**Complexity Science Doctoral Training Centre: proposal for a PhD-project October 2012**

**Social networks and health**

**Supervisors:**

Thomas House, EPSRC Career Acceleration Fellow, Mathematics Institute

Frances Griffiths, Professor of Medicine in Society, Warwick Medical School

**External advisor:**

Robert Goudie, MRC Career Development Fellow, MRC Biostatistics Unit, Cambridge

**Research objectives**

1. To increase understanding of the role of social networks in adolescent health
2. To undertake secondary analysis of a health related database (Add Health) including social network data using existing network analysis methods
3. To integrate social network analysis and interaction of attributes of individuals within the network
4. To identify the limitations of standard applications of these methods and further develop analysis methods

**Why this is interesting.**

There is evidence from the social sciences that an individual’s health and health behaviours is influenced by their social relationships. For example, after the death of a spouse the risk dying increases when other known factors are controlled for. A young person’s social network influences their health-related behaviour (such as alcohol consumption and smoking) although the mechanisms seem to be complex and contingent, and the methods of data analysis used have been limited. Even highly publicised research on social networks and health, for example the ‘spread of obesity’ through social networks (Christakis and Fowler NEJM 2007), has been subject to considerable methodological criticism. Advances in our mathematical and statistical understanding of networks and network data (particularly in a Bayesian setting) are opening up the possibility of more robust analysis methods. Understanding the impact of social networks on health raises the possibility of novel interventions, which may be targeted at specific groups such as school attenders, or more generally in society or through social media.

Few health-related datasets include data about social networks beyond immediate family members. The Add Health dataset includes data about friendship networks among adolescents in large schools in the USA (see Add Health website at: <http://www.cpc.unc.edu/projects/addhealth>). This National Longitudinal Study of Adolescent Health started in 1994 and data continues to be collected. In includes data from around 90,000 school pupils, of whom around 2100 were asked to complete a detailed survey that included details of friendship networks, health and health behaviours within 14 schools. There are many published studies from this database but few using the social network data. At the University of Warwick we have experience of analysis of data from 3 waves of data collection which include the social network data and health variables. The data is ready for analysis. Robert Goudie is very familiar with the data and willing to provide guidance.

**Starting points for the PhD project**

After familiarisation with the Add Health data and the types of analysis already undertaken, the PhD student may choose to pursue the following.

Take the Exponential Random Graph model [1, 2] as a starting point. There are then a few inter-related possibilities:

(1) Carefully fitting a model of this class to data from Add Health. There are R-packages for doing this [3, 4], or this could be done from scratch through programming in any language. Health indicators can be included as covariates. This would differ from [5] by being very much focused on health aspects, rather than structural aspects. Handling the dyad covariates (ie information about frequency of contact) might be a possible extension.

(2) There are 14 schools with "complete" social network data, and so comparing the estimates across schools would be interesting, BUT the ERGM parameters are not invariant to network size, so this is not necessarily straightforward because the schools vary in size. A solution is proposed in [6], but there might be more to think about here in more complex settings. The problem can be investigated by forward simulation.

(3) If a satisfactory solution to (2) is straightforward (eg using the approach in [6]), a fully-Bayesian model hierarchical model for the parameters of the ERGM across schools would be intriguing. Standard MCMC is not applicable because ERGM involve an intractable normalising constant, and so this is a hard problem but there are methods that may well work.

[1] Snijders, T. A. B., Pattison, P., Robins, G. L., & Handcock, M. S. (2006). New Specifications for Exponential Random Graph Models. Sociological Methodology, 36(1), 99–153. <http://dx.doi.org/10.1111/j.1467-9531.2006.00176.x>

[2] Robins, G. L., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p\*) models for social networks. Social Networks, 29(2), 173–191. <http://dx.doi.org/10.1016/j.socnet.2006.08.002>

[3] <http://cran.r-project.org/web/packages/statnet/>

[4] <http://cran.r-project.org/web/packages/Bergm/>

[5] Goodreau, S. M., Kitts, J., & Morris, M. (2009). Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. Demography, 46(1), 103–125.

[6] Krivitsky, P. N., Handcock, M. S., & Morris, M. (2010). Adjusting for network size and composition effects in exponential-family random graph models. <http://arxiv.org/abs/1004.5328>