epinet: An R Package to Analyze Epidemics Spread across Contact Network

Joanna Jabłońska

Wydział Matematyki i Nauk Informacyjnych Politechnika Warszawska



Do czego służy pakiet epinet?

Za jego pomocą można robić symulacje oraz analizy rozprzestrzeniania się epidemii w populacji.

Modelowanie epidemii

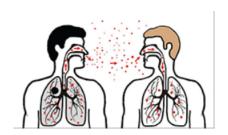
Do modelowania powiązań strukturalnych dla rozprzestrzeniania się epidemii służą sieci.

Sieć jest reprezentowana jako para (N,M), gdzie N jest liczbą węzłów, a M macierzą wierzchołków, które wyznaczają wszystkie krawędzie.



Kiedy istnieje połączenie?

Węzeł w takiej sieci jest reprezentowany przez jednostkę w populacji, a krawędź między jednostkami odnosi się do interakcji między nimi, która nastąpiła podczas trwania epidemii i która mogła doprowadzić do transmisji patogenu.



Uwaga: Istnienie kontaktu między jednostkami nie musi oznaczać, że patogen został przekazany.

Specyfikacja modelu

Podstawowe założenia dla modelowania epidemii przy użyciu pakietu **epinet**:

- Liczba węzłów w sieci jest stała, a rozmiar populacji jest znany;
- Szerzenie się epidemii jest modelowane przez tzw "SEIR compartmental model";
- Zaobserwowane dane zawierają momenty czasu, w których jednostki przechodzą z jednego stanu do kolejnego (tj. czas infekcji, czas wyzdrowienia)

SEIR Compartmental model

S (Susceptible)

Jednostki narażnone na zarażenie przez osobę zainfekowaną.

E (Exposed)

Jednostki zakażone, ale jeszcze nie zarażające.

I (Infectious)

Jednostki zakażone, które mogą przekazywać patogen osobom podatnym.

R (Removed)

Jednostki, które już nie zarażają i nie grają roli w epidemii.

Uwaga: Zaimplementowany w pakiecie model nie uwzględnia ponownego zarażenia.

Exponential-Family Random Graph Model (ERGM)

W pakiecie epinet wykorzystuje się specyfikację klasy ERGM, która służy do modelowania sieci powiązań między osobnikami, tzn: prawdopodobieństwo istnienia krawędzi pomiedzy dowolnymi jednostkami w populacji jest niezależna od istnienia innych krawędzi w sieci.

W tej klasie modeli można określić prawdopodobieństwo istnienia krawędzi powiędzy węzłami i oraz j jako $p_{\{i,j\}}$, gdzie:

$$\log\left(\frac{p_{\{i,j\}}}{1-p_{\{i,j\}}}\right) = \sum_{t=1}^k \mathbf{X}_{\{i,j\},t} \, \boldsymbol{\eta}_t.$$

X - macierz par zmiennych rozmiaru $\binom{N}{2} \ge k,$ $\eta = (\eta_1,...,\eta_k)$ - wektor, który odpowiada współczynnikom dla sieci.

Przykład

Poniżej znajduję się tabela z wyestymowanymi współczynnikami dla modelu epidemii.

Effect (Parameter)	Mean
Intercept	0.44
Classroom	7.32
House Distance	-0.31

Rozważmy parę uczniów *i* i *j* chodzących do jednej klasy, których domostwa są oddalone o 20 metrów.

Wówczas:
$$X_{\{i,j\},1} = 1$$
 i $X_{\{i,j\},2} = 20$.

Wtedy logarytm szans, że doszło do kontaktu między osobnikami wynosi: 0.44 + 1*(7.32) + 20*(-0.31) = 1.56, skąd prawdopodobieństwo, że istniał kontakt między osobnikami wynosi:

$$\frac{e^{1.56}}{e^{1.50}+1} = 0.83$$

Parametry epidemii β , θ_E , θ_I , k_E oraz k_I

- Czas oczekiwania na przekazanie zakażenia jest modelowany przy pomocy zmiennej losowej o rozkładzie wykładniczym ze średnią $1/\beta$,
- Czas, gdy jednostka znalazła się w stanie zainfekowania bez możliwości zarażania i z możliwością przekazywania patogenu jest modelowany przez zmienne losowe o rozkładzie gamma z parametrami, odpowiednio, (θ_E,k_E) oraz (θ_I,k_I) , gdzie średnia dla tak zadanego rozkładu gamma jest postaci $k\theta$, a wariancja $k\theta^2$.

Rozkłady a priori dla parametrów modelu

- ullet Parametrom η regulującym prawdopodobieństwa istnienia kontaktu między osobnikami krawędzi przypisany jest rozkład normalny.
- Rozkład a priori dla parametrów epidemii jest wybierany tak, żeby uwzględniał każdą ważną biologiczną informację posiadaną na temat patogenu (np. parametry epidemii θ_E oraz θ_I mogą być opisane odwróconym rozkładem gamma, wtedy reszta parametrów epidemii opisana jest rozkładem gamma).

Symulacja epidemii

Pakiet epinet może być użyty do symulacji rozprzestrzeniania się epidemii w populacji.

- Budujemy macierz danych,
- Tworzymy macierz zmiennych (BuildX()) dla par osobników,
- Tworzymy sieć interakcji/kontaktów między jednostkami, (SimulateDyadicLinearERGM()),
- Tworzymy symulację epidemii (SEIR.simulation()).

Przyjrzyjmy się ostatniej funkcji:

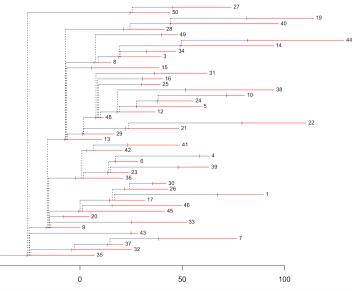
```
SEIR.simulator(M, N, beta, ki, thetai, ke = ki, thetae = thetai, latencydist = "fixed", latencyperiod = 0)
```

- M macierz kontaktów (obiekt wygenerowany przez funkcję SimulateDyadicLinearERGM()),
- beta,thetae, ke, thetai, ki odnoszą się do parametrów epidemii: $\beta, \theta_E, k_E, \theta_I, k_I$,
- latencydist="fixed" (latencydist="gamma") odnosi się do czasu spędzonego w stanie "exposed" dla każdego osobnika i może być modelowany jako stały lub przy pomocy rozkładu gamma.

Przykładowa tabela wynikowa

	Node	ID	Parent	Etime	Itime	Rtime
[1,]		35	NA	-66.14114560	-25.8757849	7.099042
[2,]		50	35	-25.61093119	24.5164631	44.023516
[3,]		9	35	-24.88898877	-16.5180717	0.000000
[4,]		43	35	-24.67232216	24.9623163	28.253567
[5,]		32	35	-23.94421137	-3.8943032	25.342128
[6,]		13	9	-15.84157623	-7.5727699	10.718028
[46,]		44	14	49.52465905	81.8946425	129.073446
[47,]		2	NA	NA	NA	NA
[48,]		11	NA	NA	NA	NA
[49,]		18	NA	NA	NA	NA
[50,]		47	NA	NA	NA	NA

Transmission Tree



-50

Funkcje

Function	Input(s)	Output
BuildX	Types of covariate(s) desired, e.g., matching, distance, etc.	A matrix of dyadic covariates of the form required by SEIR.simulator() and epinet().
SimulateDyadicLinearERGM	Network $(\pmb{\eta})$ parameters for desired network.	A simulated ERGM network in edge list matrix format.
plot.epidemic	An epidemic in the form produced by ${\tt SEIR.simulator()}.$	A graph visually depicting the epidemic.
plot.epinet	MCMC output in the form produced by $\operatorname{\mathtt{epinet}}()$.	A graph visually depicting the state of the epidemic at a particular point in the MCMC chain.
epi2newick epi2newickmcmc	An epidemic in the form produced by SEIR.simulator(). MCMC output in the form produced by epinet().	The transmission tree corresponding to the epidemic as a string in Newick format. The transmission tree corresponding to an inferred epidemic as a string in Newick format.
write.epinet	MCMC output in the form produced by $\mbox{\tt epinet}().$	Written output files consisting of the posterior samples of the epidemic parameters and transmission trees.
ess	MCMC output for a single variable, i.e., a sample from the posterior distribution of a variable.	An estimate of the ESS (effective sample size) of the sample.

Table 1: Description of some of the functions in package epinet.

Przykład 1

Dane: Mamy populację składającą się z 50 osób o identyfikatorach od 1 do 50.

Będą rozważane dwie zmienne - stała oraz odległość euklidesowa między każdą parą osób. Dla każdej osoby przypisujemy losowo dwuwymiarowe dane przestrzenne.

```
library("epinet")
set.seed(1)
N <- 50
mycov <- data.frame(id = 1:N, xpos = runif(N), ypos = runif(N))
dyadCov <- BuildX(mycov, binaryCol = list(c(2, 3)), binaryFunc = "euclidean")</pre>
```

Pierwsze wiersze wynikowe prezentują się w następujący sposób:

	node.1	node.2	(Intercept)	xpos.ypos.L2Dist
[1,]	1	2	1	0.3981306
[2,]	1	3	1	0.3098754
[3,]	1	4	1	0.6835703
[4,]	1	5	1	0.4119156
[5,]	1	6	1	0.7372506
[6,]	1	7	1	0.6980691

Mając powyższą macierz par węzłów i zmiennych możemy zasymulować sieć interakcji w populacji. Do tego posłuży nam funkcja SimulateDyadicLinearERGM().

```
eta <- c(0, -7) net <- SimulateDyadicLinearERGM(N = N, dyadiccovmat = dyadCov, eta = eta)
```

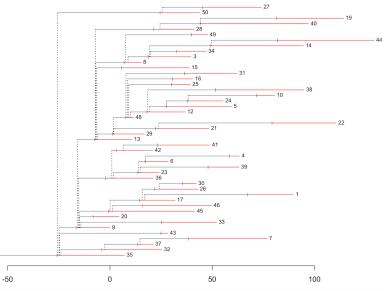
	node.1	node.2
[1,]	1	17
[2,]	1	25
[3,]	1	33
[4,]	1	34
[5,]	2	11
[6,]	3	8
[7,]	3	14
[8,]	3	28
[9,]	3	33

Symulacja epidemii w populacji, przy założeniu, że długość czasu spędzonego w stanie "exposed" jest opisana rozkładem "gamma".

```
epi <- SEIR.simulator(M = net, N = N, beta = 1, ki = 3, thetai = 7,ke = 3, latencydist = "gamma")
```

	Node	ID	Parent	Etime	Itime	Rtime
[1,]		35	NA	-66.14114560	-25.8757849	7.099042
[2,]		50	35	-25.61093119	24.5164631	44.023516
[3,]		9	35	-24.88898877	-16.5180717	0.000000
[4,]		43	35	-24.67232216	24.9623163	28.253567
[5,]		32	35	-23.94421137	-3.8943032	25.342128
[6,]		13	9	-15.84157623	-7.5727699	10.718028
[46,]		44	14	49.52465905	81.8946425	129.073446
[47,]		2	NA	NA	NA	NA
[48,]		11	NA	NA	NA	NA
[49,]		18	NA	NA	NA	NA
[50,]		47	NA	NA	NA	NA

Transmission Tree



Zaprezentujemy procedurę użycia funkcji epinet(), w celu przedstawienia wnioskowania na temat parametrów modelu, gdy ich nie znamy.

```
epinet(formula, epidata, dyadiccovmat, mcmcinput = MCMCcontrol()
  priors = priorcontrol(), verbose = TRUE)
```

Kod w skrypcie.

Funkcja epinet wykonuje wnioskowanie bayesowskie na parametrach modelu epidemii (SEIR).

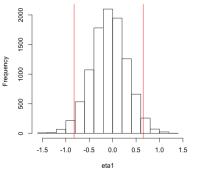
epinet(formula, epidata, dyadiccovmat, mcmcinput =
MCMCcontrol(), priors = priorcontrol(), verbose =
TRUE)

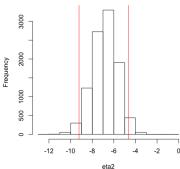
- formula obiekt, mówiący które zmienne mają zostać włączone do modelu (intercept uwzględniony domyślnie),
- epidata macierz w formacie SEIR.simulator(),
- dyadiccovmat macierz par zmiennych,
- mcmcinput lista z parametrami dla algorytmu MCMC,
- priors lista rozkładów a priori i hiperparametrów dla algorytmu MCMC,
- verbose wartość logiczna, mówiąca, czy wyświetlać informacje z procedury MCMC.

Epidemic parameter	Mean	Orginal
Beta	0.7210	1
Theta_e	10.065	7
Theta_i	8.468	7
k_e	2.257	3
k_i	2.736	3

Network parameter	Mean	Orginal
Intercept	-0.07679	0
xpos.ypos.L2Dist	-6.835	-7

Możemy zobaczyć, czy algorytm był w stanie parametry sieci $\eta.$





Przykład 2: Analiza danych dotyczących szerzenia się odry, Hegelloch, Niemcy, zima 1861

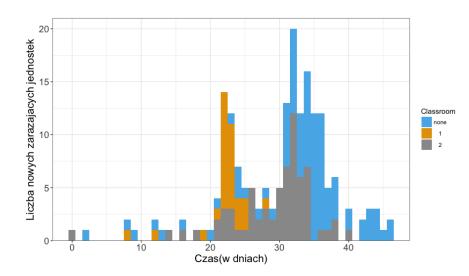
Dane składają się z:

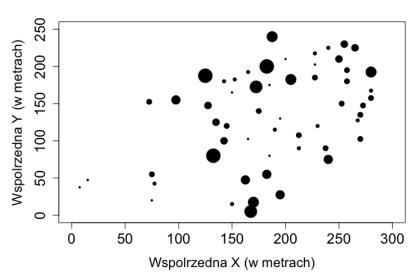
 Macierzy HagellochTimes, która zawiera informacje o jednostkach w populacji.

	NodeID	Parent	Etime	Itime	Rtime
[1,]	1	179	NA	51	59
[2,]	2	5	NA	53	61
[3,]	3	73	NA	58	66
[4,]	4	23	NA	57	62
[5,]	5	10	NA	52	61
Γ6. _]	6	21	NA	56	63

Macierzy zmiennych HagellochDyadCov między parami osobników.

	Node	ID 1	Node	ID	2 I	Household	Classroom	1	Classroom	2	House	Distance	Male	Match
[1,]		1			2	1		1		0		0.000000		0
[2,]		1			3	1		0		0		0.000000		0
[3,]		1			4	0		0		0		9.055385		0
[4,]		1			5	0		1		0		8.062258		0
[5,]		1			6	0		0		0		8.062258		0
[6,]		1			7	0		0		0	!	55.362442		0
	Femal	e Mo	ıtch A	ge D	if	f								
[1,]			1		:	1								
[2,]			1		3	3								
[3,]			0		(6								
[4,]			1		:	1								
[5,]			0			5								
[6,]			0		:	1								





Podsumowanie wyestymowanych wpółczynników dla parametrów sieci.

Effect (Parameter)	Mean	Median	95% CI
(Intercept) (η_0)	-0.44	-0.36	(-1.44, 0.51)
Classroom 1 (η_1)	7.54	7.44	(4.89, 10.60)
Classroom 2 (η_2)	0.77	0.80	(-0.26, 1.70)
House Distance (η_3)	-0.11	-0.12	(-0.17, -0.06)

Widzimy, że parametry dla klas mają średnie większe od 0, co sugeruje, że będąc w klasie zwiększa się prawdopodobieństwo, że osoby przekażą sobie wirusa. Za to efekt odległości ma negatywny wpływ na rozprzestrzenianie się epidemii. Im dalej siebie mieszkają osobniki, tym mniejsze prawdopodobieństwo, że nastąpi przekazanie wirusa.

Podsumowanie

Pakiet **epinet** pozwala nam na podstawie danych na temat epidemii otrzymać jej parametry strukury tej epidemii, a w szczegółności prawdopodobieństwo, że pomiędzy dwiema jednostkami istniał kontakt.

Ograniczenia:

- założenie niezależności w istnieniu krawędzi między jendostkami,
- zdobycie dokładnych danych na temat przechodzenia przez różne stadia epidemii jest często niemożliwe, dlatego wymaga dodatkowych założeń,
- wielkość epidemii czas procedury MCMC zachowuje się jak kwadrat wielkości populacji,
- MCMC wymaga dużej ilośći iteracji.