Decoding Segregation:

The topological landscape of segregation forms and a bottom-up ontology

Supplementary information 1: Additional analyses

Intersectionality in segregation research

How have growing concerns with the intersectionality between different social identities in sociological and cultural analysis related to segregation research? We approached this question by identifying documents that combined segregation forms addressing social identities or positions used in quantitative intersectionality analyses: (1) sex/gender/sexual orientation, (2) race/ethnicity, (3) SES (socioeconomic status)/income, (4) age/generation, (5) immigration status and nationality, (6) health and disability, (7) religion/religiosity (cf. Bauer et al., 2021). Segregation forms combining at least two of these positions were considered intersectional SFs (trigrams such as "adult gender segregation" or "racialized economic segregation"). However, these intersectional SFs were relatively infrequent in our dataset. Instead, we primarily focused on the more frequent co-occurrences of SFs that addressed one or more social positions within a single work, which we referred to as "intersectional works" in segregation research (e.g. "racial segregation" and "gender segregation"). The first intersectional co-occurrences in the sample appeared in 1977, combining economic and racial positions (namely, racial residential, socioeconomic residential and class segregation – Farley, 1977). We identified the largest intersectional co-occurrences of SFs accumulated in 2022 in the Scopus corpus (Fig. S1a). We applied Shannon Entropy as an intersectionality index to the distribution of intersectional COs relative to the total number of COs in the Scopus corpus over time (Fig. S1b). There is substantial growth in combinatorial work along intersectional lines in SR. Race/ethnicity is the most frequent position in those combinations.

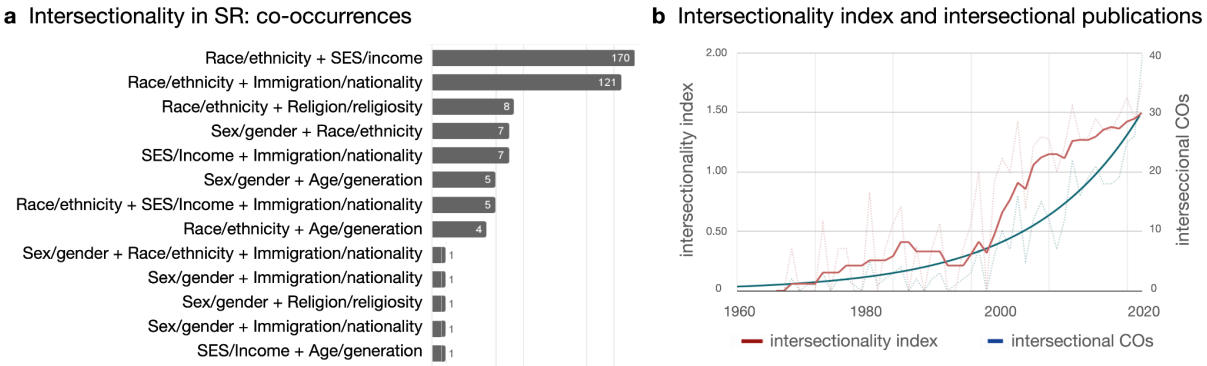

Fig. S1 – (a) Intersectional co-occurrences with the most publications accumulated in 2022 (logarithmic scale). (b) Intersectionality index based on intersectional co-occurrences measured through Shannon Entropy (red line) and the number of intersectional COs appearing in publications per year (blue line, 10 years moving average).

Our examination of SR through the lens of multidisciplinarity, transdisciplinarity, and intersectionality reveals specific trends in the field. Since the 1990s, there has been a notable increase in the number of disciplines and publications addressing segregation. However, more recent years show a concentration of research within fields such as Sociology and Political Science. The transdisciplinary nature of SFs, such as racial and residential segregation, has expanded, with many forms now crossing disciplinary boundaries. However, intersectional research – particularly studies combining multiple social identities like race, gender, and socioeconomic status – remains relatively limited and

centered around race/ethnicity. While SR demonstrates increasing multidisciplinarity and transdisciplinarity, intersectionality continues to emerge as an area of potential expansion.

Knowledge production networks

How is knowledge produced in segregation research? Now we look into how works relate to each other as they are created, the main production sites country-wise, the landscape of collaboration, and the topology of journals publishing works on the multifaceted subject.

1. Co-citation network of works in SR

The relationship between different works on segregation can be grasped by how they cite existing ones. One way to do this is to consider the collection of works cited together by documents in a corpus ("co-citation"). This differs from a direct citation network, which connects citing papers to cited papers and helps trace the chronological development of ideas. The co-citation network is undirected and built through database documents with common references. Two papers are said to be co-cited if at least one document cites both. The strength of the relationship between works is indicated by the number of times they are cited together (Osareh, 1996). The approach can capture networks of related documents. Each node represents a document, and an undirected edge between nodes A and B is formed if one or more other documents cite together both documents A and B. This implies that the network analysis goes beyond works in the Scopus corpus to cover 473,706 references – an x-ray of the field. For visualization purposes, we adopted ten as a minimum number of co-citations, i.e. the network shows only works which are co-cited ten or more times by the documents in the sample, with a limit of 1,000 most co-cited documents (i.e. nodes with the highest total link strength or the sum of the number of times the reference appears cited with other works) reduced to 998 after references were checked, and 2,000 visible links (Fig. S2). Node sizes indicate the citation number.

Works were then grouped into six clusters based on network modularity or the density of links inside communities compared to links between communities, calculated through the “smart local moving algorithm” (Waltman & Van Eck, 2013). Clusters show nodes with internal thematic affinities. However, research themes are not exclusive to specific clusters. Cluster 1 (red) is the largest and encompasses works on social, ethnic and residential segregation, including emphases on method and measurement. The article "The dimensions of residential segregation" (Massey & Denton, 1988) is the most cited in this cluster (followed by Duncan & Duncan, 1955), and in the sample as a whole. Cluster 2 (green) concentrates on works related to schools, race, intergroup contact, homophily, social networks and measurement. "The nature of prejudice" (Allport, 1954) is the most cited work. Cluster 3 (dark blue) brings works on urban segregation mostly related to poverty and race. The article "When work disappears: the world of the new urban poor" (Wilson, 1996) is the most cited document. Cluster 4 (ochre) includes a mix of works on policy, race, neighborhoods and ghettos, including the articles "The truly disadvantaged" (Wilson, 1987) and "American apartheid: segregation and the making of the underclass" (Massey, 1990), along with dynamic models of residential segregation since Schelling (1969). The most visually identifiable cluster is 5 (purple), including works on gender, sex, occupational segregation and their intersections with race, education. Cluster 6 (light blue) includes racial and ethnic features. The book "American apartheid" (Massey & Denton, 1993) is the most cited work. The full list of documents with clusters and publication data is available in Table ST13.

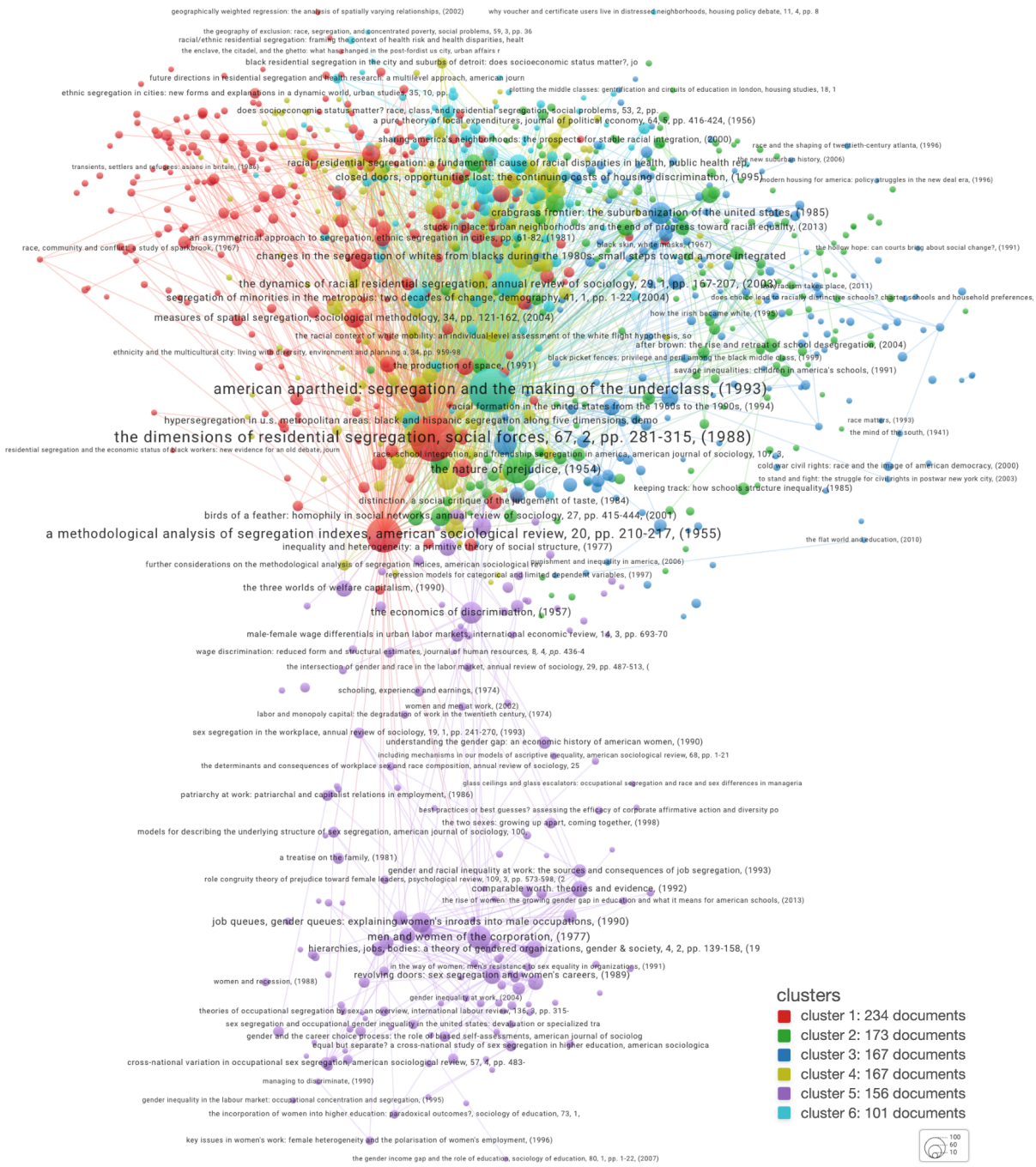

Fig. S2 – Co-citation network of documents in SR. This representation brings the most co-cited 998 documents within the network. Node sizes represent the number of citations. For a comprehensive view, refer to Table ST13 with clusters and the interactive network (layout parameters: attraction 8, repulsion 3, rotate 90 degrees; 2,000 visible links).

2. Co-authorship network across countries

We analyzed the field as an international collaboration network using a specialized type of bibliometric analysis where nodes represent countries, and edges represent co-authorship relationships between authors from different countries, identified through author affiliations. The node size (degree centrality) shows the number of unique documents in which authors from a particular country collaborate, providing insight into hierarchical structures. The weight of an edge is determined by the number of co-authored publications between authors from the two countries, reflecting the strength of

collaboration. For visualization purposes, the minimum number of documents associated with a country is five. Betweenness centrality (see network and Table ST14) confirms the enormous presence of researchers working in the US context, followed by the UK and a block of countries including Australia, Germany, Netherlands and France. Clusters show which countries appear more frequently together in the collaboration network, and are based on network modularity (Fig. S3a).

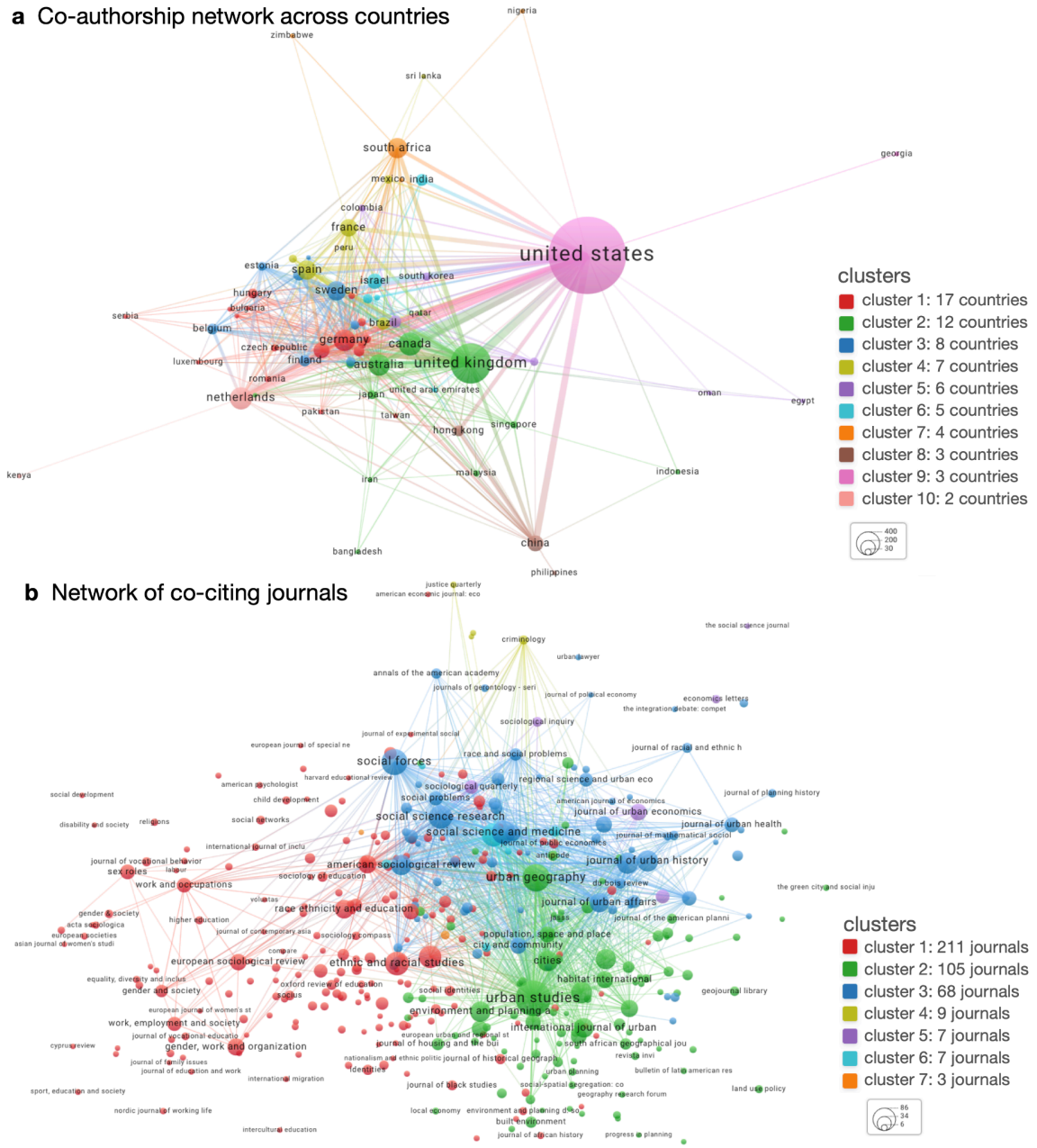

Fig. S3 – (a) Co-authorship network across countries (see the interactive network including betweenness centrality; layout parameters: attraction 6, repulsion 4, straight links). (b) Bibliographic coupling. Node size indicates the number of documents on segregation published by the journal (interactive network including betweenness centrality; attraction 5, repulsion 3).

Regarding world region clusters, they are internally diverse, showing that distance is frequently overcome in co-authorship. Cluster 1 (red) includes central European countries such as Germany, Italy, Switzerland, Eastern European countries (e.g. Hungary, Slovakia, Bulgaria), Argentina, Pakistan and Taiwan. Cluster 2 (green) includes UK, Australia and Canada as English-speaking countries, Japan, Singapore and Malaysia. Cluster 3 (dark blue) includes Sweden, Norway, Finland, Belgium and Poland; Cluster 4 (ochre) includes France, Spain and Brazil (see Table ST14).

3. Co-citing journals

When applied to journals, bibliographic coupling focuses on identifying those that share sources, revealing similarities or complementarities in their research topics and focus. Journals A and B are bibliographically coupled if they cite the same journal or set of journals. The strength of the bibliographic coupling is determined by the number of shared references between the two journals. Such connectivity generates a network where each node represents a journal. An edge between two nodes represents a bibliographic coupling relationship. The weight of an edge is determined by the number of references shared between the two journals, reflecting the strength of their bibliographic coupling. Figure S3b shows the top 1,000 journals with the greatest total link strength. The minimum number of citations of a journal considered was five. Node size indicates the number of documents on segregation published by each journal.

The three main clusters show disciplinary and thematic affinities. Cluster 1 (red) contains journals primarily associated with work, gender, ethnicity, race and education. Cluster 2 (green) includes a strong urbanistic and geographic emphasis, suggesting that journals focused on spatial dimensions of segregation tend to cite each other more strongly than those associated with other disciplines. Cluster 3 (dark blue) contains a more diverse set of journals, including sociology and urban history. In turn, the betweenness centrality analysis (interactive network) shows what journals from different disciplines emerge as the most central nodes in segregation research, namely Urban Studies, followed by Urban Geography, Social Forces, Demography and Ethnic and Racial Studies (Table ST15).

Our analysis of the knowledge production networks in segregation research illustrates a richly interconnected field driven by thematic diversity, international collaboration, and disciplinary intersections. Through co-citation, co-authorship, and bibliographic coupling analyses, we observe how various disciplines contribute to a nuanced understanding of segregation, linking concepts across social, urban, and ethnic studies. This global collaboration network reveals prominent contributions from the US and UK, with significant clusters of activity across Europe, Australia, and beyond, reflecting a collective effort to explore the complex, multifaceted nature of segregation. Together, these networks underscore the expansive, multidisciplinary reach of segregation research and its reliance on collaborative knowledge exchange.