Simulação escalável de epidemias em redes baseadas em passeios aleatórios

João Vitor B. Tavares[1], **Giulio Iacobelli**[1], **Daniel R. Figueiredo**[1]

[1]Programa de Engenharia de Sistemas e Computacão (PESC) / COPPE
Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ

{jtavares,giulio,daniel}@land.ufrj.br

Abstract. *Understanding how networks and dynamic processes relate is a important topic of research, and particularly in the context of epidemics spreading through networks. In this work, we consider the scenario where individuals move through a network. Contagion occurs when two (or more) individuals are in the same location (vertex). The aim of this work is to build a scalable simulator to this type of epidemic that can be used to analyze the behavior of epidemics in varios scenarios. Analytical and empirical results indicate that the simulator is scalable in the network size, simulation time and number of individuals.*

Resumo. *Entender como redes e processos dinâmicos se relacionam é um tema central de pesquisa, e em particular no contexto de epidemias que se desdobram sobre redes. Neste trabalho, consideramos o cenário onde indivíduos se movimentam por uma rede, que representa a estrutura do espaço de movimentação. O contágio pode ocorrer quando dois (ou mais) indivíduos se encontram em um mesmo local (vértice). O objetivo deste trabalho é projetar e implementar um simulador eficiente que possa ser utilizado para caracterizar o comportamento deste tipo epidemia em diferente cenários. Uma avaliação teórica e empírica indica que o simulador é escalável no tamanho da rede, tempo de simulação e número de indivíduos.*

1. Introdução

Ao longo de toda sua existência, a humanidade convive com epidemias de diversos tipos, e mais recentemente, com cascatas de informação em mídias sociais que podem ser representadas por epidemias.

Modelos epidêmicos em geral atribuem estados epidêmicos aos indivíduos, sendo os mais comuns: *Suscetível* (indivíduos que não estão contagiados mas que ainda podem ser contagiados), *Contraído* (indivíduos que foram contagiados mas ainda não são capazes de transmitir a epidemia), e *Infectado* (indivíduos contagiados que são capazes de propagar a epidemia).

Modelos mais recentes de epidemia em redes mais consideram a rede como sendo uma estrutura de locais sobre a qual indivíduos se movimentam [Draief and Ganesh 2011, Zhou and Liu 2009]. Neste cenário, cada vértice da rede representa um local e cada aresta indica a possibilidade de locomoção direta entre dois locais. Indivíduos possuem um estado epidêmico e o contágio pode ocorrer quando dois (ou mais) indivíduos se encontram em um mesmo local. Este trabalho estuda este tipo de modelo de epidemia em redes.

Neste trabalho projetamos e implementamos um simulador eficiente de eventos discretos para simular epidemias SCIS em redes com indivíduos se movimentando. Em particular, uma análise teórica e experimental mostra que, para um tempo fixo de simulação T, o tempo de execução não depende do tamanho da rede, que o tempo de execução cresce linearmente com tempo de simulação T, e que o tempo de execução cresce como $\mathcal{O}(K \log K)$ com o número K de indivíduos na rede.

2. Trabalhos Relacionados

Existe um grande número de modelos epidêmicos na literatura, incluindo muitos baseados em redes [Zhou and Liu 2009, Draief and Ganesh 2011, Ganesh et al. 2005, Pastor-Satorras and Vespignani 2001], além de alguns simuladores de epidemia em redes.

Barret et al. propõem um simulador escalável e paralelo de epidemias em redes, *EpiSimdemics* [Barrett et al. 2008]. Uma rede de contatos representada por um grafo bipartido que varia no tempo é avaliado, onde uma classe de vértices representa as pessoas e a outra representa as localidades e uma aresta é adicionada entre os vértices caso uma pessoa decida visitar uma localidade. Uma diferença com o presente artigo é que a movimentação dos indivíduos está restrita a rede, podendo estes se movimentarem apenas para vértices vizinhos ao vértice em que se encontram.

Alguns outros trabalham analisam a propagação de epidemia utilizando indivíduos que se movimentam em uma rede [Zhou and Liu 2009, Draief and Ganesh 2011]. Entretanto estes reportam resultados mais teóricos e assintóticos (quando o número de vértices da rede vai ao infinito), enquanto estamos interessados em resultados para redes de tamanho finito, com um número finito de indivíduos.

3. Modelo Epidêmico

Iremos considerar um modelo de epidemia em redes diferente da abordagem tradicional, onde indivíduos possuem mobilidade. Para isso, definimos $G(V, E)$, onde V é o conjunto de vértices que representam as localidades por onde os indivíduos se movimentam e E é o conjunto de arestas representando a possibilidade de mobilidade direta entre as localidades. Definimos também K como o número de indivíduos que se movimentam pela rede de acordo com passeios aleatórios independentes. O tempo que um indivíduo permanece em um vértice antes de se locomover para um vizinho é dado por uma variável aleatória exponencial com parâmetro λ. Desta forma, ao entrar em um vértice, o indivíduo permanece um tempo aleatório até se movimentar para um vértice vizinho.

A todo instante, cada indivíduo se encontra em um único estado epidêmico, podendo ser Suscetível, Contraído ou Infectado. Além disso, iremos adotar o modelo SCIS que define as transições entre os estados epidêmicos. A medida que os indivíduos se movimentam pela rede, ocorrem encontros entre indivíduos infectados e suscetíveis. Em cada encontro deste tipo o contágio ocorre de acordo com uma variável aleatória de Bernoulli de parâmetro p_u, onde u é um vértice da rede. Ou seja, com probabilidade p_u o indivíduo susceptível é contagiado pelo indivíduo infectado no instante do encontro. A associação da probabilidade de transmissão p_u ao vértice permite modelar localidades da rede com diferentes taxas de infecção.

Os indivíduos permanecem no estado suscetível até que sejam infectados e quando isso ocorre, eles passam imediatamente para o estado contraído. O tempo de permanência

no estado contraído possui distribuição exponencial com parâmetro τ. Ao final deste tempo o indivíduo muda para o estado Infectado. Uma vez infectado, o indivíduo permanece neste estado por um tempo aleatório com distribuição exponencial com parâmetro γ.

4. Simulador

Para estudar o modelo de epidemia em redes proposto, projetamos e implementamos um simulador baseado em eventos discretos. Os eventos do simulador são $Caminhar$ (C), $Infectar$ (I) e $Recuperar$ (R), sendo cada evento aplicado a um indivíduo em um determinado instante de tempo. Ou seja, cada evento é representado por um tipo de evento $\{C, I, R\}$, o indivíduo a quem pertence o evento $i \in \{1, \ldots, K\}$ e o seu tempo de ocorrência t no relógio de simulação. Os eventos da lista estão ordenados crescentemente de acordo com t.

Algorithm 1 Laço principal do simulador

```
while S_t ≠ K e tempo de simulação < T do
    e ← remover da LE
    if e.tipo == Caminhar then
        Gerar Caminhar para e.passeio
        Inserir Caminhar na LE
        v ← mover(e.passeio)
        if e.passeio.estado == infectado then
            infectar(v)
        else if e.passeio.estado == suscetivel then
            ser-infectado(v)
        end if
    else if e.tipo == Infectar then
        Gerar Recuperar
        Inserir Recuperar na LE
        infectar(e.passeio.vertice)
    else if e.tipo == Recuperar then
        ser-infectado(e.passeio.vertice)
    end if
end while
```

Ao processar o evento caminhar $\{C, i, t\}$, caso o indivíduo i esteja no estado suscetível e seja contagiado ao entrar no vértice de destino, é gerado um evento *Infectar* para o indivíduo (repare que a transição para o estado contraído é imediata, ocorrendo no mesmo instante do encontro). Por outro lado, se o indivíduo que se movimenta estiver no estado infectado, é gerado um evento *Infectar* para todo indivíduo que ele contagiar no vértice destino. Ao final, é gerado um novo evento *Caminhar* cujo tempo de ocorrência é $t_1 = t + \Delta_1$, sendo Δ_1 uma variável aleatória exponencial com parâmetro λ, definida na seção 3.

O processamento do evento $\{I, i, t\}$ muda o estado do indivíduo de contraído para infectado. Ao tornar-se infectado, o indivíduo pode infectar indivíduos suscetíveis (caso haja) que estão no mesmo vértice que ele, gerando um novo evento *Infectar* para cada

indivíduo contagiado. O processamento do evento *Infectar* gera um evento *Recuperar* para o mesmo indivíduo, ou seja, é gerado um evento $\{R, i, t_1\}$, cujo tempo de ocorrência é $t_1 = t + \Delta_2$, sendo Δ_2 uma variável aleatória exponencial com parâmetro γ.

Ao processar o evento $\{R, i, t\}$, o indivíduo i muda do estado infectado para o estado suscetível. Quando isso ocorre, o indivíduo pode ser novamente contagiado por indivíduos infectados que estejam no mesmo vértice que ele. Deste modo, no mesmo instante de tempo t, o indivíduo passa do estado infectado para o suscetível e do suscetível para o contraído, e um novo evento *Infectar* é gerado. Todo evento *Infectar* é gerado no instante em que o indivíduo muda para o estado epidêmico contraído. Em particular, o evento $\{I, i, t_1\}$ é gerado no instante t no qual o indivíduo i é contagiado, onde $t_1 = t + \Delta_3$, sendo Δ_3 uma variável aleatória exponencial com parâmetro τ.

O algoritmo 1 mostra o laço principal do simulador que processa o próximo evento da lista de eventos, gerando os eventos necessários. Repare que a simulação termina quando não há mais indivíduos infectados ou o tempo de simulação atinge o limite de T. As funções *ser-infectado()* e *infectar()* são fundamentais para a complexidade do simulador e serão discutidas em seguida.

O tamanho da lista de eventos do simulador é proporcional ao número de indivíduos K, havendo sempre um evento *Caminhar* para cada indivíduo. Os eventos *Infectar* e *Recuperar* são mutuamente exclusivos, pois caso o indivíduo esteja no estado contraído, existe um evento *Infectar* na lista de eventos e caso ele esteja no estado infectado, existe um evento *Recuperar* na lista de eventos. Além disso, no estado suscetível, o indivíduo não possui nenhum dos dois eventos. Portanto, cada indivíduo pode ter no mínimo um e no máximo dois eventos na lista de eventos. Assim, o tamanho da lista de eventos respeita a seguinte desigualdade: $K \leq |LE| \leq 2K$.

5. Desempenho computacional do simulador

A complexidade computacional do simulador depende tanto do algoritmo 1 quanto dos algoritmos que processam cada tipo de evento. Um aspecto central são as estruturas de dados utilizadas para armazenar a lista de eventos, a rede e os indivíduos. Os eventos são armazenados em uma fila de prioridade mínima, onde a complexidade das operações de remover e inserir tem custo $\mathcal{O}(\log K)$. O evento caminhar precisa escolher um vértice vizinho na rede, mas isto tem complexidade constante $\mathcal{O}(1)$, pois o grafo é armazenado é um vetor de adjacências. Desta forma a complexidade do laço principal do simulador não depende da estrutura e do tamanho da rede.

Na função *ser-infectado*, é utilizado a equação $1 - (1 - p_u)^{k_u}$ (k_u é o número de indivíduos no vértice u), ao custo $\mathcal{O}(1)$, para determinar se o indivíduo será contagiado por outros indivíduos infectados no vértice destino. Já na função *infectar*, é preciso iterar sobre todos os indivíduos suscetíveis no vértice destino para determinar individualmente se cada um será (ou não) contagiado, ao custo $\mathcal{O}(k_u)$. Sempre que houver um contágio, um novo evento *Infectar* é gerado para o indivíduo recém infectado e inserido na lista de eventos ao custo $\mathcal{O}(\log K)$. Como os indivíduos se movimentam aleatoriamente, $k_u = K$, no pior caso, e consequentemente a operação *infectar*, e portanto o laço principal do simulador, tem custo $\mathcal{O}(K \log K)$.

Repare que no laço principal, a simulação termina quando o tempo de simulação alcança o tempo limite de simulação T, de forma que a complexidade do simulador

seja linear no tempo de simulação (dado todos os outros parâmetros do modelo). Entretanto, o número de eventos processados por unidade de tempo de simulação depende dos parâmetros do modelo, tais como taxa de caminhar, taxa de recuperar, número de indivíduos e tamanho da rede. Ou seja, o tempo de execução necessário para simular um tempo T de simulação depende dos parâmetros do modelo, mas dado os parâmetros, este tempo de execução é linear em T.

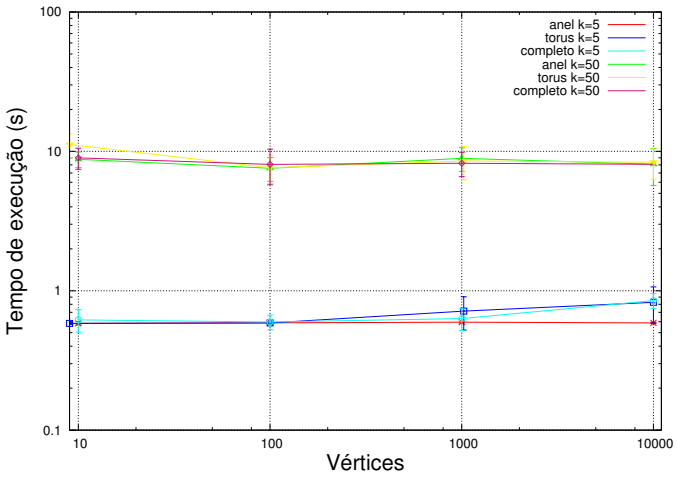

Figura 1. Tempo execução em função do tamanho da rede, n, para diferentes redes e número de indivíduos ($K = 5$, $K = 50$) em escala log-log. Parâmetros: $T = 10000, \lambda = 1, \gamma = 1, \tau = 1$.

5.1. Tempo de execução

Nesta seção iremos apresentar o tempo de execução do simulador (tempo de relógio de parede) para diferentes cenários. Os tempos foram coletados em uma CPU Intel Xeon ilustram a escalabilidade do simulador, sendo os valores de menor importância. Para cada cenário, 30 rodadas independentes foram realizadas, obtendo-se a média do tempo de execução, que é apresentada nos gráficos a seguir.

A figura 1 apresenta o tempo de execução em função do número de vértices da rede, para diferentes redes e diferentes números de indivíduos em cada cenário. Repare que o tempo de execução é constante no tamanho da rede, e independe da estrutura da rede. Pare redes com maior número de indivíduos ($K = 50$), o tempo aumenta com relação ao caso com menos indivíduos ($K = 5$), mas permanece constante com o aumento da rede.

A figura 2 apresenta o tempo de execução do simulador em função do tempo limite de simulação (T) para três tipos de rede, de epidemias com $K = 5$ e $K = 50$ indivíduos. Note que o desempenho do simulador é linear no tempo de simulação, corroborando a análise teórica, pois em escala $\log - \log$, a reta de inclinação 1 corresponde a uma relação linear na escala convencional. Os resultados também indicam que este tempo não depende da estrutura da rede (retas praticamente idênticas para as três redes). Entretanto, repare que o cenário com 50 indivíduos demanda praticamente 10 vezes mais tempo de execução

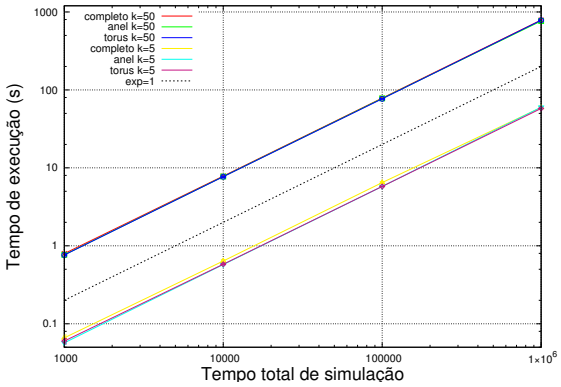

Figura 2. Tempo de execução em função do tempo de simulação (T), em escala log-log. Parâmetros: $K = \{5, 50\}, n = 25, \lambda = 1, \gamma = 1, \tau = 1$

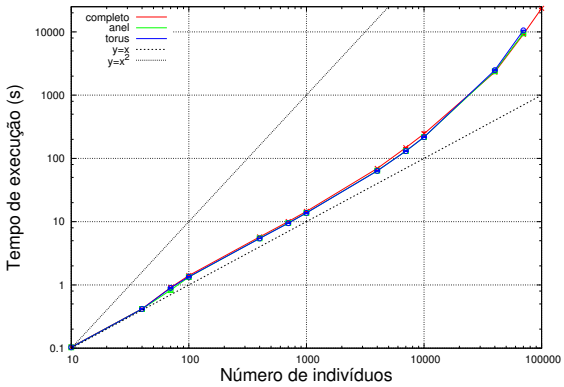

Figura 3. Tempo de execução em função do número de indivíduos na rede (K), em escala log-log (retas com inclinação 1 e 2 apenas para referência). Parâmetros: $T = 1000, n = 50, \lambda = 1, \gamma = 1, \tau = 1$

do que com 5 indivíduos. Este comportamento é esperado, pois teremos aproximadamente 10 vezes mais eventos a serem processados por unidade de tempo de simulação.

A Figura 3 apresenta o tempo de execução do simulador em função do número de indivíduos (K) para três redes. Nota-se que, para K pequeno (menor que 1000), o simulador apresenta uma escalabilidade praticamente linear (reta com inclinação 1 na escala $\log - \log$). Entretanto, conforme K aumenta (para valores maiores que 1000), a inclinação da curva representando o tempo de execução também aumenta. Este comportamento está de acordo com a análise teórica de pior caso, que tem complexidade $\mathcal{O}(K \log K)$. Repare que para valores grandes de K, o fator $\log K$ faz diferença. Novamente, observamos que o desempenho do simulador não depende da estrutura da rede, já que as curvas são praticamente idênticas para as três redes.

Referências

Barrett, C. L., Bisset, K. R., Eubank, S. G., Feng, X., and Marathe, M. V. (2008). Episimdemics: An efficient algorithm for simulating the spread of infectious disease over large social networks. In *ACM/IEEE Conf on High Performance Computing (SC)*.

Draief, M. and Ganesh, A. (2011). A random walk model for infection on graphs: spread of epidemics and rumours with mobile agents. *Discrete Event Dynamic Systems*, 21:41–61.

Ganesh, A., Massoulie, L., and Towsley, D. (2005). The effect of network topology on the spread of epidemics. In *IEEE Infocom*, pages 1455–1466.

Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203.

Zhou, J. and Liu, Z. (2009). Epidemic spreading in communities with mobile agents. *Physica A: Statistical Mechanics and its Applications*, 388:1228–1236.