sociological science

Supplement to:

Park, Barum. 2021. "Segregated in Social Space: The Spatial Structure of Acquaintanceship Networks." Sociological Science 8: 397-428.

Contents

Considerations on the Minimum Ultrametric

The "minimum ultrametric" was proposed in Watts et al. (2002) to study the nature of weak ties in spanning long distances in social space as well as the small world phenomenon (Watts 1999) that results from such ties. The rationale behind employing the minimum ultrametric was derived from the small world experiments (Travers and Milgram 1969), in which individuals were able to transmit a letter to an unknown and geographically distant person through a short chain of acquaintances. The mechanism that made this possible was, arguably, that individuals used only one social dimension—such as the occupation or region of residence of the recipient—on which they thought they are closest to the "target" person when deciding to which the letter should be sent next. The minimum ultrametric incorporates these intuitions by postulating that two individuals in social space are close to each other when they are close on at least one dimension of differentiation.

While theoretically appealing, it is impossible to use the minimum ultrametric to infer the between-group distances from a set of observed relationships that link individuals to the groups. This is because the minimum ultrametric does not obey the triangle inequality and is, therefore, not a metric in the formal sense of the word. For example, consider a two dimensional social space, $X \subseteq \mathbb{R}^2$, where positions are represented by vectors $x = (x_1, x_2) \in X$. The minimum ultrametric defines the distance between two points, $a, b \in X$, as $\rho_{\text{mu}}(a, b) = \min\{|a_1 - a_2|, |b_1 - b_2|\}$ (Watts et al. 2002). Intuitively, this means that two points are considered to be close if they are close on *at least one dimension*. Let I be a set of individuals and A and B two groups. Suppose that all the individuals in I are located at $x = (x_1, x_2)$ and let the positions of A and B be, respectively, $a = (a_1, a_2)$ and $b = (b_1, b_2)$. Further, assume that $|x_1 - a_1| = 0$ and $|x_2 - b_2| = 0$, implying that both $\rho_{\text{mu}}(x, A) = \rho_{\text{mu}}(x, B) = 0$. In words, x is as close as possible to both A and B. Even in this situation, the distance between A and B is $\rho_{\text{mu}}(a, b) = \min\{|a_1 - b_1|, |a_2 - b_2|\} = \min\{|x_1 - b_1|, |x_2 - a_2|\}$, which we can make as large as we wish by choosing arbitrarily large (or small) values for a_2 and b_1. In other words, even if all individuals in the space are simultaneously as close as possible to A and B, the distance *between* A and B can

be any positive real number and is therefore undefined. Hence, although interaction patterns at the micro-level can be modeled starting from a "minimum ultrametric space" (as in Centola 2015), modeling the macro from the bottom up is blocked by the very nature of the distance function. In fact, it is precisely the assumption that two groups being close to many individuals cannot be too far away from each other in social space that enables us to infer the distances between groups from the network composition of individuals. As the minimum ultrametric does not meet this condition, it cannot be used for the current study despite its intuitive appeal.

Details on the Estimation Procedure

Let Ω be the set of all parameters in the model. The data consists of $i = 1, 2, ..., N$ individuals and $j = 1, 2, .., J$ social groups. The number of acquaintanceships individual i has to a member of group j is denoted by y^*_{ij} which is, unfortunately, unobserved. What is observed is whether y^*_{ij} falls into any of the $k = 1, 2, ..., K$ response categories outlined in the data section. Assuming that y^*_{ij} follows a Negative Binomial distribution with rate and dispersion parameter μ_{ij} and ϕ, the likelihood of the response matrix Y given the set of parameters Ξ is

$$p(Y|\Omega) = \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{K}\left[\sum_{y^*_{ij}\in I_k}\binom{y^*_{ij}+\phi-1}{y^*_{ij}}\left(\frac{\mu_{ij}}{\mu+\phi}\right)^{y^*_{ij}}\left(\frac{\phi}{\mu_{ij}+\phi}\right)^{\phi}\right]^{\mathbb{I}(y^*_{ij}\in I_k)} \quad (1)$$

where we model the rate parameter as

$$\mu_{ij} = \gamma_i \pi_j \rho(\theta_i, \xi_j)^{-\delta} \quad (2)$$

It is clear that the model is not identified. For example, multiplying all γ_is by a constant c and multiplying all $\rho(\theta_i, \xi_j)$s by c^δ would yield the same likelihood of the data. Similarly, squaring the distances $\rho(\theta_i, \xi_j)$ and dividing δ by 2, would lead to the same likelihood as well. To obtain a unique set of parameters, I first fix $\delta = 2$ and reparameterize the model as

$$\mu_{ij} = \alpha^* \gamma^*_i \pi_j \rho^*(\theta_i, \xi_j)^{-2} \quad (3)$$

with the constraints $\prod_{i=1}^{N} \gamma_i^* = 1$ and $\prod_{i=1}^{N} \prod_{j=1}^{J} \rho_{ij}^*(\theta_i, \xi_j)^{-2} = 1$. With this parameterization, α^* can be interpreted as the "grand mean" of the responses, and γ_i^* and $\rho^*(\theta_i, \xi_j)^{-2}$ as deviations from it. This will identify the likelihood, given that we fix the coordinate system of the social space. As the positions in social space enter the likelihood only in form of their distances, the likelihood will be invariant under any isometic transformation—i.e., translation, rotation, or reflection—of the positions. In addition, the social space has no natural scale, so we have to fix the scale of the space as well, either by fixing the scale of the group positions and let the individual positions vary, or by fixing the scale of individual positions and let the group positions be estimated relative to the individual positions.

To fix the scale of the social space, I assign the individual-positions a $\theta_i \sim \text{Normal}_d(0, \Sigma)$ prior, with the constraint that $\text{diag}(\Sigma) = I_d$. That is, the prior variance of the individual positions are constrained to be equal to one for each dimension, while the correlations between the dimensions are free to vary. This "weakly" identifies the scale of the space. To fix the orientation of the space, I impose $d - 1$ hard constraints and 1 soft constraint on d group-positions. That is, for the lth group position, $\xi_l, l \leq d$, I fix $\xi_{kl} = 0$ for $k \leq d - 1$ and impose the sign constrain $\xi_{kd} > 0$. In addition, I constrain the mean of the group positions for each dimension to be equal to zero. In all models presented in the paper, these constraints were sufficient to isolate a single posterior mode to sample from.

As Σ, with the diagonal entries constraint to 1, is a correlation matrix, it was assigned a LKJ prior, i.e., $\Sigma \sim \text{LKJ}(2)$. The group positions that are unconstrained are given $\xi_{kl} \sim \text{Normal}(0, 5)$ priors, while those with sign constraints are given $\text{Gamma}(2, .1)$ priors.[1] The α^* and γ_i^* parameters are given normal priors on the log-scale: $\ln \alpha^* \sim \text{Normal}(0, 5)$ and $\ln \gamma_i^* \sim \text{Normal}(0, \sigma_\gamma)$, with $\sigma_\gamma \sim \text{Half-Normal}(3)$. Lastly, the dispersion parameter is assigned a $\phi \sim \text{Half-Normal}(3)$ prior as well. Prior distributions for the population shares are shown in Table 1 of the paper.

[1] While the Gamma(2,.1) prior puts zero density at the origin, it has a constant positive derivative at zero, which allows the likelihood to dominate the posterior near zero. It was suggested as a weakly informative prior for parameters that have to be positive when using penalized likelihood estimation "in the sense that they supply some direction but still allow inference to be driven by the data." (Chung et al. 2013: 686.)

Abusing notation, and denoting all probability distributions by $p(\cdot)$, the posterior distribution of Ω can be formulated as

$$
\begin{aligned}
p(\Omega|Y) \propto & \prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{K}\left[\sum_{y_{ij}^{*}\in I_k}\binom{y_{ij}^{*}+\phi-1}{y_{ij}^{*}}\left(\frac{\mu_{ij}}{\mu+\phi}\right)^{y_{ij}^{*}}\left(\frac{\phi}{\mu_{ij}+\phi}\right)^{\phi}\right]^{\mathbb{I}(y_{ij}^{*}\in I_k)} \\
& \times p(\sigma_\gamma)\prod_{i=1}^{N}p(\gamma_i^{*}\,|\,\sigma_\gamma) \\
& \times p(\Sigma)\prod_{i=1}^{N}\prod_{d=1}^{D}p(\theta_{id}\,|\,\Sigma) \\
& \times p(\alpha^{*})p(\phi)\prod_{j=1}^{J}p(\pi_j)\prod_{d=1}^{D}p(\xi_{jd}).
\end{aligned}
\tag{4}
$$

The code to fit the model as well as to reproduce all presented materials in the paper can be found at `https://github.com/baruuum/Replication_Code`.

Convergence Diagnostics

The potential scale reduction factor (Gelman et al. 2013: 284) and estimated effective sample sizes for the three-dimensional social space model are shown below. Due to the abundance of parameters, summary statistics are presented. The entries of Table S1 show indications of well-mixed chains. The potential scale reduction factors are close to 1 and the effective sample sizes for all parameters exceed 400. Visual inspections of the traceplots suggest as well that the chains have reached a stationary distribution and are mixing well.

Table S1: Potential Scale Reduction Factor, $\hat{R}$, and Effective Sample Size Estimates for the Three-dimensional Social Space Model

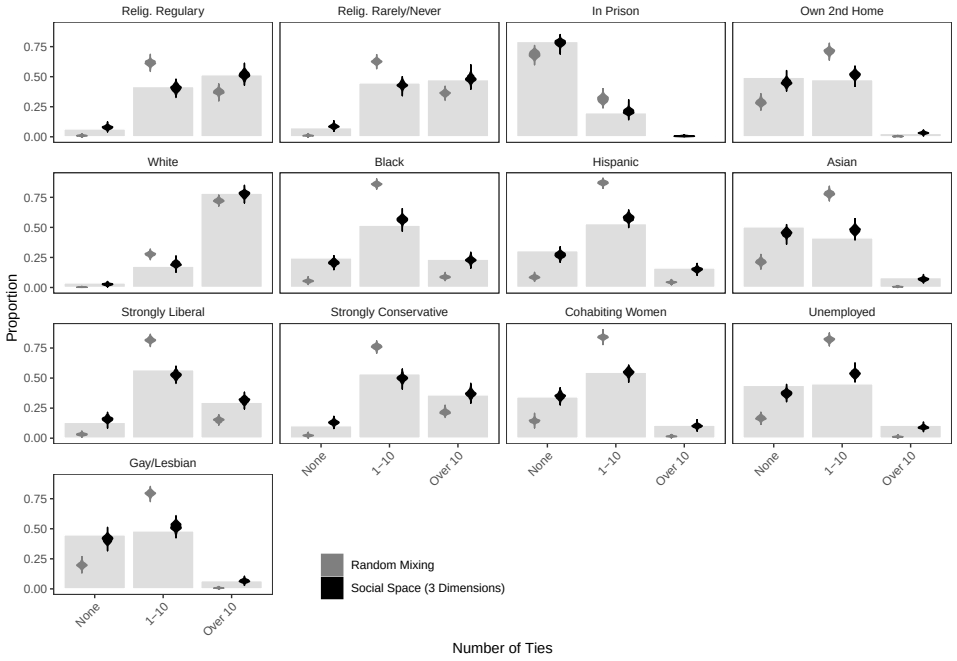
Parameter	Min	5%	25%	50%	75%	95%	Max
Potential Scale Reduction Factor, $\hat{R}$							
θ_i	0.99	0.99	0.99	0.99	1.00	1.00	1.00
ξ_j	0.99	0.99	0.99	1.00	1.00	1.00	1.00
π_j	0.99	0.99	0.99	1.00	1.00	1.00	1.00
$\log \alpha$	0.99	0.99	1.00	1.00	1.00	1.00	1.00
$\log \gamma_i$	0.99	0.99	0.99	0.99	1.00	1.00	1.00
$\sigma_{\log \gamma}$	0.99	0.99	0.99	0.99	0.99	0.99	0.99
ϕ	0.99	0.99	0.99	0.99	0.99	0.99	0.99
$\Sigma_{1,2}$				1.00			
$\Sigma_{1,3}$				1.00			
$\Sigma_{2,3}$				0.99			
Effective Sample Sizes							
Parameter	Min	5%	25%	50%	75%	95%	Max
θ_i	651.33	4227.39	6732.90	8566.82	10238.21	12488.40	17648.46
ξ_j	408.68	481.63	1664.01	7226.81	9404.25	11570.34	14713.09
π_j	596.97	896.25	4399.07	8321.92	11113.18	18757.88	20744.71
$\log \alpha$	1734.67	2268.12	4401.93	7069.20	9736.47	11870.28	12403.74
$\log \gamma_i$	654.47	4600.99	7633.14	9627.42	11087.20	13014.91	18696.84
$\sigma_{\log \gamma}$	7969.49	8084.81	8546.09	9122.70	9699.30	10160.58	10275.90
ϕ	5330.15	5671.62	7037.50	8744.86	10452.22	11818.11	12159.58
$\Sigma_{1,2}$				7991.23			
$\Sigma_{1,3}$				4605.69			
$\Sigma_{2,3}$				4956.22			

Notes: a) θ_i = individual positions, ξ_j = group positions, γ_i^* = gregariousness parameters (as deviations from grand mean on the log-scale), π_j = population share parameters, α^* = grand mean gregariousness (on the log-scale), σ_γ = standard deviation gregariousness (on the log-scale), ϕ = dispersion parameter of Negative Binomial distribution, $\Sigma_{l,k}$ = the (l,k)th element of the correlation matirx of individual positions. b) For parameters that are multidimensional summary statistics are shown.

Posterior Predictive Distributions

Figure S1 compares the posterior predictive distribution of the three-dimensional social space model, the posterior predictions of the random mixing model, and the actual distribution of the data. The light gray bars show the percentage of responses in the data. The black and gray violin plots, respectively, show the distribution of 300 draws from the posterior predictive distribution of the social space and the random mixing model. The figure shows that the model improves the fit over a

Figure S1: Distribution of Analyzed Data and Posterior Predictions from Random Mixing and Social Space Model with Three Dimensions



Notes: Light gray bars show distribution of the data. Violin plots in black and gray, respectively, show the distribution of 300 draws from the posterior predictive distribution of the social space model and the random mixing model.

random mixing model considerably, and that it performs fairly well in following the trend in the data; yet, it is also clear that the fit is not perfect. In particular, the model is better in predicting the responses above 10 acquaintanceship ties then those indicating a lower number.

Individual Distributions over the Social Space, by Family Income, Race, Ideology, and Religiosity

Kernel density estimates for the distribution of individual over estimated social space, broken down by family income, race, ideological self-identification, and religiosity are shown in Figure S2. The density estimates are fitted on the posterior median positions of the individuals. As the first two dimensions explain over 90%

of the variance in the posterior median positions, only the first two dimensions are shown.

A crude quantitative assessment of the overlap in the distributions might be conducted by using the coefficient of overlapping, which is defined as

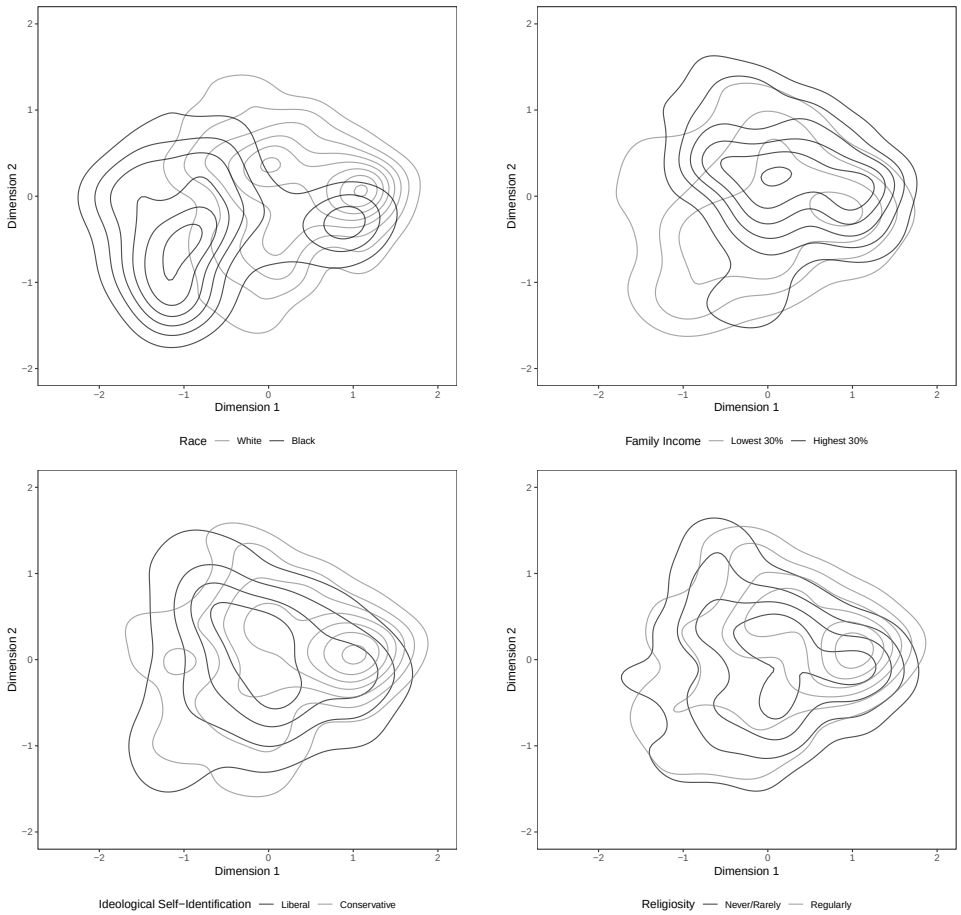
$$\text{OVL}(f, g) = \int_X \min \Big\{ f(x), g(x) \Big\} dx,$$

where f and g are the density functions of two groups both supported on the space X. I use a grid approximation to estimate $\text{OVL}(f, g)$ by calculating

$$\widehat{\text{OVL}}(f, g) = \sum_{k=1}^{K} \min \Big[\hat{f}(k), \hat{f}(k) \Big] \Delta_k$$

where K is the number of grid points, $\hat{f}(k)$ and $\hat{g}(k)$, respectively, are kernel density estimates of f and g at the point $k \in X$, and Δ_k is the size of cell to which k belongs. Using fifty equidistant grid points for each dimension, the coefficient of overlapping is estimated to be .39, .62, .68, and .71 for groups defined by race (black vs. white), family income (top 30% vs. bottom 30%), ideological self-placement (liberal or extremely liberal vs. conservative or extremely conservative), and religiosity (attending religious services nearly every week vs. once a year or less), respectively.

Figure S2: Individual Positions in Social Space by Race, Ideology, Religiosity, and Family Income

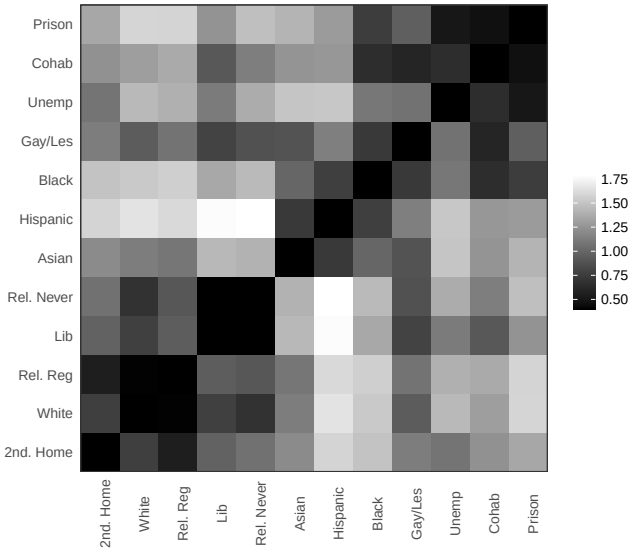


Notes: 1) Contour plots represent the distribution of posterior median positions of individuals in social space. 2) Only the first two dimensions are shown for clarity. 3) Individuals who self-identified as "liberal/conservative" or "extremely liberal/conservative" are used in the plots for ideological self-identification. 4) Individuals who report to attend services at "nearly every week" or more are used as "Regularly" category and those who report to attend services "once a year" or less are used as the "Never/Rarely" category in the analysis of Religiosity.

D-VALUES BETWEEN ALL PAIRS OF GROUPS

Figure S4 shows the posterior median of estimated D values between all pairs of groups. Lighter shades represent larger values of D and thus stronger tendencies for individuals' network compositions to vary with respect to the pair of groups.

Figure S4: Variation in Individuals' Distances to Pairs of Groups, All Groups

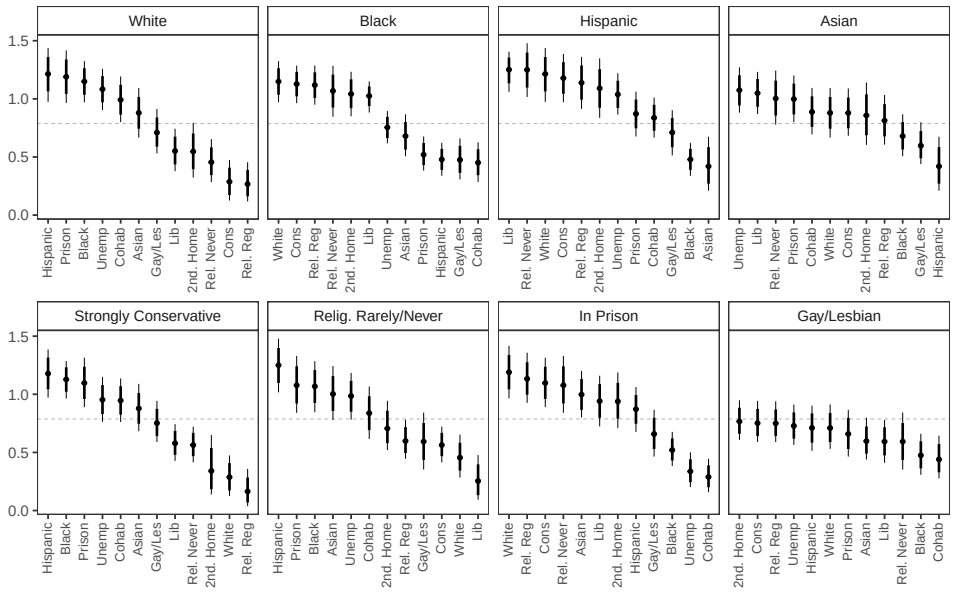


Notes: Each cell of the plot shows posterior median D_{jk} values. Higher values indicate stronger tendencies to segregate the network compositions of individuals.

Alternative Measure of Dispersion for the Calculation of *D*-values

Figure S5 shows the posterior median of estimated D values for the groups that were presented in the paper using the median absolute deviation (MAD), instead of the standard deviation, of $\Delta_{jk}(i)$ to capture the variations in network compositions.

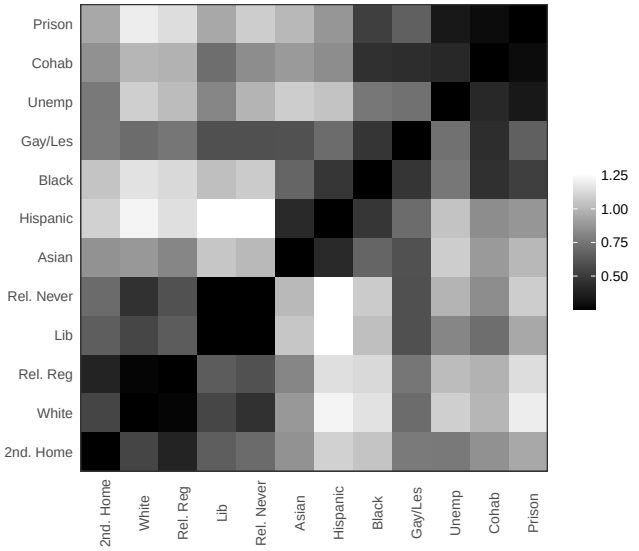
Figure S5: Variation in Individuals' Distances to Pairs of Groups



Notes: Dots represent posterior medians of estimated D values. Thin and thick vertical lines, respectively, show the 90% and 95% credible intervals. Dashed horizontal line in gray shows the posterior median of the average D values across all pairs of groups. All D-values are calculated using the MAD instead of the standard deviation of $\Delta_{jk}(i)$.

Figure S6 shows the posterior median D-values, calculated using the MAD, for all pairs of groups. Lighter shades represent larger values of D and thus stronger tendencies for individuals' network compositions to vary with respect to the pair of groups.

Figure S6: Variation in Individuals' Distances to Pairs of Groups, All Groups



Notes: Each cell of the plot shows posterior median D_{jk} values. Higher values indicate stronger tendencies to segregate the network compositions of individuals. All D-values are calculated using the MAD instead of the standard deviation of $\Delta_{jk}(i)$.

REFERENCES

Centola, Damon. 2015. "The Social Origins of Networks and Diffusion." *American Journal of Sociology* 120:1295–1338.

Chung, Yeojin, Sophia Rabe-Hesketh, Vincent Dorie, Andrew Gelman, and Jingchen Liu. 2013. "A nondegenerate penalized likelihood estimator for variance parameters in multilevel models." *Psychometrika* 78:685–709.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian data analysis (3rd ed.)*. FL: CRC Press.

Travers, Jeffrey and Stanley Milgram. 1969. "An experimental study of the small world problem." *Sociometry* pp. 425–443.

Watts, Duncan J. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105:493–527.

Watts, Duncan J., Peter Sheridan Dodds, and Mark EJ Newman. 2002. "Identity and search in social networks." *Science* 296:1302–1305.