

Estudi sobre la transició de fase per les components connexes de diferents espècies de grafs sota un procés de percolació

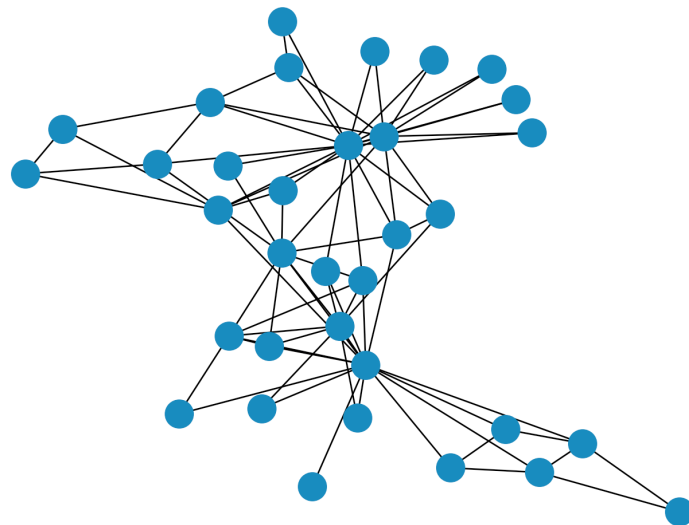
Alumnes:

Jordi Calsina Baldomà

Guido Biosca Lasa

Francisco Badia Laguillo

Àlex González Fernández



*Aquesta investigació es presenta com a part de l'assignatura d'Algorísmia
del Grau d'Enginyeria Informàtica de la UPC a la FIB*

Contents

1	Introducció	2
1.1	Objectiu	2
1.2	Metodologia	2
1.3	Problemes enfrontats	3
1.4	Generació de grafs	3
1.5	Sobre el nombre de proves necessàries	4
2	Estudi sobre els Grafs Graella	6
2.1	Percolació per vèrtexs	6
2.1.1	Dades descriptives	6
2.1.2	Anàlisi de les dades	10
2.2	Percolació per arestes	14
2.2.1	Dades descriptives:	14
2.2.2	Anàlisi de les dades	15
2.3	Conclusions:	16
3	Estudi sobre els Grafs Graella 3D	18
3.1	Percolació per vèrtexs	18
3.1.1	Dades descriptives	18
3.1.2	Anàlisi de les dades	20
3.2	Percolació per arestes	22
3.2.1	Dades descriptives	22
3.2.2	Anàlisi de les dades	24
3.3	Conclusions:	26
4	Estudi sobre els Random Geometric Graphs	27
4.1	Definicions i propietats	27
4.2	Procediment Experimental	27
4.2.1	Determinació de la k per a tenir significació estadística:	28
4.2.2	Generació de la mostra d'RGGs:	30
4.2.3	Paràmetres experimentals:	31
4.3	Resultats de l'experimentació:	31
4.4	Estudi del <i>knee of a curve</i> :	32
4.5	Conclusions experimentació:	32
5	Estudi sobre els Connected Caveman Graph	34
5.1	Estudi per vèrtexs	36
5.1.1	Dades descriptives	36
5.1.2	Anàlisi de les dades	39
5.2	Estudi per arestes	40
5.2.1	Dades descriptives	40
5.2.2	Anàlisi de les dades	43
5.3	Conclusions:	44
6	Conclusions del treball:	46

1 Introducció

1.1 Objectiu

L'objectiu de l'estudi consisteix en l'anàlisi de la relació i transicions de fase, entre el nombre de components connexos de certs grafs sota estudi, quan els vèrtexs i les arestes poden fallar.

L'estudi engloba quatre espècies de grafs: Grafs graella bidimensionals quadrats, grafs graella tridimensionals cúbics, els grafs aleatoris connexos o *Random Geometric Graphs* (RGG), i els *Connected Caveman Graphs*. L'estudi es porta a terme en les seccions (2.1), (2.2), (2.3), (2.4) respectivament.

1.2 Metodologia

Per tal d'observar la relació entre el nombre de components connexos sota un procés de percolació en vèrtexs o arestes en grafs de certes espècies, generem les següents gràfiques tant per eliminació de vèrtexs com arestes, com en la figura (1) amb un graf graella amb 1.000.000 nodes.

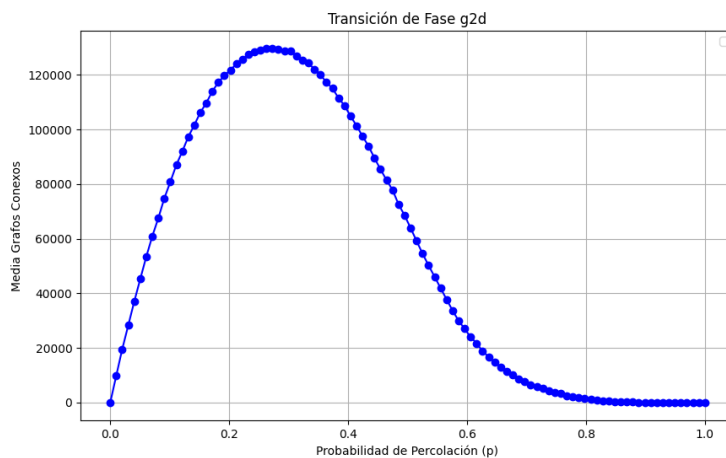


Figura 1:

L'eix de les abscisses representa la probabilitat p del procés de percolació, o sigui, la probabilitat de supervivència dels vèrtexs o arestes respectivament; i l'eix de les ordenades representa la mitjana de components connexos que té un graf de l'espècie que estem tractant.

Exactament, per tal de calcular aquesta mitjana, agafem un graf arbitrari de l'espècie amb certs paràmetres fixats, eliminem els vèrtexs o arestes amb probabilitat $1 - p$, i calculem el nombre de components connexos resultants, repetim el procés K cops en el mateix graf, i finalment en calculem la mitjana. Notem que per les diverses probabilitats

p fem servir el mateix graf per a calcular el nombre de components connexos mig.

1.3 Problemes enfrontats

Per tal de calcular les dades esmentades en el paràgraf anterior, la nostra idea original era calcular-les mitjançant un programa implementat en C++, gairebé sense fer-ne ús de cap llibreria, per tal de tenir un gran control sobre l'eficiència i estructures de dades. Però ràpidament vam veure que, tot i que el temps d'espera no era completament nefast, el temps de càlcul de les dades no era suficient.

Llavors, vam optar per escriure el codi en Python, donada la diversitat de llibreries en relació amb grafs i graficació de dades; però un factor important va ser poder admetre *multiprocessing* per tal d'accelerar el càlcul de les dades. Al següent paràgraf detallarem les llibreries de Python usades. Després de programar els codis necessaris amb Python, vam notar una millora molt considerable en el temps de càlcul de les dades.

Després de programar els codis necessaris amb Python, vam notar una millora molt considerable en el temps de càlcul de les dades. Però no prou contents amb això, vam optar, per encara millora més la velocitat de càlcul, per l'ús d'un servidor amb Google Cloud, aconseguint 32 nuclis virtuals per paral·lelitzar la generació i percolació de grafs. A més, vam disposar de 150 GB de RAM per suportar la creació de grans grafs. Encara que realment no és necessària tanta memòria, la vam assignar per si de cas, ja que un dels principals desavantatges del llenguatge Python és el seu considerable consum de memòria.

A la secció 7, la bibliografia, enumerem les diverses llibreries usades en el projecte.

1.4 Generació de grafs

Per a cada espècie de graf (que recordem són: grafs graella bidimensionals quadrats i tridimensionals cúbics, RGG, i *Connected Caveman Graphs*), generem grafs arbitraris per tal de calcular les dades.

En el cas dels grafs graella bidimensionals i tridimensionals quadrats, la implementació és prou senzilla, on l'únic paràmetre és N , que representa la mida de l'aresta del quadrat o cub; de forma que hem optat per implementar nosaltres mateixos la generació dels grafs.

Tanmateix, en el cas dels RGG i *Connected Caveman Graphs*, hem optat per generadors ja implementats, donada la llibreria NetworkX.

1.5 Sobre el nombre de proves necessàries

Donat p , notem per k el nombre de proves que fem per calcular la mitjana de components connexes del graf percolat (G, p) , que notarem per \bar{x} . Pot resultar interessant la pregunta: “Quin valor ha de tenir almenys k per tal de tenir significança estadística en \bar{x} ?” Per resoldre això aplicarem intervals de confiança, de forma que, en primer lloc, donat (G, p) , comprovarem si la variable $X =$ ‘nombre de components connexes després de percolar el graf (G, p) , compleix que $X \sim N(\mu, \sigma^2)$ amb alguns paràmetres desconeguts.

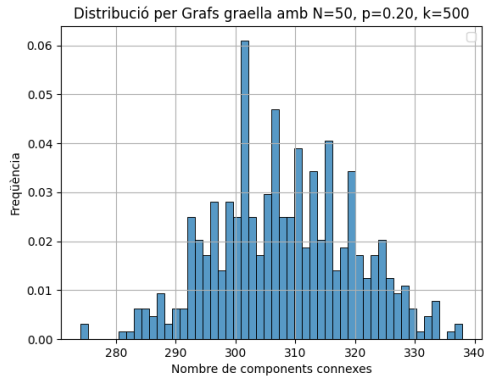
Les figures (1.1.1) fins a la (1.1.8) són histogrames, on l’eix de les abscisses es representen els ‘nombre de components connexes després de perlocar’ agrupat en intervals, i en l’eix de les ordenades la freqüència de les dades en cada interval. És clar que amb algunes k petita, com en la figura (1.1.1), la hipòtesi de Normalitat no s’acaba de complir. Tot i que amb k , com en la figura (1.1.6) sí que es compleix, però en altres casos com en la figura (1.1.8) no es compleix. De forma que no podem concloure, en general, que per a qualsevol p , la variable aleatòria X segueix una distribució Normal.

No obstant això, podem aplicar el Teorema Central Límit, que ens diu que per a qualsevol distribució d’una variable aleatòria X amb variància finita, \bar{X} tendeix a una distribució Normal. D’aquesta forma, si considerem una K gran, com $K \geq 100$, podríem calcular un interval amb confiança α , on com la desviació estàndard de la població resulta desconeguda, hauríem d’aplicar la fórmula:

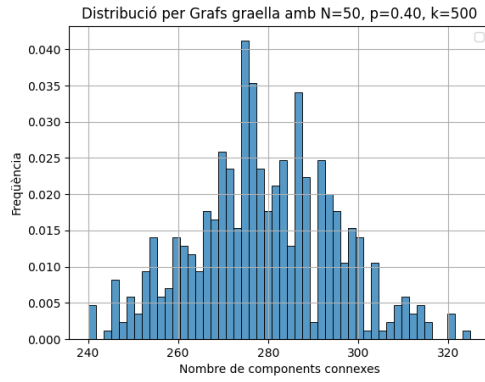
$$IC = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Si agafem K de forma que el marge d’error sigui petit, tindrem una molt bona acotació de la mitjana real amb una confiança bastant bona. Però cal notar que, evidentment, el marge d’error de cada interval amb confiança α no serà igual, ja que la desviació estàndard de la mostra depèn del paràmetre p .

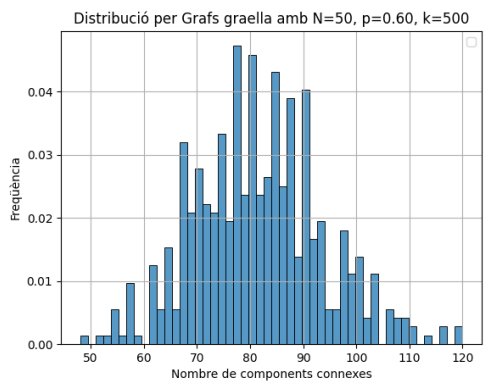
No obstant això, ometrem el càlcul dels intervals de confiança de l’estudi, donat que considerem que no aportarà cap mena de gruix a l’estudi, ja que hi ha qüestions i preguntes més interessants.



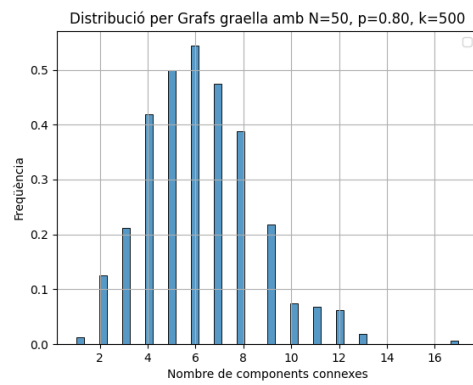
(a) Figura 1.1.1



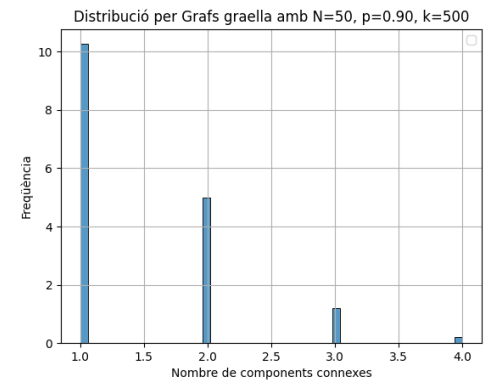
(b) Figura 1.1.2



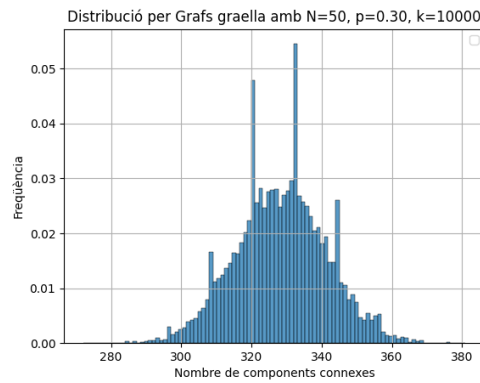
(c) Figura 1.1.3



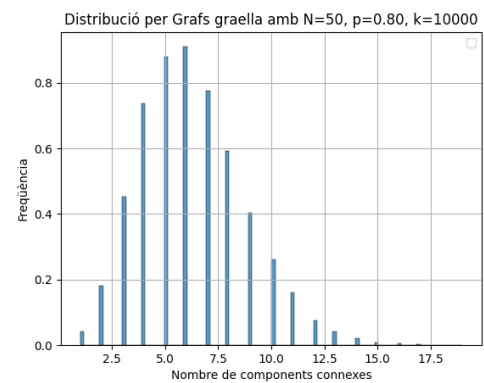
(d) Figura 1.1.4



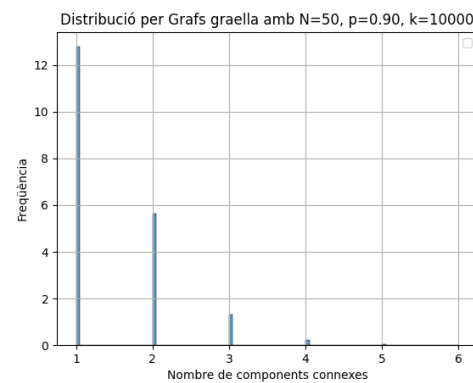
(e) Figura 1.1.5



(f) Figura 1.1.6



(g) Figura 1.1.7



(h) Figura 1.1.8

2 Estudi sobre els Grafs Graella

Definim un Graf Graella quadrat de mida N^2 com un conjunt de vèrtexs V i arestes E , $G = (V, E)$, on $V = \{a_{11}, a_{12}, \dots, a_{1N}, a_{21}, \dots, a_{2N}, \dots, a_{N1}, \dots, a_{NN}\}$, i $(a_{ij}, a_{kq}) \in E$ sí només sí $(i = k \pm 1 \text{ i } j = q)$ ó $(i = k \text{ i } j = q \pm 1)$.

La generació d'aquesta espècie de grafs és senzilla, ja que només canvia el valor del paràmetre N , que recordem que implica que el graf té N^2 nodes.

2.1 Percolació per vèrtexs

2.1.1 Dades descriptives

En les següents figures, donat un graf graella amb paràmetre N , l'eix de les abscisses representa la probabilitat amb la qual un vèrtex sobreviu, i l'eix de les ordenades representa el nombre mitjà de components connexes després de percolar el graf amb probabilitat p i calcular la mitjana a partir de k proves.

La figura (2.1.1) mostra la gràfica resultant d'aplicar el procediment esmentat sobre un graf graella amb $N = 100$, amb valors $p = 0, 0.01, 0.02, \dots, 1.0$, $k = 100$. Dit d'altra forma, calculem la mitjana de components connexes a partir de 100 proves del mateix graf percolat amb el mateix valor p .

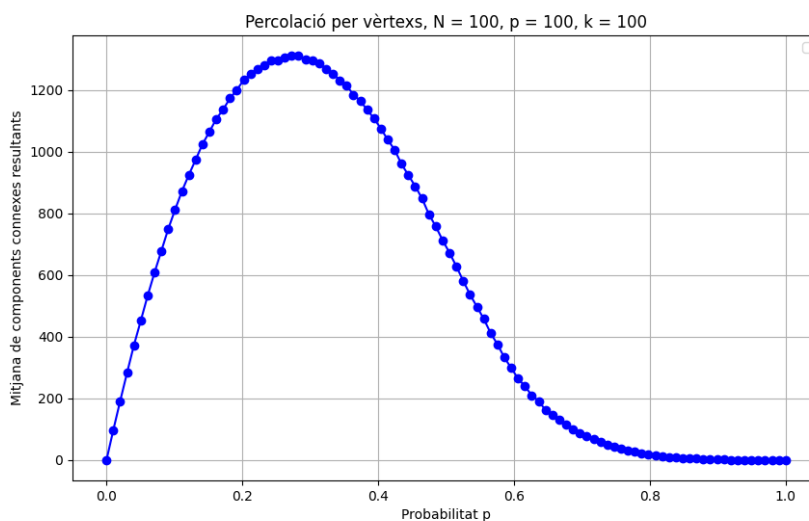


Figura 2.1.1

Pot resultar d'interès veure com evolucionen aquestes gràfiques a mesura que incrementem el paràmetre N .

La figura (2.1.2) mostra les gràfiques amb $N = 10, 20, \dots, 70, 80$.

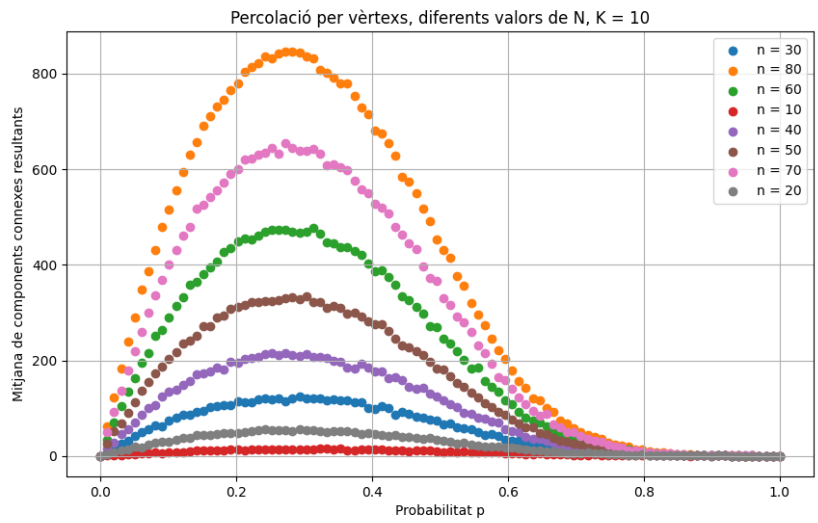


Figura 2.1.2

La figura (2.1.3) mostra les gràfiques amb $N = 5, 10, 15, \dots, 95, 100$.

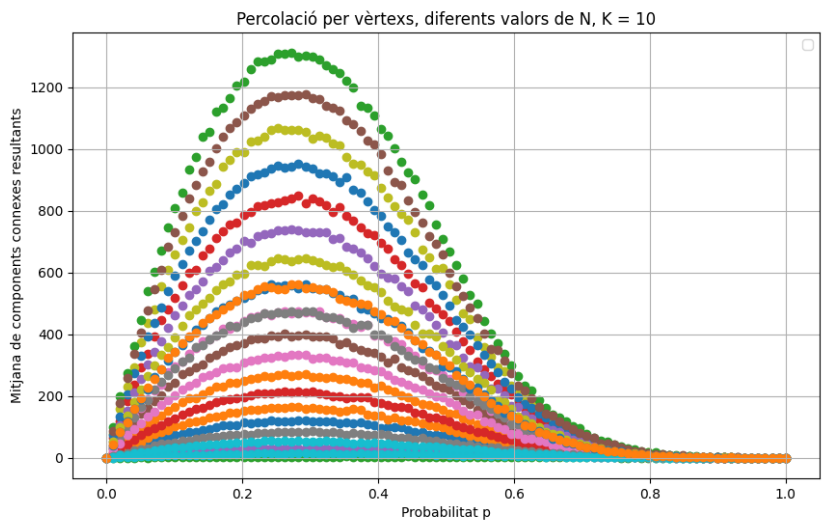


Figura 2.1.3

En últim lloc, per tal de portar les nostres màquines a l'extrem, la figura (2.1.4) mostra les gràfiques amb $N = 5, 10, 15, \dots, 295, 300$.

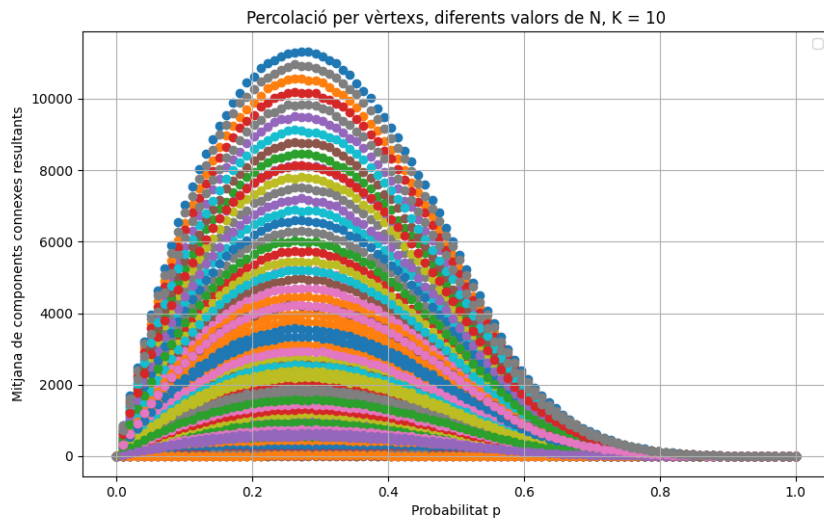


Figura 2.1.4

Sorgeixen preguntes sobre el màxim nombre de components connexes que cada gràfic arriba a assolir. La figura (2.1.5) mostra els nombres de components connexes màximes que s'assoleixen en funció de N , a partir de les dades de la figura (2.1.4).

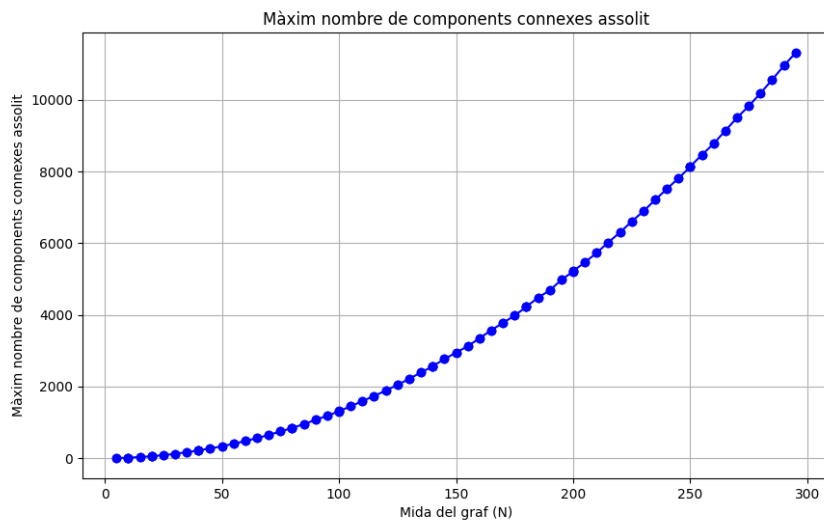


Figura 2.1.5

També, la figura (2.1.6) mostra amb quina probabilitat de supervivència p s'assoleix el màxim nombre de components connexes, en funció de N , a partir de les dades de la figura (2.1.4).

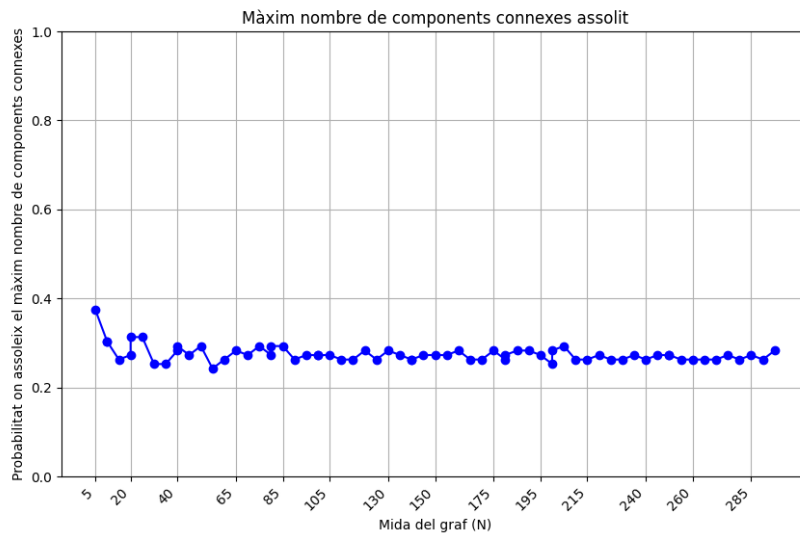


Figura 2.1.6

Notarem la probabilitat p'_N , com la probabilitat on s'assoleix aquest màxim nombre de components connexes per un graf graella de N^2 nodes. Notarem C_N com el màxim nombre de components connexes que s'assoleixen en un graf graella de N^2 nodes, després de percolar-lo.

2.1.2 Anàlisi de les dades

Sobre el comportament de la funció En les figures (2.1.1) a la (2.1.4) observem que la gràfica no és decreixent. Això pot resultar contraintuïtiu, ja que es podria pensar que a mesura que augmentem la probabilitat de supervivència, s'espera que el nombre de components connexes disminueixi. Però això no és cert. Anem a veure com és que això succeeix:

1) Per què la gràfica comença amb valors propers a 0?

Sigui p la probabilitat de supervivència, i N la mida del graf graella $G = (V, E)$. Sigui $G' = (V', E')$ el graf resultant després d'aplicar la percolació, es compleix: $\text{mitjana}(|V'|) = |V| \cdot p = N^2 \cdot p$ A més, sigui $C_{G'}$ el nombre de components connexes de G' , es compleix que:

$$\text{mitjana}(C_{G'}) \leq \text{mitjana}(|V'|) = N^2 \cdot p$$

De manera que per probabilitats de supervivència petites, necessàriament la mitjana de $C_{G'}$ és petita. En altres paraules, quan la probabilitat de supervivència és molt baixa, gairebé no queden vèrtexs actius, per tant, gairebé no tenim components connexes. Gràcies a l'argument anterior, raonablement podem admetre la següent hipòtesi:

Hipòtesi 1 En mitjana, els punts de la gràfica queden acotats superiorment per la funció $f(p) = N^2 \cdot p$.

La figura (2.1.10) i (2.1.11) amb paràmetres $N=295$ i 30 , $K=10$ ambdós, i 100 i 1000 probabilitats, respectivament; mostren la funció $f(p) = N^2 \cdot p$ graficada pel damunt de les dades. Podem concloure de forma raonable que la hipòtesi sembla ser certa.

2) Per què la gràfica creix fins p'_N ?

En aquest apartat notarem p'_N per p' . Sigui una graella de mida considerablement gran, amb $N = 100$, de manera que la graella tingui $|V| = 10.000$ nodes. És clar que, si les probabilitats de supervivència són molt baixes, com 0.05 , després de percolar només quedarien actius uns 500 nodes de mitjana. Considerem raonablement que aquests 500 nodes resulten completament aïllats entre si. Per tant, podem formular la següent hipòtesi:

Hipòtesi 2: Sigui $G' = (V', E')$ el graf percolat i $C_{G'}$ el nombre de components connexes de G' . Per a probabilitats de supervivència molt petites, es compleix que $\text{mitjana}(|V'|) \approx \text{mitjana}(C_{G'})$. Equivalentment, per probabilitats molt petites p , el comportament de la funció respecte a p és: $f(p) = N^2 \cdot p$ És a dir, el comportament és lineal.

Per tal de comprovar aquesta hipòtesi, dibuixarem la recta $f(p) = N^2 \cdot p$ damunt de les dades. En la figura (2.1.10) les dades resulten dels paràmetres $N=295$, $K=10$, 100 probabilitats, i en la figura (2.1.11) les dades resulten dels paràmetres $N=30$, $K=10$, 1000 probabilitats. La recta sembla ajustar-se només en els punts inicials amb probabilitats

molt petites, de forma que no podem concloure que la hipòtesi sigui correcta.

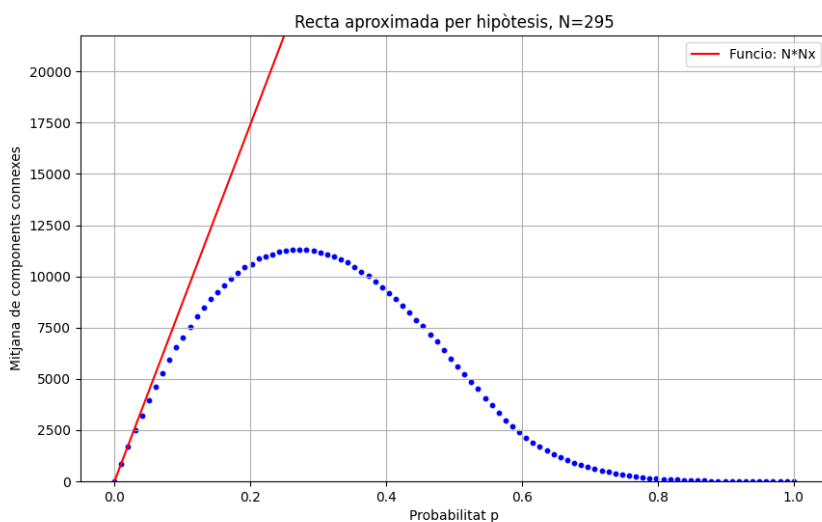


Figura 2.1.10:

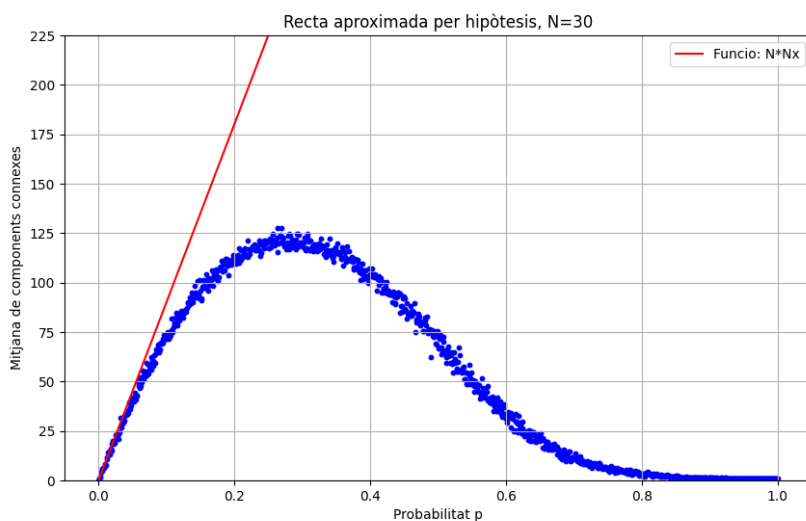


Figura 2.1.11:

Seguidament, intentem deduir el valor de p' , la probabilitat on s'assoleix el màxim nombre de components connexes. Considerem $p_1 = \frac{1}{4}$. Podem agrupar els nodes en creus, tal com es mostra en la figura (2.1.7). És raonable acceptar que el cas més probable és aquell on el node actiu és al centre de la creu. A mesura que augmentem $p > p_1$, el nombre de components connexes tendeix a disminuir, ja que dues components poden quedar unides.

Hipòtesi 3: La probabilitat on s'assoleix el màxim nombre de components connexes p' s'aproxima a $\frac{1}{4}$.

Aquesta hipòtesi la comprovem en un dels següents apartats mitjançant la figura (2.1.9).

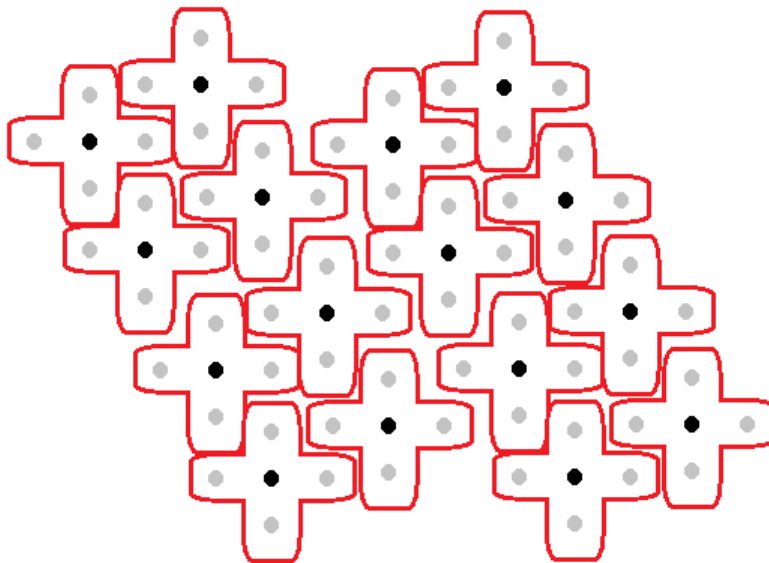


Figura 2.1.7:

Sobre el màxim nombre de components en funció de N Recordem que notem la probabilitat p'_N , com la probabilitat on s'assoleix el màxim nombre de components connexes per un graf graella de N^2 nodes, i C_N com el màxim nombre de components connexes que s'assoleixen en un graf graella de N^2 nodes després de percolar-lo.

Diverses preguntes sorgeixen sobre C_N i p'_N : Quina relació existeix entre N i C_N ? Existeix una funció senzilla que approximi aquesta relació? És p'_N constant? Resolguem aquestes qüestions:

La figura (2.1.5) mostra C_N en funció de N . Després d'aplicar un algoritme de regressió polinòmica de segon grau, obtenim la figura (2.1.8).

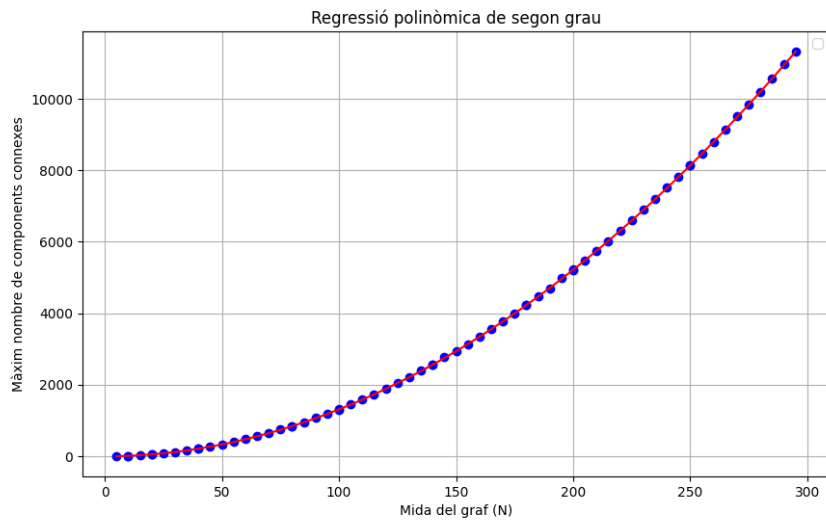


Figura 2.1.8:

El polinomi és de la forma:

$$0.12939287N^2 + 0.20368617N + 0.74172252$$

Aquest ajust sembla perfecte a les dades.

Sobre la probabilitat p'_N La figura (2.1.6) mostra la relació entre p'_N i N . Si assumim que p'_N és constant, calcularem la mitjana i la mediana. Els valors són 0.2750 (mitjana) i 0.2727 (mediana), aproximant-se a la Hipòtesi 2. La figura (2.1.9) mostra ambdues rectes.

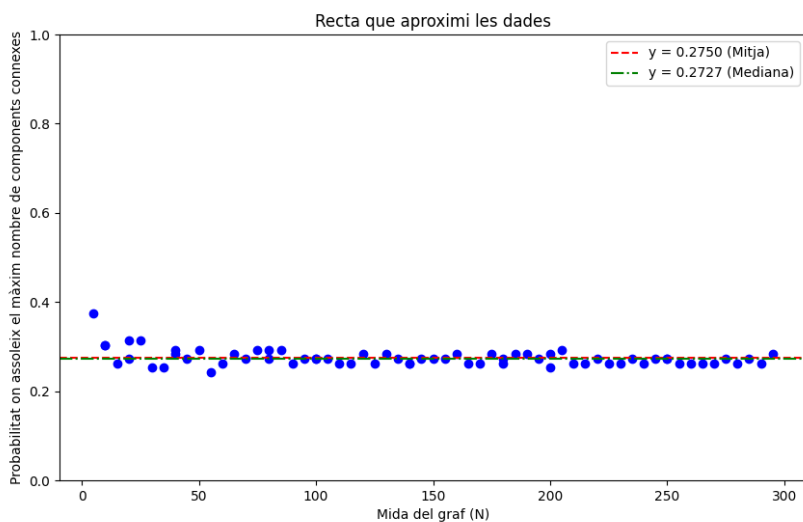


Figura 2.1.9:

2.2 Percolació per arestes

2.2.1 Dades descriptives:

La figura (2.2.1) mostra la gràfica resultant de percolar per arestes sobre un graf graella amb $N = 100$, amb valors $p = 0, 0.01, 0.02, \dots, 1.0, k = 100$.

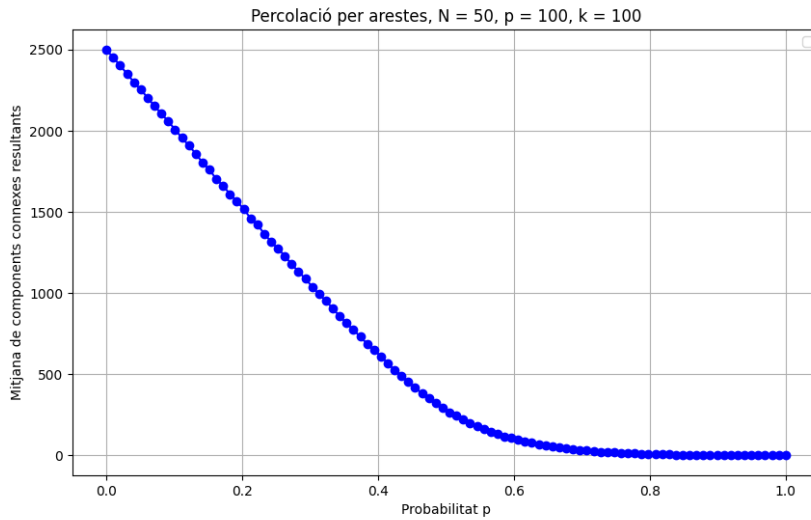


Figura 2.2.1:

Pot resultar d'interès veure com evolucionen aquestes gràfiques a mesura que incrementem el paràmetre N . En la figura (2.2.2) mostra les gràfiques amb $N = 5, 10, 15, \dots, 50$.

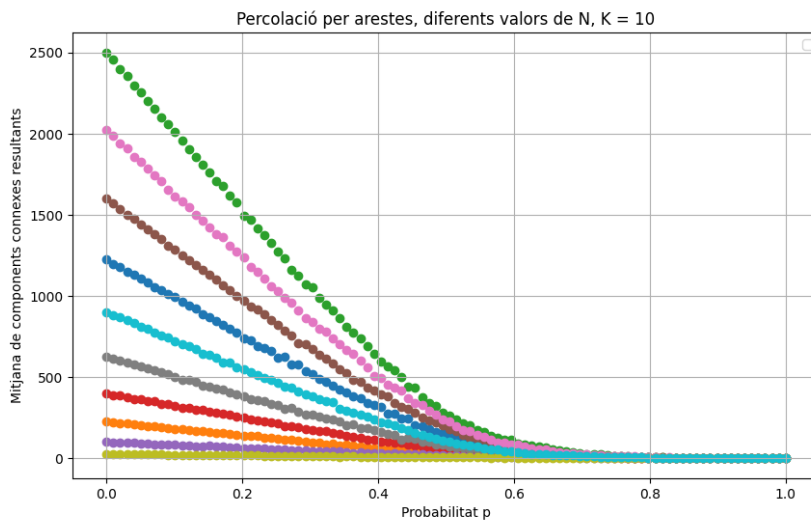


Figura 2.2.2:

2.2.2 Anàlisi de les dades

Investigarem com és que el tram inicial decreixent sembla completament lineal. També sorgeixen les preguntes: Quina funció lineal segueix aquest tram? Com evoluciona aquesta funció en funció de la mida N del graf graella? Primer enunciem dos petits lemes:

Lema 1 *Sigui $G = (V, E)$ un graf graella amb $|V| = N^2$. És clar que $|E| = 2N(N - 1)$.*

Lema 2 *Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes. Si la probabilitat de supervivència és 0, llavors el nombre de components connexes és igual a $|V| = N^2$.*

Lema 3 *Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes amb probabilitat p de supervivència de les arestes. Es compleix que: $\text{mitjana}(|E'|) = |E|p = 2N(N - 1)p$.*

Lema 4 *Sigui $G = (V, E)$ un graf amb C components connexes. Si tenim dos vèrtexs aïllats u i v , llavors el graf $G' = (V, E \cup (u, v))$ té $C - 1$ components connexes.*

L'argument de la linealitat l'enunciem de forma verbal, mitjançant un petit exemple, donat que els coneixements matemàtics i estadístics per tal de formalitzar-los queden fora de l'abast d'aquest treball. Considerem un graf graella amb $N = 100$, de forma que el graf té 10.000 nodes. Inicialment, el Lema (2) ens diu que inicialment amb probabilitat 0 hi ha 10.000 nodes, de forma que per probabilitats de supervivència petites com 0.05. Pel Lema (3), únicament tindrem 500 arestes, i és raonable acceptar que, com tenim 10.000 nodes, el més probable és que aquestes arestes no s'intersequin entre si. Llavors cada aresta uneix dos nodes aïllats, de forma que pel Lema (4), hem perdut 500 components connexes. Des d'aquest argument enunciem la hipòtesi:

Hipòtesi 1: *Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes i probabilitat p , i C les components connexes de G' . Si la probabilitat p és petita, es compleix que en mitjana, $\text{mitjana}(C) \approx |V| - |E|p = N^2 - 2N(N - 1)p$.*

Per tal de comprovar aquesta hipòtesi, en les figures (2.2.3) i (2.2.4) es dibuixen aquestes funcions damunt de les dades. En efecte, la funció calculada sembla ajustar-se en gran mesura a les dades.

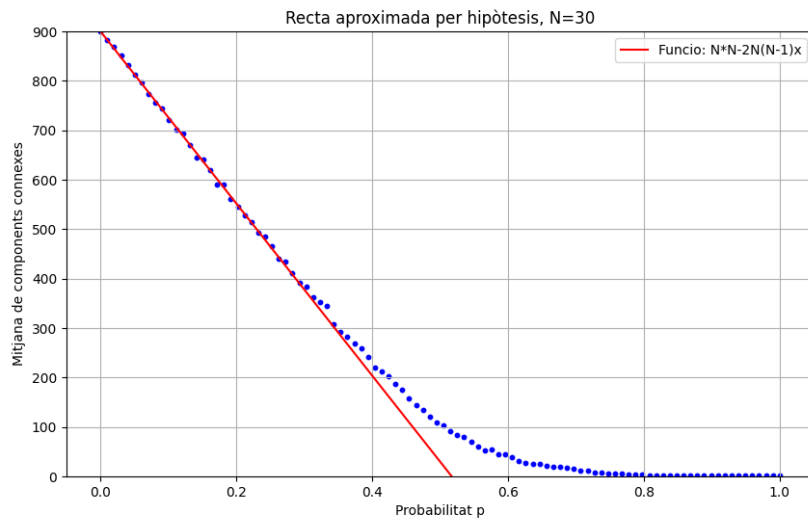


Figura 2.2.3:

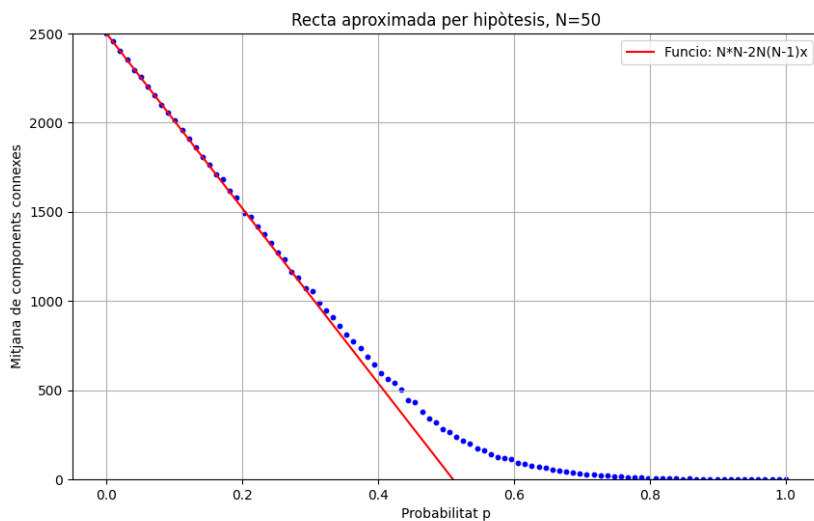


Figura 2.2.4:

2.3 Conclusions:

En relació amb la transició de fase, personalment no considerem que n'hi hagi cap que molt clara ni en la percolació per vèrtexs ni arestes; però, si considerem la transició de fase com **el punt p on el comportament de la funció per punts $p' < p$ és radicalment diferent per punts $p < p'$** , podríem concloure que la transició de fase seria el punt d'inflexió on la gràfica passa de ser creixent a decreixent.

Percolació per vèrtexs: Podem concloure que la transició de fase se situa a la probabilitat 0.27, i mantenim la hipòtesi que tendeix a $1/4$.

Percolació per arestes: No sembla existir transició de fase, ja que la funció resulta decreixent en tot el recorregut.

3 Estudi sobre els Grafs Graella 3D

En aquest estudi analitzem el comportament d'un graf graella tridimensional en un context de percolació per vèrtexs. Definim un graf de graella 3D de mida N^3 com un conjunt de vèrtexs V i arestes E , on $G = (V, E)$ amb $V = \{a_{ijk} \mid i, j, k \in \{1, 2, \dots, N\}\}$. Dos vèrtexs a_{ijk} i a_{lmn} estan connectats si i només si $(i = l \pm 1 \text{ i } j = m \text{ i } k = n)$ o si $(i = l \text{ i } j = m \pm 1 \text{ i } k = n)$, o si $(i = l \text{ i } j = m \text{ i } k = n \pm 1)$.

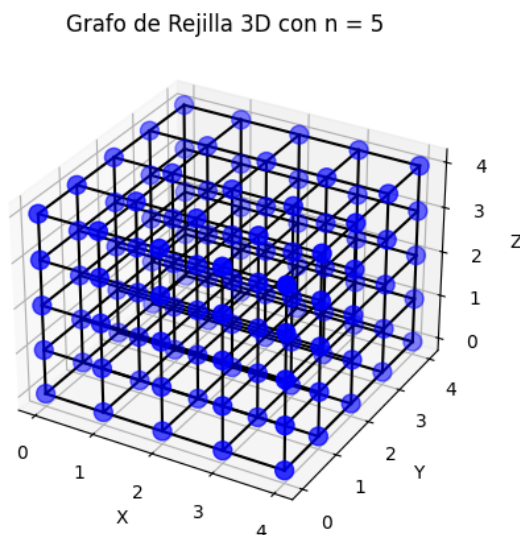


Figura 3.1: Representació gràfica del graf graella 3D abans d'aplicar la percolació.

3.1 Percolació per vèrtexs

3.1.1 Dades descriptives

Les següents figures mostren el nombre mitjà de components connexes en funció de la probabilitat de supervivència p , aplicant el procediment descrit al graf de graella 3D amb $N = 100$. L'eix horitzontal representa la probabilitat de supervivència p , mentre que l'eix vertical mostra el nombre mitjà de components connexes, calculat després de dur a terme $k = 100$ proves per a cada valor de p .

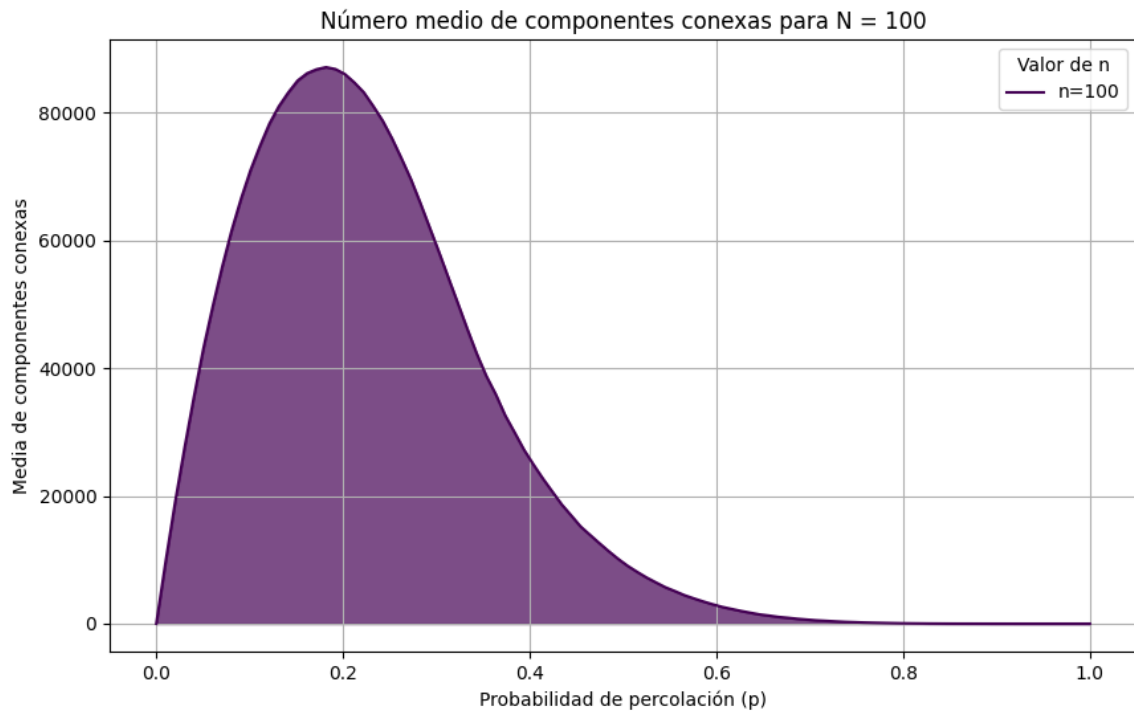


Figura 3.2: Nombre mitjà de components connexes en funció de la probabilitat de supervivència p en un graf graella 3D amb $N = 100$.

És interessant observar com varien aquests gràfics en incrementar N . A la figura 3, es mostra l'evolució del nombre de components connexes per a diferents valors de N , com $N = 10, 20, 30, \dots, 100$.

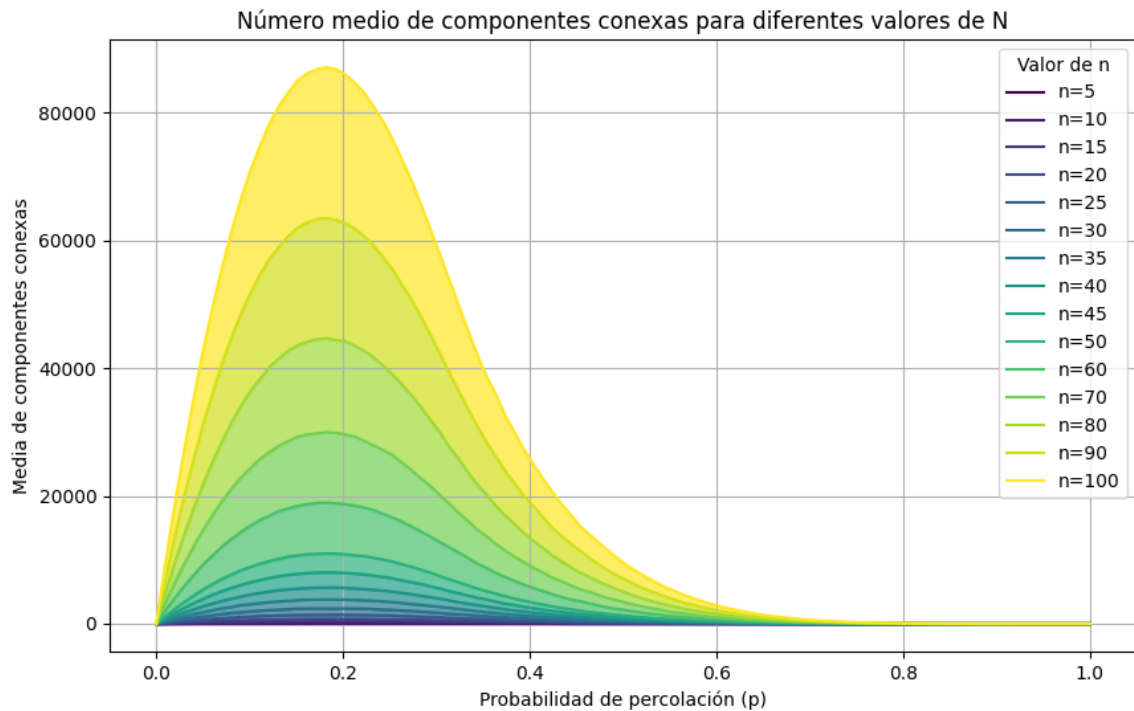


Figura 3.3: Nombre mitjà de components connexes en funció de p per a diferents mides N del graf.

3.1.2 Anàlisi de les dades

Podem observar que el nombre de components no disminueix de manera monòtona a mesura que augmenta p . A mesura que incrementem la probabilitat de supervivència, el nombre de components connexes inicialment augmenta fins a arribar a un màxim i després disminueix quan les components comencen a fusionar-se.

En particular, notem que per a p petit, el nombre de components connexes creix gairebé linealment amb p , cosa que és consistent amb el comportament esperat quan la majoria dels nodes romanen desconnectats. Podem formular la següent hipòtesi:

$$\text{Hipòtesi: Per a valors petits de } p, \mathbb{E}(C'_G) \approx N^3 p$$

Aquest comportament s'il·lustra en la següent figura, on s'observa la relació entre el nombre de components i p a la regió inicial.

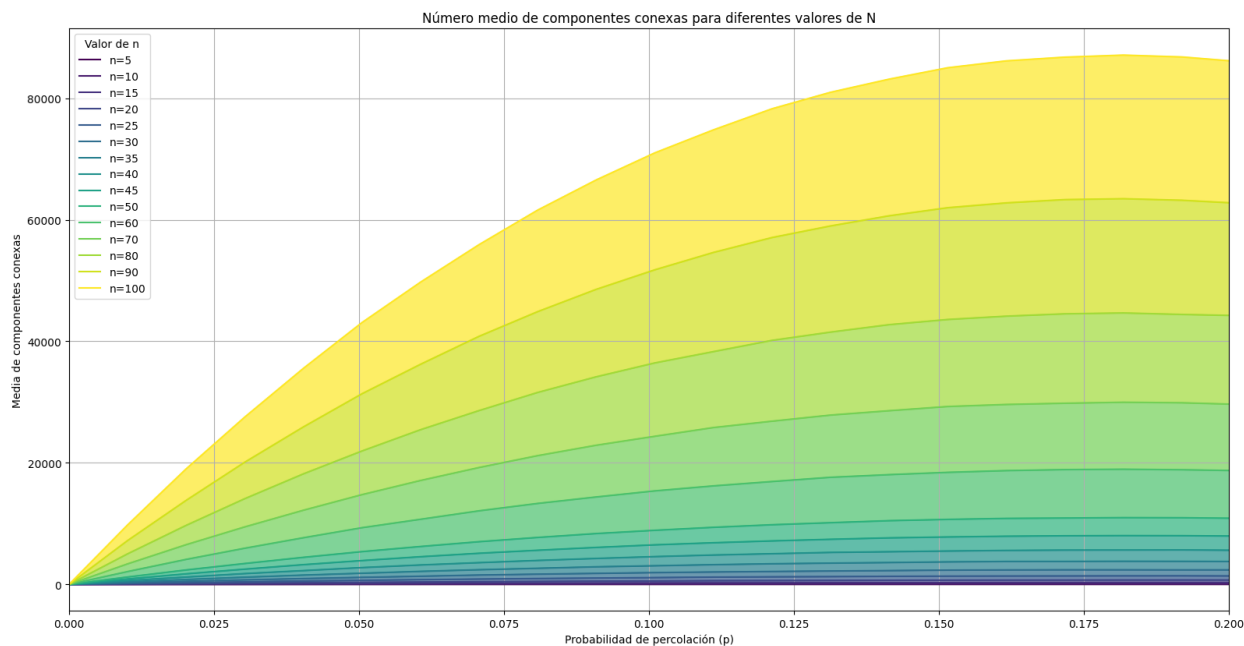


Figura 3.4: Creixement inicial del nombre de components connexes en funció de p .

A mesura que p augmenta, les components comencen a fusionar-se. Podem identificar un valor p' on s'assoleix el màxim nombre de components connexes. Aquest valor p' es manté aproximadament constant per a diferents valors de N , com es mostra a la figura 5.

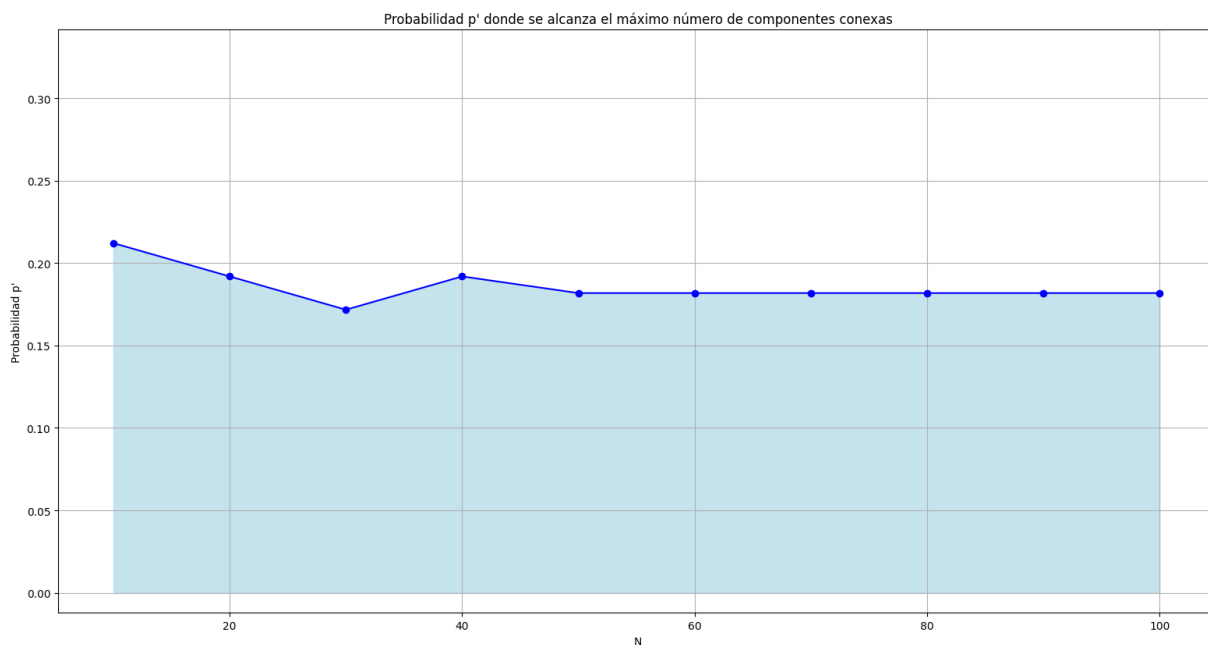


Figura 3.5: Probabilitat p' on s'assoleix el màxim nombre de components connexes per a diferents valors de N .

Relació entre C_N i N Finalment, podem estudiar com el màxim nombre de components connexes C_N es relaciona amb la mida N del graf. Aplicant una regressió polinòmica de tercer grau a les dades obtingudes, obtenim la següent expressió:

$$C_N \approx 8.7039 \times 10^{-2} \cdot x^3 - 5.3835 \times 10^{-2} \cdot x^2 + 6.1377 \cdot x - 5.7208$$

L'ajust d'aquest polinomi a les dades s'il·lustra en la figura 6.

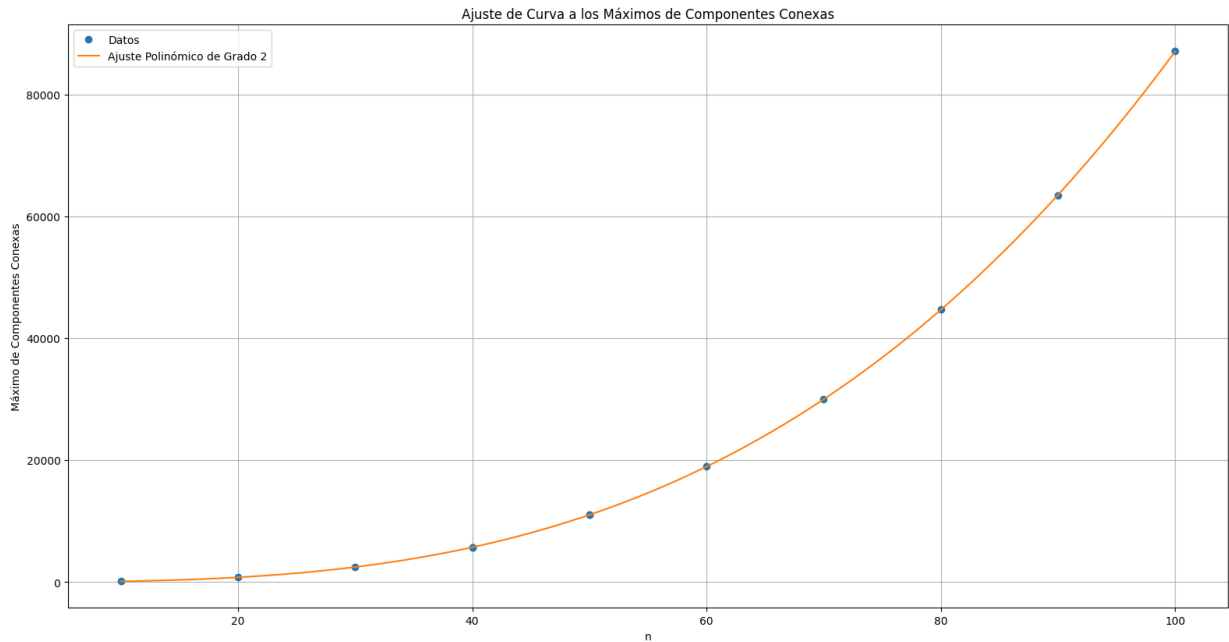


Figura 3.6: Ajust del polinomi

$C_N = 8.7039 \times 10^{-2} \cdot x^3 - 5.3835 \times 10^{-2} \cdot x^2 + 6.1377 \cdot x - 5.7208$ a les dades experimentals.

3.2 Percolació per arestes

3.2.1 Dades descriptives

La figura 7 mostra la gràfica resultant de percolar per arestes sobre un graf graella 3D amb $N = 100$, amb valors $p = 0, 0.01, 0.02, \dots, 1.0$, $k = 10$.

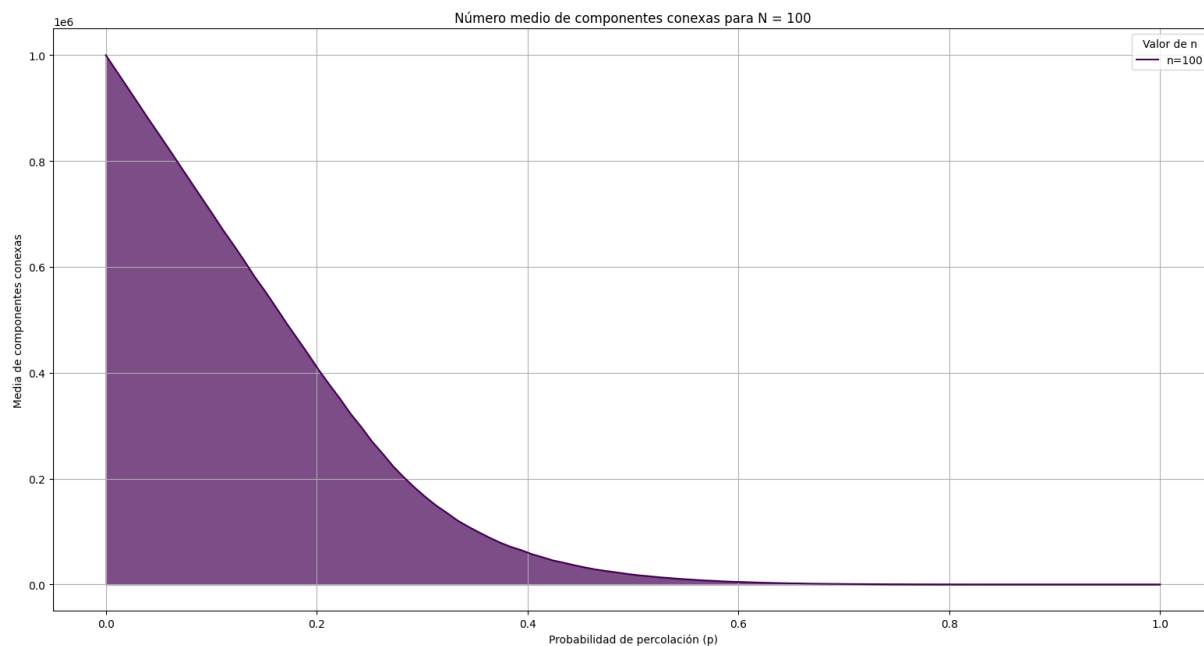


Figura 3.7: Gràfica resultant de percolació per arestes en un graf graella 3D.

Com en el cas anterior, resulta d'interès observar com evolucionen aquestes gràfiques a mesura que incrementem el paràmetre N . En la figura 8 es mostren les gràfiques amb $N = 5, 10, 15, 20, \dots, 100$.

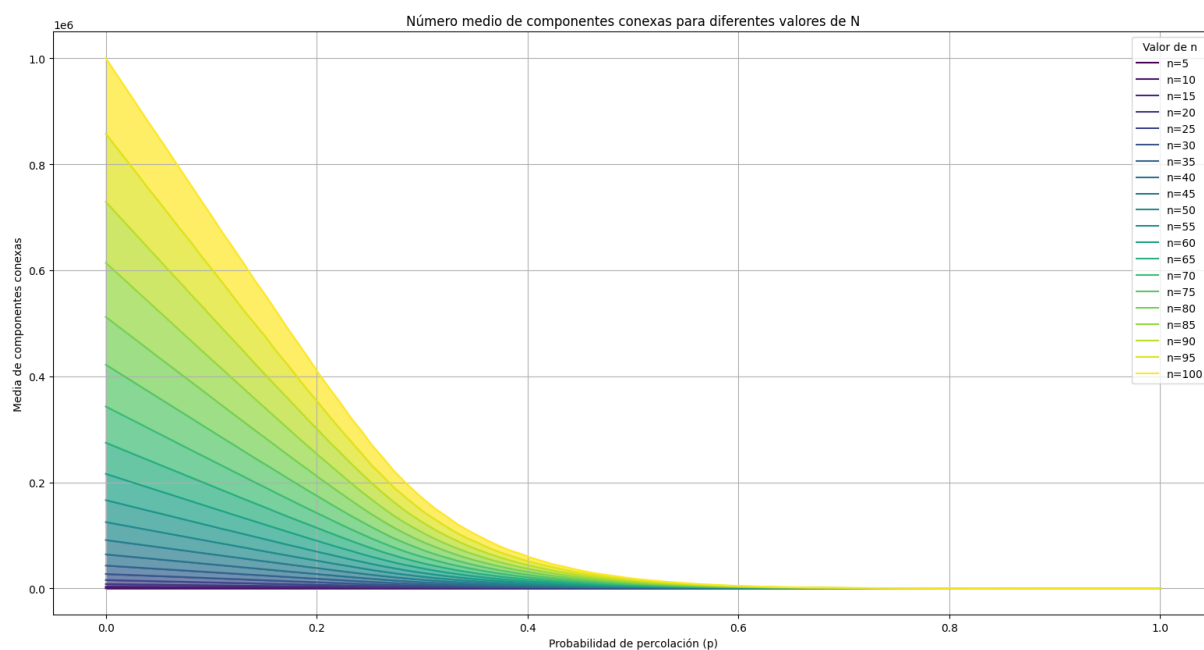


Figura 3.8: Evolució de les gràfiques amb diferents valors de N .

3.2.2 Anàlisi de les dades

Investigarem com és que el tram inicial decreixent sembla completament lineal en el context tridimensional. Les preguntes a considerar són: Quina funció lineal segueix aquest tram? Com evoluciona aquesta funció en funció de la mida N del graf graella 3D?

Lema 1: Sigui $G = (V, E)$ un graf graella amb $|V| = N^3$. És clar que $|E| = 3N^2(N - 1)$ (considerant les connexions en les tres dimensions).

Lema 2: Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes. Si la probabilitat de supervivència és 0, llavors el nombre de components connexes és igual a $|V| = N^3$.

Lema 3: Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes amb probabilitat p de supervivència de les arestes. Es compleix que: mitjana($|E'|$) = $|E|p = 3N^2(N - 1)p$.

Lema 4: Sigui $G = (V, E)$ un graf amb C components connexes. Si tenim dos vèrtexs aïllats u i v , llavors el graf $G' = (V, E \cup (u, v))$ té $C - 1$ components connexes.

L'argument de la linealitat pot ser explicat similar al cas anterior. Considerem un graf graella 3D amb $N = 100$, resultant en 1 000 000 nodes. Inicialment, el Lema (2) ens indica que amb probabilitat 0 hi ha 1 000 000 nodes. Amb petites probabilitats de supervivència com 0.05, pel Lema (3) tindrem aproximadament 300 000 arestes, i és raonable pensar que aquestes arestes no s'intersequen, resultant en la pèrdua de components connexes.

Hipòtesi 1: Sigui $G = (V, E)$ i $G' = (V', E')$ el graf percolat per arestes i probabilitat p , amb C les components connexes de G' . Si la probabilitat p és petita, es compleix que, en mitjana, mitjana(C) $\approx |V| - |E|p = N^3 - 3N^2(N - 1)p$.

Per tal de comprovar aquesta hipòtesi, en les figures 9 i 10 es dibuixen aquestes funcions damunt de les dades. Efectivament, la funció calculada sembla ajustar-se en gran mesura a les dades.

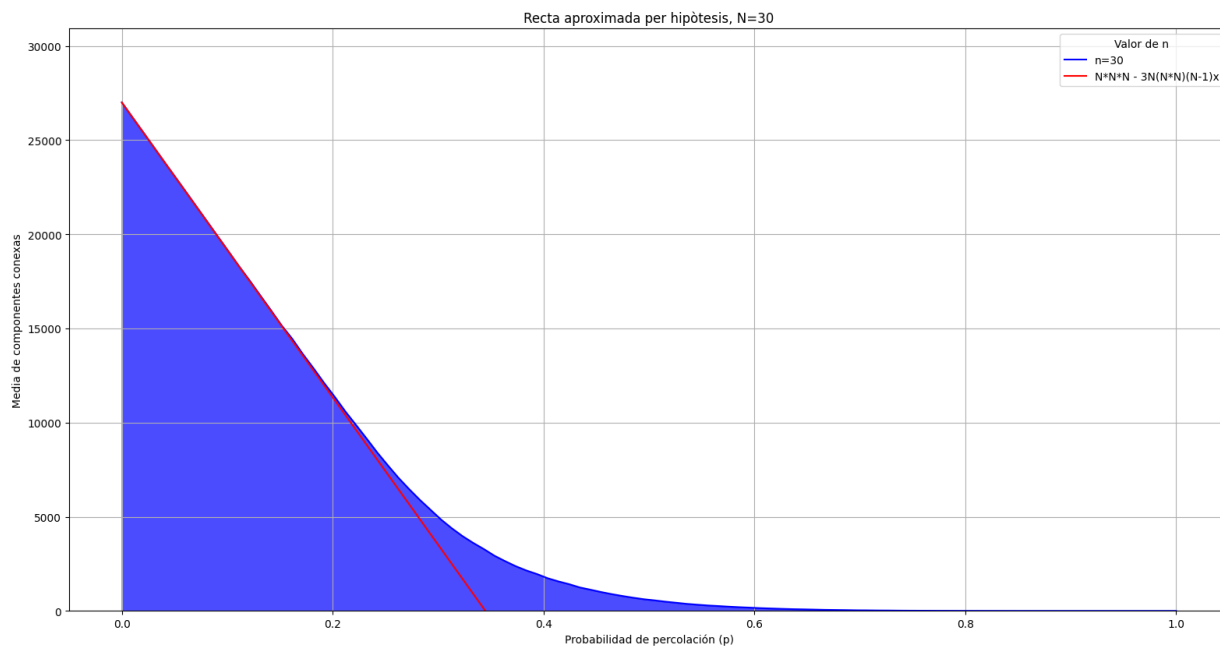


Figura 3.9: Evolució de les gràfiques amb diferents valors de N .

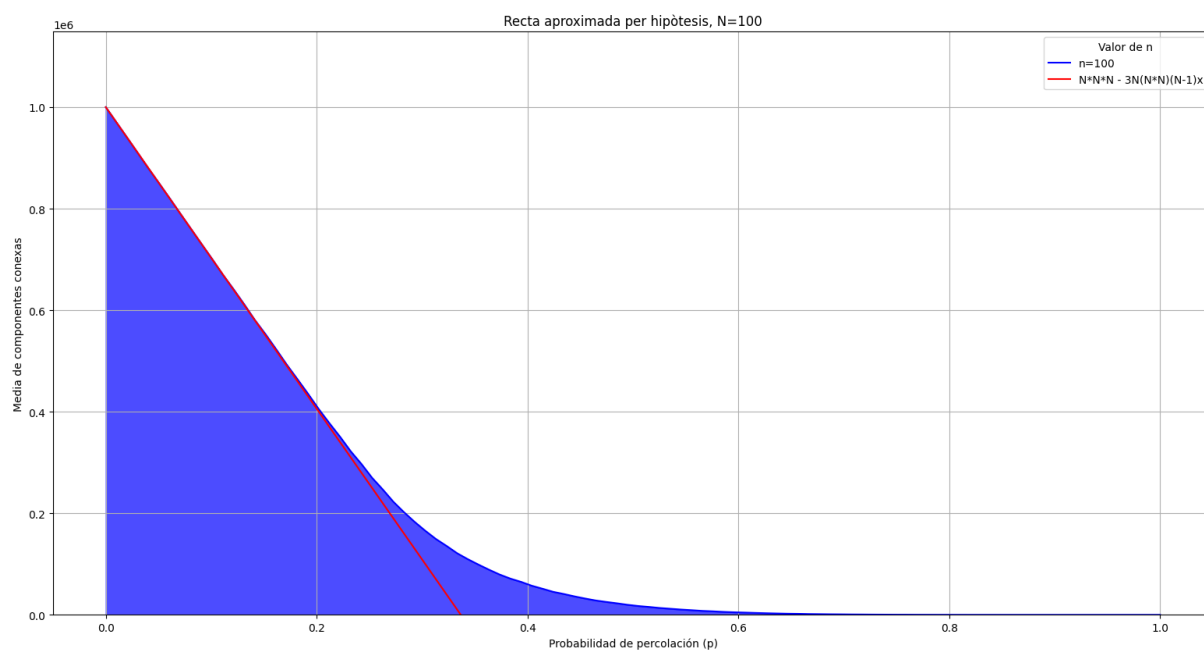


Figura 3.10: Evolució de les gràfiques amb diferents valors de N .

3.3 Conclusions:

En aquest estudi hem analitzat el comportament de la percolació en grafs graella 3D i, tot i que no hem identificat clarament una transició de fase en el sentit tradicional, els resultats suggereixen la presència d'un punt màxim de components connexes al voltant de $1/6$. Aquest punt podria ser interpretat com un índex de transició de fase, ja que a partir d'aquest valor, les components comencen a fusionar-se i es redueix el nombre total de components. No obstant això, no és clar si aquesta característica es pot considerar una transició de fase en un sentit rigorós, ja que no hem observat un canvi dràstic en les propietats globals del graf. Seria interessant aprofundir en aquest punt per determinar si realment estem davant d'una transició de fase o si es tracta simplement d'una característica particular de la graella 3D.

4 Estudi sobre els Random Geometric Graphs

En aquest apartat s'estudia el fenomen de la percolació per arestes en relació amb el nombre de components connexos per a grafs geomètrics aleatoris (Random Geometric Graphs, d'ara endavant RGG) de diverses mides. Ens posem per objectiu esbrinar si hi ha un o diversos *transicions de fase* per a l'atribut enter nombre de components connexes després de la percolació per arestes de l'RGG.

4.1 Definicions i propietats

Definició graf geomètric aleatori: Definim un graf geomètric aleatori amb N nodes i radi r , com un graf no dirigit amb N nodes distribuïts aleatòriament en un espai $[0, 1]^d$ i amb conjunt d'arestes E tal que $(x, y) \in E$ si i només si la distància euclidiana entre x i y és menor que el radi r .

Connectivitat d'un graf geomètric aleatori en dues dimensions: D'acord amb els estudis conduïts per trobar el radi r_c pel qual hi ha un *transició de fase* en la connectivitat d'un RGG generat en dues dimensions, es troba que amb una probabilitat molt alta un RGG en dues dimensions amb N nodes serà connex si el seu radi r compleix $r \geq r_c = \sqrt{\frac{\ln(n) \pm O(1)}{\pi n}}$. A la fórmula, $O(1)$ indica una quantitat constant que no augmenta en relació amb N , per simplicitat en la nostra implementació ignorem aquesta quantitat.

4.2 Procediment Experimental

Per dur a terme l'experimentació, estudiarem per a q valors discrets del paràmetre p , probabilitat d'activació d'una arista, on $p \in [0, 1]$, el nombre de components connexos per a RGGs d'ordre N inicialment connexos.

A diferència dels altres grafs estudiats en aquest treball, els grafs RGG no són deterministes sinó aleatoris. A més, mentre que els altres grafs estudiats són grafs perfectes en els quals el grau de cada node és més aviat petit, el RGG és aleatori i el grau de cada vèrtex varia molt en funció de com estan distribuïts els nodes en l'espai i quants nodes té un vèrtex en el seu radi. Aquest fet té dues conseqüències molt importants per a l'estudi de l'existència d'un *transició de fase* en la percolació per components connexos d'un RGG.

La primera és que, donat un nombre de nodes, ara no treballem amb un graf concret, sinó que ens trobem amb una infinitud de grafs aleatoris que comparteixen una sèrie de característiques i un paràmetre de radi. La segona, és que la variància en el nombre de components connexos per un cert graf augmenta significativament a causa de l'aleatorietat en la generació. Això porta a grafs que estan lluny de ser perfectes i amb graus per cada vèrtex molt diferents. Això pot comportar que les conseqüències d'eliminar dues arestes qualssevol tinguin efectes molt diferents, portant a un augment de la variància en la percolació per arestes en relació amb la variància per aquest procés en grafs deterministes.

Per mitigar el problema descrit anteriorment, per cada valor de probabilitat p aplicarem l'estudi de la percolació sobre 100 RGGs d'ordre N i radi r , així esperem tenir una mostra prou representativa de la població d'RGGs d'ordre N i radi r . Sobre cada un d'aquests RGG desactivarem o no cada una de les seves arestes amb probabilitat $1 - p$, k cops. Com que la variància de l'experiment ha augmentat significativament, tornarem a determinar k .

4.2.1 Determinació de la k per a tenir significació estadística:

En aquest estudi fixem $n = 500$ i $q = 100$ valors de probabilitat. Recordem que per a cada valor de probabilitat obtindrem la mitjana de components connexes de l'RGG després d'executar k proves sobre 100 RGGs. Per a determinar k , hem provat, en les condicions esmentades, quatre possibles valors de k : 10, 35, 70, 100.

En totes les execucions observem que la corba de components connexes per valor de probabilitat segueix la tendència d'una certa corba exponencial, però que hi ha alguns valors de probabilitat per als quals la mitjana es desvia lleugerament d'aquesta tendència. Aquestes desviacions són soroll fruit de la variància, per tant, intentem trobar una k que minimitzi el nombre de desviacions i la seva llunyania respecte a la tendència de la corba.

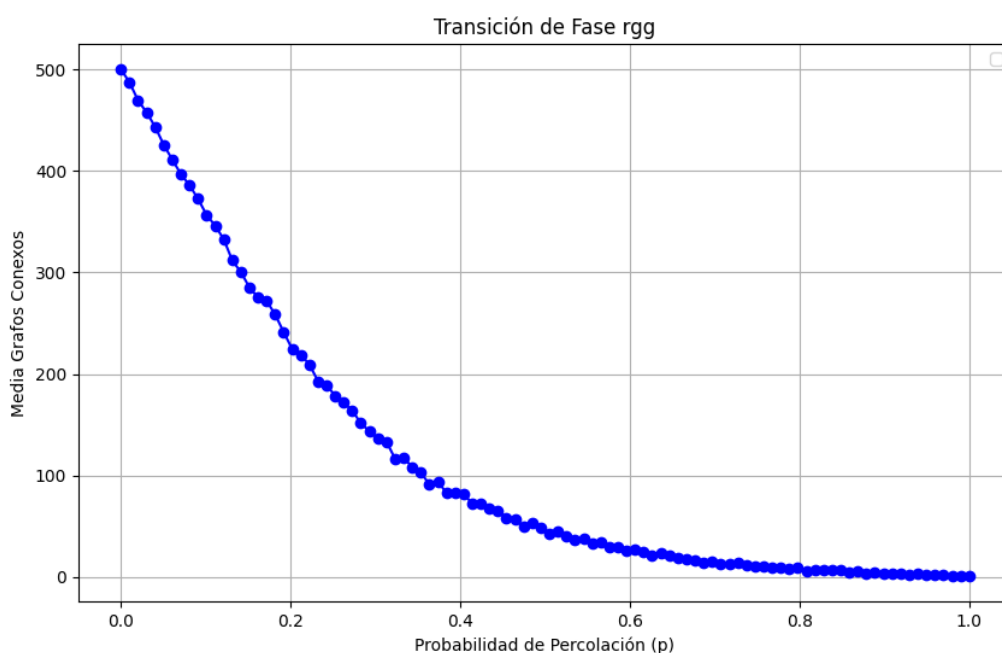


Figura 4.1: Corba de components connexes amb $k = 10$

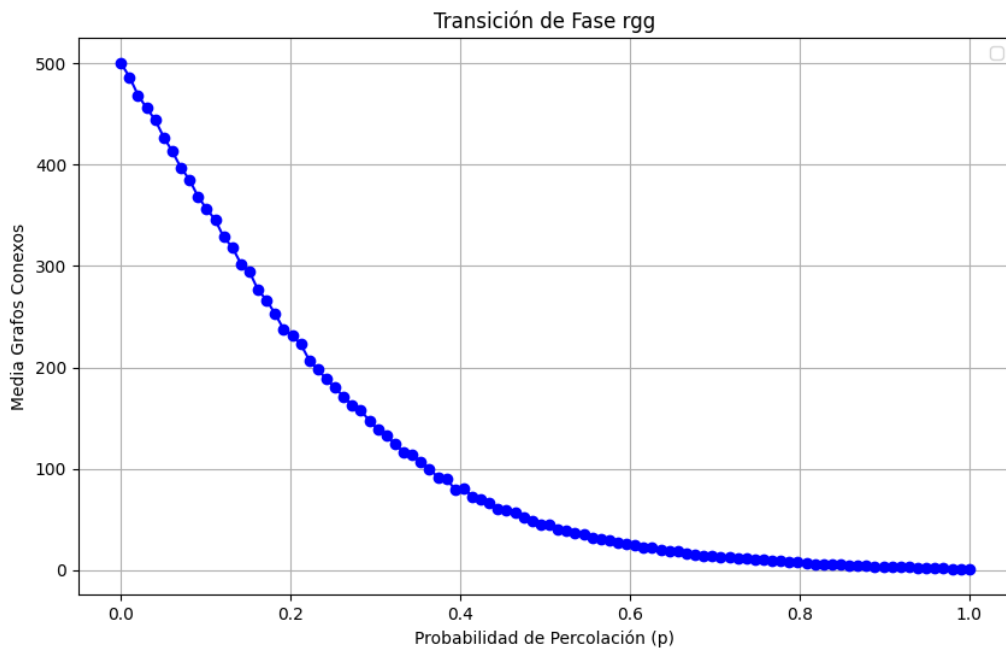


Figura 4.2: Corba de components connexes amb $k = 35$

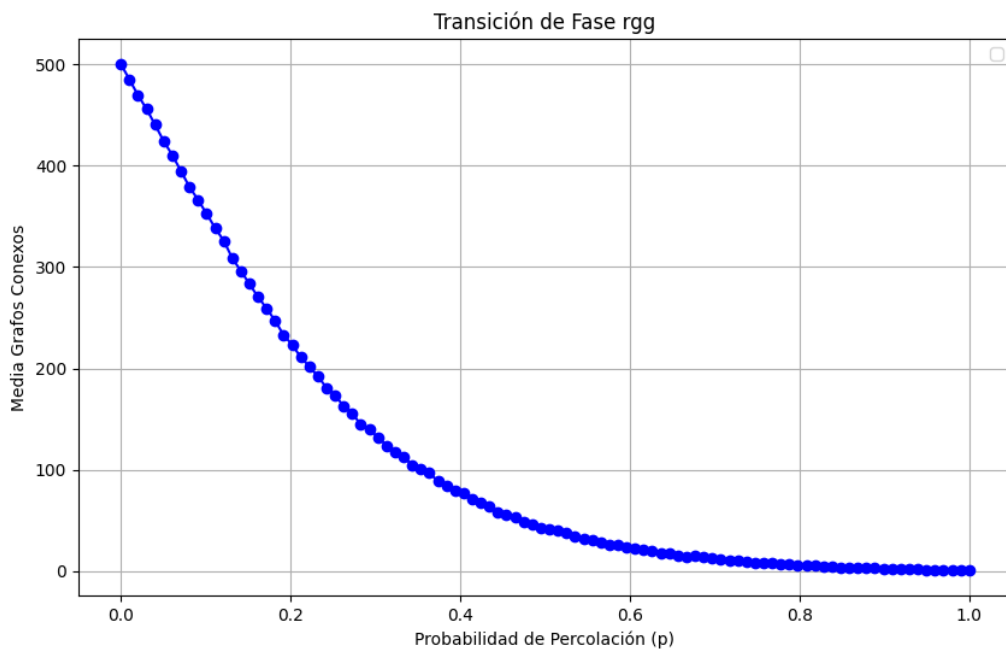


Figura 4.3: Corba de components connexes amb $k = 70$

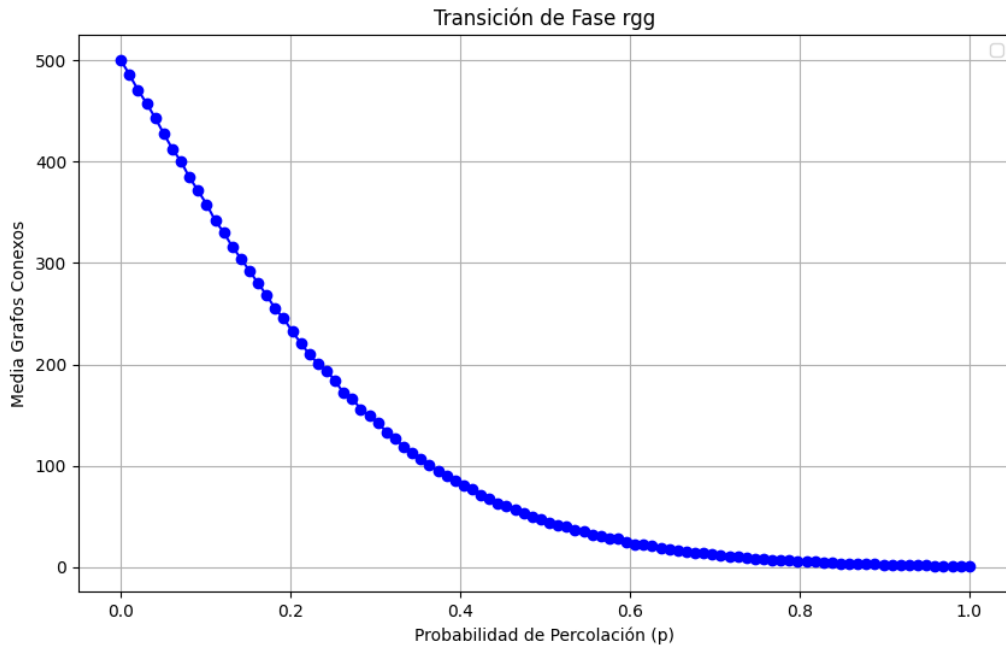


Figura 4.4: Corba de components conexas amb $k = 100$

Fem un estudi qualitatiu per a determinar la k idònia. Veiem que la mitjana de components conexas per a molts valors de p es desvia de la tendència de la corba en moltes ocasions i les desviacions tenen una magnitud important amb $k = 10$. Per $k = 35$, el nombre de desviacions i la seva magnitud disminueixen significativament. Per $k = 70$ també observem un descens destacable de les desviacions i de la seva magnitud. Finalment, per a $k = 100$ no trobem un descens significatiu de les desviacions ni de la seva magnitud.

4.2.2 Generació de la mostra d'RGGs:

Algorithm 1 Algorisme generate_connected_rgg

```

 $r \leftarrow \sqrt{\frac{\ln(n)}{\pi n}}$ 
 $G \leftarrow \text{random\_geometric\_graph}(n, r)$ 
while not is_connected( $G$ ) do
     $G \leftarrow \text{random\_geometric\_graph}(n, r)$ 
end while
return  $G$ 

```

Per generar RGGs conexas fem servir la propietat esmentada a l'apartat anterior, per la qual un RGG és connex amb una probabilitat molt alta si $r \geq r_c = \sqrt{\frac{\ln(n) \pm O(1)}{\pi n}}$. Pel nostre disseny experimental, definirem $r = r_c = \sqrt{\frac{\ln(n)}{\pi n}}$, ignorant el paràmetre $O(1)$ perquè no considerem que el seu impacte sigui prou rellevant per a fer servir *fine-tuning* per determinar-lo. Finalment, per garantir que els RGGs són conexas, després de generar cada RGG estudiem la seva connectivitat. Si és connex, procedim a usar-lo en l'experimentació; en el cas altament improbable que l'RGG no sigui connex, tornem a generar un RGG

sense variar r ni N i tornem a comprovar la connectivitat. Això és eficient perquè fem servir $r = r_c$.

Per a la comprovació de la connectivitat fem servir les funcions de la llibreria de Python NetworkX, la funció `random_geometric_graph` per a la generació dels RGG, i la funció `is_connected` per comprovar la connectivitat.

4.2.3 Paràmetres experimentals:

Finalment, hem optat per fer servir com a paràmetres experimentals $r = r_c$, $k = 70$ pels 100 RGGs generats per valor de probabilitat i $q = 100$ valors de probabilitat. Amb aquests paràmetres estudiarem la percolació per arestes en RGGs amb nombre de nodes $n \in \{1000, 2000, 3000, 5000, 7500, 10000, 12500, 15000\}$.

4.3 Resultats de l'experimentació:

Després d'obtenir i representar els resultats del procés de percolació obtenim aquest gràfic que mostra el nombre mitjà de components connexes d'un RGG en funció de la probabilitat d'activació d'una aresta.

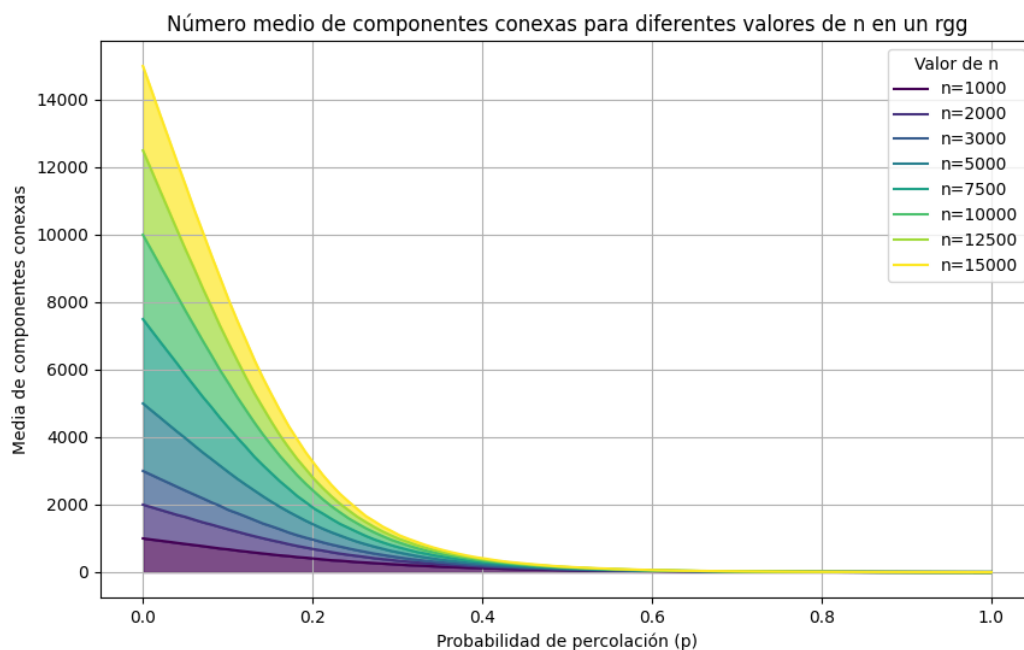


Figura 4.5: Nombre mitjà de components connexes després de la percolació en un RGG amb probabilitat q per les N triades

A partir dels resultats que es veuen al gràfic, es pot descartar que el procés de percolació per arestes en un graf d'ordre $n \leq 15000$ doni lloc a una *transició de fase* pel que fa al nombre de components connexes en funció de la probabilitat de percolació p .

És destacable, d'altra banda, que la corba *components-connexes*(p), que segueix la tendència d'una funció exponencial, és molt més pronunciada per als RGGs que per als altres

models de grafs provats. El motiu d'aquesta diferència en les característiques de la corba, queda com a objecte d'estudi per a futurs treballs.

4.4 Estudi del *knee of a curve*:

A la llum d'aquests resultats ens sorgeix una pregunta per curiositat: en cas de trobar un *transició de fase* després de dur a terme un procés de percolació en grafs amb $n > 15000$, es pot determinar quin seria el nombre de components connexes per a $p > p_{transició\ de\ fase}$.

Per descobrir-ho, estudiem el *knee of the curve* (per simplicitat *knee-point* d'ara endavant), el punt d'una funció en què aquesta passa de decreixer o créixer ràpidament a fer-ho lentament, per a la corba *components-connexes*(p) dels RGGs estudiats a l'apartat anterior. En el gràfic es pot veure com el nombre de components connexes al *knee-point*, és cada cop menor en relació amb l'ordre N del graf fins que comença a estabilitzar-se en valors d'entre $0.00001n$ i $0.00002n$ components connexes. Això pot indicar que en cas d'existir una transició de fase per $n > 15000$ el nombre de components connexes per $p > p_{transició\ de\ fase}$ seria com a molt un valor entre $0.00001n$ i $0.00002n$.

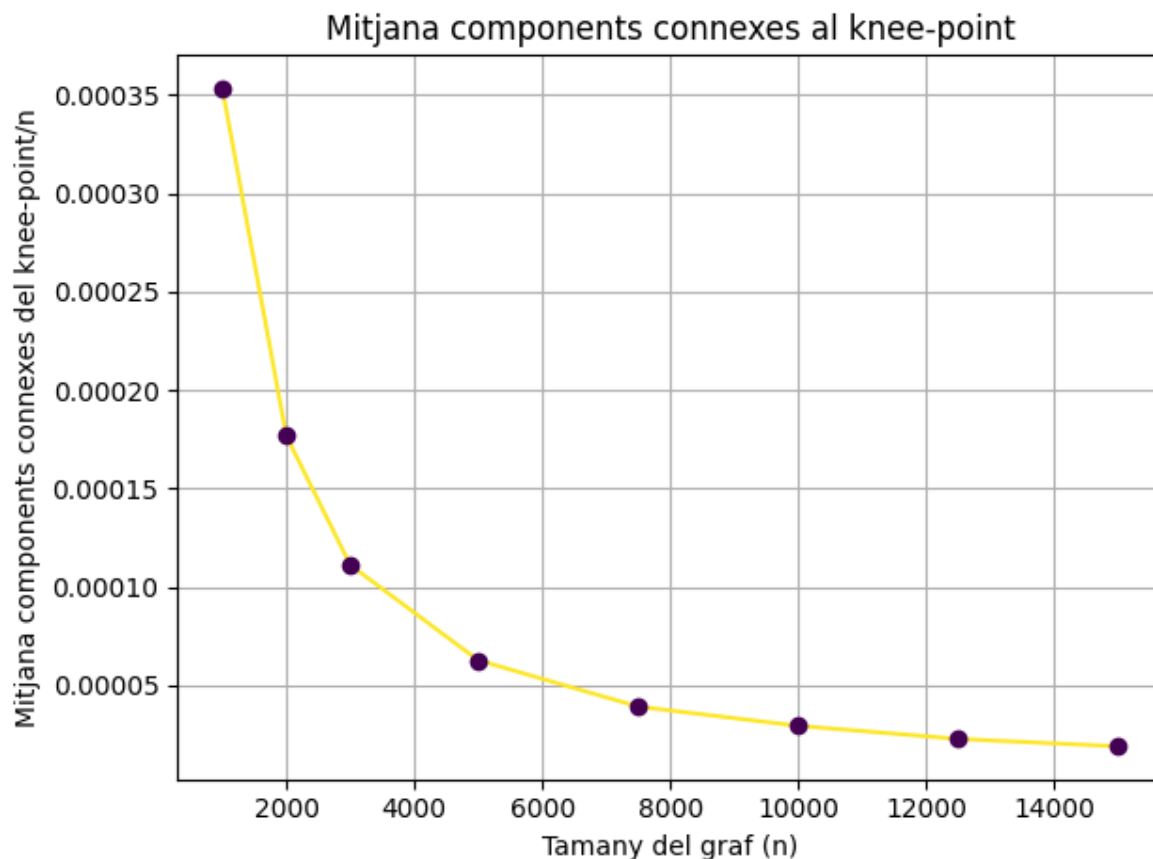


Figura 4.6: Mitjana de components al knee-point

4.5 Conclusions experimentació:

Després de sotmetre RGGs de diferents mides a un procés de percolació per arestes, no hem pogut trobar cap *transició de fase* per al nombre de components connexes del RGG

en funció de la probabilitat p d'activació de les arestes en grafs amb $n \leq 15000$ nodes. No s'han estudiat grafs més grans i, per tant, no es pot descartar l'existència d'una transició de fase en RGGs amb més nodes. A tall de curiositat, hem descobert que el *knee-point* de la mitjana de components connexes en funció de p tendeix a valors d'entre $0.00001n$ i $0.00002n$ a mesura que N es fa més gran. Queda per a treballs futurs investigar per què la corba *components-connexes*(p) per a la percolació per arestes és sensiblement més pronunciada per als RGG que per als altres grafs estudiats.

...

5 Estudi sobre els Connected Caveman Graph

Introducció

Ara estudiarem la percolació tant per arestes com per vèrtex en relació amb la connectivitat d'un Connected Caveman Graph (CCG).

El "Caveman graph" és un graf que va ser inventat per Duncan J. Watts l'any 1999 al seu llibre *"Small Worlds: The Dynamics of Networks between Order and Randomness."*, i explica com petits grups aïllats (com els homes de les caveres) poden interconnectar-se en passar el temps. El graf consisteix en un conjunt de $n/(k+1)$ cliques o grups aïllats, sent n el nombre total de nodes i $k+1$ la mida de cada clique, on cada clique representa una cova (a partir d'ara en parlarem dels cliques/grups com a coves). El problema del caveman graph original és que és fortament connex dins de cada grup, però globalment, les diferents coves estan comunicades amb les altres, i, com podem comprendre, això fa que aquest model de graf sigui una mica inútil a l'hora d'experimentar, ja que el graf mai és realment connex, quan hi han com a mínim dues coves, i, encara que per estudiar com afecta la percolació en el nombre de components connexes no afecta si un graf comença sent connex o no, ens interessa veure quan deixa de ser-ho.

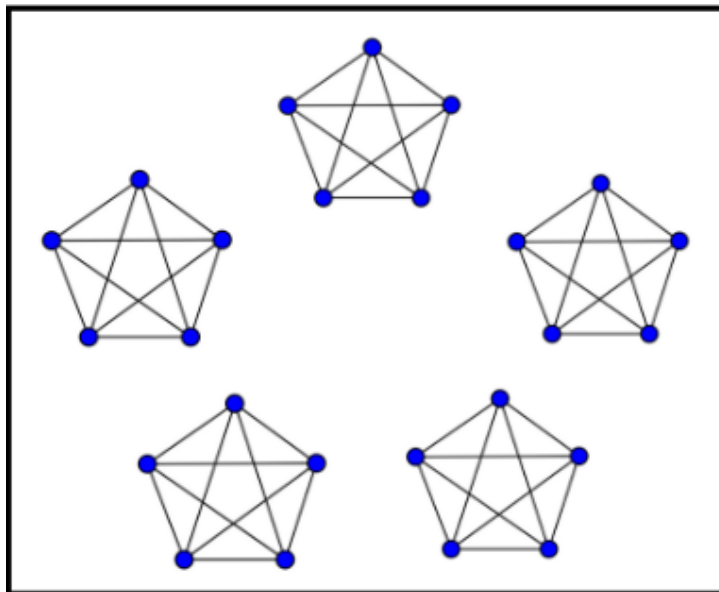
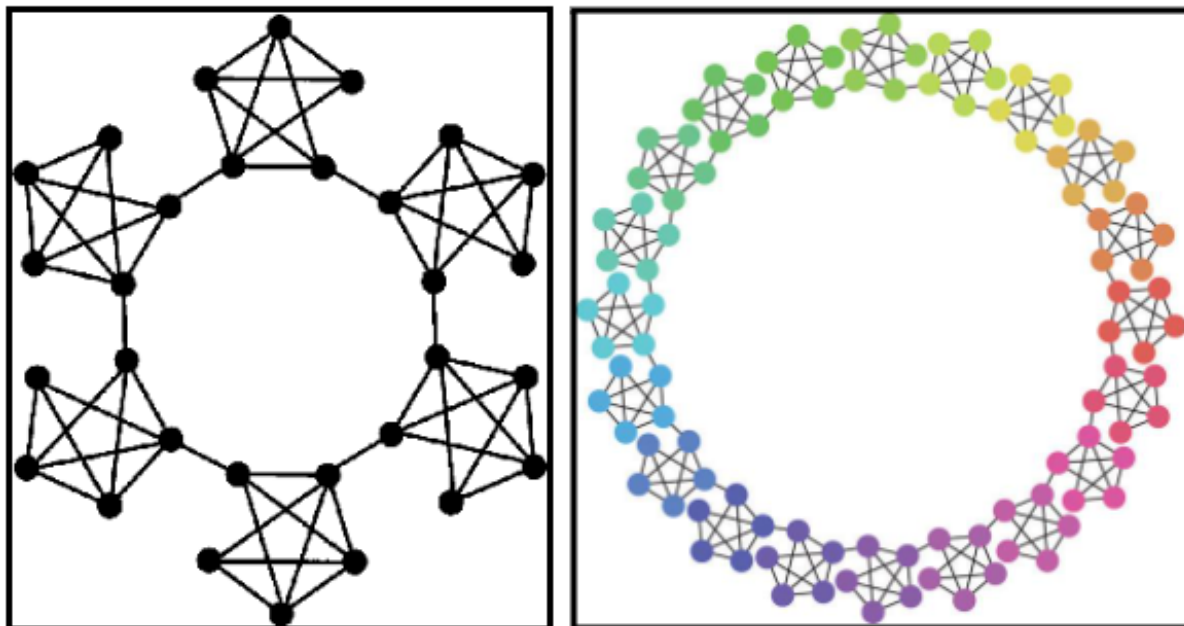


Figura 1: Imatge d'un Caveman Graph

Per tant, per a explicar com es connectaven les coves a poc a poc, J. Watts va fer que les coves anessin retirant algunes arestes dels cliques, canviant-les per arestes entre diferents coves (representant els homes de les caveres migrant de les seves coves i coneixent als altres). Si aquest procés es repeteix suficients vegades tal que cada cova visita als seus veïns més pròxims, quedem amb un graf connex, en el que es veu una societat de coves connectades i comunicades. Aquest graf s'anomena Connected Caveman Graph (a partir d'ara l'anomenarem CCG), i és amb el que experimentarem en aquest apartat de l'estudi.



Figures 2 i 3: Imatges de CCGs, amb diferents nombres de coves

Com a curiositat, en tenir un gran "clustering", és a dir, una gran connectivitat a les mateixes coves, i en estar connectades totes les coves per una aresta, la distància dels camins (tenint que $G = (V, E)$) de qualsevol parell de vèrtexs $u, v \in V$, $d(u, v)$ és mitjanament molt baixa, ja que com a màxim el camí té $2 \cdot (k - 1)$ passes, és a dir, el diàmetre del graf és $D_{\max} = 2 \times (k - 1)$, i això és relativament baix.

Motivació:

A l'aquest ser l'únic model de graf no proposat per la pràctica, hem decidit escollir un graf amb aplicacions reals en teoria de xarxes, com és el CCG, ja que és un model molt bàsic que simula xarxes socials, i està inspirat en el fenomen que explica que la gent tendeix a crear petits grups amb molta confiança i amb els que passen molt temps (coves) famílies, grups d'amics... i explica com gradualment es van connectant amb els altres grups, simulant l'evolució social.

Procediment Experimental

Per enfrontar l'experimentació, hem decidit tenir un nombre fix de persones dins d'una cova, per a facilitar l'experimentació. Hem declarat que aquest nombre fos 5, ja que hem preferit treballar amb grups petits de persones, perquè en la societat els grups aïllats solen ser més aviat petits, ja que si fossin més grans es formarien subgrups. Per tant si diem que el número de coves és 10, realment vol dir que el número de nodes és 50.

Seguint la metàfora de l'evolució social, al perlocar vèrtex i arestes volem veure com afectaria la societat de coves el fet que es morissin homes de les cavernes (percolació de vèrtex) i que acabessin les relacions entre dos homes de la caverna per qualsevol motiu (percolació d'arestes). Així podem observar com aniran quedant aïllades les coves i en quin moment tornaran a estar totalment aïllades.

Per tant, definim formalment un CCG com:

Sigui $G = (V, E)$ un graf on V és el conjunt de vèrtexs i E és el conjunt d'arestes. Un Connected Caveman Graph amb k coves, cadascuna amb n vèrtexs, està compost per:

- $V = \{v_1, v_2, \dots, v_n\}$, on cada vèrtex v_i pertany exactament a una de les k coves.
- Les coves C_1, C_2, \dots, C_k són subgrafs complets K_n on cadascuna $C_i = (V_i, E_i)$, amb $V_i \subseteq V$ i $E_i \subseteq E$, i $|V_i| = n$.

Cada cova C_i és un **subgraf complet o clique**:

$$C_i = (V_i, E_i), \quad V_i = \{v_{i1}, v_{i2}, \dots, v_{in}\}, \quad E_i = \{(v_{ij}, v_{il}) \mid 1 \leq j < l \leq n\}$$

És a dir, cada parell de vèrtexs dins de la mateixa cova C_i està connectat per una aresta.

Per estudiar la connexió entre coves, diem que almenys un vèrtex d'una cova C_i està connectat a un vèrtex d'una altra cova C_j . És a dir, existeix un conjunt d'arestes inter-coves que connecten vèrtexs de coves diferents:

$$E_{\text{inter}} = \{(v_{i1}, v_{j1}) \mid v_{i1} \in V_i, v_{j1} \in V_j, i \neq j\}$$

Aquestes connexions inter-coves asseguren que el graf sigui connex.

5.1 Estudi per vèrtexs

5.1.1 Dades descriptives

La figura següent mostra com, anàlogament als models de grafs ja descrits, augmenta o decreix el nombre mitjà de components connexes segons la probabilitat p de supervivència del vèrtex o aresta (segons l'estudi).

En la figura 4 es mostra aquest fenomen, percolant vèrtex amb una $N = 100000$, que en aquest model, N es refereix al nombre de coves, per lo tant seran 500000 vèrtexs.

Es fan $k = 10$ proves per a cada valor de p , i hi ha 100 valors de p .

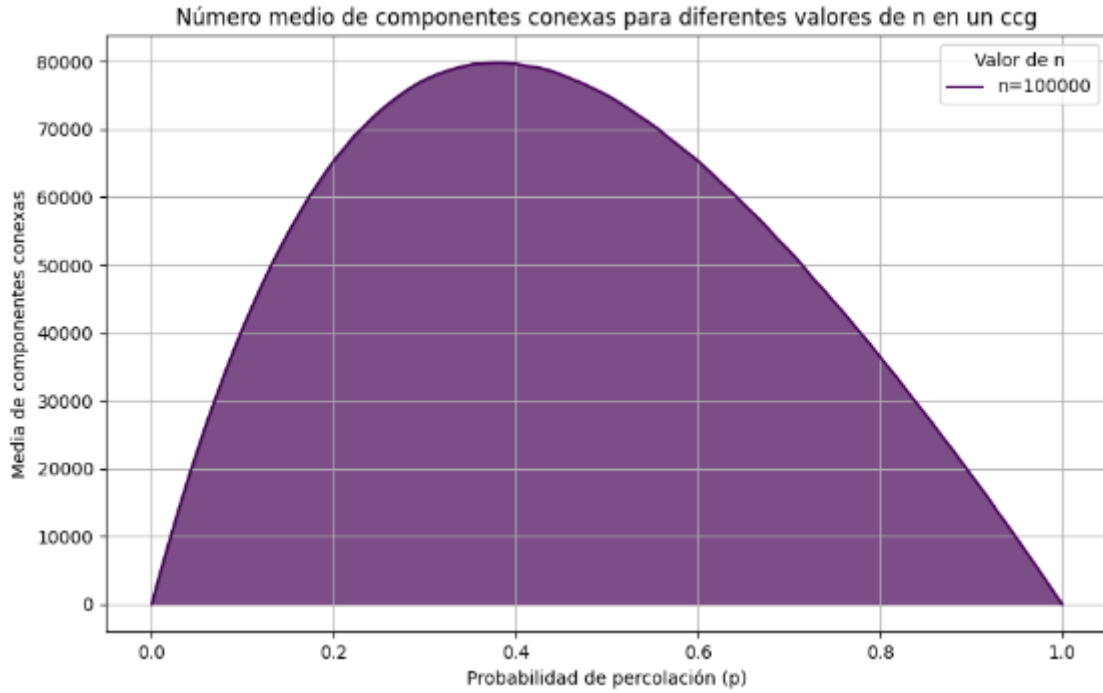


Figura 5.4

La figura 5 mostra la gràfica resultant de percolar per vèrtex sobre un CGG amb $N = 50, 100, 15, \dots, 100000$, amb valors $p = 0, 0.01, 0.02, \dots, 1.0$, $k = 100$.

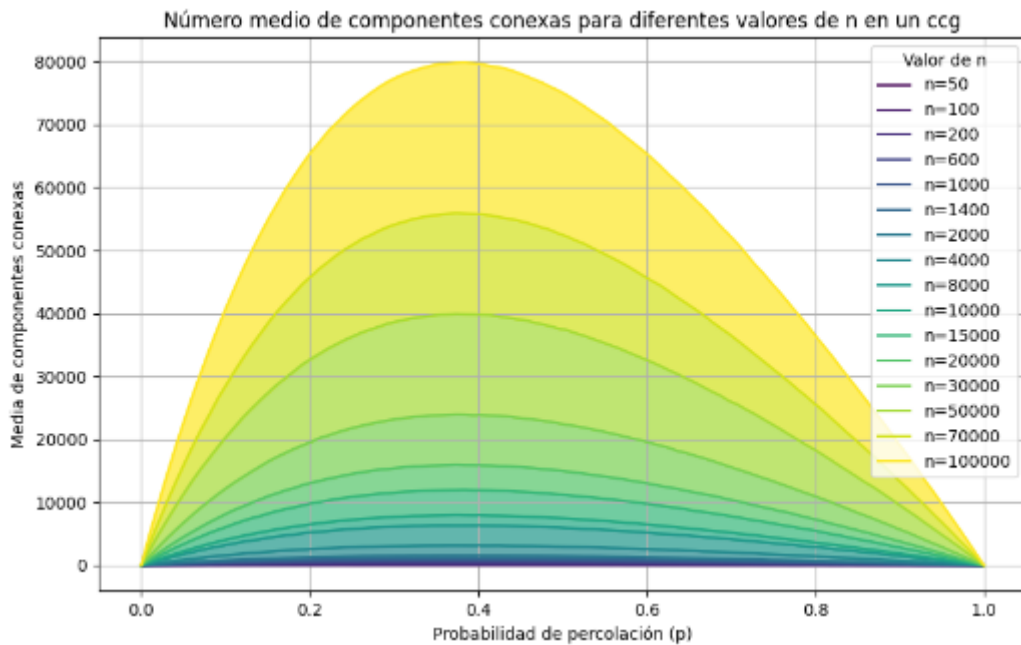


Figura 5.5

La figura 6 mostra un zoom en els primers valors de $p = 0, \dots, 0.25$ per veure el creixement inicial de les components connexes segons p .

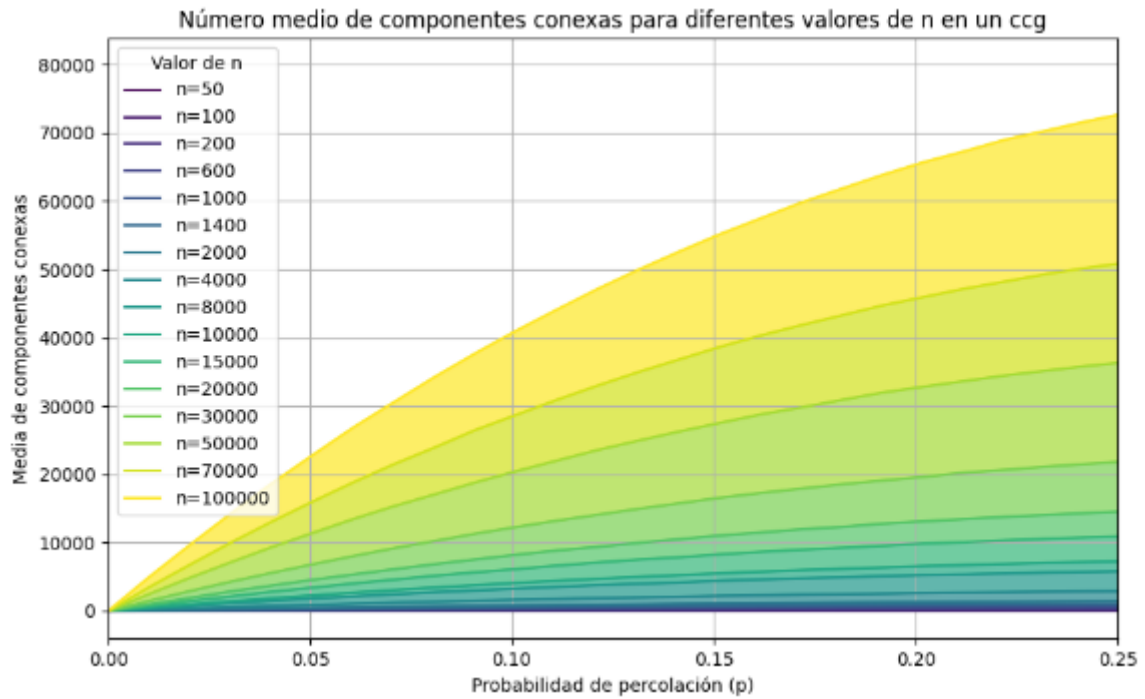


Figura 5.6

La figura 7 mostra la p on s'assoleix el màxim nombre de components. Els punts per cada valor corresponen a $N = (50, 100, \dots, 100000)$.

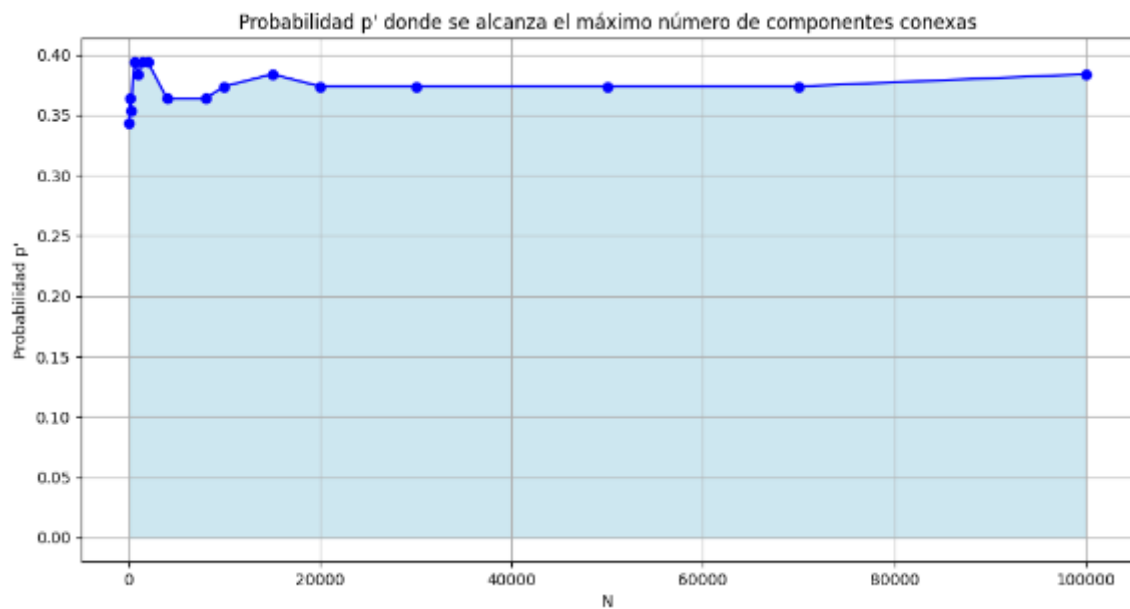


Figura 5.7

La figura 8 mostra el nombre màxim de components connexes en valors petits de N .

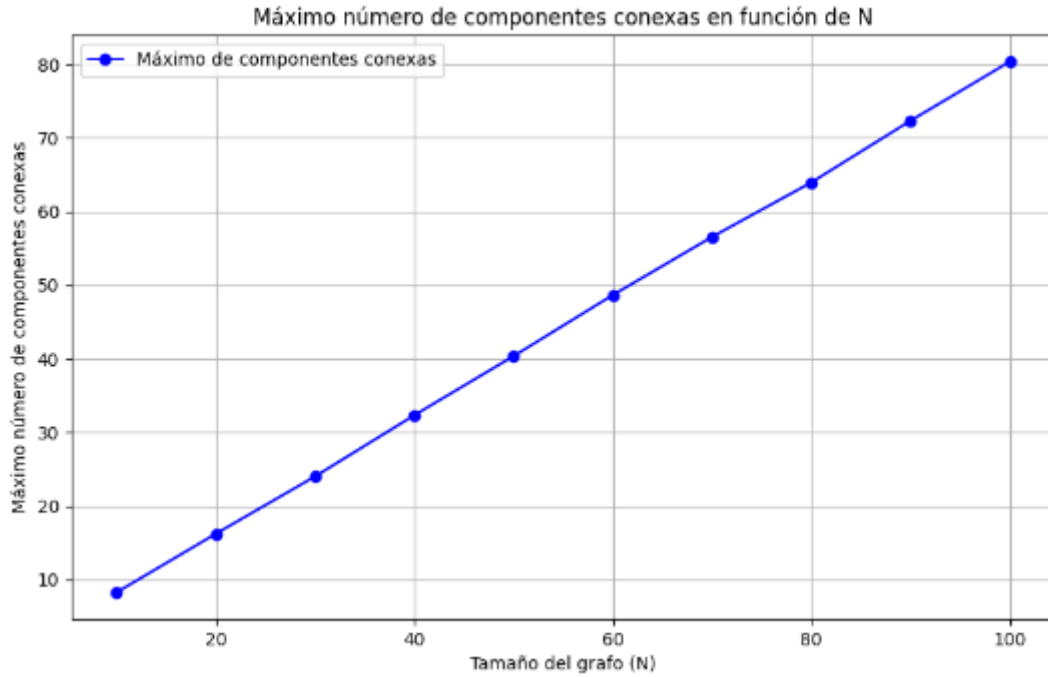


Figura 5.8

5.1.2 Anàlisi de les dades

i) Sobre el comportament de la funció de C_n respecte p

A l'haver estudiat els altres models de grafs, la meua hipòtesi:

Hipòtesi 1: Sigui $G' = (V', E')$ el graf percolat i CG' el nombre de components connexes de G' . Per a probabilitats de supervivència molt petites, es compleix que el creixement de la funció és lineal i arriba a un màxim, decreixent posteriorment de la mateixa manera, fent que CG' tendeixi a 0 per valors mitjans-alts. Penso que això serà així, ja que els CCG són grafs que depenen de les connexions binàries entre coves, i que això fa que la connexitat del graf pugui ser més fràgil.

Fixant-nos en la figura 4, veiem que la hipòtesi **NO** es compleix, ja que, alternament als altres models de grafs; tot i el mencionat al paràgraf anterior, els CCG són grafs amb una densitat de connectivitat molt gran, per tant, el creixement i decreixement és molt més suau comparat amb els altres models.

ii) Sobre la probabilitat p'_N on s'assoleix el màxim nombre de components connexes respecte a N

Com hem pogut veure en l'estudi de C_n sobre p , es dibuixa una gràfica creixent fins a un determinat valor de p , que es presenta semblant per qualsevol valor de N (figura 6), i si observem detalladament, podem deduir que p'_N s'aproxima a $4/10$; per tant:

Hipòtesi 2: La probabilitat on s'assoleix el màxim nombre de components connexes p' s'aproxima a $\frac{4}{10}$.

Estudiant la probabilitat p'_N on s'assoleixen el màxim nombre de components connexes respecte N , obtenim la gràfica de la figura 7, en la que veiem que el valor p'_N oscil·la

entre 0.32 i 0.40 aprox. per a tots els valors de $N(0 \dots 100000)$, per tant, podem dir que **es valida** la Hipòtesi 2.

iii) Sobre el màxim de components connexes en funció de N

Respecte a aquesta part de l'estudi, cal dir que amb els altres models de grafs, ja hem pogut veure que es genera una corba en estudiar el valor màxim de CG' respecte N , però per estudiar alguna cosa diferent, ens ha donat curiositat veure el fet que inicialment és una recta, i, per això, hem estudiat l'inici de la gràfica amb regressió lineal simple.

Hem trobat la recta $-0.2 + 0.81N$, que sembla que ajusta de manera quasi perfecta les dades, com es pot veure a la figura 9, en la que hem sobreposat la recta amb les dades.

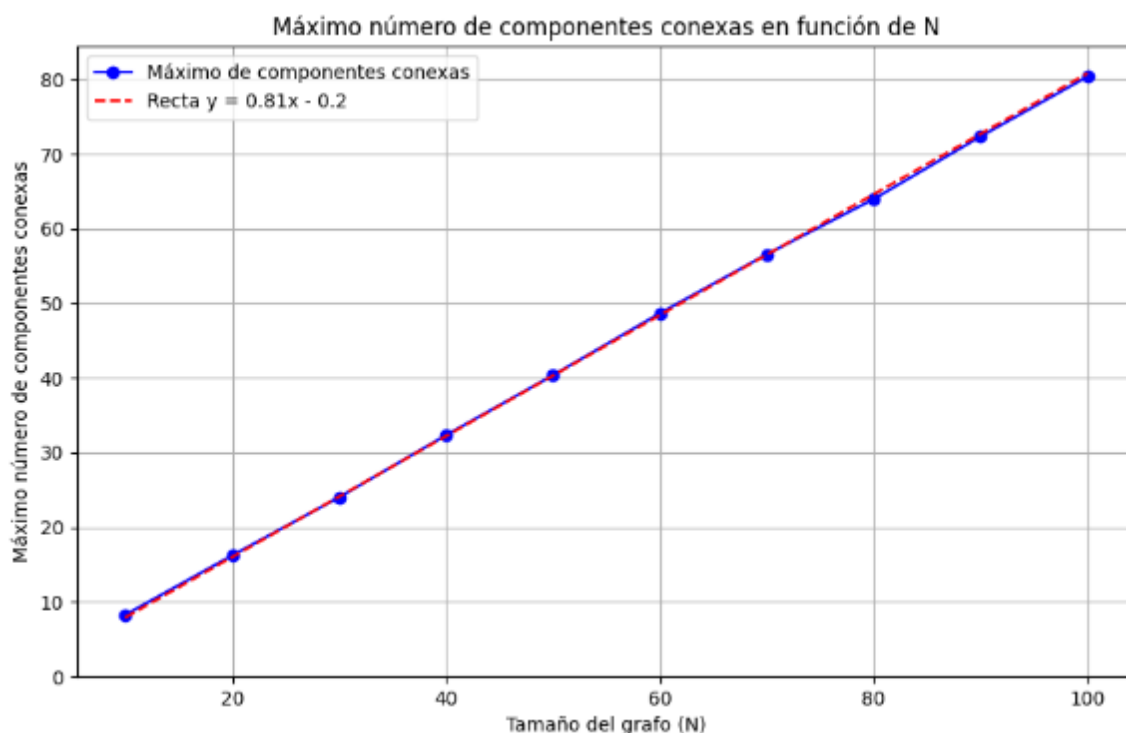


Figura 5.9

5.2 Estudi per arestes

5.2.1 Dades descriptives

Ara farem el mateix estudi, però per arestes, per tant, ja sabem quin tipus de gràfiques tenim. A la figura 10 es mostra el mateix que a la figura 4 però percolant arestes. Cal anotar que en el cas de l'estudi de vèrtex, la probabilitat $p \sim$ probabilitat que el vèrtex sobrevisqui, i en el cas de les arestes $p \sim$ probabilitat que l'aresta entre dos vèrtexs existeixi.

Igualment, es fan $k = 10$ proves per a cada valor de p , i hi ha 100 valors de p .

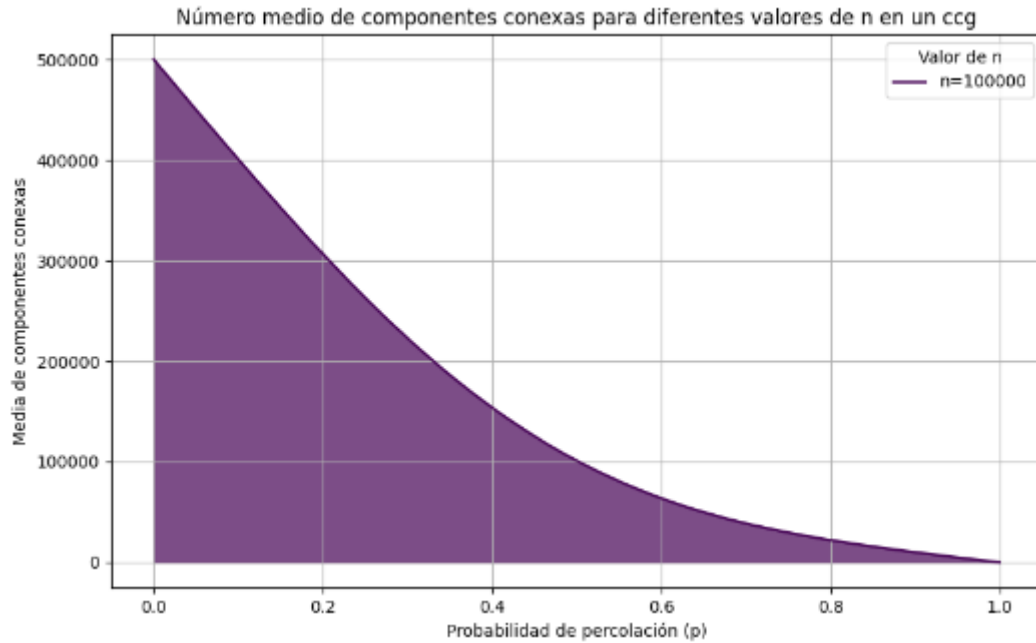


Figura 5.10

Anàlogament, a l'estudi del vèrtex, les figures 11, 12, 13 i 14 mostren el que mateix que les figures 5, 6, 7 i 8; en la que la primera mostra el nombre de components connexes segons la probabilitat de percolació, de $N = 50, 100, \dots, 100000$; la segona mostra un zoom al creixement inicial, la tercera mostra la probabilitat p on s'assoleix el màxim nombre de components connexes, de $N = 50, 100, \dots, 100000$; i l'última figura mostra el màxim nombre de components connexes respecte a N petites.

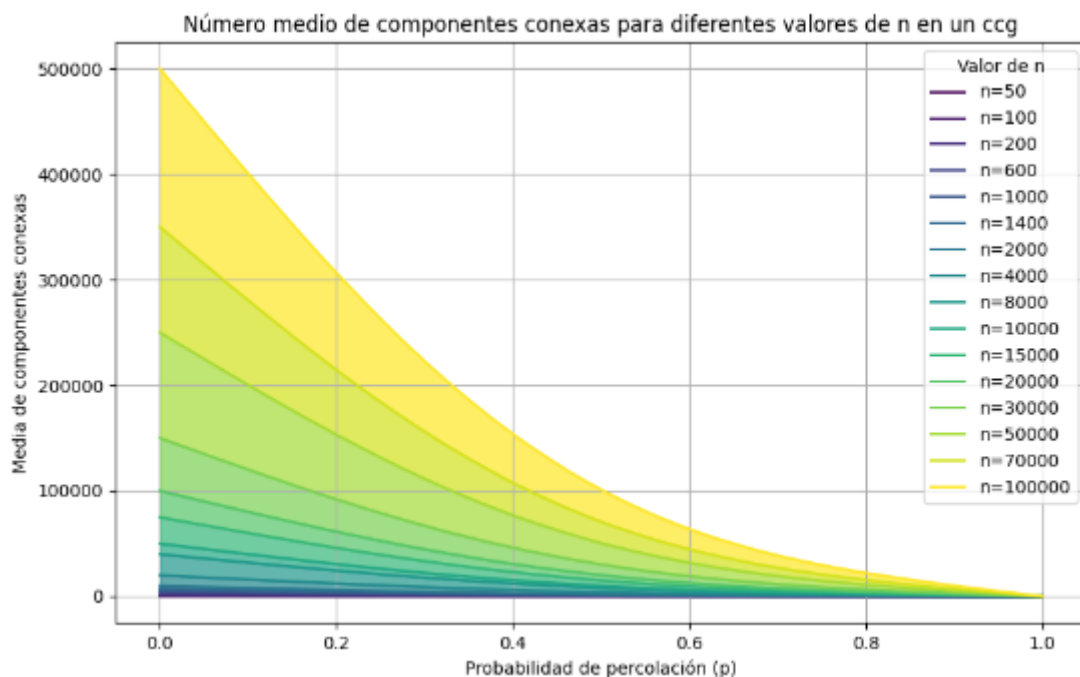


Figura 5.11

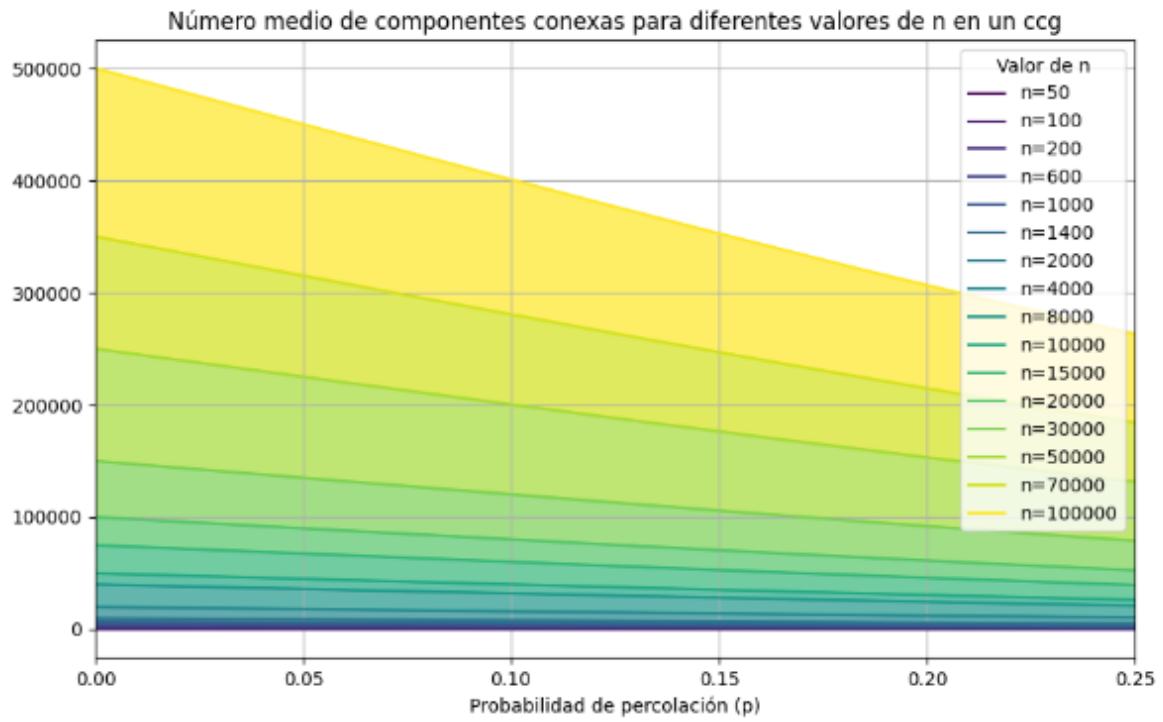


Figura 5.12

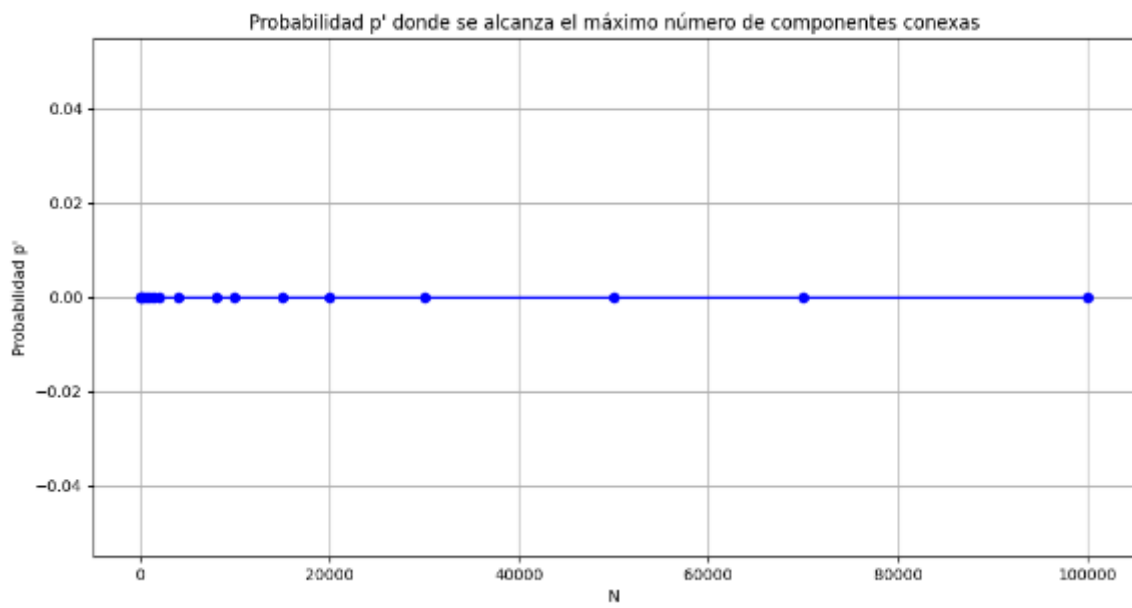


Figura 5.13

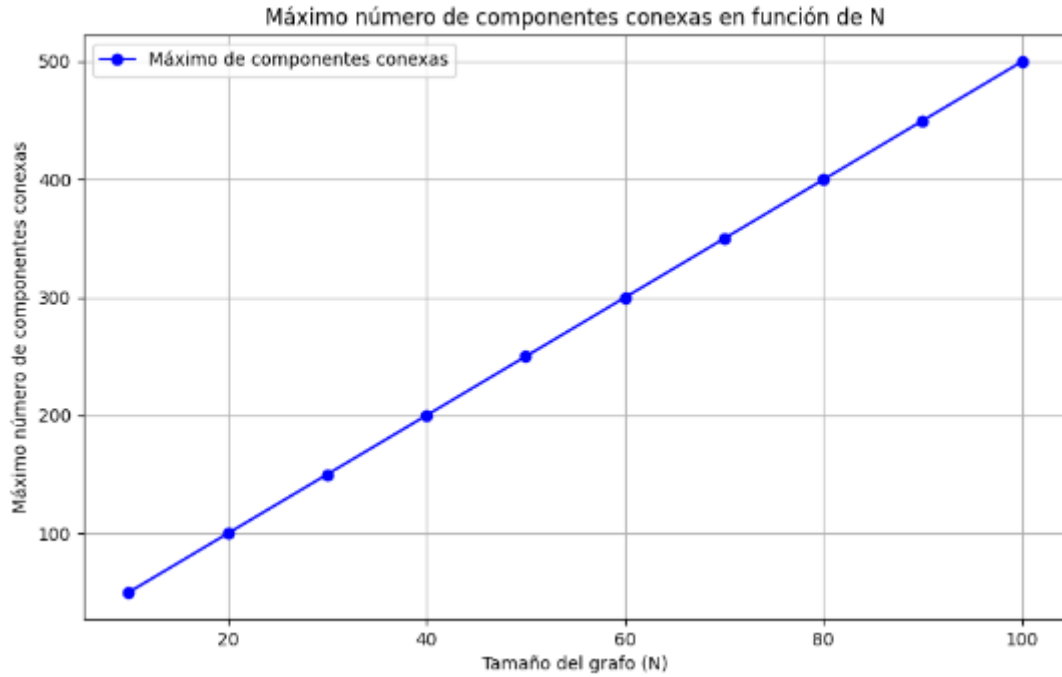


Figura 5.14

5.2.2 Anàlisi de les dades

i) Sobre el comportament de la funció de C_n respecte p

Com en estudis anteriors amb altres models de grafs, es veu que el tram inicial decreixent de la gràfica sembla totalment lineal, per valor de p'_N petits; anàlogament a l'estudi de grafs graella, tenim una recta que aproxima aquesta recta inicial (figura 15), que per $N = 50$, aquesta és $50 - 96.57p$, i es pot extrapolar a qualsevol valor de N .

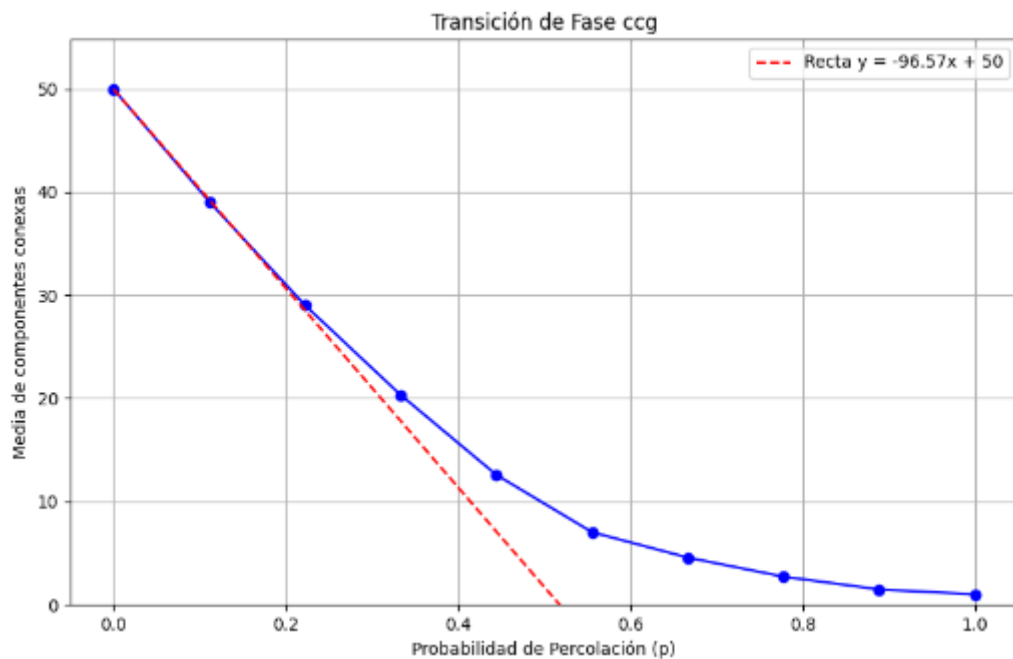


Figura 5.15

Podem veure aquest decreixement inicial i com s'aproxima a una recta per a qualsevol valor de N en la figura 12, en la que es mostra un zoom en el creixement inicial de la funció.

ii) Sobre la probabilitat p'_N on s'assoleix el màxim nombre de components connexes respecte d' N

Experimentar amb la probabilitat p'_N per tal que s'assoleix el màxim nombre de components connexes és gairebé inservible, ja que la hipòtesi quasi segura que tenim consisteix en:

Hipòtesi 1: La probabilitat on s'assoleix el màxim nombre de components connexes p' és exactament a 0.

Com veiem en la figura 13, òbviament **es valida** la hipòtesi 1, ja que per estructura de l'experiment, amb $p'_N = 0$ cap aresta sobreviu i, per tant, el número de $CG' = |V|$.

iii) Sobre el màxim de components connexes en funció de N

En aquesta part de l'estudi, anàlogament a l'estudi de percolació en vèrtex, hem volgut estudiar la recta que aproxima el creixement inicial del màxim número de CG' respecte N , i, com era esperat, compleix una recta perfecta tal que el número de CG' màxim es defineix amb $5N$, com es veu en la figura 16.

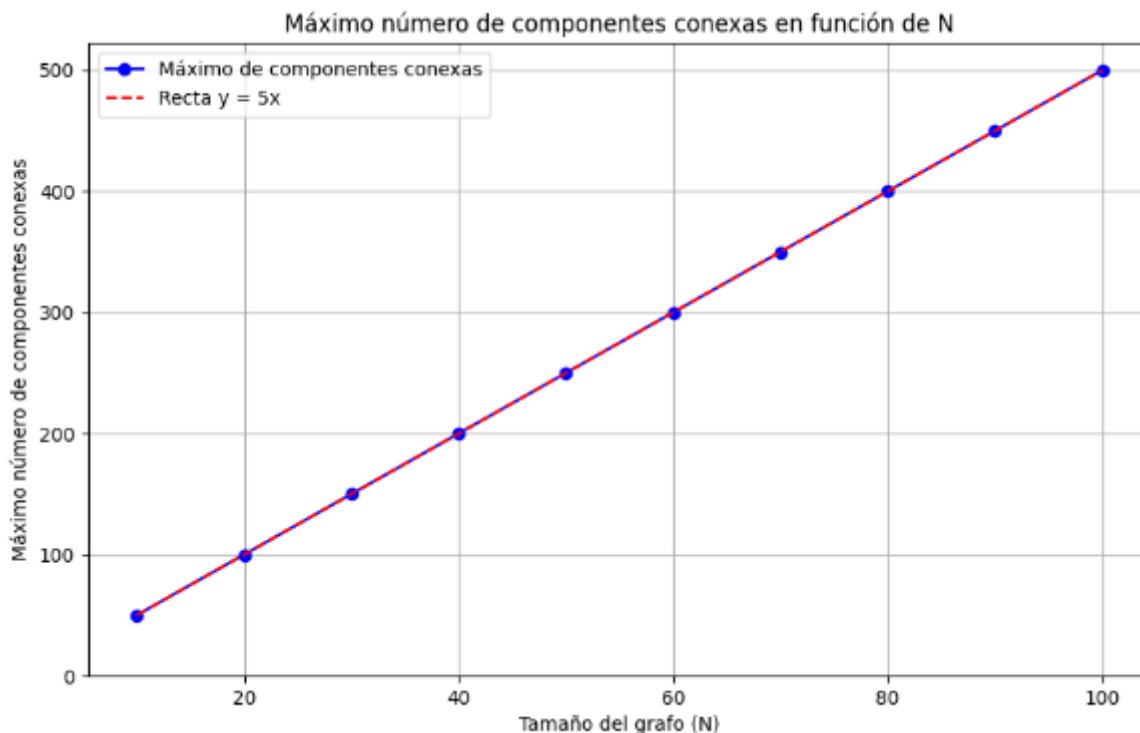


Figura 5.16

5.3 Conclusions:

Com a conclusió, un cop fet tots els estudis sobre els CCGs, podem treure conclusions amb relació a la transició de fase.

Anàlogament, als altres models de grafs prèviament estudiats, depèn de la definició que prenem per transició de fase, n'existeix o no n'existeix en el cas de l'estudi per vèrtex, ja que; si entenem la transició de fase com el valor de p_c en el qual el sistema es comporta radicalment diferent en les dues regions $p < p'$ i $p > p'$, realment no hi ha transició de fase, perquè el comportament del sistema no és abrupte, ni té un canvi radical en cap valor de $[0..1]$. En canvi, si entenem la transició de fase com el valor de p_c en el qual el sistema es comporta diferent en les regions $p < p'$ i $p > p'$, sense caldre que sigui un comportament radical, podríem veure gràficament que la transició de fase per vèrtex s'aproximaria a un valor tal que: fase aproximada a 0.38, on assolix el màxim global de nombre de components connexes la funció.

Com als altres models de grafs, en el cas de les arestes, no existeix cap valor per la transició de fase.

6 Conclusions del treball:

En aquest treball s'ha dut a terme un estudi per a determinar l'existència d'un *transició de fase* en relació amb el nombre de components connexes sota un procés de percolació. L'estudi s'ha dut a terme percolant per nodes i per arestes el graf graella en dues dimensions, el graf graella en 3 dimensions i el Connected Caveman Graph, també s'ha estudiat només sota un procés de percolació per arestes de grafs geomètrics aleatoris.

Transició de fase:

En relació amb la transició de fase, personalment no considerem que n'hi hagi cap clara ni en la percolació per vèrtexs ni arestes, en cap de les diferents espècies de grafs; però, d'altra banda, segons la definició disposada a l'enunciat, que diu el punt p tal que pels punts $p' < p$ el nombre de components connexes són radicalment diferents dels punts $p < p'$, podríem concloure que és el punt p de la gràfica amb més pendent.

Però, això no obstant, si considerem la transició de fase com el punt p on el comportament de la funció és radicalment diferent per punts $p' < p$ que per punts $p < p'$, podríem concloure que la transició de fase seria el punt d'inflexió on la gràfica passa de ser creixent a decreixent.

La conclusió respectiva de cada espècia de graf, es detalla en la conclusió de cada apartat respectiu.

Altres propietats:

Aquest estudi ha permès també observar altres propietats interessants dels grafs estudiats en sotmetre'ls a percolacions per nodes i per arestes. Pels grafs deterministes estudiats (graella 2D, graella 3D i Connected Caveman Graphs), s'ha estudiat l'evolució del nombre màxim de components connexes C_N en funció de la mida dels grafs N sota el procés de percolació i s'ha pogut establir un ajust polinòmic per a la corba C_N . També s'ha estudiat la probabilitat p'_n en què s'assoleix el màxim nombre de components connexes després per a cada graf de mida N . Pel que fa als grafs geomètrics aleatoris, s'ha dut a terme un estudi del *knee on the curve* de la corba de components connexes per als diferents valors d' N estudiats, això ha permès establir a quins valors tendeix el *knee on the curve* en RGGs.

Àmbit personal:

Finalment, en un pla més personal i com a estudiants, aquest treball ens ha introduït en el món de l'experimentació i la recerca en el camp de les Ciències de la Computació. En el transcurs d'aquest treball hem après sobre els diferents grafs estudiats, les seves propietats i com generar-los, així com a gestionar resultats no esperats inicialment com l'absència del *transició de fase*. Aquest treball ha eixamplat també el nostre horitzó de coneixement introduint-nos conceptes com la percolació o donant-nos a conèixer grafs que ni tan sols ens haguéssim pogut imaginar que existissin. Aquí volem

incloure no només els grafs estudiats com el Caveman Graph, sinó també alguns de ben curiosos que ens vam plantejar estudiar com el graf Sudoku. Per acabar, amb aquest treball hem guanyat maduresa acadèmica i autonomia a l'hora de buscar i treballar amb eines o llibreries noves així com a l'hora de plantejar-nos preguntes més enllà del que el treball demanava. Finalitzem aquest treball amb l'esperança que els propers estudis que encetem ens siguin tan profitosos pel que fa a l'aprenentatge com ho ha estat aquest projecte.

Bibliografia

En aquest apartat presentem les fonts d'on hem tret tota la informació necessària per fer el projecte.

Graella 2D & Graella 3D

- Dall, J., & Christensen, M. (2002). Random geometric graphs. *Physical Review E*, 66(1), 016121. <https://arxiv.org/pdf/cond-mat/0203026>.
- García Pérez, A. (2023). *Estadística básica con R* (2^a ed.). UNED, Col. Grado.

RGG (Random Geometric Graph)

- Wikipedia contributors. (n.d.). Random geometric graph. Wikipedia. Retrieved October 24, 2024, from https://en.wikipedia.org/wiki/Random_geometric_graph.
- Zhang, P. (2014). Random geometric graph example with collaboration distance. *ResearchGate*. https://www.researchgate.net/figure/Random-geometric-graph-example-with-collaboration-distance-rn_fig1_261470943.
- Dall, J., & Christensen, M. (2002). Random geometric graphs. *Physical Review E*, 66(1), 016121. <https://arxiv.org/pdf/cond-mat/0203026>.
- Díaz, J., Penrose, M., Petit, J., & Serna, M. (2002). Approximating layout problems on random geometric graphs. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*. <https://www.cs.upc.edu/~diaz/papersd/SODAFinal.pdf>.

CCG (Connected Caveman Graphs)

- Weisstein, E. W. (n.d.). Caveman graph. *MathWorld*. Retrieved October 24, 2024, from <https://mathworld.wolfram.com/CavemanGraph.html>.
- Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105(2), 493-527. <https://doi.org/10.1086/210318>.
- Connected caveman graph. (2023). *GraphScope Documentation*. <https://graphscope.io/>.
- Connected caveman graph documentation. (2024). Spynegroup. <https://spynegroup.github.io/2024>.

Eines útils (UTILS)

- Wikipedia contributors. (n.d.). Taula de símbols matemàtics. Viquipèdia. Retrieved October 24, 2024, from https://ca.wikipedia.org/wiki/Taula_de_símbols_matemàtics.
- NetworkX Developers. (n.d.). *NetworkX documentation*. Retrieved October 24, 2024, from <https://networkx.org/documentation/stable/reference/index.html>.
- Matplotlib Developers. (n.d.). *Matplotlib documentation*. Retrieved October 24, 2024, from <https://matplotlib.org/stable/index.html>.
- NumPy Developers. (n.d.). *NumPy documentation*. Retrieved October 24, 2024, from <https://numpy.org/doc/>.

- pandas Documentation Team. (n.d.). *Pandas documentation*. Retrieved October 24, 2024, from <https://pandas.pydata.org/docs/>.
- LaTeX Project Team. (n.d.). *LaTeX documentation*. Retrieved October 24, 2024, from <https://www.latex-project.org/help/documentation/>.