Mapping the market for remanufacturing: An application of “Big Data” analytics

DOI:
10.1016/j.ijpe.2020.107807

Mapping the market for remanufacturing: An application of "Big Data" analytics

Abstract

Remanufacturing is one of the most examined topics in the closed-loop supply chain (CLSC) literature. However, we still have limited knowledge on the characteristics of the market for remanufactured products. This study addresses this gap by using a big data analytics framework. We employ off-the-shelf, pre-trained vectors created with the Global Vectors for Word Representation (GloVe) word embedding method from a data set crawled from the Internet. The Louvain method subsequently provides us with clusters based on remanufacturing and related terms, without requiring human interactions. Our findings provide the following main insights. First, remanufacturing and related terms are associated with specific industries and products, among which printing equipment, automobiles and car parts, treadmills, consumer electronics, and household appliances. Among the terms capturing remanufacturing activity, *remanufactured*, *reconditioned*, and *rebuilt* are strongly associated with business-to-business and slow clockspeed products, while *refurbished* is mostly associated with business-to-consumer and fast clockspeed products. Second, original equipment manufacturers (OEMs) are much more salient than independent remanufacturers, and Japanese OEMs are especially well represented as players in the market for remanufacturing. Third, environmental concerns only appear weakly in the discourse surrounding product recovery, while consumers do seem to place emphasis on quality and price. In a final part of the study, we contrast the CLSC academic literature with the clusters obtained through our big data analysis, thereby identifying industries, products, and brands that are understudied. We also outline the practical implications of our work for managers involved in setting up a remanufacturing strategy, as well as regulators.

Keywords: REMANUFACTURING, REVERSE LOGISTICS, BIG DATA and CLOSED-LOOP SUPPLY CHAIN

1. Introduction

Remanufacturing recovers value from used products by replacing components or reprocessing used parts to bring the product to like-new condition (Atasu et al., 2008). Within the literature on closed-loop supply chains (CLSC), a large number of studies examine the operational, tactical, and strategic decisions associated with remanufacturing (Galbreth & Blackburn, 2010; Goodall et al., 2014; Jeihoonian et al., 2017; Samarghandi, 2017; Abbey & Guide Jr, 2018; Pazoki & Samarghandi, 2020). While these studies take a corporate (supply) perspective, a growing stream of papers examines remanufacturing from a consumer (demand) perspective, by analysing the drivers of the willingness to pay (WTP) or the size of the market for remanufactured products (Guide Jr & Li, 2010; Subramanian & Subramanyam, 2012; Quariguasi Frota Neto et al., 2016; Jakowczyk et al., 2017).

Despite the substantial academic coverage of remanufacturing in the CLSC literature, we still know relatively little about the basic characteristics of this multi-billion dollar activity. Back in 1996, Lund called the remanufacturing industry a hidden giant (Lund, 1996). Reinforcing this point, Linton (2008) argued that remanufacturing is in some respects an invisible activity, since it is not recognized as an industry by standard industry classification systems. More recently, the CEO of the Ellen MacArthur Foundation argued during his speech at the 2019 Annual World Remanufacturing Conference: “Nobody knows what reman is” (Source: `https://bit.ly/34rlwo6`).

In this paper, we aim to address the substantial gap in knowledge regarding the main characteristics of the market for remanufacturing products. We study three research questions: First, what are the key industries and products represented in the market for remanufactured products? Second, who are the most prominent players, i.e. firms and brands, represented in this market? And third, what are the most salient attributes associated with remanufactured products?

One option to examine the above questions would be to collate information from reports, newspaper articles, and academic papers, among others, and summarize these. However, this approach would be extremely time-consuming and potentially subjective, since the researcher would have to make an a priori selection of the material to review. We adopt an alternative framework based on big data analytics, allowing us to obtain a computationally inexpensive overview of the market for remanufactured products without relying on human interactions in the data collection.

More particularly, we use a novel approach that combines Natural Language Processing (NLP) and community discovery techniques. NLP, a powerful tool that attempts to make sense of human communication, has been successfully deployed in fields as diverse as bio-diversity science (Thessen et al., 2012), journalism (Cohen et al., 2011), law enforcement (Voigta et al., 2017), medicine (Hripcsak et al., 1995), and psychotherapy (Althoff et al., 2016). Community discovery, broadly speaking a mechanism for identifying individuals who exhibit similar preferences or speak the same language, in turn, has been applied in criminology, telecommunications, and sociology, among other disciplines (Ferrara et al., 2012, 2014).

To our knowledge, we are the first to adopt this approach for mapping the characteristics of the market for remanufactured products. In a first step of our analysis, we obtain off-the-shelf, pre-trained vectors from spaCy, an open-source software library for advanced natural language processing (Source: `https://bit.ly/2vorTeZ`). The vectors are created with Global Vectors for Word Representation (GloVe), one of the most well-known word embedding methods (Pennington et al., 2014), and are based on a data set obtained from Common Crawl, a nonprofit organization dedicated to providing a copy of the Internet (Source: `https://bit.ly/2xJegYJ`). In a second step, we apply the state-of-the-art Louvain method for community detection (Blondel et al., 2008), thus obtaining a series of clusters related to remanufacturing, without requiring further human input.

An analysis of these clusters provides us with the following insights regarding our three main research questions. First, with regards to the most popular industries and products linked with remanufacturing and related terms, we find evidence of strong associations with printing equipment, automobiles and car parts, treadmills, consumer electronics, and household appliances. Interestingly, among the terms capturing remanufacturing activity, *remanufactured*, *reconditioned*, and *rebuilt* are strongly associated with business-to-business (B2B), slow clockspeed products, while *refurbished* is closely linked with business-to-consumer (B2C), fast clockspeed products. The latter pattern suggests that refurbishing should not be considered a synonym for remanufacturing. Second, with regards to the key players in the market for remanufacturing, we find that original equipment manufacturers (OEMs) are much more salient than independent remanufacturers. Japanese OEMs are especially well represented, while Chinese OEMs have no salience. In a robustness test, we find that those OEMs identified by our cluster analysis as having high semantic similarity with remanufacturing tend to be effectively engaged in product remanufacturing. Third, in terms of the attributes most commonly associated with remanufacturing, the clusters suggest that environmental considerations are second to quality and price in the eyes of consumers.

Although our study is exploratory in nature and therefore does not depart from a theoretical framework, we refer to relevant theoretical concepts of isomorphism (DiMaggio & Powell, 2000), clockspeed (Fine, 2000), and resource-constrained innovation (Weiss et al., 2011) when interpreting our findings. We also position our results relative to those of related other empirical studies on remanufacturing.

Our paper has two main implications for academic research. First, our findings can serve as a point of departure for the design of future empirical academic studies on remanufacturing. More particularly, our cluster analysis may help researchers identify industries and products that are currently most closely associated to CLSC and thus warrant further research, as well as some of the industries and products that have been overlooked by previous research. This could result in more cross-sectional variation and therefore increased power in empirical tests, by enhancing the diversity of industries and products under consideration. Our findings could also inform researchers on the potential main players in each of the remanufacturing markets, which could be relevant to those interested in intra-firm case studies. Moreover, our cluster analysis may help researchers identify a list of suitable keywords associated with remanufacturing, which could serve as a useful sample collection tool for empirical studies on characteristics of remanufactured products (see, e.g., Subramanian & Subramanyam (2012)). Second, our findings on the differences between remanufacturing and refurbishing, as well as on the attributes commonly associated with remanufactured products, contribute to ongoing academic debates about these so far unresolved issues (Thierry et al., 1995; Abbey et al., 2015a,b).

Practitioners and regulators, in turn, also stand to benefit from a deeper knowledge into the characteristics of the market for remanufacturing. To give an example, OEMs not yet engaged in product remanufacturing may benefit from being able to gauge the popularity of remanufactured product types, a proxy for cannibalisation potential, which is an issue of increasing importance

(Guide Jr & Li, 2010). Firms faced with the prospect of developing a remanufacturing marketing strategy might also find it useful to know that green credentials seem to be less important than price and quality cues for these products. Finally, regulators might find our work valuable for remanufacturing policy setting purposes, as it provides them with a birds -eye overview of the main characteristics of the market. More generally, our work showcases the ability of big data analytics to generate crucial insights into CLSC market dynamics in a fast, efficient, and objective way.

The remainder of this paper is structured as follows. In the next section, we briefly outline the main strands of literature relevant to our work. Section 3 outlines the big data analytics methodology used in our study. Section 4 describes the clusters generated by the framework. Section 5 documents the main insights related to our three research questions that can be derived from these clusters. Section 6 discusses the implications of our work for academics and practitioners. Section 7 discusses limitations and avenues for future research.

2. Literature review

Our work sits at the intersection of three relevant streams of literature. Topic-wise, it is positioned within the literature on the market for remanufactured products. More broadly, it is also related to studies tapping the Internet for data, as well as studies using the NLP technique for data analysis. We now briefly describe the most relevant papers in each of these three streams and outline the incremental contributions of our study.

2.1. Understanding the market for remanufactured products

The majority of empirical studies on the marketing of remanufactured products focus on consumer perceptions on, and willingness to pay (WTP) for, these products. A large group of papers within this stream of research are based on primary data obtained from surveys or experiments. Linton (2008) surveyed college students on their WTP for remanufactured products, and introduced a procedure to evaluate the economic rationality for remanufacturing certain products. WTP was also the focus of Hazen et al. (2012), who examined its link with consumers' tolerance to ambiguity by means of an SERVQUAL survey. In Jiménez-Parra et al. (2014), questionnaires helped to identify the behavior of potential consumers of remanufactured laptops in Spain. Other scholars similarly engaged include Hamzaoui-Essoussi & Linton (2014), who surveyed consumers to study the impact of product category, perceived risk, and brand name on WTP, Abbey et al. (2015a), who conducted experimental studies in which consumers responded to product descriptions that manipulated price discount and brand equity, Harms & Linton (2016), who examined the effect of eco-certification on refurbished mobile phones, Wang & Hazen (2016), whose finding of an intrinsic link between risk perception and value and purchase intention aligns with prior research, Matsumoto et al. (2017), who used surveys to compare Japanese and US consumers' perceptions of remanufactured auto parts, and Mugge et al. (2017), who researched the effect of incentives for improving the attractiveness of refurbished smartphones. Other studies have used interviews, case studies, as well as multi-methodological approaches for examining remanufacturing market dynamics. For example, van Weelden et al. (2016) used interviews to map the factors that influence Dutch consumers' purchase of remanufactured mobile phones, Hazen et al. (2017) employed case studies and surveys in a rigorous examination of multiple dimensions of product quality, and Abbey et al. (2017) used surveys and experiments to examine the antecedents of both purchase intent and WTP for remanufactured products. In addition, a limited number of studies have gauged the drivers of the WTP for remanufactured products with the help of secondary data collected from eBay. Examples include Guide Jr & Li (2010), Subramanian & Subramanyam (2012), Quariguasi Frota Neto et al. (2016) and Xu et al. (2017).

While the above studies have WTP for remanufactured products as their key focus, a small number of papers have empirically explored other dimensions of the demand side for remanufacturing. Agrawal et al. (2015) tested how pricing and branding of remanufactured products affect the market for new products, and Jakowczyk et al. (2017) studied the determinants of search intensity and number of remanufactured products on offer, using eBay and search traffic data for electronic and electric products.

Our big data analytics framework allows us to map the basic characteristics (main industries and products, most important market players, and key attributes) of the market for remanufactured products as a whole, thus having a broader scope than these previous studies. None of the foregoing studies, nor any other study of which we are aware, has employed semantic similarities or any other NLP techniques to derive insights about product recovery.

2.2. Tapping on the cyberspace for data and answers

As we discuss in more detail in the next section, the input for our cluster analysis consists of word vectors built from a data set crawled from the Internet. Our work is, therefore, related to a growing stream of research using the abundant data residing in cyberspace. The exponential growth in information available on the Internet has been widely acknowledged, Cisco as early as 2011 estimating annual Internet traffic to be in the order of Zettabytes, the equivalent of streaming 36,000 years of high-definition movies (Guardian, 2011). Academics and practitioners alike have tapped, albeit only recently, the abundant data that resides in cyberspace (Netzer et al., 2012). This has arguably been due to the advent of high-speed Internet connections and new protocols for data sharing (e.g., application programming interfaces, or APIs) that have rendered retrieval less problematic than processing and making sense of the enormous body of data.

Management scholars have exploited, among myriad types and sources of online data, information retrieved from blogs, trading platforms like eBay (Resnick et al., 2006; Lucking-Reiley et al., 2007), online retailers (Clay et al., 2001; Smith & Brynjolfsson, 2001), review sites (Li & Wu, 2010; Kang et al., 2012), news websites (Tsagkias et al., 2009), micro-blogging (Li et al., 2018; Yearworth & White, 2018; Comrie et al., 2019), discussion forums (Liu, 2006) and security filings (Bodnaruk et al., 2015) to examine the value of feedback and other quality cues (Eaton, 2005; Cabral & Hortacsu, 2010), find weaknesses in products (Zhang et al., 2012), apply sentiment analysis to products and services (Kang et al., 2012) and forecast stock market performance (Bollen et al., 2011), among diverse other applications.

2.3. NLP

Finally, our paper is methodologically related to a growing stream of studies using NLP analysis. The function of NLP is to automate the extraction of information from human language (Trask et al., 2015). Part-of-speech tagging (determining whether a word in a sentence is a verb or a noun, for instance), named entity recognition (finding people and organisations within a text or collection of texts) and text translation are examples of the application of NLP.

Recently, NLP has found its way into the community of management scholars. In Accounting and Finance research, for instance, this technique has been deployed to examine financial constraints (Bodnaruk et al., 2015; Law & Mills, 2015; Hoberg & Maksimovic, 2014), stock-price crashes (Zhu et al., 2017), market volatility (See-To & Yang, 2017) and fraudulent financial filings (Goel et al., 2010), as well as to predict bankruptcy (Mai et al., 2019). In Operations Management, NLP techniques have been used to identify defective products (Abrahams et al., 2012, 2015) and track quality issues in the food industry (Mishra & Singh, 2016; Singh et al., 2018). To our knowledge, we are the first to apply this technique to the area of CLSC, and more particularly the mapping of the market for remanufactured products.

3. Methodology

This section outlines the methodology used for our analysis. In a first subsection, we discuss the concepts of semantic similarity and word embedding in general terms. In a second subsection, we briefly describe the GloVe word embedding method and resulting word vectors used in our paper. In a third subsection, we discuss how we use the Louvain method for community detection to obtain clusters related to remanufacturing based on the word vectors obtained through GloVe.

3.1. General notes on semantic similarity and word embedding

NLP can be used to evaluate semantic similarity, a decades-old term that has recently experienced an upsurge of interest among academics and for-profit organizations like Google. Semantic similarity, a term sometimes used interchangeably with semantic relatedness, has multiple definitions. For example, Feng et al. (2017) define semantic relatedness as a form of measurement that quantitatively identifies the relationship between two words or concepts based on the similarity or closeness of their meaning . Similarly, Li et al. (2003) argue that similarity between two words is often represented by similarity between concepts associated with the two words. This study's interpretation of similarity is nearer Li et al. (2003)'s definition, although our results are informative as well with respect to identifying terms closely associated with the meaning of remanufacturing.

One approach for measuring semantic similarity consists of considering the co-occurrence of words in text as measured by co-occurrence vectors. Take the word Toyota for example. Its co-occurrence vector will tell how many times the word Toyota appears in the same corpus of all other terms considered in the analyses.[1] The terms Car and Toyota, for example, would possibly be classified as being closely related because the two words co-occur often. The terms Toyota and Shoes will probably not. This approach may lead to the well-known curse of dimensionality. High dimensionality occurs because in a large corpus, thousands of words may neighbour other words. As a consequence, the vectors storing the relationships between a word and its neighbours will be thousands of cells long, and at times very sparse. If there are, for instance, 5,000 different words in the collection of corpora used to train the algorithm, the co-occurrence vector for the word Toyota will be 5,000 long.

An alternative approach for measuring semantic similarity applies a language processing technique called word embedding. This technique involves reducing the vector of words to a vector of much lower dimensionality, thereby allowing for off-the-shelf word embeddings to be created and easily loaded into the computer memory, without suffering the above-mentioned issues related to multidimensionality. The resulting vectors may not have a directly understandable interpretation, but the distances between them normally do. For example, the vector for France will be close to the vector for Italy, and the vector for Xbox close to that for PlayStation (Garg et al., 2018). A review on the effectiveness of word embedding algorithms is provided by Lai et al. (2016). In our analyses, we used a well-known word embedding algorithm called GloVe (Global Vectors for Word Representation), which will be summarized in the next subsection.

3.2. Obtaining word vectors using the GloVe algorithm

The GloVe algorithm uses word embedding to produce word vector representations. Due to the quality and parsimonious nature of its outputs, GloVe is one of the most popular techniques for obtaining word vector representations (Lai et al., 2016). Pennington et al. (2014), the paper in which the algorithm was introduced has been cited, to date, more than 12,000 times.

The GloVe algorithm can be summarized as follows (Source: `https://bit.ly/2QkiSuB`).[2]

1. Create a co-occurrence matrix X. Entries X_{ji} represent the number of times word j co-occurs with word i.

[1] A corpus is essentially a collection of texts.
[2] A lengthier description of the algorithm can be found in Pennington et al. (2014).

2. Calculate $X_i = \sum_k X_{ij}$. X_i represents the number of times all words present in the corpus being examined appear in the context of word i.
3. Define soft constraint for each pair of words, $w_i^T \bar{w}_k + b_i + b_j = log(X_{ij})$, where b_i and b_j are scalar biases.
4. Define a cost function: $J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \bar{w}_j + b_i + b_j - log(X_{ij}))$
5. The weighting function is defined as:

$$f(x) = \begin{cases} (x_{ij}/x_{max})^{\alpha}, & \text{if } X_{ij} < X_{max}. \\ 1, & \text{otherwise.} \end{cases} \quad (1)$$

We retrieved off-the-shelf 300-dimensional word vectors constructed with the GloVe algorithm from spaCy, an open-source software library for advanced NLP activity, written in the programming languages Python and Cython (Source: `https://bit.ly/2vorTeZ`). We did not train the vectors ourselves. The vectors came pre-trained using a corpus collected from Common Crawl. Common Crawl essentially provides a copy of the Internet to interested researchers and organizations. It contains petabytes of data collected over eight years of web crawling (Source: `https://bit.ly/2xJegYJ`)[3]. In total, the pre-trained vectors contained approximately 684,826 words.

3.3. Creating clusters related to remanufacturing using the Louvain method

In a final step of our methodology, we zoomed in on those word vectors representing remanufacturing and related terms. The main focus of our study is remanufactured products, which are denoted with the kernel term *remanufactured*. Consistent with Subramanian & Subramanyam (2012), we used three further kernel terms to capture remanufactured products: *refurbished*, *reconditioned*, and *rebuilt*, since these terms are often used interchangeably with *remanufactured*. For each kernel term, our goal was to identify words semantically close to that term, and create a network resulting in a set of clusters based on the relationships among those words. To achieve this goal, we used the Louvain method (Blondel et al., 2008). The Louvain method is a state-of-the-art method for discovering the community structure of large networks. Its key

[3]The collection of word vectors from spaCy took place in July 2017. The word vector dataset was entitled $en\_core\_web\_md - 1.2.1$.

advantages are the quality of the results generated and the low average computational time of order $\mathcal{O}(n \log n)$ (De Meo et al., 2011). More specifically, we executed the following steps:

1. *Input.* Let the ordered array $\mathbf{w} = \{w_1, \ldots, w_m\}$ represent a set of m distict words. Let $X \in \mathbb{R}_+^{m \times n}$ be a matrix, where row $\mathbf{x_i}$ is vector representation of word w_i of size k constructed using the GloVe algorithm as outlined in the previous subsection ($i = 1, \ldots, 684, 826$ and $k = 300$).
2. Let $w_{\hat{i}}$ denote the row representation for the word *remanufactured.* Calculate θ_i, which equals one minus the cosine similarity between $x_{\hat{i}}$ and x_i, for all i, which represents the similarity between the vector containing the word *remanufactured* and all other words contained in i. The vector containing the aforementioned dissimilarities is represented as $\Theta = \{\theta_1, ..., \theta_n\}$. Set $\mathbf{a}^* = \{a_i : \theta_i$ among the 200 highest values of $\Theta\}$
3. Create a weighted network $G = (\mathbf{a}^*, E)$, where E connects any two elements of $\mathbf{a}^*$, if weights are higher than 0.5.[4] The weights of the edges between any two vertices are calculated using cosine similarity, as in step 2.
4. Create clusters using the Louvain method.
5. Interpret clusters.
6. Steps 1 to 5 are repeated using the remaining three kernel terms in place of *remanufactured.*

We note that up to the point at which the clusters are formed, no interaction with the user is required, and no prior information or knowledge on the kernel terms is provided to the model, nor was the corpus constructed from documents topically related to remanufacturing.

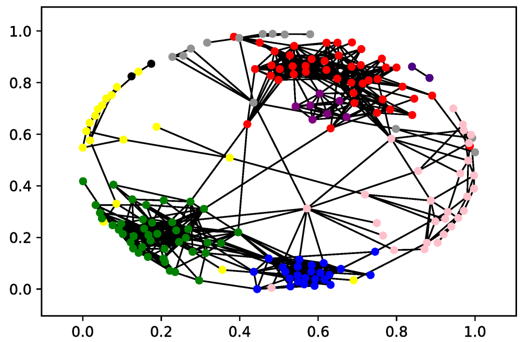
4. Summary of clusters for *remanufactured* and related terms

In this section, we document the clusters obtained for the kernel term *remanufactured*, which is the key focus of our analysis, as well as for the three related kernel terms *refurbished*, *reconditioned*, and *rebuilt.*

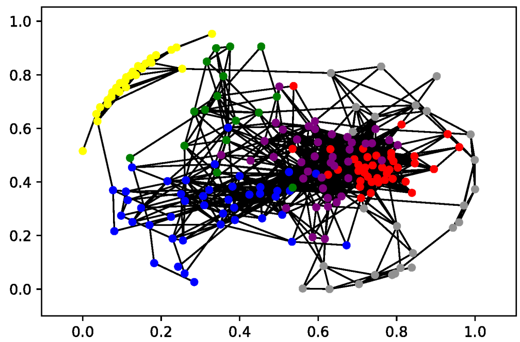
In the graphs initialized with *remanufactured*, and depicted in Figure 1a, eight clusters were formed. Each of the clusters generates a meaningful interpretation, save for the small cluster

[4] Weights vary from -1 to 1, as similarities were calculated using a cosine function. Sub-sectioning the graph in this fashion removes weak relationships.

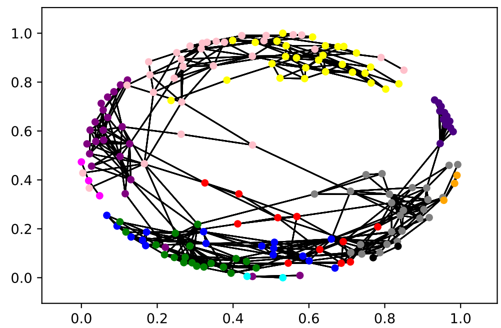
entitled *Unclear interpretation*. From the kernel term *refurbished*, presented in Figure 1b, half of the six clusters are associated with product refurbishing. The other half refer to the refurbishment of buildings, and are thus less relevant to this study and omitted from the remainder of our analysis. Similarly, the clusters formed from the kernel term *reconditioned* contain meaningful information, with the exception of one cluster entitled *Unclear interpretation* (Figure 1c). For the kernel term *rebuilt* (Figure 1d), we also obtain only one cluster without clear interpretation labelled accordingly. Tables 1,2, 3 and 4 contain information on each of the clusters.



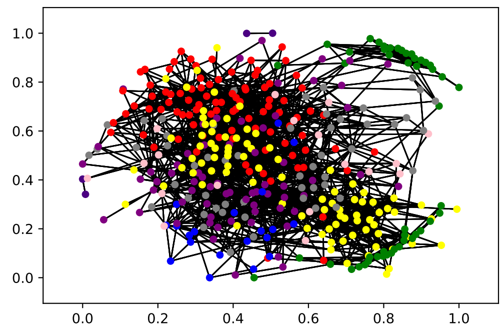
(a) Term *remanufactured*



(b) Term *refurbished*



(c) Term *reconditioned*



(d) Term *rebuilt*

FIGURE 1: Semantic graphs, where similar words are grouped together using the methodology described in Section 3. The vertices are coloured according to their clusters.

Cluster name	Color	Terms	Meaning
Printing equipment	Red	minolta, konica, cartridge, cartridges, packard, hewlett, refills, refill, printers, printer, magenta, cyan, epson, lexmark, toner, toners, inkjet, photocopier, copier, inks, ink, copiers, consumable consumables, printhead, printheads, photocopiers, refilling, refilled, ink-jet, officejet, deskjet, photosmart, laserjet, pixma, xerox, ricoh, inkjets, hp, multifunction, okidata, gestetner, tektronix, fuser.	Companies, products and parts associated with the remanufacturing of printers
Automotive	Blue	chevrolet, chevy, gmc, pontiac, buick, ford, chrysler, mazda, toyota, dodge, oldsmobile, acura, mitsubishi, 4runner, jeep, deisel, cummins, volvo, isuzu, 22re, mopar, windstar, mercruiser, toyotas, jaguar.	Car companies, car models and part retailers
Attributes	Yellow	rebadged, brand-name, name-brand, high-quality, top-quality, highest-quality, dependable, economical, off-brand, top-of-the-line, low-priced, environmentally-friendly, recycled, lowest-priced, salvaged, reliability, high-efficiency	Sellers' soft quality cues
Car parts	Green	cylinders, cylinder, carburetor, carburetors, compressor, compressors, camshaft, crankshaft, injector, injectors, turbocharger, turbochargers, o-rings, gaskets, diesels, diesel, gearbox, gearboxes, flywheel, alternator, alternators, superchargers, powertrain, drivetrain, bushings, bearings, turbo, crankshafts, idler, sprockets, engines, axles, 5-speed, horsepower, carbureted, motors, housings, cyl, outboards, transmissions, reground, unleaded	Remanufactured car parts
Warranty replacement	Grey	replacements, replacement, warranties, warranty, warrenty, warrantee, replace, batteries, chargers, li-ion, replacment, warrantied, specifications	Remanufactured products used for warranty
Cartridges	Purple	refil, cartriges, cartidges, catridge, cartrige, cartidge, catridges	Remanufactured cartridges
Synonyms	Pink	refil, refurb, refurbed, after-market, aftermarket, preowned, pre-owned, manufactures, manufactured, remanufacturing, remanufacture, refurbish, recondition, refurbished, reconditioned, reconditioning, remanufactured, distributor, disassembled, disassemble, oem, reman, spares, retrofitted, parts, rebuilt, factory, longblock, kits, re-engineered, compatable, compatible	Product conditions closely associated with remanufacturing
Unclear interpretation	Indigo	compatable and compatible	Cluster with no obvious interpretation
Treadmills	Black	precor, treadmills	Remanufactured treadmills

TABLE 1: Summary of clusters and interpretations for *remanufactured*.

Cluster name	Color	Terms	Meaning
Real estate	Red	one-bedroom, two-bedroom, two-storey, three-storey, apartment, apartments, semi-detached, bedroomed, bedrooms, bedroom, cottage, cottages, two-story, suite, terraced, storey, townhouse, bungalows, farmhouse, self-catering, penthouse, unfurnished, furnished, detached, bungalow,houses, en-suite, guesthouse, villa, sited, situated, terrace, landscaped, newly-built, family-run, duplex, new-build, annexe, seafront	Refurbished houses and apartments
Synonyms	Blue	re-branded, rebranded, remodeled, renovated, re-designed, redesigned, redecorated, refurnished revamped, derelict, re-built, rebuilt, modernized, modernised, upgraded, overhauled, rennovated, remodelled, disused, refurbed, refurb, preowned, pre-owned, repainted, refinished, refurbs, restored, repaired, redone, disassembled, dismantled, refurbished, reconditioned, refitted, rehabbed, redeveloped,newly, built, retrofitted, decommissioned, converted, salvaged, housed, brand-new	Process undergone by the product being refurbished
Consumer electronics	Yellow	wifi, wi-fi, imac, macbook, laptop, laptops, iphones, ipods, acer, compaq, toshiba, computers sony, panasonic, inspiron, fujitsu, samsung, tvs, televisions, wireless, dell, lcd, phones, cartridges printers, projectors, ac	Consumer electronics
Product conditions	Green	renovations, renovation, renovate, refurbish, upgrading, upgrade, upgrades, dilapidated refurbishment, refurbishing, repairs, newer, replaced, replacement, spares, warranty	Product conditions closely associated with refurbished
Furniture	Grey	bought, purchased, furniture, furnishings, sold, shops, shop, warehouses, warehouse, priced ,discounted, antiques, floors, kitchens, appliances, rented, leased, sale, serviced, showroom, resold, equipment, accessories, factory, manufactured, discontinued	Furniture and kitchen appliances
Quality cues	Purple	offering, offers, beautifully, superbly, well-appointed, spacious, fully-equipped, well-equipped, suites, guestrooms, equiped, equipped, air-conditioned, immaculate, immaculately, stylishly, tastefully, outfitted, accommodation, attractively, rooms, hotel, interiors, spotlessly, stylish, amenities, high-end, top-of-the-line, affordable, modernly, centrally, fitted, fully, boasts, state-of-the-art, purpose-built, well-maintained, luxury, custom-built, ultra-modern, modern	Textual quality cues associated with houses and apartments

Table 2: Summary of clusters and interpretations for *refurbished*.

Cluster name	Color	Terms	Meaning
Household appliances	Red	vaccum, vacum, washer, washers, vacuum, dishwashers, freezers, vac vacuums, appliances, dyson	Household appliances
Landscaping and mining machinery	Blue	mower, lawnmower, mowers, forklift, forklifts, lawnmowers, excavator ,excavators, backhoes, chainsaws, saws, tractors, blowers, kubota, komatsu, strimmer	Construction, landscaping and other heavy equipment.
Synonyms	Yellow	re-badged, rebadged, re-built, rebuilt, refurb, refurbed, preowned, pre-owned, repainted, refinished, referb, refurbs, salvaged, dismantled, disassembled, resprayed reconditioned, refurbished, reupholstered, overhauled, serviced, refurnished, disassemble, inspected, re-engineered, decommissioned, junked, retrofitted	Process undergone by the product be- ing reconditioned
Car Parts	Green	cylinder, cylinders, carburetor, carburetors, compressor, compressors, pneumatic, hydraulic, gearbox, gearboxes, alternator, alternators, crankshaft, bearings, machined, motors, gaskets, reciprocating, radiators	Reconditioned car parts
Construction and manufacturing machinery	Grey	lathes, lathe, 12v, 24v, grinder, grinders, makita, dewalt, nailers, nailer, ryobi, skil, bosch, bandsaw, 110v, hitachi, ridgid, planer, sander, panasonic, 18v, blenders, fujitsu, cordless, shredder	Reconditioned machinery used in construction and manufacturing.
Unclear interpretation	Purple	suppliers, supplier, manufacturers, manufacturer, acessories, accesories, equipments, equipment, manufactures, accessories distributor, distributer, manufactured, machinery, machines, factory, equipt, kitted, equiptment, stockist, consumables, dealer, fitted	Cluster with no obvious interpretation
Spare parts and core	Pink	replacement, repair, repairs, repairing, refurbishing, refurbish, after-market, aftermarket, chargers, batteries, repaired, salvage, remanufacturing, remanufacture, recondition,, reconditioning, spare, spares, rebuildable, repairable, servicing, replacment, parts, oem, defective, warranty, electrics, refurbishes, reman	Spare parts and core for reconditioning
Car manufacturers	Indigo	nissan, toyota, honda, mazda, hyundai, mitsubishi, volvo, vw, isuzu, peugeot, vauxhall, daewoo	Car brands associated with reconditioning
Synonyms 2	Black	secondhand, second-hand	Process undergone by product being reconditioned
Overstock	Orange	closeouts, overstock, overstocked	Overstocked products.
Printing equipment	Magenta	cartridges, refilled, copiers	Products associated with the reconditioned of printers.
Synonym 3	Cyan	reprocessed, reground	Process undergone by product being reconditioned

TABLE 3: Summary of clusters and interpretations for *reconditioned*.

Cluster name	Color	Terms	Meaning
Unclear interpretation	Red	reestablished, re-established, re-opened, reopened, razed, demolished, bulldozed, once, later, revived, eventually dismantled, destroyed, damaged, broke, broken, demolish, ruined, moved, relocated, intact, completely, badly, disrepair, changed, obliterated ,torn, removed, abandoned, trashed, rehabilitated, burned, decimated, wrecked, deteriorated, bombed, stripped, renamed, gutted, looted, occupied, collapsed, blown, old, resurfaced, strengthened	Cluster with no obvious interpretation
Synonym	Blue	re-installed, reinstalled, upgrading, upgrade, upgrades, upgraded, fully, installed, newer, newly	Product conditions closely associated with rebuilding
Synonym 2	Yellow	re-created, recreated, re-done, redone, remodeled, renovated, re-painted, repainted, re-designed, redesigned, redecorated, refurnished, revamped re-built, reassembled, modernised, modernized, resprayed, overhauled, remodelled, restructured, reorganized, retooled, refinished, re-constructed, reworked, re-roofed refurbished, recalibrated, retuned, remade, refitted, revitalized, reshaped, reconfigured, spruced, rehabbed, redeveloped, realigned, retrofitted, re-erected, rewired serviced, repopulated, original	Product conditions closely associated with rebuilding
Car parts	Green	engines, engine, axles, axle, brake, brakes, carburetors, carburetor, flywheel, crankshaft, driveshaft, rear, pistons, cylinder, alternator, drivetrain, gearbox, motors, parts, radiator, supercharger, aftermarket, bolted, factory	Rebuilt car parts
Real estate	Grey	renovations, renovation, towers, tower, renovate, refurbish, buildings, building, renovating, refurbishing, refurbishment, build, constructed, built, crumbling dilapidated, ruins, overhaul, refit, edifice, erected, brick, roof, fortress, palace, housed, cathedral	Rebuilt houses and apartments
Synonyms 3	Purple	replace, replacing, rebuilding, rebuild, replacements, replacement, repair, repairs restoring, restore, rebuilding, re-build, replaced, rebuilt, repairing, resurrected restoration, repaired, restored, salvage, preserved, rebuilds, recondition, reconditioned, reconstructed, historic, excavated, transformed, maintained, century, swapped, converted, junkyard	Product conditions closely associated with rebuilding
Synonym 4	Pink	disassembled, reclaimed, salvaged, cleaned, scrapped, inspected, reused, decommissioned, rusted	Product conditions closely associated with rebuilding
Ford motors	Indigo	dealership, ford	Ford motors

Table 4: Summary of clusters and interpretations for *rebuilt*.

5. Mapping the market for remanufacturing: main insights from the cluster analysis

In this section, we discuss the main conclusions that can be drawn from the clusters presented in the previous section. We have split the section in subsections capturing our three research questions.

5.1. What are the main industries and products represented in the market for remanufacturing?

We obtain two key findings related to this question. First, we observe that:

Printing equipment, automotive (including car parts), treadmills, consumer electronics, furniture, household appliances, landscaping and mining machinery, and construction and manufacturing machinery, are the industries most closely associated with product recovery.

More particularly, the clusters generated with the terms *remanufactured* and *refurbished* identify six industries as being the most closely associated with remanufacturing activity, printing equipment, automotive (including car parts) and consumer electronics being the largest ones. We now discuss a few of these clusters in some more detail.

The cluster entitled *Printing equipment* is readily identifiable as associated with the business of remanufacturing printers and cartridges, and includes OEMs in the printing business (e.g. Konica, Lexmark and Xerox), parts such as toner and cartridges, and types of printers (e.g. inkjet and laserjet). This finding is not surprising. In Europe alone, 20 million toners were sold in 2015, a substantial portion of which consisted of remanufactured OEM cartridges (Source: `https://bit.ly/2UbZyAN`). Prior research on remanufacturing has also alluded to the importance of the printing industry, although a considerable amount of the information presented in the cluster has not been fully discussed (Kerr & Ryan, 2001; Krikke et al., 2004).

Car parts, uncovered by clusters related to kernel terms *remanufactured*, *reconditioned*, and *rebuilt*, is another interesting cluster. The global market for remanufactured automotive parts is estimated to reach US$ 139.8 billion by 2020 (Source: `https://bit.ly/2SDRNB3`). This cluster mainly contains information on the perceived association between automotive parts such as carburators, crankshafts, injectors, and engines, and product recovery.

The cluster *Consumer electronics* shows that personal computers, laptops, phones, television sets, printer cartridges and projectors are the consumer electronic products most commonly associated with the term *refurbished*. Again, the prevalence of this product group is not surprising

- the market for second-hand and refurbished phones is worth US$17 billion, according to Deloitte (Source: `https://bit.ly/2cL1Hff`).

The industry- and product-specific clusters generated with the term *reconditioned*, in turn, include household appliances, as well as machinery used in mining, construction and manufacturing. Once more, this corresponds to important industries. The remanufacturing of mining equipment alone is worth USD 3.8 billion, and is projected to grow at a rate of 3.5 % in the coming years (Source: `https://bit.ly/3bcBYvB`). The market for construction machines is equally important to the economy. Hitachi alone sold JPY 20.8 billion worth of remanufactured construction machinery and parts (Source: `https://bit.ly/2H0tCsQ`).

Overall, our analysis suggests that remanufacturing activity is strongly clustered within a relatively small number of clearly identifiable industries. This industry clustering could be consistent with the theory of isomorphism. Prior research suggests that companies face pressures to conform with certain norms via isomorphic strategies (Deephouse, 1996; DiMaggio & Powell, 2000). In the context of remanufacturing, the dynamic of engagement would be as follows: an OEM that does not presently engage in the activity may feel pressured to do so if all other players in the same industry participate in some form of remanufacturing.

In addition, we also observe that:

The terms remanufactured, reconditioned and rebuilt are strongly associated with B2B market, and slow clockspeed products. The term refurbished*, in turn, is associated with B2C markets, and high clockspeed products.*

We observe a clear delineation between the clusters obtained for *refurbished* and those for the other three kernel terms. Whereas clusters generated by the terms *remanufactured, reconditioned, and rebuilt* refer to printers, cars, and treadmills, for *refurbished* we observe clusters associated with consumer electronics, furniture, and kitchen appliances.

We suggest two plausible explanations for this pattern. The first explanation draws from the concept product clockspeed, as well as terminology related to CLSC activity. Product clockspeed represents the rate at which a product changes (Fine, 2000; Peng et al., 2013). Most products associated with the kernel terms *remanufactured*, *reconditioned*, and *rebuilt* are low clockspeed products, whereas most products associated with the term *refurbished* are high clockspeed products. This result is not entirely unexpected. Product clockspeed may affect the likelihood of remanufacturing taking place - low clockspeed means that remanufacturers will more likely be returning products that are of the same generation as their new counterparts, making the activity more attractive. However, even when clockspeed is fast, it is still possible to put a product that has been collected back in the market. But that needs to be done quickly, rather than perfectly, in order not to bring an outdated product to consumers. Put differently, everything else constant, getting the recovered product back on time is more important than having it with exactly the same functional and cosmetic conditions of a new product. These different quality standard requirements might explain why the term *refurbished* is more often used for high clockspeed products. While most authors in the CLSC literature use the terms *remanufactured* and *refurbished* as synonyms (Ovchinnikov et al., 2014; Subramanian & Subramanyam, 2012; Quariguasi Frota Neto et al., 2016; Abbey et al., 2017), some authors (Thierry et al., 1995; Krikke et al., 2004) argue that refurbishing represents lower cosmetic and functional standards than remanufacturing. More particularly, in the former activity, used products are brought up to a standard of a specified quality, whilst in the latter activity, used products are brought up to standards that are as rigorous as those for new products. Thus, the observed association of *refurbished* with high clockspeed products might reflect the lower standards applicable for the recovery of these products.

An alternative explanation is largely atheoretical, and holds that certain product types are simply associated with different terminologies for remanufacturing, without there being any obvious differences in the quality or thoroughness of the product recovery process. This corresponds with Abbey & Guide Jr (2018)'s argument that remanufacturing is known as refurbishing in the electronics industry. Put differently, to the general public, and contrary to what some authors in the CLSC literature suggest (Thierry et al., 1995; Krikke et al., 2004), the terms *remanufactured*, *refurbished*, *reconditioned* and *rebuilt* may not be used to denote dissimilar product recovery standards, but rather in relation to the product to which they are referring, e.g. a car would commonly be referred to as *remanufactured*, while a phone would be referred to as *refurbished*.

5.2. What are the main firms and brands associated with remanufacturing?

Our first finding with regard to this question is that:

Country of origin	Brands	Total
United States	dell, hp, xerox, hewlett, packard, gmc, cummins,chevy, chevrolet, ford, tektronix, dewalt,	12
Japan	epson, ricoh, konica, mitsubishi, toyota, isuzu, minolta, mazda, acura, toyotas, hitachi, makita, ryobi, panasonic, honda, toshiba, fujitsu, sony, nissan, isuzu, komatsu	21
Korea	samsung, daewoo	2
Europe	chrysler, dodge, volvo, jeep , bosch, dyson	6
Taiwan	acer	1

Note: Chrysler and Dodge are based in the US, but Fiat, their parent company, in Italy.

TABLE 5: Country of origin of OEMs associated with product remanufacturing

Japanese OEMs are more salient than OEMs from elsewhere.

Table 5 lists the countries of origin of the OEMs generated by the cluster analysis for the four different kernel terms. Almost half of the companies in Table 5 are Japanese. This is surprising, and disproportionally large when compared to the proportion of large Japanese companies in the world. Of the top 100 largest companies in terms of market capitalisation, 54 are in the United States. Only two are Japanese (Source: `https://pwc.to/39Jei0L`). A possible explanation can be drawn from the theory of resource-constrained innovation (Weiss et al., 2011). Many of the Japanese companies described in the present paper have been incorporated either in the 1930s or earlier, a period when the entire country suffered from severe shortages of raw material due to extensive military expansion. Japanese companies were, therefore, more likely to suffer from severe shortages of raw materials when compared to their international counterparts (Yasuba, 1996). That constraint, in turn, may have driven Japanese companies to pursue product remanufacturing as an innovative strategy to reduce their use of natural resources. Remanufacturing an engine, for instance, arguably uses only 1% of the material used to build a new machine (Source: `https://bit.ly/2xJGXo9`).

China is also an anomaly, as it is under-represented. Surprisingly, despite the indisputable place of China as one of the most important manufacturing centres in the world, no reference is made to mainland Chinese companies in any of the clusters found in this paper, except for Volvo, which has been recently acquired by Geely, a Chinese car manufacturer (Source: `https:`

//nyti.ms/3aUQyXE). This result holds for all clusters containing references to OEMs. In sum, the public does not seem to associate product recovery to mainland Chinese manufacturers. This could be partly explained by the lack of easily recognisable Chinese brands in the printing, automotive, and consumer electronic industries.[5] Another factor could be low consumer purchase intentions for remanufactured products, which has been cited as the major bottleneck in the development of the remanufacturing industry in China (Wang et al., 2013).

A second important finding in this area is that:

OEMs are more salient than independent remanufacturers.

Whereas OEMs' names appear frequently in the clusters in this paper, independent remanufacturers (IRs) are not mentioned. This result is robust to all clusters of products. For example, for the cluster associated with cars, Chevrolet, Ford, Chrysler, Mitsubishi, Mazda, Jaguar and Toyota, among others, appear. Cluster *Consumer electronics*, in turn, contains the names of OEMs such as Acer, Compaq, Toshiba, Sony, Panasonic, Fujitsu and Samsung, but includes no mention of any independent remanufacturer.

This result may indicate that consumers pay more interest to the OEM that manufactures the products than the independent remanufacturers that recover them. It also suggests that the group of IRs may be more fragmented than that of the OEMs. Whereas the OEMs are normally larger and have more recognizable brands, IRs are typically smaller and have, overall, lower brand value.

Prior research has suggested that semantic similarities and actual engagement are associated. Garg et al. (2018), for instance, observed a correlation between the semantic similarities of occupational roles, e.g. engineer, librarian, with gender-specific terms, and the proportion of those roles effectively carried out by men or women. Accordingly, in a robustness test of the findings obtained through the cluster analysis, we examined the association between semantic similarity of OEMs with remanufacturing activity, as uncovered by the clusters, and actual OEM engagement in this activity. We hand-collected information on the OEMs participation in product remanufacturing through a search of each of their websites and other relevant online sources. We identified an OEM as engaged in product remanufacturing if it sells recovered products, either

[5] OEMs such as Hauwei or Lenovo are exceptions.

Term	Brands	Confirmed participation
Remanufactured	Hp, Xerox, Epson, Hewlett, Packard, Gmc, Ricoh, Konica, Mitsubishi, Cummins, Toyota, Chrysler, Isuzu, Chevy, Chevrolet, Minolta, Volvo, Mazda, Ford, Dodge, Acura, Tektronix, Toyotas, Jeep	22/24
Refurbished	Toshiba, Panasonic, Dell, Acer, Samsung, Fujitsu, Sony	6/7
Reconditioned	Dewalt, Bosch, Hitachi, Makita, Komatsu, Dyson, Toyota, Ryobi, Mazda, Daewoo, Mitsubishi, Kubota, Panasonic, Nissan, Isuzu, Volvo, Fujitsu, Hp, Honda	16/19
Rebuilt	Ford	1/1

Note: The term confirmed participation refers to the engagement of the OEM in product remanufacturing. We verified participation using online sources, e.g. the manufacturer's website.

TABLE 6: Brand association with product remanufacturing and actual engagement

directly to consumers, or via a single authorized dealer (i.e., an independent remanufacturer that has been authorized by the OEM to sell products and parts on its behalf), and/or if it uses remanufactured products to fulfil warranties. Table 6 lists the brands semantically most closely related to product remanufacturing, and the proportion of those that are effectively engaged in this activity. Reassuringly, we observe that most brands observed in the clusters are indeed involved in the recovery of products and parts. The brands that are not directly involved have products that are highly targeted by independent remanufacturers.

5.3. What are the main attributes and quality cues associated with product remanufacturing?

Our main finding in this area is that:

Environmental concerns only weakly appear in the discourse surrounding product remanufacturing, it being dominated by terms associated with either price or quality.

The cluster *Attributes* (Table 1) gives us some interesting insights regarding consumer perceptions on remanufactured products, without having to resort to surveys, experiments, or interviews. Essentially, it tells us how remanufactured products are described, and the main attributes associated with them. An important observation is that most terms used to describe remanufactured items are associated with either value (e.g. economical, low-priced) or quality (e.g brand-name, name-brand, high-quality, topquality, highest-quality, dependable, off-brand, top-of-the-line, reliability, high-efficiency).

Equally interesting is the absence of some expected attributes, particularly attributes alluding to the green market. Early research on the marketing of remanufactured products suggested that, in certain conditions, a green segment of consumers would prefer products in their remanufactured conditions over new products (Atasu et al., 2008). Yet, empirical evidence that such a consumer segment exists remains scarce, and consensus over this issue has not been achieved (Abbey et al., 2015b). In fact, it may be rational even for green consumers to dismiss the environmental credentials of remanufacturing. Prior research has shown that under certain market conditions, remanufacturing may actually be harmful to the environment (Ovchinnikov, 2011; Quariguasi Frota Neto & Bloemhof, 2012; Ovchinnikov et al., 2014; Raz et al., 2017). From a supply standpoint, prior research has suggested that profit, rather than environmental concerns, is the main driver for engagement in product remanufacturing (Lund & Hauser, 2010).

Our research contributes to this discussion by showing that there is very little evidence of an association between words associated with product remanufacturing and the overall environmental credentials of a product. For example, words such as green and sustainable do not even appear in the Attributes cluster. The only exception is the term environmentally-friendly.

6. Implications of our findings for academics and practitioners

Our research could be useful for academic researchers for two main reasons. First, our findings can serve as a foundation for future empirical work on remanufacturing. Second, our findings can contribute to a number of ongoing academic discussions in the area of CLSC. We now motivate these two implications in some more detail.

Our findings can serve as a starting point for future empirical studies by highlighting areas of remanufacturing activity that are currently under-researched. To make it easier for academic researchers to identify such areas, we have contrasted the CLSC academic literature with the clusters obtained in Section 4. We only considered those clusters that mention product types, brand names, or both. We retrieved information on the number of academic papers containing each of the terms that form the clusters and the term *remanufactured*, as well as at least one of the following terms: *closed-loop*, *closed loop*, *reverse logistics*, *product take back*, *product take-back*.

The use of last five terms helps filtering papers that do not fit into the closed-loop literature. The same analysis was carried out with the other three kernel terms.

Table 7 provides the results of this exercise. It outlines, for each cluster, the average number of scientific papers per term in the cluster (we do not provide the total number of scientific papers per cluster, as some of the papers for different terms overlap, so this could result in double-counting). The table also provides information on the cluster sizes, measured as the number of terms per cluster.

The table indicates that the clusters obtained by the procedure developed in the present study mirror well some of the largest sections of the academic CLSC literature. Interestingly, this implies that our big data analytics methodology can bring to light clusters of research associated with the kernel terms, without conducting a systematic literature search.

In addition to looking at clusters in their entirety, it is also useful to consider the extent to which terms within clusters are covered within the prevailing CLSC literature. For this purpose, Figure 2 provides swarm plots of the number of scientific papers associated with each kernel term. Interestingly, we find that, within clusters, there is a large dispersion of the extent to which terms are covered in the literature, despite the fact that many terms have equally strong semantic similarity. Engines, for instance, are mentioned in 5,120 scientific papers, whereas drivetrain and transmissions, respectively, 94 and 2,620 times, in spite of the three terms being roughly equally semantically close to the term *remanufactured*.

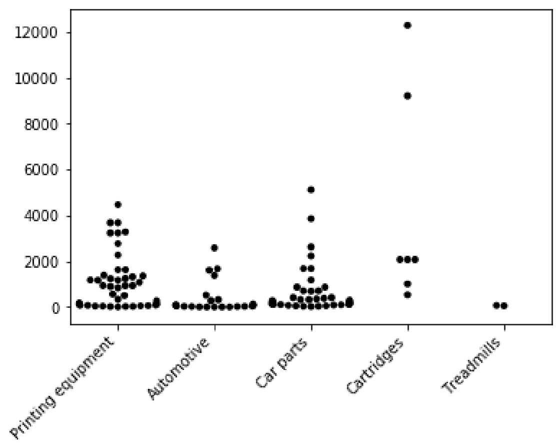
Overall, we hope that our work encourages researchers conducting empirical research on remanufacturing to consider a wider spectrum of industries and products, thereby increasing the cross-sectional variation and power to these studies. By identifying individual OEMs and brand names highly associated with remanufacturing activity, our findings could also be useful for researchers who want to conduct a single-firm case study on remanufacturing.

In addition, our cluster analysis provides relevant keywords related to remanufacturing that could be used as inputs for the sample collection of remanufacturing studies. More particularly, the clusters labelled Synonyms in Table 1 outline the many terms that have a meaning similar to product remanufacturing. For ease of reference, Table 8 outlines the list of terms associated with the four kernel terms. To the best of our knowledge, this is the first time such list is compiled.

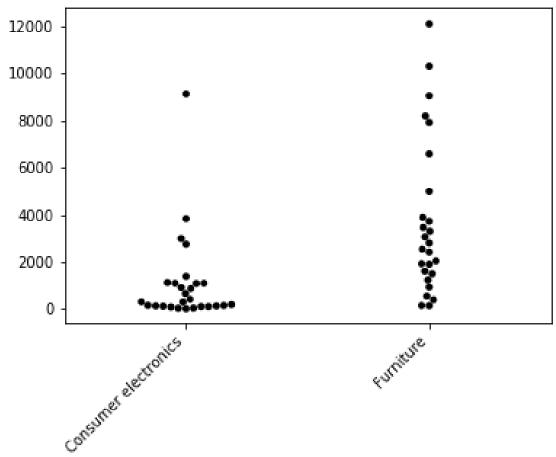
Cluster name	Cluster size	Average number of academic papers	kernel term
Printing equipment	44	1,100.36	remanufactured
Automotive	25	365.36	remanufactured
Car parts	42	658.14	remanufactured
Cartridges	7	4,182.00	remanufactured
Treadmills	2	60.50	remanufactured
Consumer electronics	27	1,077.62	refurbished
Furniture	26	3,716.73	refurbished
Household appliances	11	637.63	reconditioned
Landscaping and mining machinery	16	201.18	reconditioned
Construction and manufacturing machinery	25	414.63	reconditioned
Car manufacturers	12	87.83	reconditioned
Printing equipment	3	3,196.66	reconditioned
Car parts	24	1,535.25	rebuilt
Ford motors	2	710.00	rebuilt

Note: *Cluster size* denotes the number of terms in the cluster. *Average number of academic papers* measures the average number of papers found per term in the cluster, with papers identified using the following search strategy: $<$ kernel term $>$ AND $<$ a term that belongs to the cluster being examined $>$ AND (*closed-loop* OR *closed loop* OR *reverse logistics* OR *product take back* OR *product take-back*) carried out in Google Scholar.

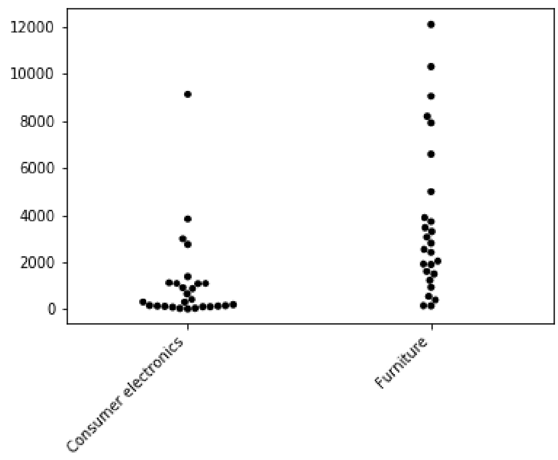
TABLE 7: Contrasting the clusters for the terms *remanufactured, refurbished, reconditioned, and rebuilt* with the existing CLSC literature.



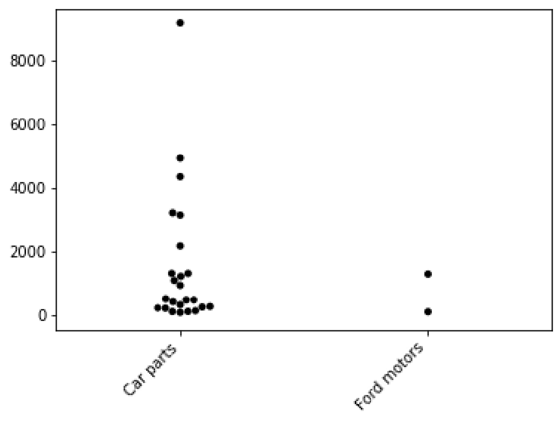
(a) kernel term: *remanufactured*



(b) kernel term: *refurbished*



(c) kernel term: *reconditioned*



(d) kernel term: *rebuilt*

FIGURE 2: Swarm plot of the number of scientific papers for the clusters originating from each of the kernel terms

A second academic implication of our work consists of the contributions it provides to a number of unresolved issues in the CLSC literature. A first unresolved issue relates to the semantic differences between terms related to remanufacturing. Most recent studies use the term *refurbished* interchangeably with other terms denoting remanufacturing activity, but our results suggest that it relates to high clockspeed, B2C products, in contrast with the other three kernel terms under study. We propose an explanation for this pattern derived from product clockspeed theory as well as earlier studies on CLSC terminology (Thierry et al., 1995; Krikke et al., 2004). The link between clockspeed and remanufacturing has been studied analytically (see Reimann et al. (2019), for instance), but not yet empirically. Future studies could explore this further. A second unresolved issue relates to the role of environmental credentials in the marketing of remanufactured products. While one of the fundamental drivers for remanufactured product sales should come from a green consumer segment, our cluster analysis suggests that environmental concerns are not strongly linked with the market for remanufactured products. This observation is consistent with a number of recent studies indicating a lack of familiarity and understanding of environmental benefits of remanufactured products among consumers (Abbey et al., 2015b), but inconsistent with other studies finding that environmental benefits matter for consumers, albeit only for the original product (Michaud & Llerena, 2011). Future research needs to examine the relative extent to which quality, affordability, and environmental credentials influence consumers of remanufactured goods. Finally, the clusters uncovered in our analysis could also potentially assist with the forecasting of consumer purchasing intentions and WTP across different types of remanufactured products, a problem to which researchers have paid considerable attention over the past years.

We hope that our mapping of the main characteristics of the market for remanufacturing is also useful for practitioners and regulators. For example, the cluster analysis' identification of major industries and products linked with remanufacturing may be relevant from the point of view of cannibalization risk. Cannibalization is an issue commonly cited as a major concern for certain OEMs, and information on products and industries potentially facing it is relevant to practitioners and academics alike (Atasu et al., 2010). Although not a sufficient condition, a sizeable market for remanufactured products is certainly necessary for significant levels of cannibalization to occur. In terms of the development of a marketing strategy for remanufactured

aftermarket, after-market, brand-new, built, century, cleaned, converted, decommissioned derelict, disassemble, disassembled, dismantled, distributor, disused, excavated, factory fully, historic, housed, inspected, installed, junked, junkyard, kits, longblock, maintained manufactured, manufactures, modernised, modernized, newer, newly, oem, original, overhauled parts, preowned, pre-owned, preserved, realigned, reassembled, rebadged, re-badged, rebranded re-branded, rebuild, re-build, rebuilding, re-building, rebuilds, rebuilt, re-built, recalibrated reclaimed, recondition, reconditioned, reconditioning, reconfigured, reconstructed, re-constructed, recreated, re-created, redecorated, redesigned, re-designed, redeveloped, redone, re-done re-engineered, re-erected, referb, refinished, refitted, refurb, refurbed, refurbish, refurbished, refurbs, refurnished, rehabbed, reinstalled, re-installed, remade, reman, remanufacture, remanufactured, remanufacturing, remodeled, remodelled, rennovated, renovated, reorganized, repainted, re-painted, repair repaired, repairing, repairs, replace, replaced, replacement, replacements, replacing, repopulated reroofed, reshaped, resprayed, restoration, restore, restored, restoring, restructured, resurrected retooled, retrofitted, retuned, reupholstered, reused, revamped, revitalized, rewired, reworked rusted, salvage, salvaged, scrapped, secondhand, second-hand, serviced, spares spruced, swapped, transformed, upgrade, upgraded, upgrades, upgrading

Table 8: Terms whose meanings are closely related to product remanufacturing

products, our findings suggest that enhancing and communicating both affordability and quality are very important when selling remanufactured products, whereas alluding to green credentials area may be ineffective, given the lack of weight given to such credentials.

7. Conclusions, limitations, and further research

This paper describes the use of pre-trained word vectors trained on an unlabelled general corpus obtained from the Internet (rather than a collection of texts derived on the topic of remanufacturing), and an application of a non-supervised algorithm to extract important information about the market for remanufactured products.

The procedure outlined here has brought into light, the industries and products that are perceived to be most closely associated with remanufacturing activity. It has also outlined the main firms and brands within such industries that are most closely associated with closed-loop supply chains, as well as the key attributes linked with remanufacturing.

This being arguably the first study on the application of big data to the area of product recovery, limitations are to be expected, and these limitations suggest additional avenues for future research.

First, as in any empirical study, the results depend on the choice of data. While the four kernel terms used in this study are based on keywords used in prior empirical studies (Subramanian & Subramanyam, 2012), adding terms like *repurposed* could certainly broaden the scope of this study's findings. Another suggestion for future research would be the analysis of other markets, such as those for recycled or fair-trade products.

Second, using words in the English language only is also a limitation of this work. The limitation arises from the obvious observation that not everyone on the Internet writes or reads in English. Take Internet users in the EU, for instance. A high proportion, namely 44% only read or watch in their native language (`https://bit.ly/2PPCGWl`). As other trained vectors containing information in other languages become available, one could study whether the same clusters re-occur in other contexts, and whether new ones can be found. Bojanowski et al. (2016), for instance, have recently published pre-trained word vectors for 294 languages.

Third, by construction, our analysis maps the market for remanufacturing in a static way, based on data crawled from the Internet over a number of recent years. The algorithm presented in this paper being of low computational intensity and easily parallelized, could be used by future research to examine the evolution of the market for remanufactured products over time, and even long term historical trends.

Overall, we hope our exploratory study illustrates the substantial academic and practical implications of using big data analysis in the context of product remanufacturing. We also hope that this study will generate further, more in-depth work.

References

Abbey, J. D., Blackburn, J. D., & Guide Jr, V. D. R. (2015a). Optimal pricing for new and remanufactured products. *Journal of Operations Management*, *36*, 130–146.

Abbey, J. D., & Guide Jr, V. D. R. (2018). A typology of remanufacturing in closed-loop supply chains. *International Journal of Production Research*, *56*, 374–384.

Abbey, J. D., Kleber, R., Souza, G. C., & Voigt, G. (2017). The role of perceived quality risk in pricing remanufactured products. *Production and Operations Management*, *26*, 100–115.

Abbey, J. D., Meloy, M. G., Blackburn, J., & Guide Jr, V. D. R. (2015b). Consumer markets for remanufactured and refurbished products. *California Management Review*, *57*, 26–42.

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, *24*, 975–990.

Abrahams, A. S., Jiao, J., Wang, G. A., & Fan, W. (2012). Vehicle defect discovery from social media. *Decision Support Systems*, *54*, 87–97.

Agrawal, V. V., Atasu, A., & Van Ittersum, K. (2015). Remanufacturing, third-party competition, and consumers' perceived value of new products. *Management Science*, *61*, 60–72.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463–476.

Atasu, A., Guide Jr, V. D. R., & Van Wassenhove, L. N. (2010). So what if remanufacturing cannibalizes my new product sales? *California Management Review*, *52*, 56–76.

Atasu, A., Sarvary, M., & Van Wassenhove, L. N. (2008). Remanufacturing as a marketing strategy. *Management Science*, *54*, 1731–1746.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, P10008.

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, *50*, 623–646.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*, .

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1–8.

Cabral, L., & Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics*, *58*, 54–78.

Clay, K., Krishnan, R., & Wolff, E. (2001). Prices and price dispersion on the web: Evidence from the online book industry. *The Journal of Industrial Economics*, *49*, 521–539.

Cohen, S., Li, C., Yang, J., & Yu, C. (2011). Computational journalism: A call to arms to database researchers. In *CIDR* (pp. 148–151). volume 2011.

Comrie, E., Burns, C., Coulson, A., Quigley, J., & Quigley, K. (2019). Rationalising the use of Twitter by official organisations during risk events: Operationalising the Social Amplification of Risk Framework through causal loop diagrams. *European Journal of Operational Research*, *272*, 792–801.

De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on* (pp. 88–93). IEEE.

Deephouse, D. L. (1996). Does isomorphism legitimate? *Academy of Management Journal*, *39*, 1024–1039.

DiMaggio, P. J., & Powell, W. W. (2000). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *Economics Meets Sociology in Strategic Management*, *17*, 143–166.

Eaton, D. H. (2005). Valuing information: Evidence from guitar auctions on eBay. *Journal of Applied Economics & Policy*, *24*, 1–19.

Feng, Y., Bagheri, E., Ensan, F., & Jovanovic, J. (2017). The state of the art in semantic relatedness: A framework for comparison. *The Knowledge Engineering Review*, *32*, 1–30.

Ferrara, E., De Meo, P., Catanese, S., & Fiumara, G. (2014). Visualizing criminal networks reconstructed from mobile phone records. *arXiv preprint arXiv:1407.2837*, .

Ferrara, E., De Meo, P., Fiumara, G., & Provetti, A. (2012). The role of strong and weak ties in Facebook: A community structure perspective. *arXiv preprint arXiv:1203.0535*, .

Fine, C. H. (2000). Clockspeed-based strategies for supply chain design. *Production and Operations Management*, *9*, 213–221.

Galbreth, M. R., & Blackburn, J. D. (2010). Optimal acquisition quantities in remanufacturing with condition uncertainty. *Production and Operations Management*, *19*, 61–69.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*, E3635–E3644.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting*, *7*, 25–46.

Goodall, P., Rosamond, E., & Harding, J. (2014). A review of the state of the art in tools and techniques used to evaluate remanufacturing feasibility. *Journal of Cleaner Production*, *81*, 1–15.

Guardian, T. (2011). What's a zettabyte? By 2015, the internet will know, says Cisco. `https://www.theguardian.com/technology/blog/2011/jun/29/zettabyte-data-internet-cisco`. [Online; accessed 14-June-2017].

Guide Jr, V. D. R., & Li, J. (2010). The potential for cannibalization of new products sales by remanufactured products. *Decision Sciences*, *41*, 547–572.

Hamzaoui-Essoussi, L., & Linton, J. D. (2014). Offering branded remanufactured/recycled products: At what price? *Journal of Remanufacturing*, *4*, 1–15.

Harms, R., & Linton, J. D. (2016). Willingness to pay for eco-certified refurbished products: The effects of environmental attitudes and knowledge. *Journal of Industrial Ecology*, *20*, 893–904.

Hazen, B. T., Boone, C. A., Wang, Y., & Khor, K. S. (2017). Perceived quality of remanufactured products: Construct and measure development. *Journal of Cleaner Production*, *142*, 716–726.

Hazen, B. T., Overstreet, R. E., Jones-Farmer, L. A., & Field, H. S. (2012). The role of ambiguity tolerance in consumer perception of remanufactured products. *International Journal of Production Economics*, *135*, 781–790.

Hoberg, G., & Maksimovic, V. (2014). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies*, *28*, 1312–1352.

Hripcsak, G., Friedman, C., Alderson, P. O., DuMouchel, W., Johnson, S. B., & Clayton, P. D. (1995). Unlocking clinical data from narrative reports: A study of natural language processing. *Annals of Internal Medicine*, *122*, 681–688.

Jakowczyk, M., Quariguasi Frota Neto, J., Gibson, A., & Van Wassenhove, L. N. (2017). Understanding the market for remanufactured products: What can we learn from online trading and Web search sites? *International Journal of Production Research*, *55*, 3465–3479.

Jeihoonian, M., Zanjani, M. K., & Gendreau, M. (2017). Closed-loop supply chain network design under uncertain quality status: Case of durable products. *International Journal of Production Economics*, *183*, 470–486.

Jiménez-Parra, B., Rubio, S., & Vicente-Molina, M.-A. (2014). Key drivers in the behavior of potential consumers of remanufactured products: A study on laptops in Spain. *Journal of Cleaner Production*, *85*, 488–496.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, *39*, 6000–6010.

Kerr, W., & Ryan, C. (2001). Eco-efficiency gains from remanufacturing: A case study of photocopier remanufacturing at Fuji Xerox Australia. *Journal of Cleaner Production*, *9*, 75–81.

Krikke, H., Blanc, I. l., & van de Velde, S. (2004). Product modularity and the design of closed-loop supply chains. *California Management Review*, *46*, 23–39.

Lai, S., Liu, K., He, S., & Zhao, J. (2016). How to generate a good word embedding. *IEEE Intelligent Systems*, *31*, 5–14.

Law, K. K., & Mills, L. F. (2015). Taxes and financial constraints: Evidence from linguistic cues. *Journal of Accounting Research*, *53*, 777–819.

Li, H., Gupta, A., Zhang, J., & Flor, N. (2018). Who will use augmented reality? An integrated approach based on text analytics and field survey. *European Journal of Operational Research*, *281*, 502–516.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48*, 354–368.

Li, Y., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, *15*, 871–882.

Linton, J. D. (2008). Assessing the economic rationality of remanufacturing products. *Journal of Product Innovation Management*, *25*, 287–302.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*, 74–89.

Lucking-Reiley, D., Bryan, D., Prasad, N., & Reeves, D. (2007). Pennies from eBay: The determinants of price in online auctions. *The Journal of Industrial Economics*, *55*, 223–233.

Lund, R. T. (1996). *The remanufacturing industry: Hidden giant*. Boston University.

Lund, R. T., & Hauser, W. M. (2010). Remanufacturing-an American perspective. In 5^{th} *International Conference on Responsive Manufacturing – Green Manufacturing* (pp. 1–6). ICRM.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*, 743–758.

Matsumoto, M., Chinen, K., & Endo, H. (2017). Comparison of US and Japanese consumers' perceptions of remanufactured auto parts. *Journal of Industrial Ecology*, *21*, 966–979.

Michaud, C., & Llerena, D. (2011). Green consumer behaviour: An experimental analysis of willingness to pay for remanufactured products. *Business Strategy and the Environment*, *20*, 408–420.

Mishra, N., & Singh, A. (2016). Use of Twitter data for waste minimisation in beef supply chain. *Annals of Operations Research*, (pp. 1–23).

Mugge, R., Jockin, B., & Bocken, N. (2017). How to sell refurbished smartphones? An investigation of different customer groups and appropriate incentives. *Journal of Cleaner Production*, *147*, 284–296.

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*, 521–543.

Ovchinnikov, A. (2011). Revenue and cost management for remanufactured products. *Production and Operations Management*, *20*, 824–840.

Ovchinnikov, A., Blass, V., & Raz, G. (2014). Economic and environmental assessment of remanufacturing strategies for product+ service firms. *Production and Operations Management*, *23*, 744–761.

Pazoki, M., & Samarghandi, H. (2020). Take-back regulation: Remanufacturing or eco-design? *International Journal of Production Economics*, *227*, 107674.

Peng, D. X., Verghese, A., Shah, R., & Schroeder, R. G. (2013). The relationships between external integration and plant improvement and innovation capabilities: The moderation effect of product clockspeed. *Journal of Supply Chain Management*, *49*, 3–24.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global Vectors for Word Representation. In *EMNLP* (pp. 1532–1543). volume 14.

Quariguasi Frota Neto, J., & Bloemhof, J. (2012). An analysis of the eco-efficiency of remanufactured personal computers and mobile phones. *Production and Operations Management*, *21*, 101–114.

Quariguasi Frota Neto, J., Bloemhof, J., & Corbett, C. (2016). Market prices of remanufactured, used and new items: Evidence from eBay. *International Journal of Production Economics*, *171*, 371–380.

Raz, G., Ovchinnikov, A., & Blass, V. (2017). Economic, environmental, and social impact of remanufacturing in a competitive setting. *IEEE Transactions on Engineering Management*, *64*, 476–490.

Reimann, M., Xiong, Y., & Zhou, Y. (2019). Managing a closed-loop supply chain with process innovation for remanufacturing. *European Journal of Operational Research*, *276*, 510–518.

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, *9*, 79–101.

Samarghandi, H. (2017). Studying the impact of merged and divided storage policies on the profitability of a remanufacturing system with deteriorating revenues. *International Journal of Production Economics*, *193*, 160–171.

See-To, E. W., & Yang, Y. (2017). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets*, *27*, 283–296.

Singh, A., Shukla, N., & Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 398–415.

Smith, M. D., & Brynjolfsson, E. (2001). Consumer decision-making at an internet shopbot: Brand still matters. *The Journal of Industrial Economics*, *49*, 541–558.

Subramanian, R., & Subramanyam, R. (2012). Key factors in the market for remanufactured products. *Manufacturing & Service Operations Management*, *14*, 315–326.

Thessen, A. E., Cui, H., & Mozzherin, D. (2012). Applications of natural language processing in biodiversity science. *Advances in Bioinformatics*, *2012*.

Thierry, M., Salomon, M., Van Nunen, J., & Van Wassenhove, L. N. (1995). Strategic issues in product recovery management. *California Management Review*, *37*, 114–136.

Trask, A., Michalak, P., & Liu, J. (2015). sense2vec - A fast and accurate method for word sense disambiguation in neural word embeddings. *arXiv preprint arXiv:1511.06388*, .

Tsagkias, M., Weerkamp, W., & De Rijke, M. (2009). Predicting the volume of comments on online news stories. In *Proceedings of the 18th ACM conference on Information and Knowledge Management* (pp. 1765–1768). ACM.

Voigta, R., Camp, N. P., Prabhakaranc, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences of the United States*, *114*, 6521—6526.

Wang, Y., & Hazen, B. T. (2016). Consumer product knowledge and intention to purchase remanufactured products. *International Journal of Production Economics*, *181*, 460–469.

Wang, Y., Wiegerinck, V., Krikke, H., & Zhang, H. (2013). Understanding the purchase intention towards remanufactured product in closed-loop supply chains. *International Journal of Physical Distribution & Logistics Management*, .

van Weelden, E., Mugge, R., & Bakker, C. (2016). Paving the way towards circular consumption: Exploring consumer acceptance of refurbished mobile phones in the Dutch market. *Journal of Cleaner Production*, *113*, 743–754.

Weiss, M., Hoegl, M., & Gibbert, M. (2011). Making virtue of necessity: The role of team climate for innovation in resource-constrained innovation projects. *Journal of Product Innovation Management*, *28*, 196–207.

Xu, X., Zeng, S., & He, Y. (2017). The influence of e-services on customer online purchasing behavior toward remanufactured products. *International Journal of Production Economics*, *187*, 113–125.

Yasuba, Y. (1996). Did Japan ever suffer from a shortage of natural resources before World War II? *The Journal of Economic History*, *56*, 543–560.

Yearworth, M., & White, L. (2018). Spontaneous emergence of community or: Self-initiating, self-organising problem structuring mediated by social media. *European Journal of Operational Research*, *268*, 809–824.

Zhang, W., Xu, H., & Wan, W. (2012). Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications*, *39*, 10283–10291.

Zhu, Y., Wu, Z., Zhang, H., & Yu, J. (2017). Media sentiment, institutional investors and probability of stock price crash: Evidence from Chinese stock markets. *Accounting & Finance*, *57*, 1635–1670.