Viral spreading on network structures

G. Canright, K. Engø-Monsen, A. Frigessi, O.Haug, K. Hellton,
R. Parviero, H. Rognebakke & I. Scheel

Department of Mathematics, Universitetet i Oslo

Abstract

One of the main questions that arise upon the launch of a new product is to predict whether it can be become popular and how. Assuming that the customers' population can be represented as a network with a known structure, analysing data on previous adoptions will permit us to estimate how new adoptions will move through said network, with the aim to detect whether or not a viral spreading is taking place.
In this work, telco data is used to build the customers' network. Subsequently, we analyse data to understand how a certain product has spread through it during a training period and provide a prediction for the popularity of said product.

Method

Let $i = 1, \dots, n$ identify each individual and $\boldsymbol{x}_i$ denote a vector of p covariates related to individual i. This vector contains typical demographic covariates such as age, gender, geographic location, and so on. These covariates are assumed not to vary over time.
For every pair of individuals on the market it is possible to define a covariate vector $\boldsymbol{x}_{ij}$ containing informations about the interaction between individual i and j with regards to the amount of phone calls and text messages the two individuals share with each other. This information is used to build the customers' network.
The rates of convincement from individual i to individual j are then modelled, and estimated by a maximum likelihood approach.

Model

The rate of convincement [1] at a given time t of the j-th individual is defined as:

$$\lambda_j(t) = \lambda_{0j}(t) + \sum_{i \in C_j(t)} \lambda_{ij}(t),$$

with $\lambda_{ij}(t)$ being the viral component:

$$\lambda_{ij}(t) = \exp\{\beta_0 + \boldsymbol{\beta}_i \boldsymbol{x}_i + \boldsymbol{\beta}_j \boldsymbol{x}_j + \boldsymbol{\beta}_{ij} \boldsymbol{x}_{ij}\}\, I_{ij}(t),$$

and $\lambda_{0j}(t)$ representing the external pressure:

$$\lambda_{0j}(t) = \exp\{\alpha_0 + \alpha_{BTL} + \alpha_{ext} \boldsymbol{x}_j\}\, I_{0j}(t).$$

The indicator variables I_{ij} are set to 1 when the neighbouring nodes of j have bought the product and are still "infectious" at time t, the set $C_j(t)$ contains said nodes. The external node is always able to convince the individual j, hence $I_{0j}(t)$ is 1 when j is susceptible.

Data

This model was tested on data regarding the adoption of a smartphone application in an European country. We had access to demographic covariates regarding the customers of a telephone provider, how they interacted with each other, and whether or not they had installed said smartphone application on their devices.

	ML estimates		CI 95%	
	Mean	SD	Inf	Sup
beta.0	−6.6965	0.2044	−7.1054	−6.2877
i.gender	−0.3320	0.0497	−0.4315	−0.2326
i.age.group2	−0.7138	0.1870	−1.0878	−0.3398
i.age.group3	−0.8791	0.1740	−1.2271	−0.5311
i.age.group4	−0.8673	0.1778	−1.2229	−0.5117
i.age.group5	−0.9810	0.1971	−1.3752	−0.5869
j.gender	0.0004	0.0466	−0.0927	0.0935
j.age.group2	0.4375	0.1699	0.0977	0.7772
j.age.group3	0.7782	0.1576	0.4631	1.0934
j.age.group4	0.6747	0.1609	0.3530	0.9964
j.age.group5	0.0443	0.1747	−0.3050	0.3936
ij.weight.voice	1.1080	0.0627	0.9825	1.2335
ij.weight.SMS	0.7783	0.0706	0.6372	0.9195
ij.same.gender	−0.4337	0.0494	−0.5326	−0.3349
ij.age.group.diff1	−0.3762	0.0557	−0.4877	−0.2647
ij.age.group.diff2up	−0.7730	0.0875	−0.9480	−0.5980
ij.same.municipality	0.2045	0.0491	0.1063	0.3028
alpha.0	−10.9630	0.0460	−11.0550	−10.8710
alpha.extBTL.J1	1.2115	0.0191	1.1734	1.2496
alpha.extBTL.J2	2.1777	0.0185	2.1407	2.2147
alpha.extBTL.J3	2.5643	0.0165	2.5313	2.5974
j.gender	0.1628	0.0130	0.1368	0.1888
j.age.group2	1.3740	0.0478	1.2784	1.4696
j.age.group3	1.6671	0.0457	1.5758	1.7584
j.age.group4	1.5396	0.0460	1.4476	1.6316
j.age.group5	0.5162	0.0482	0.4197	0.6126

Table 1: The table contains the estimated parameters of the model. The first half of the table reports the estimates of the parameters in the viral component of the model, the second half reports the estimates of the parameters that refer to the external pressure.

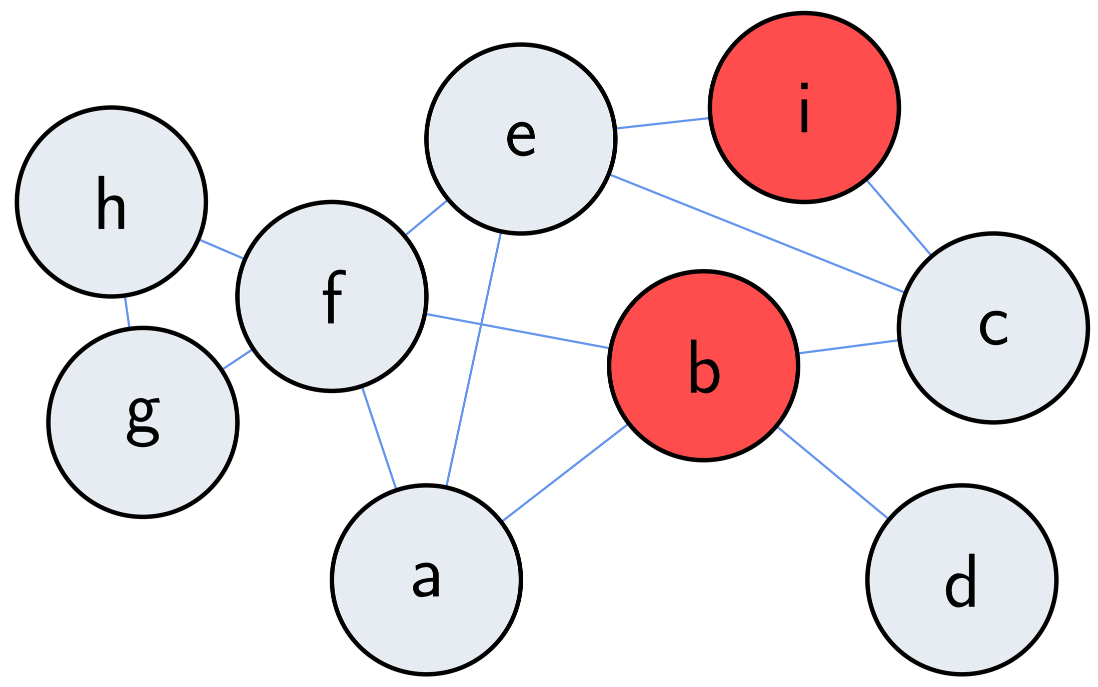

Figure 1: This simple network shows how nodes can get influenced by susceptible neighbours. For example, node **c** has two infectious nodes in its neighbourhood (**b** and **i**), whereas **g** has no infectious nodes around it. Hence, the model expects node **c** to adopt the product through the influence of nodes **b** and **i**, plus the external pressure, while **g** is only subject to external pressure.

Prediction

We also predicted the future number of adopters of the application by simulating future adoptions. After having estimated the parameters, it is possible to generate a simulated trajectory of product adoptions by extracting the next time-to-event and then assigning the event to a susceptible customer.
Each simulated trajectory is generated after sampling the parameters from a multinormal distribution, exploiting the approximation of maximum likelihood estimates. It is possible to assess uncertainty of future adoptions by simulating a fair amount of trajectories in this fashion, and then building point-wise confidence intervals.

Simulation

It is possible to show that, fixing a time point t, the next time-to-event is exponentially distributed:

$$\lambda(t) \sim \mathrm{Exp}\left[\sum_{j=1}^{n} \lambda_j(t)\right].$$

The probability p_k that the new adoption is associated with the k-th individual is then:

$$p_k = \frac{\lambda_k(t)}{\sum_{j=1}^{n} \lambda_j(t)}.$$

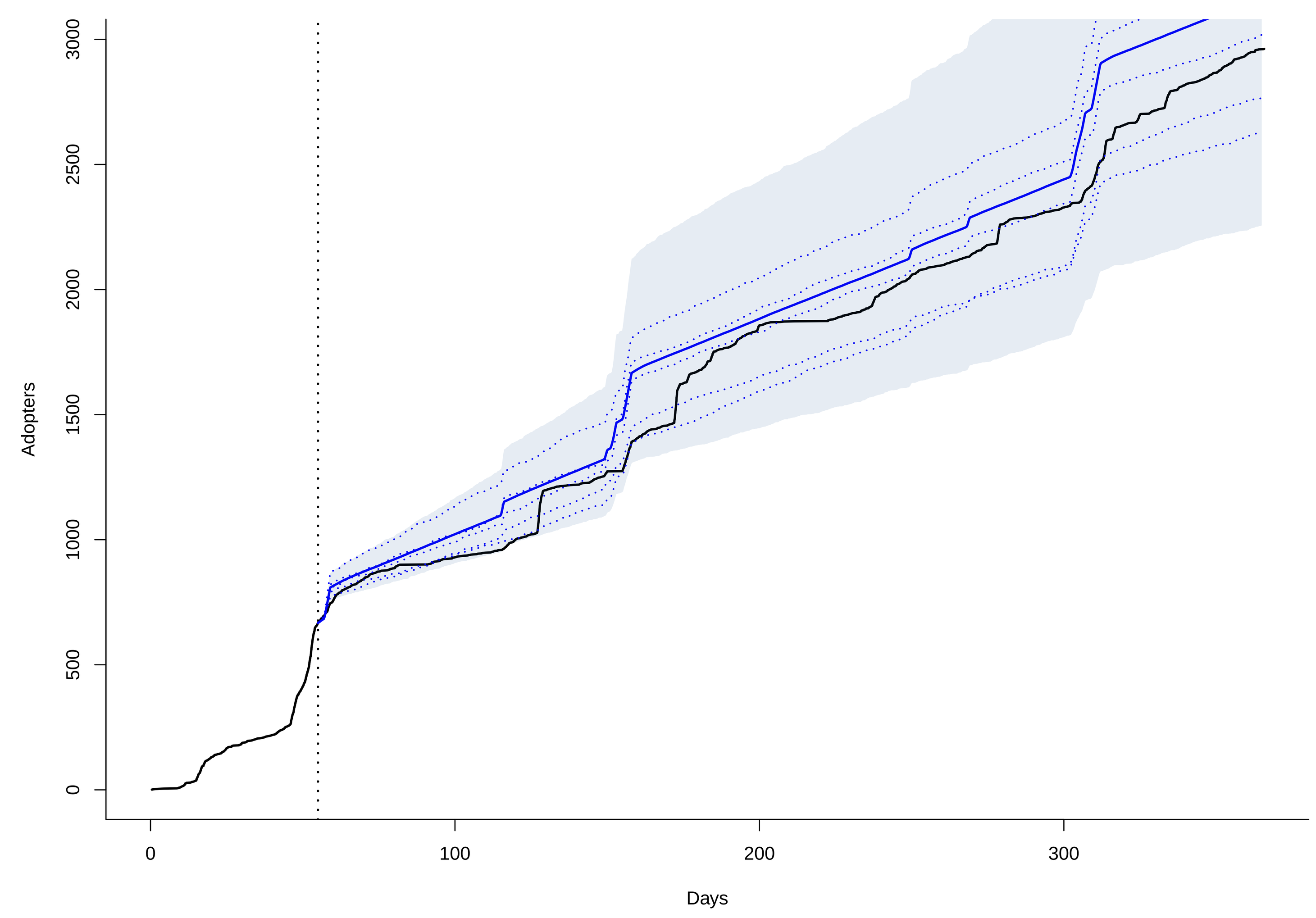

Figure 2: This figure shows some of the simulated trajectories (dashed blue lines) in regards to the actual observed data (black solid line). The solid blue line is obtained by taking the mean of all simulations. The shaded area is individuated through computing 90% point-wise confidence intervals, the dashed vertical line signals the end of the training period.

References

[1] Matt J Keeling, Mark EJ Woolhouse, Darren J Shaw, Louise Matthews, Margo Chase-Topping, Dan T Haydon, Stephen J Cornell, Jens Kappey, John Wilesmith, and Bryan T Grenfell. Dynamics of the 2001 uk foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817, 2001.

Contact Information

- Web: https://www.mn.uio.no/math/english/people/aca/riccarpa
- Email: riccarpa@math.uio.no
- Phone: +47 462 30 620