Physica A 483 (2017) 423–437

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Income inequality: A complex network analysis of US states

Periklis Gogas [a], Rangan Gupta [b], Stephen M. Miller [c,*], Theophilos Papadimitriou [a], Georgios Antonios Sarantitis [a]

[a] Department of Economics Democritus University of Thrace, Komotini, 69100, Greece
[b] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa
[c] Department of Economics, University of Nevada, Las Vegas, Las Vegas, NV, 89154-6005, USA

## HIGHLIGHTS

- Perform long-run, inter-temporal analysis of US income inequality over 1916–2012.
- Use two alternative measures of inequality: Top 1% income and Gini coefficient.
- Use descriptive analysis and Threshold-Minimum Dominating Set from Graph Theory.
- Heterogeneous evolution of inequality exists across four focal sub-periods.
- Results differ between Top 1% income and Gini coefficient.

## ARTICLE INFO

*Article history:*
Received 25 November 2016
Available online 2 May 2017

*JEL classification:*
D63

*Keywords:*
Income inequality
Graph theory
US states

## ABSTRACT

This study performs a long-run, inter-temporal analysis of income inequality in the US spanning the period 1916–2012. We employ both descriptive analysis and the Threshold-Minimum Dominating Set methodology from Graph Theory, to examine the evolution of inequality through time. In doing so, we use two alternative measures of inequality: the Top 1% share of income and the Gini coefficient. This provides new insight on the literature of income inequality across the US states. Several empirical findings emerge. First, a heterogeneous evolution of inequality exists across the four focal sub-periods. Second, the results differ between the inequality measures examined. Finally, we identify groups of similarly behaving states in terms of inequality. The US authorities can use these findings to identify inequality trends and innovations and/or examples to investigate the causes of inequality within the US and implement appropriate policies.



## 1. Introduction

The distribution of income and/or wealth between the rich and the poor has received significant research effort, attracting interest from politicians, academics, and policy makers. Most studies reach the general conclusion that high income inequality existed during the 1920s and the consequent Great Depression, followed by a period of convergence and finally divergence, once again, in more recent years, especially after the latest global financial crisis of 2007–2009.

Piketty [1] recently conducted a global analysis of income inequality. He concludes *inter alia* that for most of the developed countries, income inequality fell in the period after the two World Wars and re-surged in the 1980s. In related work on the

* Corresponding author.
*E-mail addresses:* pgkogkas@econ.duth.gr (P. Gogas), rangan.gupta@up.ac.za (R. Gupta), stephen.miller@unlv.edu (S.M. Miller), papadimi@econ.duth.gr (T. Papadimitriou), gsaranti@econ.duth.gr (G.A. Sarantitis).

http://dx.doi.org/10.1016/j.physa.2017.04.102

US states, Saez [2] concludes that 95% of the growth during the recovery from the Great Recession occurred in the Top 1% of the income distribution. Rose [3] disputes the implication of Saez's claim, arguing that Saez chose a misleading sample period. He uses Piketty's data and argues that the wealthiest 1% of Americans experienced the largest loss of income over 2007–2008 despite the gain in income over 2009–2012. Then, using Congressional Budget Office (CBO) data [4] on a broader measure of income that includes transfer income and excludes taxes paid, Rose [3] notes that although inequality measured by the Gini coefficient increases for market income over 2007–2011, it falls when considering the income measures that adjust for (a) transfer payments and (b) transfer payments and taxes.

In sum, the relevant literature does not offer a consensus due to the use of different sample periods, different measures of income, and different measures of inequality. This paper considers the inter-temporal evolution of inequality in the US states, using annual state-level data from 1916 to 2012 constructed by Frank [5]. Our sample period includes a series of "Great" episodes: the Great Depression (1929–1944), the Great Compression (1945–1979), the Great Divergence (1980–present), the Great Moderation (1982–2007), and the Great Recession (2007–2009).

Goldin and Margo [6] popularized the term Great Compression for the period following the Great Depression, an era during which the income inequality between the rich and the poor was greatly reduced in relation to prior periods (e.g., the Great Depression). Krugman [7] called the period following the Great Compression, the Great Divergence, when income inequality began to increase once again. Piketty and Saez [8] argue that in the US, the Great Compression ended in the 1970s and then reversed itself.[1]

Our study strays from the classic econometric paths and presents an empirical analysis that evolves within a Graph Theory context. In particular, we employ a new Complex Networks optimization technique called the Threshold-Minimum Dominating Set (T-MDS) to describe the evolution of income inequality in the US between 1916 and 2012. Graph Theory has met wide acceptance in the analysis of complex economic systems [10–17]. It possesses an advantage over the typical econometric analysis in that it can deliver multi-level analysis of the studied system, ranging from the network to the agent-specific level. Graph Theory can, thus, capture the dynamic, non-linear effects that take place in a complicated system of interacting agents instead of just inferring on the system as a whole (see, e.g., the studies of Hu et al. [18], Di Matteo et al. [19], and Markey-Towler and Foster [20] on income inequality through a complex network prism). The use of the T-MDS technique, in particular, allows inferences on the aggregate network's evolution as well as on the local neighborhood of each node.

Therefore, by working within a Graph Theory context and applying the T-MDS technique, we may gain new insight into the inter-relations of the US states with respect to income inequality. More specifically, we present new empirical results using (a) a data set that spans nearly 100 years and (b) two alternative inequality measures (Top 1% share of income and the Gini coefficient). We find that income inequality within the US displays heterogeneous patterns inter-temporally, reaching its peak values in the more recent years. We also report that the results differ slightly according to the selected inequality measure. We identify groups of closely behaving states that federal and state's authorities may use to design and implement more efficient tax policies and structural economic reforms. Finally, we are the first to apply the T-MDS methodology in this area.

We organize the paper into the following sections. Section 2 describes the data set and presents the descriptive data analysis. Section 3 outlines the methodological context and explains the use and possible interpretation of the T-MDS technique. Section 4 provides and explains the empirical findings. Section 5 compares the empirical results with the relevant literature. Finally, Section 6 briefly recapitulates and concludes the paper.

## 2. Data and descriptive analysis

### 2.1. Data

Frank [5] constructs inequality measures using data published in the IRS's *Statistics of Income* on the number of returns and adjusted gross income (before taxes) by state and by size of the adjusted gross income. The pre-tax adjusted gross income includes wages and salaries, capital income (dividends, interest, rents, and royalties) and entrepreneurial income (self-employment, small businesses, and partnerships). Interest on state and local bonds and transfer income from federal and state governments do not appear in this measure of income. For more details on the construction of the inequality measures, see [5, Appendix].

The IRS income data are considered problematic because of the truncation of individuals at the low-end of the income distribution. Frank [5] notes that the IRS will penalize tax payers for misreporting income, whereas [21] argue that survey-based alternatives to the IRS data introduce bias of "over-reporting of earnings by individuals in the lower tail of the income

[1] In a recent paper, Kaplan and Rauh [9] argue that economic factors provide the most logical explanation of rising income inequality. That is, "skill-based technological change, greater scale, and their interaction" (p. 53) create the necessary ingredients for demand and supply factors to generate a growing income inequality. They further reject the notion that income inequality reflects the collection of rents by individuals who "distort the economic system to extract resources in excess of their marginal products." (p. 52).

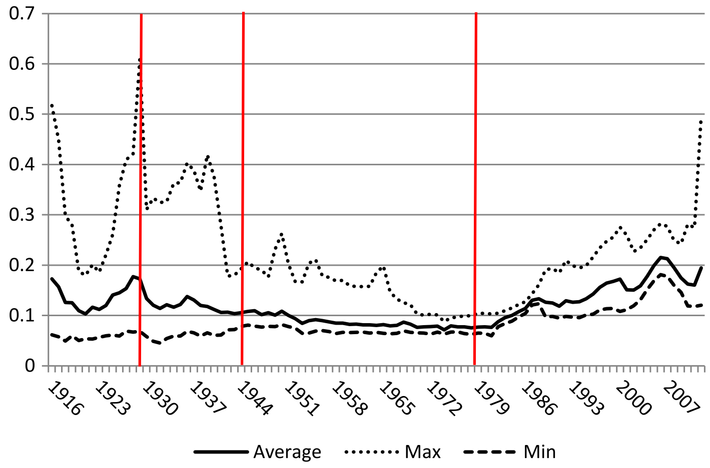


**Fig. 1.** Top 1% share of income.

distribution and under-reporting by individuals in the upper tail of the income distribution"[2] (p. 258). In our analysis, we use the Top 1% share of the income distribution, which [8,1] argue is less subject to the omission of individuals at the low end of the income distribution in the IRS data. Moreover, we also perform the same analysis using the Gini coefficient inequality measure[3] to compare the empirical findings. The IRS data afford a big advantage of reporting annual data by state for 97 years.[4]

## 2.2. Descriptive analysis

Based on the existing literature on the Great Depression, Great Compression, and Great Divergence, we identified 1929, 1944, and 1979 as the relevant focal points within the sample ranging from 1916 to 2012. Considering income inequality before the start of the Great Compression, decreases (increases) in capital income would improve (worsen) income inequality as capital income conforms to a most skewed distribution of the various components of total income. Piketty and Saez [8] argue that shocks to owners of capital during the Great Depression and World War II significantly reduced capital income. Moreover, Piketty and Saez [8] suggest that progressive income taxation provides the most probable explanation of the secular decline in capital income concentration. Krugman [7] argues that the Great Compression reflected not only progressive income taxation but also the policies of President Franklin Roosevelt that strengthened unions. Explanations for the duration of the Great Compression include the paucity of immigrants and union strength. Moreover, unions, along with social norms [8], provide an important check on excessive increases in executive pay. Analysts suggest that the ending of the Great Compression reflects technological change, globalization, and political and policy changes that reduced union strength. Krugman [7] argues that lower taxes on the rich and significant holes in the social safety net, beginning in the late 1970s and early 1980s, as well as the relative power of, and membership in, unions led to the end of the Great Compression and ushered in the Great Divergence. In addition, executive pay during this period rose considerably relative to average worker pay, reflecting relaxed social norms.

Fig. 1 plots the average of the states' Top 1% share of income from 1916 to 2012[5] along with the maximum and minimum values. We highlight the years 1929, 1944, and 1979 with vertical lines to distinguish the relevant sub-periods. Fig. 1 suggests that, on average, inequality fell during WWI and its immediate aftermath and then rose during the rest of the "roaring 20s", reflecting the downward movement in capital income that we mentioned above. Inequality then fell gradually from 1929 through 1979 and began rising through the end of the sample in 2012. Thus, we confirm the observations of the Great Compression and Great Divergence. Delaware experienced the highest inequality across all states from 1924 to 1971, achieving in 1929 the highest income share of the Top 1% that is measured in the sample, namely, 0.61.

Fig. 2 plots the standard deviation of the Top 1% share of income for each year from 1916 to 2012. In what we call the WWI+ period from 1916 to 1929, the inequality dispersion across states first converged (sigma-convergence) and then

2 The Census Bureau also provides state level data on the Gini index for every decade since 1969 and every year since 2006 for the newer American Community Survey. Unfortunately though, these sample frequencies do not provide enough observations to make a valid long-term comparison with the individual level data that we use in this study.

3 The Gini coefficient is constructed upon pretax income data.

4 We performed the same analysis using the Top 10% income share inequality measure as well. This measure yields results that are qualitatively similar to the ones of the Top 1% measure and we exclude them from the paper for brevity. These results are available upon request.

5 We also plotted the median of the Top 1%. The mean and median generally do not differ much from each other, suggesting that the asymmetry imagined from a visual inspection of Fig. 1 involves a small number of states. For example, in 1929, the Top 1% in 12 states exceeds 0.2 and in 2 states exceeds 0.3.

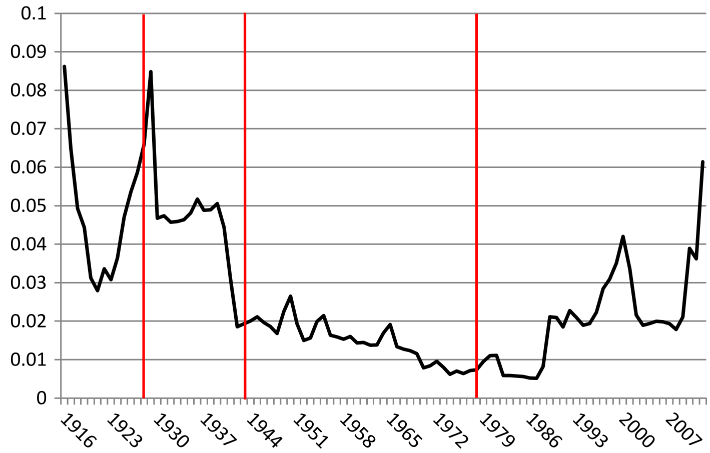


**Fig. 2.** Standard deviation of the top 1% share of income.

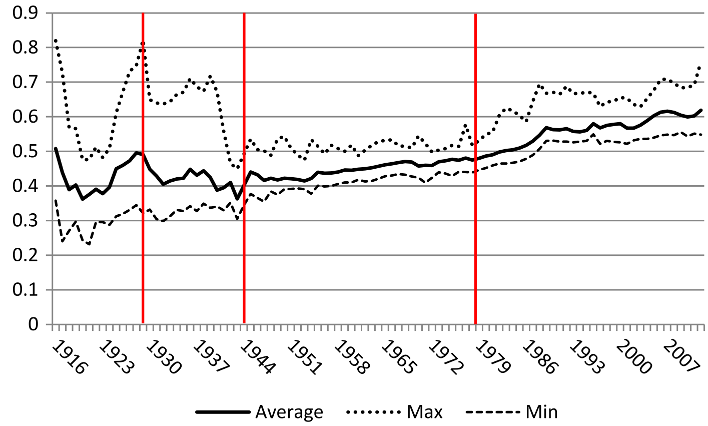


**Fig. 3.** Gini coefficient.

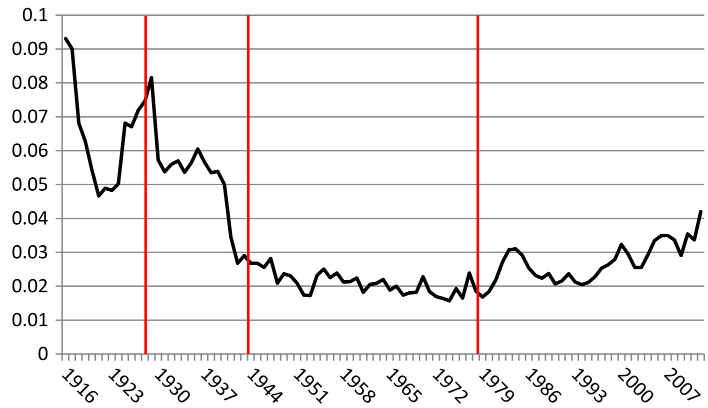


**Fig. 4.** Standard deviation of the Gini coefficient.

diverged during the 1920s. Convergence of the standard deviation of inequality among the individual states occurred during the Great Depression and the Great Compression, whereas it diverged, once again, during the Great Divergence era.

Figs. 3 and 4 plot the average Gini coefficient and its standard deviation, respectively, over the 1916–2012 period. According to these figures, inequality initially falls during WWI and its aftermath and then rises during the rest of the "roaring 20s", reflecting the corresponding movements in capital income. In the Great Depression period, the Gini coefficient falls slightly. Inequality gradually rises in the following two sub-periods, the Great Compression and the Great Divergence, as the Gini coefficient rises from 0.40 to 0.62.

According to these results, the evolution of the two inequality measures is qualitatively similar in all sub-samples with the exception of the Great Compression. This period covers 35 years of the total 97 included in our study (roughly one-third). Inequality according to the Gini coefficient increases gradually from 0.40 to 0.48 while according to the Top 1% share of income it slightly falls from 0.10 to 0.07. These results may not be contradictory as they seem at first. During the Great Compression, significant changes in the political, social, and economic norms may have driven these results. Important structural changes occurred, including progressive taxation and social resistance on excessive increases in executive pay [8], the increasing power of the unions [7], and reduced immigrant inflows. All these factors exerted a significant negative effect on the income of the Top 1% while they increased inequality in the lower income classes. Unionized labor and lower-to-medium management positions must have benefited the most from these changes to the expense of the Top 1% and the non-unionized and unskilled labor. This may be the manifestation of the "American Dream" during this period: the opportunity for upward social and economic mobility, prosperity, and success through hard work.

On the other hand, the results for the standard deviation of the Gini coefficient generally match those of the Top 1%. That is, in the WWI+ period, the inequality dispersion across states first converged and then diverged during the roaring 20s. Convergence of inequality dispersion occurred during the Great Depression and the Great Compression, whereas inequality dispersion diverged during the Great Divergence.

## 3. The methodology

### 3.1. Network construction

In representing an economic system as a Graph ($G$), we depict the economic agents as nodes ($N$) and the similarity of the nodes takes the form of edges ($E$) that link these nodes. Mathematically, we define $G = (N, E)$. In this study, the nodes of the network represent the 48 contiguous US states, excluding Alaska and Hawaii due to lack of data availability over the entire sample period, while the connecting edges reflect the similarity of the states using two inequality measures—the Top 1% share of the income distribution and the Gini coefficient. We calculate the similarity for both measures using the Pearson correlation coefficient $r$.

For both inequality measures we construct the networks that correspond to the four sub-periods of 1916–1929 (WWI+), 1930–1944 (Great Depression), 1945–1979 (Great Compression), and 1980–2012 (Great Divergence) and then we identify the T-MDS for each sub-period.[6] The use of these four sub-samples introduces a dynamic feature to our analysis.

### 3.2. Threshold-minimum dominating set

#### 3.2.1. T-MDS identification

To define the Threshold-Minimum Dominating Set (T-MDS), we must first introduce the simple Dominating Set (DS) and, then, the classic Minimum Dominating Set (MDS).

**Definition 1.** *Dominating Set* (DS) of a graph $G$ is a subset of nodes $N$ ($DS \subseteq N$) such that every node not in DS ($i \notin DS$) connects to at least one element of the DS ($\forall i \notin DS, \exists j \in DS : e_{ij} \in E.$), where $e_{ij}$ describes the edge connecting nodes $i$ and $j$.

The DS definition describes a subset of $N$, where every node in the network either lies adjacent to a DS node or is a DS node itself. Thus, since the network builds on pairwise correlations, the behavior of any non-DS node reflects on the behavior of its adjacent DS node(s).

To identify a DS, we start by creating $n$ binary variables $x_i$, $i = 1, \ldots, n$ one for each node of the network, such that:

$$x_i = \begin{cases} 0, & \text{if } i \notin DS \\ 1, & \text{if } i \in DS \end{cases}$$

to represent each node's membership status in the DS. Representing these variables in vector form produces $\mathbf{x} = [x_1 x_2, \ldots, x_n]$.

The DS notion takes the following mathematical form:

$$x_i + \sum_{j \in B(i)} x_j \geq 1, \quad i = 1, \ldots, n, \tag{1}$$

where $B(i)$ is the set of neighboring nodes of node $i$. Eq. (1) implies that each network node can either lie (a) in the DS (i.e., $x_i = 1$) or (b) adjacent to one or more DS nodes (i.e., $\exists j \in N(i) : x_i = 1$).[7]

We can identify many DSs for every network. Nonetheless, our interest focuses on the minimum sized one, which is defined as follows:

[6] The Great Compression [6] refers to the time of wage compression that occurred in the 1940s and 1950s. The reversal of this and the emergence of the Great Divergence did not occur until the late 1970s.

[7] This does not constitute a mutually exclusive relationship, as we may find nodes that verify both cases.

**Definition 2.** The *Minimum Dominating Set* (*MDS*) equals the *DS* with the smallest cardinality.

This definition conforms to the following relationship:

$$\min_x f(x) = \sum_{i=1}^{n} x_i. \tag{2}$$

Thus, the calculation of the MDS is essentially the minimization of Eq. (2) under the constraints in Eq. (1).

The MDS can adequately describe the collective behavior of an entire network by using only a minimum required subset of nodes. By studying these nodes, a researcher can infer on the topology of their neighboring ones. Nevertheless, in a correlation-based economics network, low correlation edges connect nodes with dissimilar behavior and should not participate in the identification of the MDS, since they may provide false inference and misleading results. For example, if an edge links two states and displays a correlation of $r = 0.2$, we should not consider them as adjacent (in the sense of behavior similarity), since they are, for all practical matters, uncorrelated and none of them can effectively represent the other. We overcome this inadequacy of the classic MDS optimization procedure in an economics network by imposing a threshold on the initial network's correlation values.

**Definition 3.** A *Threshold-Minimum Dominating Set* (*T-MDS*) is defined as a two-step methodology for identifying the most representative nodes in a network. These steps are defined as follows:

*Step* 1. Eliminate all edges where the correlation falls below the threshold correlation.
*Step* 2. Identify the MDS nodes on the remaining network.

The thresholding step may lead to the emergence of *isolated* nodes (i.e., nodes without any edges to connect them to the rest of the network), while Step 2 identifies the nodes that can efficiently represent the collective behavior of the interconnected network. These nodes are called *Dominant*. The T-MDS, by definition, must include every isolated node. Thus, the T-MDS typically equals the union of the isolated and the dominant node sets, T-MDS $= I \cup C$, where $I$ and $C$ are the sets of the isolated and the dominant nodes, respectively. We should not, however, consider these as a cohesive set: we must distinguish the subset of the isolated nodes from the dominant nodes' subset, since the two subsets exhibit entirely different and independent features. The states that correspond to isolated nodes exhibit highly idiosyncratic behavior and, thus, cannot represent (or be represented by) any other state.

### *3.2.2. Interpretation of the T-MDS*

The T-MDS can provide us with a manifold analysis of the US states' income inequality network. First, we can use it to infer on any convergence patterns of income inequality. In any arbitrary network, the cardinality of the T-MDS set can take values between two extremes. For complete networks, where every node connects to every other node, the T-MDS size equals 1 and each node can possibly define an MDS. For a completely disconnected network, where all nodes are isolated, the T-MDS size equals $n$ (the number of the nodes in the network). According to the above, a T-MDS cardinality close to 1 indicates a rather dense network and T-MDS cardinality close to $n$ indicates a sparse network with a lot of isolated nodes. A dense network, by definition, exhibits higher correlations between the nodes.

Second, the T-MDS identifies sub-sets of states that are defined by the neighborhood of a dominant state. These (unique or overlapping) neighborhoods are important for our analysis as they highlight states that exhibit within them strong correlations in terms of the evolution of income inequality in each sub-period. Fiscal and monetary authorities can use these neighborhoods to examine the causes of these inter-relations and possibly deal with inequality in a collective, systemic fashion. We must stress here that belonging to the same neighborhood and, thus, exhibiting strong correlations on an inequality measure does not mean that the states' income distribution is highly similar. Rather, it provides evidence in support of a highly similar evolution of inequality.[8]

Finally, the T-MDS methodology may identify certain isolated nodes. These nodes correspond to states with a completely idiosyncratic behavior with respect to the evolution of income inequality. By closely monitoring the isolated nodes in each focus period, we can draw inference on the integration process of these states in the network. For example, if a state is identified as isolated in period $t$ and in the next period it belongs to a neighborhood, then this is evidence in favor of increased income inequality evolution through time. On the other hand, if a neighborhood state becomes isolated across time, then this will indicate that this state resists the general inequality evolution patterns.

### *3.2.3. Dominant nodes and neighborhoods: discussion*

We need to clarify the semantics of the T-MDS methodology to better understand and interpret its results. The standard MDS methodology identifies the Dominating Set (i.e., the set of Dominant Nodes) through a global minimization algorithm.

---

8 For example if the evolution of state A's Gini coefficient is 0.1, 0.2, 0.3 and the respective coefficients for state B are 0.7, 0.8 and 0.9 the correlation is 1. The two states may significantly differ in terms of inequality but they have the same evolution.

**Table 1**
T-MDS metrics for the top 1% income share.

| | 1916–2012 | 1916–1929 | 1930–1944 | 1945–1979 | 1980–2012 |
|---|---|---|---|---|---|
| T-MDS cardinality | 10 | 22 | 28 | 15 | 3 |
| Isolated states | 4 | 14 | 22 | 10 | 0 |
| Dominant states | 6 | 8 | 6 | 5 | 3 |

The Dominating Set is identified such that every node that is not a member of the Dominating Set is adjacent to at least one Dominating Node; a property termed Dominance in Graph Theory.

Nonetheless, a serious shortcoming of the standard MDS approach exists when using it in networks based on similarity, such as our analysis. That is, in our case, the edges do not simply connect two nodes, but they carry a weight equal to the correlation of the two nodes in terms of the underlying variable used. The problem with the standard MDS approach is that its algorithmic process, as described above, does not take into account the weight associated to the edges. That is, the method does not consider the magnitude of similarity between the nodes, but only the number of interconnections (node degree). To overcome this inherent algorithmic problem of the standard MDS method, we use the thresholding step (T-MDS), eliminating all low similarity (unimportant) edges. Thus, after thresholding, only the high similarity (important) edges survive. On the other hand, nodes with highly idiosyncratic behavior (low correlation to all other nodes) become isolated nodes. Employing the T-MDS methodology, when we apply the MDS algorithm to identify the set of Dominant Nodes after the thresholding step, the identified Dominant Nodes connect only to highly similar (as it is measured by its correlation) nodes.

Thus, using the T-MDS method, we first eliminate all low correlation (unimportant) edges using the thresholding step and we get a reduced (in terms of the number of edges) network. Next, we identify the Dominant Nodes in the reduced network, using the global minimization algorithm. Finally, from the identified Dominant Nodes, we get their corresponding neighborhoods (i.e., a Dominant Node and all nodes connected to it define a neighborhood). A neighborhood's Dominant Node is the most interconnected node within that neighborhood. For this reason, it can represent the corresponding neighborhood. There is no causality implied between the Dominant Node and its neighbors in any direction. All nodes within a neighborhood exhibit similar behavior in terms of the underlying variable used to calculate the correlations. In that respect, any node within a neighborhood can represent it, but the best one in terms of node degree (the number of interconnections) is its Dominant Node.

## 4. Empirical results

We perform the aforementioned analysis and report the respective empirical results on both the Top 1% and the Gini coefficient measures of inequality for the case of a threshold $p = 0.90$.[9] In what follows, we examine two distinct issues with respect to inequality: the degree of inequality synchronization between the 48 US states and the evolution of convergence in inequality. Synchronization measures whether inequality in the different states moves in the same direction over time; either toward lower or greater inequality. Convergence measures whether the states move closer together over time in the degree of income inequality; either toward lower or greater inequality. Thus, a high degree of synchronization does not indicate convergence. Two perfectly synchronized states ($r = 1$) will never converge. Finally, in (Tables 4 and 5), we report in detail the dominant and isolated nodes in each sub-period, in terms of each income inequality measure. A thorough examination of the isolated nodes (that correspond to practically uncorrelated US states) and the analysis of the reasons for their appearance inter-temporally may provide policy makers with valuable information in order to successfully address the causes of income inequality within the US. We perform this analysis for both measures of inequality: the Top 1% income share and the Gini coefficient.

### *4.1. The top 1%*

In Table 1, we report the empirical results from the Top 1% inequality measure, and in Fig. 5 we plot the evolution of this measure for the dominant states over the four sub-periods. The main property of the dominant states provided by the T-MDS methodology is that they exhibit a highly similar behavior with the rest of their respective neighborhood (shown in Table 6). The domination property in Graph Theory does not imply causation or leading behavior. Thus, these states are dominant in the sense that they exhibit a high degree of similarity in the evolution of inequality with their direct neighborhood. But, the dominant state does not necessarily represent a leader within the neighborhood. That is, the neighboring states do not necessarily follow the dominant state in a causal sense.

In the first period before the Great Depression, the number of dominant states reaches its maximum of eight. This signifies the existence of several different group patterns of inequality evolution. Moreover, the number of isolated states (14) is

[9] We perform the analysis for three alternative threshold levels $p = 0.85$, $p = 0.90$ and 0.95 which all seem to yield qualitatively similar results. We do not report the $p = 0.85$ and 0.95 results for the sake of brevity. These results are available on request.

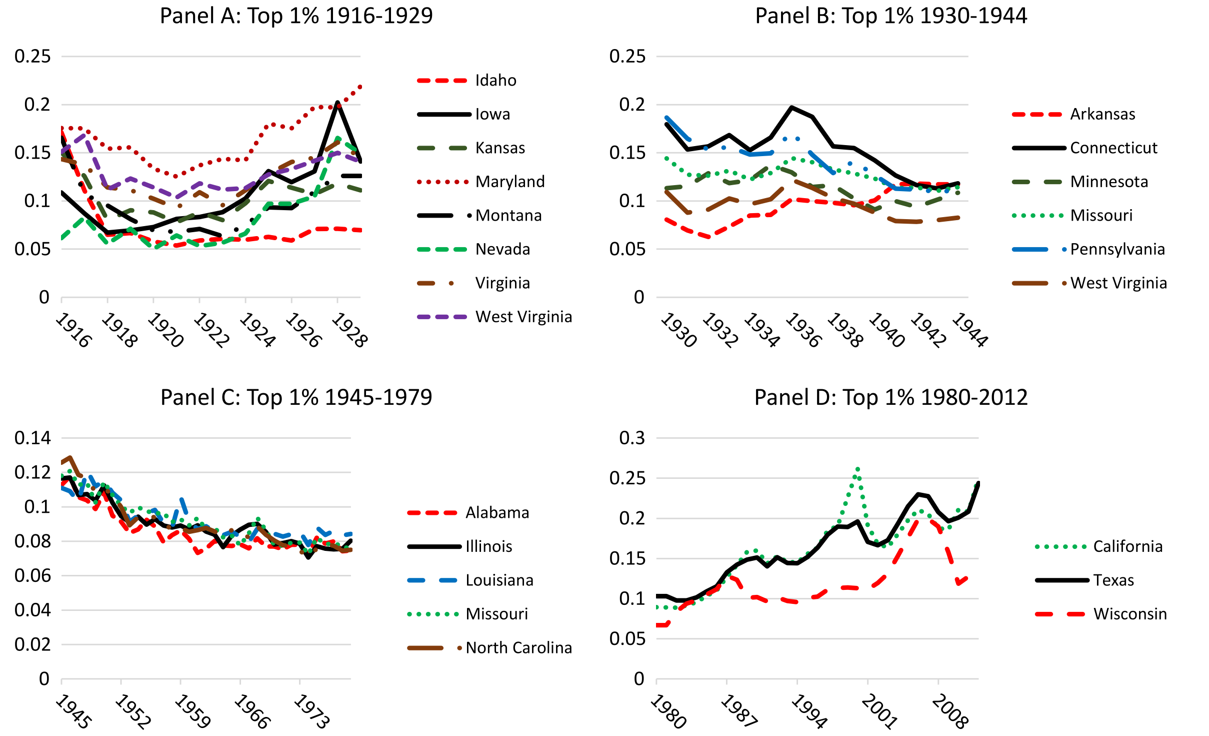


**Fig. 5.** Dominant states by sub-period: Top 1% income share.

the second highest of the four periods. Thus, the T-MDS cardinality is high, providing evidence of low synchronization in inequality. Fig. 5, Panel A, exhibits the evolution of the inequality measure in the WWI+ period for the eight dominant states. In general, we observe a U-shaped pattern for each state and the eight states maintain their distances throughout this period. The visual inspection of Fig. 5, panel A, reveals no significant convergence in inequality. The same is true when we look at the standard deviation for the total of the 48 states in our sample in Fig. 2.[10]

During the Great Depression, the number of dominant states falls to six, but the isolated states rise significantly to 22. As a result, the T-MDS reaches its maximum cardinality during this period indicating a lower degree of synchronization with respect to the previous period. Inequality, in general, as we discussed earlier, falls during this period. In Fig. 5, Panel B, the inequality measures (top 1% share) of the six dominant states appear to follow distinct to each other paths for the first half of the period. They show some convergence after 1936.

In the third period of the Great Compression, five dominant states emerge and the isolated ones fall significantly to 10. Thus, the cardinality of the T-MDS falls almost to half (from 28 to 15), indicating an increased synchronization in the evolution of inequality. From Fig. 5, Panel C, we observe there are some clear indications of convergence in inequality: all five dominant states move close together throughout this period toward lower inequality.

Finally, in the last period of the Great Divergence, we see that the dominant states fall to only three with no isolated states whatsoever. Thus, the T-MDS cardinality reduces from 15 to 3. This results provides strong evidence in support of a very high degree of synchronization of inequality within the 48 US states and synchronization reaches a maximum in this period. This high synchronization reflects the general evolution toward higher inequality in the period of the Great Divergence. In Fig. 5, Panel D, we observe that the three dominant states California, Texas, and Wisconsin (and, respectively, their neighborhoods) converge closely until 1987. Then, the California and Texas neighborhoods continue to converge throughout this period to rising patterns of inequality. On the other hand, the Wisconsin neighborhood diverges significantly from 1987 to 2001. From 2002 to 2007, it reverts toward the other two dominant states but after 2007 the Wisconsin neighborhood diverges again significantly with a distinct trend toward less inequality (i.e. the Top 1% share in total income falls to approximately half of that in the California and Texas neighborhoods).

In the last step of our analysis, we compare the T-MDS findings from each focal sub-period with the ones from the full sample. To do this we first pool the full period 1916–2012 and apply the T-MDS methodology (results shown on the first column of Table 1). Then, we examine the resemblance between these findings by calculating the degree of overlap between the neighborhoods of the full sample and each of the sub-periods. These findings are contained in Table 8 which

10 Here we provide a discussion on the properties of the dominant nodes including some inference in convergence. The whole picture though is captured in Figs. 2 and 4 for the Top 1% and the Gini coefficient measures of inequality. In these figures, we present the standard deviation of the total 48 states of our sample.

**Table 2**
T-MDS metrics for the Gini coefficient.

| | 1916–2012 | 1916–1929 | 1930–1944 | 1945–1979 | 1980–2012 |
|---|---|---|---|---|---|
| T-MDS cardinality | 4 | 13 | 25 | 27 | 5 |
| Isolated states | 1 | 8 | 22 | 21 | 2 |
| Dominant states | 3 | 5 | 3 | 6 | 3 |

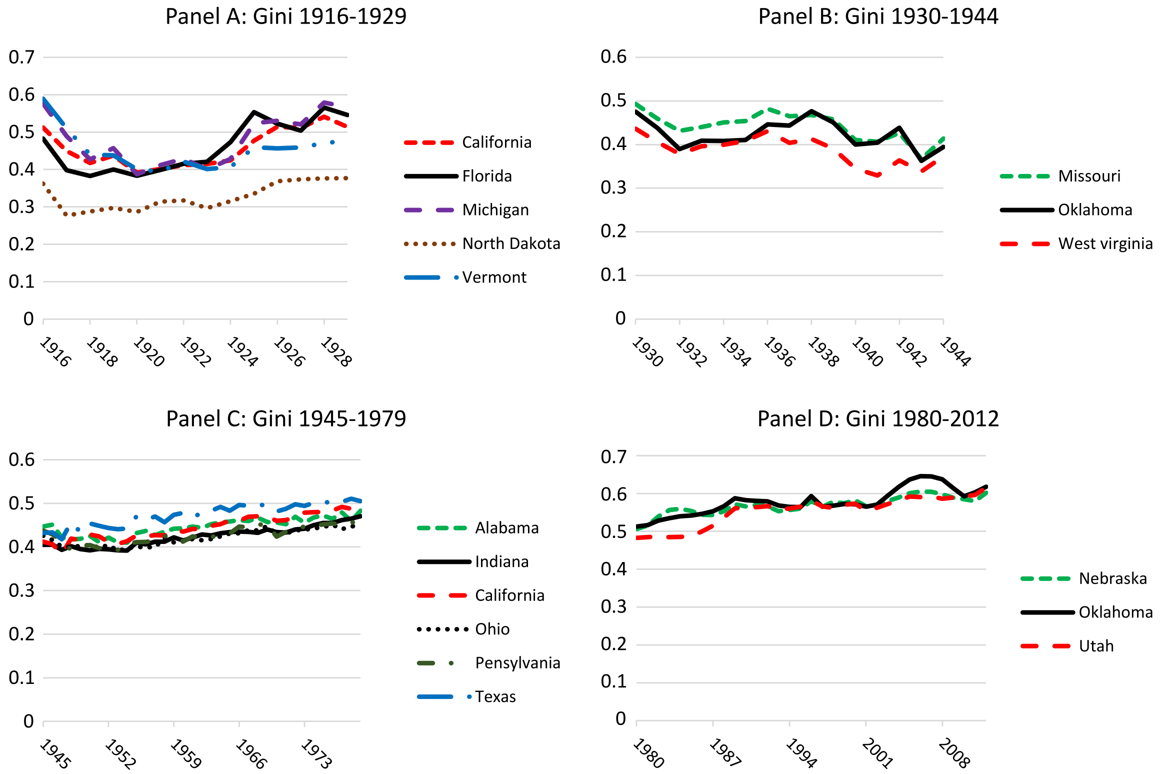


**Fig. 6.** Dominant states by sub-period: Gini coefficient.

shows the overlap between each dominant node's neighborhood in the sub-periods (columns) and the respective full sample neighborhoods (rows). In this sense a 6 × 22 table is created, corresponding to the six identified neighborhoods of the full sample and the 22 neighborhoods of the sub-period analysis. In the first period we observe that there the overlap ranges between 2% and 34%. In the second period the overlap ranges from 2% to 29%. In the third and fourth sub-periods the maximum overlap between the identified neighborhoods reaches a maximum of 41%.

### *4.2. Gini coefficient*

Table 2 reports the T-MDS results and Fig. 6 plots the Gini coefficient inequality measure over the four sub-periods for the dominant states. Once again, each dominant state captures the behavior of its direct neighbors (see Table 7). Thus, by studying only the dominant states, we can gain insight on the collective behavior of the entire network of 48 US states.

In the WWI+ period, the T-MDS methodology identifies a set of five dominant states and a set of eight isolated states. This indicates that about one in six US states presents a highly atypical behavior during the WWI+ period, while there appear to be five neighborhoods. In Fig. 6, Panel A, we plot the Gini coefficients of these five dominant states. We get the same U-shaped pattern found for the Top 1% measure of inequality. The neighborhood represented by North Dakota exhibits a significantly lower degree of inequality across this whole period. The other four dominant states seem to converge in the first half of the period but diverge again in the second.

During the Great Depression, we observe that the number of isolated states increases sharply to 22 (as it was the case with the Top 1% measure). We identify Missouri, Oklahoma, and West Virginia as the dominant states and the remaining 26 states belong to their respective neighborhoods. The high number of isolated states provides strong evidence of a low degree of inequality synchronization during the Great Depression. From Fig. 6, Panel B, we can observe that inequality for the dominant states and their neighborhoods displays a slight downward trend during this period and all three dominant states converge toward 0.4 in 1944.

**Table 3**
Inequality evolution across the 48 US states.

| Period | Synchronization | | Convergence | |
|---|---|---|---|---|
| | Top 1% | Gini | Top 1% | Gini |
| 1916–1929 | Down | Down | – | – |
| 1930–1944 | Down | Down | Increasing | Increasing |
| 1945–1979 | Up | Down | Increasing | Increasing |
| 1980–2012 | Up | Up | Decreasing | Decreasing |

**Table 4**
Dominant and isolated nodes in each sub-period: Top 1% income share.

| Period | Status | State |
|---|---|---|
| 1916–2012 | Dominant | Alabama, California, Louisiana, Maryland, South Dakota, West Virginia |
| | Isolated | Delaware, Maine, Nevada, Oklahoma |
| 1916–1929 | Dominant | Idaho, Iowa, Kansas, Maryland, Montana, Nevada, Virginia, West Virginia |
| | Isolated | Alabama, Arkansas, Georgia, Kentucky, Louisiana, Maine, Mississippi, Nebraska, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Wyoming |
| 1930–1944 | Dominant | Arkansas, Connecticut, Minnesota, Missouri, Pennsylvania, West Virginia |
| | Isolated | Alabama, Arizona, Delaware, Florida, Georgia, Idaho, Kansas, Louisiana, Maine, Montana, Nebraska, New Mexico, North Carolina, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Vermont, Washington, Wyoming |
| 1945–1979 | Dominant | Alabama, Illinois, Louisiana, Missouri, North Carolina |
| | Isolated | Delaware, Idaho, Montana, Nevada, North Dakota, Oklahoma, South Dakota, Vermont, West Virginia, Wyoming |
| 1980–2012 | Dominant | California, Texas, Wisconsin |
| | Isolated | |

For the Great Compression, the number of isolated states remains at a high level (i.e. 21 rather than 22). Additionally, the number of dominant states increases from three to six and the T-MDS reaches a cardinality of 27, the highest across all four periods. Now, the 48 US states display an even lower degree of synchronization, a result that contrasts with the findings for the Top1% in the same period. In Fig. 6, Panel C, we can see that, in general, the six neighborhoods of states do not show any distinguishable pattern of convergence or divergence.

Finally, in the last period of the Great Divergence, the isolated states fall sharply to only two. Moreover, we also find three dominant states and, consequently, the T-MDS cardinality falls from 27 to only five. This provides strong evidence in support of a high degree of inequality synchronization. From Fig. 6, Panel D, we can observe that the three dominant states do not show a stable convergence pattern through-out this period: they seem to diverge from 1980 to 1988, converge closely from 1989 to 2003 and diverge again after that.

As in the case of the Top 1% income share measure, we also examine the relation between the pooled full sample results versus the ones of the sub-periods, in the case of the Gini coefficient. The results of the Gini T-MDS metrics are contained in the first column of Table 2 while the overlapping between the dominant node neighborhoods in the sub-periods and the ones in full sample analysis, are included in Table 9. This Table has a size of $3 \times 17$, corresponding to the three identified neighborhoods of the full sample and the total of 17 neighborhoods in the sub-period analysis. In the first focal sub-period, we observe that the overlap ranges widely from 3% to 60%. In the second sub-period the overlap ranges between 3% and 40%. In the third period the smaller observed overlap is 6% while the higher overlap drops to 34%. In the last sub-period, a generally increased overlap of the identified neighborhoods is observed, starting from 9% and reaching a maximum of 71%.

Table 3 summarizes the results from both inequality measures. The synchronization results are from the T-MDS metrics and the convergence results come from the standard deviation of the Top 1% and the Gini coefficient for the four periods. For the first two periods, the qualitative results on synchronization and convergence are the same for the two inequality measures. In the period of the Great Compression, the social, economic, and political changes as we discussed earlier affected the Top 1% in a different way than the rest of the income classes. The synchronization of the states increases as there is a common trend that lowers the share of the Top 1%. The degree of synchronization for the Gini coefficient is lower this period indicating a move toward lower inequality. We essentially see a redistribution of income from the Top 1% to the middle-upper class.

## 5. Discussion

In a related set of papers, Lin and Huang [22–24] employ a series of unit-root tests to consider the convergence of income inequality measures for the 48 contiguous states using the Frank [25] annual data from 1916 to 2005.[11] Lin and Huang [24]

[11] As Lin and Huang [24] note, convergence does not necessarily mean convergence to a lower level of inequality. That is, convergence could occur around a rising level of income inequality.

**Table 5**
Dominant and isolated nodes in each sub-period: Gini coefficient.

| Period | Status | States |
|---|---|---|
| 1916–2012 | Dominant<br>Isolated | Pennsylvania, Texas, Utah<br>Delaware |
| 1916–1929 | Dominant<br>Isolated | California, Florida, Michigan, North Dakota, Vermont<br>Arkansas, Louisiana, Mississippi, Oklahoma, South Carolina, South Dakota, Washington, Wyoming |
| 1930–1944 | Dominant<br>Isolated | Missouri, Oklahoma, West Virginia<br>Alabama, Arizona, Arkansas, Delaware, Florida, Georgia, Idaho, Iowa, Kansas, Louisiana, Maine, Mississippi, Montana, Nebraska, New Mexico, North Dakota, Oregon, South Carolina, South Dakota, Tennessee, Vermont, Wyoming |
| 1945–1979 | Dominant<br>Isolated | Alabama, California, Indiana, Ohio, Pennsylvania, Texas<br>Arkansas, Colorado, Delaware, Iowa, Kansas, Kentucky, Maine, Mississippi, Missouri, Nebraska, New Hampshire, New Mexico, North Carolina, North Dakota, Oklahoma, Rhode Island, South Dakota, Tennessee, Vermont, West Virginia, Wyoming |
| 1980–2012 | Dominant<br>Isolated | Nebraska, Oklahoma, Utah<br>North Dakota, South Dakota |

ultimately use the panel unit-root test of Carrion-i-Silvestre et al. [26], which extends the Hadri [27] panel unit-root test to include an unknown number of structural breaks and cross-sectional dependence. The more conventional panel unit-root tests that they implement indicate that the inequality measures do not converge. The Carrion-i-Silvestre et al. [26] test, however, indicates convergence of the income inequality measures.[12]

While we do not test for $\beta$-convergence in this paper, our Figs. 2 and 4 do provide information on $\sigma$-convergence. For both the Top 1% and the Gini coefficient series, we observe $\sigma$-convergence from 1916 to 1980 and then $\sigma$-divergence from 1980 to 2012. That is, the convergence findings depend on the sample period examined. The different findings on convergence in [22–24] may reflect the use of the entire sample and not considering the possibility of different convergence results for the different subsamples identified in the US inequality literature WWI+ period, Great Depression, Great Compression, and Great Divergence.

An interesting finding of our analysis and more specifically the use of two alternative measures of inequality is the different results obtained for the third period under consideration, i.e. the 1945–1979 period. We observe that, in this period, the Gini coefficient reveals de-phasing between the US states while the Top 1% income share indicates increased synchronization. That is probably the direct effect of structural changes to the top of the income distribution and the concentration of wealth to higher levels of the society. As Krozer [28] notes, the Gini coefficient overemphasizes the changes in the middle of the income distribution, disregarding changes made to the top.

Two more studies that engage in income inequality measures' comparisons are Leigh [29] and Alvaredo [30]. In these papers, the authors find that the Gini coefficient is linearly related to the Top 1% measure in a statistically significant manner. This, of course, does not mean that inequality measures should always provide analogous results. In our study, we found that for most of the sample, the two measures provided similar results. Thus, overall, our findings are in line with the relevant literature on the use of income inequality measures.

## 6. Conclusion

In this paper, we examine the evolution of income inequality in 48 US states, using complex network analysis. We employ a new optimization technique, never used before within this context, called the Threshold-Minimum Dominating Set (T-MDS). We also use a long data sample that spans almost a century: the period from 1916 to 2012. Moreover, we perform a dynamic analysis and break our original sample into four consecutive sub-periods that correspond to "Great" episodes: the WWI+ period (1916–1928), the Great Depression (1929–1944), the Great Compression (1945–1979) and the Great Divergence (1980–2012).

We examine income inequality by employing two alternative measures: the Top 1% share of income and the Gini coefficient. These provide us with alternative perspectives on inequality. The Top 1% focuses on the fraction of total income held by the Top 1% of the income distribution. It does not include any information on the distribution of income amongst the remaining 99%. The Gini coefficient, on the other hand, includes information on the entire distribution of income. The inequality measures that we use, as noted above, come from IRS data, which have the problem of truncation of individuals at the low-end of the income distribution. This suggests that the Gini may incorporate more bias than the Top 1%. Thus, the analysis of both measures can offer an interesting alternative insight with respect to income inequality.

The Great Divergence, a much discussed issue, considers why inequality has increased since the ending of the Great Compression. In their analysis of this issue, Gordon and Dew-Becker [31] divide the income distribution into the top 10% and

12 Lin and Huang [24] report, however, that they can reject the null hypothesis of stationarity for 22 and 17 out of the 48 states for the Top 10% and Top 1% series, respectively, on an individual state-by-state basis.

**Table 6**
Dominant state neighborhoods: Top 1%.

| Period | Dominant state | Neighborhood |
|---|---|---|
| 1916–2012 | Alabama (AL) | Arizona, Georgia, Iowa, Kansas, Kentucky, Louisiana, Mississippi, Montana, Nebraska, Oregon, South Carolina, Tennessee, Utah |
| | California (CA) | Colorado, Connecticut, Florida, Illinois, Massachusetts, Minnesota, New Hampshire, New Jersey, Texas, Virginia, Washington |
| | Louisiana (LA) | Alabama, Arizona, Arkansas, Georgia, Kansas, Kentucky, Mississippi, Montana, Nebraska, New Mexico, Oregon, South Carolina, Tennessee, Texas, Utah |
| | Maryland (MD) | Illinois, Indiana, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, Virginia, Wisconsin |
| | South Dakota (SD) | Arkansas, Idaho, Kansas, Nebraska, North Dakota, Wyoming |
| | West Virginia (WV) | Indiana, Kentucky, Minnesota, Missouri, Ohio, Vermont, Wisconsin |
| 1916–1929 | Idaho (ID) | Vermont |
| | Iowa (IA) | Florida, Nevada |
| | Kansas (KS) | Colorado, North Dakota |
| | Maryland (MD) | Connecticut, Delaware, Illinois, Massachusetts, Michigan, Missouri, New Jersey, New York, Pennsylvania, Virginia |
| | Montana (MT) | New York, North Carolina, Ohio, Texas, Washington |
| | Nevada (NV) | Iowa, Michigan, Tennessee |
| | Virginia (VA) | Arizona, California, Connecticut, Illinois, Indiana, Maryland, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Utah, Wisconsin |
| | West Virginia (WV) | Colorado, New Mexico, Ohio, Utah |
| 1930–1944 | Arkansas (AK) | Mississippi, Oregon |
| | Connecticut (CT) | Indiana, Minnesota, Missouri, New Hampshire, New Jersey, Ohio, Rhode Island, Utah, West Virginia, Wisconsin |
| | Minnesota (MN) | California, Connecticut, Illinois, Indiana, Kentucky, Massachusetts, Missouri, New Hampshire, New Jersey, New York, Ohio, Rhode Island, West Virginia, Wisconsin |
| | Missouri (MS) | Connecticut, Indiana, Minnesota, Nevada, New Hampshire, Ohio, West Virginia, Wisconsin |
| | Pennsylvania (PA) | Illinois, Indiana, Iowa, Maryland, Massachusetts, Michigan, New Jersey, New York, Ohio, Rhode Island, Wisconsin |
| | West Virginia (WV) | Colorado, Connecticut, Minnesota, Missouri, New Hampshire, Virginia |
| 1945–1979 | Alabama (AL) | Arkansas, California, Colorado, Florida, Georgia, Illinois, Indiana, Kansas, Kentucky, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, New Hampshire, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Utah, Virginia, Washington, Wisconsin |
| | Illinois (IL) | Alabama, California, Colorado, Connecticut, Florida, Georgia, Indiana, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, Tennessee, Virginia, Wisconsin |
| | Louisiana (LA) | Arizona, California, Colorado, Kansas, New Mexico, Ohio, Oregon, Texas, Washington |
| | Missouri (MS) | Alabama, Arkansas, Colorado, Florida, Georgia, Illinois, Indiana, Iowa, Kentucky, Maryland, Massachusetts, Michigan, Minnesota, Nebraska, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, Tennessee, Virginia, Wisconsin |
| | North Carolina (NC) | Alabama, Arkansas, California, Colorado, Florida, Georgia, Illinois, Indiana, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, New Hampshire, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Virginia, Wisconsin |
| 1980–2012 | California (CA) | Colorado, Connecticut, Florida, Illinois, Maryland, Massachusetts, Nevada, New Hampshire, New Jersey, New York, Texas, Virginia, Washington |
| | Texas (TX) | Arizona, California, Colorado, Florida, Georgia, Illinois, Kansas, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, New York, Oklahoma, Pennsylvania, Rhode Island, South Dakota, Tennessee, Utah, Virginia, Washington, Wisconsin, Wyoming |
| | Wisconsin (WI) | Alabama, Arizona, Arkansas, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Hampshire, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, West Virginia |

the bottom 90%, arguing that different factors explain the movements of these two components of the income distribution. The movements in the bottom 90% largely reflect a reversal of the factors that contributed to the Great Compression, according to Golden and Margo [6]. That is, union coverage ratios declined dramatically, the import share of GDP rose significantly, and immigration rose consistently over the Great Divergence. Each of these factors helps to explain the divergence of incomes within the income distribution for the bottom 90%.

Gordon and Dew-Becker [31] also consider alternative mechanisms that can assist in explaining the divergence within the bottom 90%—the real minimum wage, lower top bracket tax rates, and skill-based technical change (SBTC). For SBTC, Gordon and Dew-Becker [31] describe the modeling of Autor, Katz, and Kearney [32] and Autor, Murname, and Levy [33]. These authors consider three tiers within the labor force—a top tier of employees doing non-routine, cognitive work; a middle tier of workers doing routine, repetitive work; and a low tier of workers doing manual, but interactive, work. The top tier includes lawyers, investment bankers, CEOs, and so on. The second tier includes bookkeepers, accountants, and

**Table 7**
Dominant state neighborhoods: Gini coefficient.

| Period | Dominant state | Neighborhood |
|---|---|---|
| 1916–2012 | Pennsylvania (PA) | California, Connecticut, Illinois, Maine, Maryland, Massachusetts, Michigan, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Rhode Island, Wisconsin |
| | Texas (TX) | Alabama, Arizona, Arkansas, California, Colorado, Georgia, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming, |
| | Utah (UT) | Alabama, Arizona, Arkansas, California, Colorado, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, Oklahoma, Oregon, South Carolina, Tennessee, Texas, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming |
| 1916–1929 | California (CA) | Alabama, Colorado, Connecticut, Georgia, Illinois, Indiana, Kentucky, Maine, Maryland, Michigan, Minnesota, Missouri, Montana, Nevada, New Hampshire, New Jersey, North Carolina, Ohio, Tennessee, Utah, Virginia, Wisconsin |
| | Florida (FL) | Alabama, Georgia, Indiana, Iowa, Nevada |
| | Michigan (MI) | Arizona, California, Colorado, Connecticut, Delaware, Georgia, Illinois, Indiana, Kansas, Maine, Maryland, Massachusetts, Minnesota, Missouri, Montana, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, Ohio, Oregon, Pennsylvania, Texas, Utah, Virginia, West Virginia, Wisconsin |
| | North Dakota (ND) | Indiana, Nebraska, Nevada |
| | Vermont (VT) | Idaho, Rhode Island |
| 1930–1944 | Missouri (MS) | California, Colorado, Connecticut, Illinois, Indiana, Kentucky, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New York, North Carolina, Ohio, Rhode Island, Utah, Washington, West Virginia |
| | Oklahoma (OK) | Texas |
| | West Virginia (WV) | Connecticut, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Utah, Virginia, Wisconsin |
| 1945–1979 | Alabama (AL) | Georgia, Illinois, Indiana, Massachusetts, Minnesota, New Jersey, Ohio, Pennsylvania, South Carolina, Utah, Virginia, Wisconsin |
| | California (CA) | Arizona, Illinois, Indiana, Louisiana, Massachusetts, Michigan, Nevada, New Jersey, Oregon, Texas, Washington, Wisconsin |
| | Indiana (IN) | Alabama, California, Illinois, Louisiana, Massachusetts, Michigan, Minnesota, Montana, New Jersey, Ohio, Oregon, Texas, Washington, Wisconsin |
| | Ohio (OH) | Alabama, Connecticut, Georgia, Illinois, Indiana, Maryland, Massachusetts, Michigan, Minnesota, New Jersey, Pennsylvania, Utah, Wisconsin |
| | Pennsylvania (PA) | Alabama, Florida, Georgia, Illinois, Maryland, Massachusetts, Michigan, Minnesota, New Jersey, New York, Ohio, South Carolina, Wisconsin |
| | Texas (TX) | Arizona, California, Idaho, Indiana, Louisiana, Oregon, Washington |
| 1980–2012 | Nebraska (NE) | Alabama, Idaho, Iowa, Kansas, Kentucky, Missouri, Montana, Ohio, Oklahoma, South Carolina, Tennessee, Vermont |
| | Oklahoma (OK) | Alabama, Arkansas, Florida, Georgia, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maine, Michigan, Mississippi, Missouri, Montana, Nebraska, New Mexico, North Carolina, Ohio, Oregon, South Carolina, Tennessee, Texas, Vermont, West Virginia, Wisconsin |
| | Utah (UT) | Alabama, Arizona, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Kansas, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nevada, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, Washington, Wisconsin |

**Table 8**
Overlap of state neighborhoods in each sub-period versus the pooled sample (Top 1% income share).

| Dominant | 1916–1929 | | | | | | | | 1930–1944 | | | | | | 1945–1979 | | | | | 1980–2012 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ID | IA | KS | MD | MT | NV | VA | WV | AK | CT | MN | MS | PA | WV | AL | IL | LA | MS | NC | CA | TX | WI |
| AL | 0.02 | 0.05 | 0.05 | 0.02 | 0.05 | 0.07 | 0.07 | 0.05 | 0.07 | 0.05 | 0.05 | 0.02 | 0.05 | 0.02 | 0.24 | 0.15 | 0.12 | 0.20 | 0.20 | 0.02 | 0.20 | 0.37 |
| CA | 0.02 | 0.05 | 0.05 | 0.15 | 0.07 | 0.02 | 0.20 | 0.05 | 0.02 | 0.12 | 0.20 | 0.10 | 0.10 | 0.15 | 0.24 | 0.27 | 0.12 | 0.17 | 0.22 | 0.29 | 0.29 | 0.20 |
| LA | 0.02 | 0.02 | 0.05 | 0.02 | 0.07 | 0.05 | 0.07 | 0.07 | 0.10 | 0.05 | 0.05 | 0.02 | 0.02 | 0.02 | 0.27 | 0.15 | 0.17 | 0.20 | 0.22 | 0.05 | 0.22 | 0.41 |
| MD | 0.02 | 0.02 | 0.02 | 0.24 | 0.10 | 0.05 | 0.34 | 0.05 | 0.02 | 0.22 | 0.29 | 0.17 | 0.29 | 0.12 | 0.34 | 0.41 | 0.05 | 0.37 | 0.39 | 0.20 | 0.34 | 0.34 |
| SD | 0.05 | 0.02 | 0.07 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.05 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.07 | 0.02 | 0.05 | 0.07 | 0.05 | 0.02 | 0.12 | 0.17 |
| WV | 0.05 | 0.02 | 0.02 | 0.05 | 0.05 | 0.02 | 0.15 | 0.07 | 0.02 | 0.17 | 0.20 | 0.17 | 0.10 | 0.10 | 0.17 | 0.15 | 0.05 | 0.17 | 0.17 | 0.02 | 0.10 | 0.22 |

Note: The left-hand column reports the dominant states for the full sample analysis. The second row lists the dominant states in each cub-period. The numbers are the fractional overlap between the two neighborhoods For example, the overlap equals 22% for the neighborhood of West Virginia in the full sample findings and the neighborhood of Wisconsin in the 1980–2012 sample analysis.

so on. Finally, the third tier includes nurses, waiters, and so on. Increased demand and decreased relative supply of SBTC workers put upward pressure on incomes in the top-tier group.

**Table 9**
Overlap of state neighborhoods in each sub-period versus the pooled sample (Gini coefficient).

| Dominant | 1916–1929 | | | | | 1930–1944 | | | 1945–1979 | | | | | | 1980–2012 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | CA | FL | MI | ND | VT | MS | OK | WV | AL | CA | IN | OH | PA | TX | NE | OK | UT |
| PA | 0.37 | 0.03 | 0.46 | 0.03 | 0.06 | 0.34 | 0.03 | 0.37 | 0.20 | 0.20 | 0.23 | 0.29 | 0.29 | 0.06 | 0.09 | 0.20 | 0.46 |
| TX | 0.43 | 0.17 | 0.51 | 0.14 | 0.09 | 0.31 | 0.09 | 0.26 | 0.26 | 0.29 | 0.31 | 0.20 | 0.17 | 0.26 | 0.37 | 0.63 | 0.57 |
| UT | 0.51 | 0.17 | 0.60 | 0.09 | 0.06 | 0.40 | 0.06 | 0.29 | 0.29 | 0.31 | 0.34 | 0.23 | 0.23 | 0.23 | 0.34 | 0.63 | 0.71 |

Note: See Table 8.

For the top 10%, Gordon and Dew-Becker [31] identify superstars, certain high paid professions (e.g., corporate lawyers, investment bankers, hedge fund managers, and so on), and corporate CEOs as receiving dramatic relative increases in compensation, leading to a higher percentage of total income accruing to the top 10% (and the top 1%).

Our findings reveal different patterns of income inequality evolution according to each focal period. For the first two periods, namely, the WWI+ period and the Great Depression, using both measures of inequality we find evidence in support of a lower degree of inequality among the 48 US states. In the third period, the Great Compression inequality is lower according to the Top 1% measure: from 0.11 to 0.08. Nonetheless, when the Gini coefficient is used in the analysis, inequality increases during this period from 0.40 to 0.48. Although these results may seem contradictory at first, we believe that they are not. The use of alternative measures of inequality allows us to detect and identify different patterns of change. During the Great Compression, significant changes in the political, social, and economic norms did occur: progressive taxation and social resistance on excessive increases in executive pay [8], the increasing power of the unions [7], and reduced immigrant inflows. These factors exerted a significant negative effect on the income of the Top 1% while they increased inequality in the lower income classes in favor of unionized labor and lower-to-medium management positions. The latter benefited from the structural changes at the expense of the Top 1% and the non-unionized and unskilled labor. This is the tangible manifestation of the ''American Dream'' and the emergence of the ''middle class'', the opportunity for upward social and economic mobility, prosperity, and success through hard work. In the last period of our study, the Great Divergence, inequality rises significantly according to both measures, reflecting the consensus in the relevant literature [8,7,1].

Finally, by employing the T-MDS methodology, we were able to highlight groups of similarly behaving states called ''neighborhoods''. Moreover, we identified the dominant and isolated states for each period and measure of inequality. The policy implications from the identification of these features of the network are obvious: (a) the policy maker either on the state or federal level can analyze the similarities within the neighborhoods as a basis for the implementation of a successful policy that aims to reduce income inequality and (b) the identification of the reason(s) that some states appear isolated is important for the policymaker. Thus, a careful examination and analysis of the specific characteristics of these states may provide significant information on the causes of inequality within the US states and the most appropriate means to implement an efficient policy mix to address it.

## References

[1] T. Piketty, Capital in the Twenty-First Century, Harvard University Press, Cambridge: MA, 2014.
[2] E. Saez, Striking it richer: the evolution of top incomes in the United States. manuscript, University of California, 2013. Berkeley: http://eml.berkeley.edu/~saez/saez-UStopincomes-2012.pdf.
[3] S. Rose, The false claim that inequality rose during the Great Recession. The Information Technology & Innovation Foundation, 2015. Report: http://www2.itif.org/2015-inequality-rose.pdf.
[4] Congressional Budget Office 2014. The distribution of household income and federal taxes, 2011. Report, https://www.cbo.gov/publication/49440.
[5] M.W. Frank, A new state-level panel of annual inequality measures over the period 1916–2005, J. Bus. Strateg. 31 (2014) 241–263.
[6] C. Goldin, R. Margo, The Great Compression: The wage structure in the United States at mid-century, Quart. J. Econ. 107 (1992) 1–34.
[7] P. Krugman, The Conscience of a Liberal, W.W. Norton & Company, New York, 2007, pp. 124–128.
[8] T. Piketty, E. Saez, Income Inequality in the United States, 1913–1998, Quart. J. Econ. 118 (2003) 1–39.
[9] S.N. Kaplan, J. Rauh, It's the market: The broad-based rise in the return on top talent, J. Econ. Perspect. 27 (2013) 35–56.
[10] R.J. Hill, Comparing price levels across countries using minimum-spanning trees, Rev. Econ. Stat. 81 (1999) 135–142.
[11] F. Allen, D. Gale, Financial contagion, J. Polit. Econ. 108 (2000) 1–33.
[12] D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, M.I. Loffredo, Interplay between topology and dynamics in the World Trade Web, Eur. Phys. J. B 57 (2007) 159–164.
[13] D.O. Cajueiro, B.M. Tabak, The role of banks in the Brazilian Interbank Market: Does bank type matter? Physica A 387 (2008) 6825–6836.
[14] S. Schiavo, J. Reyes, G. Fagiolo, International trade and financial integration: a weighted network analysis, Quant. Finance 10 (2010) 389–399.
[15] D. Acemoglu, V.M. Carvalho, A. Ozdaglar, A. Tahbaz-Salehi, The network origins of aggregate fluctuations, Econometrica 80 (2012) 1977–2016.
[16] C. Minoiu, A.J. Reyes, A network analysis of global banking: 1978–2010, J. Financ. Stab. 9 (2013) 168–184.
[17] T. Papadimitriou, P. Gogas, G.A. Sarantitis, Convergence of European Business Cycles: A Complex Networks Approach, Comput. Econ. (2014) 1–23.
[18] M.B. Hu, R. Jiang, Y.H. Wu, R. Wang, Q.S. Wu, Properties of wealth distribution in multi-agent systems of a complex network, Physica A 387 (2008) 5862–5867.
[19] T. Di Matteo, T. Aste, S.T. Hyde, Exchanges in complex networks: income and wealth distributions. 2003, arXiv preprint cond-mat/0310544.
[20] B. Markey-Towler, J. Foster, Understanding the causes of income inequality in complex economic systems (No. 478). University of Queensland, School of Economics, 2013.
[21] H. Akhand, H. Liu, Income inequality in the United States: What the individual tax files say, Appl. Econ. Lett. 9 (2002) 255–259.
[22] P.-C. Lin, H.-C. Huang, Inequality convergence in a panel of states, J. Econ. Inequal. 9 (2011) 195–206.
[23] P.-C. Lin, H.-C. Huang, Convergence of income inequality? Evidence from panel unit root tests with structural breaks, Empir. Econom. 43 (2012) 153–174.
[24] P.-C. Lin, H.-C. Huang, Inequality convergence revisited: Evidence from stationarity panel tests with breaks and cross correlation, Econ. Modell. 29 (2012) 316–326.
[25] M.W. Frank, A new state-level panel of annual inequality measures over the period 1916–2005. Working paper, Sam Houston State University, 2008.

[26] J.L.l. Carrion-i-Silvestre, T. Del Barrio-Castro, E. López-Bazo, Breaking the panels: An application to the GDP per capita, Econom. J. 8 (2005) 159–175.
[27] K. Hadri, Testing for stationarity in heterogeneous panel data, Econom. J. 3 (2000) 148–161.
[28] A. Krozer, The inequality we want: How much is too much? J. Int. Commer. 6 (03) (2015) 1550016. Economics and Policy.
[29] A. Leigh, How Closely Do Top Income Shares Track Other Measures of Inequality?*, Econom. J. 117 (524) (2007) F619–F633.
[30] F. Alvaredo, A note on the relationship between top income shares and the Gini coefficient, Econom. Lett. 110 (3) (2011) 274–277.
[31] R.J. Gordon, I. Dew-Becker, Controversies about the rise of American inequality: A survey. NBER Working Paper No. 13982, 2008.
[32] D.H. Autor, L.F. Katz, M.S. Kearney, Trends in US wage inequality: Reassessing the revisionists, Rev. Econ. Stat. 90 (2008) 300–323.
[33] D.H. Autor, R.J. Murname, F. Levy, The skill content of recent technological change: An empirical exploration, Quart. J. Econ. 118 (2003) 1279–1334.