Universitat Politècnica de Catalunya

Algorísmia

Projecte d'algorísmia

# Transició de fase i components connexes en grafs aleatoris

*Autors:*
Robert Planas
Arnau Ruana
Sven Wallin

20 d'octubre de 2019

# Índex

# 1 Introducció

Considerem dos models per a la generació de grafs no dirigits simples aleatoris.

El primer a considerar és el *Binamial Random Graph (Erdős–Rényi)*. En aquest model, en la versió seleccionada $G(n, p)$, s'afegeix una aresta entre dos nodes qualsevol amb una probabilitat $p$ independent de la resta de nodes.

El segon model a considerar és el *Random Geometric Graph*. En aquest model es generen nodes sobre un espai mètric (en el nostre cas un pla de mida 1x1) i s'afegeix una aresta entre dos nodes si la seva distància euclidiana és menor a $r$.

Hem escrit un programa en C que genera grafs seguint aquests models i troba el número o mida de les seves components connexes. Observem que, incrementant la nostra variable aleatòria, hi ha un punt on la probabilitat que el graf sigui connex es dispara. Aquest punt s'anomena transició de fase del graf, el qual l'estudiem sempre en funció d'una determinada propietat d'aquest.

Els algorismes utilitzats són principalment DFS per trobar les components connexes i algoritmes $\Theta(VE)$ per generar els grafs. Tot i que a classe s'han vist algoritmes que també podrien haver estat útils com ara *Union-Find*, els vam veure després d'escriure el codi i no els hem implementat. Un cop obtenim les dades, hem fet servir *R Studio* per analitzar-les i generar els gràfics corresponents.

No hem tingut en compte variància, mediana, distribució o altres aspectes estadísticament rellevants sinó que per simplificar hem procurat agafar mostres suficientment grans de grafs tal que l'error esperat fos assumible.

## 2 Propietats de la connectivitat

### 2.1 Connectivitat

Algunes propietats trivials dels models BRG i RGG són les següents:

- Si la probabilitat o el radi és 0 el graf mai és connex.
- Si la probabilitat o el radi és 1 el graf sempre és connex en BRG i quasi sempre en RGG.

Això és obvi, ja que en el primer cas no es genera cap aresta i en el segon es generen totes les arestes possibles.

Donat que un graf només pot ser connex si hi ha com a mínim $n-1$ arestes, i donat que el nombre de comparacions entre dos nodes és $\binom{n}{2}$, i que, idealment, multiplicant la probabilitat (només BRG) per el total de comparacions fetes obtenim el nombre d'arestes, tenim que:

$$p\frac{n(n-1)}{2} < n-1 \implies p < \frac{2(n-1)}{n(n-1)} \implies p < \frac{2}{n}$$

Que s'interpreta com, segons tendim $n \to \infty$ el graf no és connex si $p < 2/n$. Aquesta cota, que podem expressar com $\omega(1/n)$, no és gaire útil, ja que no és una cota ajustada, la propietat és necessària, però no suficient per determinar si el graf és connex. (És a dir, fins i tot amb més de $n-1$ arestes el graf podria no ser connex).

Sí que podem concloure, però, que hi ha una relació inversament proporcional entre la mida del graf i la probabilitat que sigui connex, és a dir, que segons $n$ creix el punt de $p$ on ocorre la transició de fase disminueix. Això ho comprovem experimentalment com es pot veure en les gràfiques de la secció 3, on, a mesura que augmentem $n$ disminuïm el rang de l'eix horitzontal.

Experimentalment, notem que tot i dibuixar els gràfics de forma proporcional a $n$, el canvi de fase encara es desplaça. Per tindre una idea més tangible de quin és aquest desplaçament fem l'experiment següent.

Per a 20 valors de $n$, de 5 a 100, generem 1000 grafs. Fem la mitjana de tots ells i generem les gràfiques del canvi de fase de forma homòloga als experiments de la secció 3. A continuació prenem la distància en l'eix vertical de cada punt respecte a la meitat (0.5). En total generem 20 parells de gràfiques com les de la figura 1. De cada una d'elles n'agafem el primer índex tal que la seva imatge és màxima i generem la gràfica dels índexs.

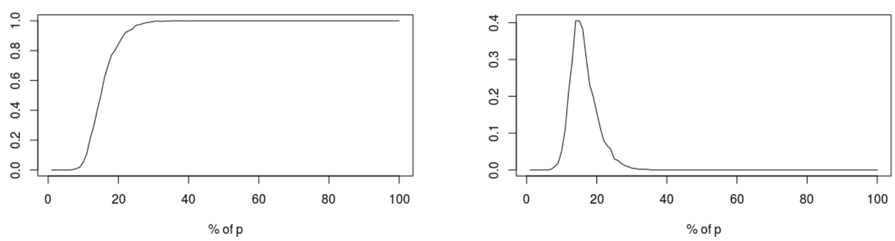


Figura 1: Transició de fase per a un BRG en N=25 i distancia respecte el punt 0.5 de l'eix vertical.

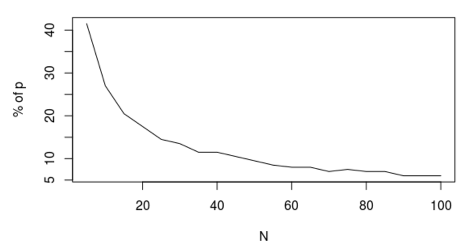


Figura 2: Desplaçament del canvi de fase per diferents valors de N.

Observem com la figura 2 ens mostra una funció de la forma $\Theta(\frac{\ln n}{n})$, que deduïm és la fita ajustada de la connectivitat del graf [1]. Hem trobat que, teòricament un graf generat amb el mètode RGG té una connectivitat fitada en $\sqrt{\frac{\ln n}{\pi n}}$, però no ho hem comprovat experimentalment.

Observem també que la mida en l'eix horitzontal de la transició de fase no és constant sinó que es comporta de forma similar.

[1]Inicialment ens ha semblat que s'ajustava a la funció $\frac{1}{\ln n}$, però hem comprovat mitjançant Geogebra (https://www.geogebra.org/) que $\frac{\ln n}{n}$ s'hi ajusta millor.

### 2.2 Component connexa gegant

Una propietat derivada de l'esmentada amb anterioritat és l'anomenada *component connexa gegant*. Aquesta propietat es dóna quan la probabilitat de formar arestes en un graf aleatori augmenta fins a un punt en el qual el graf passa de tenir moltes components connexes petites a una de molt gran, encara que es mantinguin algunes de més petites.

Tot això és degut al fet que, si quan el graf té moltes components connexes petites hi seguim afegint arestes (augmentem la probabilitat de generar-les), no necessitarem moltes connexions de més per tal que el que abans eren; per exemple, 100 components connexes ara es converteixin en 5 i una d'elles englobant la gran majoria dels nodes del graf. Aquesta component rep el nom de Component Connexa Gegant, fent honor a la seva mida.

A continuació adjuntem una imatge que il·lustra a la perfecció les dues propietats esmentades en aquesta mateixa secció.

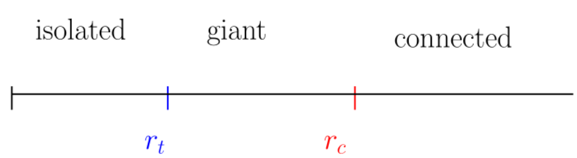


Figura 3: Imatge extreta del document adjunt *docs/random-graphs.pdf*

En la figura 3 s'hi pot observar perfectament els dos punts que impliquen la transició entre *isolated*, *giant* i *connected*. A continuació ho expliquem detalladament:

- Aïllat ⇒ El graf comença amb zero nodes i, a mesura que anem afegint arestes, comença a crear les primeres components connexes. No obstant això, el graf no és prou gran per a considerar-se gegant.
- $r_t$ ⇒ Moment en el qual una aresta fa que la component connexa més gran del graf esdevingui *gegant*.
- Gegant ⇒ El graf segueix incrementant el seu nombre d'arestes, tenint sempre una component força més gran que la resta que estan formades.
- $r_c$ ⇒ Punt on es compleix la propietat de connectivitat d'un graf i que, per tant, tots els nodes del graf passen a estar connectats per, com a mínim, una aresta.

- Connex $\Rightarrow$ A partir d'aquest punt el graf segueix creixent en nombre d'arestes però es manté connectat. El que sí que pot passar en aquesta etapa és que es reforcin les connexions, és a dir, que augmenti el grau dels nodes (adjacències que té en total amb altres nodes del graf).

Finalment adjuntem un enllaç a un vídeo de la plataforma *YouTube* que reflecteix interactivament i visualment tot el que hem esmentat sobre les propietats de connectivitat dels grafs.

# 3 Resultats experimentals sobre connectivitat

Per realitzar els experiments, executem el programa que hem dissenyat amb l'opció `-c k` que imprimeix per pantalla el nombre de components connexes de $k$ grafs de mida $n$ separats per un espai en blanc. Addicionalment assignem un rang a la propietat $p$ o $r$ per generar una taula. Aquestes comandes es poden executar automàticament cridant `make exp`.

A continuació utilitzem l'entorn de programació R per convertir les dades en valors 1 o 0 segons si hi ha o no més d'una component connexa (si només n'hi ha una el graf és connex). Després de fer la mitjana dels valors de connectivitat dels $k$ grafs generem el gràfic corresponent.

Notem que en les transicions de fase l'eix horitzontal correspon al nombre de nodes i l'eix vertical correspon a la probabilitat de **no** ser connex.

## 3.1 Temps d'execució dels experiments

El hardware utilitzat és un *i7-6500U CPU @ 2.50GHz*.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 1000 | 1.32s |
| 100 | 100 | 1000 | 16s |
| 1000 | 100 | 100 | 59s |

Taula 1: Taula del temps d'execució del model BRG.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 500 | 0.65s |
| 100 | 100 | 1000 | 24s |
| 1000 | 100 | 100 | 59s |

Taula 2: Taula del temps d'execució del model RGG.

## 3.2 Binomial Random Graph

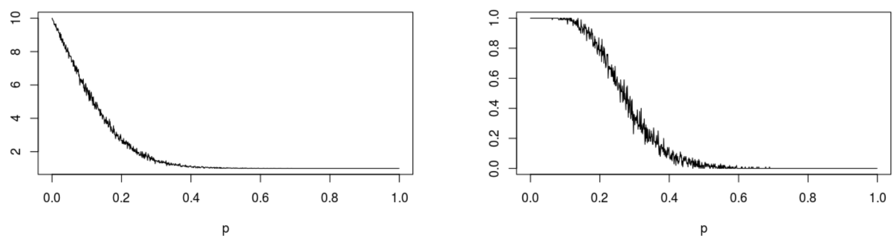


Figura 4: Número esperat de components connexes i transició de fase en un graf BRG amb N=10, K=100 i 1000 iteracions.

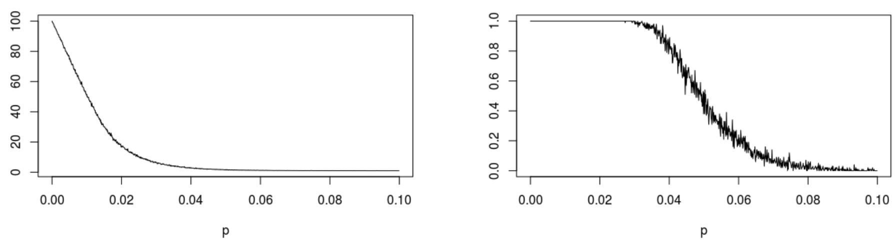


Figura 5: Número esperat de components connexes i transició de fase en un graf BRG amb N=100, K=100 i 1000 iteracions.

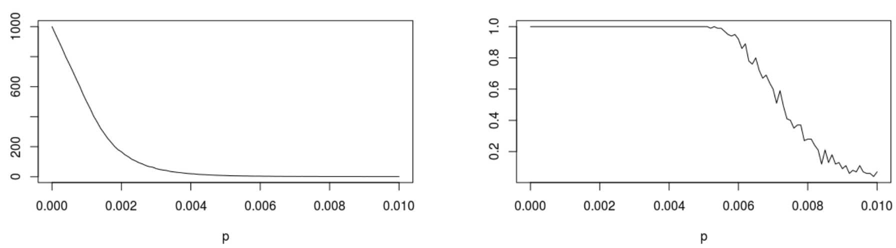


Figura 6: Número esperat de components connexes i transició de fase en un graf BRG amb N=1000, K=100 i 100 iteracions.

## 3.3 Random Geometric Graph

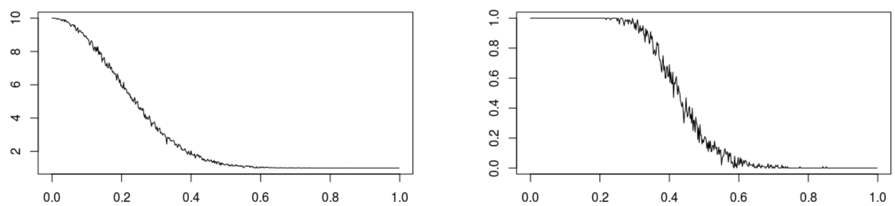


Figura 7: Número esperat de components connexes i transició de fase en un graf RGG amb N=10, K=100 i 500 iteracions.

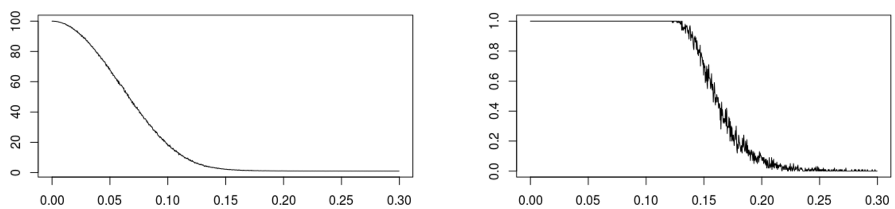


Figura 8: Número esperat de components connexes i transició de fase en un graf RGG amb N=100, K=100 i 1000 iteracions.

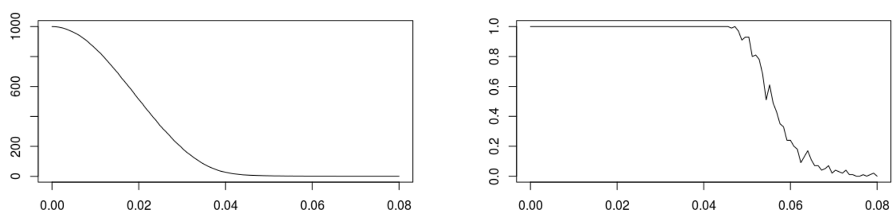


Figura 9: Número esperat de components connexes i transició de fase en un graf RGG amb N=1000, K=100 i 100 iteracions.

# 4 Resultats experimentals sobre la component connexa gegant

Volem estudiar la transició de fase de la component connexa més gran en grafs generats aleatòriament. Aquesta transició de fase permet visualitzar gràficament cóm al augmentar la probabilitat de generar arestes en els grafs hi ha un punt en el qual, sobtadament, es dispara la mida de la major component connexa, donant com a resultat el pas d'una de molt petita a una de molt gran (component connexa gegant) en comparació amb l'anterior. En la secció 2.2 s'explica amb més detall aquesta propietat.

Per a realitzar aquests experiments executem el mateix programa que hem usat en la secció 3 però, aquest cop, ho farem amb l'opció `-s k` que imprimeix per pantalla la mida de la component connexa més gran de $k$ grafs de mida $n$ separats per un espai en blanc. Addicionalment assignem un rang a la propietat $p$ o $r$ per generar una taula. Aquestes comandes també es poden executar automàticament cridant `make exp`.

Seguidament utilitzem l'entorn de programació R amb les comandes pertinents segons s'indica en el README del projecte amb la finalitat de generar les gràfiques que veurem seguidament.

Notem que en les transicions de fase l'eix horitzontal correspon a la probabilitat i l'eix vertical correspon a la mida de la major component connexa del graf.

### 4.1 Temps d'execució dels experiments

El hardware utilitzat és un *i7-6700HQ CPU @ 3.50GHz*.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 1000 | 1.74s |
| 100 | 100 | 1000 | 20.34s |
| 1000 | 100 | 100 | 48.64s |

Taula 3: Taula del temps d'execució del model BRG.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 500 | 0.88s |
| 100 | 100 | 1000 | 27.59s |
| 1000 | 100 | 100 | 68.37s |

Taula 4: Taula del temps d'execució del model RGG.

### 4.2 Binomial Geometric Graph

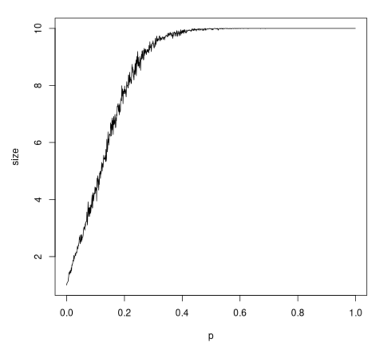


Figura 10: Transició de fase respecte la mida de la component connexa més gran amb N=10, K=100 i 1000 iteracions.

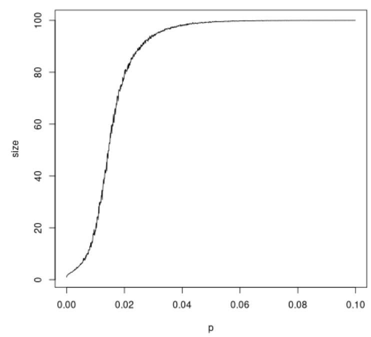


Figura 11: Transició de fase respecte la mida de la component connexa més gran amb N=100, K=100 i 1000 iteracions.

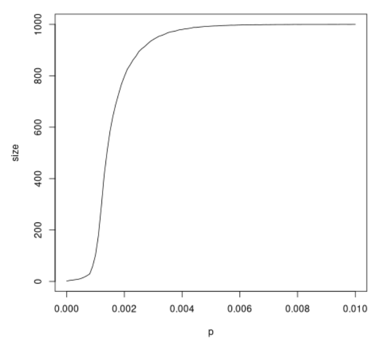


Figura 12: Transició de fase respecte la mida de la component connexa més gran amb N=1000, K=100 i 100 iteracions.

## 4.3 Random Geometric Graph

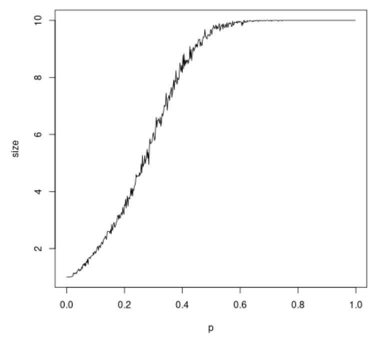


Figura 13: Transició de fase respecte la mida de la component connexa més gran amb N=10, K=100 i 1000 iteracions.

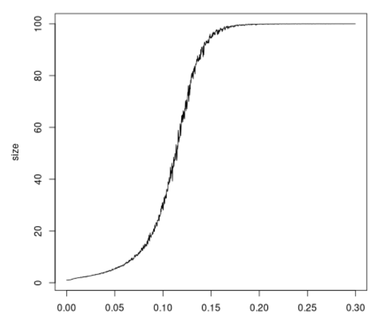


Figura 14: Transició de fase respecte la mida de la component connexa més gran amb N=100, K=100 i 1000 iteracions.

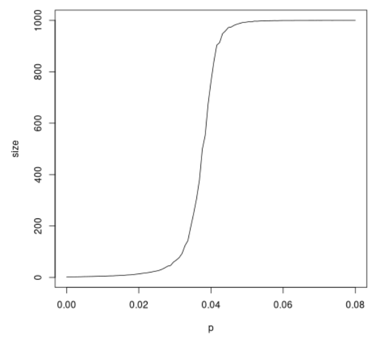


Figura 15: Transició de fase respecte la mida de la component connexa més gran amb N=1000, K=100 i 100 iteracions.

# 5 Altres propietats

A continuació, veurem algunes de les propietats addicionals que hem pogut observar mitjançant l'estudi de les dades obtingudes en les diferents proves:

## 5.1 Completesa

Tenint en compte que un graf complet és aquell on tots els seus vèrtexs estan connectats entre ells per arestes (és a dir, el grau de tots els vèrtexs és de $n - 1$, on $n$ és el nombre total de vèrtexs del graf), i que aquest sempre tindrà una única component connexa gegant, podem observar que per a BRG no hi ha transició de fase com les que hem vist, sinó que té un creixement quadràtic a l'extrem, just abans d'arribar a una probabilitat d'1. Aquest comportament és lògic, ja que només pot ser complex si totes les comparacions que es fan resulten en una aresta, i com més s'incrementi la $n$ més sobtat serà el canvi (ja que hauràs de tenir "sort"més vegades seguides).

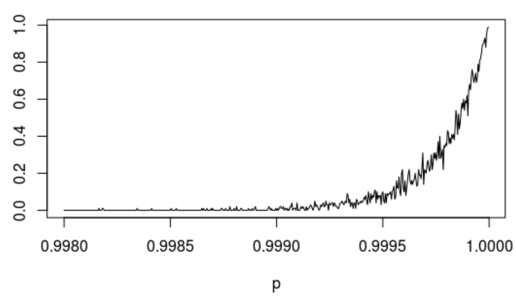


Figura 16: Completesa d'un graf BRG, en interval [0.998-1.000] i N=100.

Per al cas de RGG el comportament és similar, però s'ha d'afegir un detall. En el model RGG és possible tenir dos nodes tal que la seva distancia és superior a 1 si, com a mínim un d'ells es troba fora del cercle inscrit en el quadrat que delimita el pla de 1x1. Experimentalment doncs, veiem que fins i tot quant el radi és 1 la probabilitat de ser complet no és 1. A mesura que augmentem $n$, aquesta probabilitat disminueix.

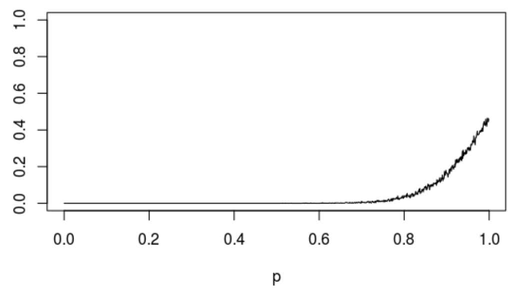


Figura 17: Completesa d'un graf RGG, amb N=10.

Durant les proves experimentals de completesa, s'ha utilitzat el següent hardware: *i7-6700HQ CPU @ 3.50GHz*.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 1000 | 1.40s |
| 100 | 100 | 1000 | 17.73s |
| 1000 | 100 | 100 | 45.94s |

Taula 5: Taula del temps d'execució del model BRG.

| N | K | Iteracions | Temps |
|---|---|---|---|
| 10 | 100 | 500 | 0.70s |
| 100 | 100 | 1000 | 23.92s |
| 1000 | 100 | 100 | 65.47s |

Taula 6: Taula del temps d'execució del model RGG.

# 6 Conclusions

Hem trobat la transició de fase esperada. Ens hem adonat de la complexitat que comporta (nombre de mostres, etc.) analitzar la transició de fase d'una propietat, el nombre de components connexes, que aparentment té una distribució senzilla (lineal en el cas de BRG fins al punt on apareix la transició de fase).

A l'hora d'analitzar la transició de fase de la component connexa gegant en un graf BRG, al principi no l'hem vist perquè aquesta no es fa aparent fins a valors grans de $n$.

No hem explorat la idea d'utilitzar un RNG que fes servir una distribució normal (o d'altres) en comptes d'una distribució uniforme. Intuïm que generaria resultats substancialment diferents en tots dos models, però hem descartat la idea perquè se sortia de l'àmbit del projecte.

Tampoc hem explorat la propietat de veure si ocorre una transició de fase respecte a la clique més gran. L'hem descartat quan hem vist que el problema de trobar la clique més gran és NP-Complet.

## 6.1 El que hem aprés

Hem après parts de C++ i R que desconeixíem, així com a escriure documents mitjançant LaTeX. Altres eines que hem aprés a fer servir inclouen: Overleaf, RStudio Cloud, WSL, Make, Bash, GitLab, GraphViz i DocTest.

També hem après com generar grafs aleatoris i com trobar les components connexes d'un graf en temps lineal mitjançant DFS. Així com implementar mitjançant llistes d'adjacència un graf no dirigit a partir de la especificació de llibre CLRS (*Introduction to algorithms*). La implementació mitjançant una matriu d'adjacència es va descartar des del primer moment.

# 7 Bibliografia

https://en.cppreference.com/w/
http://www.cplusplus.com/reference/
https://en.wikipedia.org/wiki/Erdős–Rényi_model
https://en.wikipedia.org/wiki/Random_geometric_graph
https://stackoverflow.com/
https://es.overleaf.com/learn
https://www.rdocumentation.org/
Cormen et al, Introduction to Algorithms 3rd Edition
Josep Díaz, Random Geometric Graphs (filepath: *docs/random-graphs.pdf*)