Volunteering, Crowdsourcing and Citizen Science: an Overview from the Computing Perspective

Salvatore Distefano  
Kazan Federal University, Russia  
University of Messina, Italy  
sdistefano@kpfu.ru, sdistefano@unime.it

Samuele Rodi  
Politecnico di Milano, Italy  
samuele.rodi@gmail.com

Abstract—The digital revolution alongside with the evolution of social networking and portable platform has opened a lot of possibilities. In particular, the terms crowdsourcing and citizen science have acquired special meanings, referring to a new approach of performing tedious and repetitive work which was formerly being conducted by traditional dedicated employees. In this survey, these new trends are discussed by analyzing several experimental researches and practical examples. The paper illustrates the key aspects and the multifaceted representations of the crowdsourcing phenomenon starting from definitions and underlying working principles and motivations. Some practical examples are then discussed focusing on all the most relevant forms of adoption of these approaches, such as volunteering contribution in online communities, indirect crowdsourcing and gamification.

I. INTRODUCTION

In the rise of machine computing power and the extension of computer systems interconnected over a worldwide scale have contributed to the developing of new approaches to problem solving. The possibility of reaching out every possible person on the globe has allowed the spreading of knowledge and information to exponentially grow, contributing to the development of large scale projects, which may involve a large crowd spread all over the world. This is often represented either by people from extremely different backgrounds or technicians with extraordinary specific knowledge in a given topic.

This phenomenon, enabled by the so-called Web 2.0, has drastically altered the concepts of information contribution, dissemination, and exchange [2]. In this digital revolution, social media represent a cornerstone in the information systems of the current generations. Social media allow to bypass the need to utilize centralized, authoritative agencies, and information media (e.g. news organizations), only by empowering the information value of user-generated content.

In this context, the practice of crowdsourcing has spread enormously over the last few years i.e., the delegation of tasks traditionally performed by professionals to a large community. This practice has proven to have a huge potential and applications have already been analyzed in numerous fields ranging from geographic information systems to product development, urban planning and global knowledge expansion. Wikipedia, for example, is, probably one of the most notable results of the crowdsourcing process, where people are freely allowed to take part in and to extend the world global knowledge [10].

Involving citizens in large project is, however, not a new idea. Citizen participation has a long history in the United States, and the American Institute of Certified Planners states in the year right after World War II: “... give people the opportunity to have a meaningful impact on the development of plans and programs that may affect them.” [10]. In fact, Web 2.0 does not represent the only channel employed by crowdsourcing activities. Volunteers who take part in large science project are called “citizen scientists” and participation is rarely an issue, especially in projects where people experience time spent immerse in the nature or are educationally involved in the activities.

Citizen scientists are volunteers who participate as field assistants in scientific studies, helping, for example, in the monitoring of wild animals and plants or other environmental markers. They are not paid for their assistance, nor are they necessarily even scientists. Most are amateurs with a passion for the environmental and they are concerned about its trends and problems and want to do something about them. Typically, volunteers do not analyze data or write scientific papers, but they are essential to gather the information on which studies are based [8].

This paper aims at investigating two correlated topics in present days, such as crowdsourcing and volunteering contribution, by analyzing related literature. It gathers work evaluating relevant crowdsourcing applications and notable projects involving citizen scientists. The goal of this work is to present a short survey on volunteer computing, its practices, the reasons of success, the underlying problems and some of the main possible applications. All this food for thought is meant to prove the huge potential that this rather new approach represents for both commercial and scientific communities in carrying out some large scale work needing human supervision or to extract useful information out of social data.

II. BASIC CONCEPTS

The term “Crowdsourcing” is very often used to generally describe several different terms that in recent years have become buzzwords. The common idea among these concepts...
is the principle of delegating to someone else a given task. However, there exist many different modes of deployment of such an idea.

A. Definitions

Citizen involvement in crowdsourced project is particularly strong when the purpose of a project is altruistic and generous. This happens especially in research related projects which aim at helping out nature and humanity or they have educational purposes. In such cases, universities represent the coordinating institutions, as their job is usually believed to be prompted by worthy reasons. When such activities are extended to the population and involve large crowds of people, they are categorized under the term of “Citizen Science” projects, and the volunteers taking part in these activities are referred to as citizen scientists. These persons are very rarely experts in the proposed topics, most of them are amateurs with a passion or they are simply individuals particularly concerned about environmental, educational trends or other kind of global problems and they want to do something about them [8].

In opposition, when an activity is carried out by a company in a profitable manner, different terms are adopted. For a company, bringing in some ideas from the outside has become a key aspect in order to be competitive, especially in modern days. “Open Innovation” is the paradigm assuming that a company should use also external ideas together with internal ones in order to advance technology or enhancing the perspectives accepted within the organization [14]. This practice represents a double-edged sword, because in order to privilege some benefits, firms and organization must also reveal their cards and allow the spreading of knowledge, information or ideas outside of the company [6].

By extending the Open Innovation paradigm, it is easy to hit on an advanced philosophy of product development. Usually referred to as “Open Source” development, this concept involves allowing access to the essential elements of a product to anyone for the purpose of collaborative improvement to an existing product, with the continued transparency and free distribution of the product through the various stages of open development. This practice has been widely applied to software development in the last decades through the release of the source code of a software program. Even major companies, such as IBM, Oracle, and HP, have believed in such philosophy and have largely invested in the communities that develop Open Source Software [12]. Indeed, despite the simplicity that the concept wants to prove, open innovation and open source activities are not easy task to put into practice. To get the most out of them, problems need to be stated clearly, giving nothing as foregone, and without biasing the solution with pre-considered concepts. Additionally, there exist different viable channels for their implementation, and choosing the right one for the given application while addressing the problem in a correct and clear way requires time and a lot of effort.

Within a company, however, it also happens that one or more organizational activities might be delegated to external agents, reflecting a company contracting other companies to provide services, which might otherwise be performed by in-house employees. This practice is referred to as “Outsourcing” and represent an attempt of bringing in a project some quality derived from external ideas or through tasks delegated to quality services, which might even be at a lower cost than what the company would observe by self-performing.

All the above mentioned concepts are usually intended as traditional direct crowdsourcing methods, where an activity is clearly stated and affiliated to a crowd. Nevertheless, especially in the last decade, there have been examples of indirect crowdsourcing, which consists in the gathering of information from users who unconsciously contributes to crowdsourced project, through game-like applications. These methods have particularly flourished with the spread of social media and wearable devices, when users started compulsively updating content and personal information on the network. The so-called era of the Big Data flooding the Web 2.0 represents an opportunity for data miners which are able to retrieve precious knowledge, which could not otherwise be observed in the absence of a dedicated infrastructure. This obviously requires a well architected system absolutely not simple to implement, since the Web is an ocean of noisy data. Yet, once a nice model is outlined and an elegant method of extracting information is established, this probably represents the clever and powerful crowdsourcing method nowadays.

B. Motivations

One of the most debated topics in social science regards the reasons why every year, millions of people devote substantial amounts of their time and energy into helping others. Volunteerism is one of the most important human manifestations, and a lot of people perform each day such kind of service under several contribution forms, such as assistance to needy persons, education to people, contribution for the community or even more simply when giving feedback to inform users and producers about their feelings on a product.

Clary et al. investigate on the fundamental factors that lead humans to contribute to a volunteering activity of any given kind [1]. He discovers six motivational functions collected in what he calls the Volunteer Functions Inventory, which reflects the personal attitudes and inclinations of a person as what regard volunteering contribution. He observes, through a factorial analysis on precise surveys that everyone is reactive in a different way to volunteerism when induced by different motivational aspects. These are opportunities for the individual offering payless duty and are represented by values concern, education possibilities, social opportunities, career perspectives, self-protective aspects and ego’s enhancing attempts. One more aspect, which induce to volunteerism, is the expected satisfaction that each individual receives from the proposed task and the actual outcome contributes to drive a person into retake the same job or a similar one. Clary concludes that one of the best way to convince a large crowd to participate in volunteerism is the exposure to persuasive and targeted messages matching with the individual motivational functions, together with the correct usage of feedbacks from past experiences.
Motivation in volunteering contribution is a key aspect which should be always renovated and triggered, especially in large and meticulous projects, such as Wikipedia. The notable free online encyclopedia has stepped in a development phase which has led the website to acquire more popularity than its most serious traditional rival, the Encyclopedia Britannica, by encompassing more subjects and content in less than five years from its conception. All this success is however due to an effective underlying motivational model and a straightforward interface which encourages users into creating, editing, monitoring and updating content. Stacey Kuznetsov studies this model using a Value Sensitive Design (VSD) methodology, in order to identify the values involved in the interactions between stakeholders and technology [10]. She concludes deducing five values underlying Wikipedians’ inspiration and these are: altruism, reciprocity, community, reputation, autonomy. In particular, Wikipedia gives the possibility to users to feel fulfilled by coordinating tasks, creating discussions, generate identities and profiles, receive and nominate for awards, and, most importantly, it gives the freedom of full editing and control over other users’ contribution without a centralized authority.

Investigating the motivations of volunteers is a recurrent subject, and there are even examples of platform which are meant specifically to increase the motivation of contributors. In Gilbert and Karahalios work [4], it is possible to spot the attempt of increasing the motivation of programmers working at open-source software projects by using social visualization techniques. Their platform, called CodeSaw, allows displaying the history and modifications of a project, while it records users’ activity in terms of contribution to the code and communication about the code and shows it on a visual timeline. Communication among members is considered crucial and the visual impact on users of the project development acts as an effective motivational drive to contribution in important social production projects. In their belief, social visualization and project coordination are important aspects during development, which contributes to foster motivations and represent incentives to increase social productivity.

C. Citizen Science

The practice of using citizen scientists for scientific purposes dates back of at least one century, long before the digital revolution. There exist several reasons why citizen scientists are largely employed in scientific researches. Nowadays, entities such as the National Science Foundation are funding lots of projects involving citizen scientists, in the form of grants for education. Beside the possibility of using free workforce to collect data about the natural world at a large scale, using volunteers is a way to get more citizens aware of the environment, while educating them. Cohn reports examples of such activities which have the goal of monitoring environmental trends [8]. Volunteers, in this context, are asked to perform various tasks ranging from placing cameras and equipment, counting birds and species, gather data and record observations. Collecting data is a long operation for scientists, which does not require professional skilled researchers to accomplish. Volunteers, on the other side, are usually prone to such activities as they are moved by passion and they consider it an altruistic and educational way of spending some time out enjoying the nature. In this situation, researchers are left with the responsibility of training the citizens, which sometimes might be of a headache, representing a lot of time to invest.

As a matter of fact, data collected from citizen scientists are very often of variable quality and less usable than data collected from professionals, even though they are drastically compensated by significantly greater data availability at a larger geographic scale and over a longer period. To increment their usability, training properly the volunteers and limiting their tasks is the first key measure, to adopt which could lead to surprisingly satisfying results. Other mechanisms to enhance data quality include social trust metrics. Determining the reliability of users allow to weight the quality of their provided data, and this should be performed using user reputation, which is derived from other agents that had previously interacted with the selected user. The reputation score is then calculated from a range of attributes like the direct ranking between members, inferred ranking across the social network, quality of past data, frequency of contribution and some other factors. Additional observations are brought by Hunter et al. while analyzing a volunteering contribution experiment applied to coral bleaching events and trends. Such data, filled in an online survey, includes position, reef name, coral species and color, water temperature and other parameters. He observes that over 70% of records present syntactic errors and most of this set is associated to spelling errors, GPS inaccuracy and lack of feedback [9]. These are erroneous inputs which could be easily avoided by implementing a different interface for the system, using formatted forms and adopting trust metrics.

The number of projects involving citizen scientists is, thus, increasing as they are powerful resource for research studies. Organizers benefit from citizen scientists and they also contribute to give people the chance of getting involved and develop a global education. Moreover, there is proof that a range of technologies capable of improving data of project involving citizen scientists exist, enhancing the potential reuse of such kind of contribution.

D. Crowdsourcing

Despite there exist several historical examples of obtaining needed services from a large group of people, even long before personal computers were invented, the word crowdsourcing, at present, is mainly associated with online communities’ activities. Indeed, it is of no surprise that the term itself was first coined only in 2006 by Jeff Howe in an article of the notable magazine Wired. The so-called Web 2.0, strongly represented by social networks and websites established on user-generated content, provides a suitable environment for the prosperity of crowdsourcing. On the Internet, individuals do not feel the pressure of being physically judged or scrutinized and are more inclined to share content, ideas and knowledge. In particular, contributors in such environment tend to be more focused on a project’s specific needs rather than spending time on communication with other colleagues.
Nowadays, online social interaction has become a crucial part of the worldwide scenario, and its geographical extension has soon represented a working advantage over traditional employee hiring. Cultures of participation are also a powerful tool for solving problems which cannot be handled by individuals for their size and systematic nature, or that require high level of involvement by special cliques over the world. However, the potential of such communities is strongly related to their success and worldwide diffusion. It is, thus, of extreme importance for crowdsourcing purposes assuring that a community keeps growing, always attracting new users over geographically dislocated area without ever being abandoned.

To this purpose, Alicia IriberrI and Gondy Leroy have analyzed the key aspects in an online community in order to guarantee its progress [5]. In their conclusions, the most interesting design principles include: support for different role members, encouraging new discussions, rewarding members, defining a code of behavior, allowing social recognition, introducing a feedback system, guaranteeing data privacy, website stability and different levels of anonymity. They also analyzed the right timing for the implementation of such feature within a community lifecycle. In particular, each element, once implemented, should continue to exist and never abandoned to guarantee the community continuation and evolution. In the inception phase of a community it is important to have a clear purpose and focus together with a trademark and a tag line. User-centered design and interface usability are fundamental during the build-up phase. Promotion, encouraging membership, reminders to contribution and up-to-date high-quality content are key aspects necessary to guarantee the community growth and continuation. In a mature phase is then important to recognize user contribution and community loyalty while establishing regular online events. Obviously, keeping up with updated content and attention privacy concerns is crucial to avoid premature death of participation.

With a similar analysis, Gerhard Fischer stresses out the importance in a community of accounting for richer ecologies of participation [3]. There always exists a large pool of users engaged at a different level, thus it is important to accept low thresholds of participation, including special interaction and support for different level of engagement, rewards and incentives. Additionally, it is particularly important for a successful system to pick the right interface design, providing a simple, clean and straightforward way for users to create content. This meta-design is not direct responsible of content creation, but of the development of context and of content-creating tools.

From another perspective, online communities have not been the only protagonist of the crowdsourcing evolution phenomenon of the current century. The parallel technological advancement of recent years, leading to the development of powerful portable devices has pushed the potentiality of crowdsourcing applications at a much higher level. Smartphones and other always online mobile devices equipped with embedded sensors have boosted the data availability on the web, contributing to the so-called "human-in-the-loop" paradigm, representing the human condition of being always connected to the network through any portable device [13]. Geographic Information Systems, in particular, have benefited of this massive amount of data, extracting sensitive information, useful, for instance, to monitor traffic and weather conditions, or to track health, disease spreading and outbreaks. Ambitious examples are the Common Scents project, which placed some environmental sensors on bicycles to gather data on air quality and traffic conditions around the city of Copenhagen, or the 3D town project at York University, in Canada, aiming to integrate the 3D city models with temporal information from surveillance camera, to provide a sort of augmented reality synchronized in real-time. It is also evident from such experimental applications that a large effort is needed in this field to improve the way data are collected and to reduce data saturation and redundancy to optimize the information retrieval phase. The ultimate goal is to increase the situation awareness of sensors in order to creating a human-centered computing system.

An astonishing information retrieval case is the one presented by Crooks et al. which adopt an indirect form of crowdsourcing using social media feeds to collect geographic information [2]. The strong popularity of social networking has allowed user generated content, represented by posts or tweets in popular social networks, to be considered part of a sensor system, where the user is the sensor and the application is the receiver which analyze, control and distribute the information. In the analyzed study, tweets posted during the earthquake occurred in Louisa County in August 23rd 2011 were collected from Twitter. A small subset of those feeds was harvested and filtered using the hashtag search (#quake or #earthquake keywords) over an 8-hours period, and a large part of this set contained either precise geolocation information or it was easily derivable from the IP address. From the results, it is surprising to observe that the first tweets were posted only 54 seconds after the occurrence of the event and that a dense cluster of tweets fall close to the epicenter, progressively enlarging with the propagation of the earthquake in the 2-5 subsequent minutes from the event. Despite the small subset of data, the outcome was closely comparable to the dedicated US platform DYFI (Did You Feel It?), provides a Volunteered Geographic Information service. The conclusions prove the validity of social media feeds as an alternative method to retrieve geographic information, which is very often cheaper and better distributed than a standard dedicated system. Additionally, the eventuality also exists of adopting this kind of human geographic sensor network as an early warning system for large-scale disasters or disease outbreaks.

As a matter of fact, crowdsourcing in an indirect manner demonstrates to be one of the technique in this field with the highest potential, capable of retrieving a huge amount of data without any explicit volunteering request to the crowd. On the other side, this technique undoubtedly lacks of precision if large data availability is not a counteragent, and it requires, for sure, an efficient and not always straightforward way of performing information retrieval.
E. Gamification

One of the most interesting approaches to crowdsourcing is certainly represented by the “gamification” concept, which has become of increasing popularity after the numerous successes and advancements observed by the gaming industries in the latest decades. Many companies have, thus, recently adopted “gamified” models for selling their services, given its proven efficacy to catch the attention of a large crowd of people. However, gamified applications substantially differ from full proper game, such that this practice has been contested sometimes by game industries due to the oversimplification introduced. The term “gamification”, coined in 2008, started spreading out globally no earlier than in late 2010 and it refers to the adoption of game elements and design outside of a pure gaming context aiming at enhancing engagement and motivation in order to drive participation. Gamified approaches adopt game design elements to enhance the user experience of an already existing service. These elements summarized by Reeves and Read in the “Ten Ingredients of Great Games” are identified by: self-representation with avatars; 3D environments; narrative contexts; feedbacks; reputations, rankings and levels; marketplaces and economies; competition under explicit rules; team subdivision; parallel communication systems; time pressure [15].

There exist manifold reasons inducing to game-play and their motivations can actually be of very diverse nature. In a study proposed by Nick Yee the motivational aspects of game-play are analyzed and it is established that exist three major components inducing to alienation. These are represented by an achievement factor (advancement in the game, interest in understanding the system, competition and desire to challenge other players), a social factor (helping and chatting with others, building relationships, working in a team) and an immersion factor (desire of discovery, exploring hidden objects, creating a story line, customize character and appearance, escape from reality and real problems) [7]. These factors could be of different meaning for each individual as everyone perceive the same game in different manners and mean variations can be noticed among different genders, ages and usage attitudes. For example, female players tend to score higher in the relationship sub-component in contrast with male players that usually show higher results under the achievement factor.

All these researches have lead scientists to believe that the gamification is one important motivational factor, eventually capable of attracting a large number of participants to contribute semi-consciously to a scientific investigation. Indeed, projects involving citizen scientists are drastically expanding in a way related to the attractiveness that they exhibit. Designing a system that keeps vivid the user interest will undoubtedly result in a larger contribution to data collection. Unfortunately, gamified approaches in crowdsourcing applications come at a price. Firstly, it is rarely easy to design a game-like system which fits into an already existing science study case in a well-structured and user appealing manner. Secondly, game development of proper game is a very costly and time-consuming activity, and it is almost never a viable solution to crowdsourced experiments which tends to come with a limited budget and time. Moreover, data quality is likely to be affected in such context, since players are usually keener on progressing the story than in doing science. Therefore, it is rarely recommended to apply game taskification to crowdsourcing projects, i.e. taking the effort of inserting small chunks of scientific tasks within a full game experience. On the other side, it is generally encouraged the gamification of crowdsourced tasks, the so-called task gamification, i.e. inserting few and simple game-design elements into an already existing scientific task (e.g. using scores and/or rankings).

Prestopnik and Crowston provide evidence of this fact as they analyze the crowdsourced case of species classification in biological science by adopting several systems at a different gamification level [11]. During evaluation of the different artifacts, some observations came spontaneously. Fostering motivation in citizen science project rely very often on preexisting enthusiasm of participants. Thus, systems that motivate gamers may actually be of less interest to enthusiasts and vice versa. Data quality, evaluated from the ground truth produced by expert participants, appeared to be reducing as the gamification level increased. Additionally, the most gamified approach represented by the simple game with a full storyline, Forgotten Islands, turned out to be of extremely complex endeavor for developers with much higher costs that would make it worthwhile only if the result is a widely popular game, which might cover the costs with game sales. Indeed, Forgotten Island took several months of to be completed and developers had to conclude the game development at earlier stages due to lack of resources.

To sum up, the gamification of a scientific project is seldom a viable approach as it hardly produces high quality data, but it still contributes to generate useful knowledge, to attract a larger pool of users and to develop users’ interest in the subject. For these reasons, even if these attempts might represent a failure to the scientific community, they still are a research success to the online world.

III. CONCLUSION

Currently, the web is exponentially expanding with huge amount of content generated every second. Alongside it, online communities have flourished in such an environment and they have started representing the cradle of knowledge for future generations. In this context, crowdsourcing applications have become a powerful tool to collect data, analyze trends and retrieve useful information from the volunteered help of a huge crowd dislocated on the web. The objective of this paper is to prove the potential of crowdsourcing and citizen science methods, by investigating their most crucial aspects and concerns while providing some practical examples of their application.

Right now, there is plenty of exciting work taking place, and lots of areas for additional research and activity and, the present paper only scratches the surface of its possible reach. Considering the fact that these new approaches represent an almost new discipline, which started being employed extensively only during the current decade, they are very likely to encounter soon an incredible evolution representing
an explosion of new work positions and opportunities. There are great future perspectives in these fields for both research institutes and companies and that future is only beginning to unfold in the literature and in practice.

REFERENCES


doi:10.1089/cpb.2006.9.772


