

Introduction

- ▶ We propose a statistical model to predict opinions based on social network structure.
- ▶ Our analysis is conducted on a unique data set provided by the Columbia University module of the *Cooperative Congressional Election Study (CCES)*.
- ▶ For the estimation of the social network structure, we present a variation of the overdispersed Poisson regression model introduced by Zheng et al. (2006) to be applied on an interval-based data set.
- ▶ We use these estimates to predict respondents' opinions about contemporary political issues.

Data and methods

- ▶ 1000 internet users participated in October and November 2006.
- ▶ Various questions are asked:
 - ▶ Socio-economic and political predictors: income, gender, race, age, education, ideology, party identification
 - ▶ "How many X's do you know?", where X takes 13 different values (6 names, e.g. Brenda, Kevin, and 7 groups, e.g. police officers, unemployed people.). The responses are collected as interval data (0, 1, 2-5, 6-10, more than 10 people).
 - ▶ Political opinions (for example, "Would you vote for or against a plan to start withdrawing troops this year?")

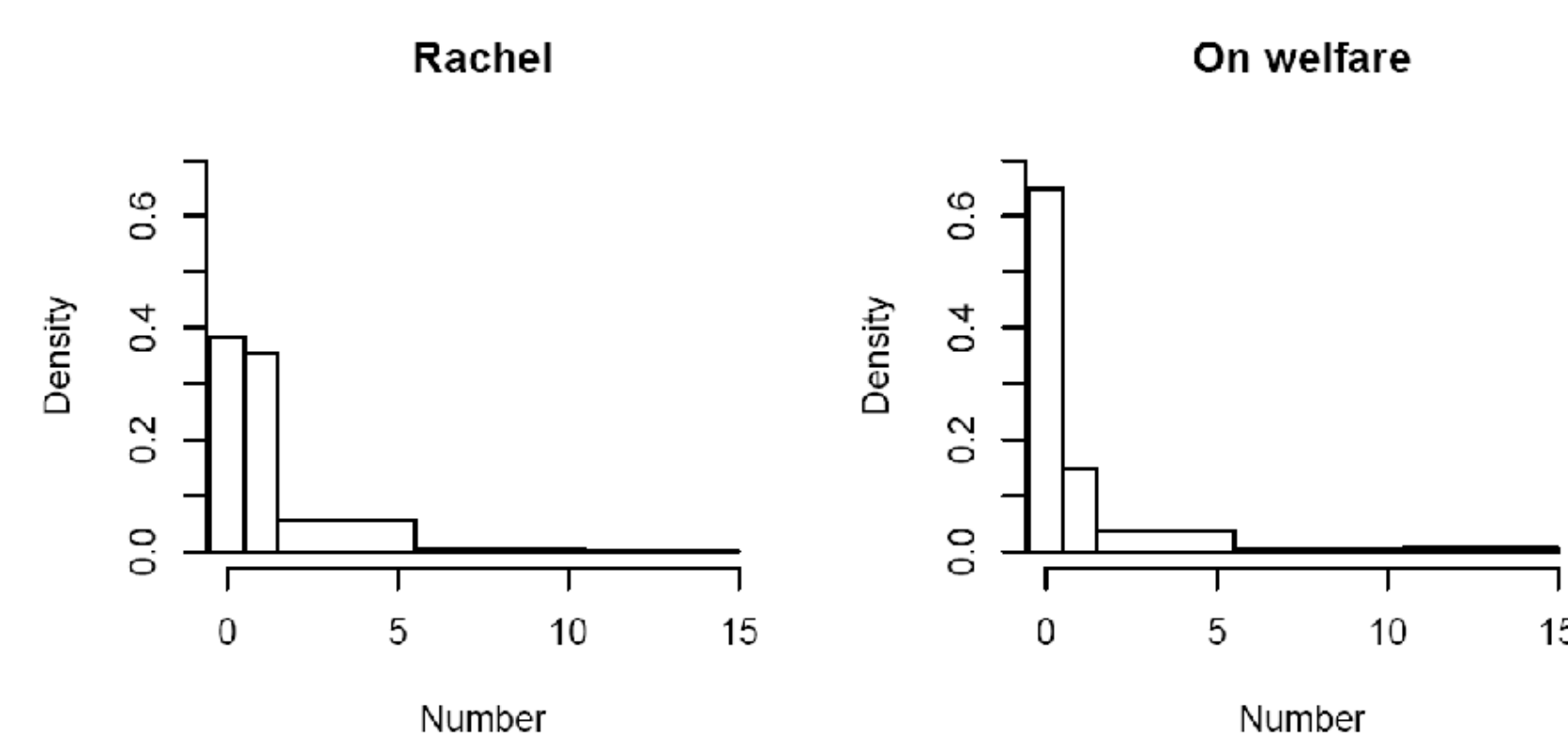


Figure: The fraction of respondents who know either no one or exactly one person is much higher for welfare than for Rachel, indicating overdispersion.

- ▶ We use an overdispersed Poisson model.
- ▶ $v_{ik} \sim \text{Poisson}(a_i b_k g_{(i,k)})$
 - ▶ v_{ik} , in the k^{th} group known by the i^{th}
 - ▶ a_i : degree of person i
 - ▶ b_k : proportion of links which involve group k
 - ▶ $g_{(i,k)}$: drawn from Gamma distribution with shape parameter $1/(w_k - 1)$, representing overdispersion in group k .
- ▶ We do not observe v_{ik} but only interval data y_{ik} .
- ▶ We use normal priors on the log of degree and on the proportion of links along with diffuse uniform priors on hyperparameters.

Estimating degree and overdispersion

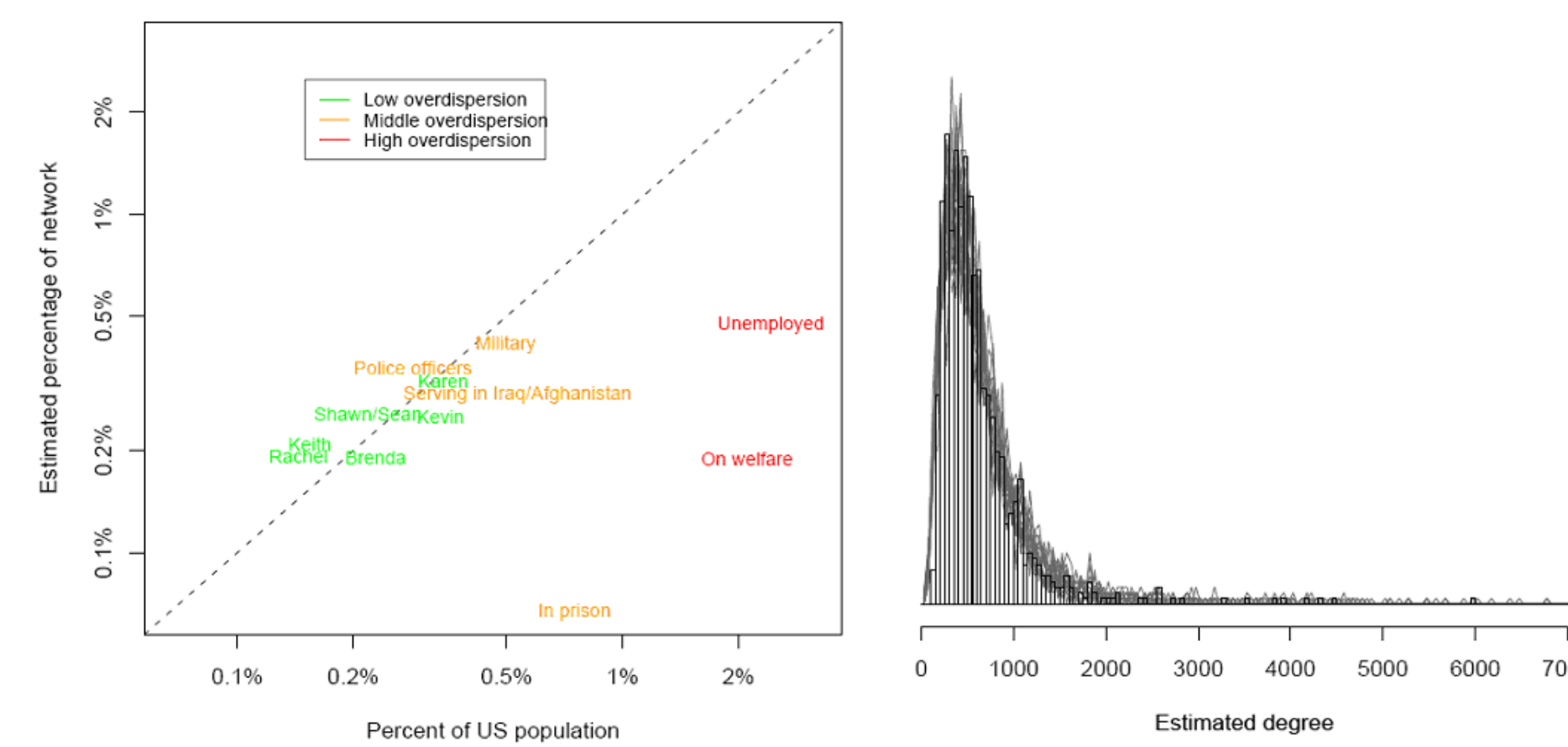


Figure: The left plot shows on the log-scale the actual and estimated group sizes, b , for the 12 groups where the true sizes are known. Since we standardize the model based on the 6 names, we expect them to lie on the diagonal. The right plot shows a histogram of estimated degree distribution a .

Using residuals to predict opinions

- ▶ $r_{i,k} := \sqrt{y_{i,k}} - E(\sqrt{Y_{i,k}})$, where $Y_{i,k} = \mathbf{1}_{\{v_{i,k}=1\}} + 3.5 \times \mathbf{1}_{\{2 \leq v_{i,k} \leq 5\}} + 8 \times \mathbf{1}_{\{6 \leq v_{i,k} \leq 10\}} + 15 \times \mathbf{1}_{\{11 \leq v_{i,k}\}}$.
- ▶ The expectation is taken under the *null model*, (without overdispersion); a positive residual means individual i knows more people in group k than the null model predicts for a person with that degree.
- ▶ Degree is generally not a significant predictor of the opinions used in this survey.
- ▶ Standardized coefficients for residuals and counts are similar.

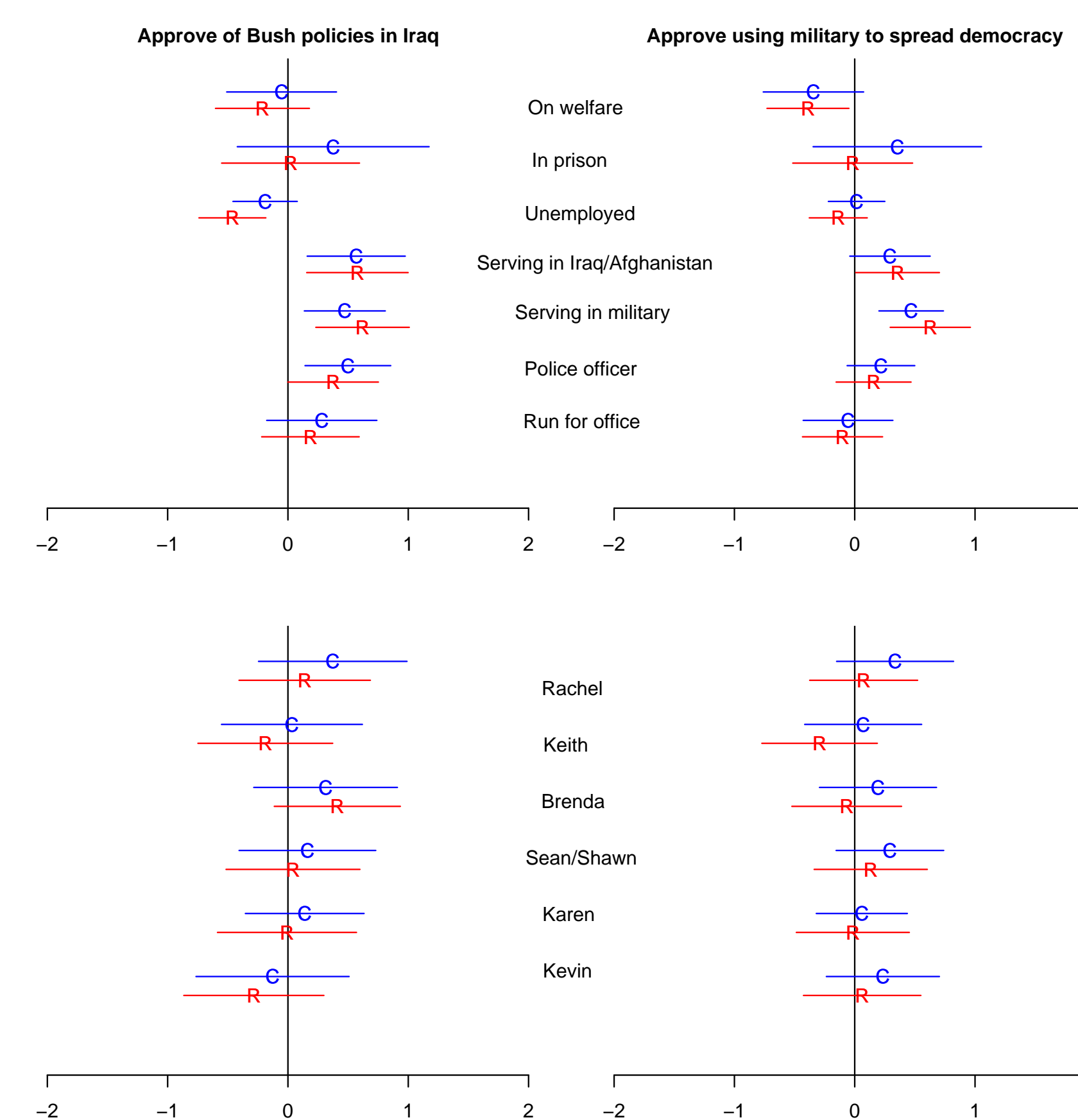


Figure: This figure compares standardized coefficient estimates for models using $y_{i,k}$, the raw counts (C), and the residuals (R) defined above.

Correlation in residuals

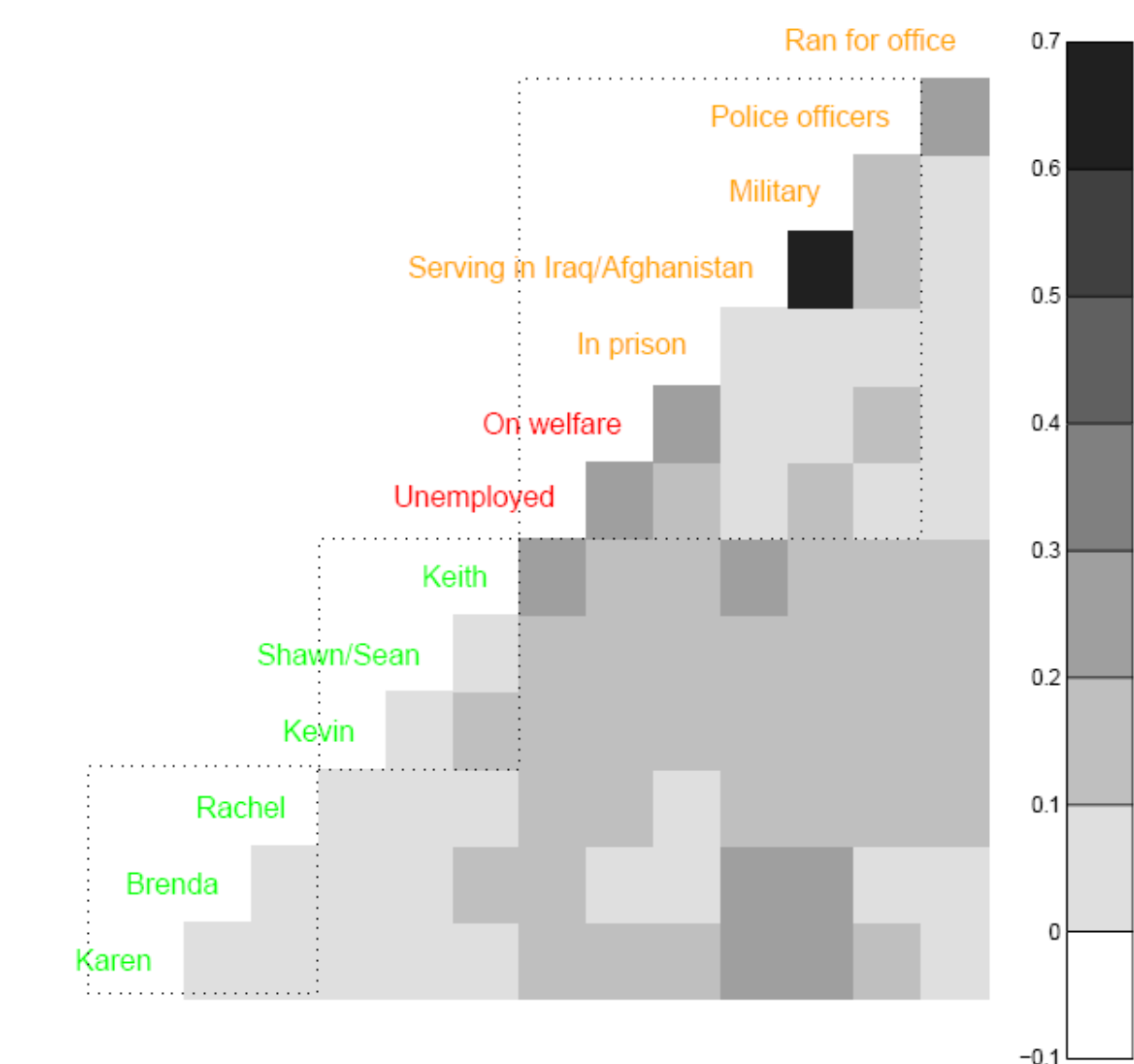


Figure: This plot shows the correlations of the residuals, $r_{i,k}$. The groups are combined in clusters of male names, female names, and other groups. Within each cluster they are ordered by real size, the larger group being on the lower left side. The true value for "Ran for office" is unknown. Green represents low, yellow represents middle, and red represent high overdispersion, respectively.

Conclusion

- ▶ We regress 25 different opinions (such as belief in the existence of weapons of mass destruction in Iraq or support of stem cell research) on the raw counts, $y_{i,k}$, and residuals, $r_{i,k}$.
- ▶ Controlled for socio-economic and political factors.
- ▶ In 22 out of these 25 opinions responses $y_{i,k}$ or residuals $r_{i,k}$ are significant for at least one subpopulation and improve the prediction.
- ▶ Knowing more individuals serving in the military or Iraq/Afghanistan is associated positively with approval of Bush's policies in Iraq and negatively with opinions of plans for troop withdrawal.
- ▶ Knowing armed service members is positively associated with supporting war on ideological grounds (spreading democracy) but not for concrete situations (supporting allies, protecting oil supply).
- ▶ Connections with these subpopulations were not important predictors of opinions on social issues (supporting tax cuts, minimum wage, stem cell research, late-term abortion).

Reference

Zheng, T., Salganik, M.J., and Gelman, A. (2006). "How many people do you know in prison?: using overdispersion in count data to estimate social structure in networks", *Journal of the American Statistical Association* 101, pp. 409-423.