Appendix A. Cultural Embeddedness and Cohesive Blocking Algorithm

The processes of cohesive blocking are recursive: the researcher first identifies the *k*-connectivity of an input graph, which is assigned the node connectivity of 0, and then removes the *k*-cutset(s) that hold(s) the network together. This procedure is then repeated on the resulting subgraphs until no further set-cutting can be done. As such, each iteration of this cohesive blocking process goes deeper into the network, as weakly connected nodes are removed first, leaving stronger connected sets.

Figure A1 illustrates the analytic processes for measuring the structural cohesion with a stylized sociogram example. Starting with the initial input graph consisting of 19 nodes (i.e., A), the network is disconnected into two components (B and C). The first component (B) consists of a simple dyad with no ties to the rest of the graph, while the second component (C) is larger and can be split into four 2-connected or bi-components (D, E, F, and G). Among them, the subgroups E and F are triads with a single tie connecting to the rest of the components, and any further cutting leads only to isolated nodes, so the cohesive blocking processes stop there. By contrast, subgroups D and G are more structurally complex and show nested patterns, so the cohesive blocking algorithm iterations continue and discover nested subgroups H and I with 3-connectivity. Figure A2 presents the tree diagram derived from the cohesive blocking algorithm and the level of the connectivity for each detected block.

Appendix B. Robust Inference with the MM-Estimator

Three types of outliers can influence the OLS estimator: (1) vertical outliers, (2) good leverage points, and (3) bad leverage points (Rousseeuw and Leroy, 2003). First, the vertical outliers are observations with outlying values for the corresponding error term (the *y* dimension) but no outlying values in the space of explanatory variables (the *x* dimension). These outliers affect the estimated intercept in the OLS estimation. Second, good leverage points are observations that are outlying in the space of explanatory variables but are located close to the regression line, and their presence causes inflation of the estimated standard errors. This reduces the efficiency but does not affect the OLS estimation. Third, bad leverage points are observations that are both outlying in the space of explanatory variables and located far from the true regression line. Their presence significantly affects the OLS estimation and leads to biased intercept and biased slope coefficients.

As such, analyzing a dataset that is contaminated with vertical outliers and bad leverage points using an OLS estimator may result in bad predictions. Instead, researchers should use robust regression methods that can provide stable results in the presence of these outliers. In our datasets, in both Studies 1 and 2, we detected several outlying cases that make our datasets unfit for an OLS estimation. To detect outliers, we use the graphical and diagnostic tools proposed by Rousseeuw and Zomeren (1990). This graphical tool is constructed by plotting the “robust standardized residual” on the vertical axis to represent outlyingness in terms of the fitted regression line or plane. The measure of the (multivariate) outlyingness of the explanatory variables is plotted on the horizontal axis and is calibrated by Mahalanobis distance. We set the limits for outlyingness in the *y* dimension as –2.25 and +2.25, representing the values of standard normal that separate the 2.5% remotest area of the distribution from the central mass. For the *x* dimension, we set the limits to $\sqrt{\chi^2_{p,0.975}}$ given that the squared Mahalanobis distance is χ^2_p distributed under normality (Verardi and Croux, 2009).

Figure B1 depicts the resulting plot from the Study 1 dataset. First, there are several bad leverage points, which are outliers in the horizontal and vertical dimensions (i.e., IDs 7, 27, 51, 60, 82, 126, 178, and 202). This means that their characteristics are very different from those of the bulk of the data and their values are much higher than they should be according to the fitted model. Second, there are vertical outliers (i.e., IDs 3 and 115) that have standard characteristics but are different from others in terms of the dependent variable. From the diagnostic plot, we can recognize that there is a serious risk that the OLS estimator is strongly distorted by the vertical outliers and bad leverage points in our dataset. Figure B2 is the diagnostic plot for Study 2, and similar to Study 1, it shows that unbiased and efficient estimation can be hindered by many outliers.

To reduce the effect of outliers, we use a robust regression method based on the MM-estimator, with the efficiency of 0.5.[1] The most commonly used robust regression methods are M-estimation, S-estimation, and MM-estimation. M-estimation was introduced by Huber (1964). The M in M-estimator stands for "maximum likelihood type," and the M-estimator is robust to vertical outliers but not to (especially bad) leverage points. By contrast, the S-estimator, introduced by Rousseeuw and Yohai (1984), is robust to bad leverage points and can theoretically withstand contamination up to 50% of the sample[2] but, at the same time, is highly

[1] The MM-estimator allows setting the efficiency level from 0.287 to 1, and the higher its value, the more efficient but the higher likelihood that the estimates are biased. Thus, MM-estimator needs to have a good compromise between robustness and efficiency. The results are essentially similar when we set the efficiency to 0.4 or 0.6.

[2] In other words, the S-estimator has a high "breakdown point," which is the proportion of incorrect observations an estimator can handle before giving an incorrect estimate. For example, the median has a breakdown point of 0.5, which is the maximum breakdown point, and that is why the early development of robust inference relied on the

inefficient compared with the M-estimator. The MM-estimator (Yohai, 1987) inherits the high breakdown point of the S-estimator while remaining almost as efficient as the OLS estimation like the M-estimator. The MM-estimator has a breakdown point as high as 0.5 and can attain an efficiency of up to 0.95 compared with OLS (Yohai, 1987). Therefore, we use the MM-estimation in our analyses.

resistant property of the median such as Rousseeuw's (1984) least median of squares regression that minimizes the median squared residual.

Appendix C. Results from CEM

Individuals with high cultural breath and embeddedness might differ systematically from those with low cultural breadth and embeddedness, and this could lead to biased estimates of regression coefficients and render them causally uninterpretable. While it is not possible to completely rule out this selection issue, one way to partially deal with it is to match individuals with high cultural breadth and embeddedness (treated) and those with low cultural breadth and embeddedness (control) on observable characteristics. This procedure results in a more balanced sample and makes a comparison between two groups more helpful.

Therefore, we used the CEM approach (Iacus, King, and Porro, 2012) to obtain covariate balance between the treatment and control sets with respect the control variables included in the main analyses: age, female, manager, education, idea length, and network constraint. The benefits of the CEM approach over other techniques have been demonstrated in several studies and across empirical settings, with CEM outperforming commonly used alternatives (Iacus, King, and Porro 2012). At the first stage, the treatment variable is dichotomized at its median value, and continuous control variables are "coarsened" into splines for the purposes of creating "strata"—or discrete mutually exclusive bins of control variables. We adjusted the bin size for each control variable (other than exact match variables such as gender) so that there is no significant difference between the treatment and control groups, and according to t-tests ($p < 0.05$), our treatment and control groups are well balanced across all control variables.[3] The matched proportion out of the initial sample is 42.9% for cultural breath and 25.5% for cultural embeddedness. After obtaining the matched sets, we ran all models on the balanced datasets by

[3] The t-test results are available on requests.

incorporating weights obtained from CEM along with control variables to reduce the standard errors.

Table C1 shows the CEM results across two treatments (cultural breadth and embeddedness) and the interaction result when the interaction term between cultural breadth and embeddedness is set as a treatment while controlling for the baseline effects for those two variables. Model 3 tests the interaction effect and finds that the interaction is statistically significant, confirming our main hypothesis.

REFERENCES

Huber, P. J.
1964 "Roust estimation of a location parameter." Annals of Mathematical Statistics, 35: 73–101.

Iacus, S. M., G. King, and G. Porro
2012 "Causal inference without balance checking: Coarsened exact matching." Political Analysis, 20: 1–24.

Rousseeuw, P. J.
1984 "Least median of squares regression." Journal of the American Statistical Association, 79: 871–880.

Rousseeuw, P. J., and A. M. Leroy
2003 Robust Regression and Outlier Detection. New York, NY: Wiley.

Rousseeuw, P. J., and B. C. Van Zomeren
1990 "Unmasking multivariate outliers and leverage points." Journal of the American Statistical Association, 85: 633–639.

Rousseeuw, P., and V. Yohai
1984 Robust Regression by Means of S-estimators. In J. Franke, W. Härdle, and R. D. Martin (eds.), Robust and Nonlinear Time Series Analysis: 256–276. New York, NY: Springer.

Verardi, V., and C. Croux
2009 "Robust regression in Stata." Stata Journal, 9: 439–453.

Yohai, V. J.
1987 "High breakdown-point and high efficiency robust estimates for regression." The Annals of Statistics, 15: 642–656.

FIGURES AND TABLES

Figure A1. An illustrative example of a cohesive blocking routine.

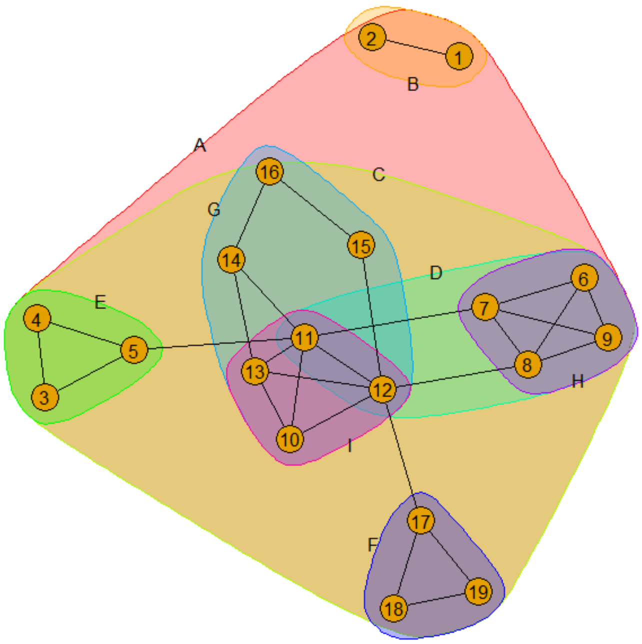

Figure A2. A tree diagram derived from the cohesive blocking algorithm.

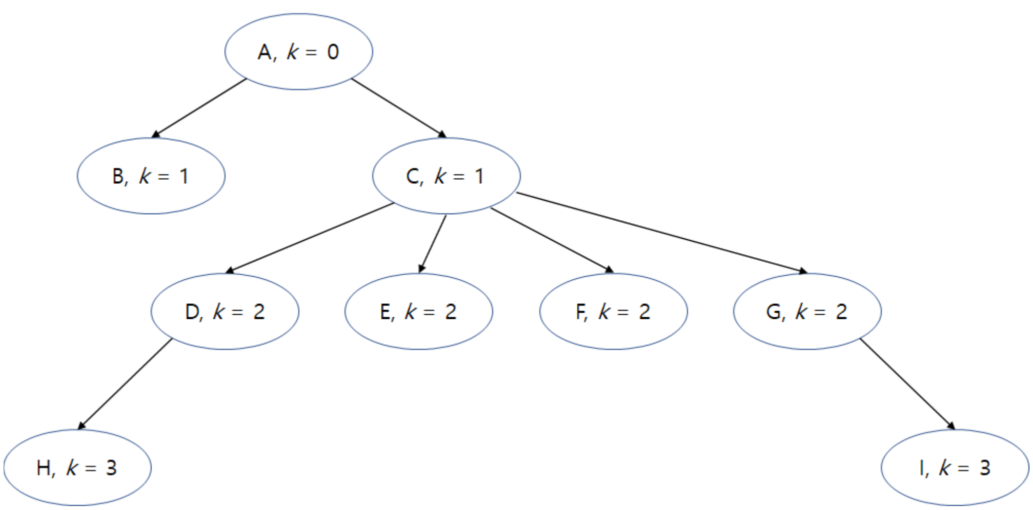

Figure B1. Diagnostic plot of standardized robust residuals versus robust Mahalanobis distances for Study 1.

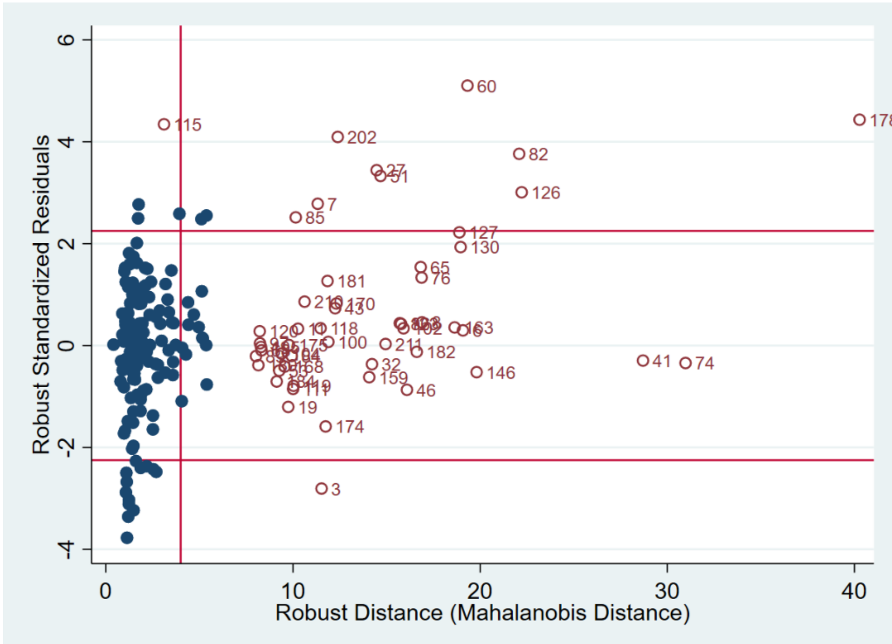

Figure B2. Diagnostic plot of standardized robust residuals versus robust Mahalanobis distances for Study 2.

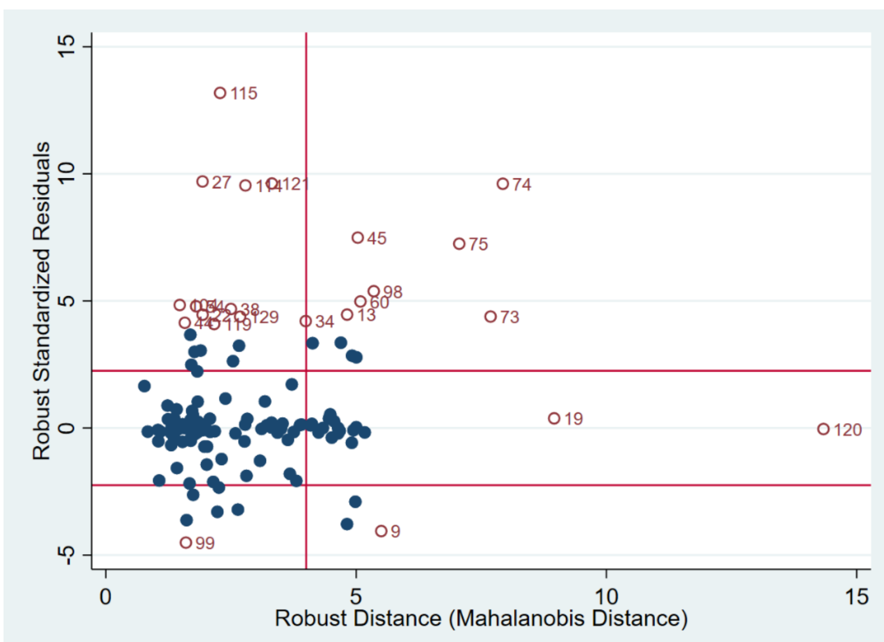

Table C1. CEM of Cultural Breadth and Cultural Embeddedness and their Interaction on Idea Creativity

	Model 1	Model 2	Model 3
Age	0.152	0.5	-2.251**
	-0.476	-0.389	-0.866
Female	-2.282	-1.798	3.903
	-3.758	-2.608	-5.439
Education	-9.848*	-3.965	-10.982
	-5.326	-4.741	-7.821
Manager	1.305	0.759	5.875
	-3.714	-2.844	-4.362
Tenure	1.926*	1.087	-4.980**
	-1.023	-0.824	-1.872
Idea Length	0.227**	0.108*	0.049
	-0.104	-0.061	-0.142
Network Constraint	11.791	-16.699**	43.696
	-16.861	-7.906	-27.769
Cultural Breadth	61.852		
	-54.347		
Cultural Breadth (dummy)		7.178***	-3.281
		-2.206	-4.534
Cultural Embeddedness		0.14	
		-0.197	
Cultural Embeddedness (dummy)	7.845**		-23.129**
	-2.928		-8.776
Cultural Breadth (dummy)			35.847***
× Cultural Embeddedness (dummy)			-9.63
Constant	-54.733	6.274	137.769**
	-37.318	-24.163	-56.18
Observations	217	217	217
R^2	0.411	0.257	0.922
Adjusted R^2	0.242	0.146	0.776
Residual Std. Error	9.800 (*df* = 42)	10.299 (*df* = 80)	5.596 (*df* = 7)
F Statistic	2.440**	2.311**	6.321**
	(*df* = 12; 42)	(*df* = 12; 80)	(*df* = 13; 7)

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$