Analisi del Contenimento delle Epidemie nella Rete Internet e nei Grafi scale-free

Giulio Bortot Matteo Stoisa Simone Soncin

June 20, 2018

Contents

1

Introduzione

Le reti sono una parte fondamentale della nostra società poichè una vasta varietà di dispositivi permette l'accesso a internet in molteplici forme e scopi: si spazia dagli orologi digitali ai dispositivi militari.

Questa eterogeneità, contemporaneamente punto di forza e tallone d'achille di tale diffusione, è vulnerabile a svariate tipologie di attacchi, che risultano quindi molto difficili da prevenire e che possono causare ingenti disagi in molteplici ambiti.

Tra le strategie di incursione nelle reti, spicca per pericolosità quella che prevede l'espansione autonoma dell'infezione tramite i collegamenti capillari tipicamente presenti in rete. Questo pericolo prende forma in programmi chiamati "worm", capaci di diffondersi ed annidarsi nei dispositivi connessi, anche senza sintomi apparenti.

Attualmente il metodo di prevenzione più diffuso è l'antivirus. Questo strumento confronta il codice di ogni elemento presente sul sistema sotto analisi con un database di virus conosciuti, al fine di riconoscere possibili minacce per la salute dell'host. Tuttavia, un worm non presente nel database non può ovviamente essere riconosciuto come tale. Per questo motivo si presenta la necessità di una strategia da adottare in una fase antecedente alla registrazione di una nuova forma virale.

Al fine di proporre una soluzione a questo problema si analizzeranno diverse strategie atte a limitare l'espansione della minaccia, lasciando ad altri strumenti il riconoscimento e la neutralizzazione della stessa.

Il presente studio riprende quello svolto da Thanh Dang Nguyen, Francois Bonnet e Xavier Defago chiamato "Mitigating the Spread of a Virus in the Internet", il quale analizza alcuni di questi metodi di contenimento. Sfrut-

tando questi concetti sono state inoltre applicate le stesse strategie ad una rete *scale free* generica, utilizzando quindi i risultati esposti nell'articolo di ispirazione come metro di confronto per valutarne l'efficacia.

2

Modello e Definizioni

Nello studio "Mitigating the Spread of a Virus in the Internet" si utilizza un modello sul quale viene simulata l'infezione di una rete al fine di applicarvi diverse strategie di contenimento.

Il modello in questione è un grafo connesso e non orientato $G(V, E)$, dove V indica il numero di vertici ed E il numero di archi.

Le simulazioni sfruttano una versione discretizzata del tempo per cadenzare l'evolversi dell'epidemia. Ad ogni step, lo stato ogni nodo determinerà il comportamento adottato da esso a seconda della strategia di contenimento scelta.

2.1 Stato dei nodi

I nodi possono assumere solo uno dei seguenti stati:

Stato	Azione
Infetto	Un nodo infetto è già stato compromesso dal virus e proverà ad infettare tutti i nodi che lo circondano
Disattivato	Un nodo disattivato, al fine di non farsi infettare dal virus ed arginarne l'espansione, interrompe ogni comunicazione con i nodi circostanti
Sano	Un nodo sano viene definito tale se non assume nessuno degli stati precedentemente descritti

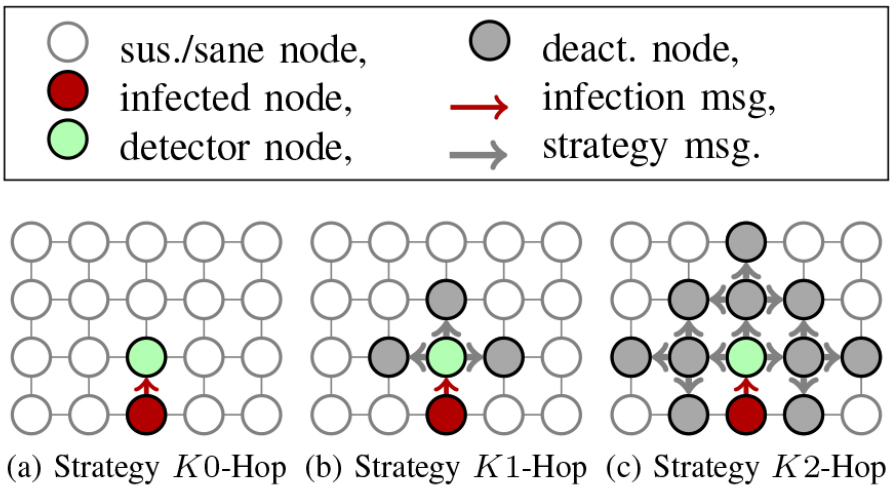


Figure 2.1: rappresentazione delle strategie K-hop

2.2 Stato degli archi

Gli archi si possono trovare solamente in due stati, attivo o tagliato. Un arco si dice attivo quando permette la comunicazione, tagliato in caso contrario.

2.3 Propagazione del virus

Nell'istante iniziale tutti i nodi sono sani tranne uno, che è infetto e designato come origine dell'epidemia. Nel susseguirsi degli step temporali ogni nodo compromesso tenterà di infettare tutti quelli a lui connessi: questo evento ha una probabilità p di andare a buon fine. Nel caso in cui un vertice non riesca ad infettare un suo vicino, questo si accorgerà del tentativo ed attuerà una delle strategie difensive esposte in seguito.

2.4 Strategie di Contenimento

Vengono ora esposte le differenti strategie di contenimento adottabili, definiamo Detector un nodo che rileva l'attacco. Vi sono due tipologie di immunizzazione:

- Strategia di taglio

  La strategia di taglio prevede che il Detector tagli l'arco di collegamento appena riconosciuto il tentativo di attacco.

- Strategia di uccisione

  Questa strategia prevede che il Detector assuma come misura di contenimento la disattivazione di se stesso o eventualmente di più nodi. A seconda di varie versioni di questa strategia, esso può mandare ai nodi adiacenti un messaggio di allarme che li invita a disattivarsi. Un nodo che riceve il messaggio d'allarme può inoltrarlo a sua volta sui propri collegamenti ma, naturalmente, i nodi già infettati o disattivati ignoreranno qualsiasi comunicazione.

La seguente tabella riassume le quattro differenti strategie che verranno simulate.

Sigla	Descrizione
C0	Il Detector cambia in *tagliato* lo stato dell'arco di collegamento al nodo che ha tentato di infettarlo
K0-Hop	Il Detector cambia il proprio stato in *disattivato*
K1-Hop	Il Detector cambia il proprio stato in *disattivato* ed invia a tutti i nodi adiacenti un ordine di disattivazione
K2-Hop	Il Detector cambia il proprio stato in *disattivato* ed invia a tutti i nodi adiacenti un ordine di disattivazione, questi ultimi inoltrano a loro volta l'ordine ai propri vicini

L'ultima strategia potrebbe essere generalizzata a Kn-Hop nel caso di inoltro n-esimo dell'ordine.

3

Simulazione ed Analisi

3.1 Caratteristiche del Modello

Il modello è stato creato con i dati forniti dalla CAIDA Project, che svolge periodicamente un censimento di tutti i router ipv4, i quali verranno utilizzati come vertici del grafo descritto in precedenza. La topologia presenta circa 3 milioni di nodi totali, di questi:

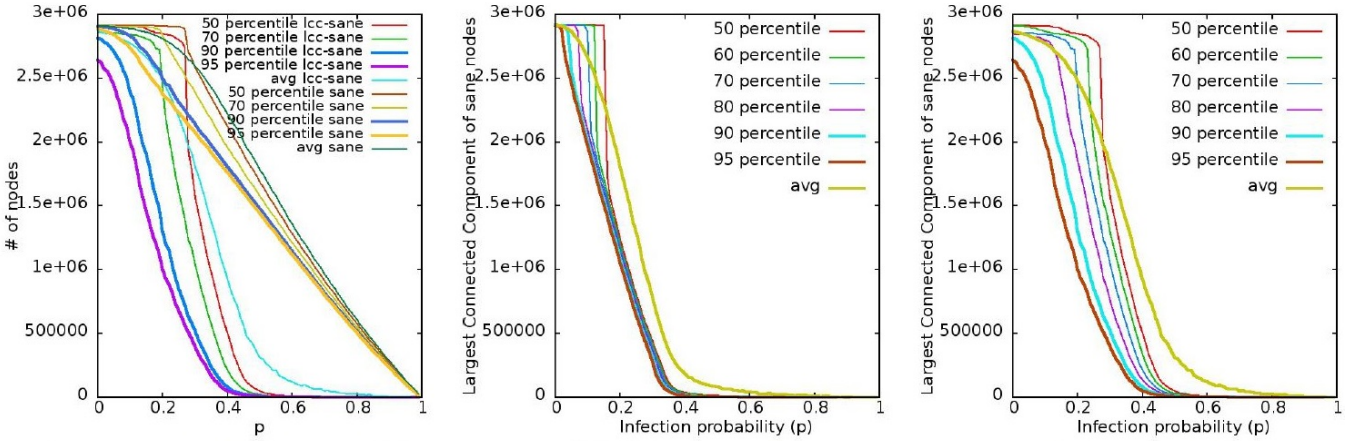
- 55 mila hanno grado maggiore di 50, in seguito chiamati *"high-degree"*
- 2 milioni hanno grado minore di 3, in seguito chiamati *"low-degree"*
  - di cui 1.5 milioni hanno grado pari a 1

Alla luce di questi dati, si è deciso di simulare una quantità di infezioni pari al numero di variazioni della p (da 0 a 1 con incrementi di 0.01), differenziando ulteriormente in base al grado del nodo dal quale viene fatta partire l'epidemia (appartente ai *low-degree* o *high-degree*).

3.2 Risultati delle Simulazioni

I grafici in figura 3.1 riportano la dimensione del più grande sotto-grafo di nodi sani e completamente connessi (LCC) al termine delle simulazioni.

Il numero di nodi sani è notevolmente maggiore della dimensione degli LCC, lasciando dedurre che a fine simulazione siano molti i sopravvissuti rimasti isolati.



(a) # of sane nodes and LCC-sane of $K1$-Hop (b) 'LCC-sane' of $K0$-Hop (c) 'LCC-sane' of $K1$-Hop

Figure 3.1: grafici simulazioni da high degree

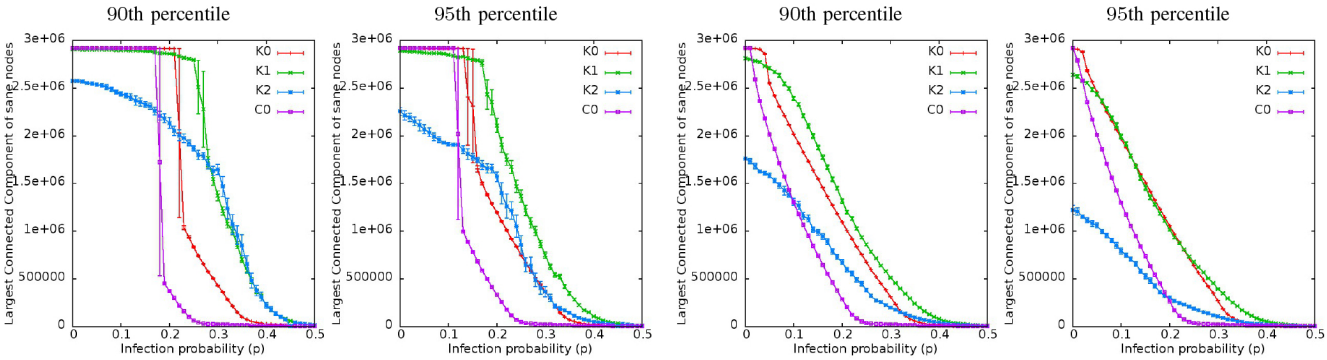
Questo è chiaramente dovuto alla topologia del grafo, il quale, come già descritto, è formato da pochi *"hub"* connessi a molte *"foglie"*. Conseguentemente infettare con successo un *hub* porta ad una grande quantità di nodi che, pur rimanendo sani, si vedono privati del collegamento al resto della rete. Lo stesso risultato si ottiene nel caso in cui, anche senza essere stato contagiato, uno degli hub venga disattivato dalle strategie K-hop. Per il medesimo principio si osservano risultati caratterizzati da una grande varianza, per questo si può notare sul grafico come sia stata affinata la curva utilizzando diversi valori percentili.

Questa particolare struttura viene dunque studiata comparando i risultati delle quattro strategie e diffierenziandoli in base al grado del nodo scelto come punto di partenza. Si ottengono i grafici in figura 3.2.

Si può notare la presenza di una soglia che determina un repentino diminuire della dimensione degli LCC sani, caratteristica meno evidente quando si parte da nodi *high-degree*.

Inoltre osservando l'andamento delle curve **C0** è possibile dedurre l'inefficacia di questa strategia, in quanto la soglia citata precedentemente si attesta a valori di p molto bassi.

Tuttavia l'approccio diametralmente opposto **K2**, pur riuscendo a smorzare la rapida diminuzione presente in **C0**, ha il consistente svantaggio di disattivare un gran numero di nodi anche a p prossime allo 0, specialmente nei casi corrispondenti agli *high-degree*.



(a) Infection starts with a 'low-degree' node

(b) Infection starts with a 'high-degree' node

Figure 3.2: infezioni con filtraggio dei risultati tramite percentile

Per p basse la strategia **K0** si rivela dunque di gran lunga migliore di **C0** e leggermente più performante rispetto a **K1**.

4

Studio su Grafo Generico Scale Free

4.1 Obiettivo

Lo studio dell'articolo ha ispirato il tentativo di rieffettuare le simulazioni per applicare ed analizzare le strategie di contenimento a grafi di topologia generica e dimensioni contenute.

La topologia della rete Internet globale, infatti, risulta di difficile gestione a causa dell'enorme numero di nodi ed archi che rendono proibitivi i mezzi computazionali richiesti, oltre alla difficile reperibilità del grafo stesso.

Si è dunque deciso di applicare le stesse simulazioni ad un grafo *scale free*, in quanto esso può modellizzare, oltre a Internet, reti di innumerevoli altri ambiti quali quello economico, sociale o il WorldWideWeb. L'obiettivo è dunque quello di offrire un'analisi più generica delle strategie di contenimento, applicandole ad un modello meno specifico.

4.2 Modello

Tramite i linguaggi C++ e Python è stato implementato un algoritmo in grado di simulare, a partire da un qualsiasi grafo, la diffusione dell'epidemia ed il suo contenimento secondo le quattro strategie precedentemente descritte.

Il compromesso tra dimensione del modello e possibilità computazionali si è attestato su un grafo avente 3000 nodi, creato a partire dall'algoritmo

Barabàsi-Albert, modificato al fine di ottenere delle caratteristiche di grado simili a quelle della rete internet, pur mantenendo una topologia generica.

In particolare il grafo creato possiede:

- l'80% dei nodi con grado minore di 3 di cui
  - il 35% di grado 1
- il 20% di nodi con grado maggiore di 3 di cui
  - l'1,5% ha grado maggiore di 15

Si ritiene che queste caratteristiche lo rendano accomunabile, seppur in scala, alla rete Internet composta di pochi hub con moltissimi collegamenti e una maggioranza di nodi foglia.

L'algoritmo prevede l'esecuzione di ogni simulazione per ogni possibile probabilità di trasmissione, da nulla (0) a certa (100), e per ognuna delle quattro strategie adottabili.

Per ottenere dati di validità statistica è stato iterato il processo per dieci volte in ogni casistica, i risultati mostrati in seguito non tengono conto del 20% dei dati che più si discostano dalla media. Le simulazioni sono state effettuate a partire da nodi di grado 2 (*low-degree*) e nodi di grado superiore a 30 (*high-degree*), scelti ogni volta casualmente per eliminare il maggior numero di *corner case*.

4.3 Simulazione

Nonostante il presente articolo si ponga come una rivisitazione di "Mitigating the Spread of a Virus in the Internet", l'utilizzo di un modello diverso ha ispirato un'analisi di punti di vista ritenuti secondari da quello originale, divantandone a questo punto una versione espansa. Di seguito verranno quindi riportati non solo grafici sulla falsa riga della versione di partenza, ma anche rappresentazioni di aspetti apprezzabili in altri campi, purchè modellizzabili da grafi scale free.

In primo luogo si è ritenuto interessante analizzare gli andamenti degli istanti temporali in cui l'epidemia si arresta, a partire da nodi *low-degree* o *high-degree*. Indipendentemente dalla completa o parziale riuscita della

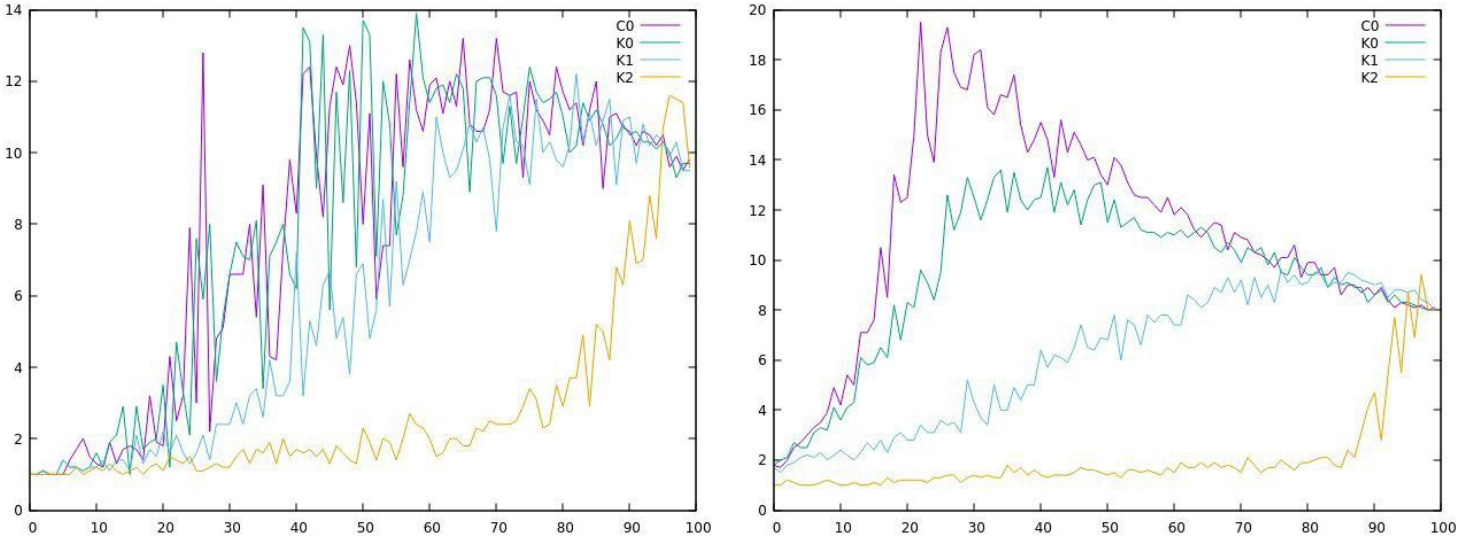


Figure 4.1: step grado 2 (sx) e grado 30 (dx)

strategia di contenimento, infatti, essa potrebbe avere semplicemente il compito di rallentare l'avanzata dell'epidemia permettendo intanto lo svilupparsi di altre contromisure.

Nei grafici riportanti questo aspetto sono state rappresentate le varie p in rapporto al numero medio di step necessari al completo isolamento del virus. Come prevedibile, l'utilizzo della strategia **K2** porta ad una situazione di stallo più rapidamente delle altre, vista la sua aggressività rispetto al numero di nodi coinvolti.

Si può inoltre notare come le simulazioni *low-degree*, sebbene presentino un andamento poco consistente, riportino dei valori massimi sull'asse degli step tendenzialmente inferiori alla controparte negli *high-degree*.

I seguenti grafici riportano l'andamento della somma tra nodi sani e nodi disattivati. Questa statistica potrebbe rivelarsi utile nel caso di applicazioni in cui la disattivazione dei nodi non sia dannosa.

Si può notare come le simulazioni *low-degree* e *high-degree* in questa statistica non si differenzino particolarmente in andamento e valori. Analizzando inoltre il dato complementare a quello rappresentato, ossia il numero di nodi infettati con successo dal virus, si può ritrovare la stessa analogia fra medesima strategia ma grado di partenza differente. L'unica eccezione osservabile

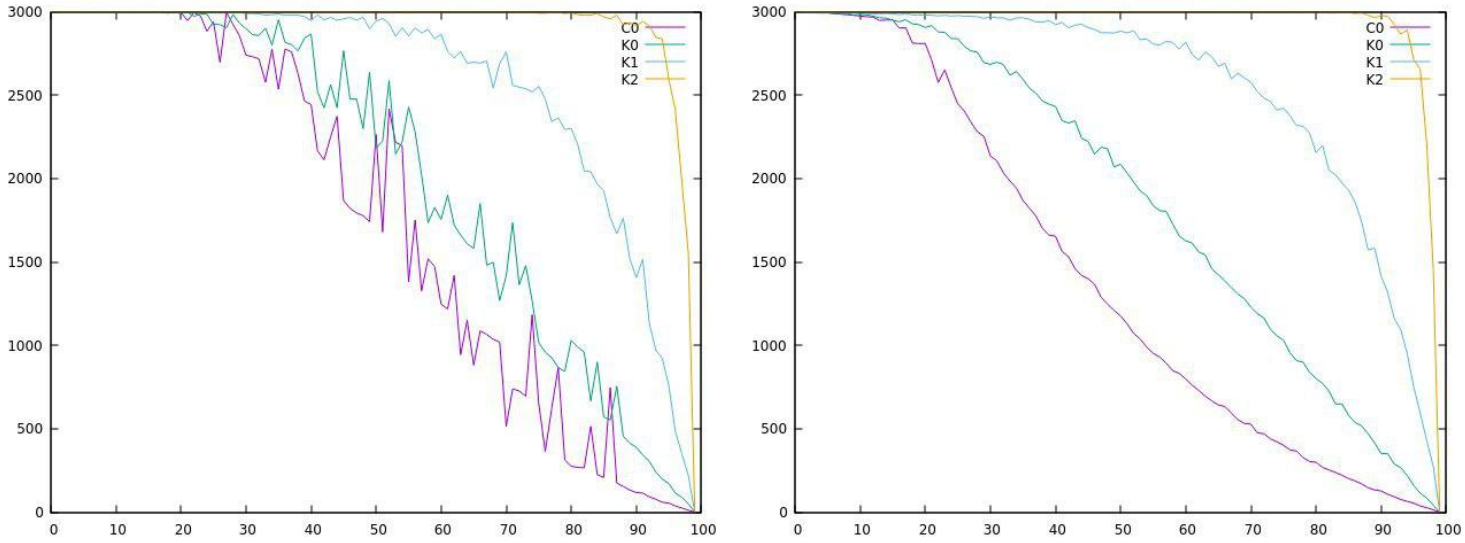


Figure 4.2: nodi disattivati grado 2 (sx) e grado 30 (dx)

rimane dunque la maggior varianza dei risultati dei nodi *low-degree*, probabilmente dovuta al non elevato numero di simulazioni ridondanti, unita alla maggior concentrazione di *corner case*.

Questi dati evidenziano come le strategie siano valutabili in efficacia indipendentemente dalla topologia della rete. Infatti la somiglianza fra le due classi di simulazioni spinge a liberarsi della distinzione, espandendo ulteriormente la validità di questi studi quando applicati a reti generiche. Si ricorda che il fattore discriminante risiede nella interconnessione tra i nodi sani, quindi, i modelli accomunabili alla rete internet dovranno comunque fare affidamento a statistiche basate sulla grandezza degli LCC, invalidando queste osservazioni.

Riportiamo infine i grafici dello studio analogo al conteggio del grafo LCC di nodi sani (4.3).

Appare subito evidente che i risultati derivanti dalle simulazioni *low-degree* hanno poca validità statistica in quanto non seguono andamenti ben definiti; anche in questo caso sembra non sufficiente il numero di simulazioni svolte. La casistica *high-degree* invece mostra andamenti analoghi a quelli dell'articolo. Vengono quindi riportati esclusivamente allo scopo di verificare la parziale analogia tra il modello e la rete originale.

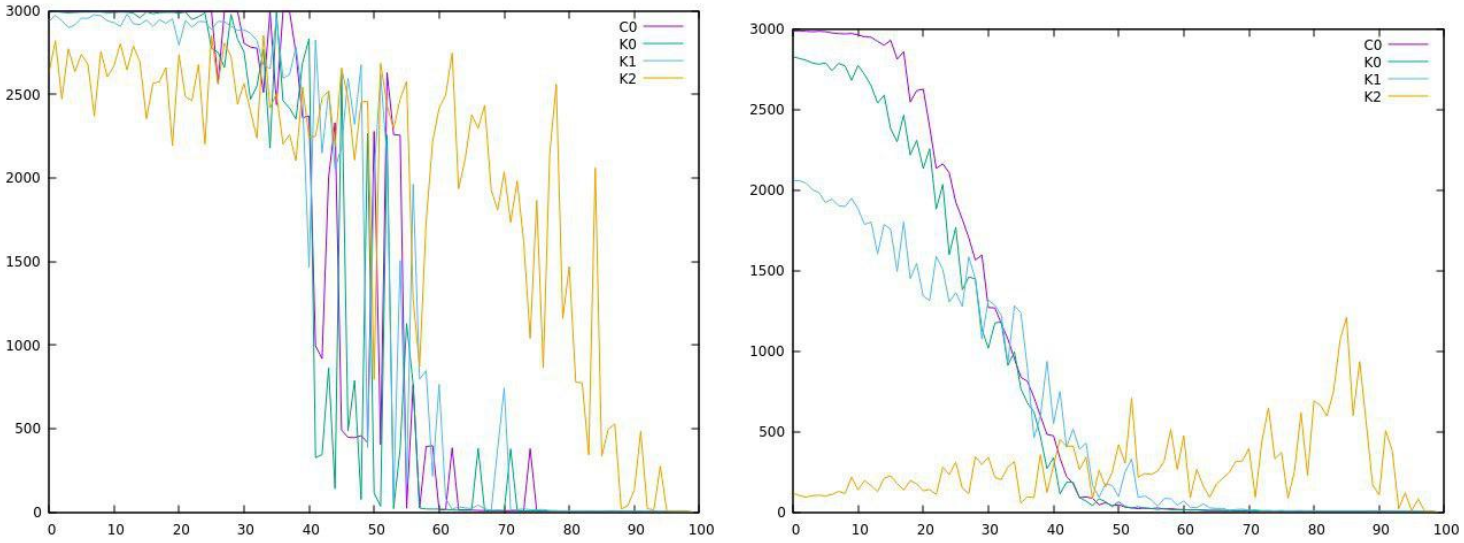


Figure 4.3: grandezza LCC partendo da grado 2 (sx) e da grado 30 (dx)

5

Conclusioni

Partendo dalle conclusioni di "Mitigating the Spread of a Virus in the Internet", riportiamo come la soglia citata nell'analisi dei grafici dell'articolo possa essere traslata a seconda della strategia utilizzata. Questo aspetto viene dunque utilizzato per valutarne l'efficacia, tuttavia nella versione generalizzata questo potrebbe non essere un fattore determinante.

Secondo questo principio viene infatti identificata come migliore la strategia **K1**, ma, come già accennato in precedenza, a seconda della natura delle reti potrebbe essere più importante valutare la rapidità di isolamento, nel qual caso **K2** spicca senza ombra di dubbio.

L'articolo di riferimento pone inoltre l'accento sull'inefficacia della strategia **C0**, giungendo alla conclusione che disattivare i nodi sia preferibile all'eliminazione dei collegamenti fra essi. Si noti come lo stato attuale dei sistemi di sicurezza preferisca invece una strategia di "cutting", implementata tramite le "blacklist", scelta apperentemente sconveniente anche secondo la versione espansa di questo studio. Tuttavia il risultato raggiunto dalle blacklist risulta meno invasivo e più facilmente realizzabile, mentre le strategie di disattivazione potrebbero provocare disagi tali da doverne giustificare l'utilizzo solamente in casi critici.

Riteniamo infine che, tenendo bene a mente la natura della rete, sia possibile sfruttare queste strategie in molteplici campi, scegliendo la più appropriata secondo le seguenti discriminanti:

- probabilità di contagio
- necessità di mantenere l'interconnessione fra i nodi

- tempo a disposizione per rispondere all'epidemia
- disagio provocato dalla disattivazione dei nodi

Dove le prime due sono state valutate da "Mitigating the Spread of a Virus in the Internet", mentre le ultime necessitano di studi più approfonditi date le limitazioni che abbiamo incontrato analizzandole.

5.1 Fonti

- "Mitigating the Spread of a Virus in the Internet", Thanh Dang Nguyen, François Bonnet and Xavier Dèfago (School of Information Science, Japan Advanced Institute of Science and Technology (JAIST))