scientific reports

OPEN

Analyzing LinkedIn data to explore the relationships between sustainable development goals, circular economy, and electoral dynamics

Konstantinos P. Tsagarakis[1,5]✉, Theodoros Daglis[1,4], Konstantinos Gkillas[2] & Amaryllis Mavragani[3]

In 2015, all United Nations Member States adopted 17 Sustainable Development Goals (SDGs) for the 2030 agenda. Addressing the issue of employing alternative data sources for exploring aspects of utilizing said goals, this paper explores the Circular Economy dimension within the SDG12 score, focusing on responsible production and consumption and the broader SDG index. Data from LinkedIn are collected, examining profiles, companies, job postings, and services using the keywords '*Sustainable Development Goals*' and '*Circular Economy*'. Furthermore, the SDG index (including the SDG12 score) for the United States is integrated in the analysis; SDG is a published metric evaluating the progress of sustainable communities within each state. Finally, data on the past five US general elections are retrieved, in order to explore the relationship between SDGs, Circular Economy, and voting behavior. Regression analyses incorporating PCA components and state election data reveal that the LinkedIn-derived SDG and circular economy components exhibit positive impacts on the corresponding indices. Notably, a state's political inclination toward the Republican or the Democratic parties highlights contrasting effects on the SDG and SDG12 indices, indicating divergent trends based on electoral choices. Overall, this study underscores LinkedIn's potential as a valuable source for assessing SDG and Circular Economy position in the US, and highlights the interplay between political factors and sustainable communities at state level.

Keywords Circular economy, LinkedIn, Web-based data, SDGs, SDG12, Social media, US elections

The 2030 Agenda of the United Nations lays out 17 Sustainable Development Goals (SDGs) adopted by all UN Member States in 2015, demanding collective action to end poverty, improve health and education, reduce inequality, foster economic growth, tackle climate change, and protect our oceans and forests[1]. The SDGs represent a global commitment to a sustainable and equitable future for all. To measure the overall performance of different countries and states, a relevant index—the SDG index—was established[1], the latter being derived by combining the individual measurements of said 17 SDGs, with crucial aspect of this index being SDG12, which is specifically assessing responsible consumption and production. Literature suggests a strong interplay between the concepts of SDGs and Circular Economy (CE)[2]. SDG12 focuses on sustainable consumption and production patterns, thus it is the most relevant SDG for assessing CE. In particular, CE aims to minimize waste and make the most of resources, directly aligning with the objectives of SDG12, which encompasses various targets that promote resource efficiency, reduction of waste, and sustainable management of natural resources, all of which are key principles of the CE. Therefore, using SDG12 as an index provides a comprehensive measure of how well an economy or system adheres to CE principles. By utilizing this index, we can evaluate the relationship between sustainable consumption and production patterns and the implementation of CE practices.

[1]School of Production Engineering and Management, Technical University of Crete, Crete, Greece. [2]Department of Management Science and Technology, University of Patras, Patras, Greece. [3]Department of Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Stirling, Scotland, UK. [4]School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Athens, Greece. [5]Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece. ✉email: ktsagarakis@tuc.gr

As indicated by Mesa et al.[3], CE plays a significant role in enhancing the sustainable communities' performance across various aspects, and provides valuable tools for promoting sustainable production and consumption, whilst establishing a close interconnection between these fundamental concepts. CE and the performance of sustainable communities are influenced by several factors, including policies, producers, consumers, and technology[4]. However, the impact of the social aspect on the aforementioned concepts remains insufficiently explored[5]. Padilla-Rivera et al.[6] underscored key social indicators for CE experts, highlighting their significant role in contributing to multiple SDGs aligned with CE strategies.

The monitoring framework for SDGs is still evolving and faces challenges in capturing essential aspects[7,8], while the assessment of the progress of countries and states toward SDGs requires considerable refinement[9]. However, to achieve sustainable communities it is essential to individually address each SDG and specify research outcomes per Goal. Both traditional (national statistical offices, international organizations, government agencies) and non-traditional data sources (earth observation, citizen-generated data) are utilized[10] towards this direction, yet numerous issues persist, e.g., cost, coverage, openness, and accuracy in quantifying the alignment of entities with the SDG framework[10]. Even though existing data and relevant metrics are updated annually, a pressing need for continuous tools and metrics are required to measure SDG progress. Consequently, unlike traditional tools such as surveys, official records, and scientific publications[11], the World Wide Web and social media platforms offer novel ways for data collection and information retrieval[12]. Regarding utilizing LinkedIn as a source of information, a few studies have already showed that it possesses a high amount of information related to CE in various dimensions[13–15], SDGs, and the relationship of CE and SDGs as well, mainly in a European countries' context[16]. Consequently, according to recent advancements in the literature, LinkedIn is an important source of information, demonstrating important strengths, such as clarity, and immediate access, rendering it appropriate for such investigations. However, no study has utilized this source of information in a US case study, while also taking into account electoral preferences.

Acknowledging these issues and gaps in the literature, this study aims at evaluating the suitability of LinkedIn-derived data in approximating measurements from the SDG index and its component related to circular economy, i.e., the SDG12 score. In addition, aiming at investigating the influence of political affiliation within each US state, we explore whether or not a state characterized by Republican or Democratic electoral tendencies affects the examined SDG metrics. This study offers crucial insights into a) the potential of utilizing constantly updated and readily available online data from LinkedIn to estimate the SDG performance related to circular economy and the SDG12 score, and b) the interplay of political dynamics in shaping the SDG performance of US states. Overall, the relationship between the SDG index with all derived variables is explored, and further analysis regarding the keyword search for SDGs within LinkedIn is performed.

Methods

This section details the data sources used. SDGs scores and US voting results were obtained from the respective referenced sources. LinkedIn data were collected manually, as described in the relevant subsection, in accordance with the platform's terms of use. Finally, the statistical analysis is presented.

SDGs scores

The SDG scores for the US were downloaded from the latest Sustainable Development Goals Achievement Report[1]. The Sustainable Development Report also includes the SDG Index and the Dashboards, and it can be regarded as a complement to the official SDG indicators and voluntary country-led review processes. This report utilizes publicly available data that are published by official providers and organizations, e.g., the World Bank, WHO, ILO, etc.[1], and was first published in 2015. Since then, an annual report has been released as well, with the year 2017 being the first year including a report for the US states SDGs index. For reference, the values of the SDGs Index vary from 30.5 (Mississippi) to 60.4 (Vermont). Lynch and Sachs[1] also provide the performance scores of states per individual SDG (the SDG12 score spans from 24 in Wyoming to 79 in Maryland and Vermont). To clarify, the SDG metrics take values from 0 to 100, with 0 denoting the worst performance and 100 the best. According to UN SDG metrics, apart from Maryland's highest SDG12 score which shows that goals have been well achieved, for other US states challenges persist—some of which are significant. For example, the least citing Wyoming faces great challenges. Similarly, the SDG index, with its best score of 60.4, indicates that there is still work to be done to fully achieve the sustainable development goals, since a score of 70 or above is generally considered good, indicating that the country is making significant progress towards achieving the SDGs, although there may still be specific targets or areas that require more attention.

LinkedIn variables and keyword selection

The LinkedIn platform offers a diverse array of tools, various profile types, and subscription tiers that harbor substantial potential for data retrieval. If properly handled, these data could provide us with valuable LinkedIn-based variables suitable for cross-sectional or long-term analyses. To monitor the variables/indicators, we must consider the multidimensional environment of social media in general, and in LinkedIn in specific. Data entries from profiles can be retrieved and treated based on keyword searches.

Data for this study were extracted directly from the LinkedIn platform, i.e., data related to users' Profiles, companies, Services, and Job postings. Users' profiles are measured based on the associated profiles, mainly with individuals, returning hits based on the personal keywords and descriptions that these individuals state. Companies can be searched based on description, industry, and keywords stated by their administrators. Jobs refer to the employment opportunities posted by companies and can be found by specific keywords that job seekers enter for their search. Finally, Services are searches resulting in what people state they can provide in different service categories. We included the total number of service categories. We selected Profiles (People),

Companies, Services, and Job postings (Jobs) for our research as they can be normalized based on geographical reference.

Since this study is focused on the most important aspects of sustainable communities, as termed by the literature review, it was important to select and examine the appropriate keywords based on their relation to SDGs. After careful consideration and combinations of searches and keywords, we selected the terms *Sustainable Development Goals* (without quotes), and *Circular Economy* (without quotes), the latter being considered a driver of SDG12, i.e., Responsible Consumption and Construction[17]. Note that the abbreviations *SDGs* and *SDG* resulted in less than 10% of the results compared to the selected keyword and were thus excluded from further consideration. LinkedIn data were manually retrieved on a single day, March 23rd, 2022, to maintain consistency and avoid significant fluctuations in the dataset. For performing the data collection, an active LinkedIn account is necessary, followed by searching for the selected keywords in the "Search" field, e.g., "circular economy", "sustainable development goals", and then filter for the geographic regions of interest, such as "Vermont", "California", "New York", etc. LinkedIn provides the total number of companies, job posts, people's profiles, and services, among others, which can be noted down by the researchers. It is essential for data collection to be performed within a short timeframe to avoid mismatches, thus in the present study data were collected within a period of a few hours. We downloaded data for all 50 US states, with queries referring to the LinkedIn profiles, job posts, services, and companies, related to SDGs, and CE, and their total number for each State, (needed for the metrics created, mentioned below, to ensure comparability). Consequently, the total volume of information derived is a matrix of 50 per 8 variables, equaling 400 observations extracted. Despite being a demanding and time-consuming task, these data are capable of being transformed into metrics with comparable properties.

For comparability and comprehensiveness, we initially crafted specific indices by dividing each variable by the overall relative information from LinkedIn pertaining to that variable for the same geographical area:

$$P_i^s = \frac{10^6 \times \text{Profile hits for Keyword}_s^i}{\text{Total People's Profiles}_s} \tag{1}$$

$$C_i^s = \frac{10^6 \times \text{Company hits for Keyword}_s^i}{\text{Total Company Profiles}_s} \tag{2}$$

$$J_i^s = \frac{10^6 \times \text{Job posts}_s^i}{\text{Total Job Posts}_s} \tag{3}$$

$$S_i^s = \frac{10^6 \times \text{Services hits for Keyword}_s^i}{\text{Total Services Profiles}_s} \tag{4}$$

where *i* denotes the keywords extracted (*sustainable development goals* and *circular economy*), while *s* declares the 50 US states examined. Using the retrieved LinkedIn data, we formulate specific ratios by dividing each data point by the corresponding total number of related profiles within that specific State. Said ratios highlight specific characteristics, allowing the comparison among states, thereby forming a cross-sectional variable.

In this context, although the variables might exhibit some noise, for large datasets a consistent effect across all states is observed, thereby mitigating the impact of these attributes, as depicted in Fig. 1 (designed by the authors).
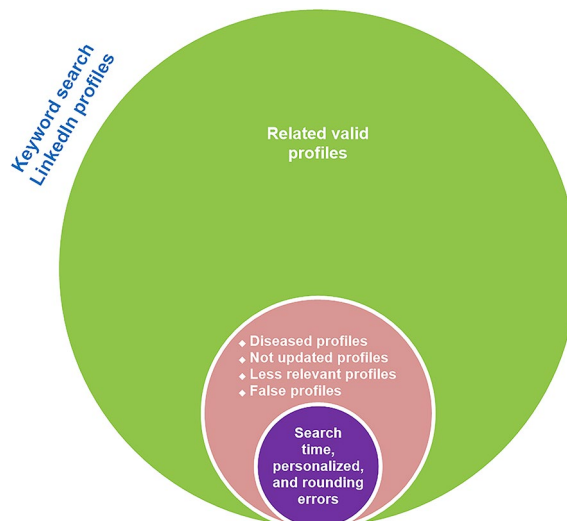

Fig. 1. Potential bias and errors in the proposed methodology.

To elaborate, these instances mainly refer to profiles that are inactive, outdated, less pertinent, or potentially inaccurate. Additionally, considering that LinkedIn employs a personalized search algorithm for users[18], search results may vary based on the individual user and the time the search is conducted[19]; however, this variability remains consistent within each State and is contingent on the user. Moreover, a rounding error occurs due to certain searches resulting in rounded numbers per hundred (if the hit count exceeds 1000) or thousands (for cases surpassing 10,000), contingent on the volume of returned hits. This signifies a rounding error of less than 5% in the hits count.

US voting behavior

To examine the effect of voting behaviors on SDG performance, data on US general elections from 2004 to 2020 are retrieved[20]. The constructed dummy variable (per State) takes the value "1" if the State voted for Republicans in at least 3 out of 5 elections, and the value "0" if the State voted for Democrats in at least 3 out of 5 elections.

Statistical analysis

At first, Spearman correlations are calculated to explore the relationship between each LinkedIn metric and its corresponding SDG index. Following this, we implement Principal Component Analysis (PCA) on the LinkedIn variables to derive the principal component holding the most substantial information, mitigating issues related to multicollinearity. Subsequently, we employ regression analysis, setting the SDG index, or SDG 12 score, as the dependent variable, and using the PCA-derived component and the US elections dummy as independent variables. Finally, we test for the best linear unbiased estimator (BLUE) to validate the effectiveness of the employed models.

Spearman correlation analysis
For a given sample of two variables *Y* and *X*, the ranks are computed by *R(X)* and *R(Y)*, respectively. Following, the correlation formula is applied to these ranks:

$$r_s = \frac{cov\left(R\left(X\right), R\left(Y\right)\right)}{\sigma_{R(X)}\sigma_{R(Y)}} \tag{5}$$

where *cov* denotes the covariance of the ranks, and σ is the standard deviation of each ranked variable. Note that this quantity can demonstrate values between -1 and 1, with -1 indicating perfect negative linear relationship, 0 indicating no linear relationship, and 1 indicating perfect positive linear relationship.

Kernel principal component analysis (PCA)
In case of presence of non-linear trends in the data, which is the case in our dataset as demonstrated by the scatterplots in the Results section, we utilize the kernel PCA. We follow the notations of Scholkopf, Smola, and Muller[21]. While the PCA has the following form:

$$C = \frac{1}{l}\sum_{j=1}^{l} x_j x_j^T \tag{6}$$

where x_j represents the data, with $\sum_{k=1}^{l} x_k = 0$.

If we now apply a non-linear mapping before the PCA, Φ: $\sum_{j=1}^{l} \Phi(x_j) = 0$, then:

$$\overline{C} = \frac{1}{l}\sum_{j=1}^{l} \Phi(x_j)\Phi(x_j)^T \tag{7}$$

and thus, we must find positive eigenvalues and eigenvectors such as $\lambda(\Phi(x_j)v) = (\Phi(x_j)\overline{C}v)$ for all data.

The coefficients should be:

$$v = \sum_{j=1}^{l} a_j \Phi(x_j) \tag{8}$$

Finally, we have:

$$v^k \Phi(x) = \sum_{j=1}^{l} a_j{}^k (\Phi(x_j)\Phi(x)) \tag{9}$$

In this work, we make use of the kernel PCA, utilizing the Laplace or Laplacian kernel function. The most suitable kernel for our case is the Laplacian kernel function, possessing the following key characteristics[22,23]:

i.it exhibits robustness to outliers, displaying less sensitivity to them compared to other kernels,

ii. Laplacian kernels tend to yield sparser solutions, necessitating fewer support vectors to delineate the decision boundary, a feature that contributes to more efficient computation and potentially enhanced generalization to new data points, and

iii. the Laplacian kernel adeptly handles non-linearly separable data by projecting it into a higher dimensional space, potentially rendering the data more linearly separable.

Given these distinctive attributes aligning with our data's characteristics, the optimal choice was employing the Laplacian kernel. Having constructed a single variable that encompasses most of the information, we may formulate a relevant model for assessing the SDG performance of the US states, based on the LinkedIn data and the election history of each state.

Regression analysis
Let Y be the dependent variable (SDG index for each State in the first model, and the SDG12 score in the second model), X_1 be the independent variable (the LinkedIn data as constructed by the PCA method referred above), and let X_2 be the dummy variable based on the election history of each state. Consequently, we have the following:

$$Y = c + \sum_{i=1}^{n} w_i X_i + \varepsilon \tag{10}$$

where c is the intercept, w_i is the coefficient of each independent variable, and finally, ε is the error term of the regression. Note that n is the number of independent variables, which in our case, corresponds to 2.

Note that:

$$X_2 = \left\{ \begin{array}{ll} 0, & if\ the\ state\ has\ voted\ for\ Democrats\ in\ three\ or\ more\ elections \\ 1, & if\ the\ state\ has\ voted\ for\ Republicans\ in\ three\ or\ more\ elections \end{array} \right.$$

Results
At the time of the data collection, LinkedIn companies in the US with the keywords "Sustainable Development Goals" and "Circular Economy" in their profiles consisted of the 20% and 11% of the registered companies in the platform worldwide with the same interests, respectively. For the category "People", the respective percentages were 11% (Sustainable Development Goals) and 6% (Circular Economy). Table 1 provides the descriptive statistics of all LinkedIn derived variables.

Correlation analysis
Sustainable development goals
Initially, we test for the correlation between the SDG index and the individual *Sustainable Development Goals* keyword variables extracted from LinkedIn (Table 2).

Figure 2 consists of the scatter plots of the relationships for the data used for the *Sustainable Development Goals* Spearman correlation analysis.

Based on the findings, the variables of profiles of individuals, companies, and services associated with the keyword *Sustainable Development Goals* on LinkedIn show a robust correlation with the SDG index ($p < 0.001$) across the examined US states, while the Job postings variable referencing SDGs also exhibit a significant correlation with the SDG index with $p < 0.05$. These results reveal that LinkedIn provides substantial information for gauging SDG adoption across various US states, while variables derived from People, Companies, and

Variable	Min	1st quartile	Mean	Standard deviation	Median	3rd quartile	Max
P_{SDGs}^s	103	211	444	375	300	585	2284
$P_{CircularEconomy}^s$	0	5	11	8	9	17	29
J_{SDGs}^s	203	293	375	142	331	413	852
$J_{CircularEconomy}^s$	150	256	292	66	278	326	599
C_{SDGs}^s	0	87	137	75	129	163	331
$C_{CircularEconomy}^s$	0	38	61	33	58	78	140
S_{SDGs}^s	1075	7548	10,321	4775	9956	12,895	24,545
$S_{CircularEconomy}^s$	0	0	304	358	209	421	1579
$SDGIndex$	30	39	46	8	45	53	60
$SDG12Index$	24	48	58	14	60	69	79

Table 1. Descriptive statistics ($\times 10^6$) of the LinkedIn derived variables created for this approach.

Index	Description	r_s	p
P^s_{SDGs}	Sustainable development goals in searches per MM people profiles	0.677	0.000
J^s_{SDGs}	Sustainable development goals in searches per MM jobs posted	0.301	0.034
C^s_{SDGs}	Sustainable development goals in searches per MM company profiles	0.602	0.000
S^s_{SDGs}	Sustainable development goals in searches per MM in services	0.682	0.000

Table 2. Spearman correlation results for the LinkedIn SDGs variables and the SDGs index.
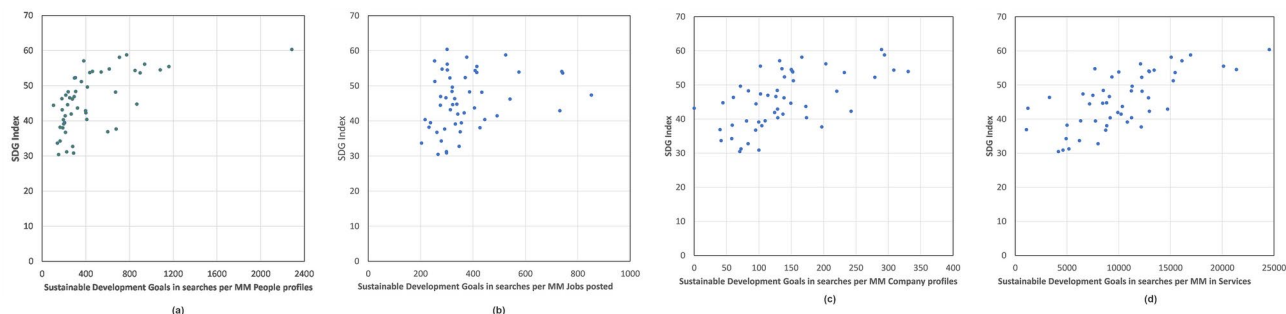
Fig. 2. Relations between the SDG index and LinkedIn activity for (**a**) Personal profiles', (**b**) Job posts, (**c**) Companies, (**d**) Services.

Index	Description	r_s	p
P^s_{CE}	Circular economy in searches per MM people profiles	0.636	0.000
J^s_{CE}	Circular economy in searches per MM jobs posted	−0.015	0.917
C^s_{CE}	Circular economy in searches per MM company profiles	0.515	0.000
S^s_{CE}	Circular economy in searches per MM in services	0.538	0.000

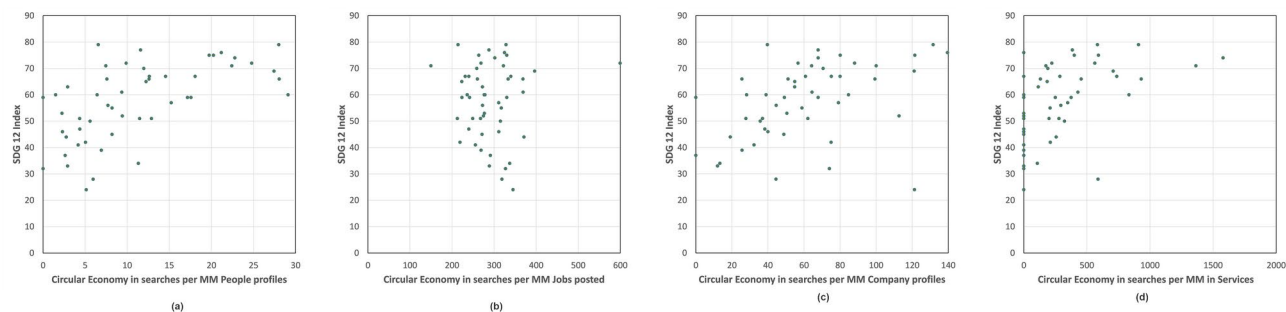**Table 3**. Spearman correlation for the LinkedIn circular economy variables and the SDG12 score.

Fig. 3. Relations among the SDG 12 index with LinkedIn activity on circular economy for (**a**) Personal profiles', (**b**) Job posts, (**c**) Companies, (**d**) Services.

Services demonstrate a cumulative advantage, displaying stronger correlations compared to indices derived from Job postings, which are time-dependent and encompass a shorter time span.

Circular economy
Following, the correlations between the SDG-12 score of each state with the *Circular Economy* keyword variables extracted from LinkedIn are examined, as described in Table 3.

Figure 3 consists of the scatter plots of the relationships for the data used for the *Circular Economy* Spearman correlation analysis.

Like the SDGs approach, the variables constructed with the circular economy keyword on LinkedIn demonstrate a positive and statistically significant correlation with the SDG12 score, apart from the variable referring to Job postings, which in this case does not demonstrate statistically significant results. Following the

same approach as for the SDGs, these results suggest that LinkedIn variables also provide substantial information for measuring SDG12 performance across US states. Similarly, the Job postings variable does not correlate as is time-dependent and encompass a shorter time span.

Principal component analysis

We then proceed with our analysis implementing kernel PCA to eliminate all correlations among the LinkedIn variables, resulting in one independent variable that explains/includes most of the information. According to the kernel PCA results, for both SDGs and CE, the first component demonstrates the most information, and thus, this component is adequate for the data representation.

Sustainable development goals

Figure 4 shows the individual heat maps of the LinkedIn created indices for People, Job postings, Companies, and Services according to the *Sustainable Development Goals* hits, while Fig. 5 illustrates the PCA component derived from the SDG variables and the corresponding SDG index. Note that Figs. 5, 6, 7, 8 were created using R statistical language, utilizing the "Usmap" R Package (P. Di Lorenzo, Usmap, R Package. https://usmap.dev/, 2024).

Circular economy

Similarly, Fig. 6 shows the individual heat maps of the LinkedIn created indices for People, Job postings, Companies, and Services based on the *Circular Economy* hits, while Fig. 7 illustrates the PCA component derived from the CE variables, and the corresponding SDG 12 score.

US voting behavior

Figure 8 depicts the voting behavior for each US State in the last 5 general elections from 2004–2020. As defined in the relative variable construction in the Methods section, the color red indicates that the State voted for Republicans at least 3 out of 5 latest general US elections, while the color blue indicates that the State voted for Democrats at least 3 out of 5 latest general US elections. As evident from the heat map, most of the states voted for Republicans (27 in total), while 23 states voted for Democrats during 2004–2020.
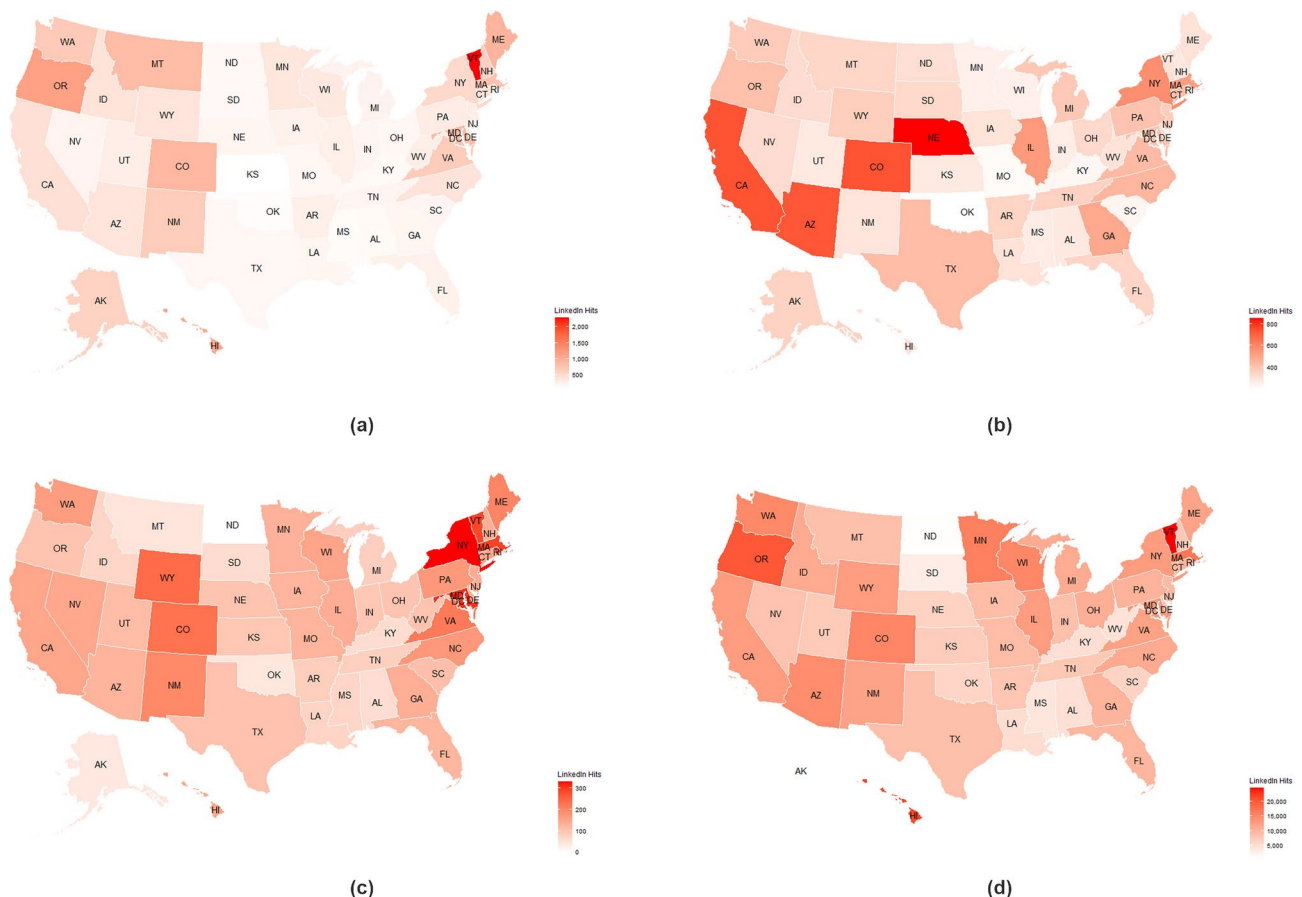

Fig. 4. Normalized LinkedIn activity for SDGs (per million) for (**a**) Personal profiles', (**b**) Job postings, (**c**) Companies, and (**d**) Services.
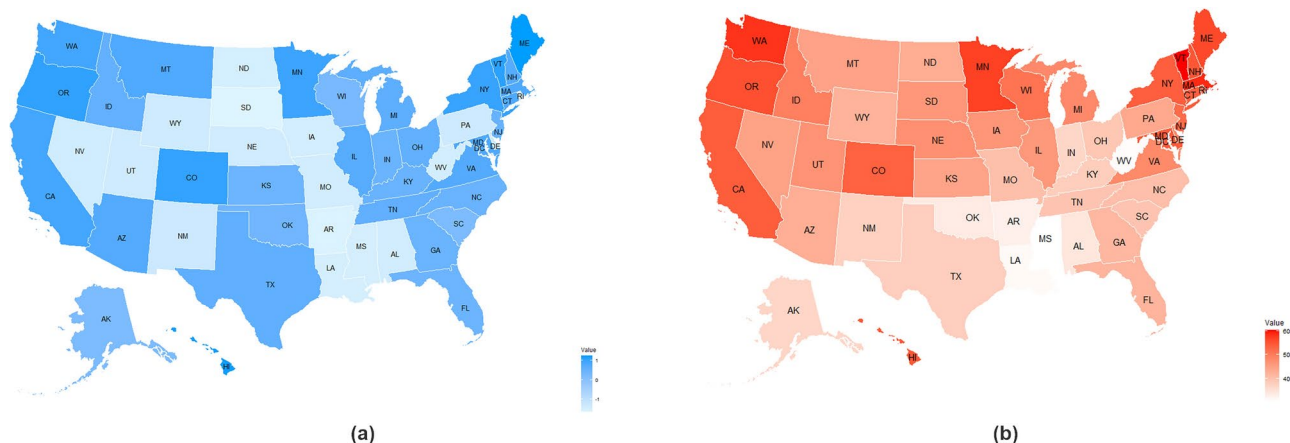
Fig. 5. States' performance: (**a**) LinkedIn PCA variable resulting from P_{SDGs}^s, J_{SDGs}^s, C_{SDGs}^s, and S_{SDGs}^s, (our calculations) and (**b**) SDGs score (adopted from[1]).
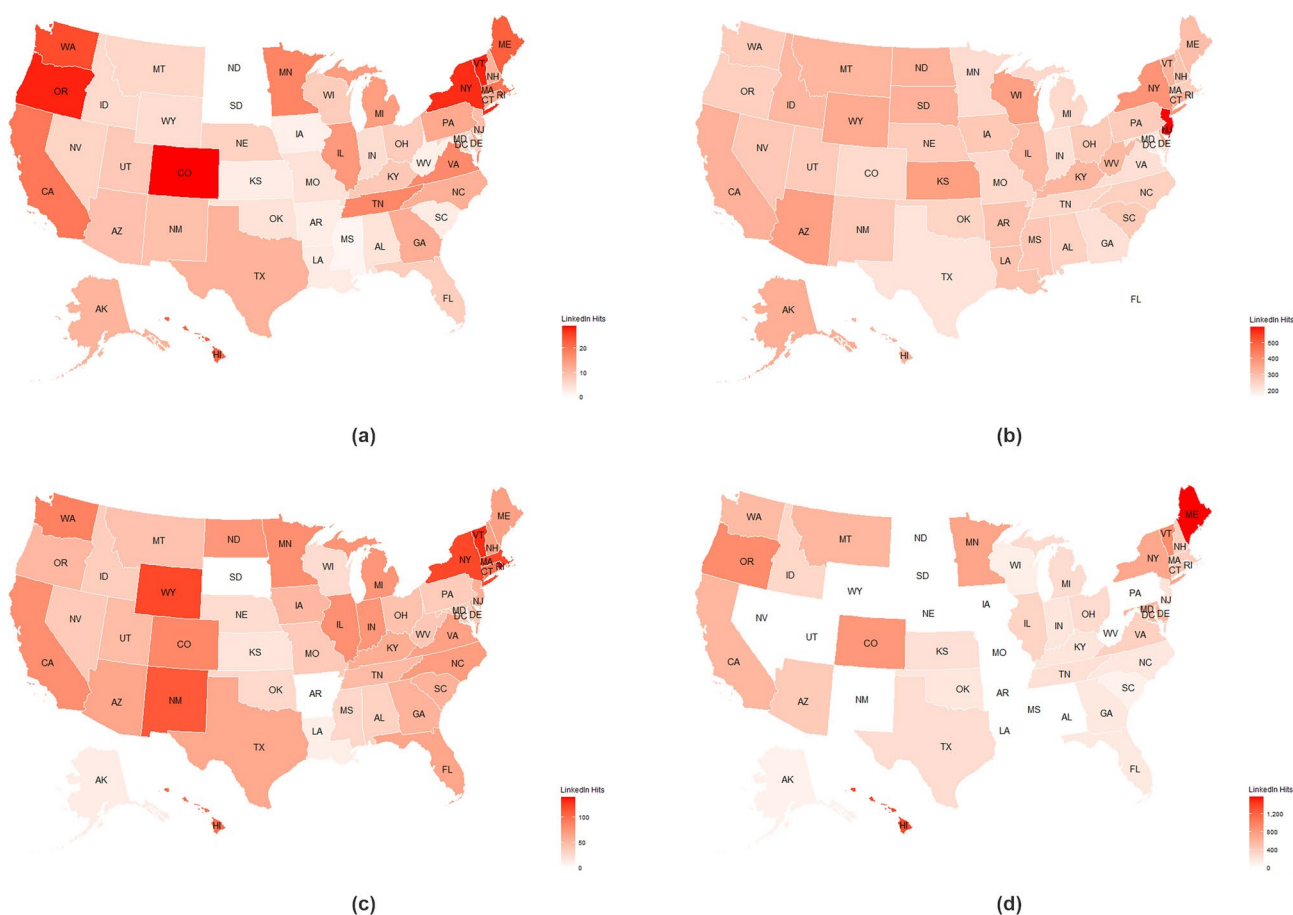

Fig. 6. Normalised LinkedIn activity for Circular Economy (per million) for (**a**) Personal profiles', (**b**) Job postings, (**c**) Companies, and (**d**) Services.

Regression analysis

Before conducting the regression analysis, it is essential to ensure the absence of multicollinearity amongst our independent variables. Utilizing the variance inflation factor (VIF), we calculated a value of 1.233 for the LinkedIn SDG approach, and 1.106 for the Circular Economy, significantly below the threshold of 5, hence, no multicollinearity is detected among the independent variables for both regression models. Consequently, we can confidently incorporate both the PCA component and the elections dummy variable into our models,
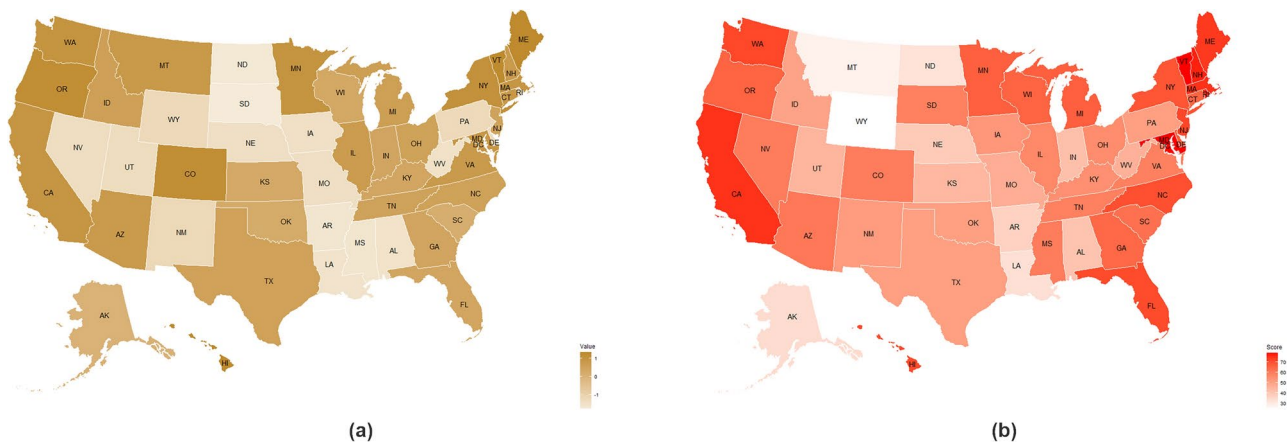
Fig. 7. States' performance: (**a**) LinkedIn PCA variable resulting from P_{CE}^s, J_{CE}^s, C_{CE}^s, and S_{CE}^s, (our calculations), (**b**) SDG 12 score (adopted from[1]).
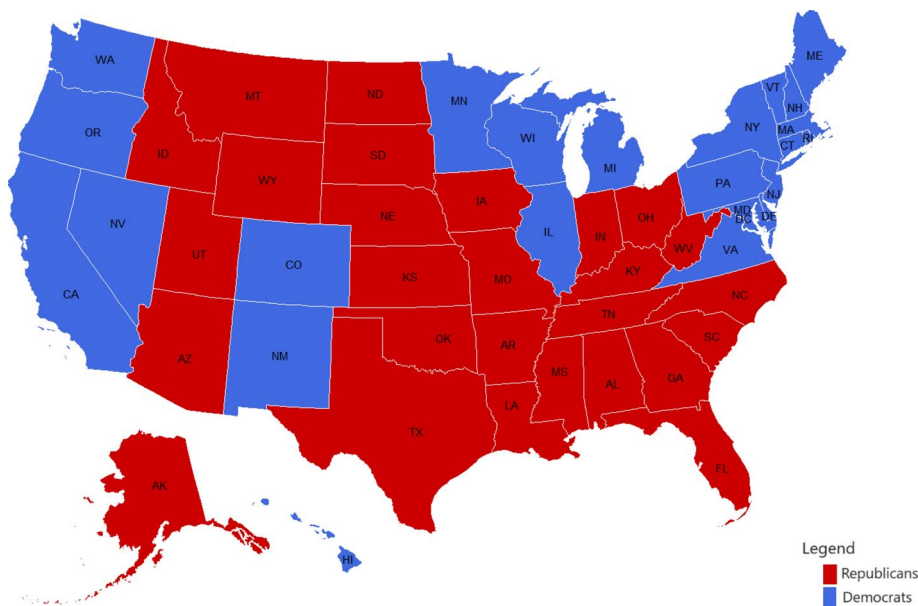

Fig. 8. USA states' classification based on the last five US general elections (2004–2020).

Variable	Beta	Standard error	t-stat	p-value	Robust standard error	Robust t-stat	Robust p-value
Intercept	3.883	0.031	126.370	0.000	0.023	168.497	0.000
Elections	− 0.148	0.048	− 3.074	0.004	0.034	− 4.389	0.000
PCA	0.059	0.018	3.344	0.002	0.014	4.177	0.000

Table 4. Regression results with robust standard errors, for the SDG model. *Note that Adjusted $R^2 = 0.580$, Multiple $R^2 = 0.597$, while the F-statistic of the model is 34.860 and its corresponding p-value < 0.001.

while, to ensure comprehensive analysis, we also present the results using robust standard errors, validating the robustness of our approach. The regression results for the SDG model are included in Table 4.

The respective tests examining for the Best Linear Unbiased Estimator (BLUE) for the SDG model are presented in Table 5.

According to the regression findings, the LinkedIn data referring to SDGs (specifically, the principal component derived from this data) demonstrates a positive impact on the SDG index. However, the US State electoral presidency exhibits a negative effect on the SDG index. In specific, when a State elects Republicans, there's an anticipated decrease in the SDG index, while the opposite trend occurs for Democrats. The regression

Test	Null Hypothesis	Test Value	p-value	Result
Breusch-Pagan	Homoscedasticity of residuals	1.19	0.552	Homoscedasticity of residuals
Goldfeld-Quandt	Homoscedasticity of residuals	1.66	0.121	Homoscedasticity of residuals
Durbin-Watson	Uncorrelated residuals	1.9625	0.4154	Uncorrelated residuals
Breusch-Godfrey	Uncorrelated residuals	0.009	0.925	Uncorrelated residuals
Kolmogorov–Smirnov	Normal residuals	0.0951	0.7203	Normal residuals
Anderson–Darling	Normal residuals	0.571	0.1315	Normal residuals

Table 5. Best linear unbiased estimator tests for the SDG model.

Variable	Beta	Standard error	t-stat	p-value	Robust standard error	Robust t-stat	Robust p-value
Intercept	4.177	0.045	93.586	0.000	0.027	155.291	0.000
Elections	− 0.301	0.062	− 4.841	0.000	0.059	− 5.107	0.000
PCA	0.034	0.011	2.937	0.005	0.012	2.870	0.006

Table 6. Regression results with robust standard errors, for the circular economy model. *Note that Adjusted $R^2 = 0.469$, Multiple $R^2 = 0.490$, while the F-statistic of the model is 22.610 and its corresponding p-value < 0.001.

Test	Null Hypothesis	Test Value	p-value	Result
Breusch-Pagan	Homoscedasticity of residuals	6.492	0.039	Homoscedasticity of residuals
Goldfeld-Quandt	Homoscedasticity of residuals	1.563	0.151	Homoscedasticity of residuals
Durbin-Watson	Uncorrelated residuals	1.995	0.461	Uncorrelated residuals
Breusch-Godfrey	Uncorrelated residuals	0.778	0.378	Uncorrelated residuals
Kolmogorov–Smirnov	Normal residuals	0.093	0.750	Normal residuals
Anderson–Darling	Normal residuals	0.581	0.124	Normal residuals

Table 7. Best linear unbiased estimator tests for the circular economy model.

results for the CE model are included in Table 6, while the relevant tests examining for BLUE are presented in Table 7.

The LinkedIn data referring to *Circular Economy* (in specific, the principal component derived from this data) demonstrates a positive impact on the SDG12 score. However, the US State electoral preference exhibits a negative effect on the SDG12 score. In specific, when a State votes Republican, there is an anticipated decrease in the SDG12 score, while the opposite trend occurs for Democrats. The findings suggest that the Republican states (as defined for this study in the respective subsection in the Methods) tend to exhibit notably lower levels of SDG and CE performance. The results of Tables 5 and 7 suggest that the estimators of the regressions are BLUE, and therefore, our results are valid for both models.

Discussion

SDGs and CE are notions closely related[2,24], affecting the macro-performance of countries and states. In this context, there are many implications deriving from examining the degree of SDGs adaptation from companies, states, and countries, especially regarding the circular economy approach[25], which should be taken into consideration. There is room for improvement when discussing about global, precise, high-frequency indicators that can represent SDGs and circular economy adaptation, necessitating the need for better explanatory data, and models[26] that can analyze, explain, and predict the adaptation and trends of different countries. Many efforts have been undergone in the past to construct indicators or even capture important information derived from them[27], even beyond conventional data sources[28] and the use of big data[29].

The monitoring framework is still maturing in measuring aspects that are important for SDG concepts[7,8], and still requires significant effort. There are traditional data sources (national statistical offices, international organizations, other ministries, and government agencies) and non-traditional data sources (earth observation, spatial data infrastructure, citizen-generated data, official sensor networks, and commercial data)[10]. However, many issues emerge in the current approaches that try to quantify and measure the adaptation of various countries, states, and organizations into the SDG framework such as being the cost to collect, coverage, openness, and accuracy amongst others[10]. Suggestions about how these issues can be addressed include the focus on a limited set of key indicators or the use of different information sources that currently are utilized[30] or the use of multi-criteria[31].

Social media platforms are regarded as non-traditional data sources, and are gaining momentum since they can provide real-time information even for rural and remote places[32]. Said information have yet to be utilized

in the SDG monitoring. In this study, we compare our results with the respective ones derived from the SDG index[33] since it is regarded as one of the best proxies for the investigation for the adoption of these concepts.

Based on our findings, the LinkedIn constructed SDG indices exhibit a strong correlation with the SDG index across the examined US states, while the corresponding index related to circular economy demonstrate a relationship with SDG12 score. These outcomes suggest that LinkedIn serves as a valuable source of information concerning their adoption across different states. Moreover, the role of the dominant political party per state also plays a significant role in the SDG performance and the SDG12 score of the respective state, and thus, understanding the influence of political affiliations on SDG performance within states, highlights the role of governance in shaping sustainable communities. In this regard, the present approach proposes that both parties should consider increasing their interest in SDG performance and adopting relevant strategies. Our findings and the proposed methodology stand to assist decision-makers within US states and other nations aligned with UN agendas, offering a pathway towards uniform progress in achieving sustainable communities.

These indices might serve as indicators for evaluating how different states are embracing and implementing sustainable practices aligned with specific SDGs. In this regard, the results indicate that the measures and metrics encapsulated within the derived indices strongly align with the objectives outlined in SDG12 as well as the overall SDG index. This insight suggests that by focusing on enhancing circular economy practices, states can effectively target and contribute to the objectives outlined in SDG12. Moreover, it implies that assessing circular economy indices might provide a window into a state's overall commitment to sustainable communities, particularly concerning responsible consumption and production, which are core aspects of SDG12. Ultimately, recognizing this relationship, policymakers and stakeholders can be guided in evaluating and improving their state's sustainable practices. The results also underscore the potential value of these indices as tools for not only measuring progress, but also steering policies and initiatives towards achieving sustainable communities, thereby facilitating a more comprehensive and targeted approach to sustainable development efforts within different states. While strategic planning is essential to achieve sustainable development[34] through circular economy, our findings can give a quantitative input for achieving the goals set.

As for the US elections, data reveal that states favoring Republicans over Democrats tend to exhibit notably lower levels of CE and SDG performance. This may be due to several reasons. First, it can be attributed to policy differences prioritizing economic growth and business interests or environmental and social issues potentially leading to different scores in CE (circularity, responsible production and consumption, etc.) and SDG performance (sustainable development goals encompassing social, economic, and environmental objectives). Furthermore, another probable reason may be resource allocation or funding towards social welfare programs or sustainable development initiatives. Similarly, the regulatory framework may be different, since some states might have less stringent regulations concerning environmental protection and social welfare, negatively impacting their CE and SDG performance. Finally, economic emphasis and orientation could be a significant reason in states prioritizing economic growth and industry over social and environmental concerns, which could result in policies that favor business interests, but might have less positive influence in CE and SDG performance.

These variables offer the opportunity for comparative analysis over time or among various concepts under investigation, and notably, the nature of LinkedIn data suggests that subsequent data mining could yield observations suitable for robust time-series analyses[16]. This could provide high-quality and potentially high-frequency data for future investigations into changes in SDGs, including wider topics such as sustainability, or more specific ones, such as Circular Economy adoption, notions related to SDGs[2,24], while, extending this approach to other countries or regions could also render it globally applicable. Moreover, LinkedIn, hosts several features worthy of exploration; for instance, it could serve as the basis for various case studies such as probing into company profiles, discerning sectoral performances[13], and leveraging specific attributes like companies' specialties and their engagement in vital concepts such as sustainability, SDGs, Circular Economy, or R-imperatives[35]. In the current intensification of climate change, an investigation employing timely data could hold significant value and relevance. Towards this direction, and since to achieve sustainable communities, it is essential to individually address each SDG and specify research outcomes per Goal[36], the World Wide Web and social media offer novel ways for data collection and information retrieval[12,37].

Limitations

This study has limitations. As a general limitation of such approaches, which does not necessarily apply to this study as it was conducted in the US, is the language barrier in non-English speaking countries, where the analysis should be performed in both English and the native language of the country. In addition, data may slightly vary depending on the time of collection. Moreover, the limited availability of data constrained our observations, precluding the possibility of employing a time-series analysis. Another limitation is that data from only 5 US general elections were employed, which may result in biases.

As is the case with all web-based data (Twitter, LinkedIn, Google Trends, etc.), one of the limitations is that representativeness and causality cannot be shown. However, these approaches have been indicated to be rigorous enough to provide valid results, especially in topics like epidemiology and public health. Most of these approaches make use of big data, which is indicating that, given the volume of the data collected and analyzed, over- or under- representation would not be risky for the reliability and validity of the results, as any outliers or errors would be within the statistical error. To address this limitation in all such approaches, as has been also shown in the case of Google Trends (e.g., with Google Flu Trends), the key is a rigor and detailed data collection methodology, where the appropriate keywords and filters are selected. Furthermore, in regions with high internet and LinkedIn penetration, like the US, where the relative normalized values per state are used, variables follow offline data[13,16].

Moreover, we have thoroughly reviewed all available retrievable information from LinkedIn, and, despite its significance, the plethora of information, and accessibility at any time, specific Sustainable Development Goals

(e.g., SDG12), do not yield sufficient results. Most companies, profiles, etc., primarily reference SDGs or the Circular Economy (which is closely related to SDG12), rather than SDG12 specifically. Therefore, we did not acquire data specifically for SDG12 due to its limited availability.

Nevertheless, if more data were to become available in the future, an updated analysis could be performed. Lastly, we did not individually consider all 17 SDG indices that comprise the overall SDG index, but only the SDG12 that relates to circular economy. This decision stemmed from our focus on assessing the overall SDG performance for each state, and due to the unavailability of adequate relevant data on the LinkedIn platform for further exploration of these individual indices.

LinkedIn profiles, jobs, etc. are dynamic, and since any changes in these categories would be instantly depicted online, data may slightly vary depending on the time of collection. Therefore, as mentioned in Methods section, data collection should be performed within a matter of hours/days cross-sectionally, in order for any variations/changes in the respective fields of search (e.g., companies) to be very limited.

Finally, a limitation that applies to all social media has to do with the fact that these platforms are dynamic, and they are constantly modifying their design or retrieved/offered information, which may make long term time data comparisons not possible.

Practical implications

Our approach proposes complementary indices as metrics for assessing SDGs, and several implications and policy proposals may arise from this work. First, the encouragement of the development of policies that balance economic growth with environmental protection and social welfare are of great significance, providing strategies that demonstrate how economic growth can be achieved, whilst promoting sustainable communities and social equity. Moreover, incentives for sustainable practices for businesses and industries, involving tax breaks, grants, or subsidies for companies that invest in renewable energy, reduce of carbon emissions, or implement socially responsible initiatives, could further contribute to an environmentally friendly business activity. Furthermore, a focus on educating the public about the benefits of sustainability and social responsibility could provide knowledge of the importance of these notions' adoption. Long-term planning would prioritize sustainability and social progress, setting clear, measurable goals aligned with SDGs and regularly assessing progress that may help drive positive change over time.

In addition, ensuring transparency in governance and accountability for policies related to CE and SDG performance, establishing clear metrics, reporting mechanisms, and regular evaluations, can hold policymakers accountable for their decisions. Moreover, encouraging policy innovation by studying successful models from both Republican and Democrat-leaning states, could help policymakers adapt and implement effective strategies, whilst considering the unique needs and contexts of their state. In this regard, a collaboration in various levels could provide notable results, for instance between government entities, private organizations, and NGOs, or even cross-partisan collaborations, working collaboratively on initiatives that promote sustainability and social welfare, developing comprehensive policies. Such initiatives could lead to more balanced and widely accepted approaches to CE and SDGs, leveraging the strengths of each unit, amplifying the impact of policies and programs.

Conclusions

The information derived from LinkedIn in the context of this work includes data collected from user profiles, companies, job postings, and services that use the keywords 'Sustainable Development Goals' and 'Circular Economy'. This data can provide insights into how individuals and organizations are engaging with these concepts, as LinkedIn profiles might highlight specific skills or experiences related to the SDGs, companies might list job postings that require knowledge of circular economy principles, and services might be described as aligning with sustainable practices.

LinkedIn addresses issues related to the SDGs and the circular economy, as job postings seeking a 'Sustainability Manager' whose role involves implementing circular economy strategies to reduce waste and promote resource efficiency would be of great interest. Furthermore, LinkedIn profiles of professionals who have completed projects or certifications in sustainable development and are now working in roles that directly contribute to achieving the SDGs, would also be another example of the usefulness of LinkedIn. Consequently, the data derived from LinkedIn helps measure how widely and effectively the SDGs and circular economy concepts are being adopted and practiced within various professional and organizational contexts in the United States.

As far as real-world indicators are concerned, different US states cite a different SDG index, demonstrating a different-level performance. The greatest performance is cited by Vermont (around 60), while the minimum (around 30) by Mississippi. The average SDG index score of the U.S. states is approximately 46, with a standard deviation of 8, which means that the U.S. has yet to achieve high performance in the SDG goals. However, this work indicates that LinkedIn data possesses valuable information for the performance of the US states in an SDG and CE frameworks, while the political tendency of each state also predicts its SDG and CE performance. Finally, this work lays the groundwork for future investigations that can investigate environmental aspects with political dimensions of the various countries in diverse contexts.

Data availability

All data are publicly available online. The raw and curated datasets are available upon reasonable request.

References

1. Lynch, A. & Sachs, J. The US State Sustainable Development Report 2021. https://www.sdgindex.org/reports/united-states-sustainable-development-report-2021/ (SDSN, 2021).
2. Belmonte-Ureña, L. J., Plaza-Úbeda, J. A., Vazquez-Brust, D. & Yakovleva, N. Circular economy, degrowth and green growth as pathways for research on sustainable development goals: A global analysis and future agenda. *Ecol. Econ.* **185**, 107050 (2021).
3. Mesa, J., Esparragoza, I. & Maury, H. Developing a set of sustainability indicators for product families based on the circular economy model. *J. Clean. Prod.* **196**, 1429–1442 (2018).
4. Mora-Contreras, R., Torres-Guevara, L. E., Mejia-Villa, A., Ormazabal, M. & Prieto-Sandoval, V. Unraveling the effect of circular economy practices on companies' sustainability performance: Evidence from a literature review. *Sustain. Prod. Consum.* **35**, 95–115 (2023).
5. Walker, A. M. et al. Assessing the social sustainability of circular economy practices: Industry perspectives from Italy and the Netherlands. *Sustain. Prod. Consum.* **27**, 831–844 (2021).
6. Padilla-Rivera, A., Telles do Carmo, B. B., Arcese, G. & Merveille, N. Social circular economy indicators: Selection through fuzzy Delphi method. *Sustain. Prod. Consum.* **26**, 101–110 (2021).
7. Ranta, V., Aarikka-Stenroos, L., Ritala, P. & Mäkinen, S. J. Exploring institutional drivers and barriers of the circular economy: A cross regional comparison of China, the US, and Europe. *Resour. Conserv. Recycl.* **135**, 70–82 (2017).
8. Prieto-Sandoval, V., Jaca, C. & Ormazabal, M. Towards a consensus on the circular economy. *J. Clean. Prod.* **179**, 605–615 (2018).
9. Chaparro-Banegas, N., Ibañez Escribano, A. M., Mas-Tur, A., & Roig-Tierno, N. Innovation facilitators and sustainable development: a country comparative approach. *Environ. Dev. Sustain.* 1–29 (2023).
10. Fritz, S. et al. Citizen science and the united nations sustainable development goals. *Nat. Sustain.* **2**(10), 922–930 (2019).
11. Yamaguchi, N. U. et al. Sustainable development goals: A bibliometric analysis of literature reviews. *Environ. Sci. Pollut. Res.* **30**(3), 5502–5515 (2023).
12. Debnath, R. et al. Social media enables people-centric climate action in the hard-to-decarbonise building sector. *Sci. Rep.* **12**(1), 19017 (2022).
13. Tsironis, G., Daglis, T. & Tsagarakis, K. P. Social media and EU companies' engagement in circular economy: A LinkedIn approach. *Sustain. Prod. Consum.* **32**, 802–816 (2022).
14. Knäble, D. & Tsagarakis, K. P. 'Made in Germany' how companies approach Circular Economy on LinkedIn. *Eur. Plan. Stud.* **32**(5), 927–951 (2024).
15. Tsironis, G., Daglis, T. & Tsagarakis, K. P. The 21 most practiced RE-s of Circular Economy from LinkedIn Company Profiles on a Global Scale. *Resour. Conserv. Recycl. Adv. Conserv. Recycl. Adv.* **21**, 200202 (2024).
16. Daglis, T., Tsironis, G. & Tsagarakis, K. P. Data mining techniques for the investigation of the circular economy. *Resour. Conserv. Recycl. Adv.* **19**, 200151 (2023).
17. Ma, W., Hoppe, T. & de Jong, M. Policy accumulation in China: a longitudinal analysis of circular economy initiatives. *Sustain. Prod. Consum.* **34**, 490–504 (2022).
18. LinkedIn. LinkedIn Search Relevance - People Search. https://www.linkedin.com/help/linkedin/answer/a524188/linkedin-search-relevance-people-search?lang=en, (Accessed 12 Dec 2023) (2023).
19. LinkedIn. LinkedIn Pages in Search Results. https://www.linkedin.com/help/linkedin/answer/a554382, (Accessed 12 Dec 2023) (2023).
20. Woolley, J. & Peters, G. *The American Presidency Project*, https://www.presidency.ucsb.edu/statistics/elections (Assessed 11 Nov 2023).
21. Scholkopf, B., Smola, A., & Muller, K. R. (n.d.). *Kernel Principal Component Analysis*. Link: https://people.eecs.berkeley.edu/~wainwrig/stat241b/scholkopf_kernel.pdf, (Accessed 22 Nov 2023).
22. Shawe-Taylor, J., Nello, C. *Kernel Methods for Pattern Analysis*, Cambridge University Press (2004).
23. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
24. Ogunmakinde, O. E., Egbelakin, T. & Sher, W. Contributions of the circular economy to the UN sustainable development goals through sustainable construction. *Resour. Conserv. Recycl.* **178**, 106023 (2022).
25. Martinho Domingues, P. J. Insights into circular economy indicators: Emphasizing dimensions of sustainability. *Environ. Sustain. Indic.* **10**, 100119 (2021).
26. Bocken, N. M., Niessen, L. & Short, S. W. The sufficiency-based circular economy—an analysis of 150 companies. *Front. Sustain.* **3**, 899289 (2022).
27. Lee, R. & Kim, J. Developing a social index for measuring the public opinion regarding the attainment of sustainable development goals. *Soc. Indic. Res.* **156**, 201–221 (2021).
28. Mishra, M. et al. A bibliometric analysis of sustainable development goals (SDGs): a review of progress, challenges, and opportunities. *Environ. Dev. Sustain.* 1–43 (2023).
29. Rathinam, F. et al. Using big data for evaluating development outcomes: A systematic map. *Campbell Syst. Rev.* **17**(3), e1149 (2021).
30. Lyytimäki, J. Seeking SDG indicators. *Nat. Sustain.* **2**, 646 (2019).
31. Anselmi, D., D'Adamo, I., Gastaldi, M., & Lombardi, G. V. A comparison of economic, environmental and social performance of European countries: a sustainable development goal index. *Environ. Dev. Sustain.* 1–25 (2023).
32. Community contributions to sustainability science. *Nat. Sustain.* **5**, 809–810 (2022).
33. Sachs, J., Traub-Schmidt, G., Kroll, C., Lafortune, G. & Fuller, G. *Sustainable Development Reports 2021* (Cambridge University Press, 2021).
34. Puglieri, F. N. et al. Strategic planning oriented to circular business models: A decision framework to promote sustainable development. *Bus. Strategy Environ.* **31**(7), (2022)
35. Circular Asia. *The Circular Rs*, https://www.circulareconomyasia.org/circular-rs/. (Accessed 22 Jan 2023) (2022).
36. Ali, S. M. et al. Development goals towards sustainability. *Sustainability* **15**, 9443 (2023).
37. Bressanelli, G., Adrodegari, F., Pigosso, D. C. A. & Parida, V. Circular economy in the digital age. *Sustainability* **14**, 5565 (2022).

Acknowledgements

Author contributions

KT conceived the idea and had the overall supervision; KG and AM designed the methodology; KT and TD performed the data collection; KG designed the statistical analysis approach; TD performed the statistical analysis; TD, KT, and KG interpreted the data; TD and KT drafted the first version of the manuscript; AM reviewed and edited the final version of the manuscript.

Declarations

Ethical considerations
This study is observations of public behaviors of research data. The study data are anonymous and de-identified. Thus, IRB approval was not sought.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to K.P.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.