Mapping Forced Marriage in Nottinghamshire using Comparative Judgement

University of Nottingham Rights Lab

Yichi Zhang[1] | **Rowland Seymour (Supervisor)**[2]

[1]Mathematical Sciences, University of Nottingham, UK [2]Rights Lab, University of Nottingham, UK

Introduction

A forced marriage is where one or both people do not or cannot consent to the marriage, and pressure or abuse is used to force them into marriage. Using comparative judgements made by 12 experts, we map the risk of forced marriage at ward level in Nottinghamshire based on the Bradley-Terry model. To explore further the comparisons made and the risk of forced marriage obtained, this project contains the relative reliability of judges, the influence of time spent on comparisons, and ward clusters with geographical information.

Comparative judgements

We assign to each ward what we call a *risk of forced marriage parameter* $\lambda_i \in \mathbb{R}$ and infer the value of each parameter using a comparative judgement model using the Bradley–Terry model. If areas i and j are compared n_{ij} times, the number of times area i is judged to be less likely to suffer from forced marriage than area j is modelled as

$$Y_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij}),$$

and we assume Y_{ij} are independent. Here the probability π_{ij} that area i is judged to be more affluent than area j depends on the difference in relative deprivation of i and j and is

$$\text{logit}(\pi_{ij}) = \lambda_i - \lambda_j \iff \pi_{ij} = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + \exp(\lambda_j)}$$

The likelihood function for the model is given by

$$\pi(\boldsymbol{y} \mid \lambda_1, \ldots, \lambda_N) = \prod_{i=1}^{N} \prod_{j<i} \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{n_{ij} - y_{ij}}.$$

The map below shows our estimated risk of forced marriage in Nottinghamshire, with red areas having the highest risk and blue the lowest.

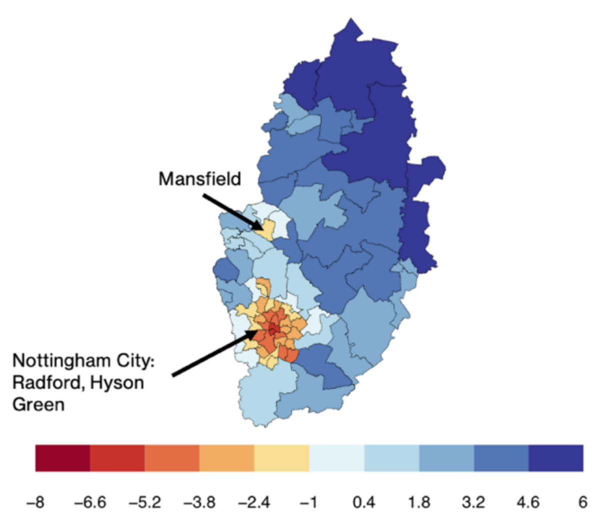

Figure 1: Estimated risk of forced marriage in Nottinghamshire

Reliability of judges

To determine the reliability of the judges, for each judge, we see how the observed comparisons differ from the expected comparisons based on the fitted model. The χ^2 statistics from Goodness-of-Fit test for each judge (j) is given by

$$\chi_j^2 = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{(O_{j,i} - E_{j,i})^2}{E_{j,i}}$$

where N_j is the number of comparisons judge j made, and $O_{j,i}$ and $E_{j,i}$ are the observed and expected outcomes of judge j's i-th comparison. There are 2 judges (30&54) with significantly high χ^2 statistics indicating they might not be trustworthy or they have a unique perspective on certain ward situations.

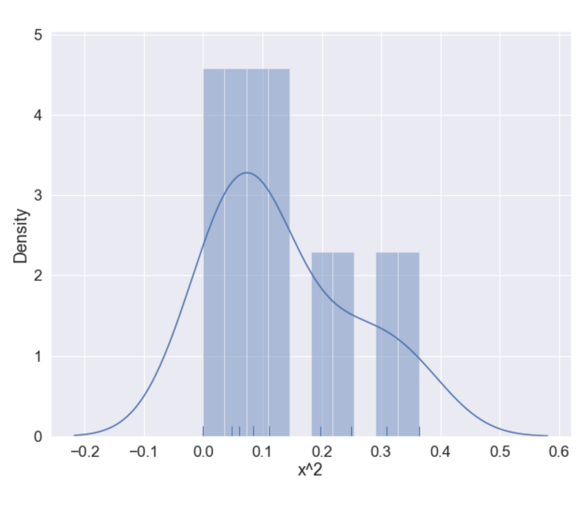

Figure 2: The distribution of χ^2 statistics for each judge

Time on comparisons

By subtracting the time for 2 consecutive comparisons, we calculate the time spent on every comparison. Ranging from less than 1 second to approaching an hour, the time spent on each judgement concentrates between 2.5 to 6 seconds and has an outlier threshold at about 11 seconds. Then, we find a negative correlation between the number of comparisons and the percentage of long-time decisions each judge made.

To reduce the effect of hasty judgements on the final results, we weight the comparisons from 0 to 1 by how long they took and fit the Bradley-Terry model again. The following chart shows how comparisons under 11 seconds are weighted, while those decisions took more than 11s are given weight 1

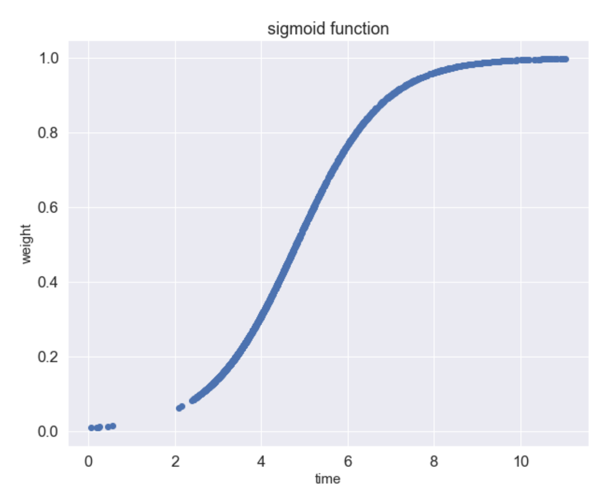

Figure 3: Logistic curve for weighting comparisons

The risk levels from weighted data give an extremely high Spearman correlation coefficient with the original ones, suggesting judgements' quality varies negligibly when the amount of time used in the changes..

Clustering Analysis

Each ward is allocated a representative point whose x and y coordinates are regarded as 2 normal features. These 2 geographical features, along with the risk levels, are inputted into various clustering models comprising K-means, DBSCAN (density-based spatial clustering of applications with noise) and GMM (Gaussian Mixture Model). Since K-means can only provide rigid boundaries to separate clusters, we expected DBSCAN and GMM to output more indicative results.

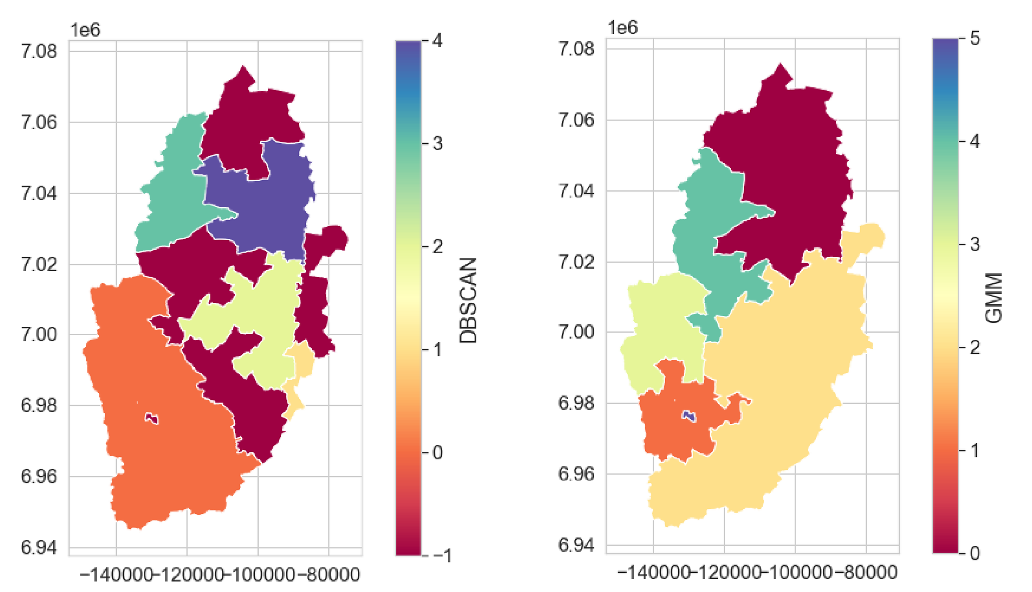

Figure 4: Clustering results from DBSCAN (left) and GMM (right).

As for DBSCAN, the wards with label -1 are the outliers that do not belong to any clusters, and we can see except the center of Nottingham city, the southwestern part of the shire is clustered together. However, in the case of GMM, Southeast wards are clustered into a group.

Since city centers are typically densely packed with small areas, with peri-urban and rural areas being larger, using Euclidean metric to cluster might be unsuitable. Hence, we first constructed a network for the wards and then defined a metric based on the shortest path length between 2 wards to iterate clusters.

References

[1] Rowland G. Seymour, David Sirl, Simon P. Preston, Ian L. Dryden, Madeleine J. A. Ellis, Bertrand Perrat, and James Goulding. The bayesian spatial bradley–terry model: Urban deprivation modelling in tanzania. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(2):288–308, 2022.