

VISUAL SENSEMAKING OF MASSIVE CROWDSOURCED DATA FOR DESIGN IDEATION

He, Yuejun (1); Camburn, Bradley (1); Luo, Jianxi (1); Yang, Maria C. (2); Wood, Kristin L. (1)

1: Singapore University of Technology and Design; 2: Massachusetts Institute of Technology

ABSTRACT

Textual idea data from online crowdsourcing contains rich information of the concepts that underlie the original ideas and can be recombined to generate new ideas. But representing such information in a way that can stimulate new ideas is not a trivial task, because crowdsourced data are often vast and in unstructured natural languages. This paper introduces a method that uses natural language processing to summarize a massive number of idea descriptions and represents the underlying concept space as word clouds with a core-periphery structure to inspire recombinations of such concepts into new ideas. We report the use of this method in a real public-sector-sponsored project to explore ideas for future transportation system design. Word clouds that represent the concept space underlying original crowdsourced ideas are used as ideation aids and stimulate many new ideas with varied novelty, usefulness and feasibility. The new ideas suggest that the proposed method helps expand the idea space. Our analysis of these ideas and a survey with the designers who generated them shed light on how people perceive and use the word clouds as ideation aids and suggest future research directions.

Keywords: Design ideation, Core-periphery structure, Crowdsourcing, Creativity, Computer Aided Design (CAD)

Contact:

He, Yuejun
Singapore University of Technology and Design
Engineering Product Development Pillar
Singapore
yuejun_he@mymail.sutd.edu.sg

Cite this article: He, Y., Camburn, B., Luo, J., Yang, M.C., Wood, K.L. (2019) 'Visual Sensemaking of Massive Crowdsourced Data for Design Ideation', in *Proceedings of the 22nd International Conference on Engineering Design (ICED19)*, Delft, The Netherlands, 5-8 August 2019. DOI:10.1017/dsi.2019.44

1 INTRODUCTION

Online idea crowdsourcing taps into collective intelligence by harvesting many diverse ideas from the crowd via the internet, and has developed into a high potential approach for engineering design and innovation projects (Kristensson *et al.*, 2004; Poetz & Schreier 2012; Schuurman *et al.*, 2012). Rich design concept information can be embedded in submitted textual idea data, and has the potential to inspire new design ideas (Goucher-Lambert & Cagan 2019). But such datasets can be vast, unstructured, and slow to process. The unstructured data must be efficiently mined, systematically summarized, and concisely represented to engage and effectively stimulate further design ideation.

We propose a methodology to summarize massive crowdsourced data in a structured way that is effective for stimulating design ideation and draws on natural language processing (NLP), graph theory and word cloud visualization. We emphasize a core-periphery structure of a complex design concept space embedded in the idea data. Such a strategy is motivated by prior studies suggesting that highly valuable designs are likely to contain both common and uncommon combinations (He & Luo 2017), i.e., cover both the core and the periphery of the space of elementary concepts underlying ideas. In turn, word clouds with a core-periphery structure represent the concept space of the prior crowdsourced ideas and are used as a visual stimulus for design ideation that recombines elementary concepts into new ideas. In this study, we applied this methodology to summarize massive sets of crowdsourced ideas on a future transportation system design and to generate additional novel, useful and feasible ideas for a real design innovation project.

2 RELATED WORK

The core of our proposed method is the extraction of key terms (as the proxy of the elementary concepts) from idea description texts, and their visualization in core-periphery structure word clouds to represent the elementary concept space and stimulate concept recombination for new ideas. We briefly review the relevant literature.

2.1 Crowdsourcing

Crowdsourcing is an approach to engage large numbers of people, often via internet, to respond to an open call for ideas to solve individual, organizational, and societal problems (Pedersen *et al.*, 2013; Howe 2008; Surowiecki & Silverman 2007). Recent advances in information and communication technologies have made it increasingly popular as one can reach out to a larger and more diverse crowd (Brabham 2009; Howe 2008; Grace *et al.*, 2015). Studies have shown the quality, quantity, and diversity of the inputs from the crowd are an improvement relative to traditional methods (Brabham 2009), possibly due to the anonymity and other benefits of electronic collaboration compared with face-to-face methods.

2.2 Natural language processing

Our method starts with Named Entity Recognition (NER) to locate and classify concepts in texts into pre-defined categories and terminology extraction to extract relevant terms from a given corpus. In the field of NLP, there have been rule-based approaches, learning-based approaches, and hybrid approaches to NER and terminology extraction (Goyal *et al.*, 2018; Paziienza *et al.*, 2005), generally followed by part-of-speech (POS) tagging, lemmatization, filtering, etc. NER and terminology extraction will be starting points for automatic text summarization and semantic analytics of natural language data of design ideas.

2.3 Combinatorial creativity

The literature suggests that uncommon recombination of prior art leads to invention and gives rise to novelty (Simonton 1999; Arthur 2007). Youn *et al.* (2015) found that the recombination of existing technologies, rather than the introduction of wholly new technologies, has been the major driver of modern inventions. Taura & Nagai (2012) revealed that in idea generation via concept blending, the highest novelty is obtained from the base concepts with high distance. Fleming (2007) found that novel combinations, lacking historical occurrences, give rise to the variability that can result in both

breakthrough and failure. Kim *et al.* (2016) and He & Luo (2017) found that highly valuable inventions combine both rarely and frequently combined technologies, based on patent data analyses.

2.4 Core-periphery network analysis

Network analysis can reveal the structure of the concept space arising from an ideation or brainstorming exercise. Fu *et al.* (2013) built a Bayesian network on the semantic similarity of patent texts to visualize a network of patents based on their analogical distances. Lim *et al.* (2016) applied structural network modelling to evaluate the interconnectivity of design concepts in massive group ideation and visualized the networks to evaluate the diversity of solutions. A core-periphery structure is embedded in many real-world networks and entails a dense core of cohesive connected vertices and a periphery of sparsely connected vertices (Borgatti & Everett 2000), as illustrated in Figure 1A. Core-periphery structures in networks have been mainly studied using a discrete model (i.e., partition of the network in the core and periphery), and a continuous model (i.e., each node is assigned a measure of “coreness” to determine the degree to which it can be classified in the core). Song *et al.* (2018) recently adopted core-periphery structure analysis for finding core and periphery functions in a product family.

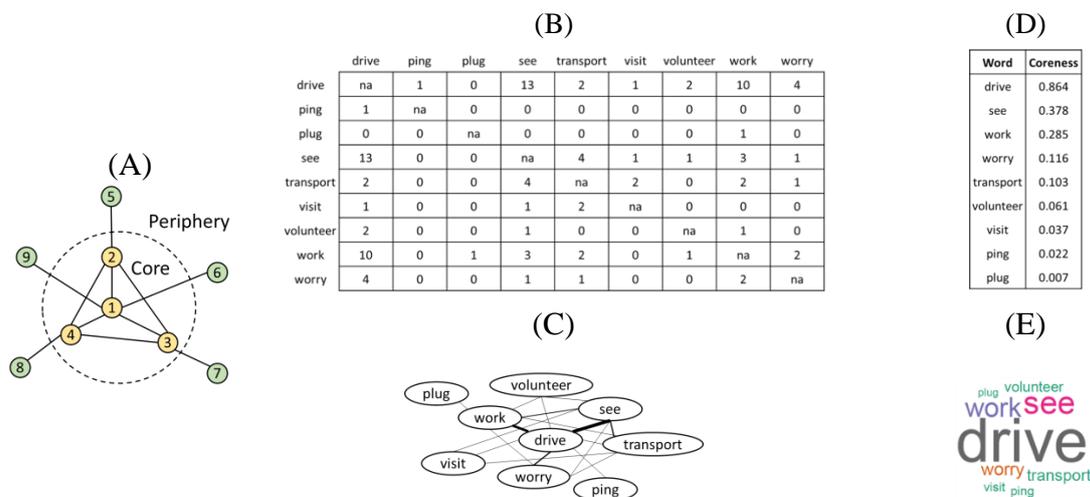


Figure 1. Core-periphery structure. A) A network with an idealized core–periphery structure. B) Co-occurrence matrix of a corpus of verbs. C) Network representation of the verbs in B). D) Coreness scores of the verbs in B). E) Word cloud visualization of the verbs in B).

2.5 Word cloud visualization

Visualization of a large network with many nodes and links can be noisy and not informative. A word cloud, as a visual summary of text data, can be more effective for showing the structure of a complex concept space. In word clouds, words are usually shown with different font sizes and colours based on the chosen properties. Semantic word clouds mostly use term frequency for selecting terms, and then use cosine similarity of terms based on co-occurrences to measure distances, multidimensional scaling (MDS) to produce the initial layout, and a force-directed graph to optimize the final layout (Schubert *et al.*, 2017). Word clouds have been proven as a useful method for text-based visual stimuli for ideation (Herring *et al.*, 2011; Taramigkou *et al.*, 2014), but tag sizes, scanning, centering, positions and layouts can affect human cognition (Lohmann *et al.*, 2009).

2.6 Text stimuli for ideation

Textual information can inspire designers. Chiu & Shu (2012) demonstrated that written language enables the exploration of the solution space during design idea generation. The ambiguity of interpretation that texts may potentially stimulate creative results. Goldschmidt & Sever (2011) found that text-stimuli yield designs that are more original compared to those without stimulation. Gonçalves *et al.* (2014) found that designers do not favour textual stimuli, but texts are somehow often used as stimuli in various ideation methods. For example, mind maps stimulate ideas around a central idea by arranging a radial diagram (Buzan & Buzan 1996), a WordTree identifies analogies by problem re-representation (Linsey *et al.*, 2012), Design Heuristics propose design strategies to facilitate idea generation (Yilmaz *et al.*, 2010), and SCAMPER offers specific ways to transform

existing ideas (e.g., combine, substitute, etc.) (Eberle 1996). Iyer *et al.* (2009) suggested that ideas emerge in a multi-level, modular semantic space, and that priming low-accessible categories in a connectionist semantic network can lead to the generation of novel ideas.

In the ideation methods mentioned above, designers are provided with a small set of stimuli at a time, which may lead to fixation. Our interest is to retrieve and represent massive sets of elementary concepts in idea texts as design stimuli in a structured way, and aid designers to perceive and get inspired by them efficiently for combinatorial creativity.

3 METHOD

Figure 2 shows the proposed workflow of steps to make sense of massive crowdsourced data for design.

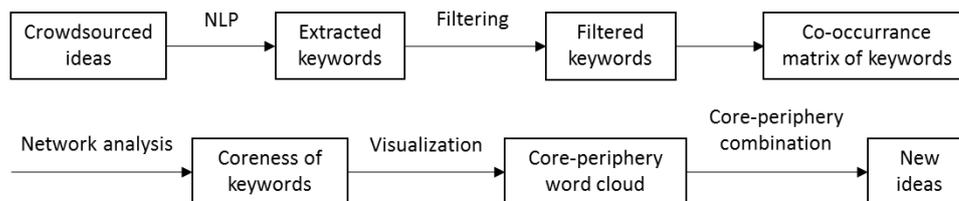


Figure 2. Workflow to make sense of massive crowdsourced data for design

Step 1: Extract keywords from crowdsourced idea texts by NLP. Noun keywords were extracted using TextRazor¹ which extracts all Named Entities (e.g., “Car,” “Uber,” and “Manhattan”) by leveraging a huge knowledge base of entity details from Wikipedia, DBpedia and Wikidata. Its entity linking precision is the highest among alternative entity linkers (Derczynski *et al.*, 2015). Verbs, adjectives, and adverbs were extracted by the natural language toolkit NLTK (Bird & Loper 2004). The textualized ideas were tokenized and tagged as part-of-speech using NLTK’s default Penn Treebank based tagger. Stop words were removed. Words were lemmatized to their base forms (e.g., cars to car) using NLTK’s WordNet library. For example, the idea “We would ride hovercraft to a monorail station. The monorail would take us to work, school, etc.”, would yield the keywords: ride, hovercraft, monorail, work.

Step 2: Filter the extracted keywords. Typos and meaningless words were removed.

Step 3: Generate the matrix of the filtered keywords based on their co-occurring frequencies in ideas, for each kind of keyword. The keyword co-occurrence matrix can be represented as a network, where the nodes are keywords and the link weights are pairwise co-occurring frequencies of keywords. For example, in Figure 1B, “drive” and “see” co-occurred in 13 ideas.

Step 4: Calculate the coreness of the keywords. The objective is to obtain the coreness values that minimize the following function, which defines the correlation between the co-occurrence matrix R and the estimated relation matrix CC^T :

$$F(R, C) = \sum_{i=1}^n \sum_{k=1}^n (r_{ik} - c_i c_k)^2, k \neq i \quad (1)$$

where C is a coreness column vector $(c_1, c_2, \dots, c_n)^T$, R is the co-occurrence matrix and r_{ik} is the co-occurrence frequency of keywords i and k . The minimal residual method (MINRES) of Comrey (1962) was used to solve the equation. $F(R, C)$ is differentiated partially with respect to each element of C and each equation is set to be equal to zero, then the following equations are obtained:

$$c_i = \frac{\sum_{k=1}^n r_{ik} c_k}{\sum_{k=1}^n c_k^2}, k \neq i \quad (2)$$

Thus, c_i is a function of the coreness values of all keywords excluding i itself and the co-occurrence frequencies of keyword i with other keywords. MINRES assumes an initial estimated coreness vector C and uses an iterative process to update the C vector by substitution using Equations 2. After desired iterations, the coreness values are normalized so that the sum of their squares is one. Figure 1D presents the coreness values of different keywords in the example co-occurrence network or matrix. “Drive” is most likely to be in the core and “plug” is the most peripheral word in the corpus.

Step 5: Generate a visual cloud of words based on their coreness degrees in the co-occurrence network. Using the package *wordcloud* (Fellows 2014) in the R environment,² words are assigned

¹ <https://www.textrazor.com>

² R is an open-source software and a programming language, <http://www.r-project.org/>

valid transportation-related ideas. Table 1 lists the 41 valid ideas and the four words from different word clouds that inspired respective ideas.

We analyzed prior occurrences of the keyword/concept combinations of these new ideas in the original set of crowdsourced ideas. 95% of the valid ideas (except for Idea 2 and Idea 24) were new four-keyword combinations and they also present new three-keyword combinations at the same time, and 85% of these ideas (except for Idea 1, 2, 13, 23, 24 and 37) present new two-keyword combinations. Thus, ideation using core-periphery word clouds from the crowdsourced ideas leads to new ideas (as the recombination of prior concepts) in the total concept combination space defined by the original crowdsourced ideas.

Table 1 also lists the evaluation of the novelty, usefulness and feasibility in three levels (high, medium and low) for the ideas. These ideas were assessed by two engineering designers with extensive experience in transportation. Their separate ratings had high Cohen's Kappa ratios (0.89 for novelty, 0.96 for usefulness, 0.93 for feasibility, and 0.93 for the pooled total). Only the assessments of 3 ideas are different between two raters for novelty, 1 idea for usefulness, and 2 ideas for feasibility. The evaluation in Table 1 are based on their consensus. The ideas are sorted first by their novelty (from high to low), then usefulness, and then feasibility. Feasibility is emphasized by the public-sector agency we worked with on this project. In working with a university, practitioners are more interested in exploring novel ideas. Thus, we prioritize novelty in sorting the ideas.

Only Idea 1 was assessed as highly novel, useful and feasible at the same time. Its novelty arises from leveraging the Electronic Road Pricing system (ERP)³ of cars for transit systems for individuals and has been implemented in both Singapore and London. It can accelerate passenger flows in train stations because passengers need not tap cards at a counter to enter a platform; they are automatically identified and charged. Like Singapore's ERP that offer adaptive road fees, the system may adjust train or bus fees adaptively to re-route users along alternate public transit routes depending on loading and congestion, leading to adaptive public transit (either routing or fees). Therefore, this is a novel low hanging fruit that could impact passenger experience significantly.

Idea 3 is highly novel and imaginative and may have appeared in the Harry Potter movies. However, it is not feasible. A feasible alternative idea is to offer virtual reality (VR) travel experiences, i.e., VR teleportation, in telephone booths. Such booths can be further synchronized with a telepresence device on the delivery end, such as the Double 2 Tele presence device.⁴ Idea 4 is novel in that it explores the integration of social services and transport, i.e. bus and church together. It is certainly feasible to have church sessions on a big bus with some custom layout designs. This general concept could be applied to integrating other social services in transport as well, counselling, exercise, etc., and redefine the future of transportation as a place where we do more than moving from place to place. On the opposite end of the spectrum are German beer-bike bars⁵ or the semi-spiritual 'hippie van' movement which defined a generation of free thinkers and is connected in some ways to the ideals of Silicon Valley and innovation.

Idea 6 is quite novel although it is similar to the Gyrosphere vehicle⁶ that appeared in the Jurassic World movie. It is feasible but might be only useful for special places, e.g., parks, and occasions, e.g., disaster relief. This idea might also relate to the Goodyear Eagle 360 sphere tire concept that takes advantage of omnidirectional motion⁷ to revolutionize the way that cars drive with smoother and more complex dynamic motion. Idea 10 is on community-based sharing of vehicles of different functions, which have existed in some places. It is certainly useful and feasible. Idea 11 is the vision of many city governments and urban planners today. There have been some flying taxi projects (e.g., Uber Elevate) and also many initiatives to construct and use more tunnels (e.g. The Boring Company of Elon Musk) which contribute to a 3-dimensional urban transportation system.

Idea 16 might be similar to the Volkswagen Maglev Car concept, which looks like a bubble, released in 2017 in a public video.⁸ The Volkswagen video is rendered but effectively shows how a hover car realistically navigates the streets, whereas its actual working requires the infrastructure underneath the road surface. A distributed maglev infrastructure would enable more impact or wider dissemination of

³ <https://www.lta.gov.sg/content/ltaweb/en/roads-and-motoring/managing-traffic-and-congestion/electronic-road-pricing-erp.html>

⁴ <https://www.amazon.com/Double-Robotics-Telepresence-accessories-included/dp/B01B3PQ0QS>

⁵ <https://www.berlincitytours.com/beer-bike-tour/>

⁶ <http://www.jurassicworld.com/intel/location/gyrosphere>

⁷ <https://www.youtube.com/watch?v=oSFYwDDVgac>

⁸ <https://www.youtube.com/watch?v=eIBjAYGwjtI>

the maglev technology which is highly efficient. Idea 18 is related to the touted concept of flying taxi or the passenger drones. Some initiatives, such as eHang passenger drones in Dubai, AirBus flying taxi and the Uber Elevate program, have been in pilot. Idea 20 might be useful for the disabled who have limited mobility. An expanded idea is to adapt one's emotional states to provide a more enjoyable ride, though today's brain control techniques do not have the accuracy needed for transportation. Idea 22 looks problematic, but actually was pursued by Ryanair some time ago.⁹ Although it is generally difficult to implement a direct tax, one can imagine incentive schemes that incorporate health-mindedness into transportation.

Some of the ideas, which were assessed with low novelty, have been implemented or prototyped by companies or regional governments in some places of the world, despite the lack of wide adoption. For instance, idea 24 has been realized with car sharing systems, such as ZipCar and Car2Go. Idea 27 has been fulfilled by the electric autonomous taxi fleet that Google and Uber are testing in certain regions. The personal rapid transit (PRT) system in Masdar City¹⁰ fulfills both ideas 24 and 27. Idea 28 is also useful and feasible, and has been implemented by Lime Scooters, an electric scooter sharing service from San Francisco.¹¹ An expanded idea is to make such last-mile mobility devices autonomous and able to return themselves back to the station after use. Likewise, idea 29 is has been implemented by dockless bike sharing systems, e.g., OFO and MoBike, which have increased bike uses in major cities in China, Singapore and a few other countries. Idea 32 essentially is an electric, autonomous and shared taxi. Grab has a large fleet of BYD all-electric taxis in Singapore.¹² Google, Uber and NuTonomy have been testing electric and autonomous taxis in the U.S. and Singapore.

Table 1. Valid ideas generated using word clouds as aids

#	Verb	Adj	Noun	Adv	New Ideas	Novel	Useful	Feasible
1	save	efficient	rapid transit	somewhere	Save time spent in the counters of rapid transit systems in peak hours by integrating a system such as ERP of Singapore with card-like transmitter-receiver in commuter-side and transmitters-receivers covering whole station entrance in station-side.	High	High	High
2	drive	small	car	enough	Make a big car into smaller units that are only suited for 1 person, so people can drive or connect to other so there will be enough space in the street.	High	High	Medium
3	teleport	universal	public transport	continuously	A continuously-operating teleportation "phone booths" in corners of streets. These booths can take you anywhere you want and operate with something like a bus card. Now the whole world is connected!	High	High	Low
4	love	big	bus	fun	A big bus is decorated as a church to communicate love	High	Medium	High
5	encourage	simple	walkability	anywhere	A shared transportation system like bus or taxi to offer an easy option that includes exploration. This was once called a "Sunday drive" a colloquial term for moving around solely to observe new places. It should be easy to adapt this route and make stops for fun.	High	Medium	High
6	envision	futuristic	ball	incredibly	I envision a futuristic hamster ball vehicle that has incredible resistance to disasters.	High	Medium	High
7	attach	soap	coffee	leisurely	The transportation in the neighbourhood will be of leisure style with daily consumption experience such as soap trials, coffee trials, and other service attachments.	High	Medium	High
8	suspend	golf-cart-only	windows media audio	along	Suspended golf carts that travel along windows! Suspended window roads (so they are see-through and like the second layer of road) that only golf carts travel along. It's a priority lane for golf carts, but it's gotta be see through so you can't carry dangerous goods inside. Semi-autonomous golf carts that keep you in suspense as you look out your windows at them whizzing dangerously by.	High	Medium	Low
9	happen	solar	street	already	Flash mobs on existing streets convert them to solar power plants inaccessible to cars.	High	Low	Low
10	anticipate	various	spiral	fast	Have a set of vehicles for a neighbourhood that only the families in the neighbourhood can access. Different types like car, truck, van, motorcycle. And the families can use them as needed.	Medium	High	High
11	fly	underground	air pollution	anywhere	In the future we could make the best use of the 3D space we leave to avoid congestion. The roads can be built anywhere in the air and underground. In the air, commuters wear customized suits to fly anywhere; while the underground tunnels are built deeper and more complex.	Medium	High	Medium
12	fly	personalized	hybrid vehicle	anywhere	A flying vehicle that can be adapted to different individuals and different conditions.	Medium	High	Low
13	travel	public	transport	enough	Share all the transportation on reasonable price.	Medium	High	Low
14	fly	uber	train	anywhere	An Uber train that can fly anywhere.	Medium	High	Low
15	hover	small	train	fast	Hover fast small train.	Medium	Medium	Medium
16	drive	bubble	maglev	forever	Bubble maglev.	Medium	Medium	Medium
17	load	mega	scooter	sharply	The new transportation system would be a convoy of scooters. You can hop on at any point and select your route. The scooters could automatically take you and your friends there, all at once. This would require a rail or something similar for scooters only.	Medium	Medium	Medium

⁹ <https://www.independent.co.uk/travel/news-and-advice/ryanair-may-charge-a-fat-tax-squo-for-its-overweight-passengers-1672979.html>

¹⁰ <https://masdar.ae/en/media/detail/masdar-citys-prt-system-celebrates-milestone-with-2-millionth-passenger;>
<https://www.youtube.com/watch?v=5G9X0voSi2Y>

¹¹ <http://fortune.com/2018/11/29/scooters-rental-startup-lime-definers>

¹² <https://cleantechnica.com/2017/02/24/byd-e6-selected-for-south-east-asias-largest-all-electric-taxi-fleet/>

18	jump	monkey	coffee	exponentially	Vehicles carry passengers travel between rooftops of buildings, like monkeys jumping between (coffee) trees. The speed of traveling increases exponentially with the distance - so it won't take long to get anywhere.	Medium	Medium	Medium
19	bypass	natural	speed limit	completely	So we need traversable wormholes.	Medium	Medium	Low
20	imagine	easy	vehicle	fast	Brain controlled car.	Medium	Medium	Low
21	light	environmental	uav	safely	Take environmental UAV to go safely with light speed.	Medium	Medium	Low
22	snow	fat	tax	basically	Tax fat people for being fat, basically	Medium	Low	Low
23	move	efficient	transport	easily	Improve public transportation service so everyone could benefit. Encourage sharing between community members. Better urban design (e.g., cluster functional buildings based on the common needs of the community) to reduce the needs for transportation.	Low	High	High
24	pay	electric	car	enough	There are enough cars in our world now. To reduce electric consumption, we could pay the car by usage.	Low	High	High
25	automate	pre-programmed	trolleybus	autonomously	Offer a few bus lines using pre-programmed autonomous trolleybuses.	Low	High	High
26	embed	viable	urban rail transit	sharply	Embed urban rail underground to link suburbs with cities.	Low	High	High
27	automate	electric	autonomous car	constantly	The next generation of car would be automated, electric and runs constantly across the neighbourhood.	Low	High	High
28	connect	electric	rapid transit	anywhere	Last mile rapid connection by using electric powered mobility rental devices.	Low	High	High
29	assume	heavy	bike rental	friendly	More bikes for transport.	Low	High	High
30	ride	easy	vehicle	anywhere	Vehicle that users ride on, easily used, easily accessible and can be used anywhere.	Low	High	Low
31	prevent	personal	drag (physics)	human error	Prevent personal physical human error automatically.	Low	High	Medium
32	automate	smart	autonomous car	automatically	An autonomous car automates our transportation automatically.	Low	High	Medium
33	power	high-traffic	share taxi	safely	Every share taxi will be renewable power charged at various stations where there is high traffic flow, so that demand for share taxis can be supplied for. The taxis should be autonomous though, I'm not sure if that makes things safer.	Low	High	Medium
34	restrict	general	transport	underground	Create an underground infrastructure for vehicles and reduce surface traffic.	Low	High	Medium
35	buzz	brilliant	blood pressure	technologically	technology bracelet that buzzes periodically to prompt you to be mindful in your busy day and measures your blood pressure too to alert you when you are stressed.	Low	Medium	High
36	ping	public	couch	along	A couch where people can ping for services?	Low	Medium	High
37	drive	public	car	basically	Basically, we use public car more than private one.	Low	Medium	High
38	encourage	free	autonomous car	efficiently	The next generation of transportation would be autonomous, free way to go, and encourage the efficiency of society.	Low	Medium	Medium
39	share	solar	public transport	technologically	Apply solar on public transport.	Low	Medium	Low
40	populate	free	train	efficiently	Train is the most popular and efficiently way to transport and is free.	Low	Medium	Low
41	see	electric-run	road surface making	already	Running track that can store electricity based on people running on it.	Low	Low	High

5 DISCUSSION

To better understand ideation process and factors that may affect the effectiveness of this method, we gathered comments from the participants on their experiences of using the core-periphery word clouds as a rapid ideation aid. 43 of the 58 participants submitted feedback, and 38 of those comments were informative about the ideation method. Participants generally thought the ideation process by word combination was interesting, inspiring, and helpful. The word clouds stimulated them to conceive new combinatorial ideas that they would not have thought of before. Meanwhile, the properties of word clouds, e.g., word sizes, positions, colors, scales, etc., affect users' attention for word selections. For instance, words with bigger sizes may attract more attention. Some suggested that it would be beneficial to reduce the number of words in each cloud, detect phrases, categorize words by meaning, and filter words by technology requirements. Also, there appear to be different patterns in the word selection process, which differ in fixation, flexibility, and randomness. There might be fixation early on with the first words. Some people pointed out they combine random words, while some others played around with the word they first noticed.

In addition to selecting words, transforming words to idea elements also plays an important role in the ideation process. First, based on the observation of the collected ideas, it is common that the participants transform selected words to different forms, e.g., connect → connection. Thus, it may be unnecessary to divide words into different groups by part of speech and let people select them in order. Second, it is common to transform selected words to their “hypernyms”, “hyponyms”, and “synonyms”, e.g., vehicle → car, imagine → brain controlled, etc. This kind of transformation could be an effective way to expand the concept space and diverge in ideation, since it could reach unlimited concepts outside the existing one. This could also be a strategy to alter a selected word and match up with the others. Third, people may use a portion of the selected words or do not use them at all, e.g., rapid transit → rapid, electric-run → electricity. Fourth, the sequence of the words does not matter. People do not use the selected words in the original sequence.

Since the initial word clouds are based on the data collected using Amazon's Mechanical Turk, we have some key observations to report. Results of ideation data drawn using Mechanical Turk can vary dramatically based on precise details of wording. A slight wording change in the prompt can dramatically vary the output. This is because, unlike other data collection methods, there is no context or implied objective. If a study was conducted inside an engineering laboratory, a participant might infer that the objectives relate to engineering. A Mechanical Turk study could relate to marketing, research in linguistics, philosophy, or psychology. Therefore, the task must be very explicit and should be piloted with small sample sizes. Secondly, a basic filtering test is often recommended to attempt to verify that the user is indeed a human and paying attention.

The exploratory study has limitations. Firstly, we introduced and demonstrated the ideation method using NLP techniques of English. In fact, different languages and lexicons will require different NLP techniques for keyword extraction and impact the results, whereas the general methodology framework would not change. Secondly, this paper uses network analysis to reveal the core-periphery structure based on co-occurrence relationships. Alternative methods, such as the deep learning-based word embedding models, can be also applied to deriving semantic relationships in future research. Thirdly, the ideation exercise only presented the core-periphery word clouds to the participants. A comparison with other representations, such as word clouds without the core-periphery structure, a linked network representation, a word list, would reveal more specific influences of the core-periphery structure representation on the ideation performance.

6 CONCLUSION

This study proposes a data-driven visualization methodology to summarize massive crowdsourced idea data, represent the underlying concept space as core-periphery word clouds, and use them as ideation stimuli for concept recombination. We applied this methodology to a real world design ideation project for a public-sector agency to explore future transportation system design ideas, and generated novel, useful and feasible ideas from the word clouds of previously-crowdsourced ideas. In addition to the ideas generated for the interest of the public-sector project for innovation, we also developed a preliminary understanding on how people might use the word cloud visualizations in different manners during ideation, how different visualization layouts may affect cognition, and thus directions for further research and development of this methodology.

These preliminary findings have suggested directions for future research on different strategies to construct and use the word clouds as design stimuli and their impact on the ideation process and outcomes. Specifically, future research should conduct controlled human experiments to test the hypotheses developed in this exploratory study about word cloud representations, word cloud contents, ideation process, and outcomes. With the further research, we anticipate the development of a potential computer tool that generates and represents core-periphery word clouds for data-driven design ideation. Such a tool will not be limited to crowdsourcing idea data, and be generally applicable to the visual sensemaking of any massive idea database for design ideation to expand the original idea space.

ACKNOWLEDGEMENTS

The research is funded by the SUTD-MIT International Design Centre (IDC, idc.sutd.edu.sg), supported by the MIT Ideation Lab and the SUTD Data-Driven Innovation Lab.

REFERENCES

- Arthur, W.B. (2007), "The structure of invention". *Research policy*, Vol. 36, pp. 274–287.
- Bird, S. and Loper, E. (2004), "NLTK: the natural language toolkit". *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, Association for Computational Linguistics, 31.
- Borgatti, S.P. and Everett, M.G. (2000), "Models of core/periphery structures". *Social networks*, Vol. 21, pp. 375–395.
- Brabham, D.C. (2009), "Crowdsourcing the public participation process for planning projects". *Planning Theory*, Vol. 8, pp. 242–262.
- Buzan, T. and Buzan, B. (1996), "The mind map book: How to use radiant thinking to maximize your brain's untapped potential", *Plume New York*.
- Chiu, I. and Shu, L. (2012), "Investigating effects of oppositely related semantic stimuli on design concept creativity". *Journal of Engineering Design*, Vol. 23, pp. 271–296.
- Comrey, A.L. (1962), "The minimum residual method of factor analysis". *Psychological Reports*, Vol. 11, pp. 15–18.

- Derczynski, L., Maynard, D., Rizzo, G., Van Erp, M., Gorrell, G., Troncy, R., Petrak, J. and Bontcheva, K. (2015), "Analysis of named entity recognition and linking for tweets". *Info. Processing & Management*, Vol. 51, pp. 32–49.
- Eberle, B. (1996), *Scamper on: Games for imagination development*, Prufrock Press Inc.
- Fellows, I. (2014), "Wordcloud: Word Clouds (2014)". *R package version*, Vol. 2.
- Fleming, L. (2007), "Breakthroughs and the "long tail" of innovation". *MIT Sloan Management Review*, Vol. 49, p. 69.
- Fu, K., Cagan, J., Kotovsky, K. and Wood, K. (2013), "Discovering structure in design databases through functional and surface based mapping". *Journal of Mechanical Design*, Vol. 135, p. 031006.
- Goldschmidt, G. and Sever, A.L. (2011), "Inspiring design ideas with texts". *Design Studies*, Vol. 32, pp. 139–155.
- Gonçalves, M., Cardoso, C. and Badke-Schaub, P. (2014), "What inspires designers? Preferences on inspirational approaches during idea generation". *Design Studies*, Vol. 35, pp. 29–53.
- Goucher-Lambert, K. and Cagan, J. (2019), "Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation". *Design Studies*, Vol. 61, pp. 1–29.
- Goyal, A., Gupta, V. and Kumar, M. (2018), "Recent Named Entity Recognition and Classification techniques: A systematic review". *Computer Science Review*, Vol. 29, pp. 21–43.
- Grace, K., Maher, M.L., Preece, J., Yeh, T., Stangle, A. and Boston, C. (2015), "A process model for crowdsourcing design: A case study in citizen science". *Design Computing and Cognition '14*. Springer.
- He, Y. and Luo, J. (2017). "The novelty 'sweet spot' of invention". *Design Science*, Vol. 3, p. e21.
- Herring, S.R., Poon, C.M., Balasi, G.A. and Bailey, B.P. (2011), "TweetSpiration: leveraging social media for design inspiration". *CHI '11 on Human Factors in Computing Systems*, ACM, pp. 2311–2316.
- Howe, J. (2008), *Crowdsourcing: How the power of the crowd is driving the future of business*, Random House.
- Iyer, L.R., Doboli, S., Minai, A.A., Brown, V.R., Levine, D.S. and Paulus, P.B. (2009), "Neural dynamics of idea generation and the effects of priming". *Neural Networks*, Vol. 22, pp. 674–686.
- Kim, D., Cerigo, D.B., Jeong, H. and Youn, H. (2016), "Technological novelty profile and invention's future impact". *EPJ Data Science*, Vol. 5, p. 8.
- Kristensson, P., Gustafsson, A. and Archer, T. (2004), "Harnessing the creative potential among users". *Journal of product innovation management*, Vol. 21, pp. 4–14.
- Lim, S.Y.C., Camburn, B.A., Moreno, D., Huang, Z. and Wood, K. (2016), "Design Concept Structures in Massive Group Ideation". *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.
- Linsey, J., Markman, A. and Wood, K. (2012), "Design by analogy: a study of the WordTree method for problem re-representation". *Journal of Mechanical Design*, Vol. 134, p. 041009.
- Lohmann, S., Ziegler, J. and Tetzlaff, L. (2009), "Comparison of tag cloud layouts: Task-related performance and visual exploration". *IFIP Conference on Human-Computer Interaction*, Springer, pp. 392–404.
- Pazienza, M.T., Pennacchiotti, M. and Zanzotto, F. M. (2005), Terminology extraction: an analysis of linguistic and statistical approaches. *Knowledge mining*. Springer.
- Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O. and De Vreede, G.-J. (2013), "Conceptual foundations of crowdsourcing: A review of IS research". *System Sciences (HICSS), 46th Hawaii International Conference on, 2013*. IEEE, pp. 579–588.
- Poetz, M. K. and Schreier, M. (2012), "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?" *Journal of product innovation management*, Vol. 29, pp. 245–256.
- Schubert, E., Spitz, A., Weiler, M., Geiß, J. and Gertz, M. (2017), "Semantic Word Clouds with Background Corpus Normalization and t-distributed Stochastic Neighbor Embedding". *arXiv preprint arXiv:Vol. 1708*. p. 03569.
- Schuurman, D., Baccarne, B., De Marez, L. and Mechant, P. (2012), "Smart ideas for smart cities: investigating crowdsourcing for generating and selecting ideas for ICT innovation in a city context". *Journal of theoretical and applied electronic commerce research*, Vol. 7, pp. 49–62.
- Simonton, D.K. (1999), "Creativity as blind variation and selective retention: Is the creative process Darwinian?" *Psychological Inquiry*, Vol. 10, pp. 309–328.
- Song, B., Luo, J., Mohan, R.E. and Wood, K.L. (2018), "Data-Driven Function Network Analysis for Product Platform Planning: A Case Study of Spherical Rolling Robots". *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2018)*.
- Surowiecki, J. and Silverman, M.P. (2007), "The wisdom of crowds". *American Journal of Physics*, Vol. 75, pp. 190–192.
- Taramigkou, M., Paraskevopoulos, F., Bothos, E., Apostolou, D. and Mentzas, G. (2014), "Leveraging user inspiration with microblogging-driven exploratory search". *International Conference on Advanced Information Systems Engineering*. Springer, pp. 238–249.
- Taura, T. and Nagai, Y. (2012), "Concept generation for design creativity: a systematized theory and methodology", Springer Science & Business Media.
- Yilmaz, S., Seifert, C.M. and Gonzalez, R. (2010), "Cognitive heuristics in design: Instructional strategies to increase creativity in idea generation". *AI EDAM*, Vol. 24, pp. 335–355.
- Youn, H., Strumsky, D., Bettencourt, L. M. and Lobo, J. (2015), "Invention as a combinatorial process: evidence from US patents". *Journal of The Royal Society Interface*, Vol. 12, p. 20150272.