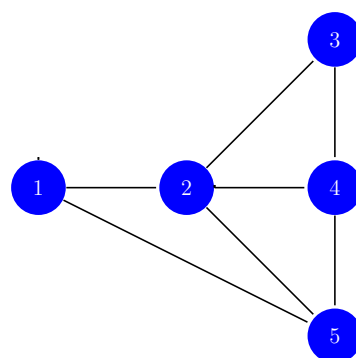
- Explica com es construeix una xarxa aleatòria usant el model Erdős-Rényi (ER).  
fixem una quantitat de nodes i una probabilitat de connexió
- Cita almenys tres propietats que diferencien una xarxa aleatòria d'una xarxa "real" i justifica el perquè.  
coeficient de cluster gran, lliure d'escala, distribució de grau,

### Problema 2 [4 punts]

Calcula els valors de centralitat de grau i el coeficient de clustering de cadascun dels nodes de la xarxa 1.



(1)



(2)

Respon les preguntes següents:

- Dos vèrtexs poden tenir el mateix grau però diferent coeficient de clustering? Justifica la teva resposta. **mesuren coses diferents (es a dir que no hi ha correlació)**
- A partir dels valors d'intermediació (*betweenness*) de les arestes de la xarxa 2, aplica un algorisme que et permeti detectar 2 comunitats.

m	cg	cc
①	1	0
②	4	$\frac{2}{\frac{1}{2} \cdot 4 \cdot 3} = \frac{2}{6} = 1/3$
③	2	$\frac{1}{\frac{1}{2} \cdot 2 \cdot 1} = \frac{1}{1} = 1$
④	3	$\frac{2}{\frac{1}{2} \cdot 3 \cdot 2} = \frac{2}{3}$
⑤	2	$\frac{1}{\frac{1}{2} \cdot 2 \cdot 1} = 1$

#connexions entre vèrtexs i

**Problema 3** [3 punts]

Et demanen de fer un recompte dels visitants **únics** d'un lloc web a partir de les dades del seu llarg fitxer de registres (*logfile*).

Suposa que el format d'aquest fitxer és el següent

DATE	siteID	action	username
26-01-2017	siteA	pageview	joan
26-01-2017	siteB	pageview	tere
26-01-2017	siteA	pageview	joan
26-01-2017	siteB	pageview	berta
26-01-2017	siteA	pageview	miquel
...			

i que es vol obtenir un llistat amb la quantitat de visitants únics de cada *site*. Descriu com resoldries aquesta tasca usant el model MapReduce.