Information Diffusion within Social Networks

Charanpal Dhanjal, Sandrine Blanchemanche, Stephan Clemençon,
Akos Rona-Tas, Fabrice Rossi

February 21, 2011

This work is motivated by the needs of state agencies which are responsible for managing various risks in social life issue advisories to the public to prevent and mitigate various hazards. We investigated how information about a common food born health hazard, known as Campylobacter, spreads once it was delivered to a random sample of individuals in France. The aim of the work is to study how properties of the individual and broader social network effect the diffusion of information within a simulated process. To this effect we modelled a social network using a set of characteristics of individuals over Erdos-Renyi [1] and Small World [3] random graph models. The social network was based on the data collected following a survey on Campylobacter conducted in France. Using this information we were also able to learn when information transmissions occurred between pairs of individuals, and could then simulate the diffusion process over the simulated network and measure certain properties of interest.

The results uncovered a strong predictability of information transmission, providing a balanced error rate of 0.092. The diffusion model was studied in the context of the effect of the network structure in the overall diffusion, and also how prediction was made for a particular social tie. Furthermore, graph visualisation methods were used in order to understand the largest connected component of the transmission graph.

In conjunction with this work we also considered a well known problem in percolation theory. In graph percolation, vertices within a graph have a binary states and a percolation process decides how activity spreads within the graph. Figure 1 demonstrates a percolation process within a graph. An interesting question about a percolation process within the social network is which k vertices should one choose in order to maximise the influence of the data. This question has been studied in [2] in the context of the Linear Threshold and Independent Cascade percolation models.

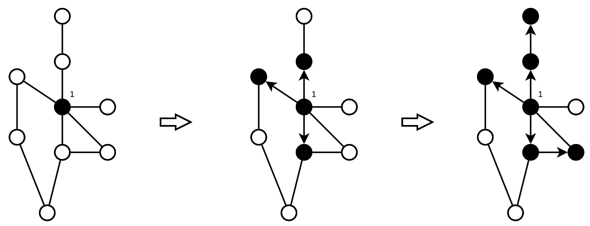

Figure 1: Percolation within a graph. Vertices in black are active and activation spreads to other vertices in an iterative manner. The initial active set is $A = \{1\}$ and 6 vertices are active at the end of the process in this case.

We generalise the problem by considering activations in $[0, 1]$, measuring the "quality" of percolation, and percolation decays along edges in the percolation graph. For a varying cost of activating each vertex, we maximise the total activation whilst keeping within a budget L. We further consider the unbudgeted version of the problem in which one simply wishes to select k initially active vertices. Both problems can be solved with greedy algorithms with a guaranteed approximation qualities, and furthermore we show a close connection to the maximal coverage problem. The unbudgeted algorithm is analysed empirically over predicted percolation graphs on a synthetic dataset and on the real dataset modelling information diffusion within a social network.

In conjunction with the research work conducted above, a significant amount of software has been written in Python for the experimental aspects. This code is available online via the following address (published in a package called Another Python Graph Library (APGL)) http://www.somethingaboutme.net/code2.php.

This work has resulted in the following publications:

- Charanpal Dhanjal, Sandrine Blanchemanche, Stephan Clemencon, Akos Rona-Tas, Fabrice Rossi, *Dissemination of Health Information within Social Networks*, to appear as a chapter in Networks in Social Policy Problems, Cambridge University Press, 2010
- Charanpal Dhanjal and Stephan Clemencon, Maximising the Quality of Influence, Accepted for publication in the *SIAM International Conference on Data Mining*, 2010
- Charanpal Dhanjal, *An Introduction to APGL*, Statistics and Applications Group, Telecom ParisTech, 2010

References

[1] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[2] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[3] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 293:420–442, 1998.