Social Networks 38 (2014) 74–87

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Statistical modelling of the group structure of social networks

Murray Aitkin[a,*], Duy Vu[b], Brian Francis[c]

[a] Department of Mathematics and Statistics, University of Melbourne, Victoria, Australia
[b] Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia
[c] Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YL, United Kingdom

ARTICLE INFO

Keywords:
Social groups
Latent classes
Bayesian model comparison
Natchez women

ABSTRACT

This research evaluates the identification of group structure in social networks through the latent class model and a new Bayesian model comparison method for the number of latent classes. The approach is applied to a well-known network of women in Natchez Mississippi. The latent class analysis reproduces the group structure of the women identified by the original sociologists.

1. Introduction

This paper investigates a statistical model for groups in a social network, which until recently has not been a major focus of analysis in the field. A recent major review of statistical modelling work in the field has been published by Goldenberg et al. (2009). In their summary Chapter 6 they conclude:

> Despite the many advances in network modeling over the last decade, there remains a host of unresolved issues. . .. We feel that, from a statistics or machine learning perspective, the biggest breakthroughs are to be made in the areas of inference and dynamic modeling. Creating a model or perhaps fixing an existing one in such a way that provides realistic generative and inference mechanisms which can identifiably infer parameters of a large real world network would make a great contribution to the statistical network modeling community.

This paper addresses the identification of *actor* groups within the framework of a two-mode or bipartite network of *actors attending events*, through a statistical model in which the groups of actors are represented by *latent classes*, which are not directly observable, but which can be probabilistically reconstructed from the event attendance patterns of the actors.

In this process several questions are of critical importance:

- how many groups can be identified;
- the nature of the membership of the groups, for example whether individuals belong to one or to many groups;
- the nature of the event attendance patterns in the groups;
- whether other non-latent class models might give a better representation of the data.

We address these questions in a Bayesian framework, and use recent developments in Bayesian model comparisons to illuminate the choice among possible models. To show the application of the approach we use a famous data set from Davis et al. (1941) analysed many times, as reported in Freeman (2003).

2. The Natchez women network

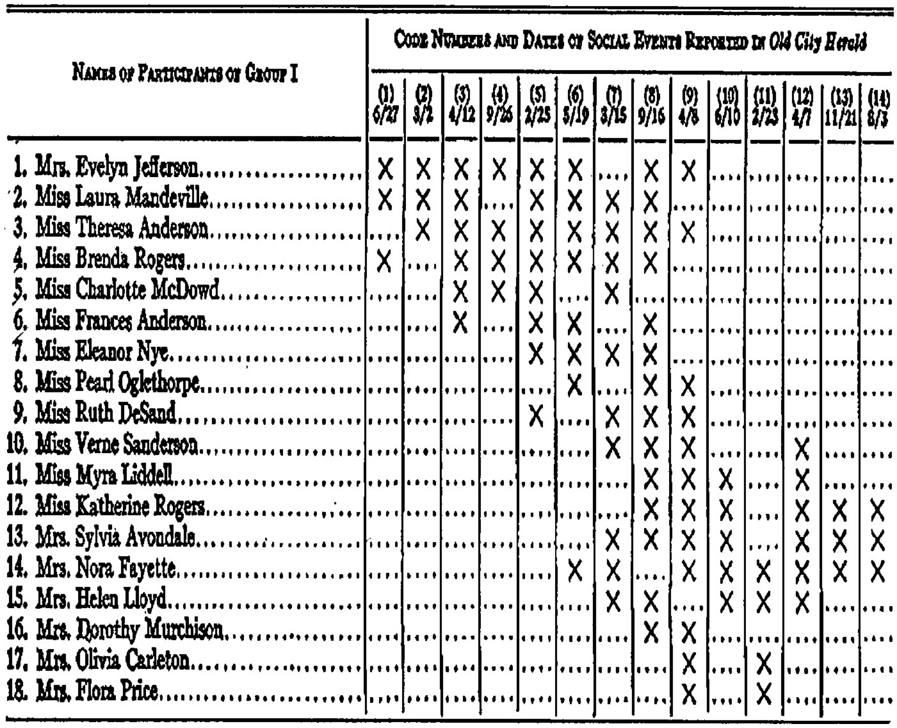
We give a detailed discussion of a simple social network which has attracted a remarkable amount of interest, and a wide variety of approaches (21 different analyses are reported and compared in Freeman, 2003).

It comes from a sociological study of social interactions among women in Natchez, Mississippi in the 1930s, reported in Davis et al. (1941, hereafter DGG). The book reported a comparative study of social class in black and white society. One aspect of the study was to assess the formation or existence of "cliques", defined by the joint participation of groups of women in attending common events. The network table (Fig. 1) which has caused so much interest to later analysts is reproduced from Davis et al; it gives the presence (×) or absence (. . .) of 18 women at 14 events. The women are named and numbered by DGG, and the events are dated from newspaper reports at the time.

The question of interest to analysts is how to describe the nature of the association among the women, and in particular to identify, as far as possible, subsets of the women which form coherent groups or cliques, using only their attendance at the events as data.

* Corresponding author at: Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia. Tel.: +61 4 0722 2040; fax: +61 3 9572 3016.
E-mail addresses: Murray.Aitkin@unimelb.edu.au (M. Aitkin), Duy.Vu@unimelb.edu.au (D. Vu), B.Francis@lancaster.ac.uk (B. Francis).

http://dx.doi.org/10.1016/j.socnet.2014.03.002
0378-8733/



Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	X	X	X	X	X	X	....	X	X	....	....	....	....	....
2. Miss Laura Mandeville	X	X	X	....	X	X	X	X	....	....	....	....	....	....
3. Miss Theresa Anderson	....	X	X	X	X	X	X	X	X	....	....	....	....	....
4. Miss Brenda Rogers	X	....	X	X	X	X	X	X	....	....	....	....	....	....
5. Miss Charlotte McDowd	....	....	X	X	X	....	X	....	....	....	....	....	....	....
6. Miss Frances Anderson	....	....	X	....	X	X	....	X	....	....	....	....	....	....
7. Miss Eleanor Nye	....	....	....	....	X	X	X	X	....	....	....	....	....	....
8. Miss Pearl Oglethorpe	....	....	....	....	....	X	....	X	X	....	....	....	....	....
9. Miss Ruth DeSand	....	....	....	....	X	....	X	X	X	....	....	....	....	....
10. Miss Verne Sanderson	....	....	....	....	....	....	X	X	X	....	....	X	....	....
11. Miss Myra Liddell	....	....	....	....	....	....	....	X	X	X	....	X	....	....
12. Miss Katherine Rogers	....	....	....	....	....	....	....	X	X	X	....	X	X	X
13. Mrs. Sylvia Avondale	....	....	....	....	....	....	X	X	X	X	....	X	X	X
14. Mrs. Nora Fayette	....	....	....	....	....	X	X	....	X	X	X	X	X	X
15. Mrs. Helen Lloyd	....	....	....	....	....	....	X	X	....	X	X	X	....	....
16. Mrs. Dorothy Murchison	....	....	....	....	....	....	....	X	X	....	....	....	....	....
17. Mrs. Olivia Carleton	....	....	....	....	....	....	....	....	X	....	X	....	....	....
18. Mrs. Flora Price	....	....	....	....	....	....	....	....	X	....	X	....	....	....

Fig. 1. DGG Table 1.

It would help this identification if we knew more about the social events than just their dates, but no further information about them is given. We do know that these events were not the unique social events of the reported days, as DGG report that other women also attended events on the same days, but not the events in this table.

2.1. The adjacency matrix

To perform any analysis we express the table elements mathematically through the link or tie variable Y_{ij}, with the presence of woman i at event j defining $Y_{ij} = 1$, and her absence from the event defining $Y_{ij} = 0$. We use n to denote the number of rows – women – and r to denote the number of columns – events. The resulting table, shown in Table 1 is the adjacency matrix, denoted by $\mathbf{Y}$.

Marginal totals (T) have been added to the table, giving the total number of events attended by each woman, and the total number of women attending each event.

From the adjacency matrix $\mathbf{Y}$ we can construct two other *symmetric* matrices of interest: the $n \times n$ "woman-by-woman" matrix $\mathbf{W} = \mathbf{YY}'$, and the $r \times r$ "event-by-event" matrix $\mathbf{E} = \mathbf{Y}'\mathbf{Y}$.

These "one-mode" networks show the connections between pairs of women through the numbers of events they have jointly attended, and the connections between pairs of events through the numbers of women jointly attending them.

The woman-by-woman matrix $\mathbf{W}$ is given in Table 2; its diagonal elements are the numbers of events attended by each woman.

The event-by-event matrix $\mathbf{E}$ is given in Table 3. The diagonal elements of $\mathbf{E}$ are the numbers of women attending each event.

Table 1
Event attendance.

W	*E*														T
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	8
2	1	1	1	0	1	1	1	1	0	0	0	0	0	0	7
3	0	1	1	1	1	1	1	1	1	0	0	0	0	0	8
4	1	0	1	1	1	1	1	1	0	0	0	0	0	0	7
5	0	0	1	1	1	0	1	0	0	0	0	0	0	0	4
6	0	0	1	0	1	1	0	1	0	0	0	0	0	0	4
7	0	0	0	0	1	1	1	1	0	0	0	0	0	0	4
8	0	0	0	0	0	1	0	1	1	0	0	0	0	0	3
9	0	0	0	0	1	0	1	1	1	0	0	0	0	0	4
10	0	0	0	0	0	0	1	1	1	0	0	1	0	0	4
11	0	0	0	0	0	0	0	1	1	1	0	1	0	0	4
12	0	0	0	0	0	0	0	1	1	1	0	1	1	1	6
13	0	0	0	0	0	0	1	1	1	1	0	1	1	1	7
14	0	0	0	0	0	1	1	0	1	1	1	1	1	1	8
15	0	0	0	0	0	0	1	1	0	1	1	1	0	0	5
16	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
17	0	0	0	0	0	0	0	0	1	0	1	0	0	0	2
18	0	0	0	0	0	0	0	0	1	0	1	0	0	0	2
T	3	3	6	4	8	8	10	14	12	5	4	6	3	3	89