Influence of Neighbours on One’s Own Opinion

Author: Ivan Bestvina Mentor: Prof. Mile Šikić, PhD
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Electronic Systems and Information Processing

1. Introduction

From Brexit to Croatian constitutional referendum, and from US to French presidential election, recent history has showed us how social processes can behave in unexpected and seemingly chaotic ways. Results prediction based on various polling methods failed repeatedly, with some authors declaring „The Death of Polling”. Also, with the rise of online social networks, effects called *echo chambers* and *information bubbles* were reported – users were surrounded almost exclusively with like-minded people. Thus, studying social influence to uncover social network patterns became crucial if we were to better understand the processes evolving on these networks. In this sense, when we talk about the social influence, we refer to both the influence friends have over each other’s opinions, and the influence of one’s opinions on the choice of his or her friends.

2. Problem

After we determined the scale of social influence in various networks, mainly the one collected through a Facebook poll before the Croatian constitutional referendum, concluding that it is possible to predict one’s opinion based on his or her neighbourhood with around 95% accuracy, we defined a two-part problem: first, to find a way of modelling these types of polarized social networks, and second, to derive a method from these models of predicting the vote share in the population when only a small, biased sample is given

3. Method

To model polarized social networks, we extended the stochastic Kronecker graph model. Kronecker graphs are graphs produced by repeatedly applying the Kronecker product, defined in equation 1, on the graph’s adjacency matrix, as shown in figure 1.

$$K = A \otimes B = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,m}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \dots & a_{n,m}\mathbf{B} \end{bmatrix}$$

Eq. 1 – Definition of the Kronecker product

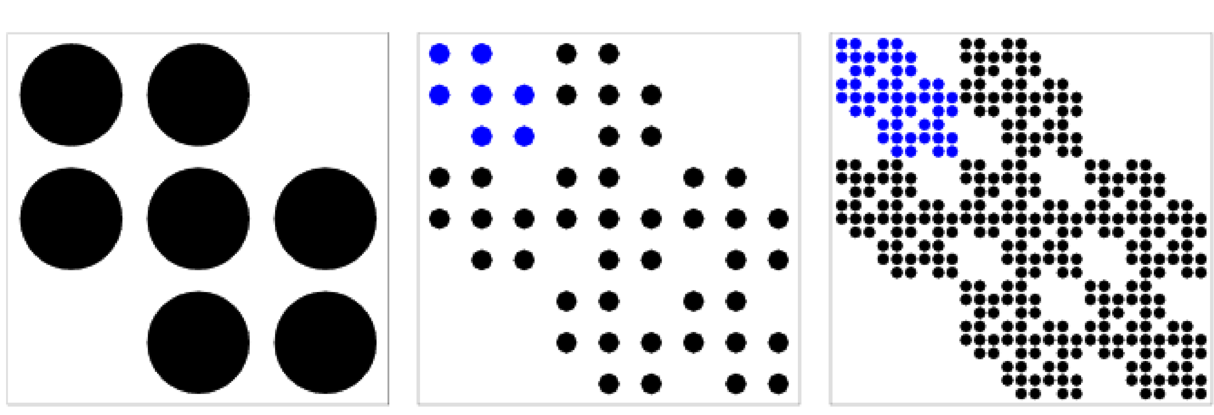

Fig. 1 – Original adjacency matrix with its second and third Kronecker power

Stochastic Kronecker graph model is the expansion of this, when instead of the initial adjacency matrix, small edge probability matrices are used. Once the k-th Kronecker power $K^{[k]}$ of the initial matrix K is calculated, each pair of nodes (i, j) is connected with probability $K_{i,j}^{[k]}$. This model provided us with great freedom, as efficient algorithms exist which fit the initial matrix to some given network, producing an infinite array of topologically similar, but random networks. Our expanded polarized Kronecker graph model is similar to this, but instead of $K^{[k]}$, it uses $K_0 \otimes K^{[k-1]}$. Matrix K_0 is the edge probability matrix of a homogenous two-cluster graph with different edge probabilities for edges inside clusters and between the clusters.

Next, we set out to find a way of inferring the vote share ratio of the population using only the small, biased sample of nodes and the information about their neighbourhoods. Because our samples are vote-biased, we based our estimations on two parameters (μ_a and μ_a), one for each group of voters (for voting options a and b respectively), representing the average ratio of neighbours who vote differently, and those who vote the same as the central node. This way, vote bias becomes unimportant, as each group is viewed separately. Solving a set of equations which arise from the model and the problem definition, we arrive at the estimator defined in equation 2.

$$Q(a) \approx \frac{-\mu_a + \sqrt{\mu_a \mu_b}}{\mu_a - \mu_b}$$

Eq. 2 – Estimator of vote share of option *a*.

4. Results

To test our method, we have simulated biased sampling choosing nodes with probabilities proportional to $\beta \cdot v(i) + (1-\beta)$, where β is the strength of bias, and $v(i)$ is the vote of node i, which is equal to 1 if node votes a, and 0 if it votes b. For each sample size we repeated the simulation 100 times to approximate the mean and variance of the estimation. In figure 2 we show the results for $\beta = 0.3, 0.8$, and for vote share of option a of 50% and 75%. Grey graph shows the trivial estimator which uses only the ratio of votes in the sample. Green graph shows the similar trivial estimator, but with sample expanded to all of the original sample’s neighbouring nodes. Blue graph shows our novel estimator.

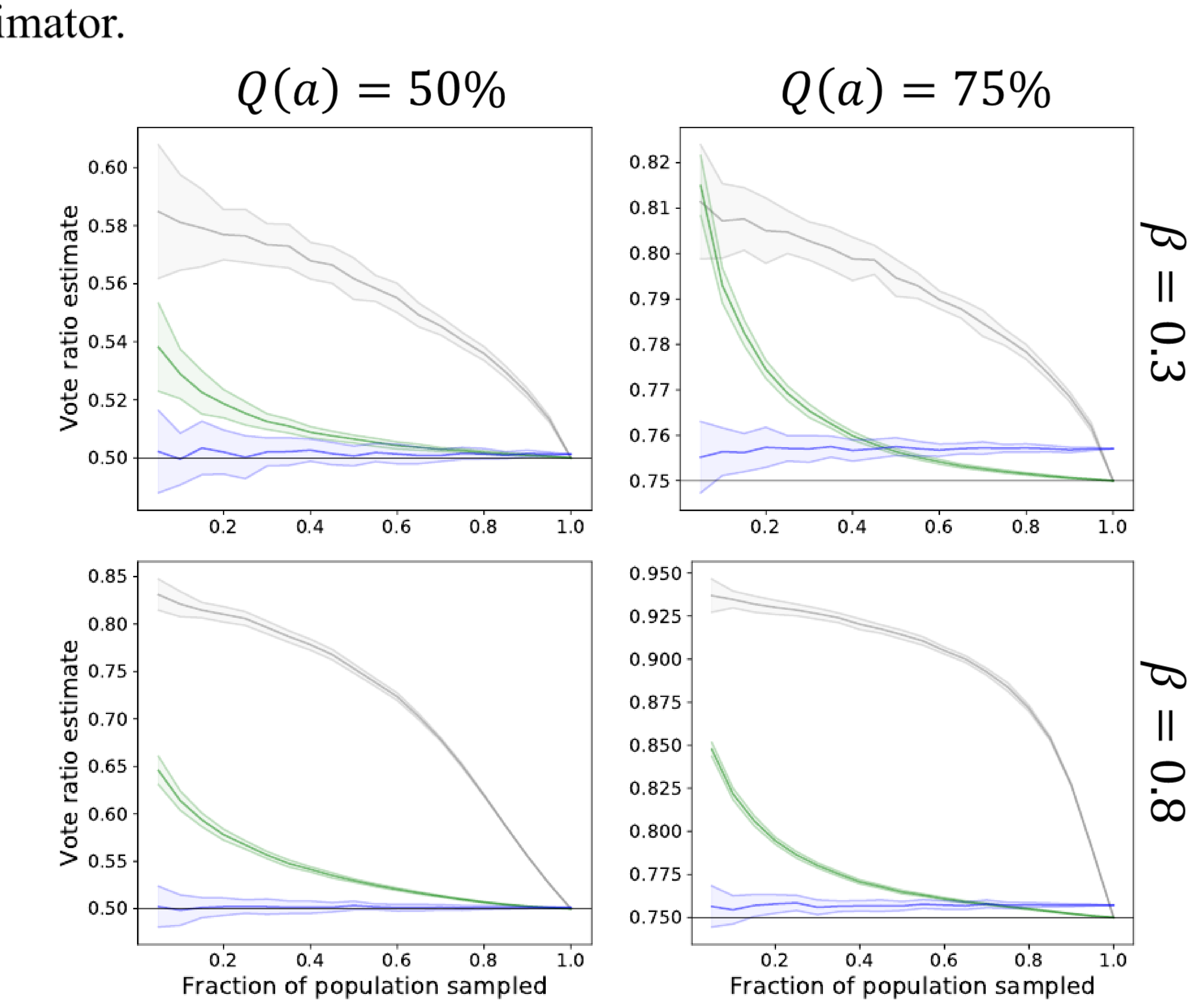

Fig. 2 – Results of vote share estimation with biased sampling

It is evident from these results that our estimator is not only superior to the trivial one, but also outperforms the trivial approach when all of the neighbourhoods of sampled nodes are given, and as such proves to be a very interesting and potentially fruitful direction of research.

5. Conclusion

Both polarized social network modelling and election prediction remain one of the most interesting problems in the research of social networks. In this work we have investigated a novel approach to both of these, and have shown how, by basing our estimates on parameters derived from the two sides of election separately, the usual problem of strong polling bias can be bypassed.

8. master thesis workshop, ZESOI, FER, Zagreb, 6 July 2017