# Supplementary Materials

## From populations to networks: relating diversity indices and frustration in signed graphs

A. Fontan[1], M. Ratta[2], and C. Altafini[3]

[1]Division of Decision and Control Systems,
School of Electrical Engineering and Computer Science,
KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
[2]Department of Mathematical Sciences "G.L. Lagrange", Politecnico di Torino, Turin, Italy
[3]Division of Automatic Control, Department of Electrical Engineering,
Linköping University, SE-58183 Linköping, Sweden

## Contents

**1 Creating the "condensed" matrix $W$ for the 3 application datasets 3**

**2 Comparison with other unbalance measures 3**

**3 Frustration on weakly balanced signed graphs: theoretical results 4**

**4 Different probabilities for positive and negative edges 8**

| Symbol | Meaning |
|---|---|
| $n$ | number of nodes in a graph. |
| $\mathcal{G}=\{\mathcal{V},\mathcal{E},A\}$ | undirected signed graph with node set $\mathcal{V}=\{1,\ldots,n\}$, edge set $\mathcal{E}=\mathcal{V}\times\mathcal{V}$, and symmetric adjacency matrix $A=[a_{ij}]$ with $a_{ij}\in\{0,\pm 1\}$. |
| $p$ | probability that an edge is included in the graph $\mathcal{G}$. |
| $m=pn^2$ | expected number of edges in $\mathcal{G}$. |
| $q$ | number of diagonal blocks of a weakly balanced graph $\mathcal{G}$. |
| $\mathcal{C}_1,\ldots,\mathcal{C}_q$ | (disjoint) groups of nodes of dimension $c_1,\ldots,c_q$; it holds that $\sum_{i=1}^{q} c_i = n$. |
| $W=[w_{ij}]$ | $q\times q$ weighted "condensed" adjacency matrix of a weakly balanced graph $\mathcal{G}$. |
| $\mathbb{1}_q$ | $q\times 1$ vector of ones. |
| $\zeta$ | frustration of a signed graph $\mathcal{G}$. |
| $S$ | signed diagonal matrix of $\pm 1$. |
| $e(S)$ | energy functional of the configuration state $S$. |
| $\{\mathcal{F}^{+}_{\text{best}},\mathcal{F}^{-}_{\text{best}}\}$ | optimal group partition obtained from the computation of $\zeta$, with cardinalities $n_{\mathcal{F}^{+}_{\text{best}}}, n_{\mathcal{F}^{-}_{\text{best}}}$. |
| $\ell_{\text{best}}$ | node excess in the best group bipartition. |
| $r_{\text{best}}=\frac{\ell_{\text{best}}}{n/2}$ | correction factor w.r.t. the best group bipartition. |
| $F$ | fractionalization index (a.k.a. Gini-Simpson index). |
| $E$ | effective number of groups (a.k.a. Laakso-Taagepera effective number of parties, or inverse Simpson index). |
| $H$ | Simpson index (a.k.a. Herfindahl-Hirschman index). |

Table S1: Notation used in the paper.

# 1 Creating the "condensed" matrix $W$ for the 3 application datasets

To build the "condensed" signed weighted adjacency matrix $W$ (introduced in Eq. (7) of the paper) we adopted the following procedure:

- A parliamentary network is modeled as an Erdös-Rényi signed graph $\mathcal{G}$, where every node is an MP. MPs from the same party are connected by a positive edge with a probability $p$, while MPs belonging to different parties are connected by a negative edge also with a probability $p$. When $i, j$ are connected, $a_{ij} = +1$ if MP $i$ and MP $j$ belong to the same party and $a_{ij} = -1$ if MP $i$ and MP $j$ belong to different parties. These and more details not included in the manuscript can be found in Ref. [33]. The matrix $W$ is obtained using Eq. (7) of the paper.

- An ethnolinguistic network is in principle modeled as a signed graph $\mathcal{G}$, where every node is an individual and each pair of individuals have a probability $p$ of being connected in $\mathcal{G}$. If $a_{ij} \neq 0$, it is assumed that $a_{ij} = +1$ if individuals $i$ and $j$ belong to the same ethnolinguistic group and $a_{ij} = -1$ otherwise, for all $i, j$. The resulting individual-level adjacency matrix $A$ is of size up to a billion, but it is not required for the analysis. Under a uniform connectivity assumption (Erdös-Rényi edge topology with edge probability $p$), the condensed matrix $W$ is obtained using Eq. (7) of the paper.

- In the mobile network application, $W$ is obtained using Eq. (7) where $q$ is the number of brands, $c_i$ represents the market fraction of brand $i$. It is assumed that different brands are connected by a negative edge.

# 2 Comparison with other unbalance measures

To measure unbalance in a signed graph, several alternatives to the frustration index $\zeta$ have been proposed in the literature, see [25] for a recent overview. Here we are interested only in quantifying the "distance to strong balance", as our signed graphs are already weakly balanced. We consider three different alternative measures:

1. Benzi-Estrada measure [27]:
$$\zeta_{\text{BE}} = \frac{1-K}{1+K},$$
where
$$K = \frac{\sum_{k=1}^{\infty} \text{Tr}\left[(P-N)^k\right]/k!}{\sum_{k=1}^{\infty} \text{Tr}\left[(P+N)^k\right]/k!},$$
and $P$ and $N$ represent resp. the positive and negative entries of the adjacency matrix $A$:
$$P_{ij} = \begin{cases} +1 & \text{if } A_{ij} = +1 \\ 0 & \text{otherwise} \end{cases}, \qquad N_{ij} = \begin{cases} +1 & \text{if } A_{ij} = -1 \\ 0 & \text{otherwise.} \end{cases}$$

2. Kirkley-Cantwell-Newman measure [28]:

$$\zeta_{\rm KCN} = \frac{1}{4} \log \frac{\det(zI - (P - N))}{\det(zI - (P + N))},$$

   where $z = \alpha\lambda^*$ with $\lambda^*$ the leading (most positive) eigenvalue of $P - N$ and $P + N$, and $\alpha$ is a parameter (chosen equal to 2, as in [28]).

3. Algebraic conflict [40]:

$$\zeta_{ac} = \min \lambda(L),$$

   where $L$ is the normalized "opposing" signed Laplacian associated to $A$ [49]: $L = I - D^{-1}A$ where $D = \text{diag}(|A|\mathbb{1})$ is the diagonal matrix having on the diagonal the row sums of the absolute values of $A$. It is know that the least eigenvalue $\min \lambda(L)$ is equal to 0 when the graph is strongly balanced, and that this eigenvalue grows with the distance to strong balance, see [30].

A numerical comparison of the four measures $\zeta$, $\zeta_{\rm KCN}$, $\zeta_{BE}$ and $\zeta_{\rm ac}$ is shown in Fig. S20 for networks of size $n = 1000$, with varying number $q = 2, \dots, 20$ of groups of uniform size, 100 instances for each $q$. As already observed in e.g. [29], the measure $\zeta_{\rm BE}$ saturates very quickly to the "completely unbalanced" value of 1. In Fig. S20**A**, $\zeta_{\rm BE} \sim 1 \; \forall \; q \geq 3$, which makes it useless for our purposes. The metric $\zeta_{\rm KCN}$ is not monotonically increasing in $q$, on the contrary, after an unclear transient it appears to decline with growing $q$. This is rather counterintuitive in our setting, as the "disorder" encoded in the signed graph grows with the number of groups $q$. The eigenvalue-based metric $\zeta_{\rm ac}$ instead behaves similarly to $\zeta$, as expected from the literature [25, 33]. In fact, the correlation between $\zeta$ and $\zeta_{\rm ac}$ is always $> 0.9$, see Fig. S20**B**.

# 3 Frustration on weakly balanced signed graphs: theoretical results

The following theorem collects 7 different expressions for the frustration $\zeta$ of a weakly balanced signed graph of Erdös-Rényi type. Some of the conditions were already obtained in [33], but only for fully connected graphs.

**Theorem 1** *Consider a Erdös-Rényi signed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ with edge probability $p$. Assume $\mathcal{G}$ is weakly balanced, with adjacency matrix $A$, of entries $A_{ij} = \{0, \pm 1\}$, having nonnegative diagonal blocks of dimension $c_1, \dots, c_q$, $\sum_{i=1}^q = n$, and nonpositive off-diagonal blocks. Let $\mathcal{C}_1, \dots, \mathcal{C}_q$ be the associated groups of nodes, $\mathcal{F}^+$, $\mathcal{F}^-$ a bipartition of such groups, and $S$ a diagonal signature matrix, $S = \text{diag}\{s_1, \dots, s_n\}$, $s_i = \pm 1$. Then the following* *expressions for the frustration are all identical:*

$$\zeta = \frac{1}{2m} \min_{s_i, s_j = \pm 1} \sum_{(i,j)\in\mathcal{E}} (1 - A_{ij} s_i s_j) \tag{S1}$$

$$= \frac{1}{2m} \min_{\substack{S=\mathrm{diag}\{s_1,\ldots,s_n\} \\ s_i=\pm 1}} \mathbb{1}^T \left(|A| - SAS\right) \mathbb{1} \tag{S2}$$

$$= \frac{1}{2m} \min_{\substack{s_i, s_j = \pm 1 \\ s_i = s_j \;\; if \\ i,j\in\mathcal{C}_k,\, k=1,\ldots,q}} \sum_{(i,j)\in\mathcal{E}} (1 - A_{ij} s_i s_j) \tag{S3}$$

$$= \frac{p}{2m} \min_{\substack{S_q=\mathrm{diag}\{s_1,\ldots,s_q\} \\ s_i=\pm 1}} \mathbb{1}_q^T \left(|W| - S_q W S_q\right) \mathbb{1}_q \tag{S4}$$

$$= \frac{p}{m} \min_{\mathcal{F}^+} \left( \sum_{\substack{i,j\in\mathcal{F}^+ \\ i\neq j}} c_i c_j + \sum_{\substack{i,j\in\mathcal{F}^- \\ i\neq j}} c_i c_j \right) \tag{S5}$$

$$= F - \frac{2}{n^2} \max_{\mathcal{F}^+} \left( n_{\mathcal{F}^+} (n - n_{\mathcal{F}^+}) \right) \tag{S6}$$

$$= F - \frac{1}{2} + \frac{1}{2} r_{\mathrm{best}}^2 \tag{S7}$$

*where $W$ is given in Eq. (7) of the paper, $m = pn^2$, $S_q = \mathrm{diag}\{s_1, \ldots, s_q\}$ is a $q \times q$ diagonal signature matrix (one $s_i$ for each cluster), $n_{\mathcal{F}^+}$ (resp. $n_{\mathcal{F}^-}$) is the size of $\mathcal{F}^+$ (resp. $\mathcal{F}^-$), $r_{\mathrm{best}} = \frac{\ell_{\mathrm{best}}}{n/2}$, and $\ell_{\mathrm{best}}$ is the least node excess with respect to $n/2$ among all possible bipartitions $\{\mathcal{F}^+, \mathcal{F}^-\}$ of $\mathcal{C}_1, \ldots, \mathcal{C}_q$.*

**Proof.** Since $|A_{ij}| \in \{0, \pm 1\}$ and $\mathbb{1}^T(|A|)\mathbb{1} = m = pn^2$, (S2) is the matrix version of (S1). The equality (S1) $\Longleftrightarrow$ (S3) means that the optimum occurs exactly in correspondence of a splitting of the groups. To show it, assume without loss of generality that the minimum energy block-wise splitting of the clusters is $\mathcal{F}^+ = \{\mathcal{C}_1, \ldots, \mathcal{C}_r\}$ and $\mathcal{F}^- = \{\mathcal{C}_{r+1}, \ldots, \mathcal{C}_q\}$.

We then have:

$$|A| - SAS = \left[\begin{array}{c|c}
\begin{array}{c|c|c|c}
\{0\} & \{0,2\} & \dots & \{0,2\} \\ \hline
\{0,2\} & \{0\} & \ddots & \vdots \\ \hline
\vdots & \ddots & \ddots & \{0,2\} \\ \hline
\{0,2\} & \dots & \{0,2\} & \{0\}
\end{array} & \{0\} \\ \hline
\{0\} & \begin{array}{c|c|c|c|c}
\{0\} & \{0,2\} & \dots & & \{0,2\} \\ \hline
\{0,2\} & \{0\} & \ddots & & \vdots \\ \hline
\vdots & \ddots & \ddots & & \\ \hline
 & & & & \{0,2\} \\ \hline
\{0,2\} & \dots & & \{0,2\} & \{0\}
\end{array}
\end{array}\right] \tag{S8}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad}_{\mathcal{F}^+}\quad\underbrace{\qquad\qquad\qquad\qquad\qquad}_{\mathcal{F}^-}$$

where $\{0\}$ means a block of size $c_i \times c_j$ (or $n_{\mathcal{F}^+} \times n_{\mathcal{F}^-}$ for the large off-diagonal blocks) of entries all equal to 0, while $\{0, 2\}$ means an equally sized block of entries 0 or 2. Denote

$$\zeta_1 = \frac{1}{2m} \min_{s_i, s_j = \pm 1} \sum_{(i,j) \in \mathcal{E}} (1 - A_{ij} s_i s_j)$$

and

$$\zeta_2 = \frac{1}{2m} \min_{\substack{s_i, s_j = \pm 1 \\ s_i = s_j \text{ if} \\ i,j \in \mathcal{C}_k,\, k=1,\dots,q}} \sum_{(i,j) \in \mathcal{E}} (1 - A_{ij} s_i s_j).$$

Clearly it is $\zeta_1 \le \zeta_2$ as the min in $\zeta_2$ is more constrained than in $\zeta_1$. By contradiction, let us assume that $\zeta_1 < \zeta_2$. Hence there must be at least a node that is misassigned in the calculation of $\zeta_2$. Assume without loss of generality it to be the first node of the first cluster (denote it $1 \in \mathcal{C}_1$). If node $i$ is a first neighbor of node 1 (i.e., if $(1,\, i) \in \mathcal{E}$), we have $A_{1i} = +1$ for $i \in \mathcal{C}_1$ and $A_{1i} = -1$ for $i \notin \mathcal{C}_1$, which in the "true" optimal assignment leads to

$$(1 - A_{1i} s_1 s_i) = \begin{cases} 0 & \text{if } i \in \mathcal{C}_1 \\ 2 & \text{if } i \in \{\mathcal{C}_2, \dots, \mathcal{C}_r\} \\ 0 & \text{if } i \in \{\mathcal{C}_{r+1}, \dots, \mathcal{C}_q\}. \end{cases}$$

In the "reassignement" induced by the contrarian assumption, all signs in the first row and column are instead switched (i.e., $s_1$ is flipped):

$$(1 - A_{1i} s_1 s_i) = \begin{cases} 2 & \text{if } i \in \mathcal{C}_1 \\ 0 & \text{if } i \in \{\mathcal{C}_2, \dots, \mathcal{C}_r\} \\ 2 & \text{if } i \in \{\mathcal{C}_{r+1}, \dots, \mathcal{C}_q\}. \end{cases}$$

Counting the contribution of node 1 to the energy functional, it is equal to $2p(c_2+\ldots+c_r)$ in the first case and $2p(c_1-1+c_{r+1}+\ldots+c_q)$ in the second case. (Recall that the minimization leads to a splitting $\{\mathcal{C}_1,\ldots,\mathcal{C}_r\}$ and $\{\mathcal{C}_{r+1},\ldots,\mathcal{C}_q\}$ s.t. the sums $c_1+\ldots+c_r$ and $c_{r+1}+\ldots+c_q$ are as equal as possible.) If (again by contradiction) it is $\zeta_1<\zeta_2$ then it must be $c_2+\ldots+c_r> c_1-1+c_{r+1}+\ldots+c_q$. But then the same consideration is true for all nodes in $\mathcal{C}_1$. Consider for instance node $2\in\mathcal{C}_1$, and assume that $A_{12}>0$ (the reasoning is analogue if $A_{12}=0$). Since now node 1 no longer is assigned to the faction of $\mathcal{C}_1$ (i.e., to $\mathcal{F}^+$), it is $s_1=-1$. In the "true" assignment, $s_2=+1$, and the contribution of node 2 to the energy functional is $2p(1+c_2+\ldots+c_r)$ , while in the reassignment ($s_2=-1$) it is $2p(c_1-2+c_{r+1}+\ldots+c_q)$. The contradictory hypothesis (which yields $c_2+\ldots+c_r>c_1-1+c_{r+1}+\ldots+c_q$) also implies that $1+c_2+\ldots+c_r>c_1-2+c_{r+1}+\ldots+c_q$, hence also node 2 should be reassigned to the $\mathcal{F}^-$ faction. Iterating the reasoning for all nodes of $\mathcal{C}_1$, this is equivalent to say that the partition $\{\mathcal{C}_1,\ldots,\mathcal{C}_r\}$ and $\{\mathcal{C}_{r+1},\ldots,\mathcal{C}_q\}$ is not an optimal one, and it should be instead $\{\mathcal{C}_2,\ldots,\mathcal{C}_r\}$ and $\{\mathcal{C}_1,\mathcal{C}_{r+1},\ldots,\mathcal{C}_q\}$, which is a contradiction. In other words, whenever flipping signs to a node improves the energy function, flipping sign to an entire group also does so, hence the minimum is always obtained in correspondence of a group partition. Therefore (S1) $\Longleftrightarrow$ (S3).

The optimization problem can then be formulated as choosing equal spin assignment to all nodes of a group. By dimension counting, the expression in (S2) can be reexpressed in terms of the weight matrix $W$ of Eq. (7) of the paper, as in (S4):

$$\begin{aligned}\zeta =&\frac{1}{2m}\min_{\substack{S=\operatorname{diag}\{s_1,\ldots,s_n\}\\ s_i=\pm1}} \mathbb{1}^T\left(|A|-SAS\right)\mathbb{1}\\ =&\frac{p}{2m}\min_{\substack{S_q=\operatorname{diag}\{s_1,\ldots,s_q\}\\ s_i=\pm1}} \mathbb{1}_q^T\left(|W|-S_qWS_q\right)\mathbb{1}_q.\end{aligned}$$

Choosing $S_q=\operatorname{diag}\{s_1,\ldots,s_q\}$ means choosing a partition $\{\mathcal{F}^+,\,\mathcal{F}^-\}$ of the clusters, which, similarly to (S8), leads to

$$|W|-S_qWS_q=$$
$$p\left[\begin{array}{cccc|cccc} 0 & 2c_1c_2 & \ldots & 2c_1c_r & & & & \\ 2c_1c_2 & 0 & \ddots & \vdots & & 0 & & \\ \vdots & \ddots & \ddots & 2c_{r-1}c_r & & & & \\ 2c_1c_r & \ldots & 2c_{r-1}c_r & 0 & & & & \\ \hline & & & & 0 & 2c_{r+1}c_{r+2} & \ldots & 2c_{r+1}c_q \\ & & & & 2c_{r+1}c_{r+2} & 0 & \ddots & \vdots \\ & 0 & & & \vdots & \ddots & \ddots & 2c_{q-1}c_q \\ & & & & 2c_{r+1}c_q & \ldots & 2c_{q-1}c_q & 0 \end{array}\right]. \qquad \text{(S9)}$$
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\mathcal{F}^+}\quad\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\mathcal{F}^-}$$

Summing over rows and columns of (S9),

$$\mathbb{1}_q^T \left(|W| - S_q W S_q\right) \mathbb{1}_q = 2p \left( \sum_{\substack{i,j \in \mathcal{F}^+ \\ i \neq j}} c_i c_j + \sum_{\substack{i,j \in \mathcal{F}^- \\ i \neq j}} c_i c_j \right),$$

hence (S5) follows from (S4). Adding and subtracting diagonal terms to (S9), replacing $\frac{p}{m}$ with $\frac{1}{n^2}$, and performing easy calculations we get:

$$\begin{aligned}
\zeta =& \frac{p}{m} \min_{\mathcal{F}^+} \left( \sum_{i,j \in \mathcal{F}^+} c_i c_j + \sum_{i,j \in \mathcal{F}^-} c_i c_j - \sum_{i=1}^q c_i^2 \right) \\
=& \frac{1}{n^2} \min_{\mathcal{F}^+} \left( n_{\mathcal{F}^+}^2 + (n - n_{\mathcal{F}^+})^2 - \sum_{i=1}^q c_i^2 \right) \\
=& \frac{1}{n^2} \left( \min_{\mathcal{F}^+} \left( 2n_{\mathcal{F}^+}^2 - 2n n_{\mathcal{F}^+} \right) + n^2 - \sum_{i=1}^q c_i^2 \right) \\
=& \frac{2}{n^2} \min_{\mathcal{F}^+} \left( n_{\mathcal{F}^+}^2 - n n_{\mathcal{F}^+} \right) + \underbrace{1 - \frac{\sum_{i=1}^q c_i^2}{n^2}}_{=F} \\
=& F - \frac{2}{n^2} \max_{\mathcal{F}^+} \left( n_{\mathcal{F}^+}(n - n_{\mathcal{F}^+}) \right)
\end{aligned}$$

which is (S6). The maximum is obtained when both $n_{\mathcal{F}^+}$ and $n - n_{\mathcal{F}^+}$ approach $n/2$. Denoting $\mathcal{F}^+_{\text{best}}$, $\mathcal{F}^-_{\text{best}}$ the best possible partition for the given groups, and denoting $\ell_{\text{best}} = n_{\mathcal{F}^+_{\text{best}}} - \frac{n}{2}$ the least "distance to equibipartition" (assuming without loss of generality that $n_{\mathcal{F}^+_{\text{best}}} \geq n_{\mathcal{F}^-_{\text{best}}}$), then

$$\zeta = F - \frac{2}{n^2} \left( n_{\mathcal{F}^+_{\text{best}}} (n - n_{\mathcal{F}^+_{\text{best}}}) \right) = F - \frac{2}{n^2} \left( \frac{n}{2} + \ell_{\text{best}} \right) \left( n - \frac{n}{2} - \ell_{\text{best}} \right) = F - \frac{2}{n^2} \left( \frac{n^2}{4} - \ell_{\text{best}}^2 \right)$$

from which (S7) is obtained. ■

# 4 Different probabilities for positive and negative edges

In this section we generalize the results to the case in which we still have a weakly balance signed graph, but the probabilities of existence of positive and negative edges are different. In the following theorem we use the same notation as in Theorem 1.

**Theorem 2** *Consider a weakly balanced signed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ in which the probability of a positive edge is $p_1$ and that of a negative edge is $p_2$. Let*

$$m = p_1 \sum_{i=1}^q c_i^2 + p_2 \sum_{\substack{i,j=1 \\ i \neq j}}^q c_i c_j \tag{S10}$$

*be the total (expected – omitted thereafter) number of edges of $\mathcal{G}$. Define $H$ (resp. $F$) as the fraction of positive (resp. negative) edges of $\mathcal{G}$,*

$$H = \frac{p_1 \sum_{i=1}^{q} c_i^2}{m}, \qquad F = \frac{p_2 \sum_{\substack{i,j=1 \\ i\neq j}}^{q} c_i c_j}{m}, \tag{S11}$$

*and $\zeta$ as in Eq.* (S1) *(i.e., Eq. (3) of the main paper). Then $F = 1 - H$, and all expressions* (S2)-(S5) *for $\zeta$ still hold provided $p$ is replaced by $p_2$, while in place of* (S6) *we have*

$$\zeta = F - \frac{2p_2}{m} \max_{\mathcal{F}^+} \left( n_{\mathcal{F}^+}(n - n_{\mathcal{F}^+}) \right).$$

*The closed-form relationship between $\zeta$ and $F$ of Eq.* (S7) *(i.e., Eq. (5) of the main paper) is replaced by*

$$\zeta = \frac{p_2}{p_1} F - \left( \frac{p_2}{p_1} - \frac{p_2}{m}\frac{n^2}{2} \right) + \frac{2p_2}{m} \ell_{\text{best}}^2. \tag{S12}$$

**Proof.**

Denote the number of positive edges $m^+ = p_1 \sum_{i=1}^{q} c_i^2$ and that of negative edges $m^- = p_2 \sum_{\substack{i,j=1 \\ i\neq j}}^{q} c_i c_j$. Then

$$H = \frac{m^+}{m} = \frac{m - m^-}{m} = 1 - F.$$

As can be deduced from the proof of Theorem 1 (see e.g., (S8)), only negative edges contribute to $\zeta$, hence the difference between $p_1$ and $p_2$ is irrelevant when computing $\zeta$, provided that $p$ is replaced with $p_2$. The expressions (S2)-(S5) follow consequently from this observation. From (S5), adding and subtracting elements and performing calculations similar to those in the proof of Theorem 1,

$$\begin{aligned} \zeta =& \frac{p_2}{m} \left( \sum_{\substack{i,j \in \mathcal{F}^+ \\ i \neq j}} c_i c_j + \sum_{\substack{i,j \in \mathcal{F}^- \\ i \neq j}} c_i c_j \right) \pm \frac{p_2}{m} \sum_{i=1}^{q} c_i^2 \\ =& \frac{p_2}{m} \min_{\mathcal{F}^+} \left( n_{\mathcal{F}^+}^2 + (n - n_{\mathcal{F}^+})^2 - \sum_{i=1}^{q} c_i^2 \right) \end{aligned}$$

from which we can observe that the minimization problem is the same as in Theorem 1, hence we get (again, adding and subtracting terms),

$$\begin{aligned} \zeta =& -\frac{2p_2}{m} \left( \frac{n^2}{4} - \ell_{\text{best}}^2 \right) + \frac{p_2 n^2}{m} - \frac{p_2}{p_1} \underbrace{\frac{p_1 \sum_{i=1}^{q} c_i^2}{m}}_{=H} \pm \frac{p_2}{p_1} \\ =& -\frac{2p_2}{m} \left( \frac{n^2}{4} - \ell_{\text{best}}^2 \right) + \frac{p_2 n^2}{m} - \frac{p_2}{p_1} + \frac{p_2}{p_1} \underbrace{(1 - H)}_{=F} \end{aligned}$$

from which (S12) follows. ■

The expression (S12) has the same structure as Eq. (S7). In fact, it reduces to (S7) when $p_1 = p_2$.

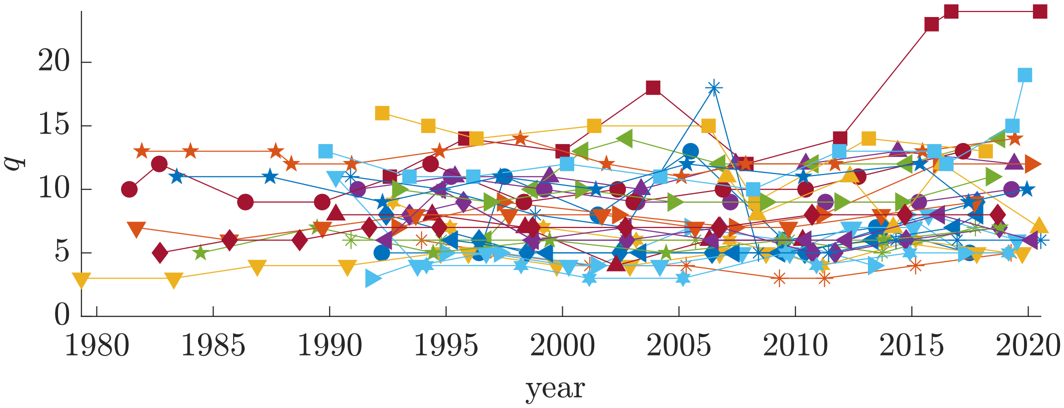


Figure S1: Number of groups (i.e., $q$) for the parliamentary networks dataset. Color code and marker shape is the same as in Fig. 4A of the main paper.

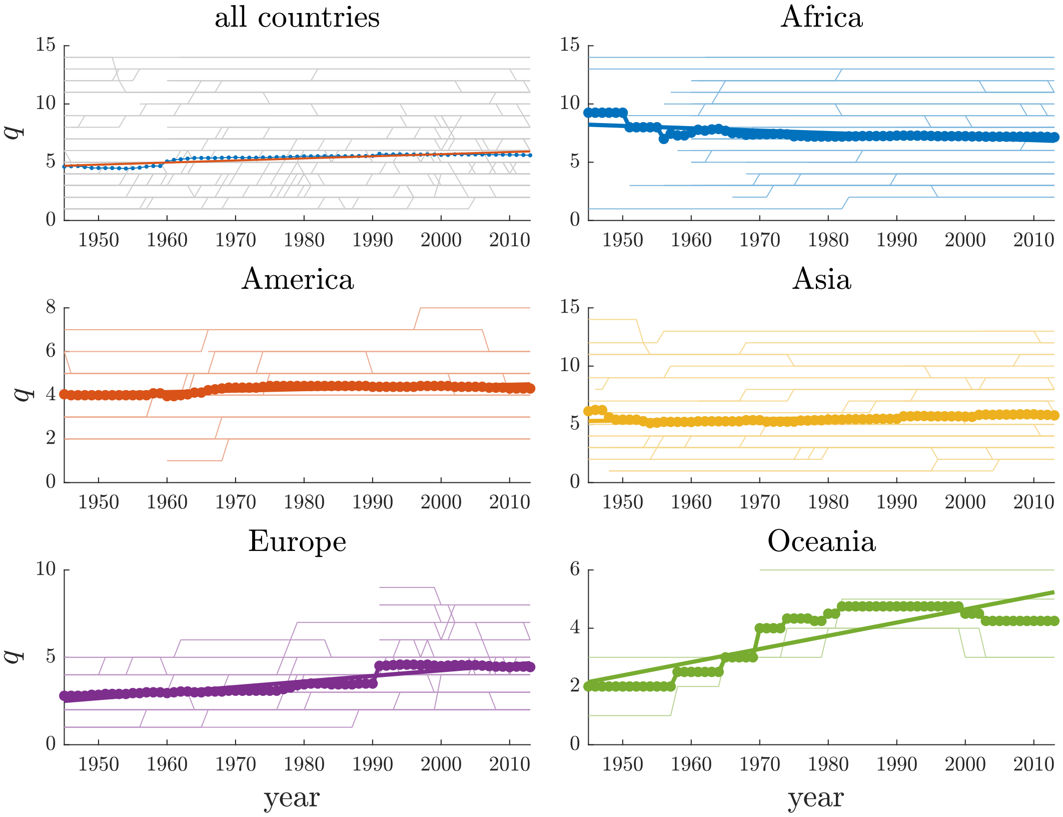


Figure S2: Number of groups (i.e., $q$) in a country for the ethnolinguistic networks dataset (overall and subdivided by continents). Color code is the same as in Fig. 4B of the main paper. Mean and linear regression line are also reported in bold.

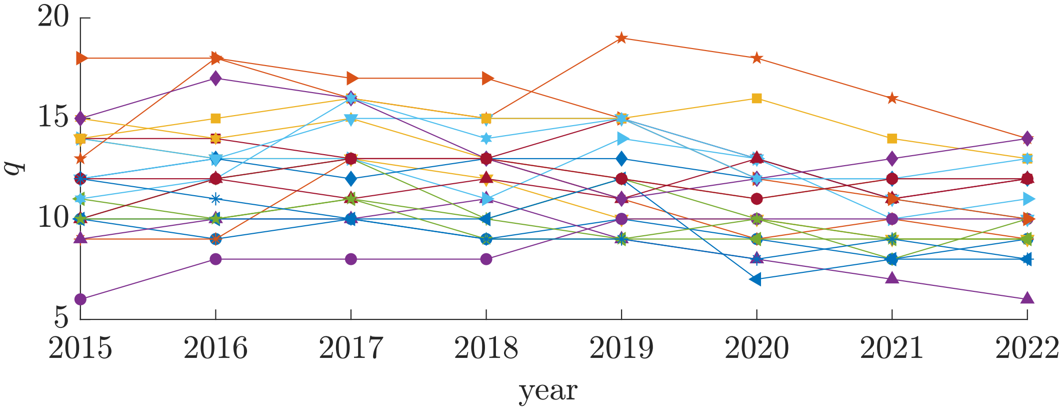


Figure S3: Number of groups (i.e., $q$) for the smartphone market shares dataset. Color code and marker shape is the same as in Fig. 4C of the main paper.

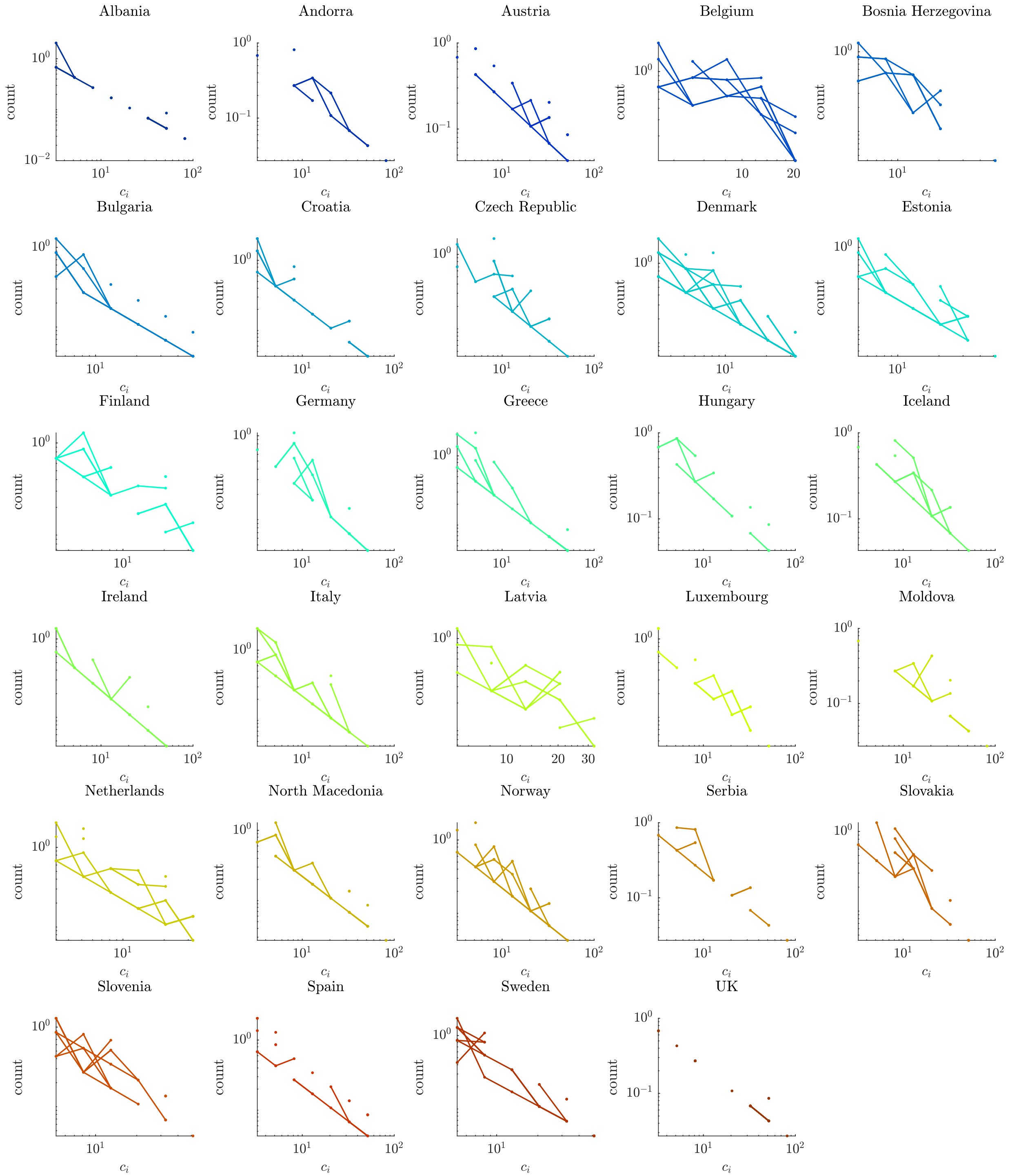


Figure S4: Group size distribution for the parliamentary networks dataset. Continuous lines identify points in neighboring bins on the same year.

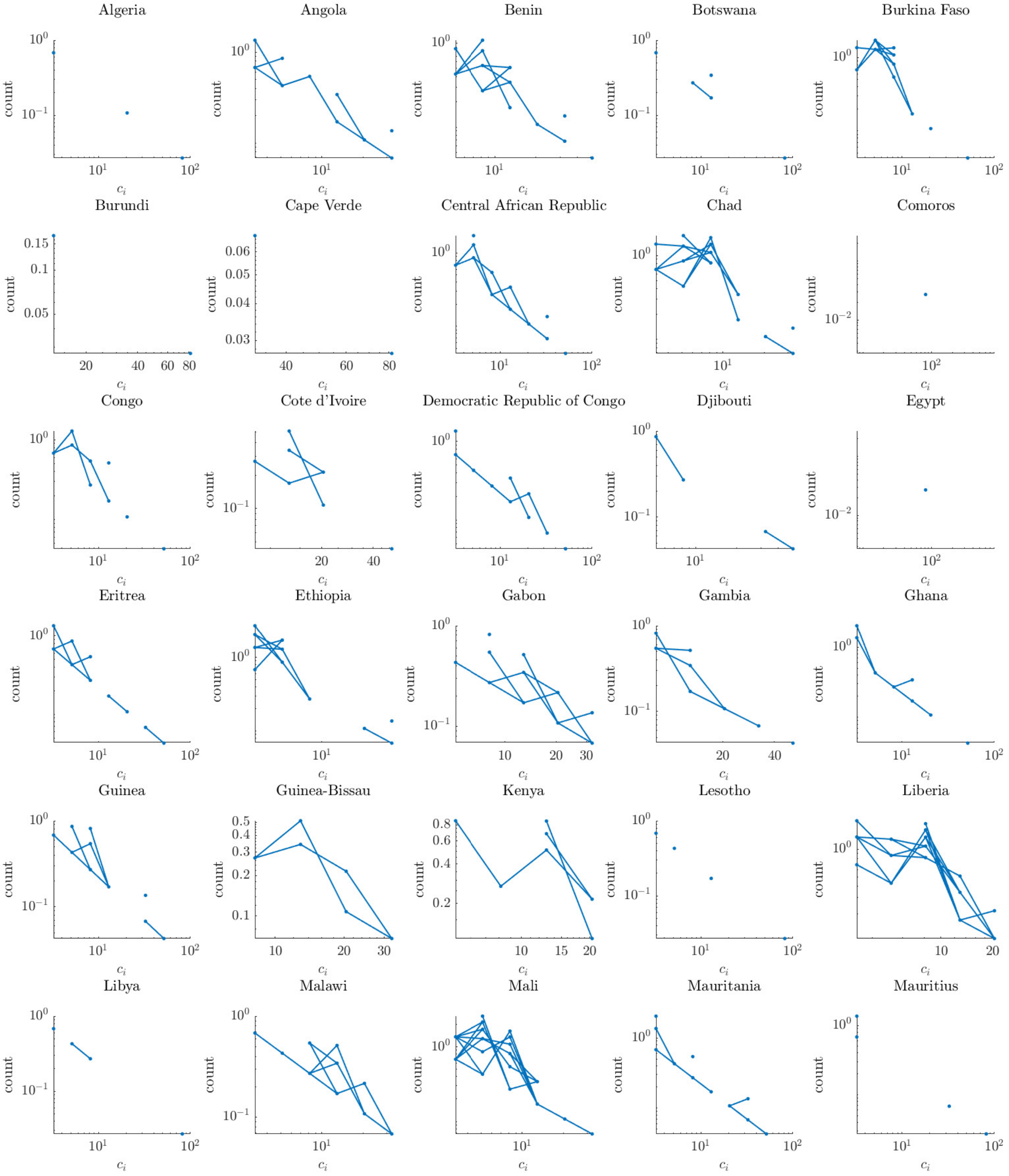


Figure S5: Group size distribution for the ethnolinguistic dataset (1/6). Continuous lines identify points in neighboring bins on the same year.

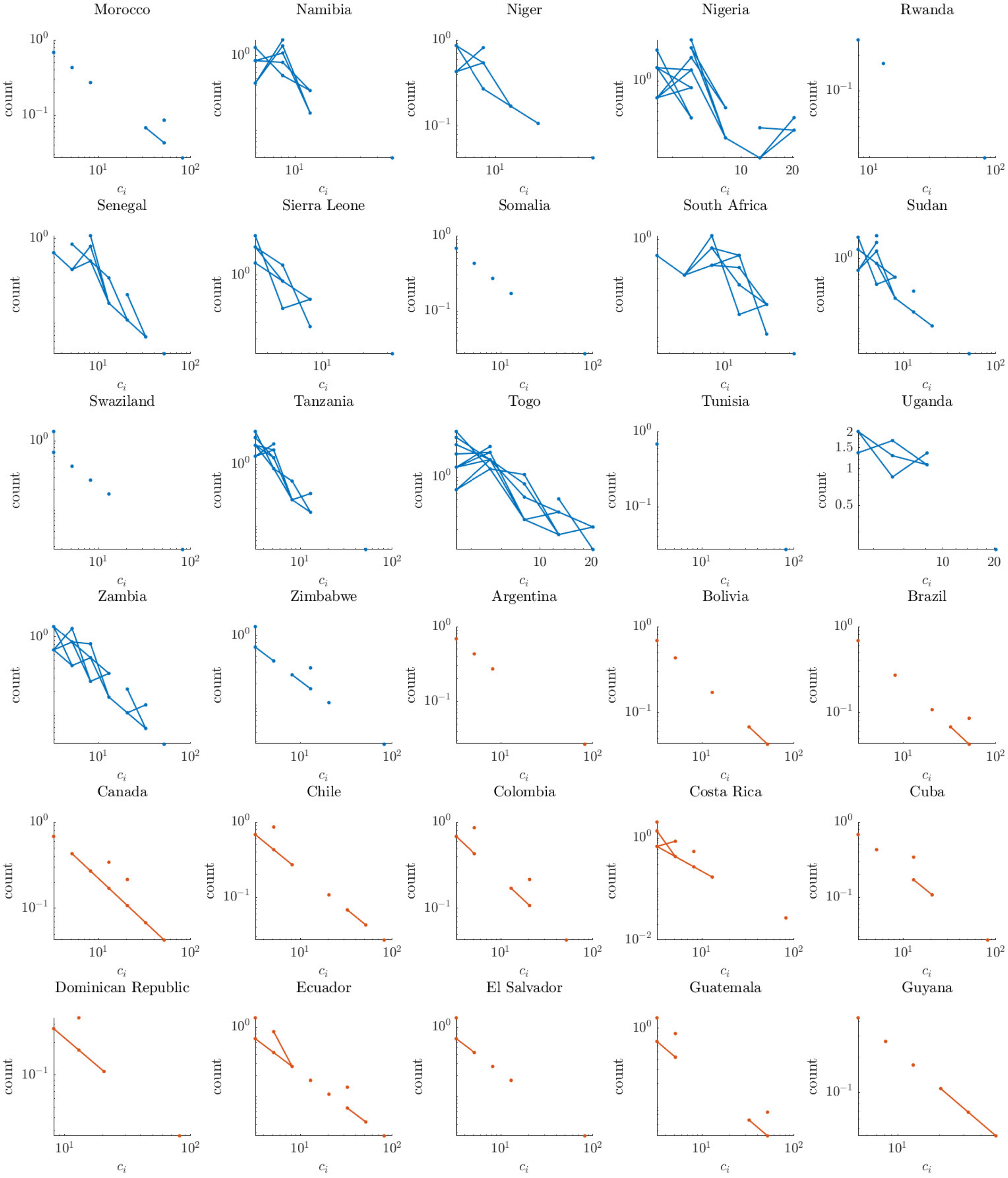


Figure S6: Group size distribution for the ethnolinguistic dataset (2/6). Continuous lines identify points in neighboring bins on the same year.

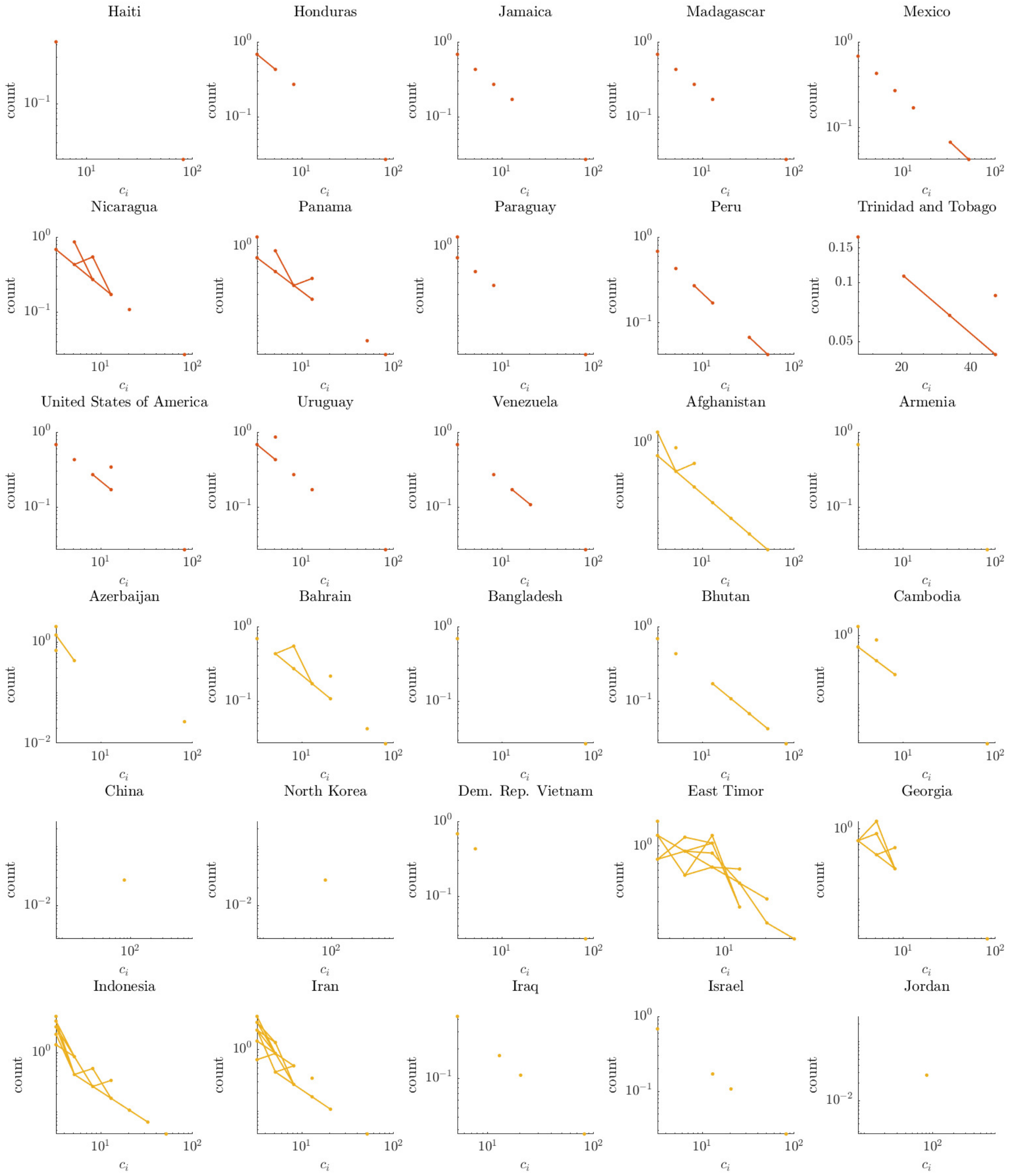


Figure S7: Group size distribution for the ethnolinguistic dataset (3/6). Continuous lines identify points in neighboring bins on the same year.

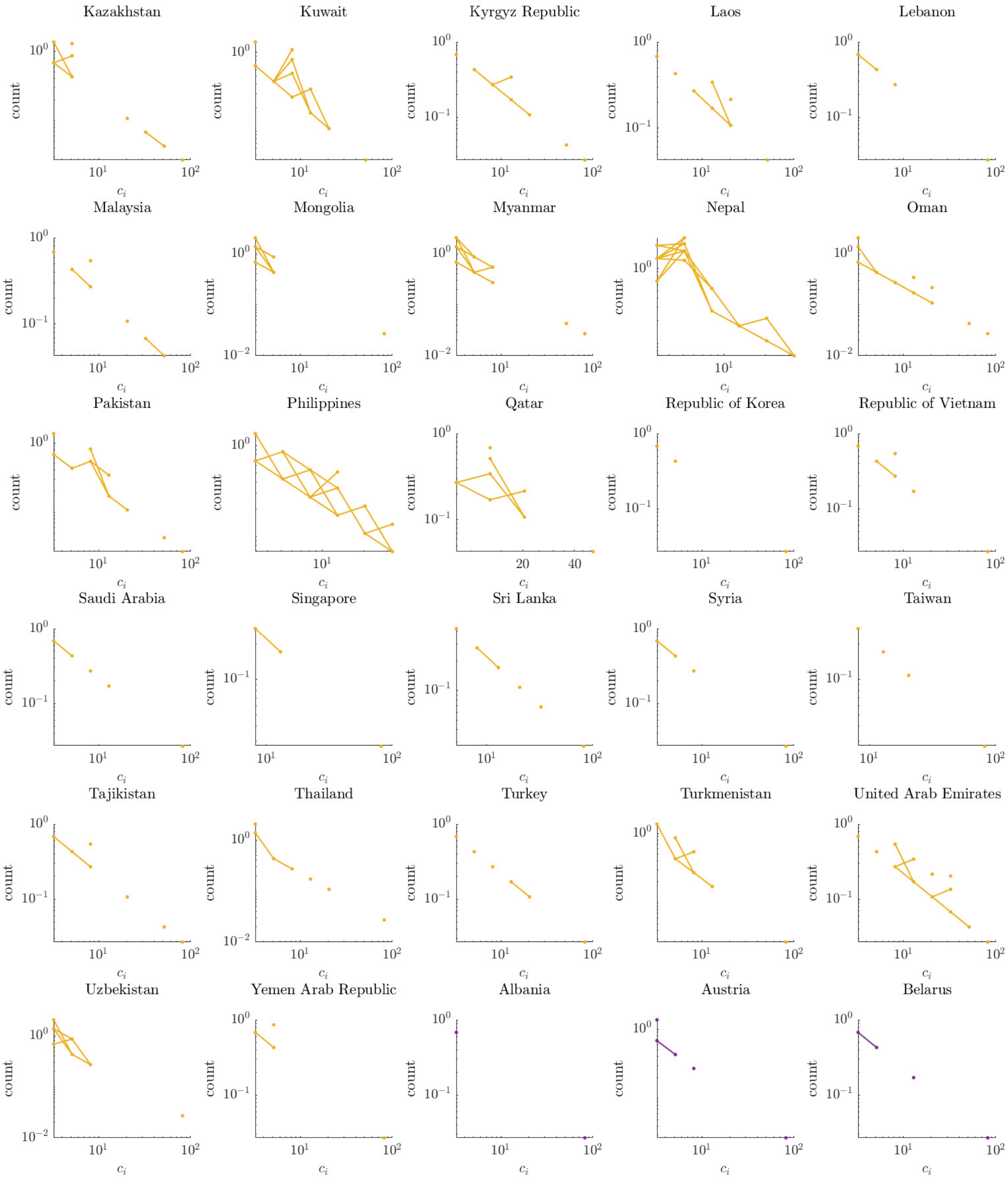


Figure S8: Group size distribution for the ethnolinguistic dataset (4/6). Continuous lines identify points in neighboring bins on the same year.

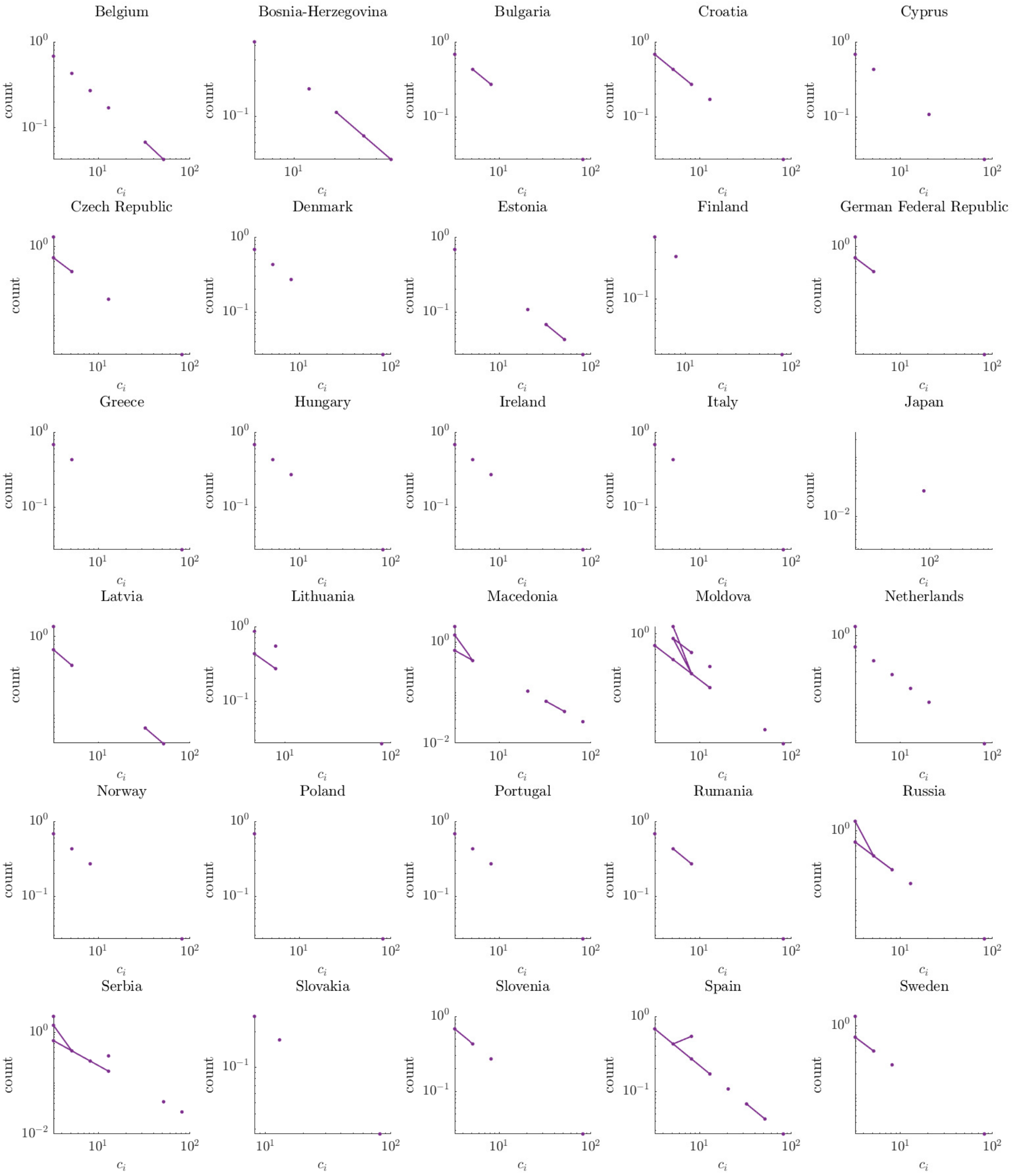


Figure S9: Group size distribution for the ethnolinguistic dataset (5/6). Continuous lines identify points in neighboring bins on the same year.

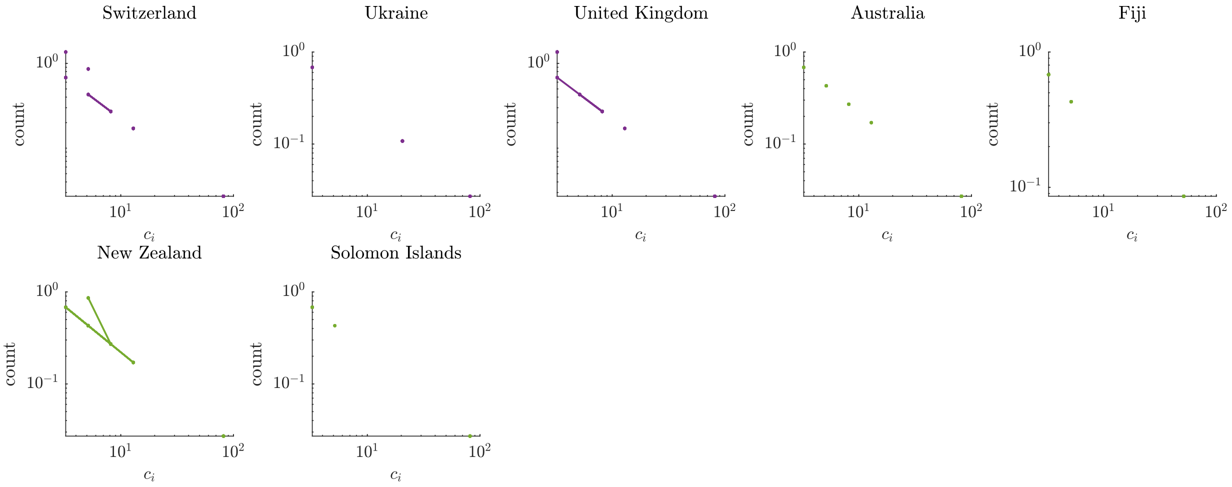


Figure S10: Group size distribution for the ethnolinguistic dataset (6/6). Continuous lines identify points in neighboring bins on the same year.

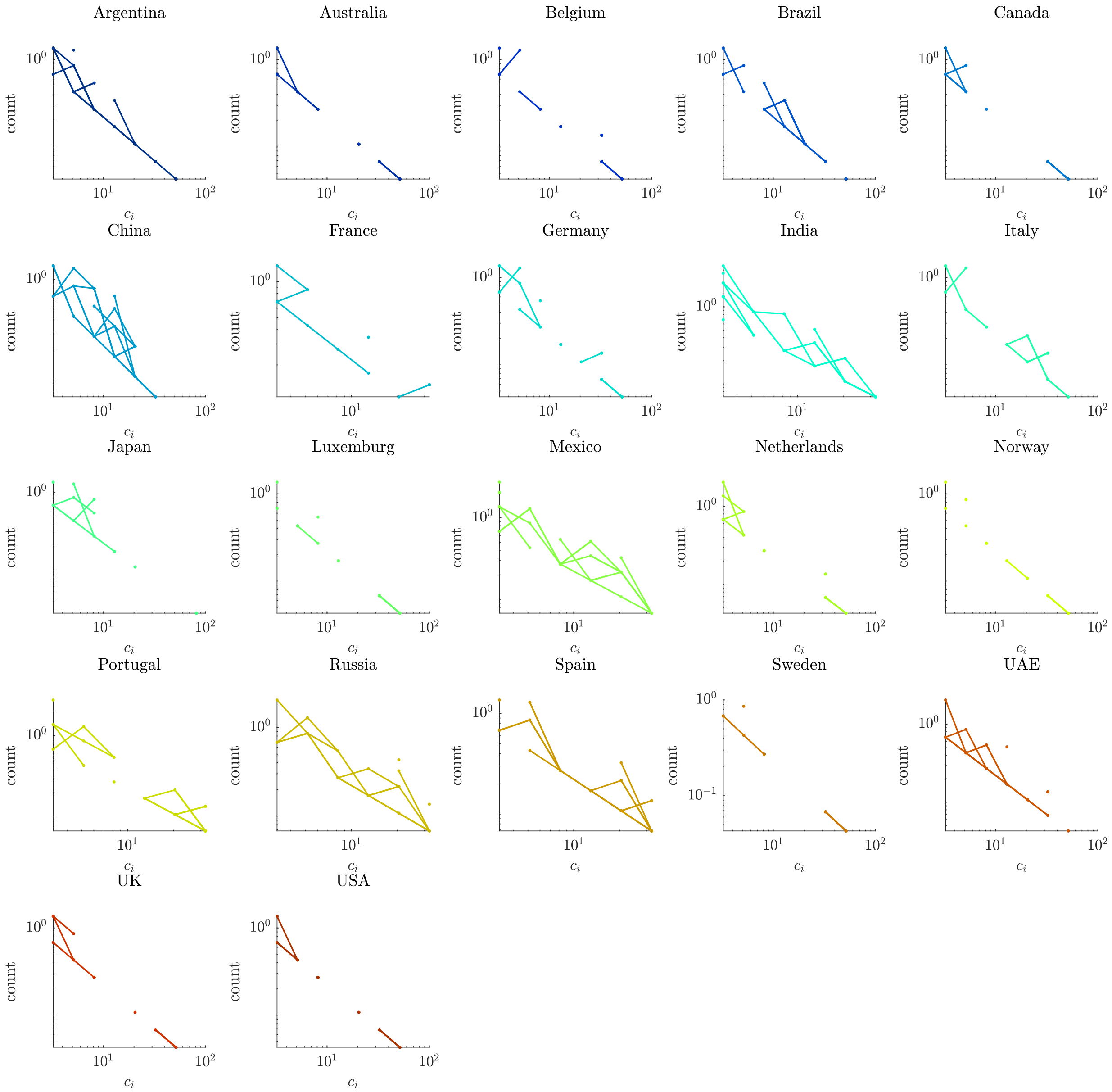


Figure S11: Group size distribution for the smartphone market share dataset. Continuous lines identify points in neighboring bins on the same year.

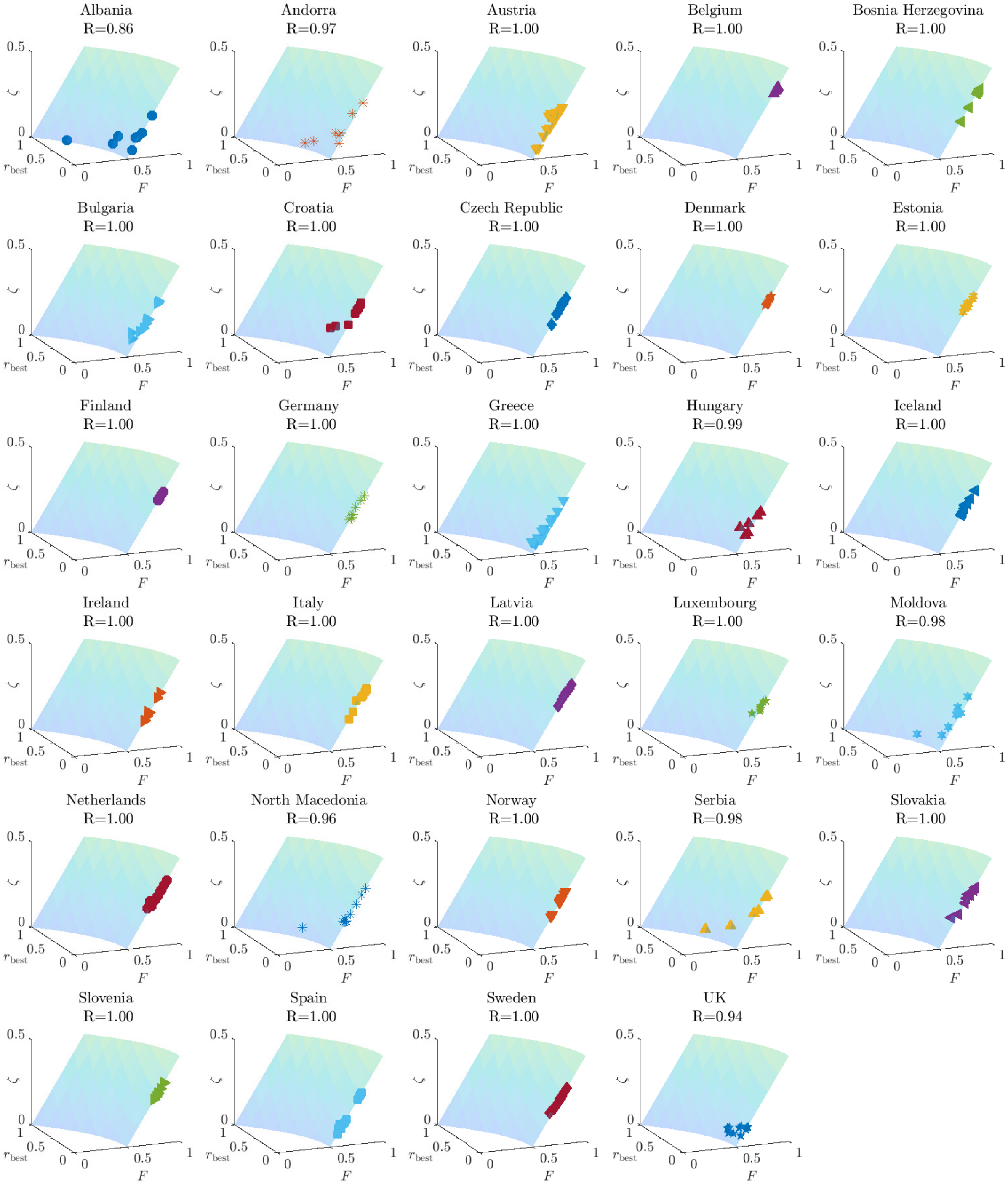


Figure S12: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the parliamentary networks dataset. $R = \text{corr}(\zeta,\, F)$.

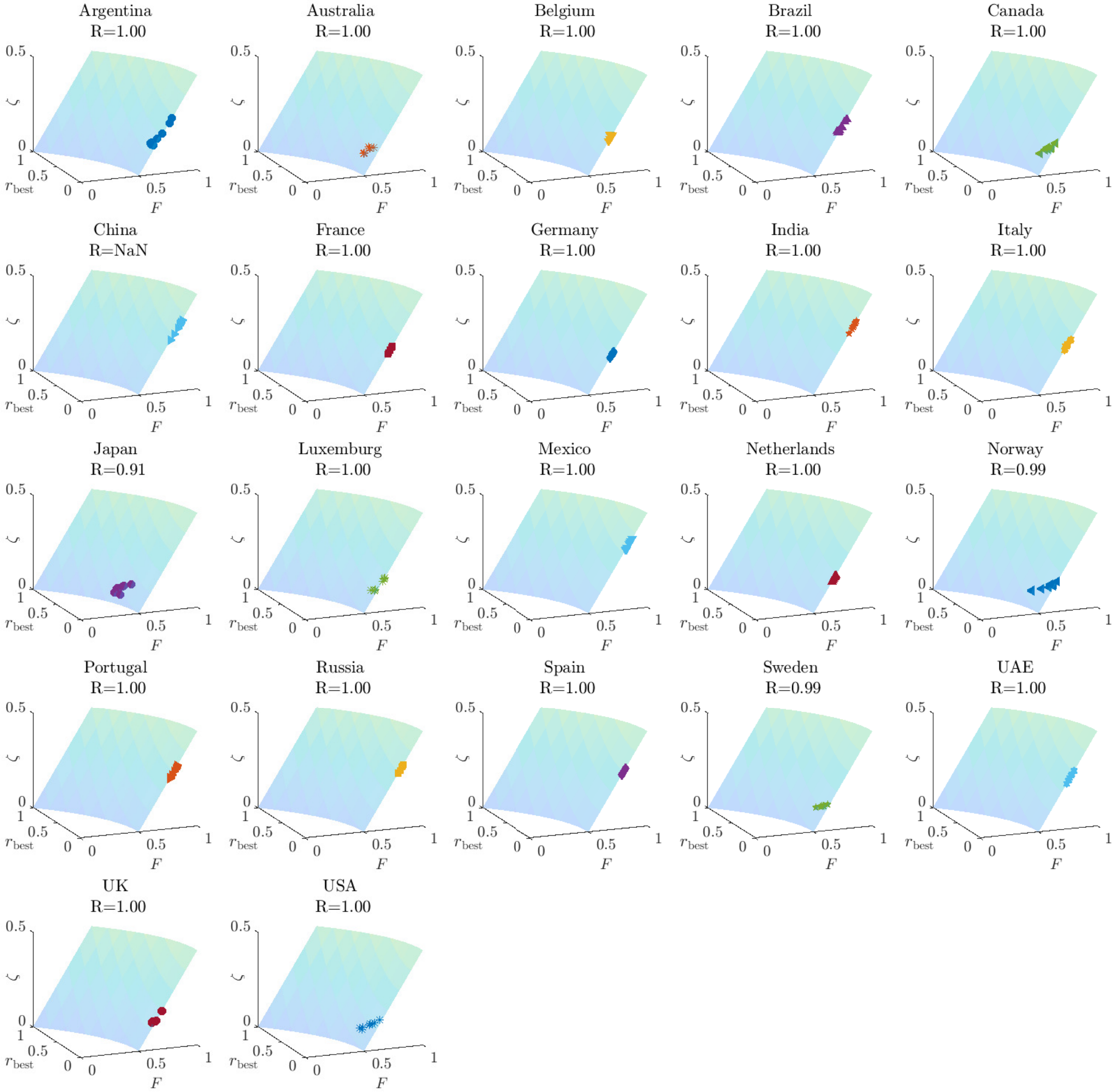


Figure S13: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the smartphone market shares dataset. $R = \text{corr}(\zeta,\, F)$.

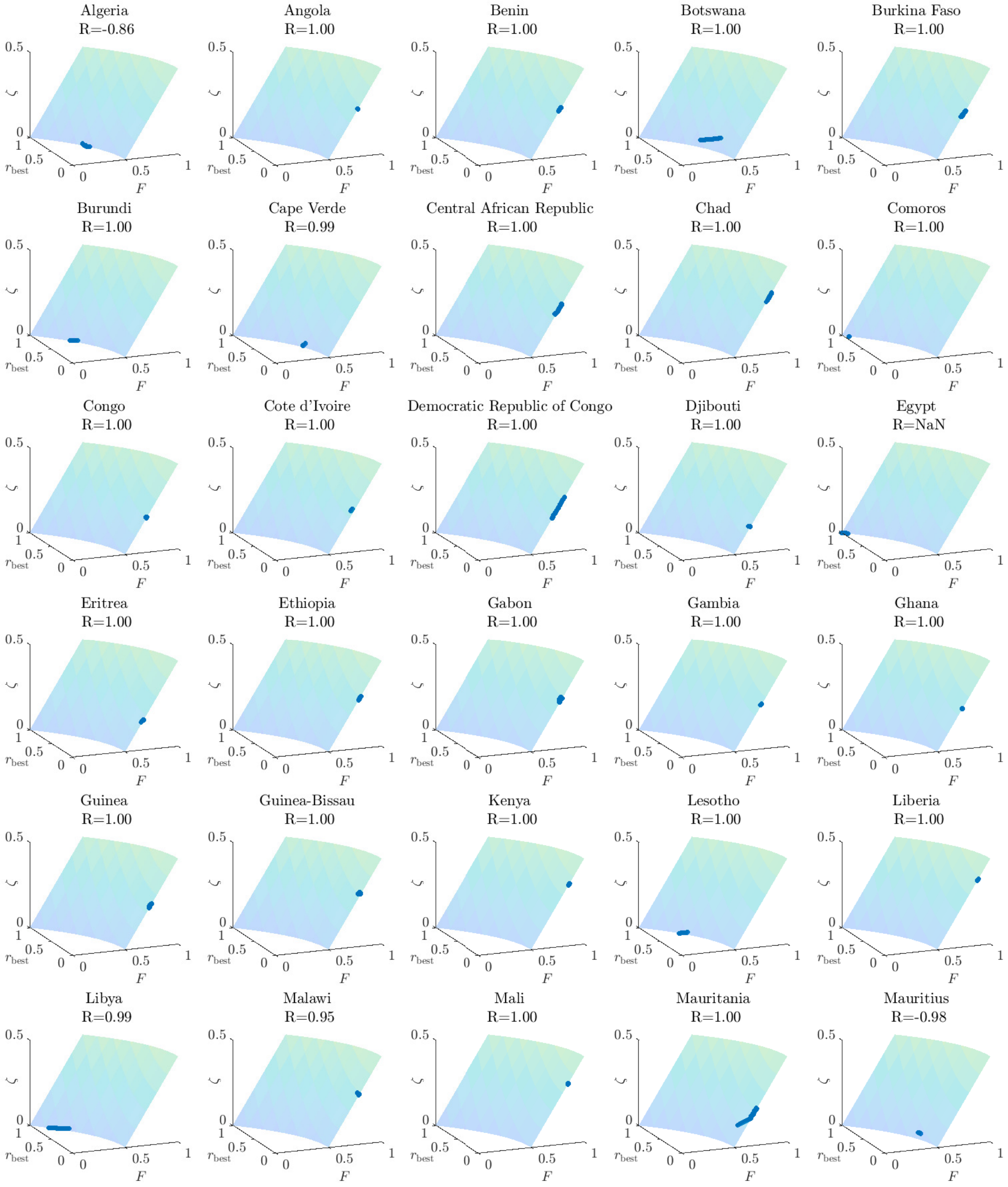


Figure S14: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (1/6). $R = \text{corr}(\zeta, F)$. $R =$ NaN occurs when a country has at most two ethnical groups, meaning that its signed graph is strongly balanced (and $\zeta = 0$).

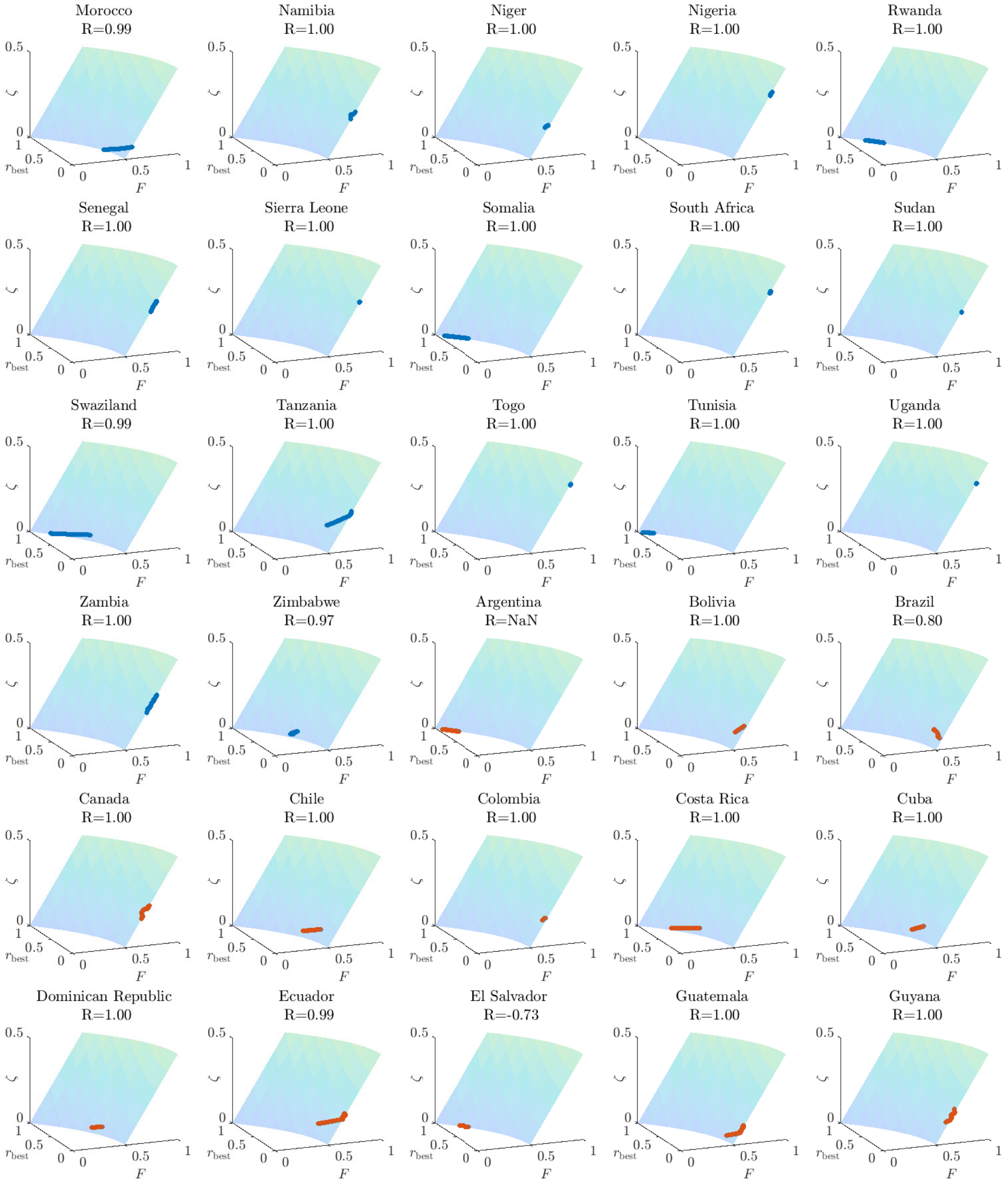


Figure S15: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (2/6). $R = \text{corr}(\zeta, F)$. $R = \text{NaN}$ occurs when a country has at most two ethnical groups, meaning that its signed graph is strongly balanced (and $\zeta = 0$).

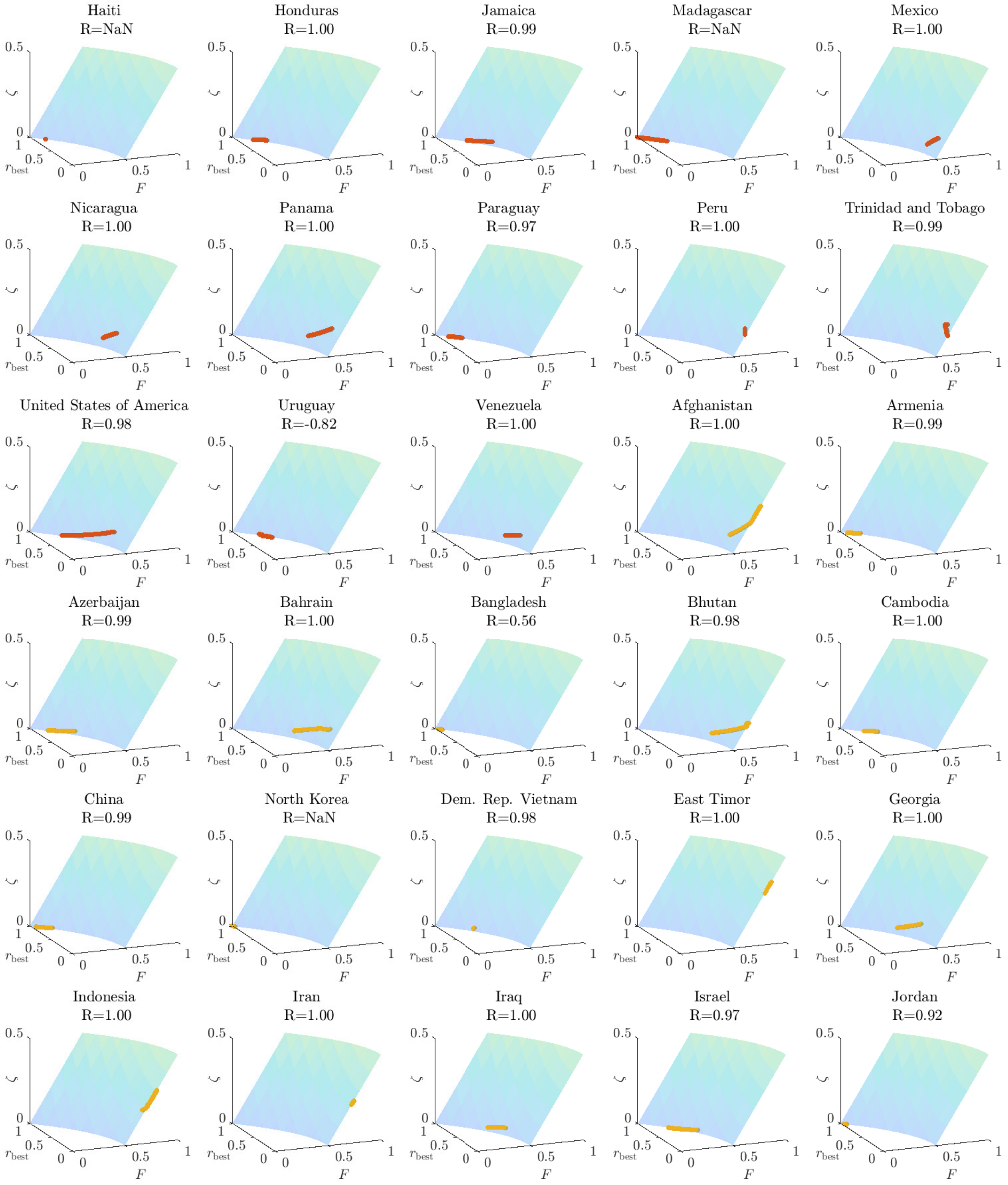


Figure S16: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (3/6). $R = \text{corr}(\zeta, F)$. $R = \text{NaN}$ occurs when a country has at most two ethnical groups, meaning that its signed graph is strongly balanced (and $\zeta = 0$).

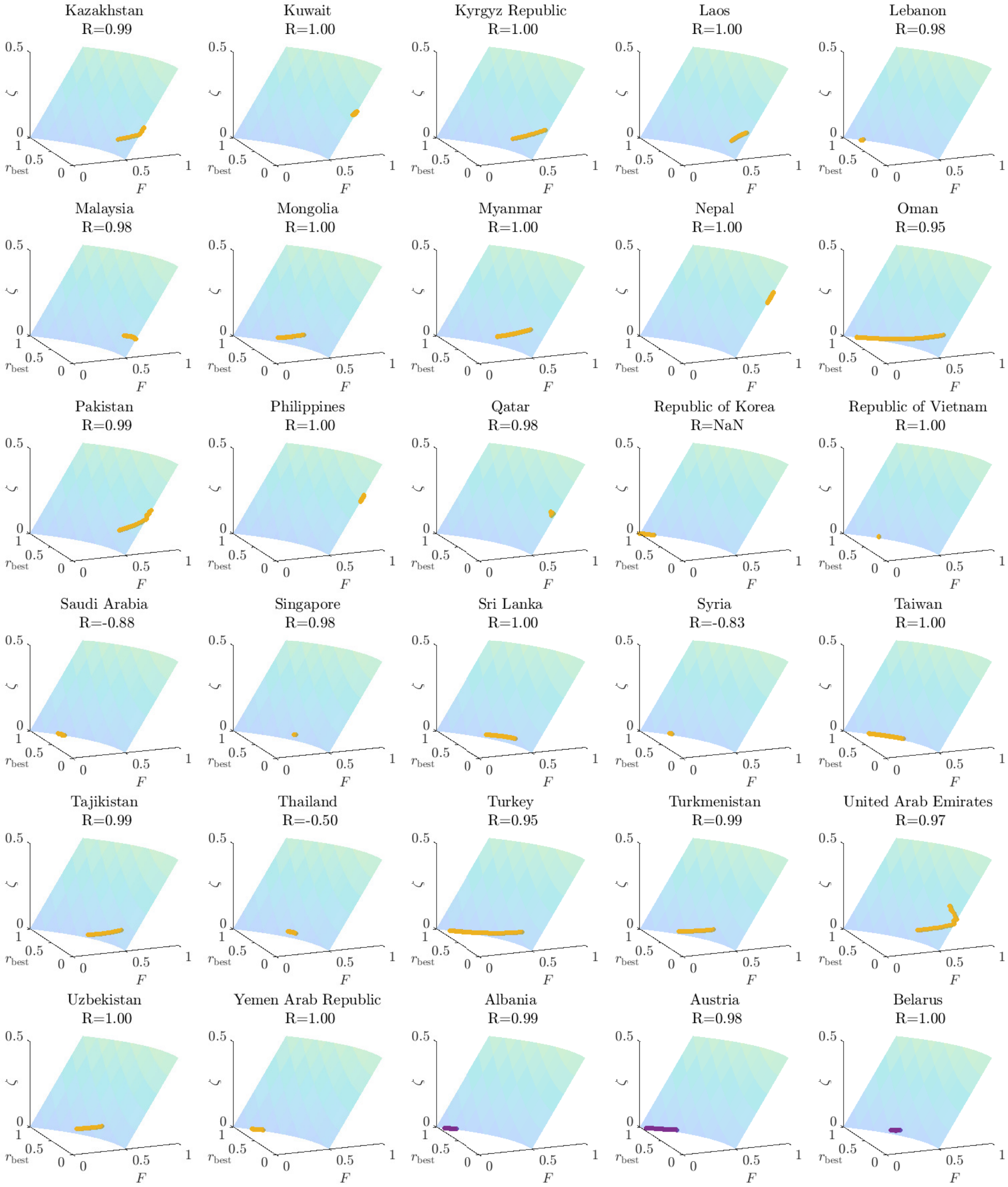


Figure S17: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (4/6). $R = \text{corr}(\zeta, F)$. $R = \text{NaN}$ occurs when a country has at most two ethnical groups, meaning that its signed graph is strongly balanced (and $\zeta = 0$).

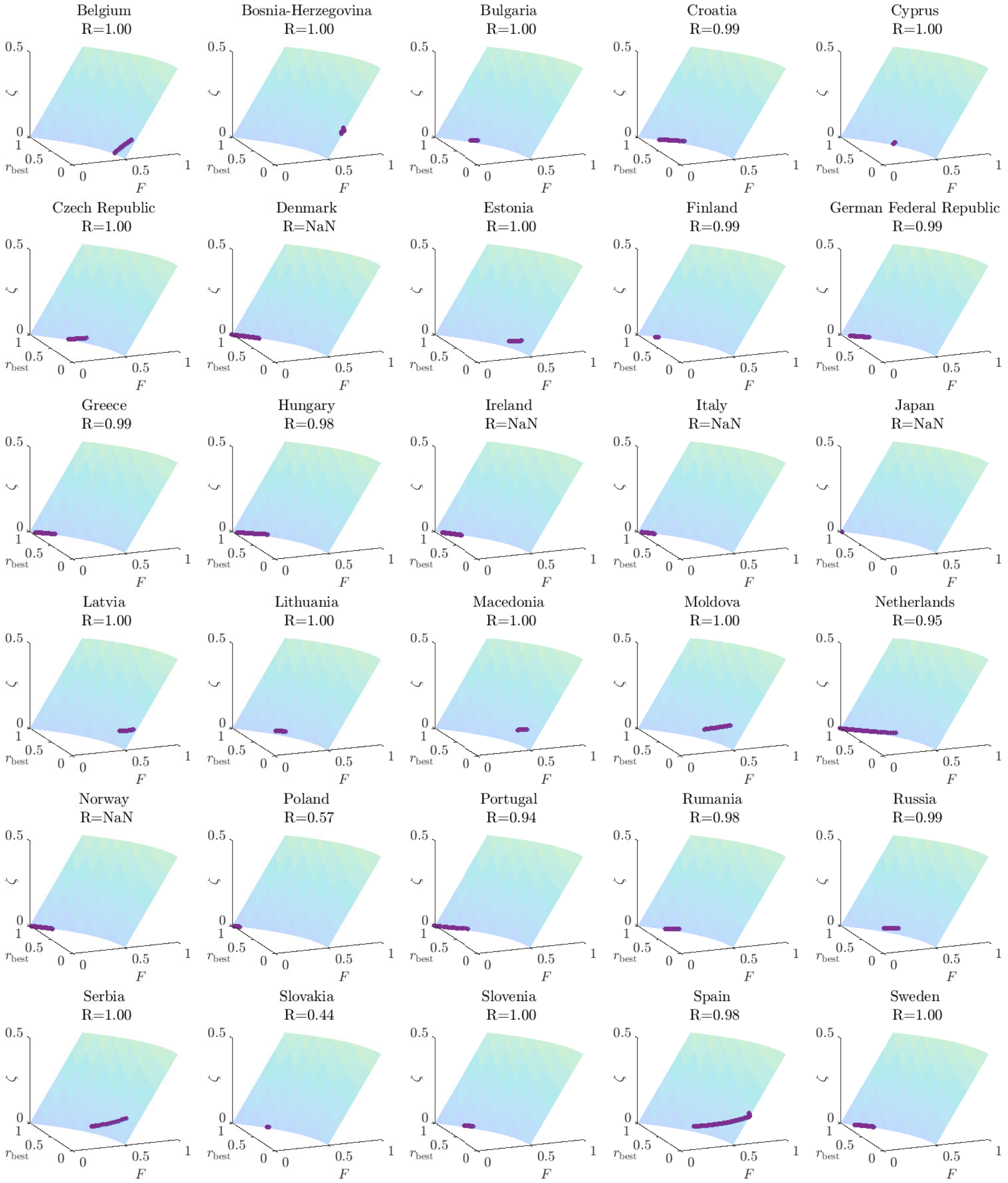
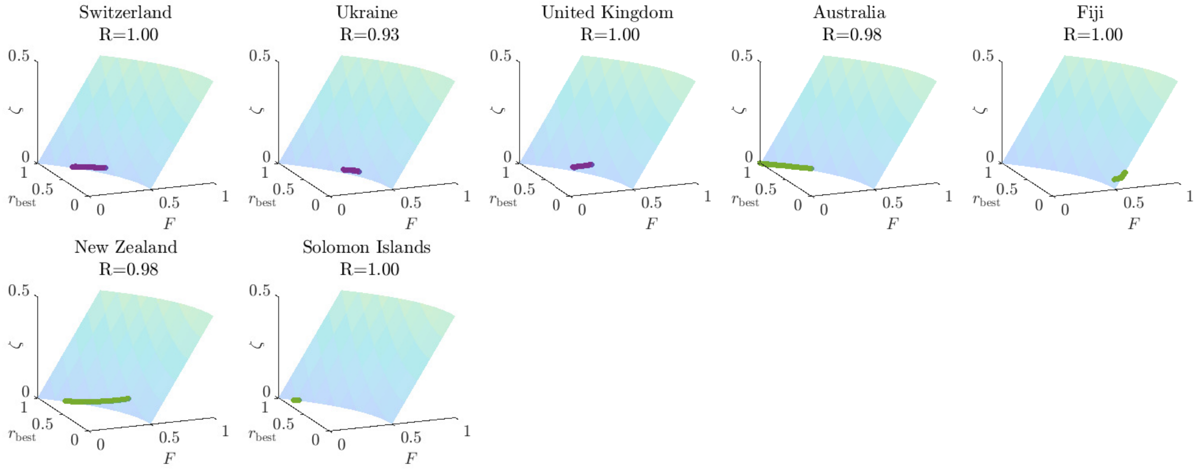


Figure S18: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (5/6). $R = \text{corr}(\zeta,\, F)$. $R = \text{NaN}$ occurs when a country has at most two ethnical groups, meaning that its signed graph is strongly balanced (and $\zeta = 0$).



Switzerland R=1.00

Ukraine R=0.93

United Kingdom R=1.00

Australia R=0.98

Fiji R=1.00

New Zealand R=0.98

Solomon Islands R=1.00

Figure S19: Frustration, fractionalization, and $r_{\text{best}}$ per country, for the ethnolinguistic networks database (6/6). $R = \text{corr}(\zeta,\, F)$.

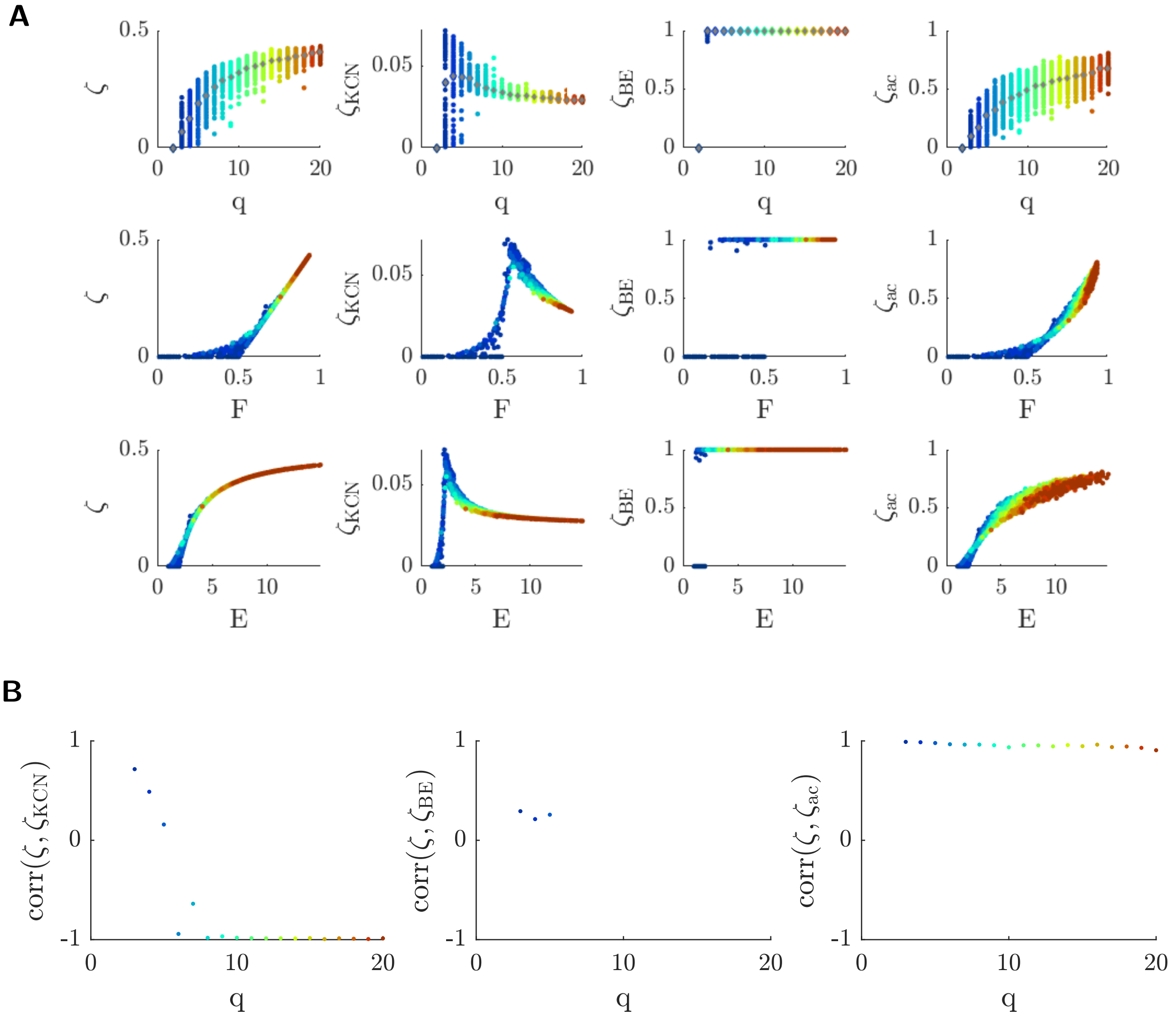


Figure S20: Frustration and other measures of unbalance: a comparison. Four measures of “distance to strong balance” are compared: $\zeta$, $\zeta_{\mathrm{KCN}}$, $\zeta_{\mathrm{BE}}$, and $\zeta_{\mathrm{ac}}$. (**A**): First row: values of the 4 measures on 100 samples of weakly balanced signed graphs with $q = 2, \ldots, 20$ groups. The color code follows the group size (as in Fig. 2 of the main paper). The grey diamonds represent the average over 100 instances. Second and third rows: scatter plots of the 4 measures vs $F$ and $E$. (**B**): Correlation between $\zeta$ and $\zeta_{\mathrm{KCN}}$, $\zeta_{\mathrm{BE}}$, and $\zeta_{\mathrm{ac}}$.