I believe that the authors have made a decent start at designing a potentially useful technique for clustering in static social networks but that users need to be aware that the technique is far from problem free and that they must be careful not to overcook the recipe and thereby to overinterpret results. It is my pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Anthony C. Atkinson (*London School of Economics and Political Science*) **and Marco Riani** (*Università di Parma*)
Over the years we have enjoyed both Adrian Raftery's talks and his flow of publications on model-based clustering. We would like to compare some results of the use of mclust with a cluster analysis that is produced by the use of the forward search.

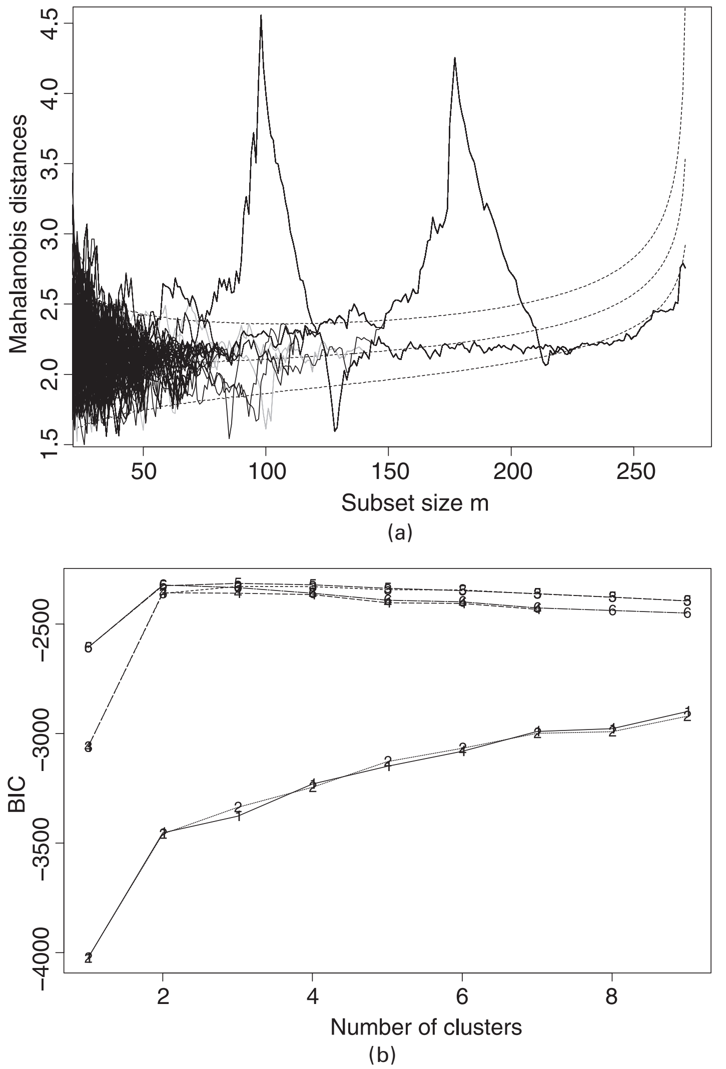

Fig. 9. Old Faithful Geyser data: (a) forward plot of minimum Mahalanobis distances from 300 random starts with 1%, 50% and 99% envelopes (two clusters are evident); (b) Bayes information criterion plot from mclust, indicating three clusters

The forward search for multivariate data is described in Atkinson *et al.* (2004). In general the search proceeds by successively fitting subsets of the data of increasing size. For a single multivariate population any outliers will enter at the end of the search with large Mahalanobis distances. If the data are clustered and the search starts in one of the, unfortunately unknown, clusters, the end of the cluster is indicated when the next observation to be added is remote from that cluster. To find clusters we have recently (Atkinson *et al.*, 2006a, b; Cerioli *et al.*, 2006) suggested running many searches from randomly selected starting-points. Some of these start in, or are attracted to, a single cluster; a forward plot of the minimum Mahalanobis distance of the observations that are not in the fitted subset then reveals the cluster structure.

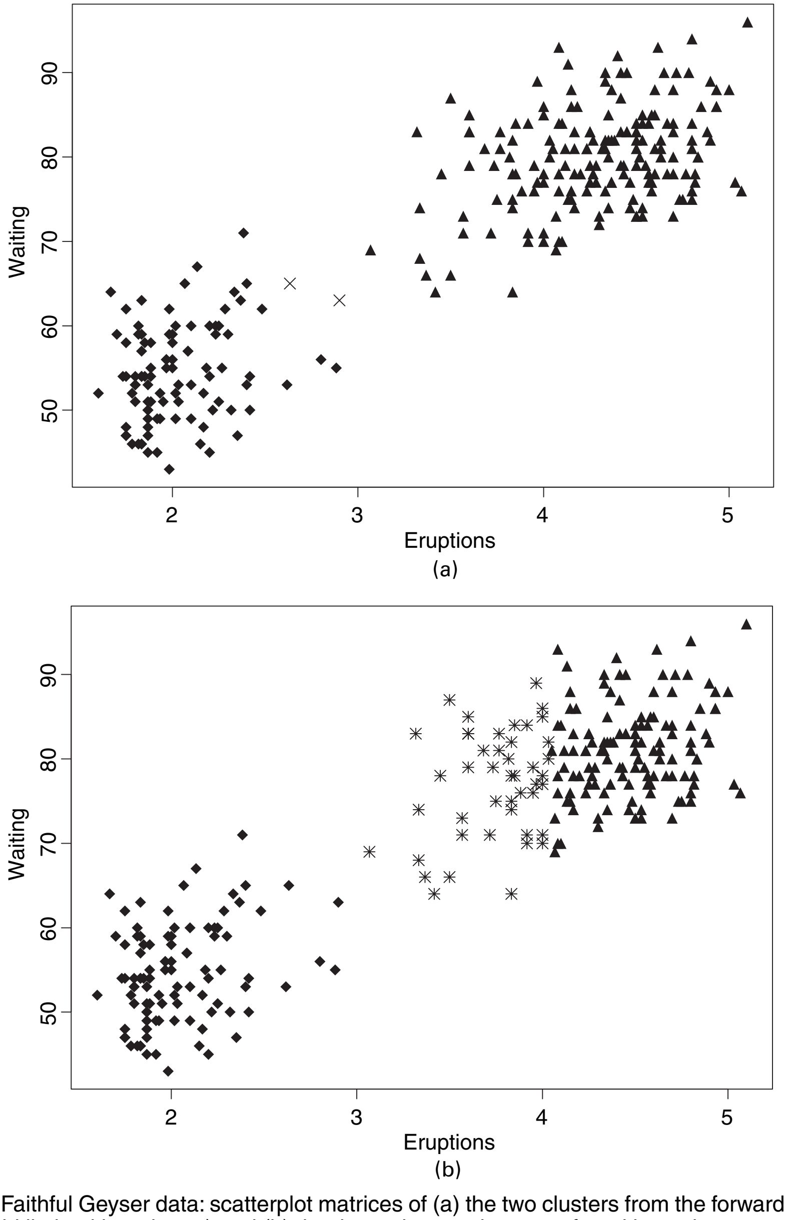

Fig. 10. Old Faithful Geyser data: scatterplot matrices of (a) the two clusters from the forward search (units marked × could lie in either cluster) and (b) the three clusters that were found by mclust

As an example we analyse 272 observations on the eruptions of the Old Faithful Geyser taken from the Modern Applied Statistics in S library (Venables and Ripley, 2002). Azzalini and Bowman (1990) described the scientific problem. Fig. 9(a), a forward plot of Mahalanobis distances from the forward search, clearly shows the two groups. Fig. 9(b) is the Bayes information criterion output of `mclust` from S-PLUS which, on the contrary, indicates three clusters. Fig. 2 of Fraley and Raftery (2006) for a slightly different set of geyser data is similar and again indicates three clusters.

We use further forward searches to establish membership of these two clusters and establish the unclustered units. A scatterplot of the resulting two clusters is shown in Fig. 10(a). The three clusters that are found by `mclust` are in Fig. 10(b). Fuller details of our analysis including further comparisons and considerations of robustness are in Atkinson and Riani (2007).

We do not want to imply that the imposing edifice, some of whose rooms we have so enjoyably visited today, is built on sand. But it does seem that there are still some fundamental problems in the foundations of clustering that need to be resolved.

Isobel Claire Gormley (*University College Dublin*) **and Thomas Brendan Murphy** (*Trinity College Dublin*)
We congratulate the authors on a thought-provoking paper. We feel that the combination of model-based clustering with latent space modelling is applicable far beyond the proposed application of the analysis of social network data.

We have recently been developing statistical models for rank data including data from Irish elections (Gormley and Murphy, 2005, 2006a) and Irish college applications (Gormley and Murphy, 2006b). Two approaches that we have taken are using mixture models (Murphy and Martin, 2003; Gormley and Murphy, 2005, 2006b) and using a latent space model (Gormley and Murphy, 2006a). The paper that is presented here provides a combination of both of these modelling approaches which we look forward to applying to the analysis of rank data.

In our work, we found that model choice is a difficult aspect of the modelling process. The number of components in a mixture model can be estimated consistently by using the Bayes information criterion (Keribin, 2000) but we found that the choice of dimensionality in our latent space model is more problematic. We were wondering whether the authors could provide us with insight into the methods for choosing the dimensionality of the latent space in their social network model.

More recently, we have considered methods for including covariates in our models. One approach that we have considered is allowing the mixture probabilities to depend on covariates (Gormley, 2006); this yields a special case of the mixture-of-experts model (Jacobs *et al.*, 1991). This model can be fitted very easily with minor changes to the mixture modelling framework. In the context of this paper this may provide an alternative method for achieving homophily by attributes.

Trevor Sweeting (*University College London*)
I would be interested to hear from the authors whether they have considered using an infinite group cluster model and, if so, what they would consider to be the relative advantages and disadvantages of such a formulation over their finite group cluster model in the context of network models. There are various possible Bayesian formulations of infinite group cluster models. A common choice of prior distribution for the group weights λ arises from a Dirichlet process prior structure for the parameters of the latent positions, since this structure automatically induces clustering. Specifically, writing $\phi_i = (m_i, s_i^2)$ the Dirichlet process mixture (DPM) structure for the latent positions would be specified as

$$z_i|\phi_i \sim \mathrm{MVN}_d(m_i, s_i^2 I_d),$$

$$\phi_i|F \overset{\mathrm{IID}}{\sim} F,$$

$$F \sim \mathrm{DP}(F_0, \gamma),$$

$$F_0 = \mathrm{MVN}_d(0, \omega^2 I_d) \times \sigma_0^2 \, \mathrm{Inv}\chi_\alpha^2.$$

Here DP$(\cdot, \cdot)$ denotes a $(d+1)$-dimensional Dirichlet process and F_0 and γ are the associated mean and precision parameters. Now, for $g = 1, 2, \ldots$, write $\theta_g = (\mu_g, \sigma_g^2)$ and $\theta = (\theta_1, \theta_2, \ldots)$. Using Sethuraman's (1994) stick breaking representation of the Dirichlet process, the above specification is equivalent to the following infinite group version of the authors' model:

$$z_i|\theta \overset{\mathrm{IID}}{\sim} \sum_{g=1}^{\infty} \lambda_g \, \mathrm{MVN}_d(\mu_g, \sigma_g^2 I_d),$$