

(aquesta no és la portada real)

Modelització i simulació de processos dinàmics en xarxes
complexes adaptatives

Arnau Duran

Enginyeria Industrial,

Escola Politècnica Superior,

Universitat de Girona

`arnauduran@gmail.com`

Octubre 2013 - Maig 2014

DOCUMENT 1: MEMÒRIA

Índex

Document 1	Memòria	1
	Índex	2
	Índex de figures	4
	Índex de taules	7
1	Introducció	8
1.1	Antecedents	8
1.2	Objecte	8
1.3	Especificacions i abast	9
2	Conceptes bàsics i topologia de les xarxes complexes	10
2.1	Conceptes bàsics de les xarxes complexes	10
2.1.1	Quantitats de mesura	11
2.1.2	Robustesa	13
2.1.3	Assortativitat	13
2.2	Topologia de xarxes	14
2.2.1	Xarxa regular	14
2.2.2	Xarxa aleatòria (xarxa d'Erdős-Rényi)	15
2.2.3	Xarxa de petit món	15
2.2.4	Xarxa exponencial	15
2.2.5	Xarxa invariant d'escala	15
3	Dinàmica de les xarxes complexes	17
3.1	Model epidèmic (SIS)	17
3.2	Models de propagació d'opinió	19
3.2.1	Descripció dels tres models	19
3.2.2	Comparació de les dinàmiques dels tres models	20
3.2.3	Variant amb reconexió	22

3.3	Algorisme de Gillespie	22
3.3.1	Descripció de l'algorisme	22
4	Programació del simulador en C++	25
4.1	Blocs del programa simulador del model epidèmic	25
4.1.1	Primer bloc: determinació dels paràmetres	25
4.1.2	Segon bloc: generació de la xarxa	26
4.1.3	Tercer bloc: algorisme de Gillespie múltiple	28
4.1.4	Quart bloc: impressió de les dades en codi llegible pel MATLAB	29
4.2	Modificació del programa per permetre reconexió (sistema de dinàmica d'enllaços)	29
4.2.1	Modificació del programa per permetre infeccions localitzades	31
4.3	Altres consideracions	31
5	Simulacions i anàlisi de resultats	32
5.1	Simulació i estudi del model epidèmic	32
5.2	Simulació i estudi de la dinàmica d'enllaços sense reconexió	38
5.3	Simulació i estudi de la dinàmica d'enllaços amb reconexió	43
5.3.1	Xarxa regular	44
5.3.2	Xarxa invariant d'escala	48
5.3.3	Xarxa invariant d'escala amb distribució inicial intencionada	52
5.4	Perspectives de futur	57
6	Resum de l'estudi econòmic de costos	58
7	Conclusions	59
8	Relació de documents	61
9	Agraïments	62
10	Bibliografia	63

Índex de figures

2.1	Exemple de distribució de graus que segueix una llei potencial.	16
2.2	Exemple d'un <i>hub</i> , és a dir, d'un node amb moltes connexions.	16
3.1	Il·lustració dels tres possibles models de regles d'actualització.[Sood et al., 2008] .	20
4.1	Exemple del sistema de llistes per emmagatzemar una xarxa. La columna <i>Posició llista</i> és la posició de cada element de la llista; cada element és un node. Dins de cada element es troba la seva llista de veïns.	28
5.1	Simulació del model epidèmic per una xarxa regular de grau 10 i per dues taxes d'infecció diferents	35
5.2	Simulació del model epidèmic per una xarxa d'Erdős-Rényi amb $p = 0,02$ i per dues taxes d'infecció diferents	36
5.3	Simulació del model epidèmic per una xarxa invariant d'escala amb $\gamma = 3$ i $m = 5$ i per dues taxes d'infecció diferents	37
5.4	Representació de la probabilitat d'èxit d'A (ε_A) en funció percentatge inicial (ρ) per una xarxa regular amb dinàmica d'enllaços sense reconexió de les diferents simulacions de la Taula 5.2a juntament amb els valors teòrics donats per l'Equació 3.5.	39
5.5	Representació de la probabilitat d'èxit d'A (ε_A) en funció percentatge inicial (ρ) per una xarxa invariant d'escala amb dinàmica d'enllaços sense reconexió de les diferents simulacions de la Taula 5.2b juntament amb els valors teòrics donats per l'Equació 3.5.	40
5.6	Distribució de Poisson de 3 graus de llibertat amb el valor de l'estadístic de contrast t obtingut mitjançant l'Equació 5.5 per cada tipus de xarxa, juntament amb el corresponent p-valor.	43
5.7	Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10%. En vermell es mostra la mitjana de les execucions. . .	45

5.8 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30%. En vermell es mostra la mitjana de les execucions. . . 46

5.9 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 50%. En vermell es mostra la mitjana de les execucions. . . 47

5.10 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10%. En vermell es mostra la mitjana de les execucions. 49

5.11 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30%. En vermell es mostra la mitjana de les execucions. 50

5.12 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 50%. En vermell es mostra la mitjana de les execucions. 51

5.13 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10% (distribució inicial intencionada al 10% dels nodes de més grau). En vermell es mostra la mitjana de les execucions. 54

5.14 Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30% (distribució inicial intencionada al 10% dels nodes de més grau). En vermell es mostra la mitjana de les execucions. 55

5.15 Diagrama de caixa en cada interval de $t = 100$ de la mitjana de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10%. (a) Distribució inicial no intencionada (corresponent a la simulació de la Figura 5.10). (b) Distribució inicial intencionada (corresponent a la simulació de la Figura 5.13). 56

- 5.16 Diagrama de caixa en cada interval de $t = 100$ de la mitjana de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0, 1$ i amb un percentatge inicial de nodes tipus A del 30%. (a) Distribució inicial no intencionada (corresponent a la simulació de la Figura 5.11). (b) Distribució inicial intencionada (corresponent a la simulació de la Figura 5.14). 56

Índex de taules

5.1	Resum de les 12 simulacions dutes a terme per cadascun dels tres tipus de xarxes, amb els paràmetres β , ρ i t	33
5.2	Probabilitat d'èxit (ε_A) segons diferent percentatge inicial (ρ) per dos tipus de xarxes diferents de 1000 nodes (50 execucions) amb dinàmica d'enllaços sense reconexió	38
5.3	Càlcul dels paràmetres estadístics necessaris per calcular l'estadístic de contrast t per cada tipus de xarxa.	42

1. INTRODUCCIÓ

1.1 Antecedents

Leonhard Euler va ser pioner en l'estudi de la teoria de grafs: publicà el primer article relacionat amb aquesta àrea d'estudi de les matemàtiques el 1736. Des de llavors, múltiples autors com Kirchoff, Seymour, etc. continuaren amb l'estudi de la teoria i topologia de grafs. La teoria de xarxes, part de la teoria de grafs, també ha estat estudiada a bastament; de fet, la solució euleriana dels *Set ponts de Königsberg* és considerada la primera demostració del camp de la teoria de xarxes.

D'altra banda, la dinàmica de xarxes fou popularitzada per Dan Gillespie en el seu article de 1977 Gillespie [1977]. L'algorisme proposat per Gillespie permet la simulació discreta i estocàstica d'un sistema de partícules. Aquest algorisme és el que s'utilitzarà en el present treball per dur a terme les simulacions de les xarxes. El camp de l'anàlisi de la dinàmica de xarxes, de fet, és un camp emergent en l'actualitat; comprèn tant l'anàlisi estadística com la utilització de simulacions per solucionar problemes de la mateixa dinàmica.

Les xarxes complexes (xarxes de característiques complexes, sovint xarxes reals) també són objecte d'estudi de l'actualitat, sobretot a causa de l'aparició de les xarxes socials. S'han convertit en un paradigma per l'estudi de processos dinàmics en sistemes formats per molts components que interactuen entre si de manera molt homogènia. La topologia, les relacions entre nodes i la mateixa dinàmica de les xarxes complexes són uns estudis avui en dia emergents.

1.2 Objecte

L'objectiu del present treball de final de carrera és triple:

1. Estudiar i entendre els conceptes bàsics i la topologia de les xarxes complexes, així com diferents tipus de dinàmiques de processos sobre elles.
2. Programar un simulador estocàstic en llenguatge C++ capaç de generar trajectòries mitjançant l'algorisme de Gillespie tant pel model epidèmic com pel model de dinàmica d'enllaços amb reconexió.
3. Utilitzar el simulador tant per estudiar casos que ja han estat tractats en la literatura com casos nous que no han estat tractats.

1.3 Especificacions i abast

El treball s'inclou dins la línia de recerca sobre xarxes complexes que desenvolupen professors del grup "Equacions diferencials, modelització i aplicacions" del departament d'Informàtica, Matemàtica Aplicada i Estadística de la Universitat de Girona.

El treball es divideix en les següents quatre etapes, dutes a terme per ordre cronològic:

1. Estudi, comprensió i anàlisi dels conceptes bàsics de xarxes complexes, quantitats de mesura i topologia de xarxes.
2. Estudi, comprensió i anàlisi de la dinàmica de les xarxes complexes, concretament de tres temes: el model epidèmic (SIS), els models de propagació d'opinió i l'algorisme de Gillespie.
3. Programació del simulador en C++ mitjançant l'algorisme de Gillespie, tant pel model epidèmic com el model amb reconexió (adaptatiu).
4. Anàlisi dels resultats obtinguts amb les simulacions per diferents casos.
 - 4.1 Model epidèmic: es comprovarà que les trajectòries simulades coincideixen amb la solució analítica obtinguda a partir de la integració numèrica de les equacions diferencials del model.
 - 4.2 Model de dinàmica d'enllaços sense reconexió: es comprovarà que la probabilitat d'èxit predita per la literatura existent sobre el tema coincideix amb l'obtinguda per mitjà del simulador.
 - 4.3 Model de dinàmica d'enllaços amb reconexió: els resultats d'aquest apartat són novedosos, ja que no existeix literatura sobre aquest cas concret (xarxes complexes adaptatives). Es farà especial incís en les xarxes de topologia assimilable a xarxes reals com, per exemple, les xarxes socials.

2. CONCEPTES BÀSICS I TOPOLOGIA DE LES XARXES COMPLEXES

En matemàtiques i, més concretament, en teoria de grafs, un graf és una representació abstracta d'un conjunt d'elements interconnectats entre alguns d'ells dos a dos. Els elements s'anomenen *vèrtexs* o *nodes*, i els enllaços que els uneixen, *arestes*. Segons les seves característiques, hi ha diferents tipus de grafs, com poden ser: multigrafs (amb arestes repetides), grafs dirigits (amb arestes que només van en un sentit), grafs ponderats (les arestes dels quals tenen associat un pes o cost), etc. Aquest treball se centrarà en l'estudi de grafs (o xarxes) sense arestes repetides i amb arestes no ponderades ni dirigitades, i sense llaços (arestes que van d'un vèrtex a si mateix).

El terme "xarxa" –en el context d'aquest treball– comporta una mica de confusió. L'opinió més generalitzada és que el graf és una representació matemàtica d'una xarxa, la qual s'entén com un concepte més físic; més o menys, doncs, això equival a dir que graf i xarxa són el mateix. En aquest treball s'usaran les dues paraules indistintament.

El terme "xarxa complexa" s'utilitza per designar una xarxa que presenta característiques topològiques no trivials o, dit d'una altra manera, característiques que no presenten xarxes simples com retícules o grafs aleatoris però que es poden observar en xarxes reals. Més concretament, el terme designa una xarxa que presenta fenòmens que són producte de la interacció espontània entre els seus elements. A més a més, una altra característica que posseeix és que si s'estudia una part aïllada no es pot obtenir un coneixement general de tota la xarxa, ja que la seva composició és resultat d'una dinàmica participada per molts elements. Finalment, el darrer tret que presenten la majoria de xarxes complexes és la presència de complicacions a qualsevol nivell [Barrat et al., 2008]. Són exemples de xarxes complexes les xarxes socials, les xarxes de regulació gènica, les xarxes de telecomunicacions, etc.

És important distingir entre dos termes que podrien resultar confusos: *xarxa complexa* (tal com s'ha presentat al paràgraf anterior) i una *xarxa complicada* (entesa com una xarxa, per exemple, amb un gran nombre de components).[Barrat et al., 2008]

2.1 Conceptes bàsics de les xarxes complexes

Entrant ja en l'estudi de les xarxes complexes, convé descriure alguns dels conceptes bàsics que serveixen per quantificar i mesurar-les. Encara que se n'han proposat moltes, les tres quantitats de mesura que tenen el paper més important en l'estudi recent de les xarxes complexes són la

longitud mitjana de camí, el coeficient de clusterització i la distribució de graus.[Wang, 2002]

2.1.1 Quantitats de mesura

Longitud mitjana de camí

Per començar, es defineix la distància d_{ij} entre dos nodes de la xarxa com la distància més curta que els separa (comptada amb arestes). D'altra banda, es defineix el diàmetre D d'una xarxa com la distància màxima que es pot trobar entre dos dels seus nodes qualssevol.

La longitud mitjana de camí (en anglès *average path length*), denotada per L , es defineix com la mitjana de les distàncies d_{ij} de totes les parelles de nodes i, j de la xarxa.

S'observa que $L/D < 1$, excepte quan es tracta d'un graf complet, que llavors és $L/D = 1$. En un estudi dut a terme entre 1997 i 1999, Pastor-Satorras et al. [2001] trobà que la longitud mitjana de camí d'Internet era de $L = 4$, quan en el graf aleatori corresponent li correspondria una $L = 10$.

Coeficient de clusterització

El coeficient de clusterització (en anglès *clustering coefficient*), denotat per C , és una mesura que explica com de relacionats estan els veïns d'un cert node; per exemple, en una xarxa social, quants dels amics d'una persona són també amics entre si. Matemàticament, el coeficient de clusterització d'una xarxa es calcula fent la mitjana per tots els i nodes de cadascun dels seus C_i corresponents calculats de la següent manera:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2.1)$$

On k_i és el nombre de veïns del node i , i E_i és el nombre d'arestes entre els k_i veïns.

S'observa que $C_i \leq 1$, i $C_i = 1$ si es tracta d'un graf complet. A més, s'ha demostrat que la majoria de xarxes reals de gran escala tenen tendència a la clusterització. Pastor-Satorras et al. [2001] determinà el C_i d'Internet en un valor d'entre 0,18 i 0,3. En una xarxa completament aleatòria, $C_i \simeq 0,001$.

Distribució de graus

El grau d'un vèrtex i d'un graf, denotat per k_i , és el nombre total d'arestes que incideixen sobre seu; dit d'una altra manera, és el nombre de connexions que té. La mitjana de tots els k_i del graf, anomenada grau mitjà, es denota per $\langle k \rangle$.

La distribució de graus (en anglès *degree distribution*), denotada per la funció $P(k)$, representa la possibilitat que un vèrtex seleccionat aleatòriament de la xarxa tingui grau k . En el cas d'una xarxa totalment aleatòria, la probabilitat $P(k)$ que k arestes incideixin sobre un node segueix la distribució de Poisson:

$$P(k) \simeq e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad (2.2)$$

A causa de les característiques de la distribució de Poisson, per $k \gg \langle k \rangle$ la probabilitat $P(k)$ de trobar un node que tingui k arestes incidents esdevé negligiblement petita.

S'ha descrit una altra distribució que explica millor el comportament de xarxes reals de gran escala: es tracta de la distribució potencial, també anomenada "distribució invariant d'escala". Aquesta distribució decreix més gradualment que l'Equació 2.2, la qual cosa fa que la possibilitat de trobar nodes amb graus elevats sigui més alta:

$$P(k) \sim k^{-\gamma} \quad (2.3)$$

El valor de la γ depèn de la xarxa. Per exemple, l'estudi dut a terme per Faloutsos et al. [1999] sobre la topologia d'Internet entre els anys 1997 i 1998 el determinà, en aquest cas concret, en un valor d'entre 2,2 i 2,4.

D'altra banda, una igualtat fonamental de tota xarxa és la següent, on L és el nombre d'arestes i N el nombre de vèrtexs:

$$N \cdot \langle k \rangle = 2 \cdot L \quad (2.4)$$

Una altra distribució important, lleugerament diferent que la $P(k)$, és la $Q(k)$, la qual depèn de la primera, i que representa la probabilitat que una arista dirigida "acabi" o "comenci" (es tria un dels dos extrems a l'atzar) en un node de grau k .

$$Q(k) = \frac{k \cdot P(k)}{\langle k \rangle} \quad (2.5)$$

Cal observar la diferència entre $P(k)$ i $Q(k)$: en el primer cas, es tria un node a l'atzar i se n'observa el grau; en el segon, s'escull una aresta a l'atzar i es mira el grau del node on comença o on acaba.

Les distribucions $P(k)$ i $Q(k)$ són característiques estadístiques bàsiques d'una xarxa, molt útils per xarxes molt grans però no tant per xarxes reals, en les quals no determinen prou bé l'estructura de la xarxa.

2.1.2 Robustesa

La robustesa es defineix com la capacitat que té una xarxa per resistir-se a l'eliminació d'arestes o vèrtexs. Per exemple, una xarxa d'ordinadors robusta respecte a eliminacions aleatòries és capaç de continuar amb el seu funcionament normal encara que succeeixin fallades aleatòries. Una xarxa també pot ser robusta respecte a eliminacions dirigides (o intencionades): per exemple, un epidemiòleg que vol eradicar els vèrtexs infectats per una malaltia.

La robustesa d'una xarxa depèn en gran mesura de la seva estructura i, especialment, de la seva distribució de graus. Per exemple, l'eliminació de vèrtexs d'una xarxa amb una distribució de graus de Poisson farà que aquesta xarxa se separi en grups petits força ràpidament. [Jamakovic and Uhlig, 2007]

Actualment s'està fent recerca en el sentit de determinar una manera de quantificar la robustesa d'una xarxa.

2.1.3 Assortativitat

L'assortativitat, en general, és un concepte que fa referència a la preferència dels nodes de la xarxa per estar enllaçats amb altres nodes que, en algun aspecte, són similars a ells. Normalment, aquest aspecte es tracta del grau del node. És a dir, en una xarxa assortativa respecte als graus dels nodes, els nodes de grau gran estaran més relacionats amb altres nodes de grau gran, i els de grau petit amb altres de grau petit.

Existeix un coeficient d'assortativitat R , anomenat coeficient de Newman, que mesura la correlació entre nodes de grau similar. De fet, es tracta del coeficient de correlació de Pearson entre parelles de nodes enllaçats. [Newman, 2002] Aquest coeficient es troba en l'interval $[-1, 1]$. Un coeficient de -1 indica que es tracta d'una xarxa completament disassortativa, i un coeficient d' 1 indica que és completament assortativa.

2.2 Topologia de xarxes

La topologia consisteix en l'estudi de l'estructura de les xarxes, és a dir, la disposició dels seus nodes i com els interrelacionen les arestes. A continuació es presenten alguns dels models més típics de xarxes, alguns dels quals seran utilitzats en la modelització duta a terme a la segona part del present treball. Els algorismes per la seva generació es presentaran a l'apartat corresponent (Segon bloc: generació de la xarxa).

És important conèixer la topologia d'una xarxa ja que sol tenir un paper crucial a l'hora de determinar-ne les característiques dinàmiques. Per entendre bé l'estructura, el primer pas sol consistir en la mesura de les propietats descrites a la secció Quantitats de mesura: la longitud mitjana de camí, el coeficient de clusterització i la distribució de graus.

Cal tenir present que en tot el present treball es contemplen les següents condicions pel que fa a la topologia de totes les xarxes tractades:

- La xarxa no té llaços ("autoarestes")
- Les arestes són no dirigides i no ponderades
- La xarxa no té arestes dobles (repetides)

2.2.1 Xarxa regular

Una xarxa regular és aquella que tots els seus nodes tenen el mateix grau (mateix nombre de veïns).

La xarxa regular de grau màxim s'anomena xarxa regular completa (o, simplement, xarxa completa). En aquest tipus de xarxa, tots els nodes tenen el grau màxim (és a dir, $k_i = N - 1$, on N és el nombre de nodes de la xarxa) i el nombre d'arestes és $N(N - 1)/2$. A més, la seva longitud mitjana de camí és $L = 1$ i el seu coeficient de clusterització és $C_i = 1$. Tanmateix, les xarxes completes no es poden assimilar a cap xarxa del món real.

Un tipus especial de xarxa regular és la xarxa reticular (en anglès *lattice network*), en la qual els nodes es disposen en forma de retícula i es connecten amb els seus veïns més pròxims. Per tant, els nodes que no es troben a la part exterior de la retícula tenen grau 4, mentre que els nodes de la part exterior tenen grau 3 o 2.

2.2.2 Xarxa aleatòria (xarxa d'Erdős-Rényi)

Al pol oposat de les xarxes regulars es troben els grafs totalment aleatoris, estudiats per primer cop a Erdős and Rényi [1960]. El graf d'Erdős-Rényi (*ER graph*), introduït el 1960, ha estat el model de xarxa aleatòria més estudiat. El principi bàsic que governa aquest tipus de xarxa és que cada aresta entre dos dels N nodes de la xarxa existeix o no existeix segons una mateixa probabilitat p ; resulta, doncs, que una xarxa aleatòria d' N nodes té $pN(N-1)/2$ arestes. Erdős i Rényi també demostraren que el grau mitjà d'una xarxa aleatòria és $\langle k \rangle = pN$, que la longitud mitjana de camí és $L \sim \frac{\ln N}{\ln \langle k \rangle}$ i que les xarxes aleatòries no mostren clusterització.

Watts and Strogatz [1998] conceberen un mètode per transformar una xarxa regular a xarxa aleatòria mitjançant un algorisme amb reconexions.

2.2.3 Xarxa de petit món

L'efecte del petit món fou introduït per Watts and Strogatz [1998] Segons ells mateixos, un exemple quotidià seria quan una persona es troba amb un desconegut i junts descobreixen que tenen una amigat en comú. Una xarxa que presenta aquesta característica és aquella en la qual la majoria de nodes no són veïns entre ells, però es pot arribar a la majoria de nodes a partir de tots els altres amb només un petit nombre de passes.

S'ha demostrat que aquest efecte està present en moltes xarxes del món real.[Wang, 2002]

2.2.4 Xarxa exponencial

Una xarxa exponencial és aquella que la seva distribució de graus té un pic en un cert valor i decreix exponencialment. Una xarxa exponencial és homogènia: tots els nodes tenen més o menys el mateix nombre de connexions.[Wang, 2002]

2.2.5 Xarxa invariant d'escala

Una xarxa invariant d'escala, o xarxa lliure d'escala (en anglès *scale-free network*) és aquella que la seva distribució de graus segueix una llei potencial com la descrita a l'Equació 2.3, normalment amb una $\gamma > 2$ (la Figura 2.1 mostra un exemple de distribució de graus que segueix una llei potencial). A més, se sol establir un grau mínim dels nodes, de tal manera que el grau obtingut per cada node ha de ser forçosament igual o superior a aquest valor.

Una xarxa invariant d'escala és no homogènia, en el sentit que la majoria de nodes tenen

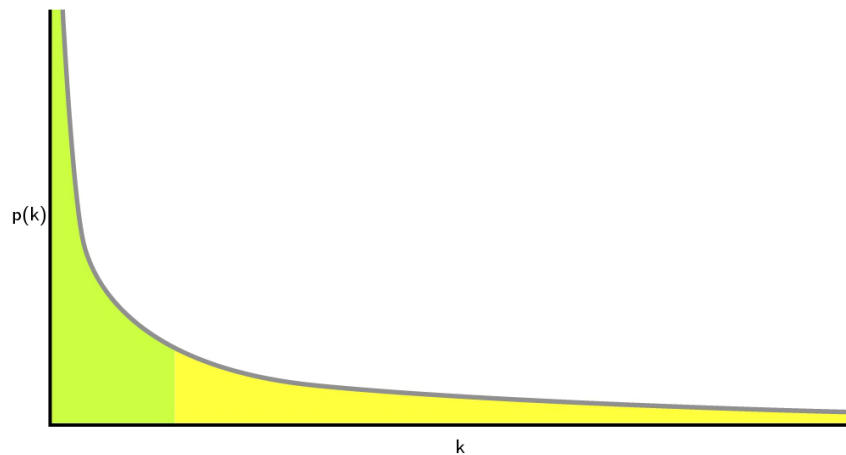


Figura 2.1: Exemple de distribució de graus que segueix una llei potencial.

Les zones de diferent color representen la regla 80-20. La zona de l'esquerra és la que domina, i la de la dreta s'anomena "cua llarga". És d'aquesta darrera zona d'on provenen els pocs nodes amb moltes connexions.

poques connexions i tan sols uns pocs nodes tenen moltes connexions. No s'ha de confondre amb la xarxa exponencial, que és molt més homogènia. Aquest tipus de xarxa sol tenir nodes amb moltes connexions (de vegades anomenats *hubs*; exemple a la Figura 2.2) els quals, segons el model d'interacció aplicat a la xarxa, poden tenir una gran importància i poden determinar en gran mesura l'evolució i el comportament d'aquesta.

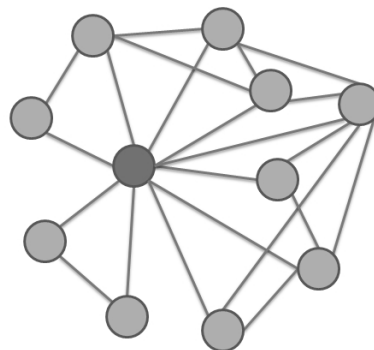


Figura 2.2: Exemple d'un *hub*, és a dir, d'un node amb moltes connexions.

Segons s'ha pogut comprovar amb resultats numèrics en comparació amb les xarxes aleatòries, la longitud mitjana de camí de les xarxes invariants d'escala és menor, però el seu coeficient de clusterització és molt més gran.[Wang, 2002] Tanmateix, no existeix cap expressió analítica per trobar cap d'aquests dos valors.

Les xarxes invariants d'escala són un descobriment relativament recent, estudiat per primer cop per Barabási and Albert [1999]. Aquests mateixos observaren que, per exemple, l'Internet segueix la distribució de les xarxes invariants d'escala.

3. DINÀMICA DE LES XARXES COMPLEXES

La dinàmica de xarxes estudia l'evolució de les xarxes al llarg del temps, ja sigui pel que fa a l'estat dels seus nodes o arestes, o a l'addició, eliminació o reconexió de nodes al llarg del temps, fent un canvi a cada interval infinitesimal. En el cas del present treball aquesta dinàmica està governada per l'algorisme estocàstic de Gillespie (vegeu la secció Algorisme de Gillespie).

3.1 Model epidèmic (SIS)

El model epidèmic és un model simplificat que intenta descriure la transmissió d'una malaltia a través d'individus (nodes) d'una xarxa. En el cas del present treball, s'estudiarà i modelitzarà el model epidèmic SIS (*susceptible-infectat* o, sinònimament, *sa-infectat*).

El model epidèmic SIS és simple: contempla dos estats discrets pels individus (nodes) de la xarxa: "sa" (o "susceptible") i "infectat".¹ Cada aresta de la xarxa significa una relació binària, un enllaç a través del qual la infecció pot propagar-se d'un individu a un altre. L'estudi d'aquest tipus de model se centra en dos paràmetres: el nombre d'infectats i la seva localització en la xarxa. Aquest treball se centra en el primer dels dos i, sobretot, en la seva evolució al llarg del temps.

A cada instant de temps, un node sa té una taxa β (*taxa de transmissió*, en anglès *spreading rate*) de tornar-se infectat a causa d'un veí infectat en un interval de temps infinitesimal dt . D'altra banda, cada node infectat té una taxa μ (*taxa de recuperació*, en anglès *recovery rate*) de recuperar-se a cada interval de temps dt , independentment dels seus veïns.

Tal com s'explica amb detall a Barrat et al. [2008], i tal com es pot deduir intuïtivament, cada node sa té una probabilitat d'adquirir una infecció de $\beta k i dt$, on k és el grau del node (nombre de veïns) i i és la fracció de nodes infectats de la xarxa en tant per u. Cal adonar-se en aquest punt de la important dependència de l'evolució del sistema respecte el grau dels nodes; se suposa, per començar, que tots els nodes de la xarxa tenen el mateix grau $k = \langle k \rangle$ (xarxa regular o homogènia). Es pot escriure, doncs, la següent equació diferencial que descriu l'evolució d'un model SIS, de moment sense recuperació:

$$\frac{di(t)}{dt} = \beta \langle k \rangle i (1 - i) \tag{3.1}$$

¹Existeixen models similars que contemplin més estats pels individus, com per exemple el model SIR, que contempla tres estats: "susceptible", "infectat" i "immune". Tanmateix, però, aquests models no són objectiu d'anàlisi en el present treball.

L'equació anterior és intuïtivament fàcil d'entendre: tots els nodes susceptibles de la xarxa ($1 - i$) tenen la probabilitat $\beta k i dt$ comentada anteriorment d'infectar-se (canviant k per $\langle k \rangle$, ja que es consideren tots els graus dels nodes de la xarxa idèntics).

Òbviament, un model d'aquest tipus (que de vegades s'anomena simplement model SI, en contraposició al model SIS) acabarà tard o d'hora amb tota la xarxa infectada; falta tenir-hi en compte la taxa de recuperació μ . L'Equació 3.1 queda modificada doncs com a:

$$\frac{di(t)}{dt} = \beta \langle k \rangle i (1 - i) - \mu i \quad (3.2)$$

Com es pot veure, a l'equació diferencial s'hi ha afegit el terme $-\mu i$ que sostreu a la taxa d'increment d'infectats la taxa de recuperació dels nodes infectats que passen a ser susceptibles.

En el model SIS es pot prendre la següent combinació de paràmetres:

$$\lambda = \langle k \rangle \frac{\beta}{\mu} \quad (3.3)$$

El valor $\lambda = 1$ és una combinació crítica de paràmetres (o condició llindar) del model SIS.² Si $\lambda > 1$, és a dir, si $\beta \langle k \rangle > \mu$, a llarg termini hi haurà una fracció positiva d'infectats estable que serà:

$$i = 1 - \frac{\mu}{\beta \langle k \rangle} \quad (3.4)$$

D'altra banda, si $\lambda < 1$, és a dir, si $\beta \langle k \rangle < \mu$, la infecció s'extingirà a llarg termini i tots els nodes de la xarxa quedaran sans. El cas particular de $\lambda = 1$ a la pràctica no es donarà mai.

A més, falta saber quin serà el comportament teòric de l'epidèmia en xarxes heterogènies (en el sentit d'una gran dispersió pel que fa als graus dels nodes; per exemple, una xarxa invariant d'escala). Dels paràmetres de l'Equació 3.2, l'únic que dóna informació sobre la xarxa (encara que poca) és el grau mitjà $\langle k \rangle$. Resulta, doncs, que per molt diferent que sigui la topologia de la xarxa (xarxa regular, xarxa invariant d'escala, etc.), si dues xarxes tenen el mateix grau mitjà, el seu comportament predit per les equacions diferencials explicades anteriorment serà el mateix. Aquest fet és contraintuïtiu i, de fet, existeixen equacions que prenen més paràmetres i són més acurats; no obstant, a la secció Simulació i estudi del model epidèmic es veurà com aquestes EDOs descriuen prou bé l'evolució del nombre d'infectats de la xarxa al llarg del temps.

²De fet, λ és el nombre mitjà d'infectats produït per un infectat al principi de l'epidèmia.

3.2 Models de propagació d'opinió

Els models de propagació d'opinió són models de processos que tenen lloc sobre xarxes de contacte (per exemple, xarxes socials) en què el mateix procés dinàmic intervé en la creació i/o eliminació d'arestes. Per tant, són processos més complexos que l'exemple del model epidèmic de la secció anterior, en què la xarxa de contactes és fixa i no canvia durant el procés de propagació de la malaltia. En canvi, en els processos que es consideraran a continuació hi ha una doble interacció: l'estructura de la xarxa influeix en l'estat dels nodes (tal com passava en el model SIS) però, a la inversa, l'estat que es troben els nodes a cada instant provoca canvis en la topologia de la xarxa mitjançant trencaments i reconexions d'enllaços. Així doncs, els processos que s'explicaran en aquesta secció provoquen que la xarxa ja no tingui una distribució de graus fixa durant tota l'evolució del procés. Aquesta mena de xarxes que evolucionen al llarg del temps s'anomenen xarxes adaptatives i, en general, és un problema obert calcular la variació de la topologia de la xarxa al llarg del temps (és a dir, trobar la distribució de graus $p_k(t)$ a cada instant de temps t).

Existeixen tres models de propagació d'opinió bàsics: el *model del votant*, el *procés d'invasió* i la *dinàmica d'enllaços*. En aquesta secció es considera una xarxa en la qual els nodes poden ser de tipus A o B (és a dir, que adopten l'opinió A o B).³ L'estat de la xarxa canviarà amb el temps segons les regles d'actualització determinades pel model que segueixi la xarxa. Aquesta actualització de l'estat dels nodes, en tots els casos, depèn dels veïns de cada node i , per tant, depèn en gran importància de la topologia de la xarxa. La Figura 3.1 il·lustra de manera resumida els tres processos descrits.

3.2.1 Descripció dels tres models

El primer model de propagació d'opinió estudiat, el model del votant, defineix una regla d'actualització que consisteix en dues passes:

1. S'escull un node a l'atzar
2. El node adopta el mateix estat que un veí seu a l'atzar

En segon lloc, el procés d'invasió és recíproc respecte el model del votant. En aquest cas, no és el node qui adopta l'estat d'un veí sinó que aquest node "contagia" el seu estat a un veí. Les

³Això es podria assimilar a dir que els nodes estan infectats o sans, però per claredat de nomenclatura s'opta per l'altra opció.

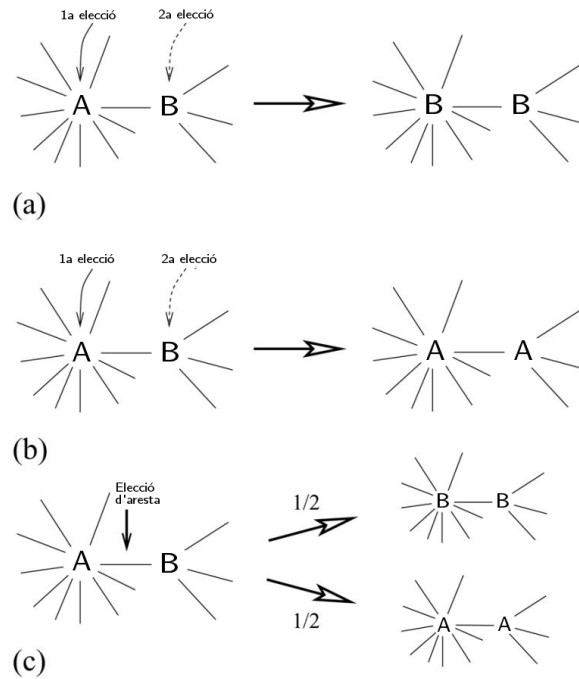


Figura 3.1: Il·lustració dels tres possibles models de regles d'actualització.[Sood et al., 2008]

a) model del votant; b) procés d'invasió; c) dinàmica d'enllaços.

passes són les següents:

1. S'escull un node a l'atzar
2. El node exporta el seu estat a un veí seu a l'atzar

En el cas d'una xarxa homogènia (per exemple, una xarxa de grau regular) és indiferent usar qualsevol dels dos sistemes descrits fins ara (model del votant i procés d'invasió); tanmateix, en xarxes heterogènies, la diferència és substancial.

Finalment, el model de dinàmica d'enllaços és un cas intermedi entre el model del votant i el procés d'invasió. En aquest cas, el procediment és el següent:

1. S'escull un enllaç a l'atzar
2. Un dels nodes de l'enllaç, escollit a l'atzar, adopta l'estat de l'altre node

3.2.2 Comparació de les dinàmiques dels tres models

Les dinàmiques corresponents als tres models de propagació d'opinió són diferents si la xarxa és heterogènia (com, per exemple, una xarxa invariant d'escala). Per analitzar com es comporta una xarxa sota la governació de cadascun dels models, cal definir un parell de conceptes:

- La densitat inicial ρ , nombre de nodes inicials que són del tipus A (distribuïts aleatòriament).
- La probabilitat d'èxit $\varepsilon_A(\rho)$, probabilitat que un sistema finit amb una densitat inicial ρ de nodes A assolixi un consens A . S'entén per *consens* el fet que tots els nodes de la xarxa siguin del mateix tipus. També s'anomena probabilitat de fixació.

Segons Sood et al. [2008], la probabilitat d'èxit ε_A del model de dinàmica d'enllaços (independentment de la topologia de la xarxa) és, simplement:

$$\varepsilon_A(\rho) = \rho \tag{3.5}$$

És a dir, la probabilitat d'èxit d' A correspon a la densitat inicial de nodes A . Si, per exemple, es pensa en A i B com dues opinions diferents, si una xarxa comença amb un 10% de nodes que opinen A , nou de cada deu cops aquesta opinió A quedarà eradicada per la B , i un de cada deu l'opinió A s'imposarà a la B . D'altra banda, és obvi que si la densitat inicial de cadascun dels dos tipus és $\rho = 0,5$ (50%), la probabilitat d'èxit és la mateixa tant per A com per B .

L'Equació 3.5 també és vàlida per al model del votant i el procés d'invasió, sempre i quan es tracti de xarxes regulars.

Convé recordar que, en cada actualització de la xarxa, determinada segons les regles del model escollit, el temps s'incrementa un cert valor, ja que es tracta d'un procés continu en el temps. Aquest valor, en el cas del present treball, ve determinat per l'algorisme de Gillespie. Existeixen estudis que discuteixen el temps necessari per assolir el consens en una xarxa per alguns casos concrets. Per exemple, Sood et al. [2008] proposa la següent equació per determinar el temps de consens en una xarxa regular pel cas del model del votant:

$$T(\rho) = N \left[(1 - \rho) \ln \frac{1}{1 - \rho} + \rho \ln \frac{1}{\rho} \right] \tag{3.6}$$

Un parell de casos notables són els següents: si hi ha inicialment la meitat de nodes de cada tipus ($\rho = 0,5$, llavors $T(0,5) = N \ln 2$) i si hi ha inicialment un node d'un tipus i la resta de l'altre ($\rho = 1/N$, llavors $T(1/N) \approx 1/N$).

No s'ha trobat cap equació similar a l'Equació 3.6 pels altres dos models de propagació d'opinió. Tanmateix, la citada equació pot ser útil per obtenir una estimació *a priori* del temps necessari per assolir un consens en qualsevol tipus de xarxa.

3.2.3 Variant amb reconexió

Una variant que es pot implementar és que un node, en comptes de canviar d'estat, abandoni la seva relació amb l'altre node i es reconnecti a un altre node del seu mateix tipus segons una probabilitat p . Les xarxes que segueixen una llei d'aquest tipus s'anomenen xarxes adaptatives i no han estat gaire tractades en la literatura de les xarxes complexes. En el present treball s'estudiaran a la secció Simulació i estudi de la dinàmica d'enllaços amb reconexió.

3.3 Algorisme de Gillespie

El model determinista "tradicional" presentat a la secció Model epidèmic (SIS), modelitzat per una equació diferencial, no és prou útil ja que li manca exactitud i és difícil calcular el temps transcorregut, sobretot quan es tracta de xarxes amb molts nodes.

L'algorisme de Gillespie és una variant matemàtica del mètode dinàmic de Monte-Carlo que permet generar una possible solució d'una equació estocàstica. Fou popularitzat per Dan Gillespie l'any 1977 en el seu article Gillespie [1977], en el qual utilitza l'algorisme per simular reaccions moleculars en un sistema químic de manera eficient per mitjà d'una potència de càlcul limitada.

L'algorisme de Gillespie permet una simulació discreta i estocàstica d'un sistema format per un cert nombre de partícules de tipus diferents (molècules, si es fa l'analogia amb el sistema químic proposat per Gillespie). La clau de l'algorisme és que fa l'assumpció que la gran majoria de reaccions que tenen lloc en un cert volum de partícules no tenen un efecte net sobre la població de cada tipus de partícules i, per tant, es poden ignorar; l'algorisme es pot centrar, doncs, només en les reaccions que afecten les poblacions dels tipus de partícules. Això simplifica el problema i redueix enormement el temps de càlcul. A més, també s'assumeix que totes les reaccions durant la simulació han de comprendre com a màxim dos nodes, ja que les reaccions que en comprenen tres són extremadament rares. El model, doncs, és una seqüència de reaccions binàries que simplifiquen la situació real.

3.3.1 Descripció de l'algorisme

A continuació es descriu breument l'algorisme de Gillespie en base a l'article Gillespie [2007].⁴ No es pretén fer una justificació matemàtica de l'algorisme en profunditat, ja que això no està

⁴En l'article es parla de "molècules", "espècies moleculars" i "reaccions químiques". Per coherència amb aquest treball, aquí es parlarà d'"individus", "tipus d'individus" i "reaccions entre individus".

comprès dins de l'abast d'aquest treball, però sí que se'n donarà una visió general.

Es considera un sistema amb N tipus d'individus $\{S_1, \dots, S_N\}$ (en el cas del model epidèmic SIS, dos tipus: "sa" i "infectat"). $X_i(t)$ és el nombre d'individus del tipus i , i $X(t) = (X_1(t), \dots, X_n(t))$ és el vector d'estat que explica la situació del sistema en un instant de temps t , és a dir, com estan repartits tots els individus pel que fa als tipus (el nombre d'individus del sistema complet és constant). L'objectiu de la simulació mitjanant l'algorisme de Gillespie és estimar el vector d'estat $X(t)$ a partir d'un vector d'estat inicial $X(t_0) = x_0$ en un temps inicial t_0 .

Els individus del sistema estan relacionats entre si per M reaccions $\{R_1, \dots, R_M\}$. Cada reacció R_j relaciona dos individus (és una relació binària, com s'ha explicat anteriorment) i està caracteritzada matemàticament per dues quantitats:

- Vector de canvi d'estat $v_j = \{v_{1j}, \dots, v_{Nj}\}$, on v_{ij} és el canvi en la població molecular de S_i causat per la reacció R_j , de tal manera que un sistema en l'estat x passa a l'estat $x + v_j$.
- Funció de propensió a_j , definida de tal manera que $a_j(x)dt$ és la probabilitat, donat un estat x del sistema, que la reacció R_j tingui lloc en algun lloc del sistema en el proper interval infinitesimal de temps dt .

Cal recordar que el que es vol obtenir és el següent:

$$P(x, t | x_0, t_0) = \text{Prob}\{X(t) = x \text{ donat un vector d'estat inicial } X(t_0) = x_0\} \quad (3.7)$$

De l'equació anterior se'n pot obtenir una equació que dóna l'evolució durant el temps, la qual determina la probabilitat que cada tipus d'individu tingui una població donada al cap d'un cert temps. Tanmateix, aquesta equació només és resoluble per alguns casos molt senzills i, per tant, no serveix pels propòsits desitjats. Cal optar, doncs, per un mètode que generi trajectòries simulades (realitzacions numèriques) de $X(t)$ versus t . No s'obindrà, doncs, una funció de probabilitat de densitat (que s'obtindria amb la resolució numèrica de l'equació derivada de l'Equació 3.7) sinó una mostra aleatòria de $X(t)$. Per generar aquestes trajectòries cal definir una nova funció de probabilitat similar, però no igual, que l'Equació 3.7:

$$P(\tau, j | x, t) = \text{Probabilitat que, per un vector d'estat } X(t) = x, \text{ la següent reacció} \\ \text{del sistema sigui } R_j \text{ i tingui lloc en el proper interval de temps } [t + \tau, t + \tau + d\tau) \quad (3.8)$$

Aquesta funció de probabilitat comprèn dues variables aleatòries: τ (el temps fins a la propera reacció) i j (l'índex de la reacció que toca). Aplicant lleis de probabilitat a l'Equació 3.8 es

pot obtenir la fórmula exacta següent, que és la clau de la simulació estocàstica que persegueix l'algorisme de Gillespie:

$$P(\tau, j | x, t) = a_j \exp(-a_0(x) \tau) \quad (3.9)$$

On:

$$a_0(x) = \sum_{k=1}^M a_k(x) \quad (3.10)$$

En altres paraules, es pot pensar en $a_0(x)$ com una recta formada per M segments, cadascun de longitud a_j . També es pot veure, doncs, que τ és una variable aleatòria exponencial i j una variable aleatòria independent sencera.

Per generar mostres de τ i j existeixen molts mètodes de Monte-Carlo segons el tipus de variable; es farà servir el mètode directe, que consisteix en el següent: s'extreuen dos nombres aleatoris r_1, r_2 a partir d'una distribució uniforme unitària i llavors es pren:

$$\tau = \frac{1}{a_0(x)} \ln\left(\frac{1}{r_1}\right) \quad (3.11)$$

$$j = \text{l'enter més petit que satisfaci } \sum_{k=1}^M a_k(x) > r_2 a_0(x) \quad (3.12)$$

En resum: l'algorisme de simulació estocàstica (ASE) del procés $X(t)$, anomenat algorisme de Gillespie, és el següent:

1. Inicialitzar l'estat del sistema $x = x_0$ i el temps $t = t_0$.
2. Amb el sistema en l'estat x i en el temps t , avaluar tots els $a_j(x)$ i la seva suma $a_0(x)$.
3. Generar el valor de τ i j mitjançant l'Equació 3.11 i l'Equació 3.12, respectivament.
4. Efectuar la reacció escollida ($x \rightarrow x + v_j$) i actualitzar el temps ($t \rightarrow t + \tau$).
5. Guardar x i t , i tornar al pas 2 excepte que es compleixi la condició de final.

La condició de final, en el cas del present treball, serà un temps màxim que, quan es compleixi, posarà fi a la simulació.

Com que l'algorisme de Gillespie proporciona una mostra aleatòria (estocàstica) de l'evolució de $X(t)$ en el temps, convindrà fer diverses simulacions a partir del mateix estat inicial per obtenir-ne una mitjana fiable.

4. PROGRAMACIÓ DEL SIMULADOR EN C++

Un cop compresa la topologia i la dinàmica de les xarxes complexes, ja es pot procedir a l'objectiu principal d'aquest treball, que no és altre que programar un simulador que, mitjançant l'algorisme de Gillespie estocàstic descrit anteriorment, permeti obtenir dades que, analitzades degudament, proveeixin informació sobre l'evolució, distribució de tipus de nodes, etc. al llarg del temps.

En aquest capítol no es pretén centrar-se en el codi sinó explicar conceptualment els diferents blocs del programa. El codi complet es pot trobar annexat en forma de CD al treball o a l'adreça web http://www.arnauduran.net/codi_tfc en llicència CC-BY-SA. A continuació s'expliquen les etapes del programa que utilitza l'algorisme de Gillespie per simular l'evolució d'un model epidèmic (SIS) (vegeu Models de propagació d'opinió) i, després de les quatre etapes, es comenta la modificació duta a terme al programa per permetre simular un model de dinàmica d'enllaços amb reconexió o sense. Les quatre etapes del programa descrites a continuació són les següents:

1. Determinació dels paràmetres
2. Generació de la xarxa
3. Algorisme de Gillespie múltiple
4. Impressió de les dades en codi llegible pel Matlab

La programació s'ha fet en l'entorn de desenvolupament integrat Microsoft Visual Studio Express 2013. La màquina utilitzada per realitzar les simulacions ha estat un Lenovo IdeaPad amb 8 GB de RAM, Windows 8.1 de 64 bits i un processador Intel Core i7 a 2,10 GHz.

4.1 Blocs del programa simulador del model epidèmic

4.1.1 Primer bloc: determinació dels paràmetres

El primer pas que cal fer en el programa és establir els valors dels paràmetres que determinaran la topologia de la xarxa i la seva evolució. Concretament, aquests paràmetres, que es demanen a l'usuari quan executa el programa, són els següents:

- Nombre de nodes
- Taxa d'infecció (β)

- Taxa de recuperació (μ)
- Percentatge de nodes infectats d'inici (ρ)
- Tipus de xarxa (retícula, completa, invariant d'escala, regular o d'Erdős-Rényi). Segons el tipus de xarxa, es demanen els següents paràmetres addicionals:
 - Retícula: valors d'amplada i d'alçada de la retícula (a i b)
 - Completa: cap
 - Invariant d'escala: valors de la γ i la m
 - Regular: valor del grau k
 - Erdős-Rényi: probabilitat d'aresta p

Els valors que no es demanen a l'usuari, però que es poden variar retocant el codi del programa de manera fàcil, són els següents:

- Nombre d'execucions de l'algorisme de Gillespie
- Temps final d'execució
- Temps de mostratge

4.1.2 Segon bloc: generació de la xarxa

Un cop determinats els paràmetres, es pot procedir a generar la xarxa desitjada per l'usuari. El procés de generació de la xarxa difereix segons el tipus. Cal recordar que el programa admet cinc tipus de xarxes (retícula, completa, invariant d'escala, regular i d'Erdős-Rényi). Les dues primeres i l'última es generen amb un mètode propi simple, mentre que la invariant d'escala i la regular es construeixen mitjançant un mètode anomenat *configuration model* [Newman, 2003]. Aquest mètode serveix per construir una xarxa a partir d'una distribució de graus $p(k)$ coneguda; és un mod molt incorrelacionat, és a dir, la correlació de graus que s'obté és molt baixa, així com el clústering. El *configuration model* consisteix en el següent:

1. Les entrades són el nombre de nodes N i la distribució de graus $p(k)$. El grau de cada node ($K(i)$) s'emmagatzema en una llista i ve donat per una fórmula que difereix segons el tipus de xarxa desitjada. Aquesta fórmula conté una variable aleatòria $y \in [0, 1]$. Un cop obtingut el valor donat per la fórmula cal fer ús de la funció sostre (en el cas del C++, `ceil()`) perquè sigui un nombre sencer.

2. Una vegada obtinguda la llista de graus, es fa la suma k_{total} de tots els graus de la xarxa. Per exemple, en una xarxa de quatre nodes de graus $k(1) = 2, k(2) = 3, k(3) = 1, k(4) = 4$ la suma de tots els graus seria 10.¹
3. Es crea una llista V de k_{total} elements on apareix, de manera ordenada, cada node tantes vegades com grau tingui. Seguint l'exemple anterior, aquesta llista seria $V = [1, 1, 2, 2, 2, 3, 4, 4, 4, 4]$.
4. Es fa una permutació a l'atzar de la llista (en C++ existeix la funció `random_shuffle()`, proporcionada per la llibreria `algorithm`).
5. S'agafen parelles de la llista, començant pel principi, les quals defineixen les arestes de la xarxa. Per exemple, suposant que la permutació obtinguda de la llista V anterior fos $V' = [4, 1, 3, 4, 2, 4, 2, 2, 1, 4]$, les arestes que caldria col·locar a la xarxa serien la $(4, 1)$, la $(3, 4)$, etc.²

Els procediments per generar cadascun dels tipus de xarxes, doncs, són els següents:

- Retícula: un cop comprovat que $a * b = N$, es fa un bucle per passar per totes les files i , a dins d'aquest, un bucle per passar per totes les columnes. S'enllaça cada element amb els seus veïns de sobre, sota, dreta i esquerra, excepte que es tracti d'un node que està als límits de la xarxa.
- Completa: s'enllacen tots els nodes amb tots.
- Invariant d'escala: en primer lloc, es comprova que $\gamma > 2$ i $m > 0$. Un cop fet això, es fa ús del *configuration model* per generar la xarxa; la fórmula per obtenir el grau d'un node i és la següent:

$$K(i) = m \left(1 - \frac{(N-1)}{N} * y \right)^{1/(1-\gamma)} \quad (4.1)$$

- Regular: un cop comprovat que el grau desitjat $k > 0$, es fa ús del *configuration model*. En aquest cas, no cal cap fórmula, ja que $K(i) = k$ per tot i .
- Erdős-Rényi: amb una $p \in [0, 1]$, es passa per totes les arestes possibles que correspondrien a una xarxa completa i es decideix si cadascuna d'elles existeix o no segons p .

¹La suma k_{total} ha de ser parell. En cas que fos senar, s'agafa un node a l'atzar i se li suma 1 al seu grau.

²En aquest punt cal tenir en compte dos aspectes. En primer lloc, cal descartar els llaços (parelles de nombres iguals, com per exemple la $(2, 2)$). En segon lloc, cal ignorar les arestes repetides que ja han sortit abans a la llista. Quan el nombre de nodes N és suficientment gran, aquests dos casos es donen molt poc.

Les arestes, en el programa, s'emmagatzemen en una llista de llistes. La llista mare és una llista d' N posicions, és a dir, conté tants elements com nodes té la xarxa. Cada posició conté, al seu torn, una llista amb tots els veïns del node corresponent a aquesta posició. Per exemple, si una xarxa té 10 nodes, l'estructura seria similar al que es pot veure a la Figura 4.1.³

Posició llista	Node	Veïns
0	1	3 6 7
1	2	3 4
2	3	9
3	4	0 1 7
4	5	1
5	6	6
6	7	0 5
7	8	0 3 8 9
8	9	7 9
9	10	2 7 8

Figura 4.1: Exemple del sistema de llistes per emmagatzemar una xarxa. La columna *Posició llista* és la posició de cada element de la llista; cada element és un node. Dins de cada element es troba la seva llista de veïns.

Aquest tipus d'emmagatzematge de xarxa permet afegir i treure arestes fàcilment, cosa necessària per posteriors modificacions del programa que requeriran processos de reconexió dels nodes. També es pot veure que, òbviament, si dos nodes són veïns, forçosament cadascun apareixerà a la llista de veïns de l'altre. Finalment, cal notar que les llistes de veïns no tenen per què estar ordenades: pot ser que ho estiguin segons el mètode de generació de xarxa emprat, o pot ser que no ho estiguin per a aquest mateix motiu o bé perquè s'han fet modificacions a la xarxa com, per exemple, reconexions.

4.1.3 Tercer bloc: algorisme de Gillespie múltiple

Una vegada la xarxa ha estat generada ja es pot procedir a simular estocàsticament la seva trajectòria segons els paràmetres β i μ que defineixen la interacció entre nodes. Per fer-ho es fa ús de l'algorisme de Gillespie, però no es fa una sola execució de l'algorisme si no que se'n fan múltiples (tantes com l'usuari vulgui); això es necessari ja que es tracta d'un algorisme estocàstic. Al final del programa es calcula la mitjana aritmètica de totes les execucions en cada instant de temps amb l'objectiu de veure'n l'evolució general.

³En C++, les posicions de llistes comencen sempre per l'índex 0. En el present treball, quan s'enumeren nodes pel seu "nom" es comença per l'1, tal com es pot veure a la Figura 4.1. Tanmateix, això només és un problema de nomenclatura que cal tenir present a l'hora de programar.

El que es descriu a continuació es du a terme per cada execució de l'algorisme de Gillespie; totes les execucions tenen lloc sobre la mateixa topografia de xarxa inicial. En primer lloc, s'infecten tants nodes com ha decidit l'usuari; la distribució és aleatòria en tota la xarxa. Un cop feta la infecció inicial, ja es pot tirar endavant l'algorisme de Gillespie, el qual farà iteracions fins que s'arribi al temps màxim determinat per l'usuari o bé fins que la xarxa assoleixi un estat de consens (tots els nodes sans o tots infectats). A l'inici de cada iteració, es construeix una llista amb tots els possibles esdeveniments a_j : aquests comprenen tant els esdeveniments de probabilitat μ (quan un node infectat es pot recuperar) com els esdeveniments de probabilitat β (quan un node sa es pot infectar per un dels veïns). Amb la llista feta, s'escull la reacció j que té lloc amb la fórmula corresponent descrita anteriorment, es calcula el temps τ que passarà fins la següent iteració i s'afecta la xarxa segons la reacció escollida (és a dir, s'infecta o es cura el node corresponent).⁴ A part de tot això, cal comprovar si en aquesta iteració toca guardar el mostratge de dades d'infectats (l'usuari, a l'inici, determina l'interval de temps de mostratge);⁵ el mostratge es desa en una llista que, posteriorment, serveix per calcular la mitjana de totes les execucions en cada instant de mostreig. Finalment, al final de l'iteració s'imprimeixen dades com ara el percentatge final d'infectats, el nombre d'iteracions i el temps real que ha tardat l'execució.

4.1.4 Quart bloc: impressió de les dades en codi llegible pel MATLAB

En aquest punt el programa ja ha obtingut les trajectòries de la xarxa per cada execució, les quals estan emmagatzemades en una llista. Interessa, però, disposar de les dades en un format llegible pel MATLAB, de tal manera que puguin ser analitzades amb més comoditat pel paquet matemàtic. Amb aquest objectiu, el programa en C++ obre i escriu un fitxer `.m` amb la sintaxi pròpia del MATLAB que conté la mitjana del nombre d'infectats en cada instant de mostreig. Obrint el fitxer amb el MATLAB i executant-lo es pot obtenir directament la gràfica de l'evolució de la xarxa al llarg del temps sense fer-hi més modificacions.

4.2 Modificació del programa per permetre reconexió (sistema de dinàmica d'enllaços)

El programa descrit a la secció anterior és capaç de simular l'evolució d'un model epidèmic, però per simular un sistema de dinàmica d'enllaços (vegeu la secció Models de propagació d'opinió)

⁴Els càlculs comprenen la intervenció de variables aleatòries proporcionades per la llibreria `stdlib.h`.

⁵Cal tenir en compte que entre iteració i iteració pot passar més d'un instant de mostreig.

cal modificar-lo.⁶ Al mateix temps, s'aprofitarà per modificar el programa per permetre reconexions dels nodes. De fet, els dos casos s'engloben en un mateix, ja que una xarxa que segueix el model de dinàmica d'enllaços és una xarxa adaptativa (amb reconexió) amb una probabilitat de reconexió $p = 0$. Per tant, cal només modificar el programa anterior perquè permeti reconexió dels nodes amb una probabilitat p determinada per l'usuari.

El que cal modificar, gairebé de manera exclusiva, respecte l'algorisme del model epidèmic és l'algorisme de Gillespie. En aquest cas l'evolució dels nodes ja no ve determinat per una β i una μ , sinó que depèn exclusivament del paràmetre p . En el model que es vol programar, es tria una aresta, a continuació es tria un dels dos nodes de l'aresta a l'atzar, i un cop escollit es decideix si aquest node trenca l'enllaç i es reconnecta a un node del seu mateix tipus (amb probabilitat p) o bé adopta l'estat de l'altre node (amb probabilitat $1 - p$). Els esdeveniments a_j de l'algorisme de Gillespie, doncs, són tots els veïns de tipus diferent de cada node de la xarxa. Llavors, un cop escollit el node sobre el que cal actuar mitjançant l'algorisme de Gillespie, es tria segons la probabilitat p si cal reconnectar-lo o cal canviar-lo d'estat. La reconexió, si és el cas, es fa de manera aleatòria amb un altre node del mateix tipus de la xarxa amb el qual, òbviament, el node en qüestió encara no estigui connectat (no es contemplen arestes dobles). La resta de l'algorisme és gairebé idèntica a la del cas del model epidèmic.

El tema de la reconexió planteja dues qüestions addicionals que cal tenir en compte:

1. Pot ser que, en voler reconnectar un node, aquest ja estigui connectat amb tots els altres nodes del seu tipus i que, per tant, sigui impossible reconnectar-lo. En aquest cas, es decideix saltar a la següent iteració sense fer cap afectació a la xarxa. A més, el programa imprimeix un avís per pantalla en cas que això passi.
2. Com que hi ha reconexions, és possible (i, certament, ocorre sovint) que la xarxa se separi en dues xarxes inconnexes, formades cadascuna exclusivament per nodes del tipus A o del tipus B. Això sol passar quan el nombre de tipus A o B s'aproxima a zero. En el cas que hi hagi separació, el programa fa una comprovació i emet un avís per pantalla.

El procés d'impressió de dades pel MATLAB és el mateix, només que en aquest programa s'afegeix la impressió del mostratge de cada execució per poder-ne fer un anàlisi més detallat, ja que en el cas de xarxes adaptatives la mitjana no és gaire representativa de la dinàmica del sistema.

⁶Arribats a aquest punt, ja no es parlarà de "node infectat" sinó de "node tipus A", per evitar confusions. Tanmateix, es tracta tan sols d'un tema de nomenclatura.

4.2.1 Modificació del programa per permetre infeccions localitzades

En el cas de les xarxes adaptatives, també interessarà començar amb nodes de tipus A localitzats, és a dir, fer una "infecció" localitzada –també anomenada intencionada o dirigida– en aquells nodes que tinguin el grau més alt. Per fer-ho, l'usuari determina el nombre k de nodes de major grau que vol que siguin de tipus A de manera intencionada, i la resta de tipus A s'establiran de manera aleatòria fins arribar al percentatge inicial desitjat. Cal recordar que en cada execució es crida la rutina per establir els nodes de tipus A i B, de tal manera que en cada execució la distribució inicial serà diferent (però amb el mateix percentatge inicial).

4.3 Altres consideracions

A continuació es descriuen alguns temes que no han estat coberts en les seccions anteriors:

- Per emmagatzemar les diferents llistes necessàries per la simulació s'ha utilitzat la llibreria `vector`. Aquesta llibreria permet tenir contenidors dinàmics de mida il·limitada, al contrari que les llistes "normals" del C++, que no permeten emmagatzemar gaires elements (cal recordar que en el cas de xarxes se sol tractar amb xarxes de l'ordre de milers de nodes, com a mínim). Aquesta llibreria permet un nou tipus de constructor anomenat `vector` que permet definir llistes de qualsevol tipus (fins i tot de tuples). Mitjançant la instrucció `push_back` s'afegeixen elements al final de llista, i també es disposa de les funcions `insert`, `erase` i `size`, entre d'altres, per dur a terme diferents operacions.
- El temps d'execució del programa quan es tracten xarxes de mida gran és força elevat en comparació amb programes similars en altres llenguatges com, per exemple, Python. Això és degut al mateix C++. S'ha intentat optimitzar el temps d'execució al màxim i, fins i tot, s'ha intentat fer programació en paral·lel, encara que no ha sigut possible per diverses raons com, per exemple, que en la majoria de bucles del programa la condició de final del bucle no està definida de manera exacta abans de començar-lo.
- Durant el procés de programació s'ha fet ús de funcions auxiliars per fer provatures (*testing*) que, per exemple, mostren per pantalla la llista de nodes i d'adjacències, si estan infectats o no, etc. Això ha servit per corregir errors i assegurar-se que es cobrien tots els escenaris possibles (separació de grups, nodes sense veïns per reconnectar, etc.).

5. SIMULACIONS I ANÀLISI DE RESULTATS

Un cop programat el simulador estocàstic en C++ ja és possible tirar endavant les simulacions desitjades. En aquest capítol es pretén estudiar tres casos de dinàmiques diferents:

1. Model epidèmic: es comprovarà que les trajectòries simulades estocàsticament coincideixen amb la solució analítica obtinguda a partir de la integració numèrica de les equacions diferencials.
2. Model de dinàmica d'enllaços sense reconexió: es comprovarà que la probabilitat d'èxit predita per la literatura existent sobre el tema coincideix amb l'obtinguda per mitjà del simulador, i es farà el corresponent contrast estadístic per verificar-ho.
3. Model de dinàmica d'enllaços amb reconexió: els resultats d'aquest apartat són coneixement nou, ja que no existeix literatura sobre el cas concret de xarxes complexes adaptatives. Se centrarà especialment en xarxes invariants d'escala, que són de topologia assimilable a xarxes reals com, per exemple, les xarxes socials.

5.1 Simulació i estudi del model epidèmic

L'objectiu de l'estudi d'aquesta secció és utilitzar el simulador programat en C++ per comprovar que l'evolució de la quantitat d'infectats donada pel programa s'ajusta correctament al model analític. A més, també s'observarà que les equacions diferencials també s'ajusten força bé al comportament d'una xarxa heterogènia com és una xarxa invariant d'escala.

Per fer l'estudi del model epidèmic es tractaran tres topologies de xarxa diferents (xarxa regular, xarxa d'Erdős-Renyi i xarxa invariant d'escala). La xarxa estarà composta per 5000 nodes, amb una taxa de recuperació $\mu = 1$, un grau mitjà $\langle k \rangle = 10$ i, pel que fa a la taxa de dispersió, es diferenciarien dos casos per cada tipus de xarxa:

1. $\beta = 0,3$. Com que, en aquest cas, $\beta > \frac{\mu}{\langle k \rangle}$, a llarg termini hi haurà una certa fracció d'infectats. Analíticament, segons l'Equació 3.4, aquesta fracció és del 66,6%.
2. $\beta = 0,05$. Com que, en aquest cas, $\beta < \frac{\mu}{\langle k \rangle}$, a llarg termini la infecció s'extingirà.

A més a més, per cada xarxa i per cadascuna de les dues taxes de dispersió β se simularà el comportament de la xarxa per dos fraccions inicials d'infectats ρ diferents, una de força petita i una de força gran:

1. $\rho = 0,01$ (o, el que és el mateix, 50 nodes).
2. $\rho = 0,9$ (o, el que és el mateix, 4500 nodes).

Els paràmetres per generar cadascun dels tipus de xarxes durant la simulació, de tal manera que tinguin un grau mitjà $\langle k \rangle = 10$, són els següents:

- Xarxa regular: $k = 10$.
- Xarxa aleatòria (xarxa d'Erdős-Rényi): $p = \frac{\langle k \rangle}{N} = \frac{10}{5000} = 0,002$.
- Xarxa invariant d'escala: el grau mitjà d'aquesta xarxa ve donat per l'Equació 5.1 de sota. Fixant una $\gamma = 3$, s'obté un grau mínim $m = 5$.

$$\langle k \rangle = \frac{\gamma - 1}{\gamma - 2} m (1 - N^{\frac{2-\gamma}{\gamma-1}}) \approx \frac{\gamma - 1}{\gamma - 2} m \quad \text{per a } N \text{ grans} \quad (5.1)$$

Amb les dades anteriors ja es poden simular tots els 12 casos, resumits a la Taula 5.1. El temps màxim de cadascuna de les 12 simulacions es pot aproximar per l'Equació 3.6; en algun dels casos, s'ha hagut de repetir la simulació ja que s'havia pres un temps massa curt per permetre que el model s'estabilitzés. Per cada simulació s'han fet 50 execucions de l'algorisme, prenent la mitjana de cada instant de temps al final.

	Xarxa regular	Xarxa ER	Xarxa invariant d'escala
$\beta = 0,3$	$\rho = 0,01, t = 8$	$\rho = 0,01, t = 8$	$\rho = 0,01, t = 16$
	$\rho = 0,9, t = 8$	$\rho = 0,9, t = 8$	$\rho = 0,9, t = 16$
$\beta = 0,05$	$\rho = 0,01, t = 16$	$\rho = 0,01, t = 16$	$\rho = 0,01, t = 50$
	$\rho = 0,9, t = 16$	$\rho = 0,9, t = 16$	$\rho = 0,9, t = 50$

Taula 5.1: Resum de les 12 simulacions dutes a terme per cadascun dels tres tipus de xarxes, amb els paràmetres β , ρ i t .

Paral·lelament a les simulacions obtingudes mitjançant l'algorisme de Gillespie programat en C++, també s'han representat les corbes teòriques resultants d'integrar l'equació diferencial Equació 3.2 amb les condicions inicials donades per cada cas:

$$\frac{di(t)}{dt} = \beta \langle k \rangle i (1 - i) - \mu i \quad (3.2 \text{ revisitada})$$

La integració de l'equació diferencial s'ha dut a terme mitjançant la instrucció `ode45` del Matlab.

Un cop feta la simulació i la representació del model teòric, doncs, s'han obtingut les següents gràfiques: Figura 5.1 (xarxa regular), Figura 5.2 (xarxa d'Erdős-Rényi) i Figura 5.3 (xarxa

invariant d'escala). Com es pot veure, les simulacions i l'evolució predita per l'equació diferencial segueixen corbes molt similars, amb petites diferències en el temps de creixement i en la fracció final d'infectats en el cas de $\beta = 0,3$. Cal recordar que en el cas de $\beta = 0,3$ s'esperava una fracció final d'infectats del 66,6% (segons l'Equació 3.4) i, tal com es veu en els gràfics, en els tres casos la fracció d'infectats un cop estabilitzat el sistema és lleugerament inferior. El model d'equacions sobreestima el nombre d'infectats a l'equilibri perquè suposa una barreja homogènia dels tipus de nodes, ergo ignora els casos que un node infectat ja té tots els veïns infectats i, per tant, no pot transmetre la infecció: cal recordar que a les equacions se suposa que hi ha $\langle k \rangle (1 - i)$ nodes susceptibles (sans). Com més baix és el grau mitjà $\langle k \rangle$, més important esdevé aquesta negligència de les equacions. D'altra banda, per una β petita que fa que el nombre d'infectats tendeixi a zero al cap del temps, aquesta sobreestimació del model analític deixa de ser important, tal com es pot comprovar a les gràfiques: les simulacions i la corba teòrica, quan tendeixen a zero, coincideixen quasi perfectament.

En el cas de la xarxa heterogènia (la xarxa invariant d'escala), encara que a priori la coincidència entre les simulacions i el model teòric havia de ser menor que en els altres casos, es pot veure que s'ajusten força bé, igual que els altres dos casos, per tant es pot confirmar que l'Equació 3.2 és també vàlida per predir el comportament de xarxes heterogènies.

En els annexos del present treball es poden trobar els resultats de les 12 simulacions dutes a terme.

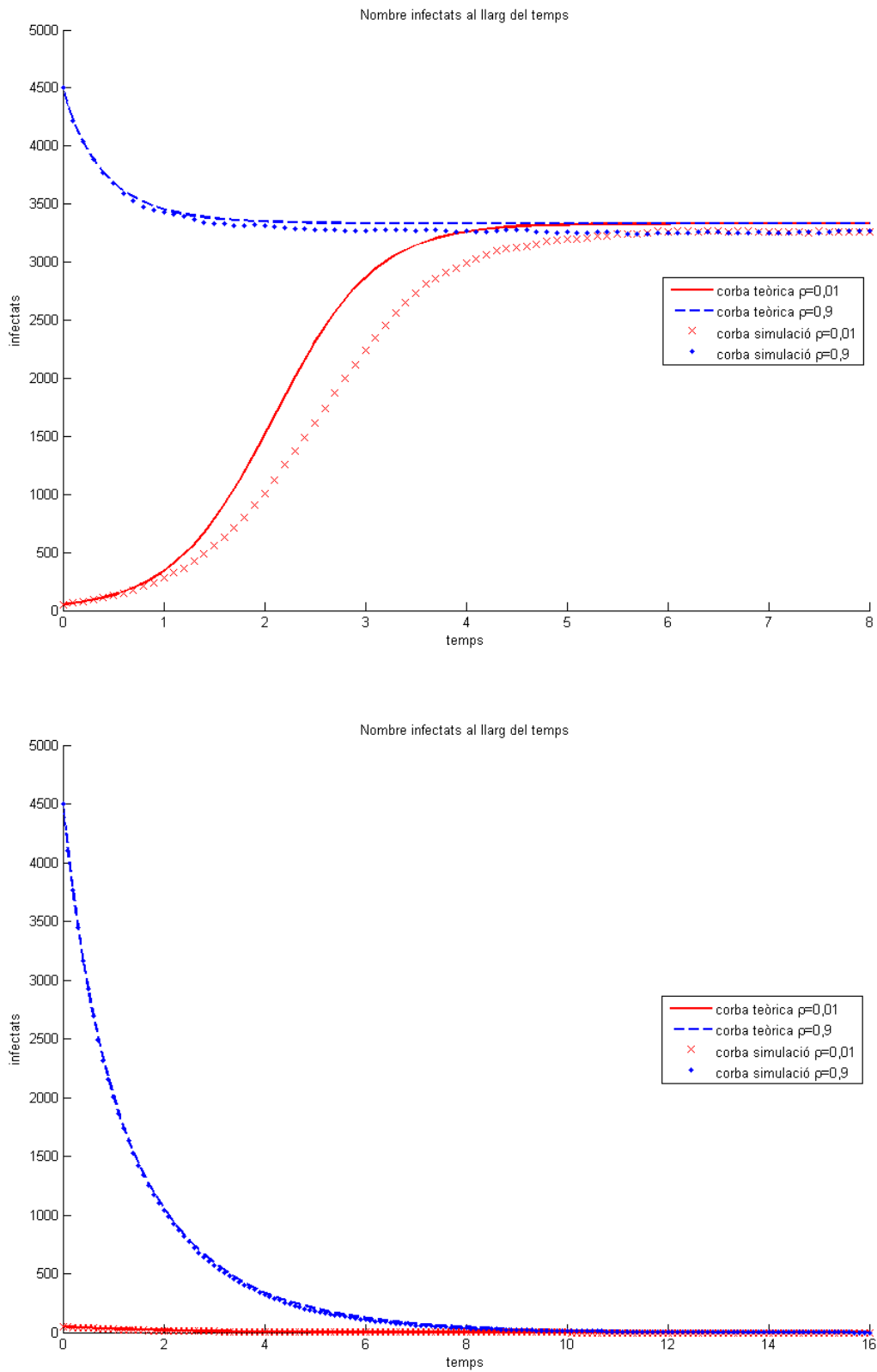


Figura 5.1: Simulació del model epidèmic per una xarxa regular de grau 10 i per dues taxes d'infecció diferents

Imatge superior: $\beta = 0,3$; imatge inferior: $\beta = 0,05$.

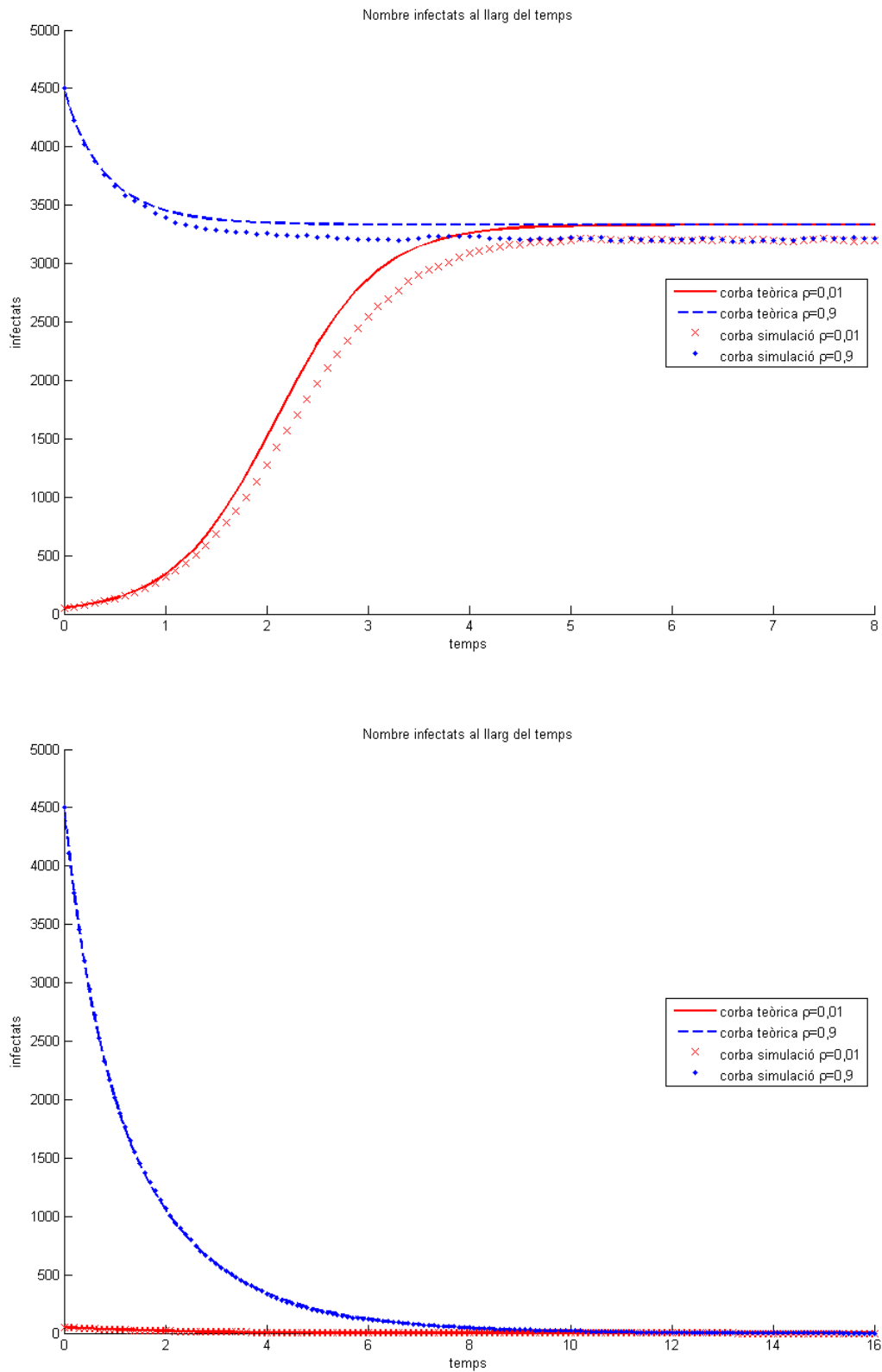


Figura 5.2: Simulació del model epidèmic per una xarxa d'Erdős-Rényi amb $p = 0,02$ i per dues taxes d'infecció diferents

Imatge superior: $\beta = 0,3$; imatge inferior: $\beta = 0,05$.

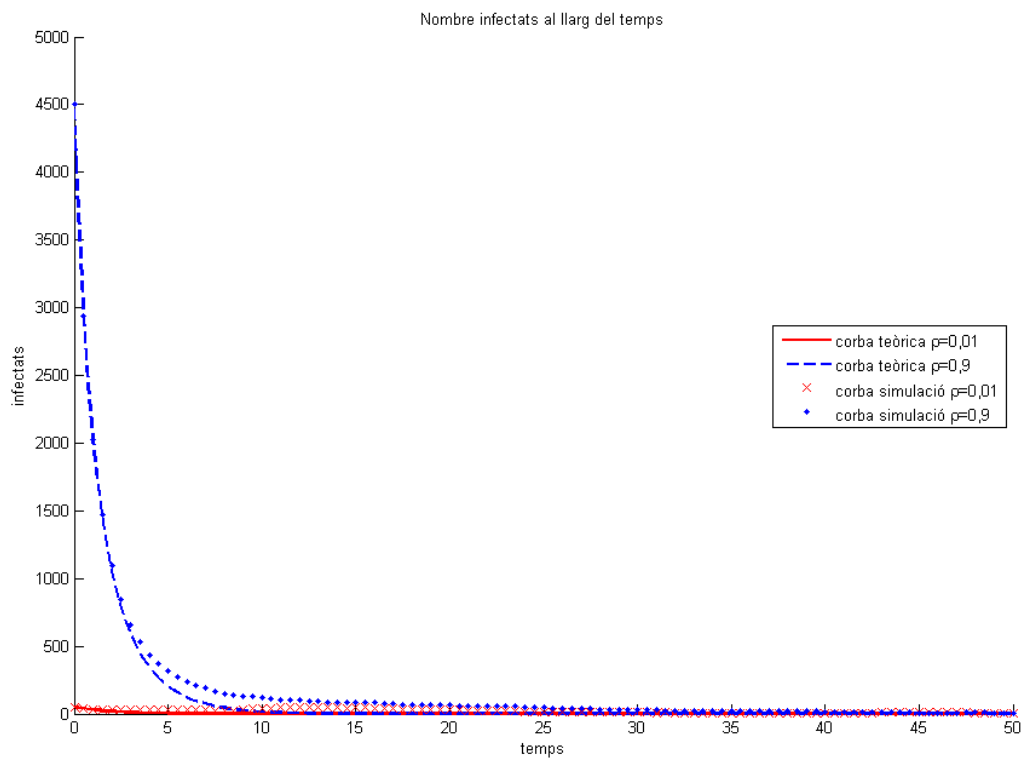
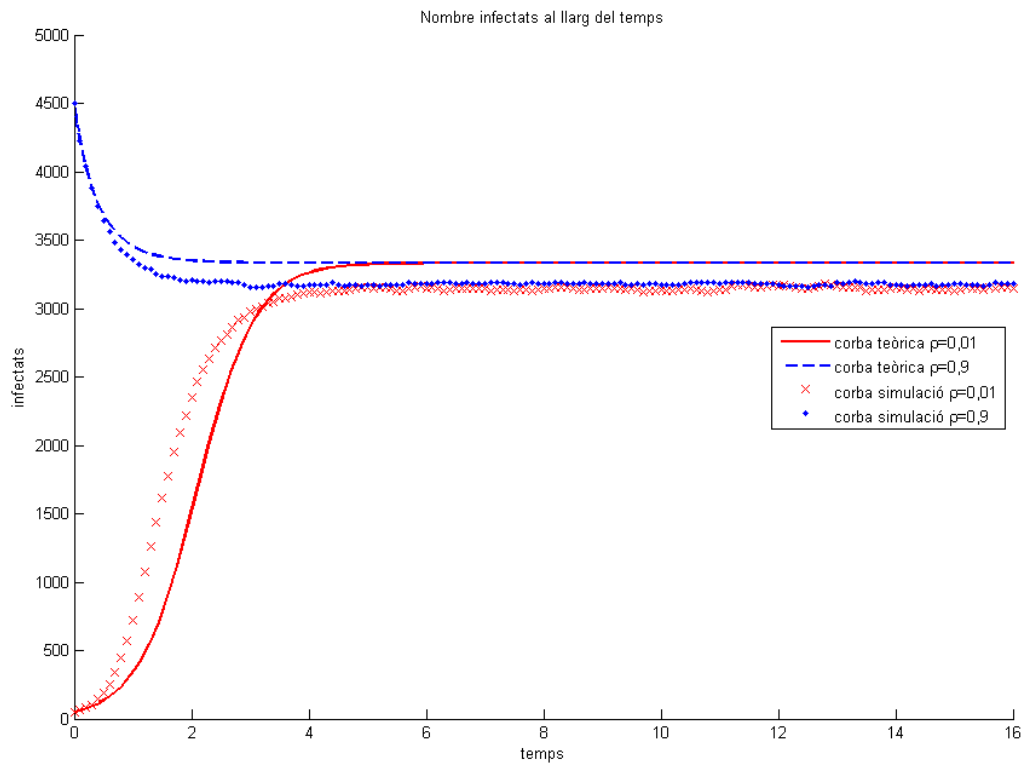


Figura 5.3: Simulació del model epidèmic per una xarxa invariant d'escala amb $\gamma = 3$ i $m = 5$ i per dues taxes d'infecció diferents

Imatge superior: $\beta = 0,3$; imatge inferior: $\beta = 0,05$. En el cas de la imatge inferior, s'han dibuixat només un cinquè dels punts obtinguts per facilitar la visualització del gràfic.

5.2 Simulació i estudi de la dinàmica d'enllaços sense reconexió

A la secció Comparació de les dinàmiques dels tres models del present treball s'ha analitzat el comportament dels tres models de propagació d'opinió (model del votant, procés d'invasió i dinàmica d'enllaços). En aquest apartat es pretén estudiar el tercer d'aquests models sense reconexió i, en particular, demostrar que la probabilitat d'èxit ε_A del model de dinàmica d'enllaços és, independentment de la topologia de la xarxa però amb els nodes A i B distribuïts aleatòriament, el descrit per l'Equació 3.5:

$$\varepsilon_A(\rho) = \rho \quad (3.5 \text{ revisitada})$$

L'estudi es durà a terme de la següent manera: es faran 50 execucions per cinc densitats inicials (ρ) de nodes tipus A, i es comptarà quin percentatge de vegades s'assoleix un consens de A; en altres paraules, es trobarà la probabilitat d'èxit ε_A per cadascuna de les densitats inicials.¹ Aquest procés es durà a terme en dos tipus de xarxes de 1000 nodes: una xarxa regular de grau 4 i una xarxa invariant d'escala amb $\gamma = 3$ i $m = 1$. Pel que fa al temps màxim de la simulació, s'ha optat per fer servir l'equació Equació 3.6 per trobar-ne una aproximació, encara que seria només aplicable al model del votant.

Els resultats de les simulacions es poden veure a la Taula 5.2a (per la xarxa regular) i a la Taula 5.2b (per la xarxa invariant d'escala). A més a més, a les imatges Figura 5.4 i Figura 5.5 es mostra, respectivament, com s'ajusten els resultats obtinguts mitjançant la simulació al comportament predit per l'Equació 3.5.

ρ	ε_A	ρ	ε_A
10%	12,00%	10%	7,32%
20%	18,00%	20%	17,78%
30%	27,08%	30%	23,81%
40%	44,68%	40%	40,00%
50%	47,37%	50%	56,10%

(a) Xarxa regular de grau 4

(b) Xarxa invariant d'escala amb $\gamma = 3$ i $m = 1$

Taula 5.2: Probabilitat d'èxit (ε_A) segons diferent percentatge inicial (ρ) per dos tipus de xarxes diferents de 1000 nodes (50 execucions) amb dinàmica d'enllaços sense reconexió

¹Tot i que el temps màxim establert és força gran, algunes de les simulacions, tal com es pot veure a l'annex corresponent, no arriben a assolir un consens en cap de les dues direccions al cap del temps màxim. Aquestes simulacions es descarten de cares al càlcul de ε_A .

Probabilitat d'èxit en funció de la densitat inicial per una xarxa regular

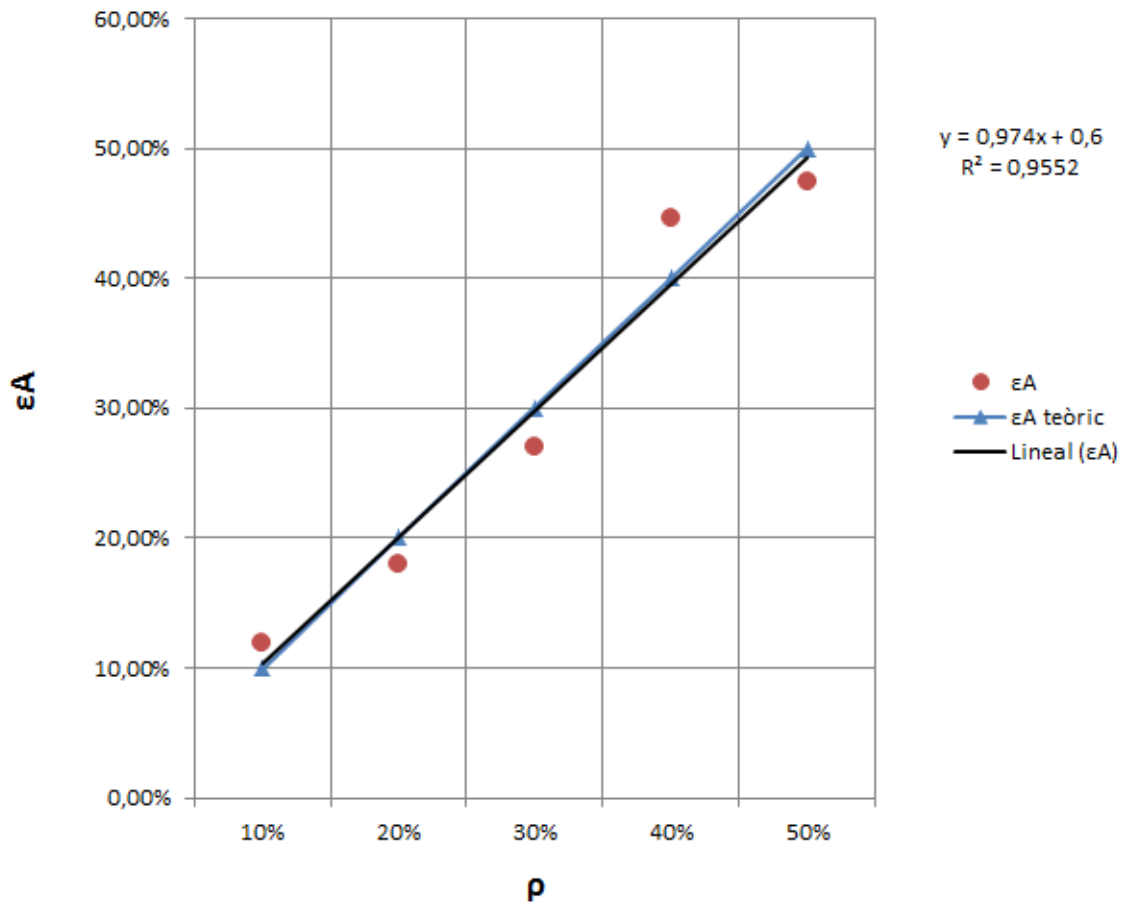


Figura 5.4: Representació de la probabilitat d'èxit d'A (ε_A) en funció percentatge inicial (ρ) per una xarxa regular amb dinàmica d'enllaços sense reconexió de les diferents simulacions de la Taula 5.2a juntament amb els valors teòrics donats per l'Equació 3.5.

Probabilitat d'èxit en funció de la densitat inicial per una xarxa invariant d'escala

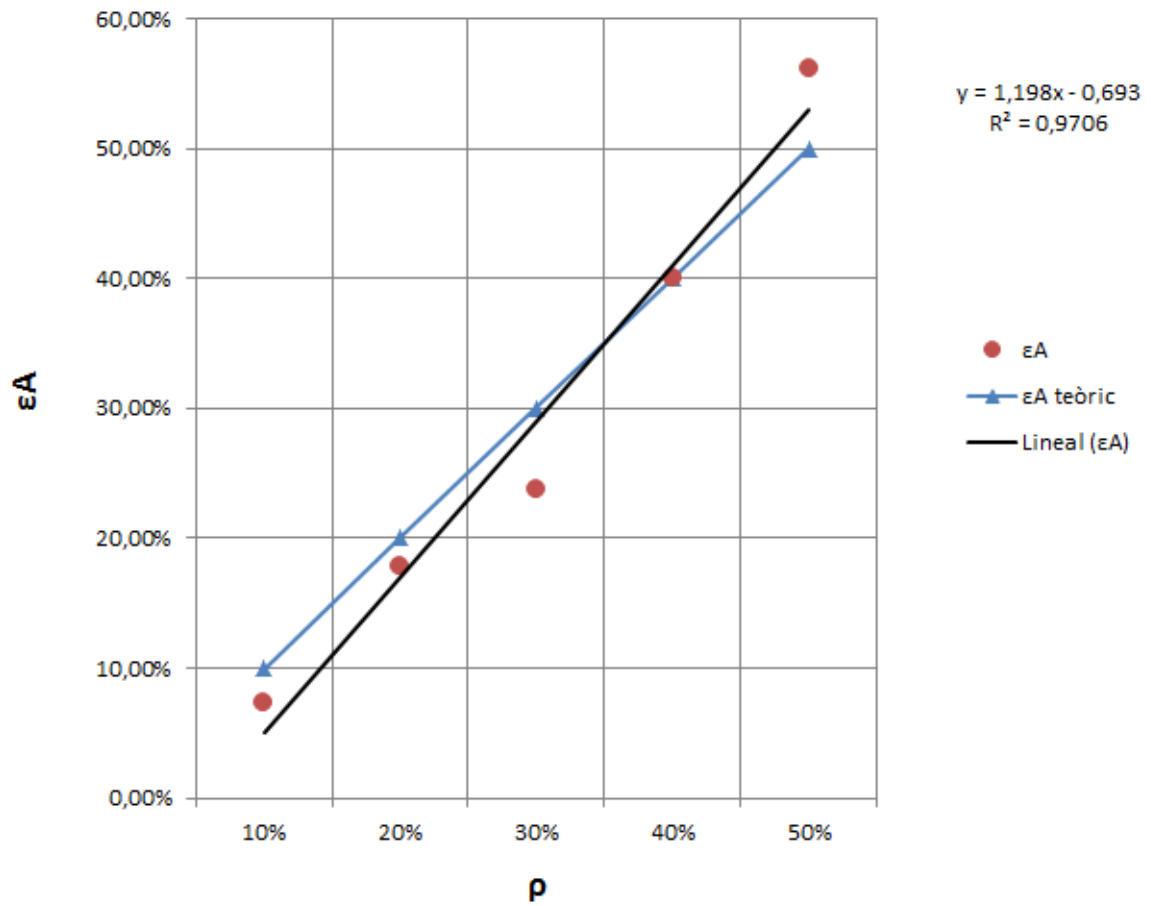


Figura 5.5: Representació de la probabilitat d'èxit d'A (ε_A) en funció percentatge inicial (ρ) per una xarxa invariant d'escala amb dinàmica d'enllaços sense reconexió de les diferents simulacions de la Taula 5.2b juntament amb els valors teòrics donats per l'Equació 3.5.

La valoració dels resultats obtinguts a partir d'un simple anàlisi visual és positiva, ja que sembla que el comportament predit es compleix en el cas de les dues xarxes. Tanmateix, convé avaluar com de bé s'ajusten els punts obtinguts a la recta prevista mitjanant una mesura estadística. El paràmetre estadístic que cal utilitzar en aquest cas és el coeficient de determinació, altrament dit coeficient R^2 ($R^2 \in [0, 1]$) que es defineix de la següent manera:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (5.2)$$

On SS_{res} i SS_{tot} són la suma de quadrats residual i total, respectivament, i es defineixen per:

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (5.3)$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (5.4)$$

On y_i són els punts observats (obtinguts amb les simulacions), \bar{y} és la seva mitjana i f_i són els punts reals (predits pel model teòric).

Si es duen a terme els càlculs per cadascun dels dos casos de l'estudi (xarxa regular i xarxa invariant d'escala) s'obtenen els següents valors R^2 :

- Xarxa regular: $R^2 = 0,9544$
- Xarxa invariant d'escala: $R^2 = 0,9407$

Tenint en compte que $R^2 \in [0, 1]$ i que R^2 és millor com més proper a 1 és, es pot dir que els valors simulats s'ajusten molt bé als valors predits pel model teòric de l'Equació 3.5, tant per una xarxa regular com per una xarxa invariant d'escala.

A part del coeficient de determinació, per acabar de comprovar si el comportament s'ajusta al predit es pot dur a terme una anàlisi estadística de contrast d'hipòtesis, que consisteix en el següent:

- Hipòtesi nul·la H_0 : el pendent de la recta de les simulacions és igual que el pendent de la recta teòrica, és a dir, $m = 1$.
- Hipòtesi alternativa H_1 : els pendents són diferents, és a dir, $m \neq 1$.

Si H_0 és certa, l'estadístic següent segueix una distribució t de student amb $n - 2$ graus de

llibertat:

$$t = \frac{s_x \sqrt{n-2}}{s_y \sqrt{1-r^2}} (b_{21} - \beta_{21}^o) \tag{5.5}$$

On:

- s_x i s_y són les variàncies mostrals de les variables x, y , respectivament.
- n és el nombre de dades, en aquest cas $n = 5$.
- r és la correlació mostral, que ve donada per $\frac{s_{xy}}{s_x s_y}$, on s_{xy} és la covariància de les dues variables x, y .
- b_{21} és el coeficient de regressió mostral, que ve donat per $\frac{s_{xy}}{s_x^2}$.
- β_{21}^o és el valor del pendent que volem comparar, és a dir, $\beta_{21}^o = 1$.

Les dades obtingudes mitjançant les fórmules anteriors per cadascun dels dos casos d'estudi són les següents (obtingudes mitjançant el programari estadístic Minitab 9):

s_x	250	s_x	250
s_y	248	s_t	370
s_{xy}	243,55	s_{xy}	299,45
r	0,003928226	r	0,003237297
b_{21}	0,0038968	b_{21}	0,0047912
t	-1,739228492	t	-1,164703539
(a) Xarxa regular		(b) Xarxa invariant d'escala	

Taula 5.3: Càlcul dels paràmetres estadístics necessaris per calcular l'estadístic de contrast t per cada tipus de xarxa.

Si es fa el contrast bilateral s'obté, doncs, en cada cas (amb un nivell de significació $\alpha = 0,05$):

- Xarxa regular: p-valor = 0,09019 > 0,025 (Figura 5.6a), ergo s'accepta la hipòtesi nul·la $H_0: m = 1$ (amb un possible error de tipus II). La simulació segueix el model predit correctament.
- Xarxa invariant d'escala: p-valor = 0,1642 > 0,025 (Figura 5.6b), ergo s'accepta la hipòtesi nul·la $H_0: m = 1$ (amb un possible error de tipus II). La simulació segueix el model predit correctament.

Seria bo disposar de més dades de simulacions per fer el contrast d'hipòtesis, ja que 3 graus de llibertat per una distribució de Poisson són pocs.

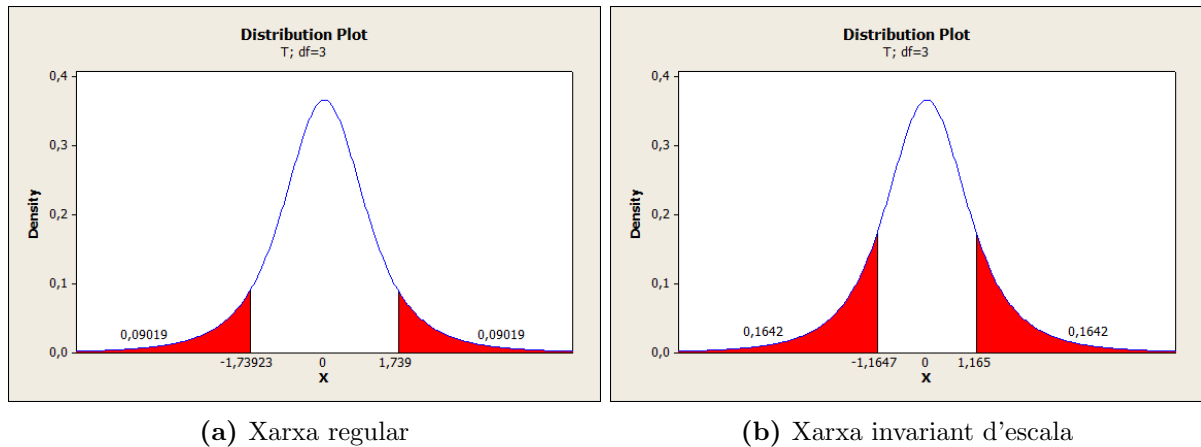


Figura 5.6: Distribució de Poisson de 3 graus de llibertat amb el valor de l'estadístic de contrast t obtingut mitjançant l'Equació 5.5 per cada tipus de xarxa, juntament amb el corresponent p-valor.

5.3 Simulació i estudi de la dinàmica d'enllaços amb reconexió

En aquest darrer apartat de simulacions es vol estudiar el cas d'una xarxa que segueix un model de dinàmica d'enllaços –igual que en la secció anterior– però, en aquest cas, el node escollit té probabilitat p de reconnectar-se a un altre node del seu mateix tipus i probabilitat $1 - p$ d'adoptar l'estat del seu node veí. A més a més, les regularitats observades en la secció anterior no es compleixen a causa de la reconexió dels nodes. Cal notar també que per aquest tipus de xarxes adaptatives no existeix literatura amb la qual poder contrastar els resultats, per la qual cosa l'anàlisi dels *outputs* de les simulacions serà més aviat qualitatiu que no pas quantitatiu. Els resultats obtinguts, doncs, són investigació original.

L'objectiu és estudiar dos tipus de xarxes que segueixen el model de dinàmica d'enllaços amb reconexió, la regular i la invariant d'escala, amb diferents percentatges inicials de nodes de tipus A, amb una probabilitat de reconexió fixa de $p = 0, 1$ i de 1000 nodes.² El simulador C++ permet veure la trajectòria de cadascuna de les execucions; se'n faran 30 de cada cas, per poder observar de manera qualitativa com es comporta el sistema. Cal destacar que, en aquest apartat, la mitjana de totes les execucions, a diferència de quan s'estudiava el model epidèmic a la secció Simulació i estudi del model epidèmic, no dóna gaire informació sobre la dinàmica de la xarxa, ja que l'evolució pot variar molt d'una execució a l'altra, tal com es veurà a continuació.

²S'escull una p molt baixa perquè, en una xarxa de 1000 nodes (relativament pocs), si s'estableix una p elevada és fàcil que se separi la xarxa en dos grups no conexas, cadascun de diferent tipus de node.

5.3.1 Xarxa regular

En primer lloc se simularà l'evolució d'una xarxa regular de grau $k = 10$ per tres poblacions inicials de nodes A diferents ($\rho = 10\%$, $\rho = 30\%$, $\rho = 50\%$). Les gràfiques obtingudes es poden veure a la Figura 5.7, a la Figura 5.8 i a la Figura 5.9. Com es pot veure, per exemple, el comportament de la xarxa en el cas $\rho = 10\%$ és similar al que s'esperaria per una xarxa d'identica topologia amb una $p = 0$: aproximadament el 90% de les simulacions acaben amb un 0% de nodes de tipus A (o gairebé el 0%: quan es veuen a la gràfica línies estables horitzontals és que la xarxa s'ha separat en dos grups inconnexos i, per tant, ja no pot evolucionar més). Les gràfiques pels casos de $\rho = 30\%$ i $\rho = 50\%$ a primera vista no aporten gaire informació, ja que hi ha massa soroll, però serviran per fer comparacions amb la xarxa invariant d'escala. En les gràfiques també es pot observar l'evolució de la mitjana de les execucions que, encara que no és gaire representativa de l'evolució, sí que dóna una idea de quina és la tendència general de la dinàmica de la xarxa.

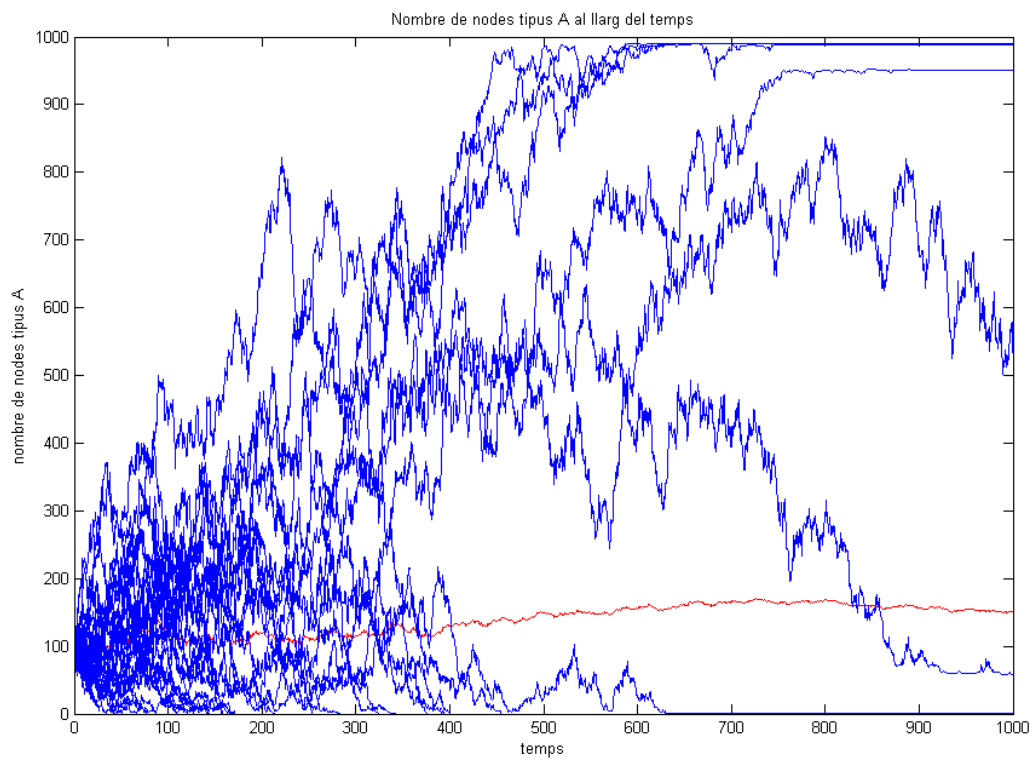


Figura 5.7: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0, 1$ i amb un percentatge inicial de nodes tipus A del 10%. En vermell es mostra la mitjana de les execucions.

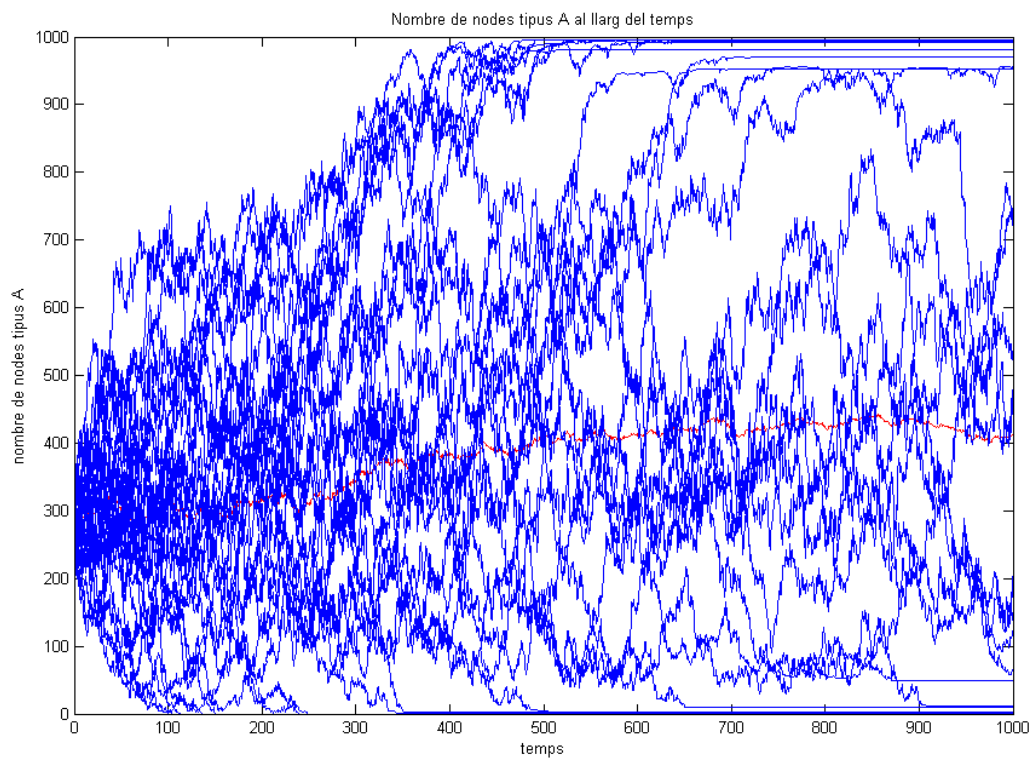


Figura 5.8: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30%. En vermell es mostra la mitjana de les execucions.

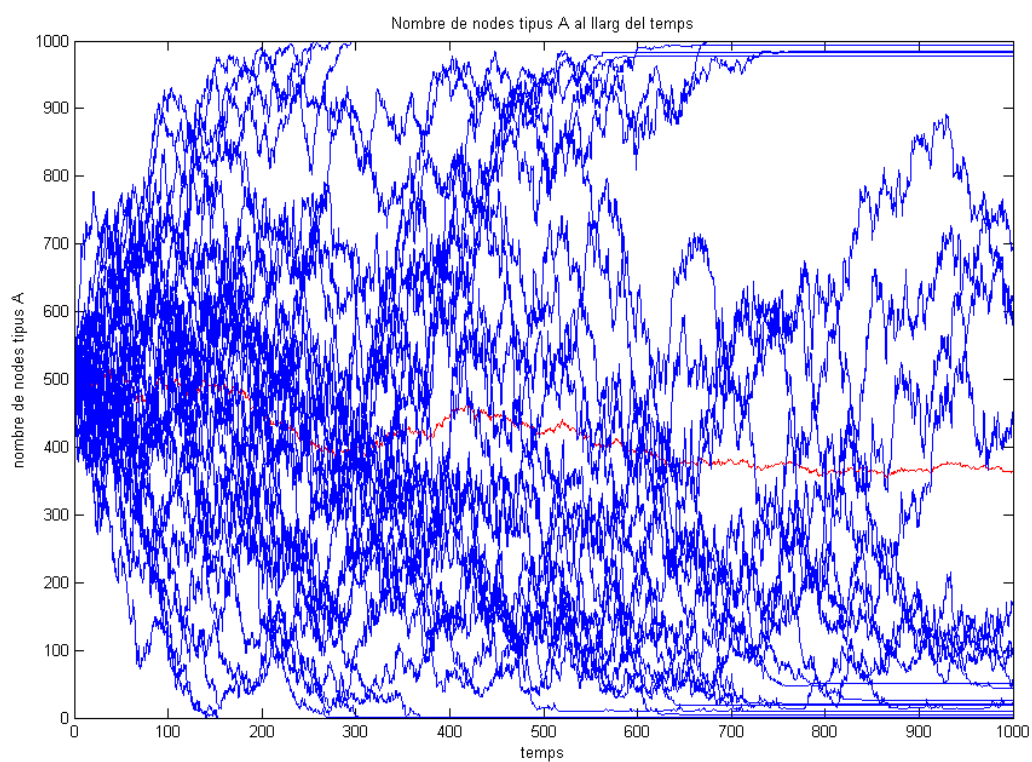


Figura 5.9: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa regular de grau 10 amb reconexió de $p = 0, 1$ i amb un percentatge inicial de nodes tipus A del 50%. En vermell es mostra la mitjana de les execucions.

5.3.2 Xarxa invariant d'escala

Després de la xarxa regular se simularà la dinàmica d'una xarxa invariant d'escala amb paràmetres $\gamma = 3$ i $m = 5$, que són equivalents a un grau mitjà $k = 10$,³ i així es podran comparar les gràfiques amb les de la secció anterior. La característica principal d'una xarxa invariant d'escala és que existeixen uns pocs nodes amb moltes connexions (*hubs*) que tenen una important influència sobre la dinàmica de la xarxa. Les gràfiques de les simulacions es poden veure a la Figura 5.10, a la Figura 5.11 i a la Figura 5.12.

Tal com es pot veure a primer cop d'ull a la gràfica de la simulació per un percentatge inicial $\rho = 10\%$ observant l'esdevenir de les execucions i la mitjana comparant-la amb la gràfica de la xarxa regular, sembla que per la xarxa invariant d'escala les trajectòries tendeixen a assolir un consens del 0% un un temps molt menor que en el cas de la xarxa regular. Això s'explica per la següent raó: la xarxa invariant d'escala de grau mitjà 10 i grau mínim 5 té molts nodes de grau menor que 10,⁴ al contrari que la xarxa regular, en la qual tots els nodes tenen grau 10; per tant, quan una població petita de la xarxa invariant d'escala és inicialment de tipus A, aquesta població és de nodes de grau menor que si fos una xarxa regular, i aquests nodes tenen menys capacitat per propagar-se, per la qual cosa s'extingeixen més ràpidament. També s'observa que, a part d'atènyer més ràpidament el consens, també s'assoleix més freqüentment (pel mateix raonament). Això ho confirma la mitjana, la qual és més baixa que en el cas de la xarxa regular.

A les simulacions amb $\rho = 30\%$ i $\rho = 50\%$, d'altra banda, no es pot apreciar gaire diferència entre la xarxa regular i la xarxa invariant d'escala. Això s'explica de manera similar que el paràgraf anterior: com que ja no es tracta d'una fracció d'inici petita sinó més gran de nodes de tipus A, en una xarxa invariant d'escala els nodes d'aquesta fracció inicial ara comprendran nodes de grau superior, la qual cosa farà que siguin capaços de propagar l'opinió i que no s'extingeixi tan ràpidament.

En resum, quan s'introdueix a l'atzar una nova opinió en una població, s'ha observat que les diferències en la topologia de les xarxes tenen un paper més important en l'evolució de l'opinió quan la fracció de nodes de tipus A inicial és petita. En canvi, quan la fracció inicial augmenta, la rellevància de la topologia de les xarxes disminueix.

³Aquests paràmetres han estat determinats en la secció anterior Simulació i estudi del model epidèmic; vegeu l'Equació 5.1.

⁴Per xarxes de mida prou gran, la fracció esperada de nodes de grau entre el grau mínim m i el grau mitjà $2m$ és de $3/4$.

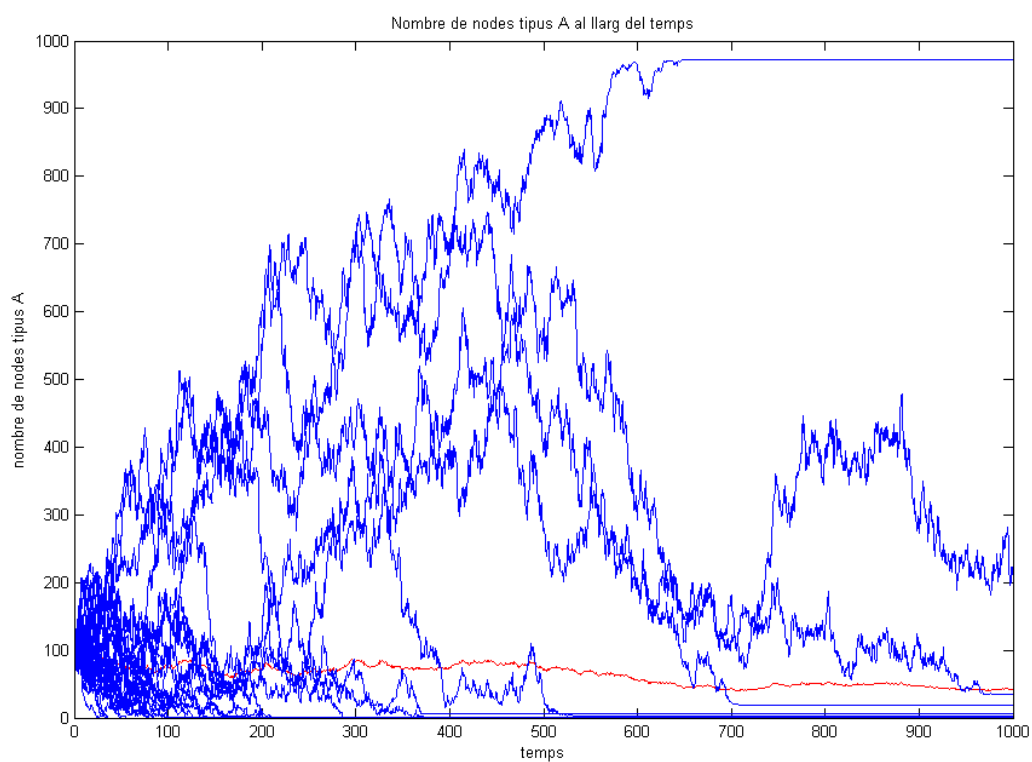


Figura 5.10: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10%. En vermell es mostra la mitjana de les execucions.

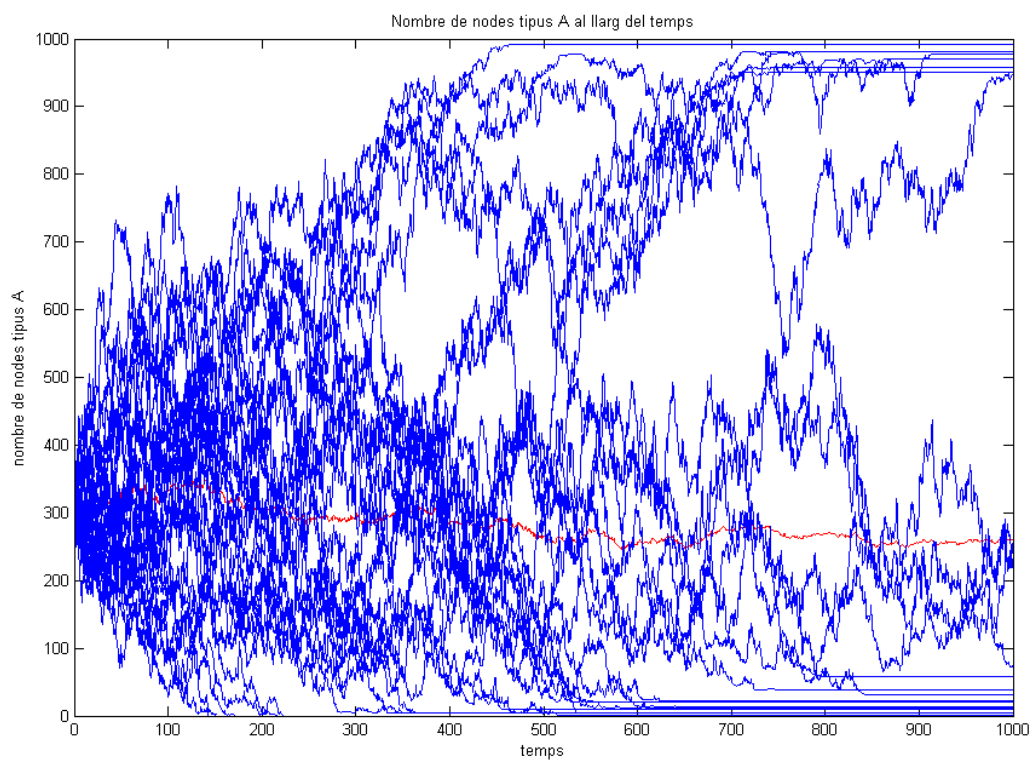


Figura 5.11: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30%. En vermell es mostra la mitjana de les execucions.

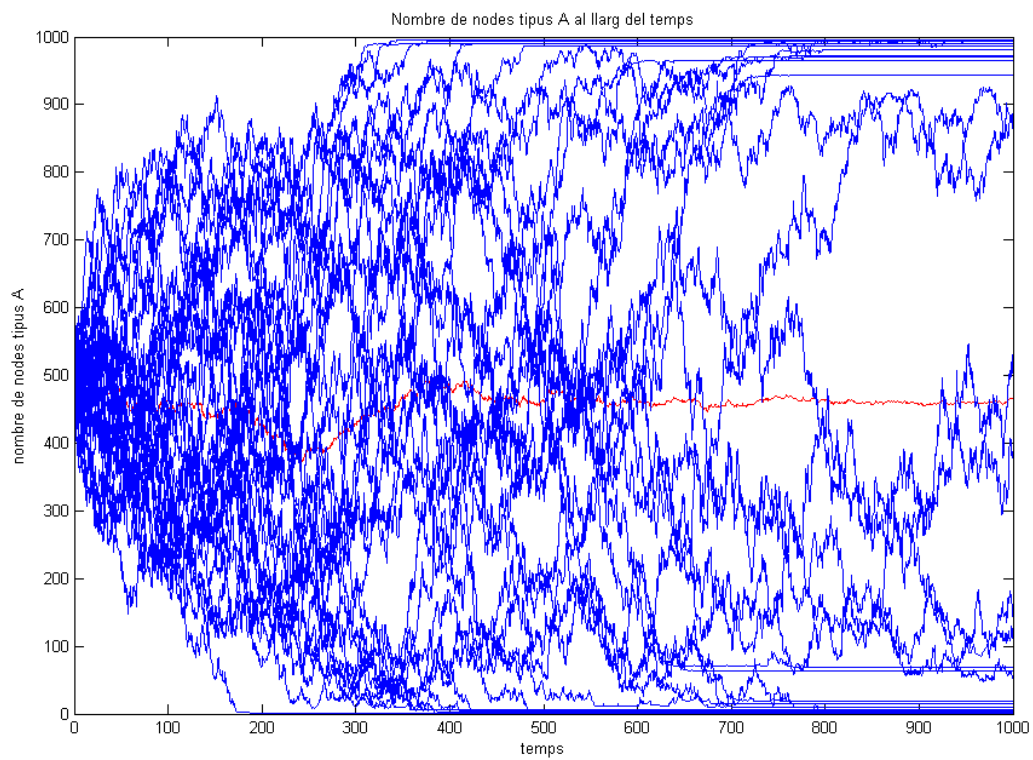


Figura 5.12: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 50%. En vermell es mostra la mitjana de les execucions.

5.3.3 Xarxa invariant d'escala amb distribució inicial intencionada

En la secció anterior s'ha observat que, quan la fracció inicial de nodes d'opinió A és petita, les diferències entre les xarxes regulars i les xarxes invariants d'escala són degudes als graus dels nodes que inicialment adopten l'opinió A. Per demostrar que aquest factor és realment determinant per l'evolució de l'opinió en la població, es proposa que la nova opinió s'introdueixi selectivament en els nodes de més grau de la xarxa invariant d'escala.

S'estableix la següent condició d'inici: que el 10% dels nodes de més grau de la xarxa (nodes amb més connexions) siguin de tipus (opinió) A. En cas que es vulgui un percentatge d'inici superior al 10%, la fracció restant s'establirà de manera aleatòria com fins ara. D'aquesta manera s'assegurarà que els nodes més influents de la xarxa siguin de tipus A. Això es podria assimilar al fet que, en una xarxa social, les persones amb més contactes (més influents) introduïssin una nova opinió diferent que la de la resta. En aquest cas, només es faran simulacions per percentatges inicials de $\rho = 10\%$ i $\rho = 30\%$, ja que en el cas de $\rho = 50\%$ no té gaire sentit començar amb una distribució inicial intencionada perquè hi ha la meitat de cada tipus. Les simulacions obtingudes es poden veure a la Figura 5.13 i a la Figura 5.14.

Si es fa la comparació de la Figura 5.13 (xarxa invariant d'escala, $\rho = 10\%$, distribució inicial intencionada) amb la Figura 5.10 (ídem però distribució inicial no intencionada) es pot veure clarament que en el primer cas, quan els nodes de més grau són del mateix tipus, el seu tipus s'expandeix moltes més vegades que no pas en l'altre cas, és a dir, s'arriba a un consens de nodes de tipus A més cops, i hi ha més fluctuacions: la dinàmica té un aspecte més transitori, i segurament si es deixés més temps les trajectòries s'acabarien fixant al consens de tipus A molts més cops que en l'altre cas. Això ho confirma la mitjana, que és més elevada en el primer cas. A més a més, també en el primer cas s'observa que la tendència cap al consens és molt més ràpida que en l'altre; això també és degut, segurament, a la presència d'aquests *hubs* que tenen una elevada influència en la xarxa.

D'altra banda, si es comparen la Figura 5.14 (xarxa invariant d'escala, $\rho = 30\%$, distribució inicial intencionada) amb la Figura 5.11 (ídem però distribució inicial no intencionada) es pot observar el mateix, encara que no de manera tan clara, ja que hi ha moltes més fluctuacions. En el cas de l'inici intencionat, la mitjana és més elevada que en el cas normal i el temps d'establiment sembla que és més ràpid.

Per intentar fer una comparació addicional de les dinàmiques en el cas d'inici intencionat i no intencionat, s'ha fet un anàlisi de diagrama de caixa cada cert interval de temps, el qual

es pot veure a la Figura 5.15 (per una $\rho = 10\%$) i a la Figura 5.16 (per una $\rho = 30\%$). Un diagrama de caixa mostra força més informació que la mitjana i és molt més representativa en el cas de les presents simulacions: la línia vermella horitzontal dintre de la caixa representa la mediana de les dades, és a dir, aquella dada que deixa per sota seu i per sobre seu el mateix nombre de dades, i els límits de la caixa són els quartils del 25% i del 75%. Si s'analitzen, per exemple, la Figura 5.15a i la Figura 5.15b, corresponents a la simulació amb $\rho = 10\%$ no intencionada i intencionada, respectivament, en el primer cas es pot veure com les caixes de seguida queden col·lapsades, la qual cosa significa que les execucions tendeixen a anar cap al consens de desaparició dels nodes de tipus A; en canvi, en el segon cas –l'intencional–, els diagrames de caixa demostren que els nodes de tipus A tendeixen a perpetuar-se més en el temps que en l'altre cas, ja que el tercer quartil és molt ample, encara que les medianes també de seguida cauen cap al zero, cosa que significa que moltes de les execucions acaben també en el consens de desaparició dels nodes A. Per l'altre costat, en el cas de la simulació per $\rho = 30\%$ no intencionada i intencionada (Figura 5.16a i Figura 5.16b), es pot observar exactament el mateix: en el primer cas, la mediana decau força ràpidament cap al zero i el segon quartil queda molt reduït, mentre que en el segon cas la mediana és superior, el segon quartil és una mica més ample i el tercer quartil ho és significativament molt més, especialment cap al final del temps establert, la qual cosa representa que aquesta simulació conté moltes trajectòries que no tendeixen a la desaparició dels nodes de tipus A.

S'ha vist, doncs, que quan s'introdueix una nova opinió en un àmbit on no existia anteriorment (és a dir, quan una petita fracció de nodes adopta aquesta nova opinió), les opinions d'aquells nodes (persones) que tenen més influència (connexions) són d'una gran importància a l'hora de determinar l'evolució de tal opinió. Això és especialment rellevant en les xarxes reals heterogènies (assimilables a una xarxa invariant d'escala) com, per exemple, una xarxa social.

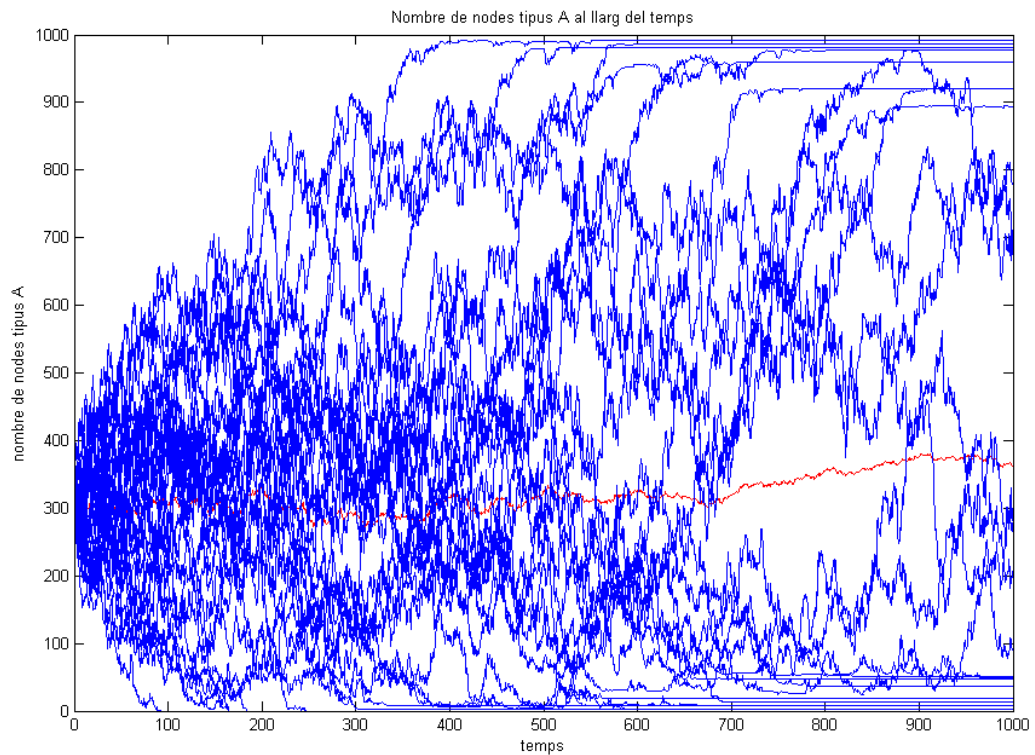


Figura 5.13: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 10% (distribució inicial intencionada al 10% dels nodes de més grau). En vermell es mostra la mitjana de les execucions.

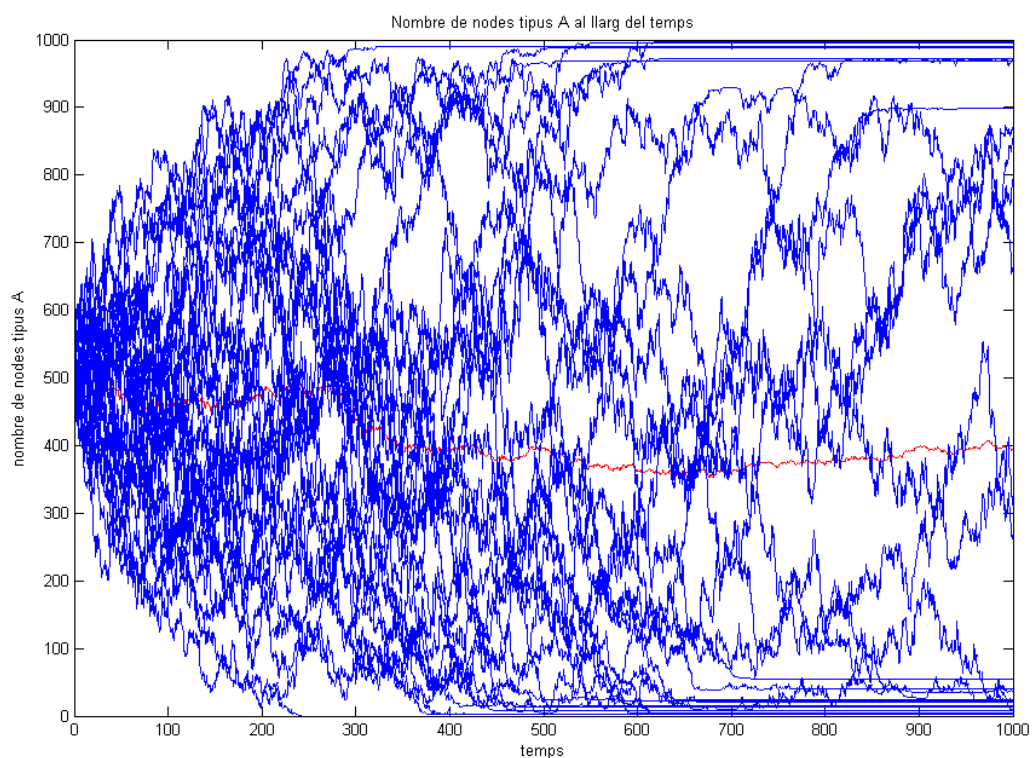


Figura 5.14: Gràfica de l'evolució de cadascuna de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0,1$ i amb un percentatge inicial de nodes tipus A del 30% (distribució inicial intencionada al 10% dels nodes de més grau). En vermell es mostra la mitjana de les execucions.

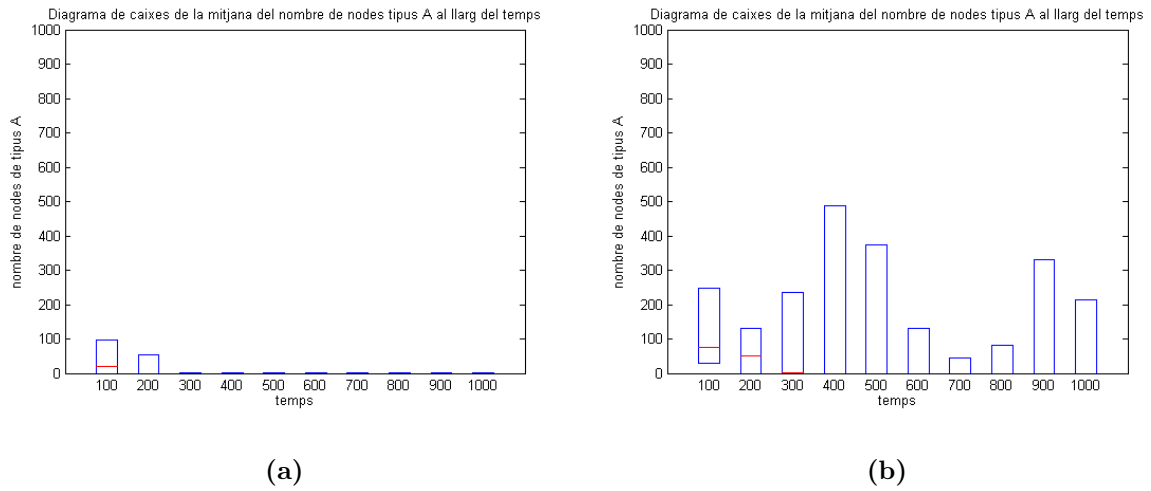


Figura 5.15: Diagrama de caixa en cada interval de $t = 100$ de la mitjana de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0, 1$ i amb un percentatge inicial de nodes tipus A del 10%. (a) Distribució inicial no intencionada (corresponent a la simulació de la Figura 5.10). (b) Distribució inicial intencionada (corresponent a la simulació de la Figura 5.13).

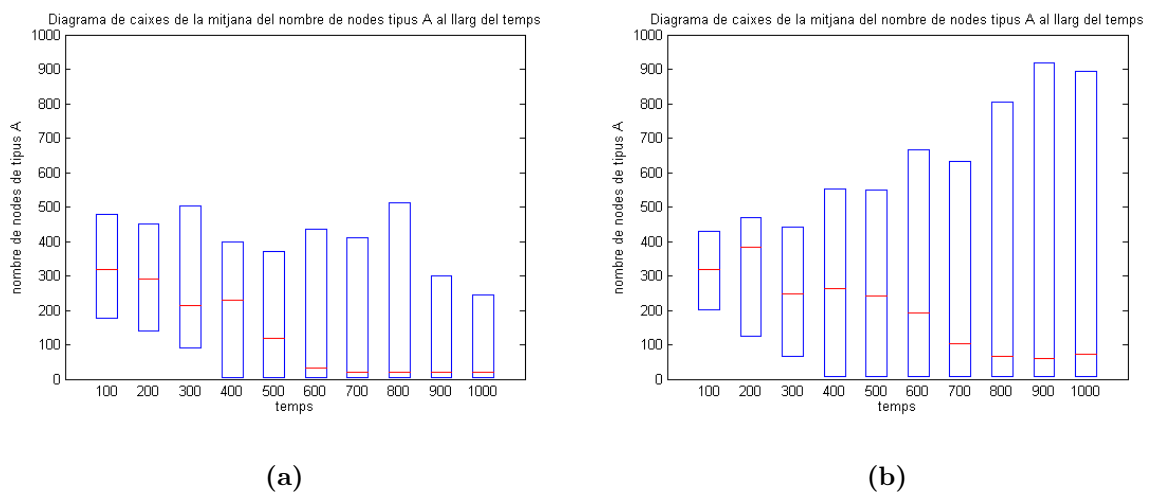


Figura 5.16: Diagrama de caixa en cada interval de $t = 100$ de la mitjana de les 30 execucions de la simulació d'una xarxa invariant d'escala de grau mitjà 10 amb reconexió de $p = 0, 1$ i amb un percentatge inicial de nodes tipus A del 30%. (a) Distribució inicial no intencionada (corresponent a la simulació de la Figura 5.11). (b) Distribució inicial intencionada (corresponent a la simulació de la Figura 5.14).

5.4 Perspectives de futur

El fet de disposar del simulador estocàstic en C++ deixa oberta la possibilitat de seguir explorant en el futur. En el present treball s'ha estudiat, en la secció Simulació i estudi de la dinàmica d'enllaços amb reconexió, l'evolució d'una xarxa invariant d'escala adaptativa, la qual és assimilable a una xarxa real com, per exemple, una xarxa social; la simulació es podria dur a terme, per exemple, amb més nodes, o se'n podrien fer més execucions per treure resultats més acurats. D'igual manera, es podrien explorar altres topologies de xarxa com per exemple xarxes amb una distribució de graus de Poisson o, directament, xarxes extremes de xarxes socials. Fins i tot es podria predir l'evolució d'una o múltiples opinions durant, per exemple, un període electoral.

Tanmateix, seria convenient, abans d'obrir noves vies, optimitzar el programa en C++ perquè faci les simulacions en menys temps, ja sigui canviant de sistema operatiu (per exemple, a Linux) o utilitzant una màquina més potent.

6. RESUM DE L'ESTUDI ECONÒMIC DE COSTOS

Segons l'estudi econòmic de costos, el cost total de l'elaboració d'aquest treball és de CINC MIL TRES-CENTS TRETZE EUROS amb CINQUANTA-NOU CÈNTIMS (5.313,59 €).

En referència a les tasques realitzades durant l'elaboració del present treball es presenta a continuació la valoració econòmica efectuada:

Codi	Descripció	Import
1	Estudi previ	- €
2	Desenvolupament i redacció del treball (87,5 h)	3.500,00 €
3	Amortització del maquinari	160,00 €
4	Amortització del programari	
4.1	Matlab 11	142,86 €
4.2	Microsoft Visual Studio 2013	- €
4.3	Minitab 9	83,33 €
	Total	3.886,19 €
	Costos indirectes (13%)	505,20 €
	Subtotal	4.391,40 €
	21% IVA	922,19 €
	TOTAL	5.313,59 €

7. CONCLUSIONS

En primer lloc, el present treball ha servit per comprendre els trets fonamentals de les xarxes complexes, des de les seves quantitats de mesura fins la seva topologia, la qual és un dels trets principals quan es tracta de determinar la dinàmica de les xarxes. En segon lloc, també s'ha estudiat diferents tipus de dinàmica de xarxes, començant pel model epidèmic, passant pels diferents models de propagació d'opinió i acabant per les xarxes adaptatives. Finalment, l'etapa d'estudi ha conclòs amb l'anàlisi i comprensió de l'algorisme de Gillespie, el qual és una eina útil per simular trajectòries estocàstiques de xarxes, les quals altrament serien difícils de determinar a partir de l'anàlisi matemàtica d'equacions diferencials.

La programació del simulador en llenguatge C++ ha estat necessària com a prèvia per poder dur a terme les diferents simulacions objectiu del present treball. Aquest llenguatge d'alt nivell té totes les prestacions necessàries per implementar les diferents funcions que calen per l'algorisme de Gillespie i la resta de passos necessaris per simular tant el model epidèmic com la dinàmica d'una xarxa adaptativa, encara que ha estat una mica limitat pel que fa al seu rendiment (les simulacions tardaven molta estona).

Pel que fa a les simulacions i l'anàlisi dels seus resultats s'han contemplat tres casos diferents:

1. Simulació i estudi del model epidèmic: s'ha comprovat que el resultat de la simulació estocàstica del model epidèmic amb uns paràmetres determinats s'ajustava a l'evolució predita per les equacions diferencials del model, en alguns casos (xarxa regular, xarxa d'Erdős-Rényi) millor que d'altres (xarxa invariant d'escala).
2. Simulació i estudi de la dinàmica d'enllaços sense reconexió: s'ha comprovat que la probabilitat d'èxit dels elements de la xarxa en funció del nombre inicial d'elements d'un cert tipus predita per la literatura existent sobre la dinàmica de les xarxes complexes s'ajustava als resultats obtinguts amb el simulador programat en el present treball. Això s'ha vist fent una regressió lineal i també duent a terme el corresponent contrast estadístic. Els resultats teòrics i pràctics s'han ajustat de la manera esperada. S'ha conclòs que la topologia de la xarxa no té un paper important pel que fa a les probabilitats d'èxit.
3. Simulació i estudi de la dinàmica d'enllaços amb reconexió: en aquest cas, com que no existeix literatura publicada sobre el tema, els anàlisis de resultats han estat més qualitius que no pas quantitius. S'han simulat processos sobre dues topologies de xarxa diferents, la xarxa regular i la xarxa invariant d'escala, i s'ha observat que quan s'introdueix a l'atzar una nova opinió en una població, les diferències en la topologia de les xarxa té un paper

més important en l'evolució de l'opinió quan la fracció de nodes d'opinió A inicial és petita. En canvi, quan la fracció inicial augmenta, la importància de la topologia de les xarxes disminueix. Per demostrar que aquest factor és realment determinant per l'evolució de l'opinió en la població, s'ha proposat que la nova opinió s'introdueix selectivament en els nodes de més grau de la xarxa invariant d'escala, i s'ha vist que quan s'introdueix una nova opinió en un àmbit on no existia anteriorment (és a dir, quan una petita fracció de nodes adopta aquesta nova opinió), les opinions d'aquells nodes (persones) que tenen més influència (connexions) són d'una gran importància a l'hora de determinar la seva evolució. Això és especialment rellevant en les xarxes reals heterogènies (assimilables a una xarxa invariant d'escala) com, per exemple, una xarxa social, en la qual alguns nodes (persones) tenen moltes connexions i, per tant, molta influència sobre les opinions dels seus veïns.

8. RELACIÓ DE DOCUMENTS

1. Memòria
2. Codi informàtic (en suport CD)¹
3. Resultats de les simulacions (en suport CD)

¹El codi també es pot trobar a l'adreça web http://www.arnauduran.net/codi_tfc en llicència CC-BY-SA.

9. AGRAÏMENTS

M'agradaria dedicar aquestes línies als meus dos tutors del Treball final de carrera, els Drs. Joan Saldaña i David Juher, que des del dia que van acceptar dur-me el treball hi han dedicat moltes hores per resoldre'm dubtes, respondre'm qüestions, compartir opinions i aportar-hi idees per enriquir-lo. Sense la seva vàlua i paciència aquest treball no hauria estat possible.

10. Bibliografia

- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barrat, A., Barthélemy, M., and Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, UK; New York.
- Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, page 17–61.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29(4):251–262.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.
- Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55.
- Jamakovic, A. and Uhlig, S. (2007). Influence of the network structure on robustness. In *in Networks, 2007. ICON 2007. 15th IEEE International Conference*, page 278–283.
- Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20). arXiv:cond-mat/0205405.
- Pastor-Satorras, R., Vázquez, A., and Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87(25):258701.
- Sood, V., Antal, T., and Redner, S. (2008). Voter models on heterogeneous networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 77(4 Pt 1):041121.
- Wang, X. F. (2002). Complex networks: topology, dynamics and synchronization. *International Journal of Bifurcation and Chaos*, 12(05):885–916.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442.