

Elementos de Programação

Projecto de Biocomputação

Departamento de Matemática, IST

Novembro de 2018

Propagação de epidemias

O estudo da propagação de epidemias é uma área de investigação importante e actual, nomeadamente em medicina, no que diz respeito aos fenómenos de contágio associados a doenças infecciosas, mas também, por exemplo, na análise de processos de disseminação de informação em redes sociais.

Existem diversas abordagens à modelação de epidemias, algumas delas estocásticas, mas na sua maioria recorrendo a modelos determinísticos [1, 2, 3]. Os modelos mais usados são usualmente denominados SIS, SIR e SEIR, consoante a forma de compartimentação dos indivíduos. Nestes modelos epidémicos os indivíduos da população são compartimentados de acordo com o seu estado em relação à doença: *susceptível*, *exposto*, *infectado* e *recuperado*. No modelo SIS um indivíduo *susceptível* pode ser contagiado por um indivíduo *infectado* passando ele próprio a *infectado*, mas não desenvolve imunidade à doença. Assim, após recuperar, regressando ao estado *susceptível*, pode voltar a ser infectado. No modelo SIR os indivíduos, depois de infectados, desenvolvem sempre imunidade à doença passando para o estado *recuperado*, não voltando ao estado *susceptível*. No modelo SEIR, que é baseado no modelo SIR, um indivíduo *susceptível* que entre em contacto com um indivíduo *infectado* passa primeiro ao estado *exposto* que irá mais tarde evoluir para o estado *infectado* mas durante esta fase não pode contagiar. A Figura 1 ilustra as possíveis transições de estado de um indivíduo de acordo com o modelo SEIR. Um indivíduo pode morrer em qualquer um dos estados, situação que se encontra representada a tracejado.

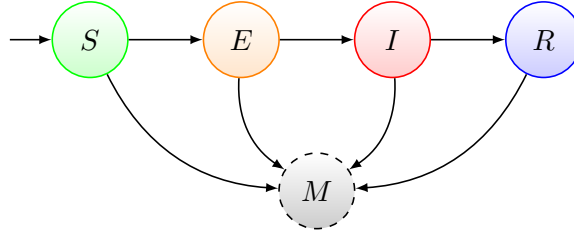


Figura 1: Diagrama de transição de um indivíduo.

Neste projecto, pretende-se desenvolver em Python um simulador de propagação de epidemias baseado num modelo SEIR, enriquecido com mecanismos de vida artificial, de acordo com os princípios de simulação discreta estocástica.

Neste modelo, considera-se uma população inicial de indivíduos *susceptíveis*, exceptuando uma percentagem pequena de *infectados*. Os indivíduos encontram-se dispostos numa grelha de dimensão N , como se descreve adiante. Cada indivíduo pode evoluir de acordo com vários eventos que podem afectar o seu estado.

Descrição do modelo

Uma grelha de dimensão N é uma grelha quadriculada com $(2N + 1) \times (2N + 1)$ quadriculas, como se ilustra na figura seguinte.

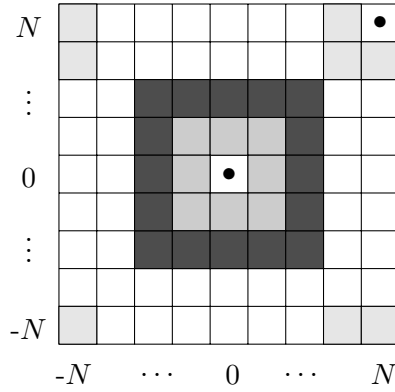


Figura 2: Grelha e vizinhanças.

Cada quadrícula é identificada pelas suas duas coordenadas inteiras (x, y) com $x, y \in \{-N, \dots, 0, \dots, N\}$. As 8 quadrículas contíguas a uma quadrícula (a, b) são a vizinhança 1 de (a, b) . As quadrículas a cinzento claro na Figura 2 são a vizinhança 1 de $(0, 0)$. De forma semelhante, as 16 quadrículas que envolvem a vizinhança 1 de (a, b) são a vizinhança 2 de (a, b) . As quadrículas a cinzento escuro nesta figura são a vizinhança 2 de $(0, 0)$. Considera-se que a primeira e a última linhas são contíguas. Da mesma forma considera-se que a primeira e a última colunas também são contíguas. Esta situação também está ilustrada na Figura 2 para a quadrícula (N, N) em que as quadrículas sombreadas são a vizinhança 1 de (N, N) .

Na grelha podem existir obstáculos, ou seja, quadrículas que não podem ser ocupadas por indivíduos. Na Figura 3 ilustra-se uma grelha com um obstáculo, constituído pelas três quadrículas a preto. Mais tarde será relevante saber se dois indivíduos estão em contacto para efeitos de contágio. Na ausência de obstáculos dois indivíduos estão em contacto se estiverem ou na vizinhança 1 ou na vizinhança 2 um do outro. Se existirem obstáculos, para dois indivíduos na vizinhança 2 um do outro estarem em contacto terá de existir uma quadrícula livre de obstáculos comum às vizinhanças 1 de cada um deles. Na figura 3 ilustram-se dois cenários possíveis. Nesta figura, os indivíduos 1 e 2 não estão em contacto mas os indivíduos 1 e 3 estão.

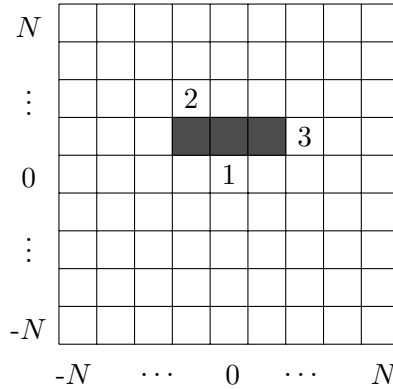


Figura 3: Grelha e obstáculos.

Um indivíduo é caracterizado por uma identificação (um número natural único para cada indivíduo) e um estado (“S”, “E”, “I”, e “R”). No início da simulação distribuem-se P_s indivíduos susceptíveis e P_i indivíduos infectados aleatoriamente pelas quadrículas livres de obstáculos da grelha, garantindo que cada quadrícula é ocupada por, no máximo, um indivíduo.

Cada indivíduo evolui de acordo com os seguintes mecanismos aleatórios, até ser atingido um *tempo limite de simulação* T_h :

- **Deslocamento**, com tempo médio entre deslocamentos dado por uma variável aleatória exponencial com valor médio T_d . O indivíduo desloca-se para uma posição livre* na sua vizinhança 1, se existir alguma, com probabilidade p_d . No entanto, se na vizinhança 1 existirem pelo menos 3 indivíduos infectados e posições livres o indivíduo desloca-se sempre. A nova posição é escolhida uniformemente de entre as posições livres na vizinhança 1.
- **Reprodução**, com tempo médio entre reproduções dado por uma variável aleatória exponencial com valor médio T_r . A reprodução efectiva-se, com probabilidade p_r , desde que na vizinhança 1 do indivíduo existam pelo menos duas posições livres e pelo menos uma ocupada por outro indivíduo. O novo indivíduo é criado no estado *susceptível* (independentemente do estado do progenitor) e colocado uniformemente numa das posições livres na vizinhança 1 do progenitor.
- **Morte**, com tempo de morte dado por uma variável aleatória exponencial com valor médio T_m . O indivíduo morre (é eliminado da simulação) com probabilidade p_m . Se estiver infectado a probabilidade de morte aumenta 10% (até um máximo de 1).
- **Avaliação** de estado, com tempo médio entre avaliações dado por uma variável aleatória exponencial com valor médio $T_a(s)$ que depende do estado s do indivíduo. Este evento gere a evolução do estado do indivíduo, podendo alterá-lo.

- Se o estado do indivíduo for *susceptível* e existirem indivíduos infectados na sua vizinhança 1 ou na sua vizinhança 2, o estado muda para *exposto* com probabilidade p_e , em que p_e depende do valor de $x = 2n_1 + n_2$ onde n_1 é o número de infectados na vizinhança 1 e n_2 o número n_2 de infectados *em contacto* na vizinhança 2, de acordo com a expressão

$$\frac{1}{2 \log(1.8)} - \frac{1}{2 \log(\frac{x(x-1)}{5} + 1.8)}.$$

Havendo mudança do estado para *exposto*, a próxima avaliação dar-se-á em tempo médio $T_a(\text{"E"})$. Se não existir mudança de estado, a próxima avaliação dar-se-á em tempo médio $T_a(\text{"S"})$.

- Se o estado do indivíduo for *exposto* este muda para o estado *infectado* e a nova avaliação realiza-se com tempo médio $T_a(\text{"I"})$.

*Não ocupada por indivíduos ou obstáculos.

- Se o estado do indivíduo for *infectado* este muda para o estado *recuperado* do qual não sairá (excepto por morte), não sendo necessárias novas avaliações.

Simulador

O simulador deve ser desenvolvido seguindo o *método da programação modular por camadas centrado nos dados*.

1. Comece por identificar os tipos de dados relevantes e suas operações. Nomeadamente, os tipos de dados relevantes são **evento**, **cap**, **indivíduo** e **grelha**.

Note que estes tipos de dados devem incluir, entre outras, operações para construir valores, consultar a informação associada, e manipular os valores de forma a implementar o simulador.

No caso do tipo **grelha**, devem ser consideradas operações que permitam colocar e retirar indivíduos da grelha; determinar a posição de um indivíduo; determinar se uma posição está livre ou ocupada; calcular as posições ocupadas adjacentes a uma posição dada (vizinhança 1); determinar o número de *infectados* nas vizinhanças 1 e 2 (em contacto); determinar o número total de *infectados* na grelha; determinar as listas das coordenadas dos indivíduos *susceptíveis*, *expostos*, *infectados* e *recuperados* na grelha, entre outras que considere relevantes.

2. Implemente estas camadas sobre a camada básica do Python.
3. Desenvolva o programa abstracto pretendido sobre a camada que disponibiliza estes tipos de dados.
4. Integre o programa obtido em 3 com os módulos desenvolvidos em 2.
5. Experimente o programa desenvolvido com diversos conjuntos de dados à sua escolha, apresentando os respectivos resultados. Em particular, considere o conjunto:

- $N = 10, P_s = 96, P_i = 4, T_h = 50, T_d = 1, T_r = 10, T_m = 20, p_d = 0.6, p_r = 0.3, p_m = 0.5, T_a(\text{"S"}) = 1, T_a(\text{"E"}) = 1, T_a(\text{"I"}) = 10$ e lista de obstáculos $[(2,-10),(2,-9),\dots,(2,0),(2,1),\dots,(2,10)]$.

6. Considere a possibilidade de gerar diferentes configurações iniciais :
 - (a) os indivíduos encontram-se distribuídos aleatoriamente na grelha;
 - (b) os indivíduos encontram-se distribuídos aleatoriamente na grelha, mas em que p por cento deles se encontram em posições adjacentes.

Note que o programa a desenvolver deve:

- receber os seguintes dados
 - a dimensão N da grelha;
 - os números P_s e P_i de indivíduos *susceptíveis* e *infectados* inicialmente na população;
 - o tempo limite de simulação T_h ;
 - os tempos médios T_d de deslocamento, T_r de reprodução e T_m de morte;
 - as probabilidades p_d de deslocamento, p_r de reprodução e p_m de morte;
 - os tempos médios $T_a(\text{“S”})$, $T_a(\text{“E”})$, $T_a(\text{“I”})$ necessários para a função T_a ;
 - a lista de quadrículas, identificadas pelas suas coordenadas, que constituem os obstáculos,
- e devolver a seguinte informação
 - um gráfico com a evolução do número de infectados ao longo do tempo;
 - um ficheiro `resultados.txt` que, na primeira linha, tem a dimensão N da grelha, na segunda linha tem a lista de obstáculos, e em cada uma das linhas seguintes tem uma lista com 4 listas com as coordenadas dos indivíduos *susceptíveis*, *expostos*, *infectados* e *recuperados*, respectivamente, correspondendo à evolução da grelha ao longo do tempo. Será disponibilizado na página da disciplina um *notebook* Python que lê o conteúdo do ficheiro `resultados.txt` e gera uma animação da evolução do sistema. Para que a animação seja gerada correctamente, é necessário que a estrutura do ficheiro `resultados.txt` seja respeitada.

Referências

- [1] W. Kermack, A. McKendrick. A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society A*:115, p. 700–721, London 1927.
- [2] L. Allen. An introduction to stochastic epidemic models. In F. Brauer, P. van den Driessche, J. Wu (eds), *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, p. 81–130, Springer 2008.
- [3] P. Sanchez, S. Sanchez. A scalable discrete event stochastic agent-based model of infectious disease propagation. In L. Yilmaz *et al.* (eds), *Proceedings of the 2015 Winter Simulation Conference*, p. 151–158, IEEE 2015.

Entrega do projecto

O projecto é entregue através do sistema Fenix, após a inscrição do respectivo grupo. A entrega do projecto está dividida em duas partes.

Parte 1

Na primeira parte do projecto, cada grupo deve submeter os módulos desenvolvidos no ponto 2 da secção anterior. A entrega deve consistir de um único arquivo (zip ou rar) contendo os módulos desenvolvidos, e um pequeno relatório descrevendo as operações dos tipos de dados e explicando as principais opções tomadas para a sua implementação, bem como exemplos ilustrando o seu correcto funcionamento.

Data limite de submissão: **23h59m do dia 1 de Dezembro de 2018.**

Parte 2

Na segunda parte do projecto, cada grupo deve submeter o simulador. Para esta fase, serão disponibilizados na página da disciplina módulos com implementações dos tipos de dados. Cada grupo pode optar por desenvolver o simulador recorrendo aos módulos disponibilizados ou recorrendo aos seus próprios módulos, possivelmente alterados após a primeira submissão. Tal opção deve estar claramente identificada no relatório, implicando a ressubmissão dos elementos da Parte 1, caso os tipos de dados tenham sido alterados e utilizados. A entrega deve consistir de um único arquivo (zip ou rar) contendo o simulador e eventuais módulos adicionais que tenham sido desenvolvidos, e um pequeno relatório explicando as principais opções tomadas para a implementação do simulador e exemplos que ilustrem e permitam analisar o comportamento do modelo proposto.

Data limite de submissão: **23h59m do dia 15 de Dezembro de 2018.**