CIRCULAR ECONOMY: NEW PARADIGM OR JUST RELABELLING? A QUANTITATIVE TEXT AND SOCIAL NETWORK ANALYSIS ON WIKIPEDIA WEBPAGES

DOI:10.13140/RG.2.2.32576.99843

CIRCULAR ECONOMY: NEW PARADIGM OR JUST RELABELLING? A QUANTITATIVE TEXT AND SOCIAL NETWORK ANALYSIS ON WIKIPEDIA WEBPAGES A PREPRINT

Preprint · May 2019

DOI: 10.13140/RG.2.2.32576.99843

Dario Cottafava
Università degli Studi di Torino

Grazia Sveva Ascione
Università degli Studi di Torino

Ilaria Allori
Università di Pisa

CIRCULAR ECONOMY: NEW PARADIGM OR JUST RELABELLING? A QUANTITATIVE TEXT AND SOCIAL NETWORK ANALYSIS ON WIKIPEDIA WEBPAGES

A PREPRINT

Dario Cottafava*
Department of Physics
University of Turin
Turin, 10100, Italy,
dario.cottafava@unito.it

Grazia Sveva Ascione
Department of Economics and Statistics
University of Turin
Turin, 10100, Italy
grazia.ascione@unito.it

Ilaria Allori
Department of Engineering of Energy, Systems, Territory and Construction
University of Pisa
Pisa, 56122, Italy
ilaria.allori@hotmail.it

May 30, 2019

ABSTRACT

Circular Economy has definitely gained a momentum among both, researchers and practitioners. The result is the production of multifarious knowledge, involving different disciplines such as economics, chemistry, design and Industrial Ecology. However, despite the steep growth of articles in Scopus, critics hold that blurriness in defining the Circular Economy is occurring and, due to its interdisciplinary but ambiguous nature, a systemic approach is required in order to set its boundaries. In this work, an innovative methodology is presented and discussed to identify which network of keywords exhaustively depicts the Circular Economy domain, in order to assess if the Circular Economy might be considered as a new paradigm or it represents a relabelling of already existent knowledge. The methodology consists in four steps: first, an initial seed list of keywords related to the Circular Economy is extracted from the Scopus database. Secondly, the network of the overall Wikipedia pages is discussed and, thirdly, two sub-networks, one representing the relevant fields and one the crucial technologies, are pointed out by classifying Wikipedia pages depending on the presence of a specific term in the first sentence. Finally, centrality indices are discussed and the most important clusters are highlighted, revealing the six main fields related to the Circular Economy: Sustainability, Material Flow Analysis, Waste Management, Water Management, Waste Electrical and Electronic Equipment (WEEE) and Bioeconomy.

Keywords Circular Economy · Social Network Analysis · Wikipedia

1 Introduction

The linear model of production – based on the "take, make, and dispose" motto, which relies on natural resources and generates wastes and emissions – appears nowadays outdated. Environmental problems, such as pollution of air, soil and water, as well as biodiversity loss, resources' reduction and extensive land use are progressively putting the earth's ecosystem in jeopardy [1, 2, 3], hence requiring the adoption of a novel approach. To address this issue, the paradigm of Circular Economy (CE) has gained a momentum among researchers and practitioners [4], aiming to develop new

*https://inno-ce.campusnet.unito.it/do/docenti.pl/Show?_id=dcottafa sustainable technologies, practices and methodologies. However, despite the steep growth in Scopus publications (around 350 in 2016 and more than 1’000 in 2018) and consultancy reports, as well as the attention received from important policy makers [5], critics hold that blurriness occur when a concept operate in significantly diverse worlds of thoughts.

Much of the blurriness is generated by the different fields which involve CE as well as by the overlapping with the wider issue of Sustainability. According to Geissdoerfer et al. [6], the concepts present several differences. For instance, they have different origins, because the idea of Sustainability is noticeably older and institutionalised by environmental movements and supranational bodies, while the modern understanding of the Circular Economy spilled from the efforts of Ellen MacArthur Foundation (EMF 2013) and different schools of Industrial Ecology. Moreover, CE and Sustainability also differ in terms of goals because the Circular Economy aims to generate a closed loop, getting rid of resources inputs, as well of waste, emissions and leakages of the system, while Sustainability has several aims. In that sense, Circular Economy seems to give priority to the economic system, generating, at the same time, considerable benefits for the environment, but only accessory improvement for the social system, while Sustainability was conceived in order to include, holistically, all three dimensions – environment, economy and society - as equally significant. Some authors tried to overcome the blurriness around CE through bibliometric analysis [6, 7, 8] and systematic literature reviews [9]. These studies resulted in an increased transparency about the understanding of CE and its fundamental principles. On the other hand, they could not provide any information about terms definition and the linkages between them nor they could define a dictionary about CE. Considering the necessity of systematically analysing the CE concept, it is clear that a methodology which goes beyond the analysis and comparison of different definitions would be needed. This research work could help not only to define the boundaries of the CE, but also to analyse the relationships between different disciplines and spheres of knowledge, trying to establish the presence or not of elements of novelty in the CE paradigma.

In this paper we use a novel approach, based on a previous work by Chiarello et al. [10]. The key features of this approach are:

1. a description of the Circular Economy is provided through an enriched dictionary and a collection of lemmas with their linkages;
2. the obtained dictionary is expert independent, as it has been created using Wikipedia;
3. the boundaries of Circular Economy are defined endogenously;

The paper is structured as follows. Section 2 provides a theoretical description of CE in its complexity. In Section 3 the methodology used in this research work is presented in detail. In Section 4 the results are presented and discussed. Section 5 acknowledges the limitations and the hints for further research. Section 6 concludes the paper.

2 Circular economy: relabelling of old knowledge?

The idea of the Circular Economy is attributed to Pearce and Turner [11] in a study which analysed the connections between the environment and economic activities [12]. The scholars assessed the characteristics of the linear and open-ended economic system evaluating the role of natural resources in the economy for the production and consumption system and generating outputs in the form of waste. Their vision was inspired by Boulding [13] who perceived the Earth as a spaceship, a closed and circular system with limited assimilative capacity; the author argued it should coexist in equilibrium with the environment. Afterwards, the Circular Economy has been brought closer to the concept of industrial economies by Stahel and Reday [14]. They viewed loop economy in relation to the industrial strategies for waste prevention, along with regional job creation, resource efficiency, and dematerialisation of the industrial economy. Stahel [15] also highlighted the opportunities of selling utilisation of products instead of their ownership, i.e. "product as a service", hence promoting a sustainable business model in terms of loop economy, allowing value and profit creation for companies without the externalization of costs and risks related to waste.

As shown in Figure 1, the query of Circular Economy in Scopus generates 4’583 results, depicting an upwards trend in the production of papers from 2010 to 2018. In 2010, for instance, papers produced about CE were just 154, while in 2018 they reached the number of 1’287; the steep increase is around 736 %. This confirms the relevance of the topic. Due to governmental policies and strategies, chinese and european academics were the most interested by CE, defining methodologies and practical tools to support decision-making processes and awareness among the actors involved. Over the previous decades, the concept of Circular Economy emerged as blend of several theoretical influences such as cradle-to-cradle [16], laws of Ecology [17], regenerative design [18], Industrial Ecology [19], looped and performance economy [20], Biomimicry [21] and the blue economy [22]. Nowadays, CE studies follow three main lines of action [9]: the first targets the social and economic dynamics at macro and administrative level; the second aims to to encourage firms’ participation in circular practices implementation to encourage the diffusion of new kinds of product design; the

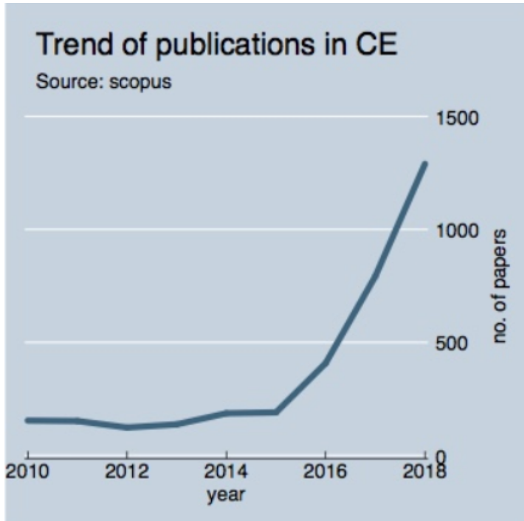

Figure 1: Trend of publication of articles about CE in Scopus (2010-2018).

Table 1: Breakdown of CE papers per research field (top 10)

Subject Area	No. of publications
Environmental Science	1'911
Engineering	1'671
Energy	964
Business, Management and Accounting	806
Social Sciences	781
Economics, Econometrics and Finance	464
Computer Science	365
Materials Science	364
Chemical Engineering	318
Earth and Planetary Sciences	287

third, developed at meso-level, relates to industrial symbiosis experiences. Another relevant sub-sector for CE is waste management [23]. As stated in the Introduction, CE dealt with the issue of Sustainability [6] as well, in the attempt to provide ways to operationalise its implementation at both, the environmental and economic level, because it might be challenging to take into account the social and institutional implications as well. For what concerns business model strategies, scholars mainly focus on studying closing material loops strategy, while slowing the loops, which would need a radical change in the production and consumption habits, is only marginally considered in CE implementation practices. [24]. On the other hand, other studies about CE acknowledge the relevance of changing the process of production and consumption and those pay attention to strategies for social change. [25].

Table 1 confirms that the scientific production about CE is split among different research fields. The top 3 is constituted by Environmental Science, Engineering and Energy, directly followed by Business, showing an increasing interest from the managerial field for circular practices. It is clear then that different research streams come from diverse epistemological fields, such as Ecology, Biology and Economics [8]. Hence, the increasing relevance of the idea of Circular Economy as a way to achieve the general purpose of "sustainable development" encouraged scholars to come up with diverse ways to understand it through different "field-specific" interpretations. Consequently, the definition of CE is not static and it includes an extensive range of principles and proposals. The richness of epistemologic nuances of CE leads critics to question its potential, holding that it lacks of elements of conceptual clarity [26]. For that reason, Circular Economy has been defined as "an idea and an ideal" (Gregson et al., 2015) [27] as well as Blomsma and Brennan (2017)[28] stress that "theoretical or paradigmatic clarity regarding the concept of CE has yet to emerge". Scholars and private actors showed a complete different attitude towards the phenomenon. On the one hand, scholars tried to overcome the lack of clarity through meticulous literature studies and comparisons. D'Amato et al. [7] performed a comparative analysis of Circular Economy, Green Economy and Bioeconomy, finding out that Bioeconomy could be understood as a part of Green Economy, confirming the results of Kleinschmit et al. (2014) [29]. Furthermore, Homrich et al. [8] in their study used a combined approach, using bibliometrics, semantic and content analysis, in the attempt to identify the main research streams of CE. Reike et al. [23] tried to overcome the confusion and the ambiguity about CE analysing the historical development of CE and the different value Retention Options (ROs) for products with an increased circularity.

On the other hand, consultancy and advocacy framed CE as a new phenomenon, envisaging a stark contrast between the previous and circular's paradigma [30, 31], spreading this idea among firms and practitioners. Moreover, Ellen MacArthur Foundation, aware of the complexity and versatility of the concept, tried to operationalize it through the ReSOLVE framework [32] as well as aiming to develop a comprehensive categorization of CE practices. Unfortunately, as demonstrated by Homrich et al. [8], the framework proved itself not effective for its purpose as only nearly 40% of the articles analysed by the authors were enclosed in one of its categories. This demonstrates again the complexity of the phenomenon and its multi-faceted nature which is worth of being further analysed.

Thus, it can be said that CE has been associated with a wide variety of concepts during the last thirty years and this multiple affiliations might have hindered the search of both, theoretical and operational clarity. Considering the large number of researches about Circular Economy, their broad spectrum of approaches and their multiple applications, a unified perspective of the keywords of CE could boost CE implementation and therefore enhance the adoption of sustainable practices.

3 Research design

3.1 Wikipedia: opportunities and threats

Wikipedia presents itself as a powerful tool for knowledge creation and interpretation; nowadays it is the largest as well as most visited encyclopedia in existence. It is densely structured and the articles proposed are rich in linkages with other articles and online pages [33]. It covers a wide range of fields such as the arts, history, geography, sports, science, music and games. Therefore, Wikipedia could be seen as a “small world”, where it takes, on average, just 4.5 clicks to move from one article to another [34]. Since it is becoming a "database storing all human knowledge", Wikipedia mining is a promising approach which integrates semantic purposes with Web 2.0 [35]. On the one hand, the pages of Wikipedia are created and checked regularly by expert users who create a large and reliable peer-to-peer community. Moreover, as a corpus for knowledge extraction, Wikipedia's peculiar characteristics are not limited to its scale, but they entail a dense link structure, disambiguation based on URL, brief link texts and well structured sentences as well. The fact that these characteristics are relevant to extract accurate knowledge from Wikipedia is strongly highlighted by a number of previous researches on Wikipedia mining, mainly focused on semantic relatedness measurements among concepts, such as Gabrilovich and Markovitch [36] and Strube and Ponzetto [37]. On the other hand, Wikipedia does not allow the user to distinguish between equivalent/synonym relations and hierarchical/hyper-hyponomy relations [33]. In addition, critics often question the reliability of the content generated in Wikipedia as well as they stress the complexity of the rules for new contributors and the presence of an “élite” of contributors enforcing these rules. Each page in Wikipedia is the product of a number of different social forces, due to corresponding people (or bots), who cause fluctuations of information, and give in every instant a visible representation for every chosen argument. [38]. The theme of “edit wars” in Wikipedia is largely treated by the literature [39, 40, 41]. An even bigger issue for this research might lie in the fact that many articles in Wikipedia are just few lines long, hence not giving complete information about the topic considered. Considering the literature on field delineation, beyond the research of Chiarello et al. [10], Wikipedia was never used to analyse, through the network created by its pages, the emergence of new, complex fields. Nonetheless, this provide an interesting opportunity to exploit the pages of Wikipedia considering they include links on a content relationship basis, hence the linkages between the pages could be considered as hint for a semantic relation, acknowledging, at the same time, the limitations of Wikipedia explained above.

3.2 Methodology

This section explains the steps made in order to analyse the network of words related to Circular Economy.

The first step consists in the generation of a seed list. The seed list was generated by extracting keywords from the results of a query with the words “Circular Economy” in Scopus. The keywords include both, those selected by the authors of the papers and those generated by Scopus itself. Then, the keywords were kept only if their occurrence was greater than 1. Afterwards, the seed list, which originally contained around 1000 words, was manually parsed, in order to choose only relevant words regarding the CE. The manual check is aimed at excluding off-topic words such as names of countries or cities (e.g. Netherlands, China, Beijing, . . .), polysemic words (e.g. deconstruction) and general concepts (e.g. flow, experiment, crisis, analysis, . . .). This resulted in a list of more than 100 words, obtained after the parsing and cleaning process, the final seed list. Moreover, 70 words from the final seed list were labeled as enabling technologies for the CE and 50 words were labeled as fields and branches of study related to the CE. These other two sub-classifications were made aiming to generate two networks, where the technologies and the fields of CE were respectively described.

The second step was related to the creation of two small dictionaries, the type keywords, useful to label each node as relevant field or technology. Around 15 words related to the concept of field and 10 words related to technology were selected. The selected keywords were chosen manually in order to include each possible synonym used in Wikipedia to describe a research field or a technology, methodology or technique. The field dictionary includes keywords such as field, model, branch, framework, system, subject while the technology dictionary includes keywords as technolog, methodolog, techniqu, analys and so on.

The final seed list was adopted as keywords' seed list for the initialization phase represented on the right side in Figure 2. Figure 2 shows the flow of the developed software for, on the right, the initialization phase and, on the left side, for the Wikipedia scraper and the network builder. Within the initialization phase, the initial node list for the network is created with a triple check on, firstly, the existence of a related Wikipedia webpage, secondly, the existence of the node and, finally, the type of the page. The first check consists in automatically building the modular Wikipedia url string such as `https://en.wikipedia.org/wiki/[FIRSTWORD]_[SECONDWORD]` and then check its existence. If more than two words appear, it is necessary to add an underscore to separate each word. The second check is the one regarding the existence of the node, to ensure that, after a possible redirect, only a unique node exists for each page. The node existence check was performed on the url of the page. Finally, the third check consists in parsing and storing the first sentence of each Wikipedia page and check if one, or more, type keywords appear. If they appear, the node is added to the node list while, if not, the node is added to the blacklist. The blacklisted nodes were labelled with an ad hoc flag. Furthermore, a check on the redirect was added, in order to guarantee the uniqueness of the added nodes due to url redirection (for instance, Biofuels is redirected to Biofuel).

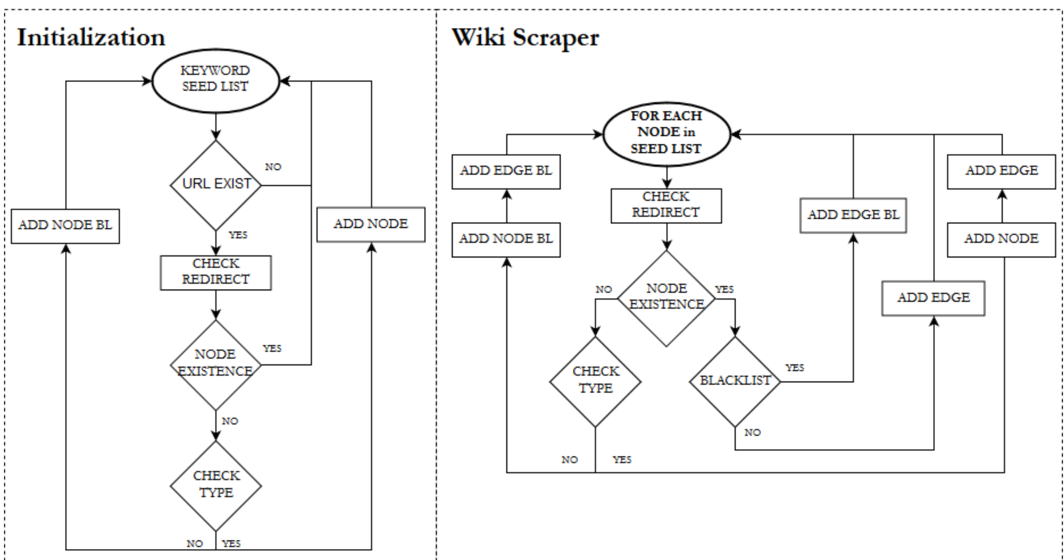

Figure 2: Flowchart of the Wikipedia scraper software developed in R.

After the initialization phase, the Wikipedia scraper module starts to parse all links (i.e. urls) from the seed list nodes. Each developed module performs a single task such as a control on url redirection, on the node existence and so on. The Wikipedia scraper performs, at each step, the same control of the initialization phase, except for the url exist control. This is explained by the fact that when the scraping algorithm starts from an existing Wikipedia page it follows only links directed to other Wikipedia pages.

Finally, a data manipulation module was developed to analyse and classify easily the dataset of Wikipedia webpages and their interconnections. The data manipulation module was necessary to label correctly the single pages and to update all nodes and edges information, such as the blacklist flag. The data manipulation module was used to create the different networks discussed in Section 4. The Social Network Analysis of the four networks was performed thanks to the software Gephi (ver. 0.9.2) analysing the Betweenness centrality [42], the Authority index [43], the PageRank index [44] and identifying the clusters thanks to the modularity, whose algorithm implemented in Gephi looks for the nodes that are more densely linked together than to the rest of the network [45]. Betweenness is an index of centrality. The intuition about centrality is that it denotes an order of importance on the vertices or edges of a graph by assigning real

Table 2: CE overall network measures

Graph name	Average degree	Modularity	Network diameter	Graph Density
All	12.75	0.351	5	0.021
Tech	8.83	0.294	5	0.55
Field	6.34	0.278	5	0.047

values to them [42]. Betweenness centrality of a node measures its ease to act as a bridge between each couple of nodes of the network. Indeed, the definition of Betweenness centrality is given by the expression

$g\left(v\right)=\sum_{s\neq v\neq t}\frac{\sigma_{st}(v)}{\sigma_{st}}$ where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths which pass through v. The PageRank index defines the importance of a node in a network considering how many links it receives weighted on the PageRank score of each neighbour. It is a measure of the likelihood of reaching a certain node starting from a random node of the graph. The PageRank index has an intrinsic limitation, i.e. it doesn't weight recursively the importance of the nodes which point to the selected node. This limitation is overcome by the Authority index. Authority centrality is one of the two generated by the algorithm HITS (Hyperlink-Induced Topic Search), together with Hubs. Authority, in particular, weights the inbound link pointing to a node depending on the importance of the source node. The importance of the source node is given by the Hub index, which depends on the number of outbound links and on the importance of the target nodes. The values of the two indices are defined recursively; therefore, a higher value for Hubs takes place if the node is pointing to many nodes with a high Authority score. In that sense, a node with a high Authority score is supposed to hold useful information because it is pointed to by many nodes with an high Hub score.

4 Results and Discussion

The results include 4 dictionaries: the first one (Figure 3) is generated using the seed list identified above, without filters; the second one (Figure 4) points out the technologies related to CE and it has been created keeping only the pages of Wikipedia having in their first sentence at least one of the words belonging to the technology dictionary. The third graph (Figure 5) shows the branches and the fields of study related to CE and it has been generated keeping the pages of Wikipedia having in their first sentence words belonging to the field dictionary. The dictionaries generated can be represented with a graph in which each node represents a word and each edge represents a link of Wikipedia. In addition, these networks allow to generate metrics, which could be used to generate useful indicators such as Betweenness, PageRank and Authority. The nodes with a degree lower than 5 are dropped out. The dimension of the label reflects the Betweenness index while the dimension of the node reflects the Degree of the node (inbound plus outbound links). The colours represent the Modularity.

The first graph (Figure 3) shows the network of words generated from the seed list without using any filter in Wikipedia. This provides an overall image of the adopted dictionary, the final seed list, related to the CE. The modularity algorithm, with resolution 1.0, [45] identifies 6 main clusters, represented by the different colours in Figure 3. Light green represents the cluster related to Sustainability, Sustainable Development, Environmental Economics and Ecology. It aggregates previous broad and general disciplines related to the environment. The orange cluster aggregates the Circular Economy, Life-cycle Assessment, Material Flow Analysis. The blue one is related to waste management and extended producer responsibility. The purple cluster shows a focus on water reuse, Anaerobic Digestion and waste water. The red cluster is focused on bioenergy, Biofuels, biogas and Bioeconomy. Finally, between the purple and the red one a small yellow cluster represents the WEEE, the rare-earth elements and other metals. Thus, the six identified clusters are: 1) Sustainability, 2) Material Flow Analysis, 3) Waste Management, 4) Water Management, 5) Waste Electrical and Electronic Equipment and 6) Bioeconomy. Other two networks have been analysed: the technology network and the field one. The two networks were obtained by selecting only the Wikipedia pages which contain in the first sentence one, or more, of the words representing the types as described in Section 3. Table 2 summarizes the global network parameters for each graph. The average degree, the modularity, the network diameter and the graph density are reported. As shown, the overall network is the least dense graph but it has the highest average degree. The most dense graph is the technology graph while the field one is the network with the lower average degree.

Figure 4 exhibits the most relevant technologies related to the Circular Economy, while Figure 5 shows the related fields. From the technology network four of the six clusters described in the overall network can be recognized. The Sustainability cluster (the green one), the Waste Management cluster (the blue one), the WEEE cluster (the yellow one) and the Bioeconomy cluster (the red one). The orange and the purple ones were in part adsorbed and aggregate to the other clusters and in part they are not considered because of the type filter. Finally, Figure 3 shows the related fields.

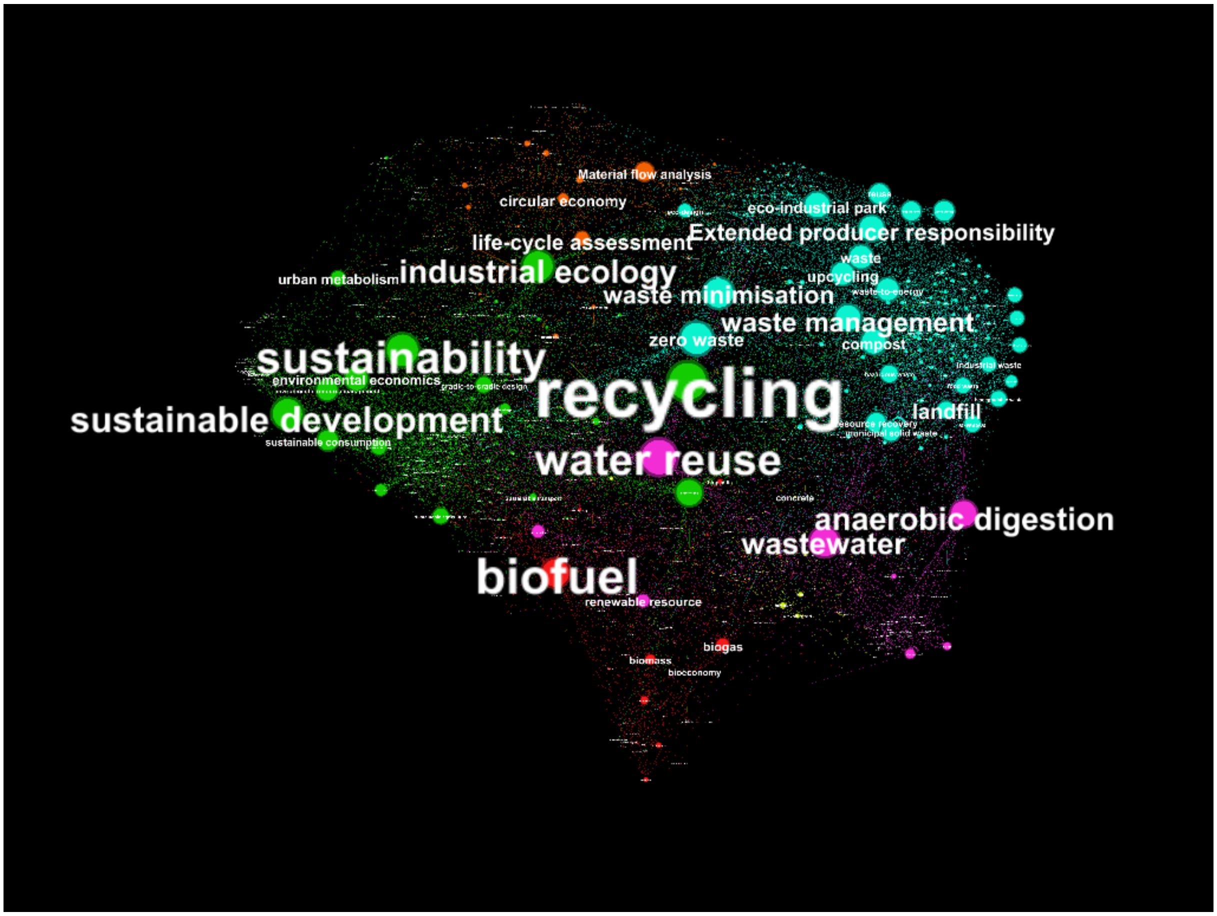

Figure 3: CE overall network.

Two of the six clusters can be identified: the Sustainability cluster (the green one) and the Material Flow Analysis cluster (the orange one). The violet cluster, instead, is a mix of fields from different clusters of Figure 3 and it aggregates keywords such as urban metabolism, Industrial Ecology and Ecology.

Table 3, 4 and 5 show the top five nodes with respect to three centrality indices: Betweenness, the PageRank and the Authority, respectively. Betweenness centrality analysis shows that fundamental nodes for the CE are Recycling, Biofuel and Sustainability. This is not at all a surprising result, because the idea of Recycling is strongly linked to the CE concept through the 3R imperative " Reduce, reuse and recycle". According to Reike et al. [23], there are three research articles, from 2010 onward, which acknowledge the connection between CE and the 3R imperative. On the other hand, this paradigm was already existent and mentioned in other 7 papers about Industrial Ecology and 19 about Reverse Logistics and Closed-Loop Supply Chain Management. Furthermore, Murray et al. [25] underline that the urge for Recycling started many years before the upsurge of CE, encouraged by consistent policy interventions in different countries, such as *Basic Law for Establishing a Sound Material-cycle Society* in Japan and *The Waste Avoidance and Management Act* in Germany, both enacted in 2002. This entails that the field of CE could not be originally and directly associated with the 3R paradigma, including therefore the practice of Recycling [23]. Biofuel, a fundamental keyword for the dictionary of CE, dates back its production to the Seventies and it experienced a steady growth since then [46]. Comprehensive bibliometric reviews about CE [6, 23, 47] do not mention Biofuel at all. Moreover, D'Amato et al. [7] argue Biofuel is a concept related to the idea of Bioeconomy but not associated with Green and Circular Economy. In that sense, even if it exhibits a high Betweenness, Biofuel seems an older topic than CE itself, whose importance has been growing over the last decade driven by governmental policies more than the emergence of the idea of CE [48]. The concept of Sustainability has its roots in the idea of Sustainable Development[49] and it seems more "open ended" than the concept of CE [50]. Sustainability as a topic is experiencing a clear growth trend, considering the query of Sustainability in Scopus generates 9'142 results for 2010 and 22'250 in 2018, with an increase of 143 %. The upwards trend is real but lower than the one of CE. In addition, Recycling and Sustainability were included among the most important words for number of occurences in selected articles about CE in Geissodoerfer et al.[6], confirming

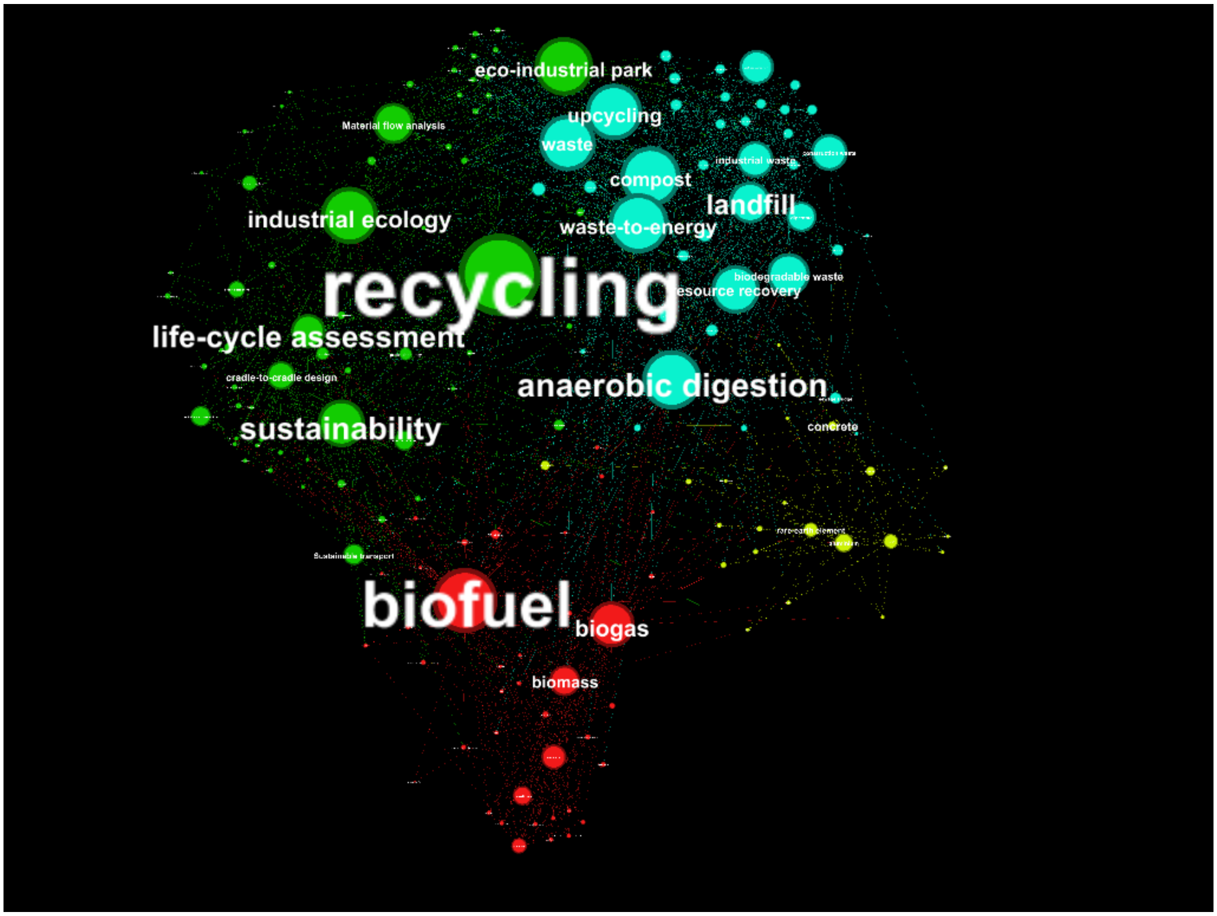

Figure 4: CE technologies network.

the validity of the methodology developed in this paper. For what concerns the main keywords for the fields related to CE, it is no surprise to find Industrial Ecology with the highest value of Betweenness (0.036), because, as stated in the Introduction, CE emerged as a separate field from Industrial Ecology itself. On the other hand, Cradle-to-Cradle design (C2C), defined also as regenerative design, occupies the 4th position for Betweenness and this is an interesting result because of, as argued by Reike et al. [23], CE only very recently started to be associated with this idea. The result obtained could then entail that CE might be pushing towards the emergence of Cradle-to-Cradle, linking it to other fields of expertise. The CE technology network demonstrates again the importance of Recycling and Biofuel, but giving more detail about the key processes and technologies for CE. Anaerobic Digestion (third position in terms of Betweenness in the network) is a fermentation process where organic raw materials such as food waste, sewage sludge and other industrial wastes are converted into biogas [51]. Due to its close connection with waste, from a cluster point of view, Anaerobic Digestion is closer to the "waste" cluster. According to Fagerström et al. (2019) [52], the process of Anaerobic Digestion is directly linked to the production of biogas, therefore to the concept of Bioeconomy. In addition, in their report the authors demonstrate that Anaerobic Digestion and the production of biogas is closely related to CE through four case studies. To sum, the most influencial nodes of the three networks, defined by Betweenness index, might not be considered per se as elements of novelty directly associated to the idea of CE. On the other hand, the presence of concepts such as C2C, for instance, might be a sign of cross-fertilization among diverse concepts, glued and enhanced by the growing paradigma of CE.

The PageRank top values are consistent with the Betweenness centrality ones. Recycling occupies again the first position in the overall network and in the technology one. Other important words detected by the PageRank algorithm are Sustainability, Biofuel and LCA, already identified by Betweenness. On the other hand, Permaculture, first ranked by the PageRank algorithm for the field is an interesting finding. The goal of Permaculture is to manage the urbanized ecosystem allowing the satisfaction of population's needs, while preserving the stability of natural ecosystem [53]. Permaculture could then be defined as a mixture of fields, such as architecture, biology, silviculture and zootechnics. The captivating fact about Permaculture is that it seems to have a different nuance compared to Sustainability: Permaculture

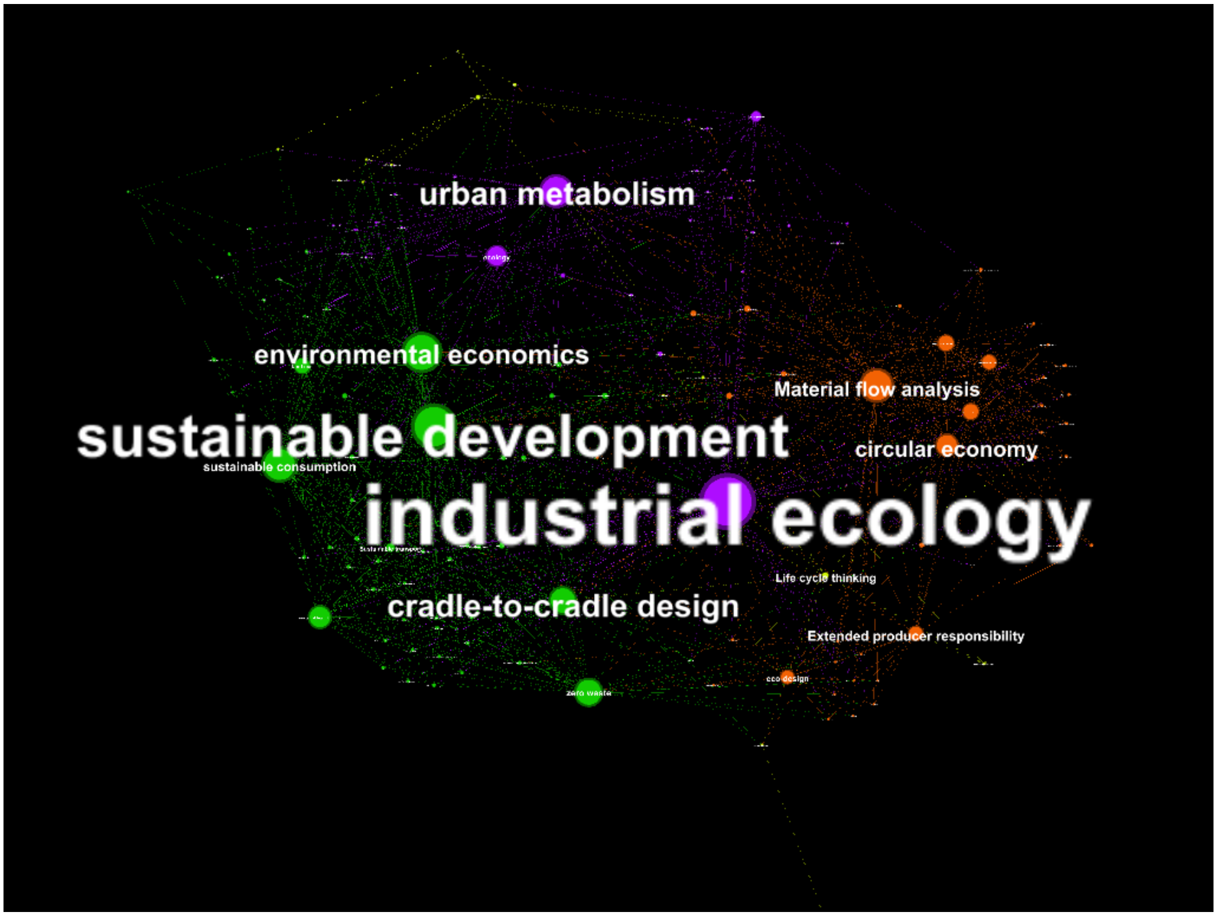

Figure 5: CE field network.

Table 3: Betweenness, top 5

	All		Field		Tech	
Rank	Keyword	Value	Keyword	Value	Keyword	Value
1	Recycling	0.014	Industrial Ecology	0.036	Recycling	0.460
2	Biofuel	0.011	Sustainable Development	0.025	Biofuel	0.358
3	Sustainability	0.009	Urban metabolism	0.013	Anaerobic Digestion	0.018
4	Water reuse	0.009	Cradle-to-cradle design	0.013	Sustainability	0.017
5	Sustainable Development	0.007	Environmental economics	0.011	Life-cycle assessment	0.016

has a "value-added factor which extends beyond what might be merely maintained or sustained" [53]. Thus, while Sustainability aims to maintain what already exists, Permaculture's purpose is to be regenerative. For instance, a regenerative product must not only be 100% recycled but it also has to improve environmental conditions at any stage of its production and use. Since there is no waste in nature, this idea of regeneration seem to upgrade the concept of Sustainability towards a real Circular Economy.

The highest values for the Authority scores are linked to the idea of waste (waste management, waste hierarchy, waste minimization etc). This result emerged from the structure of the analysed network and it can be noticed, in a qualitative way, by observing the densities of links and the number of high degree nodes of the Waste Management cluster in Figure 3. Waste has been the very first topic faced, in the 90s, by the CE but the interest for waste management started in the 70s, confirmed by a steep increase of the topic in the literature of Industrial Ecology and cleaner production.

Table 4: PageRank, top 5

	All		Field		Tech	
Rank	Keyword	Value	Keyword	Value	Keyword	Value
1	Recycling	0.003	Permaculture	0.013	Recycling	0.0102
2	Regenerative Design	0.0028	Regenerative Design	0.012	Biofuel	0.0095
3	Greenhouse gases	0.0024	Cradle-to-cradle design	0.010	Life-cycle assessment	0.0093
4	Circular Economy	0.0024	Sustainable Development	0.0099	Sustainability	0.0088
5	Sustainability	0.0024	Infrastructure	0.0096	Efficient energy use	0.0087

Table 5: Authority, top 5

	All		Field		Tech	
Rank	Keyword	Value	Keyword	Value	Keyword	Value
1	Waste Management	0.109	Sustainable Development	0.161	Recycling	0.165
2	Recycling	0.104	Rebound Effect	0.150	Waste hierarchy	0.163
3	Waste hierarchy	0.099	Industrial Ecology	0.147	Heat waste	0.158
4	Sewage treatment	0.099	Anthropogenic	0.143	Downcycling	0.157
5	Waste Minimization	0.098	Natural Resource	0.143	Scrap	0.154

5 Limitations and further research

The discussed methodology showed how, with a general process, fields and technologies related to a precise field, in this research to the Circular Economy, can be extracted, identified and analysed thanks to a Social Network Analysis approach with a non-intensive, computationally speaking, algorithm. The main potentiality of this methodology lies on the analysis of Wikipedia, which, currently, is the greatest and most accurate encyclopedia in the world. The modular layout of each page allows easily to extract information thanks to a data scraping algorithm and a simple bot. Moreover, the “small world” structure permits to easily explore the whole network. In this work, the Social Network Analysis, supported by a text analysis, has been proven to be an excellent approach to identify and extract relevant concepts related to the Circular Economy. The same approach can be easily generalized to several other fields such as Artificial Intelligence, Industry 4.0, Sustainable Development Goals, Green Chemistry and so on.

This research work is a first step towards the creation of a robust, systematic and scalable methodology for network analysis using Wikipedia. Nevertheless, further improvements need to be implemented: in first place, the seed list creation process has to be empowered in order to obtain an automatic and reproducible method, that could then be applied to a diverse set of research fields. The manual nature of this activity deeply influenced this work’s results. Indeed, despite its potentiality, this methodology revealed several criticisms and limitations which must be further investigated. The obtained final networks, i.e. the field and the technology ones, are strongly affected by the choices made during the selection of the final seed list of keywords, as well as they are sensitive to parameters, such as the degree filter, the modularity resolution or the definition of the field and technology dictionaries. When varying anyone of these setting parameters, the final results are considerably affected.

First of all, considering the final seed list of keywords, a more general algorithmic approach has to be developed. Indeed, almost all the most important and crucial nodes revealed by the Social Network Analysis belong to the initial seed list because the graph analysed corresponds to the nearest neighbourhood of the initial list. A possible solution could reside in improving the seed list creation process, aiming at an automatic expert-independent approach, providing a reproducible methodology; this goal might be obtained by filtering the initial keywords list retrieved from Scopus by a pre-fixed occurrence threshold or by fixing a criterium in the paper selection activity in first place (i.e. filtering by top-cited papers).

Secondly, the "type" dictionaries used to “filter” the fields or the technologies strongly affect the final graph. Adding, or removing, a single word might lead to a completely unpredictable result. Although the type keyword approach used in this work is a great starting point towards the automatization of the dictionary creation process, it needs to be refined.

Thirdly, the degree filter also has a noteworthy effect on the final graph. Without the degree filter, i.e. considering also nodes with only one link, an higher number of clusters were obtained but the centrality indices were strongly affected by hundreds of nodes with no importance at all for the Circular Economy. Thus, the choice of filter nodes with a degree lower than five was necessary to drop out not strongly interconnected nodes in order to minimize biases in subsequent analyses of centrality degrees due to off topic nodes. In addition, the degree filter allows to minimize biases due to the total number of links of a Wikipedia page, which is not an interesting information and it can strongly affect centrality results simply because of some Wikipedia pages can have more than one thousand links to other pages. On the contrary, analysing the network without the degree filter allowed to identified more precise and bounded clusters with respect to the discussed networks. This aspect can be partially overcome by varying the resolution parameter of the modularity. Indeed, values lower than one allow to identify smaller communities.

Finally, the networks obtained from Wikipedia are user-generated and some research fields or knowledge might not appear at all or they might be under-evaluated. Newest, most innovative and recent research fields may be under-represented or may be penalized by the directionality of links, if the directed network is considered. For instance, a newer field cites an older one, e.g. Circular Economy points to Biomimicry, but not viceversa.

Future studies and investigations are necessary to solve some of the discussed limits. For instance, to explore deeper the Wikipedia network, stronger and more powerful type dictionaries are necessary in order to avoid the polynomial explosion of the number of nodes and edges. Indeed, the number of nodes increases by, at least, two orders of magnitude since each page has on average one hundred links and some Wikipedia pages can have more than one thousand links. It's trivial to show how at each step of the Wikipedia scraper algorithm, i.e. nearest neighbours, 2nd order neighbourhood, the number of nodes increases according to the following rule $n_{t+1} \sim n_t^2$, where n_{t+1} is the number of nodes at the step t+1 and n_t at step t. In order to avoid this explosion, a blacklist or a stop list of keywords can be implemented. On the contrary, to contain the network explosion, only pages which contain certain crucial keywords can be considered. For instance, considering this work, pages which do not contain words such as Sustainability, waste, environment, or which do not contain couple or triplet of words, could be excluded. Otherwise, a ranking system, based on keywords, can be implemented by assigning a score to each page and selecting only pages with a score greater than a determined thresholds. Score, for instance, can be based on machine learning algorithms which assign to each word a weight. Furthermore a topic modeling analysis on single pages can be implemented in order to compute the "distance" between pages and the initial seed list or the nearest neighbours in order to drop out too far away Wikipedia pages. Otherwise, an algorithm based on the network can be developed in order to exclude nodes, or entire branches, which do not connect with the initial seed list after two or more steps. Finally, further analyses and considerations could be done by comparing the clusters obtained implementing the methodology presented in this work (bottom-up) and the clusters that would be expected (top-down). In addition, it would also be of great interest, once the above-mentioned improvements have been applied and a more accurate technology dictionary with reference to the CE has been created, performing a crossover with Industry 4.0, Artificial Intelligence, Sustainable Development and so on.

6 Conclusion

Circular Economy is an emerging field and in recent years the number of publications related to the CE is rapidly growing. Despite the great interest in circular practices, innovations and applications from both researchers and practitioners, blurriness in the CE definition and in identifying precise boundaries is still occurring. In this paper, an innovative methodology has been set up and discussed in order to identify relevant fields and technologies related to the CE, their interconnections and the relation of the CE with existing fields. The methodology presented is based on analysing the network of pages of Wikipedia starting from a dictionary, a seed list, of relevant keywords extracted from the Scopus database related to the Circular Economy. The network, then, has been analysed through to two other lists of words, one for the technologies and one for the fields, hence two sub-networks were built. To obtain the two sub-networks, a filter was applied to select only pages containing words such as "field", "technolog", "treatment" in the first sentence of the Wikipedia page. The graphs selected, finally, were analysed in terms of centrality degrees (Betweenness, PageRank, Authority) and modularity. Six main clusters emerged: 1) Sustainability, 2) Material Flow Analysis, 3) Waste Management, 4) Water Management, 5) WEEE, 6) Bioeconomy. The most crucial nodes are Wikipedia pages such as Recycling, Biofuel, Anaerobic Digestion, Sustainability, Industrial Ecology or Water Reuse.

The presented methodology shows how Wikipedia can be used as a "small world" and it allows to analyse, in detail, a precise field extracting useful information for researchers, practitioners and policy-makers. Despite its powerful potentiality further investigations are needed in order to better tune the sub-networks identification and to avoid possible biases in centrality degree analysis.

Is the Circular Economy a new paradigm or is it a relabeling of existing knowledge? From the analysis of the obtained graphs Circular Economy results to be a quite important node and it has a dedicated cluster, with existing fields such as Material Flow Analysis and life-cycle assessment, for instance, but it does not seems nor to connect far-away fields neither to open a new emerging branch. Moreover, despite having its own cluster, Betweenness centrality demonstrated that the most influencial keywords, such as Recycling, Biofuel and Sustainability, at the present time, were not originally conceived within the framework of the CE.

On the contrary, analysing the field network, Industrial Ecology, Sustainable Development as well as urban metabolism have an important role in terms of Betweenness. Urban metabolism and Industrial Ecology seems to belong to a new cluster of fields connecting, on one side, the CE cluster and, on the other side, the Sustainability cluster. The presence in the network of fields only recently associated with CE, such as C2C, might be interpreted as an interesting signal of cross-fertilization promoted by the CE paradigma. The emergence of Permaculture as first-ranked result by the Pagerank algorithm, could be interpreted as a further sign of this process.

References

[1] Johan Rockström, Will L Steffen, Kevin Noone, Åsa Persson, F Stuart Chapin III, Eric Lambin, Timothy M Lenton, Marten Scheffer, Carl Folke, Hans Joachim Schellnhuber, et al. Planetary boundaries: exploring the safe operating space for humanity. *Ecology and society*, 2009.

[2] Tim Jackson. *Prosperity without growth: Economics for a finite planet*. Routledge, 2009.

[3] Dennis Meadows and Jorgan Randers. *The limits to growth: the 30-year update*. Routledge, 2012.

[4] Patrizia Ghisellini, Catia Cialani, and Sergio Ulgiati. A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. *Journal of Cleaner production*, 114:11–32, 2016.

[5] M Tennant, G Brennan, and F Blomsma. Business and production solutions: Closing the loop. *Sustainability: Key issues. New York: Routledge*, 2015.

[6] Martin Geissdoerfer, Paulo Savaget, Nancy MP Bocken, and Erik Jan Hultink. The circular economy–a new sustainability paradigm? *Journal of cleaner production*, 143:757–768, 2017.

[7] Dalia D'Amato, N Droste, B Allen, M Kettunen, K Lähtinen, J Korhonen, Pekka Leskinen, BDf Matthies, and A Toppinen. Green, circular, bio economy: A comparative analysis of sustainability avenues. *Journal of Cleaner Production*, 168:716–734, 2017.

[8] Aline Sacchi Homrich, Graziela Galvao, Lorena Gamboa Abadia, and Marly M Carvalho. The circular economy umbrella: Trends and gaps on integrating pathways. *Journal of Cleaner Production*, 175:525–543, 2018.

[9] Roberto Merli, Michele Preziosi, and Alessia Acampora. How do scholars approach the circular economy? a systematic literature review. *Journal of Cleaner Production*, 178:703–722, 2018.

[10] Filippo Chiarello, Leonello Trivelli, Andrea Bonaccorsi, and Gualtiero Fantoni. Extracting and mapping industry 4.0 technologies using wikipedia. *Computers in Industry*, 100:244–257, 2018.

[11] David W Pearce and R Kerry Turner. *Economics of natural resources and the environment*. JHU Press, 1990.

[12] Mikael Skou Andersen. An introductory note on the environmental economics of the circular economy. *Sustainability Science*, 2(1):133–140, 2007.

[13] Kenneth Boulding. E., 1966, the economics of the coming spaceship earth. *New York*, 1966.

[14] W Stahel. Jobs for tomorrow: The potential for substituting manpower for energy, report to the commission of the european communities (now european commission), brussels, 1976.

[15] Walter R Stahel. The product life factor. *An Inquiry into the Nature of Sustainable Societies: The Role of the Private Sector (Series: 1982 Mitchell Prize Papers), NARC*, 1982.

[16] William McDonough and Michael Braungart. *Cradle to cradle: Remaking the way we make things*. North point press, 2010.

[17] Barry Commoner. *The closing circle: nature, man, and technology*. Knopf, 2014.

[18] John Tillman Lyle. *Regenerative design for sustainable development*. John Wiley & Sons, 1996.

[19] TE Graedel and Braden R Allenby. *Industrial Ecology and Sustainable Engineering: International Edition*. Pearson Education Inc., Upper Saddle River, Prentice Hall, 2010.

[20] Walter Stahel. *The performance economy*. Springer, 2010.

[21] Janine M Benyus. Biomimicry: Innovation inspired by nature, 1997.

[22] Gunter A Pauli. *The blue economy: 10 years, 100 innovations, 100 million jobs*. Paradigm publications, 2010.

[23] Julian Kirchherr, Denise Reike, and Marko Hekkert. Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling*, 127:221–232, 2017.

[24] Nancy MP Bocken, Ingrid de Pauw, Conny Bakker, and Bram van der Grinten. Product design and business model strategies for a circular economy. *Journal of Industrial and Production Engineering*, 33(5):308–320, 2016.

[25] Alan Murray, Keith Skene, and Kathryn Haynes. The circular economy: An interdisciplinary exploration of the concept and application in a global context. *Journal of Business Ethics*, 140(3):369–380, 2017.

[26] Michael Lieder and Amir Rashid. Towards circular economy implementation: a comprehensive review in context of manufacturing industry. *Journal of cleaner production*, 115:36–51, 2016.

[27] Nicky Gregson, Mike Crang, Sara Fuller, and Helen Holmes. Interrogating the circular economy: the moral economy of resource recovery in the eu. *Economy and Society*, 44(2):218–243, 2015.

[28] Fenna Blomsma and Geraldine Brennan. The emergence of circular economy: A new framing around prolonging resource productivity. *Journal of Industrial Ecology*, 21(3):603–614, 2017.

[29] Daniela Kleinschmit, Berit Hauger Lindstad, Bo Jellesmark Thorsen, Anne Toppinen, Anders Roos, and Sjur Baardsen. Shades of green: a social scientific view on bioeconomy in the forest sector. *Scandinavian journal of forest research*, 29(4):402–410, 2014.

[30] Accenture circle economy, 2016.

[31] Ellen MacArthur. Towards the circular economy. *Journal of Industrial Ecology*, 2:23–44, 2013.

[32] Ellen MacArthur Foundation. Delivering the circular economy: A toolkit for policymakers, 2015.

[33] David Milne and Ian H Witten. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 509–518. ACM, 2008.

[34] Stephen Dolan. Six degrees of wikipedia. *Retrieved June*, 2008.

[35] Kotaro Nakayama, Takahiro Hara, and Shojiro Nishio. Wikipedia link structure and text mining for semantic relation extraction. In *SemSearch*, pages 59–73, 2008.

[36] Evgeniy Gabrilovich, Shaul Markovitch, et al. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJcAI*, volume 7, pages 1606–1611, 2007.

[37] Simone Paolo Ponzetto and Michael Strube. Knowledge derived from wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, 30:181–212, 2007.

[38] Massimo Marchiori and Enrico Bonetti Vieno. Negapedia: The negative side of wikipedia. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1–4. IEEE, 2018.

[39] Róbert Sumi, Taha Yasseri, et al. Edit wars in wikipedia. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 724–727. IEEE, 2011.

[40] Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. Dynamics of conflicts in wikipedia. *PloS one*, 7(6):e38869, 2012.

[41] Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. Societal controversies in wikipedia articles. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 193–196. ACM, 2015.

[42] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

[43] Jon M Kleinberg. Hubs, authorities, and communities. *ACM computing surveys (CSUR)*, 31(4es):5, 1999.

[44] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[45] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[46] Harvey Lapan and GianCarlo Moschini. Second-best biofuel policies and the welfare effects of quantity mandates and subsidies. *Journal of Environmental Economics and Management*, 63(2):224–241, 2012.

[47] Yuliya Kalmykova, Madumita Sadagopan, and Leonardo Rosado. Circular economy–from review of theories and practices to development of implementation tools. *Resources, Conservation and Recycling*, 135:190–201, 2018.

[48] Giovanni Sorda, Martin Banse, and Claudia Kemfert. An overview of biofuel policies across the world. *Energy policy*, 38(11):6977–6988, 2010.

[49] RH Cassen. Our common future: Report of the world commission on environment and development., 1987.

[50] Zengwei Yuan, Jun Bi, and Yuichi Moriguichi. The circular economy: A new development strategy in china. *Journal of Industrial Ecology*, 10(1-2):4–8, 2006.

[51] Jens Bo Holm-Nielsen, Teodorita Al Seadi, and Piotr Oleskowicz-Popiel. The future of anaerobic digestion and biogas utilization. *Bioresource technology*, 100(22):5478–5484, 2009.

[52] Anton Fagerström, Teodorita Al Seadi, Saija Rasi, Tormod Briseid, et al. *The role of Anaerobic Digestion and Biogas in the Circular Economy*. IEA Bioenergy, 2019.

[53] Christopher J Rhodes. Permaculture: Regenerative–not merely sustainable. *Science progress*, 98(4):403–412, 2015.