Crowdsourcing and its relationship to wisdom of the crowd and insight building: a bibliometric study

Avishag Gordon[1] ⓘ

Abstract

First coined in 2006, the term 'crowdsourcing' began to appear as a concept in academic literature around 2009. Since then academic research on this topic has been showing a rising trend, most prominently in the field of computer science and its subfields. Analysis of the types of motivations for using crowdsourcing shows that extrinsic motivations are mentioned more often in the literature, compared to intrinsic ones. This serves to establish the perception of crowdsourcing as a research method for gaining robust insight on an issue, rather than as a recreational activity for personal development. Finally, the findings of this study reveal that China is emerging as the main funder and producer of papers on crowdsourcing.

Keywords Crowdsourcing · Insight · Motivations · Bibliometrics

Introduction

The main task of crowdsourcing is harvesting data and processing it with the help of the crowd (Amsterdamer & Milo, 2014). The term was first coined in 2006 by *Wired* writer Jeff Howe (Drabham, 2013), although earlier, in 2004, reference was made to 'wisdom of the crowd' as a collective intelligence for analyzing problems (Heckman, 2004, Surowieck, 2004). The emergence of crowdsourcing, which impacts the evolution of knowledge management, was generated by the proliferation of the internet. (Ghezzi et al., 2018).

Crowdsourcing is designed to arrive at an insight regarding a certain issue, so that 'crowdsourcing' and 'insight building' can practically be viewed as synonyms. Crowdsourcing aids in knowledge creation and acquisition by converting the external environment into a significant knowledge source (Dimitrova & Scarso, 2017), and is currently considered effective for increasing the ability to solve complex tasks (Oishi et al., 2014).

In the context of this article, an insight is defined as an exploratory data analysis that seeks to understand what the data has to say about an issue based on accumulating X number of opinions. An insight is derived from a voluntary response sample comprising people

---

✉ Avishag Gordon
avishag@g.technion.ac.il

[1] The University of Haifa, Haifa, Israel

who responded to a general appeal (Moore, 1997, 119). In this sense it has similar attributes to those of crowdsourcing (Fig. 1).

An insight that consists of X number of responses is like an unstructured opinion poll that aims at gathering information on the opinion of a large group of people on a subject. Similar to an inferential statistical sample, an insight aims at extending knowledge and information beyond a specific authoritative decision. An example of seeking insight would be the attempt of a municipal authority to arrive at a decision on how urban neighborhoods should be designed in the future based on the residents' opinion (Drabham, 2009). In this process, residents' opinions are added to the municipal authority ability, to arrive at a decision.

Crowdsourcing aims at reaching an insight about an issue with the help of people's opinions. It shares attributes with inferential statistics in that it relies on evidence collected
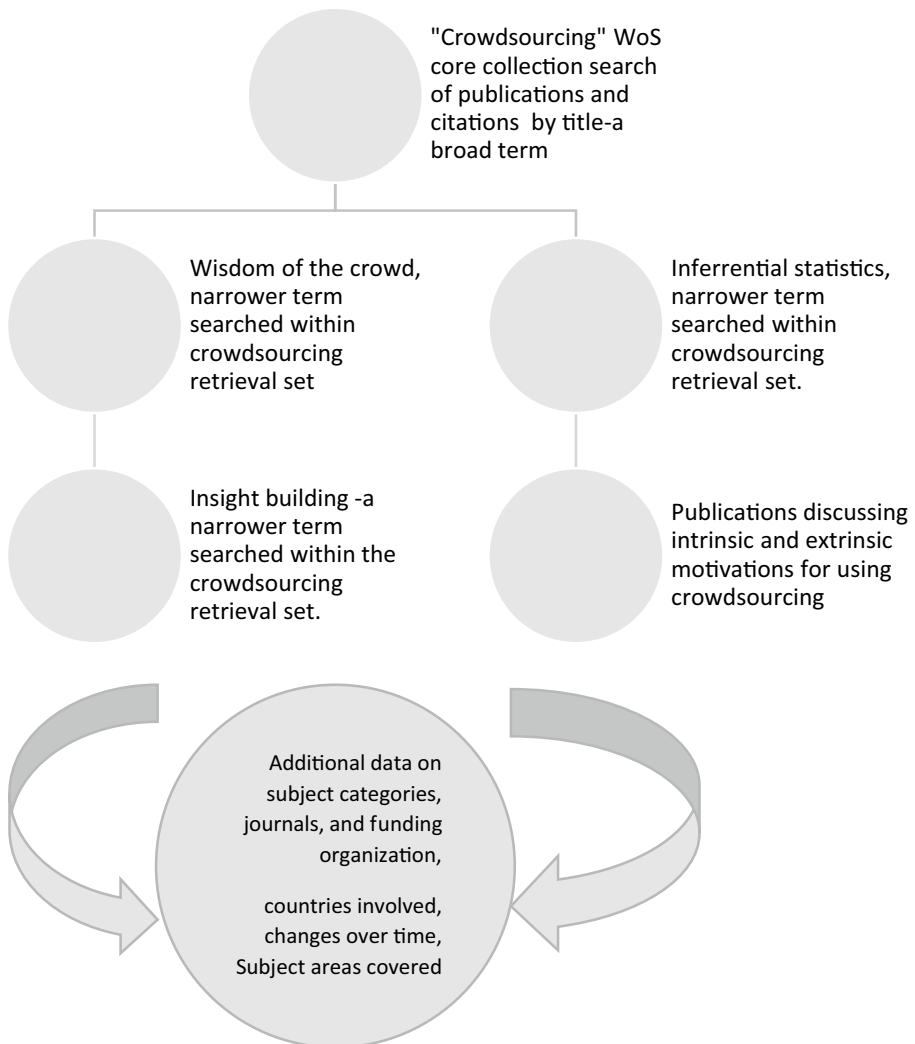

Fig. 1 Schematic representation of data retrieval on crowdsourcing

from a sample of the population to assess claims concerning the entire population (Moore, 1997). However, while crowdsourcing is not a new idea, it seems it is now being used in novel ways across many research fields and disciplines and for a variety of purposes (Bassi et al., 2020). According to Bassi et al., several conditions need to be present for crowdsourcing to take place. These include a defined audience, a defined goal, clearly defined compensation for the participants and for the process to involve an open call operated and managed via the internet (Ibid., 292). In light of the above, it is clear that while crowdsourcing has been taking place for several decades now, in the form of inferential statistics used for insight building, the main change since the start of the current millennium has been the channel through which it is performed, i.e. the internet, which enables crowdsourcing at a higher level of operation and management.

The research question

Does the literature on crowdsourcing reflect the evolution of this concept in academic research and the motivations behind its use as an insight-building tool?

Methodology

The data for the analyses was drawn from the Clarivate Analytics Web of Science database. The growth rate of the literature on this subject as well as the evolution of crowdsourcing as a concept were examined. The data was searched by the title "crowdsourcing" on July 20, 2020 and yielded 3677 papers. Publications added after that date were not included in the analyses. The main subject categories covering crowdsourcing were also examined and the changes occurring over time in this respect are discussed. In addition, the possible use of terms such as 'insight building,' 'inferential statistics,' and 'wisdom of the crowd' in the literature as synonyms for crowdsourcing was also investigated.

Publications discussing intrinsic and extrinsic motivations for using crowdsourcing were examined to establish whether usage of the term reflected a perception of it as a recreational activity for personal development or as a research method. The terms used in this analysis were drawn from two articles discussing the motivations for using crowdsourcing (Pinto et al., 2018, Feng et al., 2018).

Finally, the journals and funding bodies contributing to the literature on crowdsourcing were also examine.

Findings

Figure 2 shows an upward trend in publication numbers and a downward trend in citation numbers. These results drove the author to investigate the reason behind the decrease in citation numbers, which seems to be in opposition not only to the general trend of citations in other fields, but also to the increased number of publications on crowdsourcing over the same period of time. The assumption here is that the reliance on proceeding papers on crowdsourcing publications is the reason behind this decline.

Figure 3 shows that the general decline in citations to papers on crowdsourcing is due to the specific decline in citations to conference papers on this subject. Proceeding papers
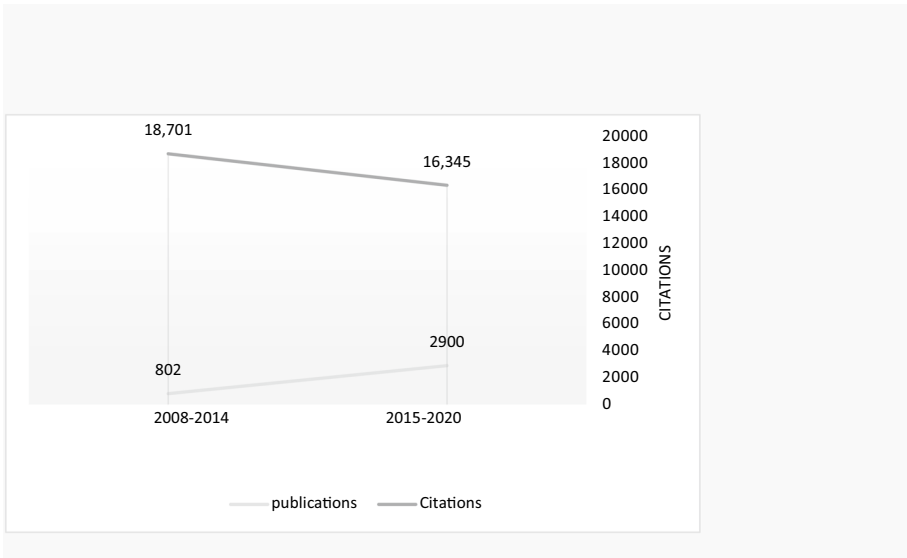
Fig. 2 The number of appearances of crowdsourcing as a concept in publications and citations, 2008–2020
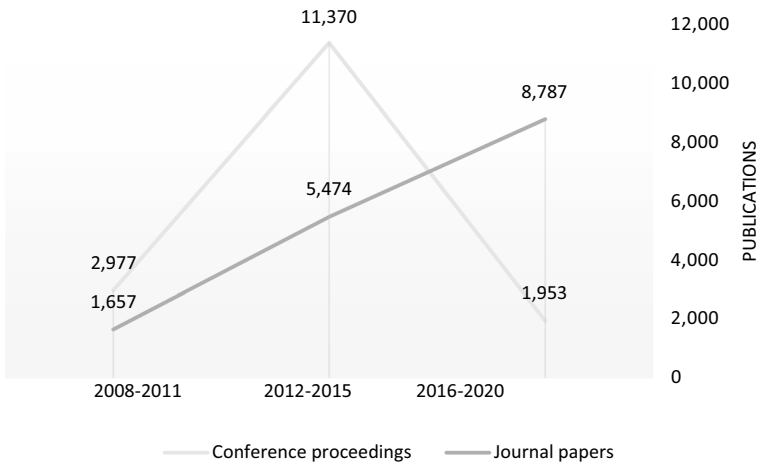*source*: Clarivate Analytics

Fig. 3 Citations to conference proceedings and journal articles on crowdsourcing over time

often receive less citations than journal papers (Gonzales-Albo & Bordons, 2011; Ingw-ersen et al., 2014).

Crowdsourcing, insight building, and inferential statics all elaborate on the same issue, namely, drawing perceptive conclusions by accessing the wisdom of the crowd. The data obtained from the Clarivate Analytics database shows that the term 'insight' was used 165 times in literature on crowdsourcing, the term 'wisdom of the crowd' was used 101 times and the term 'referential statistics' was used only in two papers.

Although inferential statistics aims at inferring results of a statistical test on a large population and in that sense is similar in nature to crowdsourcing, it was not perceived as such in the literature, and will be set aside at this point.

Though the term 'crowdsourcing' was coined in 2006 (Drabham, 2013), the actual use of the term as a concept first appeared in the literature in 2009. Previously this concept was referred to in publication titles as 'crowding,' 'crowded' and 'crowd.' Since 2009 the publication titles on this subject use the terms 'crowdsourcing' and 'crowd-sourcing' interchangeably.

The next figure describes the relationship between the three terms: 'crowdsourcing,' 'insight building' and 'wisdom of the crowd.'

Figure 5 describes the crowdsourcing as a method at arriving at an insight on an issue. The target in this process is the wisdom of the crowd, and the purpose is building an insight on a given course of action.

Figure 5 is a more up to date definition on the crowdsourcing process, that shortens the lengthy definition of Arolas and Gonzales (2012, 198) who defined crowdsourcing by the following elements: "the crowd; the task at hand; the recompense obtained; the initiator of the crowdsourcing activity; what is obtained by them following the crowdsourcing process; the type of process; the call to participate; and the medium." The authors admit, however, that there is no consensus as to the definition of crowdsourcing.

Crowdsourcing is a dynamic field, judging by the number of highly cited papers (29), and relies heavily on open source journals, with 838 out of 2461 contributing journals being open source ones (34.05%).

Overall, the USA is the highest contributor of papers on this subject, followed by China, who has the highest number of contributing funding institutions in this field.

Subject areas covered

Crowdsourcing as an insight building method covers many research areas and serves various purposes. It is especially useful where there is a need for large group participation in order to arrive at a collaborative insight, e.g. for knowledge gathering, geolocation-based data collection, sound pattern recognition (Bassi et al., 2020) and urban planning (Liao et al., 2019). Many studies present crowdsourcing as a research method or technique for data collation (Ibid., 298). The leading subject areas focusing on crowdsourcing, according to the Clarivate Analytics database, are the various areas of computer science, electric and electronic engineering and telecommunication. These emerged as the leading subjects from 2008 to 2020, although Certoma et al. (2015) and Randhawa et al. (2019) note that interest in crowdsourcing has expanded from computing researches to social and environmental applications. These principal coverage areas were followed by business and management, as well as other subject areas in the fields of social sciences, medicine and the humanities, urban planning, visualization design and more. The main coverage derives from computer science as an umbrella discipline, and this did not change from 2008 to 2020. The type of material also did not change since 2008, with the number of proceeding papers on crowdsourcing consistently exceeding that of journal articles. This type of publication is typical of the computer science subfields, and apparently affected the number of citations to papers on crowdsourcing. The most covered category of crowdsourcing is in computer information systems, which includes 538 published proceeding papers indexed in the Clarivate

Analytics database with 3212 citations, compared to 464 articles in this field that were cited 6162 times.

Aside from the increased number of publications on this subject, the other change taking place since 2008 is the growing focus of Chinese authors about crowdsourcing, compared with contributions from western countries. This is evident in the number of contributing authors, funding bodies and organizations and institutions from China. While publications from China have not taken the lead yet, it is evident that this is the trend regarding this subject.

Countries contributing papers on crowdsourcing: growth and shift

Table 1 shows that while China is closing the gap in terms of number of publications and funding bodies, it is still lagging behind the USA in number of authors involved, and especially in the number of organizations involved in publishing on crowdsourcing. Nevertheless, the increase in the number of Chinese funding bodies is indicative of the effort they are investing in this direction.

Intrinsic and extrinsic motivations behind the use of crowdsourcing

The literature on crowdsourcing also goes into a description of the motivations behind the use of crowdsourcing as a method. Ferraz et al. (2017) identify several motivations for the use of crowdsourcing. The current paper examines the space allocated to literature on these motivations within the literature on crowdsourcing.

Table 2 shows much higher incidence of extrinsic motivations, compared to intrinsic ones, leading to the conclusion that crowdsourcing is perceived to be more of a research method rather than a recreational activity for personal development. This is incongruent with Pinto and Denner's (2018) findings, according to which intrinsic motivations play a more significant driver compared to extrinsic ones for using crowdsourcing.

However, it is important to note that Pinto and Denner's study was based on an online questionnaire to which a sample of users responded, while the current study examines the

Table 1 The shift in the number of published papers on crowdsourcing, authors, institutions and funding bodies between the USA and China *Source* Clarivate Analytics core collection

Countries	2008–2011	2012–2015	2016–2020
The number of publications on crowdsourcing			
USA	62/121	458/1185	799/2387
China	4/121	183/1185	680/2387
The number of authors writing on crowdsourcing			
USA	215/318	1540/3331	2989/6895
China	9/318	601/3331	1861/6895
The number of organizations involved in publishing on crowdsourcing			
USA	132/200	652/1198	1063/2123
China	21/200	241/1198	666/2123
The number of bodies funding papers on crowdsourcing			
USA	12/29	234/482	**637/1533**
China	1/29	102/482	**670/1533**

Table 2 The intrinsic and extrinsic motivations shown in the literature about the uses of crowdsourcing *source* Clarivate Analytics 2020

Extrinsic motivations	Number of publications	Intrinsic motivations	Number of publications
Knowledge	508	Reputation	96
Learning*	461	Satisfaction	55
Problem solving	373	Contribute to a project	44
Interest	301	Self-development	35
Incentive	276	Develop skills	35
Share with others	268	Enjoyment	19
Data collection	260	Self-efficacy	16
Contribution	245	Autonomy	16
Gaming	225	Altruism	6
Reward	202	Fun	10
Collaboration	187	Self-presentation	4
Decision making	184	Reciprocity	4
Insight	165	Socializing	3
Reliability	153	Playfulness	3
Engagement	147	Social bonding	2
Feedback	135	Prestige	2

*Several motivations such as 'learning' could appear on either side, as the situation dictates. In this study 'learning' was included among the extrinsic motivations, mainly because crowdsourcing is designed for insight building on a subject, and this means learning more about it

publication trend in this respect. In the published literature on crowdsourcing extrinsic values of crowdsourcing usage were much more numerous than the intrinsic ones. These results could reflect a difference between what people consider important when asked about a subject, and what they focus on when they write papers about it.

Journals covering crowdsourcing and funding organizations

Most of the journals that cover crowdsourcing are from the computer science disciplines and subfields. *Lecture Notes of Computer Science* (LNCS) is in the lead with 156 papers, followed by various IEEE journals, who have more papers on this subject than journals from other fields. The one exception here is *Plos One,* which published 29 papers on crowdsourcing. These findings are compatible with those regarding subject coverage of this field.

China is the main country funding papers on crowdsourcing, having funded 670 papers on this subject, as opposed to 637 papers funded by American funding organizations (Table 1).

Discussion and conclusions

A search by title in the Clarivate Analytics database reveals how insight building by using the wisdom of the crowd evolved from the use of terms such as 'crowd,' 'crowding' and 'crowded' to 'crowdsourcing.' The data shows an increase in the number of
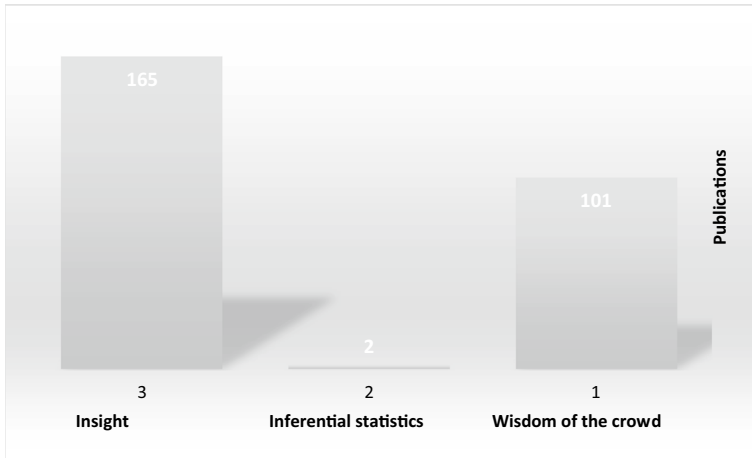
Fig. 4 Usages of the terms 'insight,' 'inferential statistics' and 'wisdom of the crowd' within publications on crowdsourcing
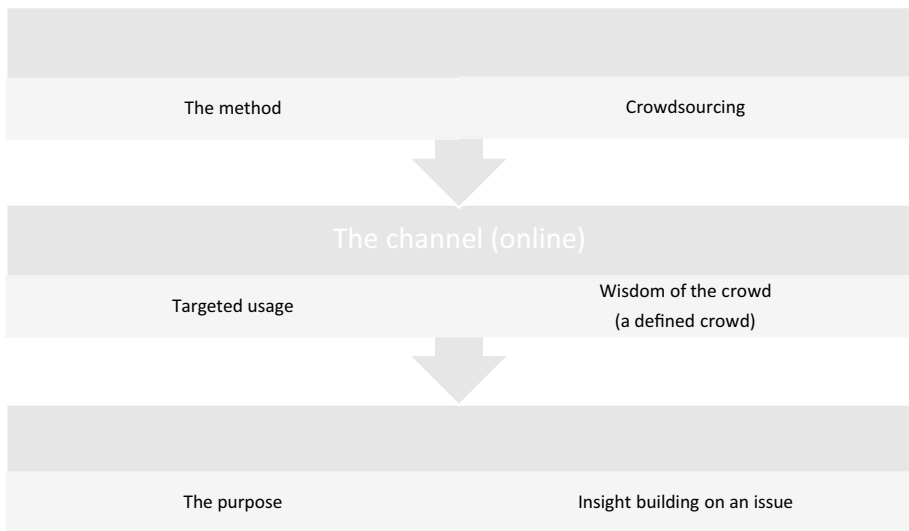

Fig. 5 The process of insight building using crowdsourcing that targets the wisdom of the crowd

publications on the subject alongside a decrease in the number of citations (Fig. 2). This is explained by the nature of the fields covering crowdsourcing, namely the various subfields of computer science that rely heavily on proceeding papers that often receive fewer citations compared to papers in scientific journals. Growth can also be seen in the number of authors and countries contributing papers on crowdsourcing (Table 1). The analysis also shows the changes occurring in crowdsourcing coverage over time, e.g. the growing number of Chinese publications on this subject, compared to publications from the USA. Other factors have remained unchanged during the last 12 years, among

them the highest coverage categories in this area, i.e. the subfields of computer science. Figures 4 and 5 show the relationship between the concepts of crowdsourcing, insight building and wisdom of the crowd, which together constitute the elements that form crowdsourcing.

As is the case in many other research areas, the literature on crowdsourcing is growing, and so is the number of countries involved in contributing papers on this subject, as shown in Fig. 6. This trend has been most notable in the last five years.

The motivations for using crowdsourcing are addressed in the literature, and are divided into intrinsic and extrinsic motivations, although sometimes a motivation, by definition could appear on both sides of the divide. This is true regarding the concept of learning, for example. In the current study learning was regarded as an extrinsic motivation, which is the case when one aims to learn something new about a subject by using crowdsourcing as a research method. However, if learning fulfils some inner desire to delve deeper into a situation through crowdsourcing, with no other practical purpose in mind, it would be considered an intrinsic motivation for using crowdsourcing.

An analysis of the motivations shows that most papers on crowdsourcing focused on extrinsic motivations (Table 2). In other words, crowdsourcing is perceived to be a research method, rather than a recreational activity for personal development. These findings directly oppose those found by Pinto and Denner (2018), according to which intrinsic motivations were more significant drivers for using crowdsourcing. However, it appears this difference originates from the research method used by these authors, which was a questionnaire that surveyed a sample of people, as opposed to the current study that examined the literature on crowdsourcing published by researchers.

Crowdsourcing coverage is expected to increase in the future, incorporating more fields and subfields and an increasing number of papers published on the subject. This could balance the effect of publications that rely mainly on proceeding papers as a principal publication source, and consequently lead to a growth trend in crowdsourcing related citations.
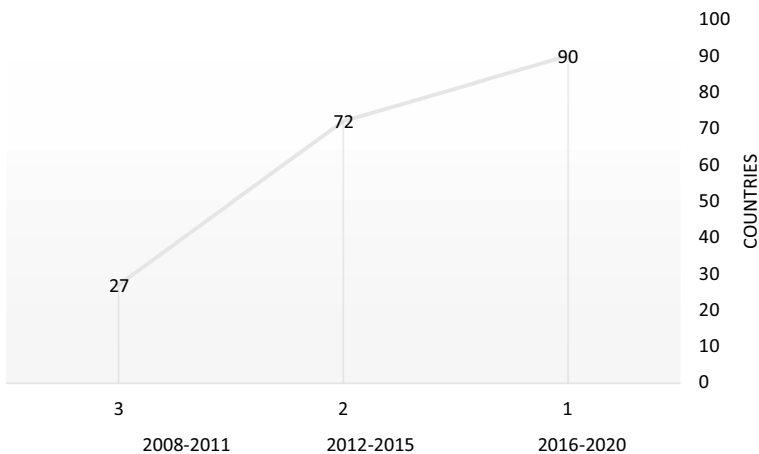

Fig. 6 The number of countries contributing papers on crowdsourcing over time. The graph shows a leap between 2008 and 2015 and a continual increase since then, albeit at a slower pace

References

Amsterdamer, Y., & Milo, T. (2014). Foundations of crowd data sourcing. *SIGMOD Record, 43*(4), 5–14.

Bassi, H., Christopher, J. L., Misener, L., & Johnson, A. M. (2020). Exploring the characteristics of crowdsourcing: An online observational study. *Journal of Information Science, 46*(3), 291–312. https://doi.org/10.1177/0165551519828626

Certoma, C., Corsini, F., & Rizzi, F. (2015). Crowdsourcing urban sustainability: Data people and technologies in participatory governance. *Futures, 74,* 93–106.

Dimitrova, S., & Scarso, E. (2017). The impact of crowdsourcing on the evolution of knowledge management: Insights from a case study. *Knowledge and Process Management, 24*(4), 287–295.

Drabham, D. C. (2009). Crowdsourcing the public participation process for planning projects. *Planning Theory, 8*(3), 242–262.

Drabham, D. C. (2013). *Crowdsourcing.* The MIT Press.

Estelle´s-Arolas, E., & Gonza´lez-EFL, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science, 38*(2), 189–200.

Feng, Y. Y., Hua, J., Yu, Y., Yang, C., & Cui, T. (2018). Gamification artifacts and crowdsourcing participation: Examining the mediating role of intrinsic motivations. *Computers in Human Behavior, 81,* 124–136.

Ferraz, I. R., Gouvea, M. A., & Barreto, I. F. (2017). Determinants factors of participation in crowdsourcing initiatives. *Journal of Management and Technology, 17*(1), 107–129.

Ghezzi, A., Gabelloni, D., Martini, A., & Natalicchio, A. (2018). Crowdsourcing: A review and suggestions for future research. *International Journal of Management Reviews, 20*(2), 343–363.

Gonzalez-Albo, B., & Bordons, M. (2011). Articles vs proceeding papers: Do they differ in research relevance and impact? A case study in the library and Information science field. *Journal of Informetrics, 5*(3), 369–381.

Heckman, L. (2004). Wisdom of the Crowd (a book review). *Library Journal, 129*(6), 105–105.

Ingwersen, P., Larsen, B., Garcia-Zorita, J. C., Serrano-López, A. E., & Sanz-Casado, E. (2014). Influence of proceeding papers on citation impact in seven sub-fields of sustainable energy research 2005–2011. *Scientometrics, 101*(2), 1273–1292.

Liao, P., Wan, Y., Tang, P., Wu, C., Hu, Y., & Zhang, S. (2019). Applying crowdsourcing techniques in urban planning: A bibliometric analysis of research and practice prospects. *Cities, 94,* 33–43.

Moore, D. S. (1997). *Active practice of statistics* (p. 233). W.H. Freeman company.

Oishi, K., Cebrian, M., Abeliuk, A., Abeliuk, A., & Masuda, N. (2014). Iterated crowdsourcing dilemma game. *Scientific Reports, 4,* 4100.

Pinto, S. F. L., & Denner, D. S. C. (2018). Motivations of crowdsourcing contributors. *Innovation and Management Review, 15*(1), 58–72.

Randhawa, K., Wilden, R., & West, J. (2019). Crowdsourcing without profit: The role of the seeker in open social innovation. *R&D Management, 49*(30), 298–317.

Surowieck, J. (2004). *The wisdom of the crowd.* Anchor.