

# Proyecto de Algoritmia: Conectividad y Percolación

Alejandro Alarcón y Martí Vall

G1307

Otoño 2020/2021

## ÍNDICE

<b>INTRODUCCIÓN</b>	<b>3</b>
<b>GENERADORES</b>	<b>4</b>
GENERADOR ALEATORIO BÁSICO SEGÚN $G^*$	4
GENERADOR DE GRAFOS GEOMÉTRICOS ALEATORIOS	7
<b>PERCOLACIÓN DE ARISTAS</b>	<b>9</b>
<b>PERCOLACIÓN DE NODOS</b>	<b>11</b>
<b>PERCOLACIÓN EN MATRICES <math>N \times N</math></b>	<b>13</b>
<b>ALGORITMOS</b>	<b>15</b>
<b>CONCLUSIÓN</b>	<b>15</b>

## INTRODUCCIÓN

En este proyecto el objetivo será el estudio experimental de la transición de fase en el número de componentes conexas de un grafo sometido a un proceso de percolación.

Este proceso se puede aplicar tanto a aristas como a nodos, y consiste en su eliminación dado un coeficiente  $q$ , que representa la probabilidad de que dicha arista o nodo en cuestión sea eliminado.

De este modo, una vez realizado el proceso de percolación de aristas de un grafo  $G = (V, E)$ , obtenemos un grafo  $G' = (V', E')$  tal que  $V' = V$  y  $E'$  es un subconjunto de  $E$ .

Por otro lado, en el caso de la percolación de nodos, el hecho de que desaparezca un nodo, implica que también desaparecerán todas las aristas que están conectadas a él. de modo que para un grafo  $G = (V, E)$  obtendremos otro grafo  $G' = (V', E')$  tal que  $V'$  es un subconjunto de  $V$  y  $E'$  un subconjunto de  $E$ .

## GENERADORES

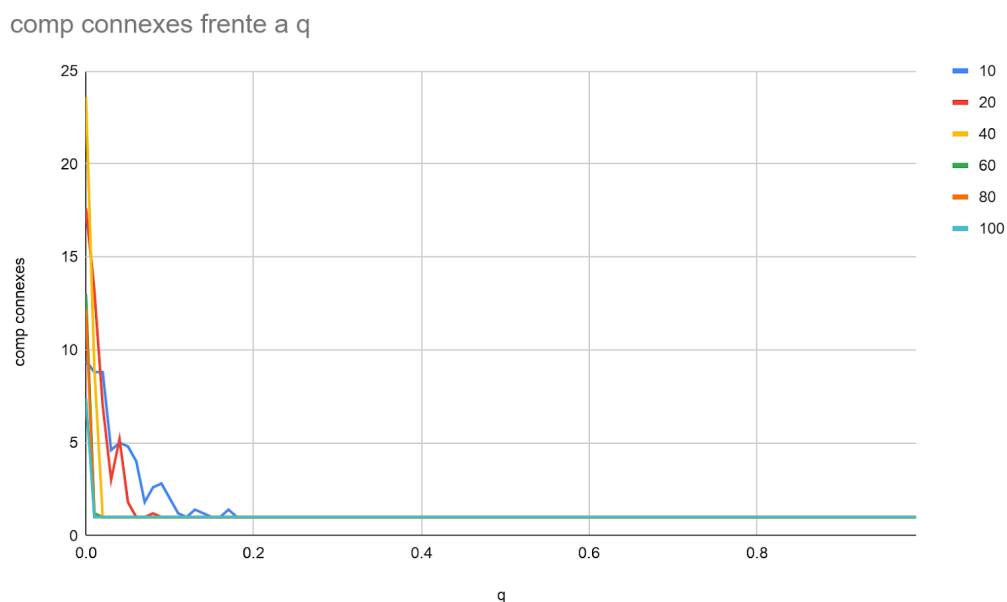
En primer lugar, empezaremos hablando de los generadores de grafos aleatorios que hemos usado. Aunque se han estudiado muchos modelos, al final se decidió, para mantener la simplicidad del proyecto y teniendo en cuenta los requisitos del enunciado, seleccionar el modelo aleatorio basado en  $G^*$  y el modelo de grafos geométricos aleatorios que se menciona en el enunciado.

Sin embargo, también se estudiaron modelos como el Erdős-Rényi, Watts–Strogatz o Barabási–Albert, además de otros de grafos bipartitos o regulares. Se consideró que las propiedades de estos grafos no eran las adecuadas para la experimentación en este proyecto.

### GENERADOR ALEATORIO BÁSICO SEGÚN $G^*$

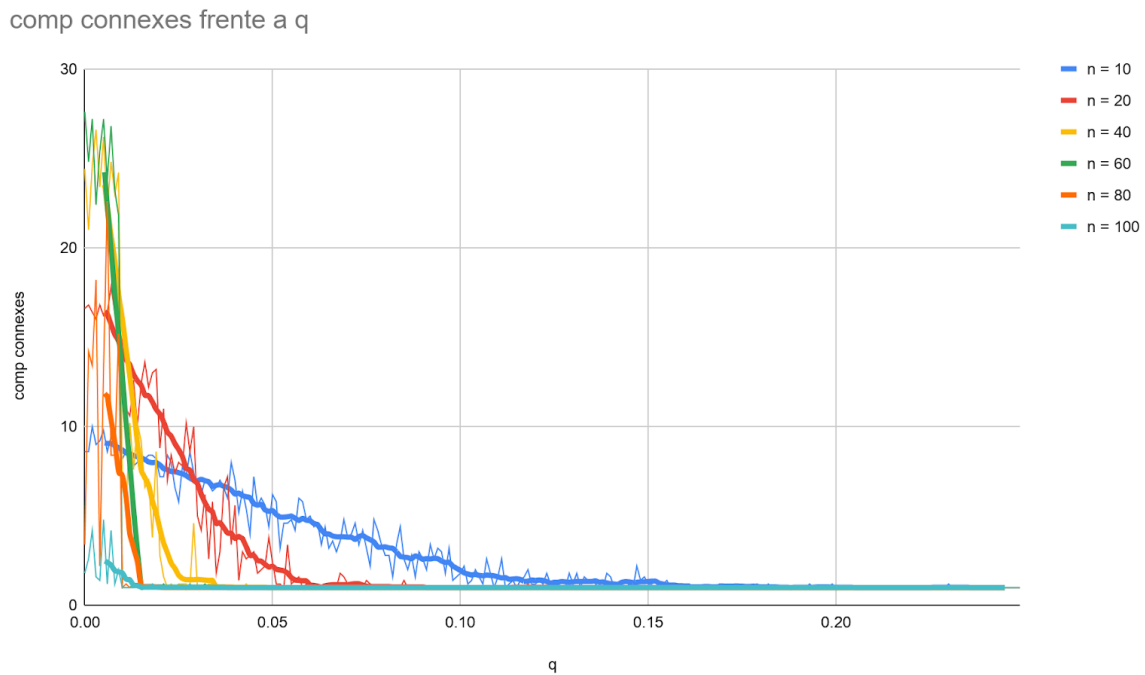
Este generador consiste en generar un grafo completo (o vacío) con un tamaño que se pasa como parámetro  $y$ , para cada par de vértices  $u,v$ , generar un coeficiente  $q'$  que se comparará con el parámetro  $q$  para determinar si la arista existe o no. Básicamente es un generador totalmente aleatorio que no tiene en cuenta nada más que un valor generado aleatoriamente para decidir si una arista existe o no. Cabe destacar que debido a que hemos representado los grafos como matrices de adyacencia el algoritmo que genera la arista se aplica dos veces por cada arista, una para la posición  $[u][v]$  y otra para la posición  $[v][u]$ . Lo que como se podrá ver reflejado en los resultados hace que los grafos tiendan a ser más conexos de lo que deberían.

A partir de este generador se han realizado dos estudios, uno relacionado con el nombre de componentes conexos que aparecen en el grafo generado según el parámetro  $q$  y otro que hace referencia a la probabilidad de que el grafo sea conexo según el parámetro  $q$ .



En este primer gráfico se muestra la variación en la cantidad de componentes conexas con incrementos de 0.01 en el valor del parámetro  $q$  por iteración. Como se puede apreciar, el número de componentes conexas se reduce drásticamente desde valores muy cercanos a 0 de  $q$ . Así que decidimos realizar una segunda recogida de datos con incrementos del parámetro  $q$  de 0.001 para poder apreciar mejor la transición en el número de componentes conexas.

Así pues, obtenemos el siguiente gráfico:

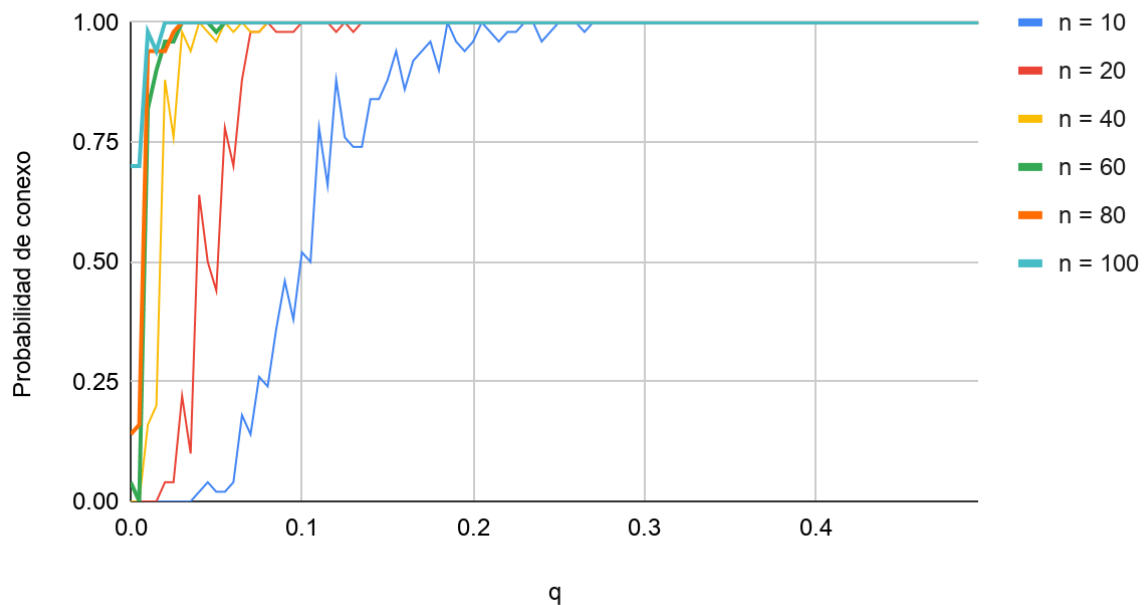


En este segundo gráfico que recoge un intervalo más pequeño, de 0 a 0.25, se puede apreciar mejor la tendencia para cada valor de  $n$ , y podemos ver cómo a medida que aumenta el tamaño del grafo se observa una posible transición de fase más clara, sin embargo, nada lo suficientemente claro como para sacar conclusiones.

En definitiva, con los grafos generados con el generador básico no obtuvimos datos relevantes como para llegar a una conclusión.

En lo referente al estudio de la probabilidad de que el grafo sea conexo obtuvimos el siguiente gráfico:

Probabilidad de conexo frente a  $q$



{Gráfico que muestra la probabilidad de que un grafo sea conexo vs la variación del parámetro  $q$ , con saltos de 0.005, generador básico.}

Aquí sí que se puede apreciar mejor una clara transición de fase, que como se puede observar se vuelve más prematura a medida que aumentamos el tamaño del grafo. Sin embargo, para todos los tamaños se ve claramente como existe una transición de fase aunque no podemos dar un valor de  $x$  generalizado para esta transición con los datos obtenidos

Con estos primeros dos estudios utilizando el generador básico pudimos hacernos una idea del comportamiento que debería apreciarse en los resultados, aunque no conseguimos sacar ninguna conclusión clara. Aun así nos fue de mucha utilidad como referencia para el estudio realmente relevante en este proyecto en el que utilizamos el generador de grafos geométricos aleatorios.

## GENERADOR DE GRAFOS GEOMÉTRICOS ALEATORIOS

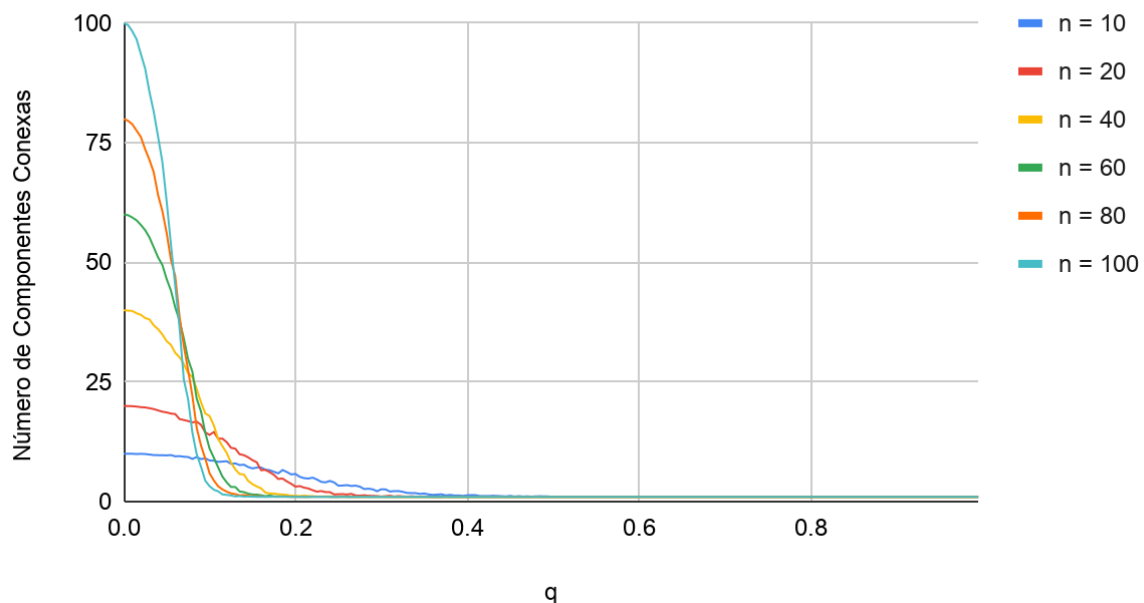
Este generador consiste en asignar tantos puntos como nodos deseemos tener en nuestro grafo a un espacio  $n$ -dimensional con unas coordenadas aleatorias pertenecientes a  $[0,1]$  para cada dimensión. Posteriormente, se calcula la distancia euclidiana para cada par de vértices  $u,v$  pertenecientes a  $V$  y, si esta distancia es menor que un radio  $r$  pasado como parámetro, se genera una arista que unirá  $u$  con  $v$ .

Este modelo tiende a generar conjuntos o clusters de nodos muy relacionados entre sí y, a medida que va subiendo el radio se empiezan a interconectar estos conjuntos. Esto es debido a que al depender el factor que genera las aristas de la distancia, primero se crean estos clusters de alta modularidad en grupos de vértices cercanos, y a medida que aumente el factor  $r$  se conectan estos clusters por los extremos de estos hasta que se termina interconectando todo el grafo.

Este es el modelo que se ha usado para realizar la mayoría de los estudios. Sin embargo, en algunos casos también se plasman los resultados obtenidos con el modelo anterior.

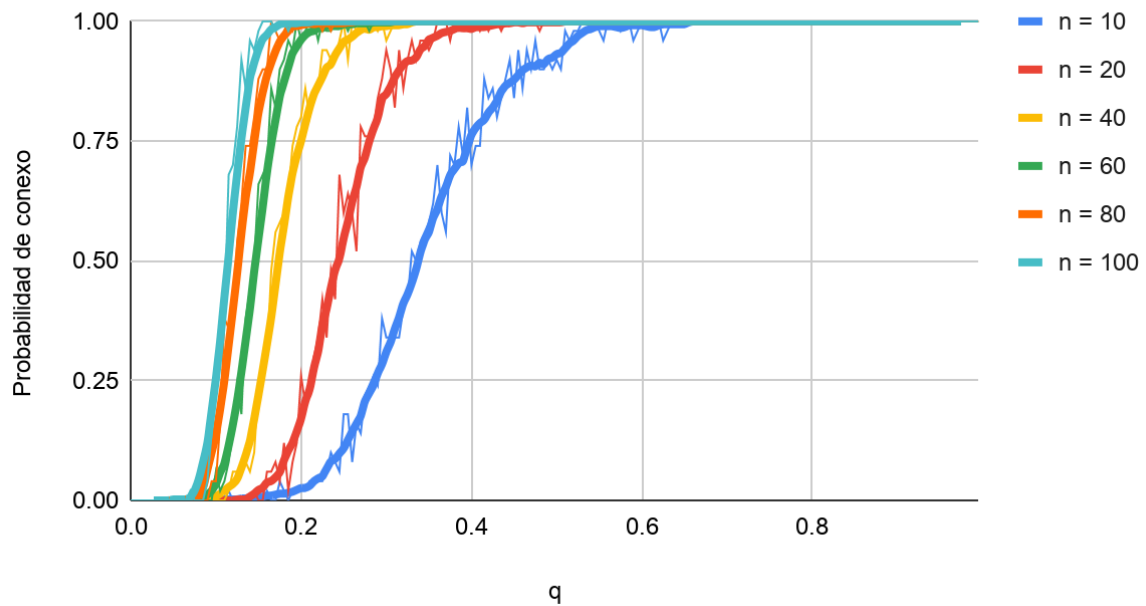
En las siguientes gráficas se pueden observar los resultados obtenidos para la cantidad de componentes conexas y la probabilidad de que el grafo sea conexo, ambos según el parámetro  $r$  (radio).

### Número de Componentes Conexas frente a $q$



{Gráfica que muestra la cantidad de componentes conexas que aparecen en un grafo geométrico aleatorio según la variación del parámetro  $r$ .}

## Probabilidad de conexo frente a $q$



{Gráfica que muestra la probabilidad de que el grafo resultante sea conexo con el modelo geométrico aleatorio según la variación del parámetro  $r$ }

En estos dos gráficos, a diferencia de los gráficos obtenidos con el generador básico, vemos cómo aparece una clara transición de fase, sobretodo en el gráfico de la conectividad, donde también podemos apreciar que a medida que aumenta el tamaño del grafo la transición de fase es más prematura, viendo como cada vez se van acercando más las gráficas de los diferentes tamaño hacia un punto de transición de fase de más o menos  $x = 0.1$ .



## PERCOLACIÓN DE ARISTAS

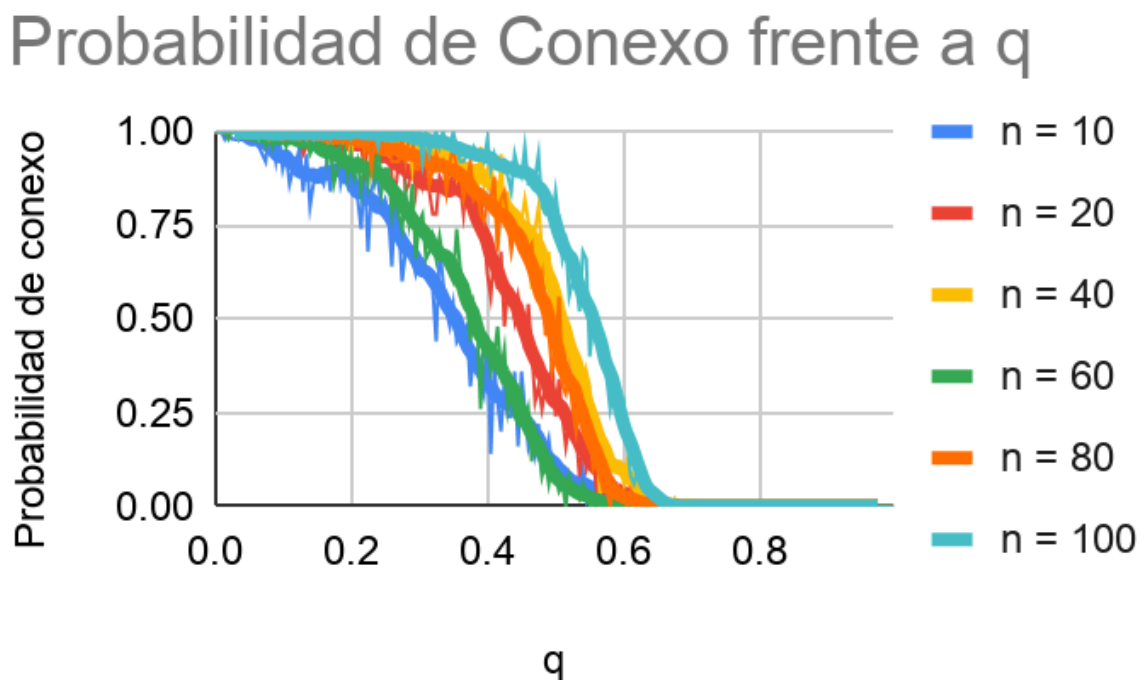
Como se ha mencionado previamente, el proceso de percolación por aristas, consiste en eliminar aristas de un grafo según una probabilidad  $q$  de que la arista sea eliminada.

Este proceso no puede crear nuevas aristas, simplemente puede eliminar las que existían en el grafo original. De este modo, después de este proceso, tendremos, como mínimo, el mismo número de componentes conexas que tenía el grafo original.

Para los estudios realizados en este apartado y el siguiente, se ha partido siempre de grafos conexos. De manera, que al estudiar la conectividad del grafo tras la percolación, el único parámetro que influya en la conectividad del grafo sea dicha percolación.

También es importante tener en cuenta que los resultados pueden variar mucho según la densidad del grafo original, pues a más aristas tenga el grafo más difícil será que la percolación afecte a su conectividad. Por este motivo, se ha realizado el estudio para grafos creados con distintas densidades. En el primer caso se ha estudiado un coeficiente para el generador que genera grafos conexos poco densos. Para ello utilizando los datos de los estudios anteriores hemos usado para cada tamaño el coeficiente más bajo que hiciera que el grafo fuera conexo en un 99% de los casos (después de generar el grafo comprobamos siempre que fuera conexo y si no lo era se generaba otro grafo). En el segundo caso se ha aplicado un radio de 0.6 y en el tercero uno de 0.8, lo que genera grafos muy densos.

Del primer caso hemos obtenido el siguiente resultado:

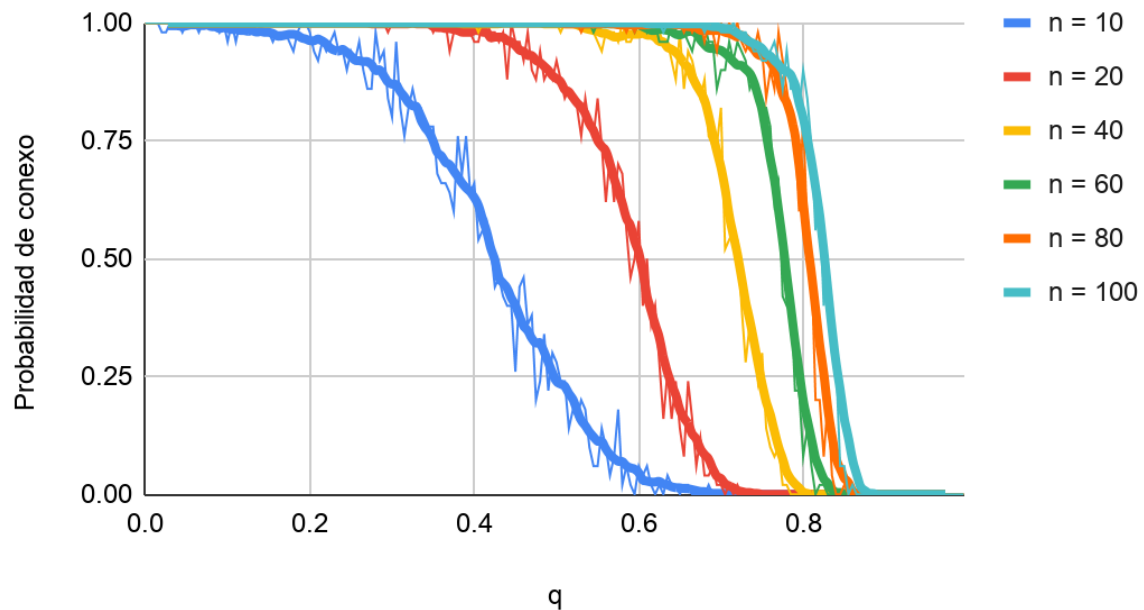


En este primer caso podemos volver a ver un resultado parecido al del estudio anterior. Se observa claramente como entre los valores de 0.4 y 0.6 del parámetro  $q$  de probabilidad hay

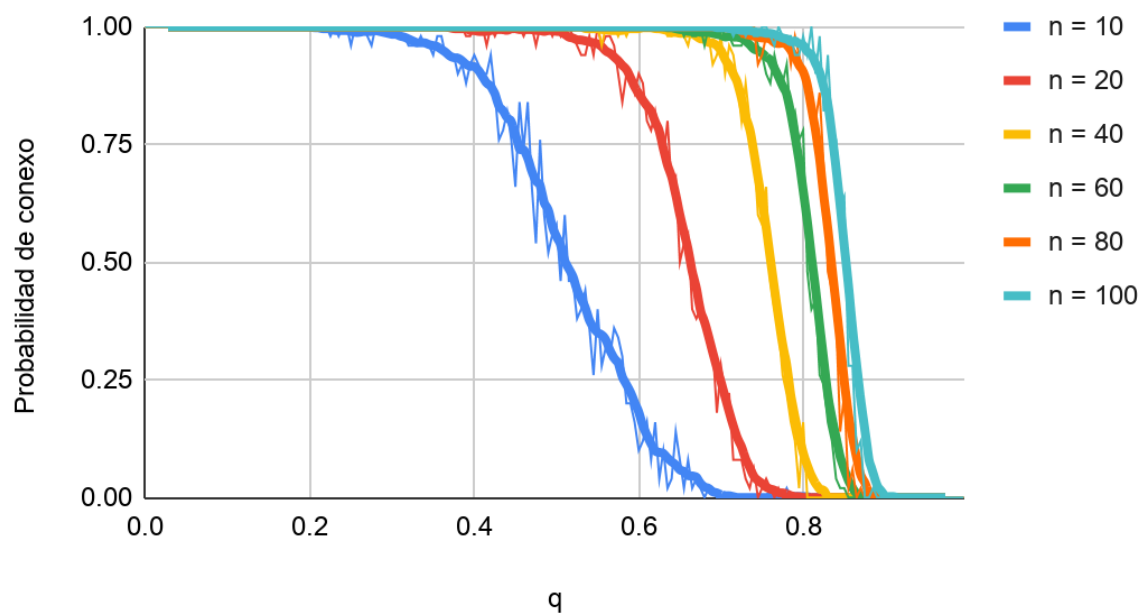
una transición de fase en la conectividad de los grafos, pasando de ser más probable que sea conexo a que sea más probable que sea inconexo. Sin embargo, del mismo modo que en el estudio anterior no se termina de apreciar un valor exacto para esa transición, sino que sigue siendo un intervalo en el que se mueven dependiendo del tamaño del grafo.

Por otro lado, en el segundo y tercer caso obtenemos resultados muy parecidos entre ellos como se puede observar en los gráficos a continuación:

### Probabilidad de Conexo frente a $q$ (coef = 0.6)



### Probabilidad de Conexo frente a $q$ (coef = 0.8)



En estos dos casos con grafos más densos podemos ver que aunque al principio la transición de fase depende del tamaño del grafo, para tamaños más grandes ya se empieza a observar un valor más preciso para la transición de fase. Así pues, vemos que para estos grafos más densos el valor umbral de  $x$  (threshold  $f(x) = 0.5$ ) de la transición de fase es alrededor de 0.85.

## PERCOLACIÓN DE NODOS

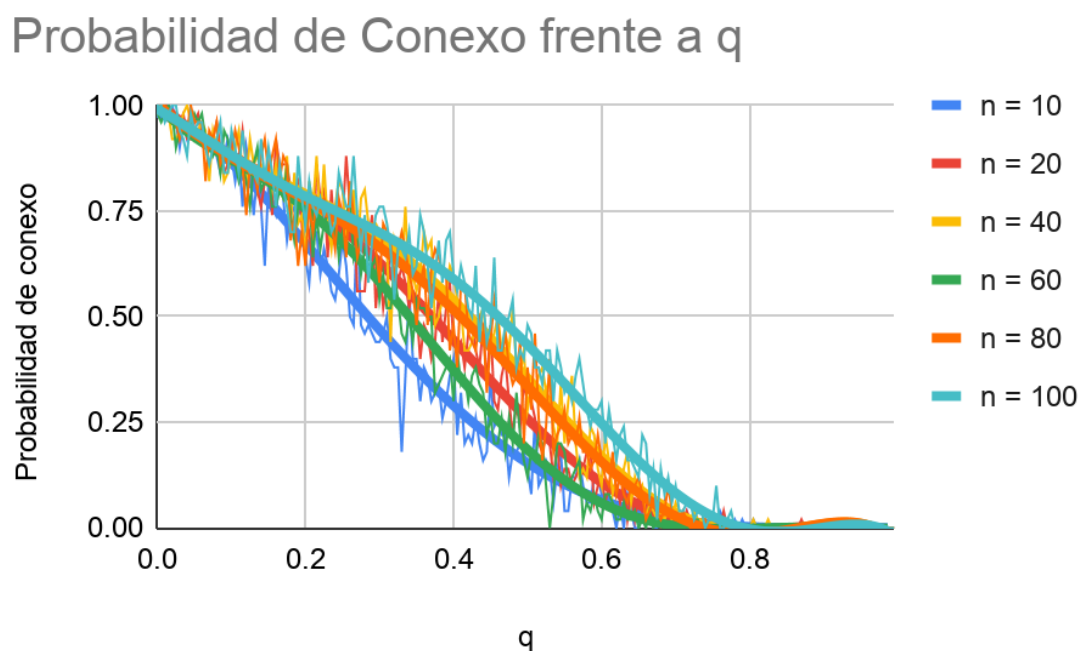
Este proceso es muy similar a la percolación por aristas, sin embargo, en este caso, lo que se tratan son los nodos, y el hecho de que un nodo  $u$  perteneciente a  $V$  desaparezca, implica que todas las aristas que tienen a  $u$  como componente, también desaparecerán.

Cabe destacar, que en este caso lo que determina si un grafo sigue siendo conexo o no después de haber aplicado la percolación, es el hecho de si se ha eliminado algún nodo bisagra o no. Los nodos bisagra son nodos que unen dos subconjuntos del grafo, de manera que si se eliminara ese nodo quedarían los dos subconjuntos separados y el grafo ya no sería conexo.

Este modelo tampoco incrementará en número de nodos, por lo que el conjunto de vértices del grafo resultante siempre será un subconjunto de los del grafo original.

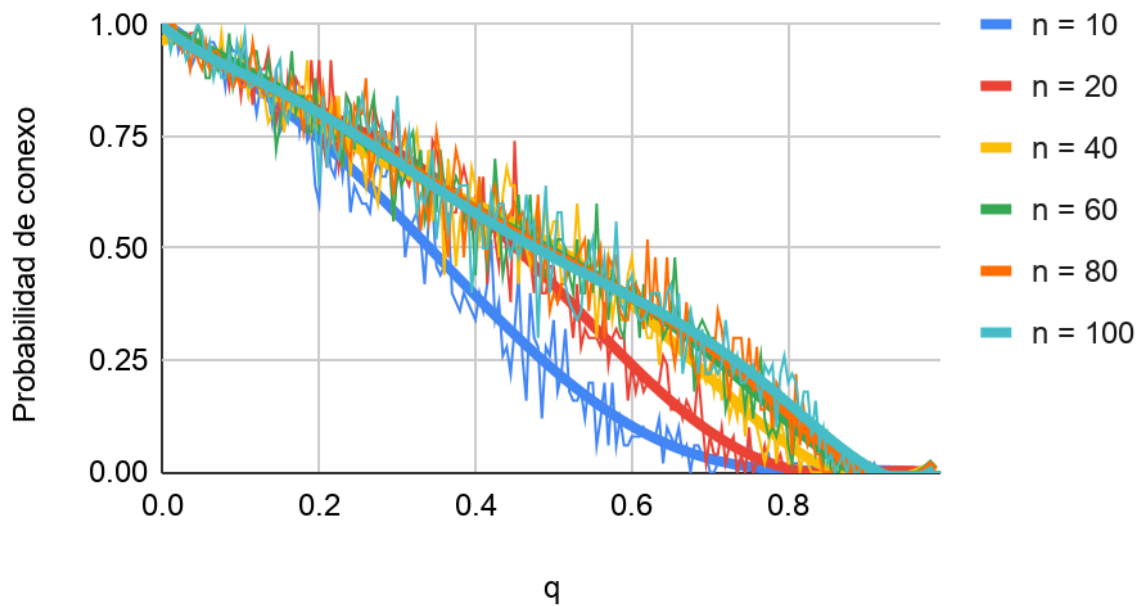
Para este apartado realizaremos los tres mismos casos que para el apartado anterior, un primer caso con grafos poco densos, y otros dos casos con grafos cada vez más densos (concretamente con los mismos coeficientes 0.6 y 0.8 que en los casos del apartado anterior).

De este modo obtenemos los siguientes resultados:

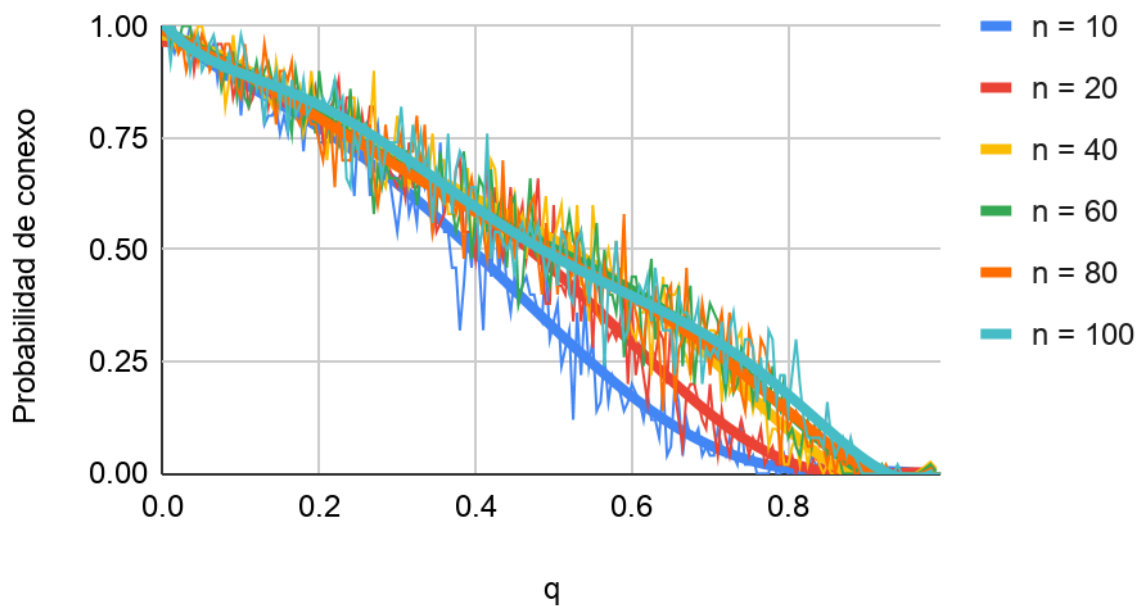


{Gráfico creado utilizando valores umbral}

### Probabilidad de Conexo frente a $q$ (coef = 0.6)



### Probabilidad de Conexo frente a $q$ (coef = 0.8)



En este apartado los tres casos dan resultados muy parecidos entre sí, y por otro lado, muy diferentes a los resultados obtenidos en los apartados anteriores. Aquí podemos ver como no existe ningún tipo de transición de fase exagerada, sino que por el contrario el se observa un comportamiento con una clara tendencia lineal decreciente. Es evidente, que este comportamiento tan distinto se debe al hecho de que estamos eliminando nodos (y sus

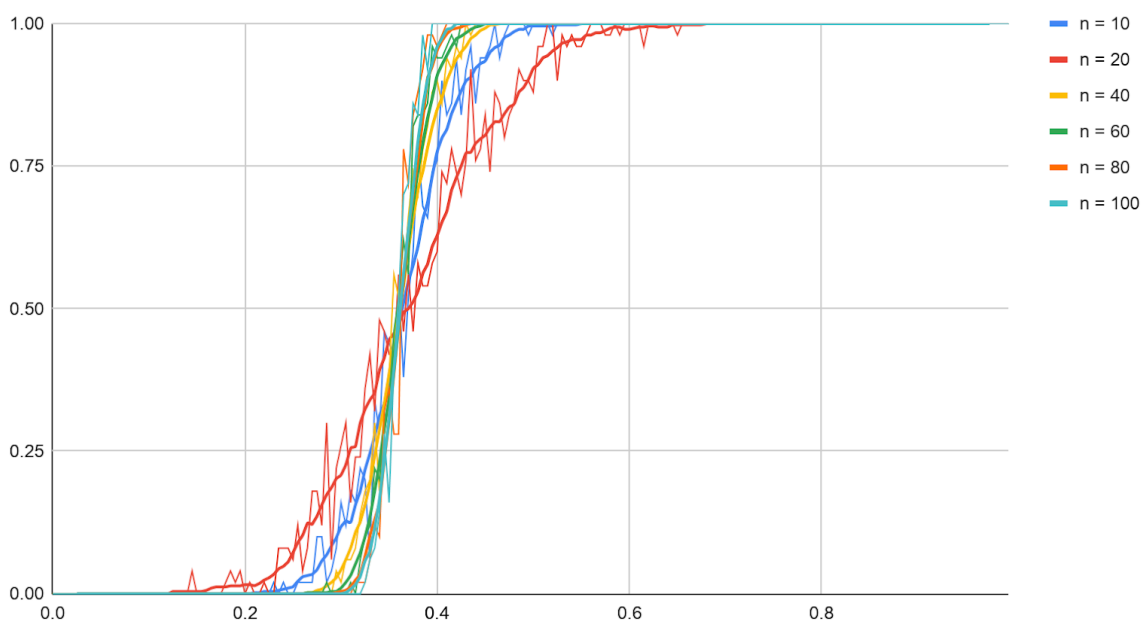
respectivas aristas) en vez de aristas, de manera que a diferencia de en la percolación de aristas, donde al eliminar una arista siempre se reducía la densidad del grafo y por lo tanto se disminuía la probabilidad de que fuera conexo, en este caso no tiene porqué, pues se puede eliminar perfectamente un nodo que solo estuviera conectado con otro nodo, reduciendo el tamaño del grafo pero sin afectar en nada a su conectividad. Además de esto, el hecho de eliminar nodos implica que para que el grafo sea conexo, menos nodos van a tener que estar conectados.

## PERCOLACIÓN EN MATRICES $N \times N$

Finalmente, en este último estudio trabajamos con otra interpretación. En este caso, representamos los grafos como matrices  $n \times n$  donde cada posición de la matriz es un nodo, y cada nodo está abierto con una probabilidad  $q$  o cerrado con una probabilidad de  $1-q$ . Si un nodo está abierto significa que está conectado con todos los nodos abiertos de su alrededor (entendiendo los de su alrededor como los directamente colindantes). Así pues, en este apartado vamos a estudiar la probabilidad de que después de la percolación siga existiendo un camino entre la parte superior de la matriz y la parte inferior.

Para generar estas matrices usamos como base el código de los generadores y recogimos datos para ambos generadores obteniendo los siguientes gráficos:

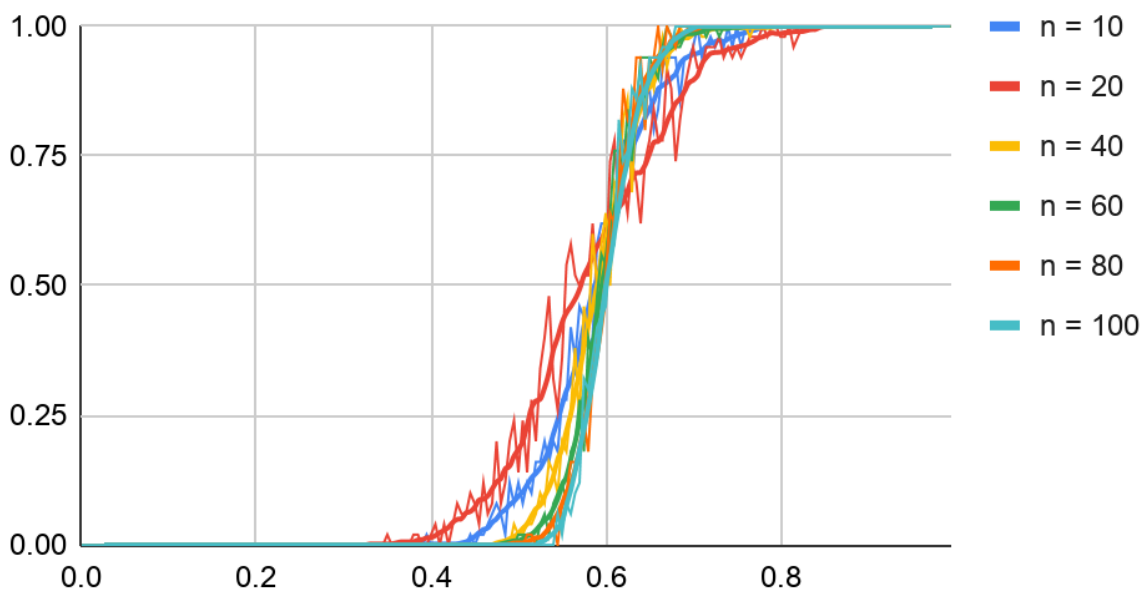
Probabilidad de percolación frente a  $q$



{Matrices generadas con el generador básico}

En este primer gráfico vemos como ya se puede apreciar una clara transición de fase, sin embargo, si lo comparamos con los datos de otros trabajos<sup>1</sup> [1] vemos como el valor exacto para el umbral de la transición de fase ( $f(x) = 0.5$ ) en este caso debería ser de  $x = 0.593$  y en cambio en nuestro gráfico ese valor es de alrededor de  $x = 0.35$ . Sin embargo, con el generador de grafos geométricos aleatorios obtuvimos resultados mucho mejores.

## Probabilidad de percolación frente a $q$



{Matrices generadas con el generador de grafos geométricos aleatorios}

Aquí podemos ver como, igual que en el otro gráfico, se ve claramente que hay una transición de fase, pero en este caso sí que se ve que el valor umbral de la transición es mucho más próximo al correcto, se puede observar a ojo que es alrededor de  $x = 0.6$ . Siendo el valor real  $x = 0.593$ , podemos decir que hemos obtenido un resultado bastante correcto.

---

<sup>1</sup> <https://algs4.cs.princeton.edu/lectures/keynote/15UnionFind-2x2.pdf>

## ALGORITMOS

Para la implementación de los experimentos que se han tenido que llevar a cabo para la realización del proyecto, se han usado principalmente algoritmos enfocados a grafos no dirigidos. Los principales han sido DFS (Depth-First Search) y BFS (Breadth-First Search).

Adicionalmente, se ha usado un algoritmo basado en BFS (se podría usar también DFS) para contar el número de componentes conexas de un grafo. Este algoritmo también nos ayuda a determinar si un grafo es conexo, ya que un grafo conexo tiene una única componente conexa.

Por último en lo que se refiere a algoritmos, también se han implementado dos generadores de grafos aleatorios. El primero, que etiquetamos como Generador Básico Basado en  $G^*$ , se basa en determinar para cada par de vértices del grafo (representados con los puntos) si existirá o no una arista que los una.

El segundo, crea un espacio  $n$ -dimensional (en nuestro caso bidimensional) y sitúa puntos aleatoriamente en ese espacio. Posteriormente, se determinará si existe una arista entre dos puntos según la distancia entre ellos. Como se mencionó previamente, este generador lleva asociado un fenómeno conocido como clusterización.

Para terminar, en cuanto a las estructuras de datos, se han creado algunos *structs* para almacenar los datos que se iban generando en los experimentos. También se ha definido el grafo como un vector< vector<bool> >. En cuanto a las estructuras de la librería estándar, se han usado vectores, pilas y colas.

## CONCLUSIÓN

En este estudio hemos podido observar en varias ocasiones la aparición de una transición de fase en la conectividad de varios tipos de grafo. Con el primer estudio usando el generador básico no hemos obtenido resultados concluyentes, pero con la experimentación con el generador de grafos geométricos aleatorios hemos podido ver que aparecía una clara transición de fase en el estudio de la percolación de aristas y en el de matrices de  $n \times n$ . Aun así, en el estudio de percolación de nodos no sólo no ha aparecido ninguna transición de fase sino que hemos observado una clara tendencia lineal en los resultados.

Así pues, vemos que solo aparece transición de fase en la percolación de aristas, puesto que como se ha comentado en el estudio de la percolación de nodos, el efecto de eliminar los nodos no tiene porque implicar la disminución de la conectividad del grafo, por lo que mitiga el efecto y no aparece transición de fase.

## BIBLIOGRAFÍA

Cormen, Thomas H., et al. *Introduction to Algorithms*.

“NetworkX Documentation.” *NetworkX*, [networkx.org/](https://networkx.org/).

“Random Geometric Graph.” *Wikipedia*, Wikimedia Foundation, 15 Oct. 2020, [en.wikipedia.org/wiki/Random\\_geometric\\_graph](https://en.wikipedia.org/wiki/Random_geometric_graph).

“Teoría De Grafos.” *Wikipedia*, Wikimedia Foundation, 15 Sept. 2020, [es.wikipedia.org/wiki/Teoría\\_de\\_grafos](https://es.wikipedia.org/wiki/Teoría_de_grafos).